



HAL
open science

Optimisation adaptative basée sur les métamodèles

Nathalie Bartoli

► **To cite this version:**

Nathalie Bartoli. Optimisation adaptative basée sur les métamodèles. Optimisation et contrôle [math.OC]. Université Toulouse III, 2019. tel-02499582

HAL Id: tel-02499582

<https://hal.science/tel-02499582>

Submitted on 5 Mar 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Habilitation à Diriger des Recherches



Délivrée par : *l'Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)*

Soutenu le 16 décembre 2019 par :
Nathalie BARTOLI

Optimisation adaptative basée sur les métamodèles

JURY

CHARLES AUDET	Polytechnique Montréal	Rapporteur
NICOLAS GAYTON	SIGMA, Clermont-Ferrand	Rapporteur
RODOLPHE LERICHE	CNRS, Saint-Etienne	Rapporteur
JEAN-BAPTISTE HIRIART-URRUTY	Université Toulouse III	Président du Jury
BÉATRICE LAURENT	INSA Toulouse	Examineur
JOAQUIM R.R.A. MARTINS	Université du Michigan	Examineur
JOSEPH MORLIER	ISAE-SUPAERO	Examineur
MARCEL MONGEAU	ENAC	Examineur

Section 26 :

Mathématiques appliquées et applications des mathématiques

Unité de Recherche :

Equipe MOIS, Modélisation et Ingénierie des Systèmes ISAE-SUPAERO et ONERA

Parrains de Recherche :

Marcel MONGEAU et Joseph MORLIER

Remerciements

Je tiens tout d'abord à remercier les personnes qui ont accepté de faire partie du jury. En particulier, les trois rapporteurs, Charles, Nicolas et Rodolphe qui ont pris de leur temps pour relire ce manuscrit et assister à la soutenance.

Je suis très reconnaissante à Jean-Baptiste d'avoir présidé ce jury et de m'avoir encouragée dans cette démarche depuis quelques années. Merci Béatrice d'avoir représenté lors de ce jury, l'INSA GMM qui a marqué le début de mes études. Joaquim, merci beaucoup de m'avoir motivée dans cette démarche depuis le début de notre collaboration en 2015. Marcel, merci d'avoir accepté d'être mon parrain lors de mes premières démarches en 2018 afin de mener à bien ce projet de HdR. Merci également de l'aide financière de l'ENAC pour la venue du jury.

Un grand merci Joseph, parrain non officiel depuis des années, grâce à qui j'ai pu organiser ma recherche à travers notre collaboration fructueuse depuis 2011.

Merci à Catherine et à l'école doctorale ED AA de m'avoir soutenue. Merci à l'ED MITT d'avoir accepté mon dossier.

Une grande reconnaissance à mes collègues pour leur soutien au quotidien lors des montages de collaboration, de mes démarches administratives toujours un peu compliquées.... Mon parcours est riche de projets, de collaborations, d'encadrements et d'échanges via des mobilités d'étudiants mais le chemin pour y arriver est souvent très sinueux, leur soutien est donc une aide indéniable! Merci beaucoup.

Merci aux membres du club footing et/ou du club "Muscu" pendant les moments euphoriques et plus difficiles! Merci aux membres de la salle Café pour les échanges informels mais très instructifs! Merci à Thierry, Peter, Sylvain, Thomas, Rémi, Antoine, Judicäel, Sébastien, Luis, et Jérôme.

Merci aux différents étudiants, doctorants, post-doctorants qui ont accepté mon encadrement et avec qui j'ai pu échanger au quotidien. Le résultat de leurs travaux est un des piliers de ce manuscrit. Merci à vous!

L'optimisation multidisciplinaire n'est possible qu'avec de multiples experts. J'ai eu la chance de collaborer avec nombre d'entre eux et de partager des idées et des réalisations. Merci à eux ! La liste de ces experts est longue et s'allonge au cours des années : Thierry, Peter, Sylvain, Sébastien, Jérôme, Joseph, Youssef, Michel, Emmanuel, Christian, Joaquim, ...

Un manuscrit, une soutenance, un jury, un système de visio et d'audiconférence, une audience et des visiteurs ... tout cela est impossible sans les services de soutien de l'ONERA, de l'ISAE-SUPAERO et de l'Université de Toulouse. Merci à Claire, Nadine, Colette, Jean-Paul, Frédéric, Marc, Stéphane et Michelle pour leur aide durant la préparation et la logistique.

Merci à mes copines d'avoir fait le déplacement pour entendre parler de métamodèles et non de "topmodèles". Merci à elles !

J'ai également bénéficié d'un soutien familial sans faille dans cette aventure. Merci à Nicolas, Alexis, Fanny et Thomas !

Résumé — Ce document est une synthèse de mon activité de recherche autour de l’optimisation et des métamodèles lors de ces 10 dernières années. Cette activité de recherche est très liée à mon investissement dans le domaine de l’optimisation multidisciplinaire appliquée à la conception avion avant-projet et en particulier aux nouvelles configurations. La thématique de recherche associée concerne le développement de méthodes numériques pour l’optimisation efficace de fonctions de type ”boîtes noires”. Les méthodes proposées, fondées essentiellement sur le développement de métamodèles, tentent de répondre aux problématiques liées au temps de calcul, au grand nombre de variables de conception et/ou de contraintes, mais également à la prise en compte des incertitudes de modèles dans le processus de conception. Différents types de métamodèles sont détaillés dans ce manuscrit, notamment les mélanges d’experts et les processus gaussiens avec une adaptation pour la grande dimension au travers des articles publiés sur le sujet. Ces modèles sont actuellement disponibles dans la librairie Python *open source* SMT. Un processus d’optimisation adaptative basée sur ces métamodèles est ensuite détaillé et validé sur de nombreux cas d’application académiques et industriels. La prise en compte des modèles à fidélité variable est également abordée à travers l’utilisation de métamodèles spécifiques et l’adaptation du processus d’optimisation, ce qui permet de trouver un compromis entre la qualité de la solution obtenue et le temps de restitution. Enfin, l’utilisation de métamodèles disciplinaires dans un processus de conception est présentée. Celle-ci introduit des incertitudes de modèles qui se propagent lors de l’analyse multidisciplinaire et affectent ensuite la fonction objectif et/ou les contraintes. L’adaptation du processus d’optimisation pour prendre en compte ces incertitudes a permis d’obtenir des premiers résultats très prometteurs. Les travaux présentés dans ce manuscrit sont le fruit de collaborations internes et externes, régionales et internationales lors de projets européens ou d’encadrements de stagiaires, doctorants et post-doctorants.

Mots clés : Métamodèles, Processus gaussien, Optimisation Multidisciplinaire, Optimisation bayésienne

Table des matières

1	Curriculum Vitae	3
1.1	Actuellement	4
1.2	Formation et diplômes	4
1.3	Parcours	5
1.4	Synthèse de l'activité de recherche	6
1.4.1	Publications et Communications	6
1.4.2	Six publications les plus significatives	6
1.4.3	Développement logiciel	7
1.4.4	Collaborations	7
1.4.5	Organisation de workshops	7
1.4.6	Encadrement	8
1.4.7	Enseignements	8
1.4.8	Actions de Formation	8
1.4.9	Rayonnement scientifique	8
1.4.10	Montage de projets	9
2	Activités de recherche	11
2.1	Résumé	12
2.2	MDO pour la conception	12
2.2.1	Introduire des outils haute fidélité	14
2.2.2	Introduire de nouvelles disciplines	17
2.2.3	Introduire des modèles de substitution	19

2.2.3.1	Un plan d'expériences optimal	20
2.2.3.2	Un plan d'expériences adaptatif	21
2.2.3.3	La technique des mélanges d'experts	22
2.2.3.4	Etude du modèle de krigeage et ses hyperparamètres	24
2.2.3.5	Des modèles dynamiques	27
2.2.4	Introduire des modèles de fidélité variée	28
2.2.5	Disposer d'optimiseurs efficaces pour les boîtes noires	30
2.2.5.1	Optimisation directe basée sur des métamodèles . .	31
2.2.5.2	Optimisation adaptative basée sur des métamodèles	32
2.2.5.3	Amélioration du conditionnement de la matrice du krigeage au cours des enrichissements	36
2.2.5.4	Optimisation multi-fidélité	36
2.2.6	Formulations MDO adaptées	38
2.2.7	Prise en compte des incertitudes	39
2.2.7.1	Gestion des incertitudes de modèles	39
2.2.7.2	Gestion des contraintes de fiabilité	41
2.2.8	Analyse de sensibilité	42
2.3	Mon projet de recherche à l'ONERA	43
3	Librairies informatiques	45
3.1	Développement de librairies en interne	46
3.2	Développement de librairies en opensource	47
4	Contribution de recherche	49
4.1	Développement de métamodèles	51
4.2	Optimisation basée sur les métamodèles	52
4.3	Capitalisation dans une librairie informatique open source	55
4.4	Sélection des publications	55

4.4.1	Surrogate modeling approximation using a mixture of experts based on EM joint estimation, Structural and Multidisciplinary Optimization 2011	56
4.4.2	Improving kriging surrogates of high-dimensional design models by Partial Least Squares dimension reduction, Structural and Multidisciplinary Optimization 2016	75
4.4.3	An Improved Approach for Estimating the Hyperparameters of the Kriging Model for High-Dimensional Problems through the Partial Least Squares Method, Mathematical Problems in Engineering 2016	94
4.4.4	Approximation of the critical buckling factor for composite panels, Structural and Multidisciplinary Optimization 2012 .	106
4.4.5	Efficient global optimization for high-dimensional constrained problems by using the Kriging models combined with the partial least squares method, Engineering Optimization 2018	131
4.4.6	Adaptive modeling strategy for constrained global optimization with application to aerodynamic wing design, Aerospace Science and Technology Journal 2019	158
4.4.7	A Python surrogate modeling framework with derivatives, Advances in Engineering Software 2019	177
5	Liste des encadrements	191
5.1	Activités d’encadrement d’étudiants	192
5.1.1	Encadrements de projets élèves	192
5.1.2	Encadrements de stage	192
5.2	Encadrements de thèse	195
5.3	Encadrements de post-doc	196
6	Encadrements thèses et post-docs	199
6.1	Thèse de R. Priem	200
6.2	Thèse de M. Menz	201
6.3	Thèse de A. Sgueglia	203

6.4	Thèse de J. Mas Colomer	204
6.5	Thèse de P. Schmollgruber	206
6.6	Thèse de M. Gruet	207
6.7	Thèse de M.-A. Bouhleb	209
6.8	Thèse de D. Bettebghor	211
6.9	Thèse de J. Clément	213
6.10	Post-Doc de S. Dubreuil	214
6.11	Post-Doc de F. Luraghi	216
7	Projets	219
7.1	Projets internes ONERA	220
7.2	Projets régionaux et nationaux	221
7.3	Projets européens	222
7.4	Projets en cours de montage	223
8	Collaborations	225
8.1	Collaborations régionales	226
8.2	Collaborations européennes	228
8.3	Collaborations internationales	230
9	Dissémination	235
9.1	Organisation de workshops	236
9.2	Participation à des formations	238
9.2.1	Formation Matlab	238
9.2.2	Formation métamodèles - Machine Learning	238
9.3	Publications	239
9.3.1	Revue internationale avec comité de lecture	239
9.3.1.1	Articles soumis	240
9.3.1.2	Articles en révision	240
9.3.1.3	Articles parus	241

9.3.2	Ouvrage	242
9.3.3	Chapitres de livre	243
9.3.4	Mémoires de DEA et de thèse	243
9.4	Conférences	244
9.4.1	Conférences internationales avec actes	244
9.4.2	Autres conférences et workshops sans acte	250
9.4.3	Conférences invitées et séminaires	253
9.5	Brevets	254
9.6	Administratif et Scientifique	255
9.6.1	Participation à des activités de relecture	255
9.6.2	Participation à des jurys	255
10	Enseignements	257
10.1	Introduction	258
10.2	Optimisation mono et multidisciplinaire	258
10.2.1	Optimisation multidisciplinaire, ISAE-SUPAERO	258
10.2.2	Optimisation stochastique, ISAE-SUPAERO	259
10.2.3	Fiabilité, ISAE-SUPAERO	259
10.2.4	Optimisation de structure multidisciplinaire, ISAE-SUPAERO	260
10.2.5	Optimisation filière Masters spécialisés, ISAE-SUPAERO	260
10.2.6	Cours d'optimisation non linéaire, ISAE-SUPAERO	260
10.2.7	Projet en optimisation, ISAE-SUPAERO	260
10.2.8	Travaux dirigés de probabilités et statistiques, INSA	260
10.3	Calcul scientifique	261
10.3.1	Travaux dirigés d'introduction à la simulation numérique, ISAE-SUPAERO	261
10.3.2	Travaux dirigés de calcul scientifique, ISAE-SUPAERO	261
10.3.3	Cours pour les EDP, ISAE-ENSICA	261
10.3.4	Travaux pratiques Matlab - DEUG MASS	261

10.3.5	Travaux pratiques Internet - DEUG MASS	262
10.3.6	Travaux dirigés et projets en Analyse Numérique - première année ENSICA	262
10.3.7	Travaux pratiques de Fortran 90 - INSA, option Génie Mathématique et Modélisation	262
11	Conclusion et perspectives de recherche	263
12	Appendix	269
12.1	Fonction G07	270
12.2	Fonction de Griewank	270
12.3	Fonction de Branin Modifiée	270
	Bibliographie	273

Motivations

Ce document est une synthèse de mon activité de recherche dans le domaine de l'optimisation et des métamodèles lors de ces 10 dernières années environ. De part ma formation en mathématiques appliquées, j'ai travaillé sur différents sujets autour des schémas numériques en tant qu'ingénieur de recherche ONERA dans l'unité Modélisation Mathématique et Simulation Numérique. Cette activité n'est pas décrite dans ce manuscrit. Mon changement de département au sein de l'ONERA en 2013 (Département de Conception et évaluation des Performances des Systèmes, unité Systèmes Aéronautique) m'a permis de concentrer mon activité autour de l'optimisation multidisciplinaire pour la conception avion. Par motivation personnelle, c'est ce travail qu'il m'a paru intéressant de regrouper dans ce dossier visant à l'obtention d'une Habilitation à Diriger des Recherches.

Cette activité de recherche est très liée à mon investissement autour de l'optimisation multidisciplinaire appliquée à la conception avion avant-projet et donc je me suis intéressée à l'évolution du processus de conception pour traiter les nouvelles configurations comme par exemple l'aile volante avec propulsion distribuée.

Depuis 2013, j'ai intégré une équipe multidisciplinaire et dont la complémentarité entre les chercheurs est force de proposition. C'est la raison pour laquelle, mon travail de recherche est très multidisciplinaire.

Ce rapport n'est pas destiné à être technique, les détails, les preuves, ainsi que les résultats numériques détaillés se trouvent dans les articles présentés dans le chapitre 6.

Onze chapitres composent ce dossier : les trois premiers chapitres détaillent respectivement mon CV, mes activités de recherche et les développements informatiques réalisés. Le quatrième chapitre présente plus spécifiquement ma contribution de recherche autour des métamodèles et de l'optimisation à partir des sept articles qui ont été sélectionnées. Les chapitres 5 et 6 détaillent les encadrements réalisés (stages, thèses et post-docs). Ma participation aux projets internes et externes est décrite dans le chapitre 7 et les collaborations mises en place dans le chapitre 8.

Le chapitre 9 présente mes activités de dissémination et le chapitre 10 détaille mes activités d'enseignement. Le dernier chapitre vient conclure et présenter les perspectives de recherche pour les prochaines années.

Chapitre 1

Curriculum Vitae

Sommaire

1.1	Actuellement	4
1.2	Formation et diplômes	4
1.3	Parcours	5
1.4	Synthèse de l'activité de recherche	6
1.4.1	Publications et Communications	6
1.4.2	Six publications les plus significatives	6
1.4.3	Développement logiciel	7
1.4.4	Collaborations	7
1.4.5	Organisation de workshops	7
1.4.6	Encadrement	8
1.4.7	Enseignements	8
1.4.8	Actions de Formation	8
1.4.9	Rayonnement scientifique	8
1.4.10	Montage de projets	9

Ce chapitre présente un résumé de ma formation, mon parcours et mes activités de recherche.

Nathalie BARTOLI, Maître de recherche à l'ONERA

Née le 4 juin 1974

Pacsée, 3 enfants

Tél : 05 62 25 26 44

Email : Nathalie.Bartoli@onera.fr

Adresse professionnelle : ONERA, DTIS/M2CI, 2 avenue Edouard Belin, 31055
Toulouse France

Page [web ONERA](#)

Page [Google Scholar](#)

1.1 Actuellement

Je suis actuellement maître de recherche de niveau 1 dans l'équipe Méthodes Multidisciplinaires et Concepts Intégrés du Département Traitement de l'Information et Systèmes (DTIS/M2CI). En parallèle de mes activités de recherche, j'ai une activité d'enseignements à l'ISAE-SUPAERO (environ 60 heures par an).

1.2 Formation et diplômes

- **2001** Qualification aux fonctions de Maître de Conférences - section 26
- **1997-2000** Doctorat de Mathématiques Appliquées, INSA Toulouse, *Modèles pour la diffraction d'ondes par des obstacles revêtus de couches minces. Résolution de problèmes de diffraction d'ondes via une formulation intégrale de type point-selle*, soutenue le 21 décembre 2000, Directeur de thèse : A. Bendali
- **1996-1997** DEA en Mathématiques Appliquées, INSA Toulouse - Université Paul Sabatier, Mention Bien
- **1992-1997** Ingénieur en Génie Mathématique et Modélisation, INSA Toulouse Modélisation (aéronautique et espace) et calcul scientifique (mécanique des structures, mécanique des fluides, électromagnétisme), analyse numérique, automatique, optimisation, probabilités et statistiques, informatique
- **1992** Baccalauréat série C (Mathématiques et sciences physiques, Mention très bien)

1.3 Parcours

- **Depuis 2018 Maître de recherche à l'ONERA** (Toulouse), département Traitement de l'Information et Systèmes, Unité Méthodes Multidisciplinaires et Concepts Intégrés
Thèmes de recherche : algorithmes d'optimisation, formulations MDO, modèles de substitution, gestion des incertitudes
- **2015-2018 Ingénieur de recherche détaché à l'IRT Saint-Exupéry** dans le projet MDA-MDO (mise à disposition 20% du temps)
Thèmes de recherche : formulations MDO adaptées au couplage avion global et processus haute-fidélité
- **2012-2017 Ingénieur de recherche à l'ONERA** (Toulouse), département Conception et évaluation des Performances des Systèmes, Unité Systèmes Aéronautiques
Thèmes de recherche : optimisation multi-disciplinaire adaptée à la conception avion avant-projet, modèles réduits pour l'optimisation en grande dimension
- **2005-2012 Ingénieur de recherche à l'ONERA** (Toulouse), département Traitement de l'Information et Modélisation, Unité Modélisation Mathématique et Simulation Numérique.
Thèmes de recherche : optimisation, schémas de décomposition-coordination, analyse post-optimale, modèles de substitution et mélanges d'experts, partitionnement (clustering), schémas numériques pour le calcul scientifique.
Relais de l'adjoint qualité du département et animateur qualité de l'unité M2SN
- **2003-2004 Chercheur au CERFACS** (Toulouse), équipe Electromagnétisme.
Thèmes de recherche : équations intégrales, éléments finis, méthodes multipolaires rapides et itératives.
- **2002 Post-doctorat INRIA Sophia-Antipolis** en collaboration avec Alcatel Space (Cannes), *Optimisation et validation d'un logiciel de calcul d'antennes*
- **2001 Chercheur au CERFACS** (Toulouse), équipe Electromagnétisme.
Thèmes de recherche : équations intégrales, éléments finis, méthodes multipolaires rapides et itératives.
- **1997-2000 Doctorant CERFACS** (Toulouse), équipe Electromagnétisme.
Etude d'une nouvelle formulation intégrale et de modèles numériques pour la diffraction d'obstacles revêtus. Bourse ministérielle.
- **1997 (6 mois) Stage de fin d'études-DEA au CERT-ONERA** (Toulouse),

équipe électromagnétisme : *Simulations numériques pour la détection et l'identification d'objets enfouis.*

1.4 Synthèse de l'activité de recherche

Mon activité de recherche est résumée dans les différents points listés ci-dessous. Chacun de ces points sera repris avec plus de détails dans les chapitres 2.1 à 10.

1.4.1 Publications et Communications

- 10 publications acceptées dans des revues à comité de lecture, 7 en cours de révision et 3 viennent d'être soumis.
- 1 chapitre de livre accepté chez l'éditeur Springer
- 1 chapitre de livre en révision dans la Collection Science, Editions ISTE
- 60 actes de conférences avec comités de lecture
- 30 conférences sans comité de lecture
- 1 brevet accepté
- 1 ouvrage scientifique aux Editions Cépaduès

1.4.2 Six publications les plus significatives

- N. Bartoli, T. Lefebvre, S. Dubreuil, R. Olivanti, R. Priem, N. Bons, J.R.R.A. Martins and J. Morlier, *Adaptive modeling strategy for high-dimensional constrained global optimization*, Aerospace Science and Technology Journal, March 2019, 5year-IF 2.439, 2017-IF 2.228
- M. A. Bouhlel, J. T. Hwang, N. Bartoli, R. Lafage, J. Morlier and J.R.R.A. Martins, *SMT : A Surrogate Modeling Toolbox in Python*, Advances in Engineering Software, Elsevier 2019, 5year-IF 2.948, 2017-IF 3.198
- M.-A. Bouhlel, N. Bartoli, R. Regis, A. Otsmane and J. Morlier, *Efficient Global Optimization for high-dimensional constrained problems by using the kriging models combined with the Partial Least Squares method*, Engineering optimization, 2018, 5year-IF 1.908, 2017-IF 1.622
- M.-A. Bouhlel, N. Bartoli, A. Otsmane and J. Morlier, *Improving kriging surrogates of high-dimensional design models by Partial Least Squares dimension reduction*, Structural and Multidisciplinary Optimization, vol 53, n°5, pp 935-952, 2016, 5year-IF 3.269, 2017-IF 2.876, Q1
- D. Bettebghor and N. Bartoli. *Approximation of the critical buckling factor for*

composite panels, Structural and Multidisciplinary Optimization, vol 46, n°4, pp 561-584, 2012, 5year-IF 3.269, 2017-IF 2.876, Q1

• D. Bettebghor, N. Bartoli, S. Grihon, J. Morlier and M. Samuelides. *Surrogate modeling approximation using a mixture of experts based on EM joint estimation*, Structural and Multidisciplinary Optimization, vol 43, n°2, pp 243-259, 2011, 5year-IF 3.269, 2017-IF 2.876, Q1

1.4.3 Développement logiciel

- Participation au développement de la toolbox open source python Surrogate Model Toolbox [SMT](#) qui propose des algorithmes de construction de modèles réduits (surrogate models) en collaboration avec l'Université du Michigan, la NASA, l'ONERA et l'ISAE-SUPAERO.

1.4.4 Collaborations

- Instituts académiques français (ISAE-SUPAERO, Institut Clément Ader, IRT Saint-Exupéry, ENAC, INRIA)
- Industriels (AIRBUS, SNECMA, BOMBARDIER, ATR)
- Instituts internationaux (MDO Lab/Université du Michigan (US), Université de St Joseph-Philadelphie (US), Université de Hong-Kong, Université de Floride (US), Université de Southampton (UK), DLR (Allemagne), NLR (Pays-Bas), Université de Naples (Italie), TsAGI (Russie), TUDelft (Pays-Bas), Nasa-Glenn (US))

1.4.5 Organisation de workshops

- Organisation de séminaires autour de la MDO dans le cadre de la Fédération de Recherche ONERA-ISAE-ENAC depuis 2018 (2 à 3 par an)
- Organisation d'une session spéciale pour le projet H2020 AGILE à l'AIAA 2018 à Atlanta.
- Organisation du premier workshop européen pour la plate-forme OpenM-DAO en collaboration avec l'ISAE-SUPAERO et la NASA Glenn, ONERA Toulouse octobre 2017, 60 participants [lien](#)
- Organisation du premier workshop européen sur la MDO ([lien agenda](#)) en collaboration avec l'IRT Saint-Exupéry et le DLR, Braunschweig Germany,

October 2017, 110 participants. La seconde édition devrait avoir lieu à l'automne 2019 à Toulouse

- Organisation de séminaires MDO avec l'ISAE-SUPAERO (environ 1 à 2 par an depuis 2015).

1.4.6 Encadrement

- 9 co-encadrements de thèse dont 6 thèses soutenues entre 2009 et 2018
- 2 thèses en fin de montage pour un démarrage courant 2019
- 29 co-encadrements de stage dont 19 en stage de fin d'études (niveau MR2)
- 2 co-encadrements de post-doc
- 12 co-encadrements de projets étudiants
- Participation à des jurys de thèse

1.4.7 Enseignements

- Responsable du module optimisation en 1ere année à l'ISAE-SUPAERO (amphis, PC et BE) depuis 2012
- ISAE-SUPAERO/ENSICA (Optimisation mono et multidisciplinaire, Optimisation linéaire et non-linéaire, Calcul scientifique, Equations aux dérivées partielles, Krigeage, Analyse numérique, Introduction à Matlab) depuis 2005
- INSA (Probabilités & Statistiques, Programmation Fortran 90) de 1997 à 2004
- Université Sciences Sociales (Statistiques, Algèbre linéaire) de 2000 à 2004

1.4.8 Actions de Formation

- Organisation et participation à la Formation EUROSAE *Machine Learning dans un contexte aéronautique : motivation, présentation et applications*, MTS 003, (2018 et 2019)
- Organisation et participation à la formation interne ONERA sur les métamodèles (2014, 2015 et 2019)

1.4.9 Rayonnement scientifique

- Animatrice du thème X-OAD (eXtended-Overall Aircraft Design, MDO et ingénierie système) au sein de la fédération ONERA-ISAE-ENAC depuis

2018

- Membre du comité technique de l'AIAA pour la thématique Multidisciplinary Design Optimization (MDO-TC) depuis 2017
- Membre de la Société *International Society of Global Optimization* depuis 2019
- Animatrice thématique Conception et Optimisation des Systèmes au sein du département, 2016 et 2017
- Responsable de la thématique "Optimisation Multidisciplinaire" dans l'Equipe Commune de Recherche ONERA-ISAE (ECR MOIS), 2014 à 2017
- Membre du GdR MascotNum depuis 2006
- Activité de relecture pour deux revues scientifiques : Structural Multidisciplinary Optimization, Transactions on Mathematical Software

1.4.10 Montage de projets

- Participation au montage de l'ANR France & Hong-Kong University of Science and Technology (mars 2018) avec ENAC (coordination), demande d'une bourse de voyage obtenue pour l'année 2019
- Participations aux montages de projet RTRA (OSYCAF, TFC, TigDISC)
- Participation à différents montages de projet H2020 (FlyAhead, AGILE, SIBCAT, AGILE 4.0)
- Projet MDA-MDO avec l'IRT Saint-Exupéry (2011-2014)
- Participation au montage de l'ANR STARGATE avec l'ENAC (coordination) entre 2010 et 2012

Chapitre 2

Présentation générale des activités de recherche

Sommaire

2.1	Résumé	12
2.2	MDO pour la conception	12
2.2.1	Introduire des outils haute fidélité	14
2.2.2	Introduire de nouvelles disciplines	17
2.2.3	Introduire des modèles de substitution	19
2.2.4	Introduire des modèles de fidélité variée	28
2.2.5	Disposer d'optimiseurs efficaces pour les boîtes noires	30
2.2.6	Formulations MDO adaptées	38
2.2.7	Prise en compte des incertitudes	39
2.2.8	Analyse de sensibilité	42
2.3	Mon projet de recherche à l'ONERA	43

Ce chapitre présente mes différentes activités de recherche autour de l'optimisation multidisciplinaire pour la conception avion et fait le lien avec la liste des publications sélectionnées dans le chapitre 4.

2.1 Résumé de mon parcours de recherche

De 1997 à 2005, ma recherche était concentrée autour des équations intégrales pour l'électromagnétisme, la modélisation mathématique des problèmes de diffraction d'ondes et la mise en place de méthodes rapides pour la résolution des systèmes obtenus (préconditionnement, méthode multipôle, ...). Ce travail a été réalisé au CERFACS et à l'INRIA Sophia Antipolis, ponctué par une activité de recherche contractuelle pour le CEA-CESTA, le CNES et Alcatel-Space (Cannes). A la suite de mon embauche à l'ONERA en 2005, j'ai poursuivi mon activité sur les schémas numériques pour les équations aux dérivées partielles en travaillant sur les méthodes éléments finis d'ordre élevé (via des projets internes PRF MAHPSO [18], ARF DMORPH [87]) tout en démarrant une nouvelle activité autour de l'optimisation multidisciplinaire et les modèles de substitution via des projets internes (PRF DOOM, PRF HEPHAIISTOS), des projets européens (FP6 INTELLECT D.M. et FP7 MAAXIMUS), un projet DPAC ARTEMIS et les co-encadrements des thèses de Joël Clément et de Dimitri Bettebghor.

Depuis 2013, j'ai changé de département et concentré mon activité de recherche autour de l'optimisation multidisciplinaire pour la conception avion avant-projet en participant à différents projets internes (PRF CREATION, PR VESTA, PRF ACADIA, PR MUFIN), le projet européen H2020 AGILE et le projet MDA-MDO avec l'IRT Saint-Exupéry.

Dans ce manuscrit, je vais donc essentiellement décrire cette dernière partie de mon activité relative à l'optimisation multidisciplinaire et aux modèles de substitution appliqués à la conception.

2.2 Optimisation multidisciplinaire pour la conception avion avant-projet

La conception de produits est plus que jamais collaborative. Les équipes de concepteurs, bien souvent experts dans une discipline (fluide, structure, propulsion,...) doivent donc disposer à la fois de logiciels qui puissent interagir entre eux mais aussi d'une plateforme qui centralise dans un même espace l'ensemble de calculs nécessaires à la conception du produit. On parle alors de conception multidisciplinaire. Par exemple, dans la conception d'une aile d'avion, la discipline structure a pour souhait d'augmenter l'épaisseur pour avoir une meilleure résistance, alors que le but de la discipline aérodynamique est au contraire de réduire l'épaisseur de l'aile afin de diminuer la traînée. L'optimisation multidisci-

plinaire est donc un moyen de trouver le meilleur compromis entre les différentes disciplines. Ces dernières décennies ont vu naître un intérêt significatif et croissant pour améliorer l'efficacité des processus de conception de véhicule via le développement d'outils et de techniques dans le domaine de l'optimisation multidisciplinaire (MDO pour Multidisciplinary Design and Optimization). La conception de véhicules aérospatiaux est une compétence clé de l'ONERA et est au coeur de ses préoccupations scientifiques et applicatives. Cette compétence mêle donc des aspects scientifiques bien précis comme la modélisation de phénomènes physiques, le développement de modèles de substitution (appelés métamodèles ou surrogate models) ou la formulation de problèmes d'optimisation, et des aspects techniques tels que la disponibilité d'outils informatiques partagés adéquats permettant une automatisation du processus de MDO. En effet, de nombreuses optimisations ont été effectuées sur la configuration classique "tube and wing" validées par des campagnes d'essais depuis des années afin de diminuer de manière significative la consommation de carburant de ces configurations dites "classiques". Afin de continuer à améliorer les performances des avions, il devient à présent nécessaire d'explorer d'autres configurations non-conventionnelles, telles que l'aile volante, l'aile haubanée ou l'aile jointive, couplées à des technologies innovantes comme la propulsion distribuée, ou l'ingestion de couche limite, configurations pour lesquelles l'absence de connaissances accumulées ne permet pas de disposer de modélisations simplifiées et impose de revenir aux principes physiques fondamentaux pour en analyser les caractéristiques particulières. L'optimisation multidisciplinaire apparaît depuis quelques années comme l'outil incontournable pour concevoir ces nouveaux avions.

Plusieurs points découlent de cette problématique liée aux nouveaux concepts :

- intégrer des d'outils de simulation haute fidélité,
- introduire de nouvelles disciplines liées aux nouveaux concepts,
- réduire les temps de calcul en utilisant des modèles de substitution,
- prendre en compte des outils de fidélité variée et assurer le couplage,
- être capable de traiter un grand nombre de variables et de contraintes,
- disposer d'optimiseurs efficaces qui minimisent le nombre d'évaluations des fonctions coûteuses,
- développer des approches pour exploiter les modèles de fidélité variée
- s'assurer de la convergence du problème en proposant une formulation adaptée,
- prendre en compte les incertitudes de modèles introduites par la simulation et/ou les modèles réduits.

Mon travail de recherche depuis 2005 s'articule autour de ces différents points représentés sur la Figure 2.1 que je vais détailler dans les sections qui vont suivre.

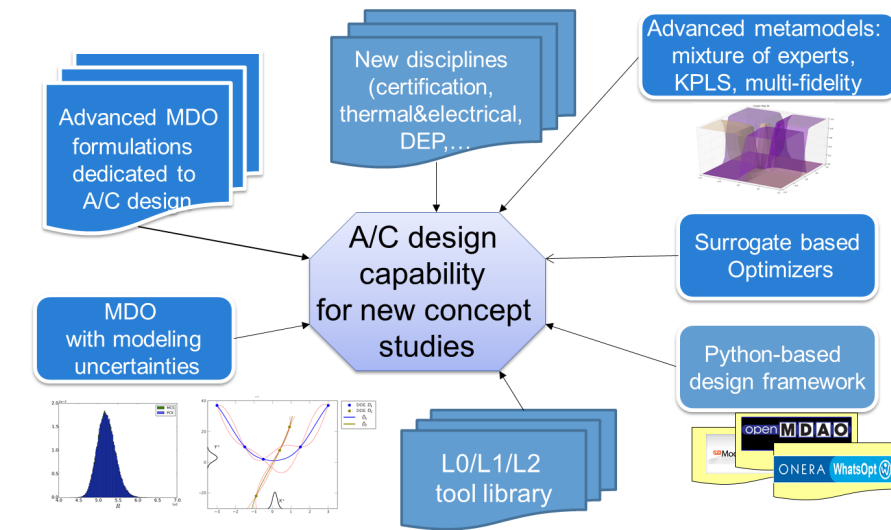


FIGURE 2.1 – Axes de recherche autour de la conception avion.

L’aspect capitalisation des méthodes et outils développés est un enjeu très important pour améliorer le processus de conception, j’essaie donc de m’assurer depuis quelques années que tous les développements réalisés soient faits en python et intégrables dans la mesure du possible dans la plateforme OpenMDAO [94] de la NASA, dans la toolbox opensource SMT [194] et dans l’environnement de conception ”WhatsOpt” [123] développé en interne ONERA. Ce travail essentiel de capitalisation est fait en collaboration avec les personnes de l’équipe avec l’aide de R. Lafage (ONERA/DTIS), expert en développement logiciel.

2.2.1 Introduire des outils haute fidélité

L’amélioration du processus de conception avion nécessite à la fois d’augmenter la précision des modèles utilisés et d’augmenter le spectre des disciplines à prendre en compte le plus tôt possible dans le processus de design, notamment pour l’étude de nouvelles configurations. Dans le cadre de la thèse de J. Mas Colomer qui a débuté en octobre 2015 en collaboration avec l’ISAE, nous nous intéressons à la conception de démonstrateurs en vol dans le but d’étudier les nouvelles configurations et de valider les outils numériques, tout en évitant les coûts de conception, construction et exploitation d’un avion à taille réelle. Pour pouvoir utiliser les résultats obtenus à travers l’utilisation de démonstrateurs en vol, il faut garantir certains aspects de similitude entre la maquette à échelle et l’avion complet. Si l’on

considère la similitude aéroélastique (objet de la thèse de J. Mas Colomer), tant statique que dynamique, celle-ci se traduit par l'obtention de la même déformée de la voilure, ainsi que les mêmes modes, vitesses et fréquences de flottement (à un facteur d'échelle près). Dans le cas où la similitude aérodynamique n'est pas garantie (domaine de vol du drone limité) où la structure de la maquette n'est pas une copie homothétique de l'avion de référence, l'optimisation se présente comme une solution pour trouver les paramètres de conception qui respectent au mieux ces critères de similitude. L'obtention de ces critères a nécessité l'introduction de modèles de plus haute fidélité, tels que PANAIR [83] pour la partie aérodynamique ou NASTRAN [141] pour la partie structure, au sein du processus de conception. Un travail a été réalisé sur le couplage aéro-structure en statique pour traiter des problèmes de déformées fixes et en dynamique pour traiter les problèmes de flottement comme l'illustre la Figure 2.2. Ces études ont fait l'objet de rapports techniques internes [57, 58], d'une publication dans la conférence AIAA SCITECH 2017 [139], d'une communication à WCSMO en 2017 [60], d'une communication à ENGOPT2018 [61] et d'un article soumis [140] avec le Pr J.R.R.A. Martins, responsable du MDOLab à l'Université du Michigan. Un séjour de 3 mois a été effectué par J. Mas Colomer au MDOLab dans le cadre de cette collaboration (septembre à décembre 2017) avec le support de l'Université Fédérale Toulouse Midi-Pyrénées et l'Ecole des Docteurs. Les développements réalisés sont effectués dans la plateforme OpenMDAO développée par la NASA [94, 95].

Pour l'estimation simplifiée des performances d'un avion, il est possible d'utiliser la formule de Bréguet [8] bien connue en aéronautique pour calculer le rayon d'action (niveau L0 [53]). Dans le cadre d'études de conception avion, où le dimensionnement de l'appareil est effectué, une analyse basée sur intégration temporelle est généralement utilisée (niveau L1 [53]). Dans les travaux de thèse de P. Schmollgruber, afin de simuler une mission complète d'un aéronef, un modèle de simulation plus évolué est intégré afin d'acquérir plus d'information sur le comportement de l'appareil (niveau L2 [53]). Pour ce modèle plus évolué, le choix a été fait d'utiliser le simulateur opensource JSBSIM [3] basé sur un modèle à six degrés de liberté (trois translations et trois rotations) qui possède différentes boucles de contrôle pour assurer la stabilité et la navigation. Pour cet usage, des coefficients aérodynamiques doivent être fournis en entrée et des modifications dans le processus de dimensionnement ont dû être apportées dans l'outil de conception avion avant-projet FAST (Fixed wing Aircraft Sizing Tool) [182]. La Figure 2.3 illustre les différents modèles possibles pour représenter la mission d'un avion : de la formule de Bréguet à un découpage en segments de mission pour arriver à une trajectoire simulée dans JSBSIM. La validation de ce couplage entre des modules avant projet

et des modules plus évolués pour les lois de contrôle et de navigation a été faite en comparant des données simulées et des données issues de trajectoires d'A320 réalisées (données enregistrées par une antenne ADSB à l'ONERA). Ces comparaisons menées dans le cadre de la thèse de P. Schmollgruber [180] ont été réalisées avec l'aide de J. Bedouet (ONERA/DTIS) expert en Air Traffic Management. Ce travail a donné lieu à deux articles à AIAA/SCITECH (2017 et 2018) [182, 181]

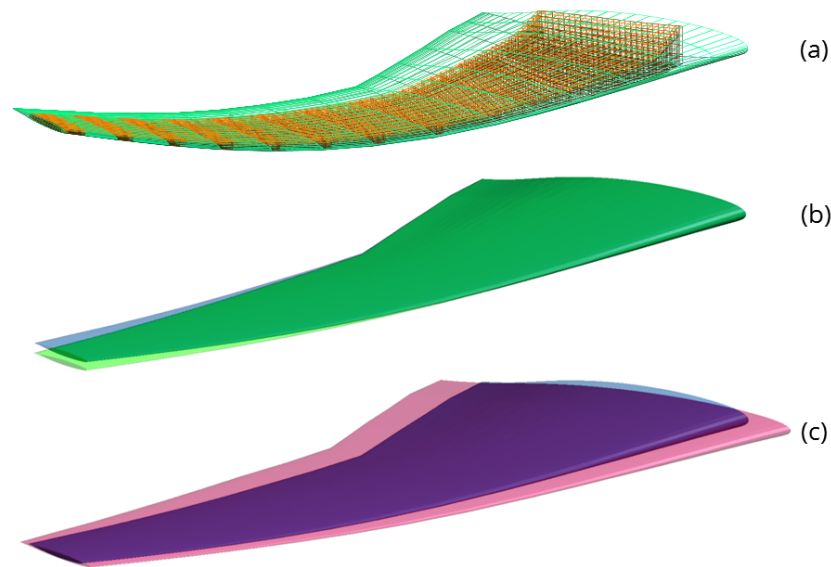


FIGURE 2.2 – Exemple de couplage aéroélastique entre un modèle structures (orange) et un modèle aérodynamique (vert) (a). Effet de la différence des effets de compressibilité sur la déformation de l'aile (b). Aile optimisée pour le même déplacement en bout d'aile malgré des conditions de compressibilité différentes (c) [59].

Bilan : 3 articles à la conférence internationale AIAA/SCITECH (2017, 2018), 2 conférences européennes (2017, 2018) et une collaboration avec le MDOLab (séjour du doctorant J. Mas Colomer au MDOLab en 2017 pour 3 mois). Un manuscrit de thèse a été rédigé sur cette thématique [59] et un chapitre de thèse [180].

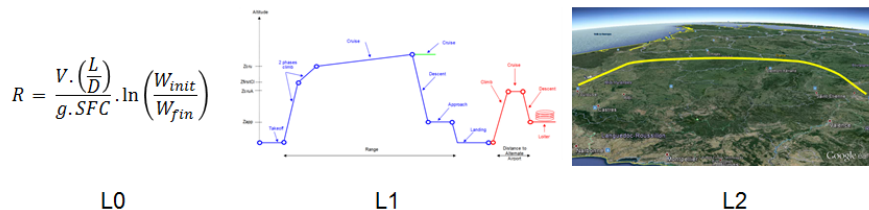


FIGURE 2.3 – Exemples de différents modèles pour représenter la mission de l’avion [180].

2.2.2 Introduire de nouvelles disciplines

Pour concevoir de nouveaux aéronefs, comme l’aile haubanée, l’aile volante, l’aile jointive ou bien de nouveaux systèmes opérationnels (assistance au décollage et à l’atterrissage), plusieurs études récentes ont montré l’intérêt de prendre en compte les contraintes de certification au niveau avant-projet étant donné leurs forts impacts sur le design d’un appareil [179]. De même, prendre en considération le système d’augmentation de stabilité au même niveau de conception permettrait d’ouvrir l’espace de design et d’augmenter les gains en performance [162]. Ce travail a fait l’objet de la thèse de P. Schmollgruber [180], ingénieur de recherche au DTIS, que j’ai accompagné dans sa démarche. L’idée est donc de repenser le processus MDAO (Multidisciplinary Design Analysis and Optimization) afin d’intégrer de manière efficace ces nouvelles disciplines nécessitant des modèles plus complexes et/ou de nouvelles données. Ainsi, on propose de considérer la certification via l’intégration d’un module spécifique contenant les contraintes de la CS-25 (Certification Spécifications fournies en Europe par l’EASA). En ce qui concerne la prise en compte du système d’augmentation de la stabilité, l’ajout d’un modèle de plus haute fidélité du comportement de l’avion (i.e. un modèle avec 6 degrés de liberté) s’est avéré indispensable comme discuté dans la Section 2.2.1. Les premiers résultats obtenus conduisent à poursuivre ces activités en se concentrant sur l’implémentation efficace de l’appel au modèle de comportement de l’avion. Ce travail a été présenté dans plusieurs conférences [183, 184, 26, 182] et il a fait l’objet du développement python d’un outil de conception préliminaire avion appelé FAST pour Fixed-Wing Aircraft Sizing Tool (collaboration ONERA-ISAE-SUPAERO). En parallèle de cette activité sur la certification, il est apparu intéressant d’associer de nouvelles technologies comme la propulsion distribuée, l’ingestion de la couche limite,... pour la conception de ces nouveaux aéronefs [201, 1]. La thèse de A. Sgueglia en collaboration avec l’ISAE qui a démarré en janvier 2017 a pour but

de définir un processus de conception et d'optimisation multidisciplinaire pour une aile volante à propulsion distribuée. Au cours des deux premières années de thèse, les points suivants ont été abordés : intégration de la chaîne propulsion distribuée et étude des nouveaux paramètres liées aux nouvelles technologies (batteries, ...), gestion des couplages et des priorités au sein du processus de dimensionnement, mise en place de la formulation MDO avec une boucle d'optimisation (mono et multi-objectif). La dernière année sera consacrée à l'intégration de modèles physiques dédiés à mieux représenter les phénomènes aérodynamiques avec l'ingestion de la couche limite due aux moteurs enterrés. Les développements sont réalisés dans la version FAST couplée à la plate forme OpenMDAO suite à la collaboration faite avec la NASA et l'Université du Michigan qui ont accueilli A. Sgueglia pendant 3 mois (janvier à avril 2018). Le passage du code FAST dans OpenMDAO a nécessité le codage des dérivées analytiques de chacun des modules utilisés ce qui a permis d'observer des gains en temps de calcul dans les boucles de convergence (résolution de la MDA), un facteur 5 environ a été obtenu [187]. Ce travail de collaboration fait l'objet d'un article soumis au Journal of Aircraft [187]. Les premiers résultats ont permis la comparaison d'un avion conventionnel type A320 avec un avion à propulsion hybride distribuée (de 20 à 40 moteurs comme illustré sur la Figure 2.4) [186] ou avec une aile volante avec des moteurs conventionnels [188]. La dernière étape de la thèse sera l'application à une aile volante avec une propulsion hybride distribuée.

Plus récemment, dans le cadre d'une collaboration avec C. Gogu de l'Institut Clément Ader, nous avons lancé un projet étudiant avec des élèves de Master de l'ISAE-SUPAERO pour considérer dans le processus de conception les différentes interactions avec les autres acteurs du milieu tels la concurrence, les utilisateurs finaux (compagnies aériennes) et les organismes de réglementation. La stratégie mise en place dans la thèse de G. Waycaster [203] est basée sur la théorie des jeux afin d'intégrer directement l'effet de ces interactions dans un problème d'optimisation de conception. Dans un premier temps l'idée serait de considérer des modèles physiques un peu plus complexes que les modèles analytiques initialement utilisés en considérant le logiciel FAST [182] dans un processus d'optimisation. Dans un second temps, on pourra quantifier le gain de cette approche intégrée par rapport à une approche plus classique. Le projet des étudiants de Master se terminant en février 2020, des premiers résultats devraient donc être obtenus courant 2019.

Bilan : 3 articles à la conférence internationale AIAA/SCITECH (2015, 2017, 2018), 3 conférences européennes (2014, 2015, 2018), un article soumis [187], une collaboration avec le MDOLab et la NASA (séjour du doctorant A. Sgueglia en 2018 pour 3 mois), une collaboration avec l'ICA. Un chapitre de thèse a été rédigé

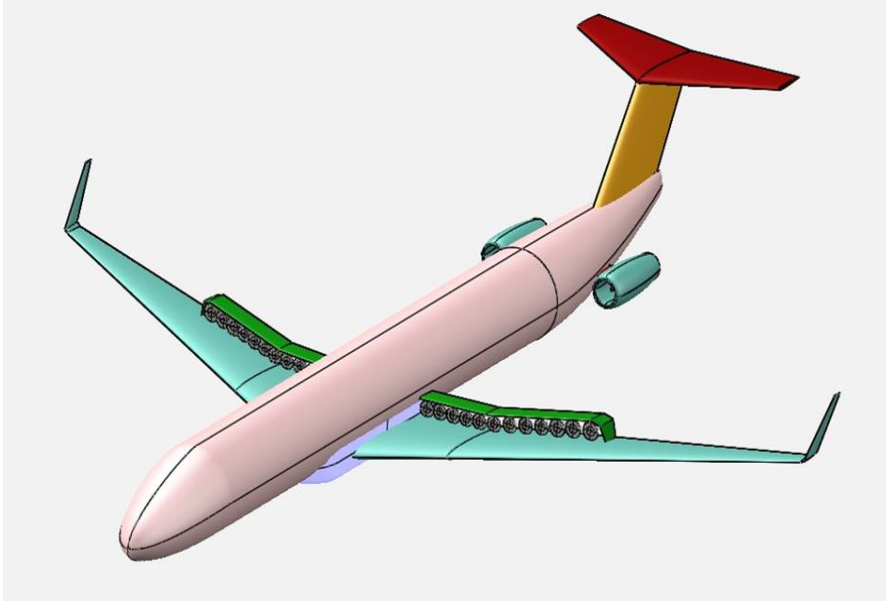


FIGURE 2.4 – Avion à propulsion distribuée [186].

sur cette thématique [180].

2.2.3 Introduire des modèles de substitution

La difficulté des problèmes de conception à résoudre réside alors dans la conjonction de plusieurs facteurs : nombre élevé de variables de conception et parfois d'objectifs ou de contraintes à satisfaire, mélange de variables de nature à la fois discrète et continue. De plus un grand nombre de simulations du système peut être requis et le coût en temps calcul par évaluation peut alors être une limite à la résolution du problème d'optimisation. Le fait de complexifier le problème de conception en rajoutant des outils haute fidélité comme PANAIR ou NASTRAN (cf Section 2.2.2) se fait au prix de temps de calcul importants et a tendance à transformer le simulateur en une véritable boîte noire. La stratégie actuelle pour réduire de façon significative le temps de calcul consiste à construire une surface de réponse représentative du modèle, appelée modèle réduit (ou métamodèle ou encore modèle de substitution) et d'utiliser cette approximation analytique dans le processus de conception ou d'optimisation [102]. Il existe de nombreuses façons de construire un métamodèle, faisant appel à certaines hypothèses adaptées au type de données disponibles et/ou à la taille de la base de données [89, 104]. Dans le cas de dépendances simples, des modèles linéaires ou polynomiaux peuvent être

utilisés, tandis que des interactions plus complexes requièrent la construction de modèles plus élaborés, tels que les réseaux de neurones, les fonctions à base radiales, les splines ou encore la régression par processus gaussiens (également appelée krigeage). J’ai commencé mes activités sur ce domaine dans le cadre d’un projet interne à l’ONERA (PRF DOOM) où j’ai mené différentes études pour recenser les modèles existants afin de les décrire [20, 14] et également comparer leurs performances [12]. Mes premiers travaux entre 2005 et 2008 concernaient l’application des réseaux de neurones (Projet FP7 INTELLECT D.M. et Projet interne PRF HEPHAIISTOS) pour une chambre à combustion en utilisant le code de calcul CEDRE de l’ONERA dédié à la simulation numérique des phénomènes multi-physiques dans le domaine de l’énergétique et de la propulsion [119, 118, 120].

Depuis 2011, ma contribution dans ce domaine des métamodèles concerne différents points :

- la génération de plans d’expériences optimaux pour générer les bases de données,
- la mise en place de nouveaux modèles (mélanges d’experts, modèles de krigeage couplés à la méthode des moindres carrés partiels KPLS et KPLS+K, co-krigeage)
- la capitalisation de ces nouveaux modèles dans une toolbox python,
- l’optimisation des hyperparamètres du krigeage,
- l’application des réseaux de neurones (statiques et dynamiques) à la météo spatiale.

L’ensemble des développements réalisés est capitalisé depuis 4 ans dans différentes toolboxes python gérées via svn, Redmine (Projet ONERAMDAO en interne ONERA) ou github ([OneraHub](#)). De plus depuis 2014, je participe à la formation Machine Learning qui est faite en interne ONERA (2014, 2015 et 2019) ou bien dans le cadre de EUROSAB (2018, 2019) en décrivant ce qui est fait autour des processus gaussiens d’un point de vue théorique et applicatif via des notebooks en python [earlyML](#).

Je vais donc décrire dans la suite l’ensemble de ces contributions.

2.2.3.1 Un plan d’expériences optimal

Avant toute chose, il faut choisir des points pour la construction d’un métamodèle. Chaque observation de la fonction objectif peut induire un temps de calcul considérable et pour cette raison, il est important de bien choisir le plan d’expériences (Design of Experiments) pour obtenir le meilleur modèle possible. L’enjeu revient alors à couvrir la plus grande partie du domaine de définition de la fonction. Il faut

aussi que le plan ait de bonnes qualités en projection, afin d'éviter d'éventuelles répétitions de données dans le cas où une variable ne serait pas influente. Parmi les plans à remplissage d'espace, les LHD (Latin Hypercube Designs) sont les plus utilisés du fait de leur simplicité de fabrication et leurs qualités générales [143] et il est commun de les construire à partir d'un nombre de points n égal à 10 fois la dimension de l'espace [132]. Il existe différents critères pour mesurer la qualité d'un plan d'expériences [152]. En utilisant l'algorithme ESE (Enhanced Stochastic Evolutionary [111]), il est possible d'améliorer la qualité du plan d'expériences en optimisant un critère basé sur le calcul de distance entre chacun des points du plan comme le montre la qualité des différents plans d'expériences LHS (pour Latin Hypercube Sampling) sur la Figure 2.5. À partir du calcul du critère de distance pour un plan initial de n points, il est facile de le mettre à jour après permutation de deux points, le tout avec une complexité temporelle en $\mathcal{O}(n)$. La permutation est le cœur de l'algorithme ESE ; le calcul complet du critère ne se fait alors qu'une seule fois, d'où une économie en temps de calcul importante. Cet algorithme développé dans le cadre du stage de R. Vauclin (mars-septembre 2014) est implémenté dans la toolbox python *doe* disponible en interne ONERA et depuis peu dans la toolbox opensource [SMT](#).

Bilan : 1 rapport de stage, un module disponible en python dans la toolbox interne et dans la toolbox opensource SMT.

2.2.3.2 Un plan d'expériences adaptatif

Une fois le plan d'expériences construit à partir d'un nombre de points fixés par l'utilisateur (il est commun de prendre 10 fois plus de points que la dimension de l'espace des variables de design [132]), il peut être intéressant de l'enrichir en rajoutant des points bien choisis. A partir des travaux de [165, 173, 131], nous avons testé durant l'été 2018 avec le stage de T. Lahire (Etudiant ISAE-SUPAERO) différents critères d'enrichissement basés sur l'erreur de variance (variance-to-mean ratio). Les résultats très préliminaires ont permis de montrer l'effet de cet enrichissement pour venir réduire globalement l'erreur globale par rapport à une stratégie à nombre de points fixés. Pour une erreur donnée, il faut environ 20% de points en moins en choisissant de manière adaptative les points plutôt que via un LHS obtenu à nombre de points fixe. Un exemple d'application pour approcher la fonction G07 (décrite dans l'Annexe 12.1) en dimension 10 est illustré sur la Figure 2.6 où le plan d'expériences initial de 40 points est enrichi itérativement jusqu'à 100 points. A partir de 30 points ajoutés, l'erreur de validation calculée via une base de test devient inférieure à l'erreur obtenue avec un LHS fixe de 100 points, soit

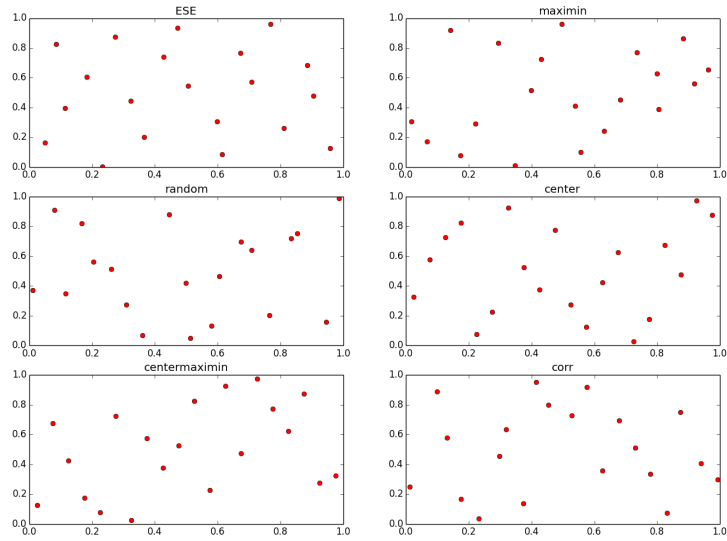


FIGURE 2.5 – Comparaisons de différentes stratégies pour remplir un espace de dimension 2.

une économie de 30% de points sur cet exemple.

Bilan : 1 rapport de stage, un code et un notebook disponible dans le package python *doe* de l'ONERA.

2.2.3.3 La technique des mélanges d'experts

Pour approcher des fonctions de plus en plus complexes, nous avons développé des modèles appelés mélanges d'experts. Adaptée des travaux de Jordan et Jacobs [114] pour des modèles linéaires, cette technique développée en 2011 dans le cadre de la thèse de D. Bettebghor (ONERA-ISAE-AIRBUS) [34] permet d'approcher des fonctions discontinues ou à dérivées discontinues. Plutôt que de construire un seul modèle réduit, on construit sur la base d'apprentissage un certain nombre de modèles appelés alors experts locaux qu'on combine en un seul afin d'améliorer la précision. Cette combinaison se fait sur la base d'un clustering ou partitionnement automatique de la base d'apprentissage au moyen de mélange de gaussiennes comme le montre la Figure 2.7b où 5 sous-domaines ont été identifiés automatiquement par l'algorithme. Des travaux sur le choix optimal du nombre de clusters ont été effectués à partir de différents critères d'erreur calculée par validation croisée

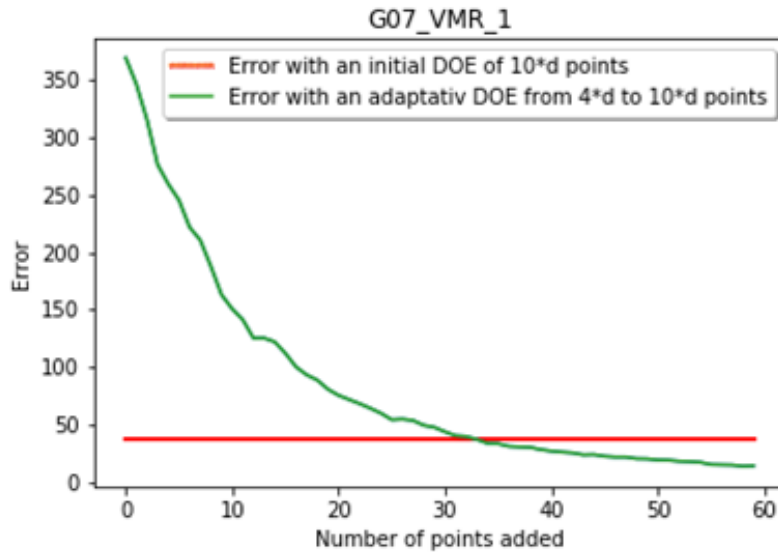
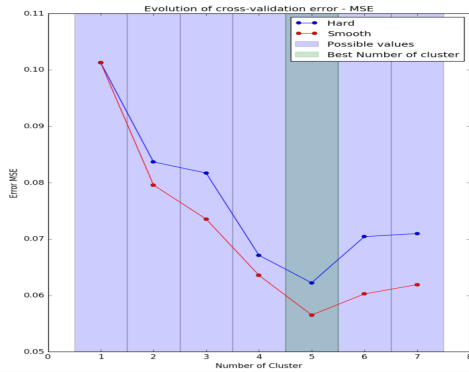


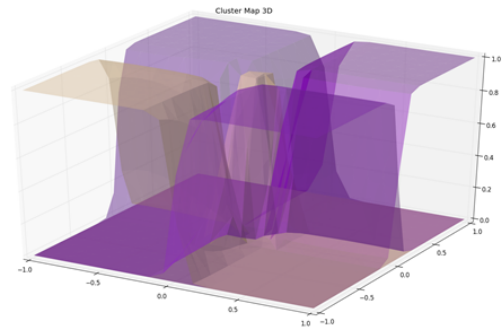
FIGURE 2.6 – Plan d’expériences adaptatif pour approcher la fonction G07.

comme l’illustre la Figure 2.7a. Le modèle final ainsi obtenu peut être continu (avec une dérivée analytique) ou discontinu suivant le choix de l’utilisateur. Ce travail a fait l’objet de deux articles [36, 35] et a été appliqué à différentes problématiques : problèmes de flambement de panneaux raidis [15] et [16], dimensionnement de panneaux avec des composites courbes, dimensionnement de rotors d’hélicoptères dans le cadre du projet interne PRF CREATION [27] ou bien pour générer des empreintes acoustiques [25]. Ces travaux ont été présentés à différentes manifestations externes [176, 33, 17, 27] et internes [30, 32]. Les outils informatiques initialement développés en Matlab avec la librairie IMAGE (Improved Metamodelling Approximation through Gaussian mixture of Experts) [29] ont ensuite été poursuivis en python. De plus, l’encadrement de différents stagiaires sur le sujet a permis de développer, améliorer et valider les modules python et de les intégrer dans la toolbox opensource SMT en collaboration avec R. Lafage (ONERA/DTIS). Le module *moe* est disponible en interne depuis 2016 et en externe depuis janvier 2018 dans SMT. Des comparaisons avec la méthode de mélanges d’experts développée par R. Liem dans le cadre de sa thèse au MDOLab [130, 129] ont été effectuées pour approcher les coefficients aérodynamiques d’une aile volante dans le cadre du stage de R. Priem [169] via une collaboration avec l’Université de Hong-Kong.

Ce travail sur les mélanges d’experts est présenté à partir de l’article [36] dans la Section 4.4.1.



(a) Sélection du nombre optimal de clusters



(b) 5 clusters

FIGURE 2.7 – Exemple de mélanges avec 5 experts sur la fonction Norm1.

Bilan : 2 articles, 1 chapitre de thèse, 4 conférences, 2 rapports de stage, 1 jupyter notebook, 3 rapports ONERA, une collaboration ONERA/ISAE/MDOLab/-NASA/Univ Hong-Kong et un module *moe* disponible en python sous svn dans la toolbox interne ONERA ou dans la toolbox opensource SMT (version 0.2).

2.2.3.4 Etude du modèle de krigeage et ses hyperparamètres

Parmi les métamodèles les plus couramment utilisés, le krigeage est le seul qui fournit de manière intrinsèque une estimation de l’erreur de prédiction. Cette estimation d’erreur est une information clé utilisée dans les algorithmes d’optimisation avec enrichissement (méthode EGO pour Efficient Global Optimisation [112]) ou dans des approches de calcul de fiabilité (méthode AK-MCS pour Active learning reliability method combining Kriging and Monte Carlo Simulation [81]), c’est la raison pour laquelle ce métamodèle est largement utilisé aujourd’hui.

Le krigeage fait intervenir un lot de paramètres inconnus $(\theta, \sigma^2, \beta)$. La méthode du maximum de vraisemblance [157] se prête particulièrement bien pour les calculer : on obtient une solution analytique simple pour les paramètres σ^2 et β tout en garantissant rapidité et précision. Il reste à estimer les paramètres θ (appelés encore hyperparamètres caractérisés par un vecteur de dimension d égale à la taille de l’espace de conception) pour minimiser la log-vraisemblance. Il n’y a pas de solution analytique à ce problème ; il faut utiliser un optimiseur efficace car cette fonction est fréquemment multimodale ; il est impossible d’utiliser un algorithme d’optimisation local. À noter qu’avec la montée en dimension, ce problème devient rapidement très difficile et coûteux en temps, ce qui nuit à la performance du

krigeage. Cette étape est d’autant plus longue et délicate dans la construction du modèle qu’il y a de paramètres de conception (a priori d paramètres) et qu’il y a de points dans la base de données (observations) : il faut donc définir judicieusement la méthode d’optimisation. Différents développements ont donc été menés pour améliorer les performances du krigeage en collaboration avec l’ISAE via des stages ou des thèses en co-encadrement [197, 40, 169, 93].

- Une première contribution a été d’implémenter le gradient de la fonction de log-vraisemblance par rapport à ses différents paramètres à partir des travaux réalisés dans [193]. De ce fait, l’algorithme L-BFGS-B (une méthode de quasi-Newton dans un domaine borné avec approximation de la matrice hessienne via la méthode Broyden-Fletcher-Goldfarb-Shanno), basé sur le calcul analytique des gradients, a ainsi pu être utilisé pour l’optimisation des hyperparamètres du krigeage. Le code disponible permet également d’utiliser les algorithmes qui n’utilisent pas l’information du gradient comme COBYLA (Constrained Optimization BY Linear Approximation [168]) ou CMA-ES [103] mais qui dans ce cas précis se sont montrés moins efficaces. Le gradient analytique de la fonction de vraisemblance a été validé pour différents types de noyaux stationnaires (exponentiel, Matérn 3/2 et 5/2).
- La contribution principale concerne l’estimation des paramètres d’un modèle de krigeage en grande dimension. Cette estimation peut être coûteuse en temps, notamment si la dimension du problème en entrée est importante. Le problème d’optimisation de la vraisemblance se fait dans un espace de dimension d . Nous nous plaçons dans le cas où il n’est pas possible de réduire a priori la dimension du problème. Il faut alors définir des techniques qui permettent de réduire le temps de calcul intrinsèque du modèle. L’idée, qui a été développée dans le cadre de la thèse de M.-A. Bouhrel en collaboration avec la SNECMA et l’ISAE, consiste à combiner la méthode Partial Least Squares (PLS) et le krigeage [44] et ainsi utiliser des informations sur la contribution de chacune des variables sur la réponse pour pondérer les paramètres de corrélation θ du krigeage. Le nombre de paramètres θ_i est considérablement diminué, ce qui réduit considérablement le nombre de calculs à mettre en œuvre (notamment dans la maximisation de la vraisemblance) et donc augmente la vitesse de construction des métamodèles de krigeage. Enfin les propriétés de la matrice de corrélation classique sont conservées avec cette nouvelle expression de la fonction de corrélation. L’algorithme KPLS (appelé ainsi car il combine le krigeage et la méthode PLS) présente l’avantage d’être très facile à implémenter informatiquement. La méthode peut être facilement adaptée aux différents types

de noyaux de covariances stationnaires (Matérn 3/2 et 5/2). Un deuxième modèle (KPLS+K) a été proposé, il utilise la méthode KPLS pour initialiser les hyperparamètres d'un krigeage classique et ainsi réduire le temps lié à l'optimisation des paramètres θ tout en travaillant dans la dimension initiale d . Ces modèles ont été validés sur des fonctions analytiques jusqu'à la dimension 124 (fonction MOPTA [113]) et sur des codes de calcul SNECMA avec 98 variables, les gains en temps CPU sont considérables par rapport à un krigeage classique (facteur de gain de l'ordre de 30 à 50). Pour exemple, le Tableau 2.1 présente les résultats sur la fonction analytique G07 en dimension 10 (fonction objectif du problème décrit en Annexe 12.1) où le modèle de krigeage ordinaire de scikit-learn (version 0.14 [159]) est comparé au modèle KPLS avec différents nombres de composantes principales retenues dans la décomposition PLS ($h = 1$ à $h = 3$). L'apprentissage est fait à partir de 100 points LHS et la validation sur un échantillon de 5000 points sur lequel l'erreur relative est calculée à l'aide de la formule suivante

$$ER = \frac{\|\hat{y} - y_{test}\|_2}{\|y_{test}\|_2} 100.$$

Une deuxième comparaison est faite sur la fonction de Griewank en dimen-

Métamodèles	ER (%)	temps CPU
krigeage ordinaire	0.013	5.14 s
KPLS1	0.014	0.11 s
KPLS2	0.0015	0.43 s
KPLS3	0.0008	0.44 s

TABLE 2.1 – Erreur de prédiction de la fonction G07 en dimension 10, avec un plan d'expériences de 100 points entre le krigeage ordinaire de scikit learn (v0.14) et la méthode KPLS. Le nombre de composantes principales retenues pour KPLS varie de 1 à 3 [40]

sion 60 (voir la description de la fonction dans l'Annexe 12.2) dans l'intervalle $[-600, 600]$ pour comparer le krigeage ordinaire (version 0.14 [159]), le modèle KPLS et le modèle KPLS+K dans le Tableau 2.2. On dispose dans ce cas de 300 points d'apprentissage de type LHS et 5000 points de validation pour calculer l'erreur relative ER .

Métamodèles	ER (%)	temps CPU
krigeage	0.65	2894.56 s
KPLS1	0.79	0.86 s
KPLS2	0.74	1.85 s
KPLS3	0.70	20.01 s
KPLS1+K	0.66	22.00 s
KPLS2+K	0.60	23.03 s
KPLS3+K	0.61	41.13 s

TABLE 2.2 – Erreur de prédiction de la fonction Griewank en dimension 60, avec un plan d’expériences de 100 points entre le krigeage ordinaire de scikit learn (v0.14), les méthodes KPLS et KPLS+K. Le nombre de composantes principales retenues pour KPLS varie de 1 à 3 [40].

L’amélioration à la fois de l’erreur et des temps de calcul est très nette sur ces deux exemples analytiques.

Ces travaux ont fait l’objet de plusieurs conférences [44, 45], de rapports [41, 42], d’un dépôt de brevet (SNECMA/ISAE/ONERA en juillet 2015 [156]) et de deux articles [47, 46]. Depuis 2014, les modèles KPLS et KPLS-K sont disponibles en interne et depuis 2017 dans la toolbox opensource SMT.

Ce travail sur les modèles de krigeage est présenté à partir des deux articles [47, 46] dans les Sections 4.4.2 et 4.4.3.

Bilan : 2 articles, un brevet, 2 conférences, 1 manuscrit de thèse, 2 rapports ONERA, une librairie python *kpls* disponible en interne ou dans la toolbox opensource SMT.

2.2.3.5 Des modèles dynamiques

En parallèle de ces activités sur les modèles dits statiques, j’ai participé activement au co-encadrement de la thèse de M. Gruet (CNES/ONERA-DESP, octobre 2015-2018 [96]) sur la modélisation des indices magnétiques pour la prédiction des événements solaires. L’intensité de l’activité magnétique est caractérisée par des réseaux de neurones dynamiques pour lesquels la notion de temps est introduite avec une récurrence. Différents types de réseaux de neurones ont été utilisés et comparés pour prédire l’indice magnétique global appelé *am* à partir des données du vent solaire. Les réseaux dit ”classiques” comme le perceptron multicouches ont été comparés au réseau à retard de temps (TDNN pour Time Delay Neural

Network) et aux réseaux récurrents NARX (Nonlinear AutoRegressive model with eXogenous inputs) et LSTM (pour Long Short Term Memory). Les prédictions de *am* obtenues à 1h et 3h de temps ont montré des avancées très prometteuses dans le domaine de la météo de l'espace. Ma contribution dans ce travail de thèse a été d'assurer le suivi d'un point de vue mathématique (analyse de sensibilité, réseau de neurones statiques et dynamiques) pendant les trois ans. Les résultats ont été présentés dans des conférences [98, 100, 99, 148] et un article est en cours de révision [97].

Bilan : 4 conférences, 1 manuscrit de thèse, 1 article en cours de révision.

L'ensemble des développements menés autour des modèles de substitution statistiques est capitalisé au sein de la toolbox interne *moe* développée en python [169, 93] et dans la toolbox SMT [194] décrite dans un article récent [49]. Cette capitalisation est détaillée dans la Section 3. Une formation autour des métamodèles est également proposée en interne et en externe comme détaillée dans la Section 9.2.

2.2.4 Introduire des modèles de fidélité variée

Nous nous intéressons ici à la problématique de construction d'un modèle réduit dans un contexte où plusieurs niveaux de précision (ou niveaux de fidélité) sont disponibles. On suppose par exemple que l'on dispose d'une méthode de résolution analytique formulée sous certaines hypothèses, d'une méthode semi-analytique, d'un code de calcul éléments finis plus précis mais lent à calculer,... comme introduit dans la Section 2.2.2. Le co-krigeage est une méthode qui permet d'approcher une fonction à partir de plusieurs jeux de données différents mais corrélés entre eux. L'idée du co-krigeage avec données multi-fidélité consiste à apprendre la relation entre les deux niveaux de code pour pouvoir exploiter au mieux les données basse fidélité, plus rapides à calculer. L'enjeu principal consiste au développement et à l'intégration en Python d'une méthode de construction de modèles réduits multi-fidélité. En particulier, la formulation récursive du co-krigeage multi-fidélité, décrite dans les travaux récents de [125] et qui présente à la fois des avantages en termes de vitesse de calcul, de lisibilité et d'adaptabilité, a été étudiée. Cette méthode a été enrichie en y ajoutant la possibilité d'estimer un bruit d'observation et en adaptant une technique combinant krigeage et PLS appelée KPLS et KPLS+K pour traiter le cas de la grande dimension. Plusieurs techniques de construction de plans d'expériences multi-fidélité ont été développées [197]. Les premiers travaux sur ce sujet ont été réalisés lors du stage de fin d'études de R. Vauclin (co-encadrement ISAE-SUPAERO et ONERA) en 2014. La première

contribution concerne la construction des plans d'expériences imbriqués, hypothèse sur laquelle repose la formulation récursive de L. Le Gratiet [125]. Dans ses travaux, L. Le Gratiet reprend le concept du modèle de co-krigeage [117] pour un nombre quelconque de niveaux de fidélité et montre qu'il est possible de calculer la moyenne et la variance de krigeage de façon récursive. Ainsi, non seulement les paramètres de chacun des niveaux de codes peuvent être estimés indépendamment, mais en plus il n'est plus nécessaire de reconstruire une matrice de covariance de grande taille (liée aux nombres de points par niveau de code) qui devra être inversée dans la boucle d'optimisation. Construire un modèle de krigeage avec s niveaux de fidélité devient aussi coûteux en temps que construire s modèles séparément. La formulation du co-krigeage définie ci-dessus requiert l'imbrication des plans d'expériences de chacun des niveaux de code. Une nouvelle stratégie a donc été introduite en n'effectuant des permutations de l'algorithme ESE [111] qu'entre des points d'un même niveau de fidélité, ce qui diffère de la littérature dans le sens où l'optimisation du plan se fait niveau après niveau. En termes de temps de calcul, le remplissage des plans niveau par niveau est aussi rapide, tout en étant plus performant. Les indicateurs sont meilleurs dans tous les cas, en particulier en petite dimension où ils s'approchent même de la qualité d'un plan classique. La procédure qui a été développée n'est réalisable que s'il y a au moins k fois plus de points Basse Fidélité que de points Haute Fidélité (k entier). Deux approches pour construire cette imbrication ont été comparées : la recherche des plus proches voisins ainsi que le remplissage du LHD avec une optimisation niveau par niveau par l'algorithme ESE. La Figure 2.8 donne un aperçu de plans 2D à trois niveaux construits avec la méthode par plus proche voisin (Figure 2.8a) ou la méthode par niveau (Figure 2.8b). Le module de co-krigeage développé en Python (toolbox *mfk*) a été intégré à la version 0.12 d'OpenMDAO (ce qui a fait l'objet d'un fait marquant ONERA en 2015) et plus récemment à la toolbox SMT (ce qui a fait l'objet d'un fait marquant ONERA en 2017). Plus de détails sont donnés dans les Sections 3.1 et 3.2. En interne ONERA, la toolbox a été utilisée au DTIS (anciennement DCPS/SSD) pour des applications lanceurs et au DEMR en collaboration avec P. Borderies pour des applications en problèmes inverses. Plus récemment en 2017, dans la thèse de M.G. Hernández-Godino [84], une collaboration a été mise en place avec le professeur R. T. Haftka du Dept. of Mechanical and Aerospace Engineering de l'Université de Floride. Nous construisons des métamodèles multi-fidélité pour prédire l'évolution de particules dans le cadre de détonations cylindriques multiphasés (gaz et particules) avec des maillages pouvant être très denses (de l'ordre de 500 000 cellules) et les temps de calcul du code de simulation très coûteux (plusieurs heures à plusieurs jours). Des comparaisons sur l'amplification des particules ont été effectuées entre les modèles de co-krigeage

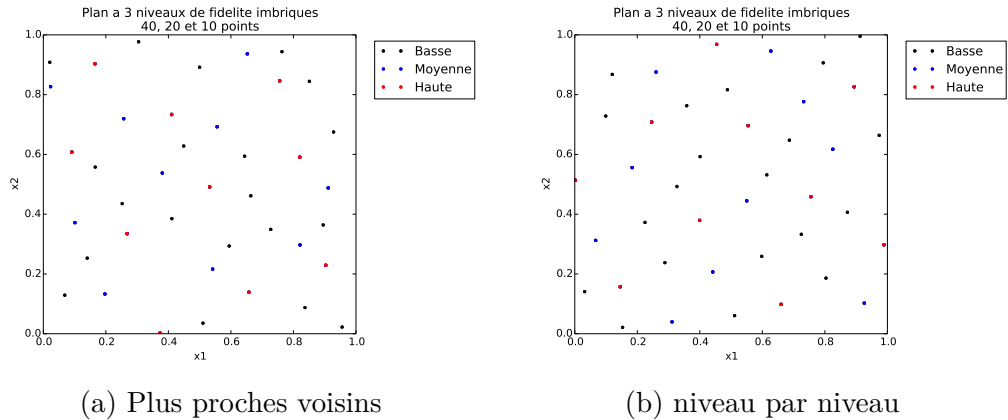


FIGURE 2.8 – Plans d’expériences 2D à trois niveaux de modèle avec respectivement 10, 20 et 30 points.

de SMT [86] et des modèles polynomiaux multifidélité basés sur une régression au premier ordre [204]. Un article a été soumis récemment (mars 2019) dans le journal *Structural and Multidisciplinary Optimization* [85].

Bilan : 2 rapports de stage, 1 article soumis, 1 toolbox *mfk* disponible en python en interne et en externe via SMT et intégrée à la dernière version d’OpenM-DAO, 1 collaboration avec la NASA-Glenn/Université du Michigan et 1 collaboration avec l’Université de Floride.

2.2.5 Disposer d’optimiseurs efficaces pour les boîtes noires

L’application visée concerne l’optimisation des nouveaux concepts d’aéronef pour lesquels l’utilisation de codes de calcul (haute fidélité) est rapidement indispensable mais dont l’appel direct reste trop coûteux. Les difficultés rencontrées sont de deux types : la dimensionnalité du problème d’une part en terme de nombre de variables de conception (quelques centaines) et de contraintes (quelques centaines) et d’autre part en terme de temps de calcul (plusieurs heures voir plusieurs jours pour l’obtention de la fonction objectif ou d’une contrainte). Le principe réside à construire des modèles de substitution à partir d’une base de données parcimonieuse. Deux possibilités sont envisageables :

- construire une fois pour toutes les métamodèles (objectif et contraintes) à partir d’une base de données avec un grand nombre de points (en pratique, la règle est souvent de prendre un minimum de 10 points par dimension de l’espace [132]) et d’utiliser un optimiseur qui appelle ces métamodèles

pour déterminer la solution optimale. Une autre alternative pour construire le métamodèle précis est d'utiliser un plan d'expériences adaptatif comme cela a été décrit dans la Section 2.2.3.2.

- construire une base de données avec très peu de points et enrichir cette base en ajoutant des points susceptibles d'être solution du problème. Cela nécessite la minimisation d'un critère (critère d'amélioration espérée, ...) basé sur un modèle de type krigeage et l'appel aux vrais codes de calcul à chaque itération pour enrichir la base de données avec la solution optimale potentielle. C'est la stratégie adoptée par l'optimisation bayésienne.

Suivant les cadres d'application, je me suis intéressée à ces deux stratégies qui dans les deux cas optimisent le métamodèle plutôt que la fonction objectif, afin de réaliser d'importantes économies en temps de calcul. **C'est sur cette thématique que ma contribution est la plus importante via la méthodologie développée, les co-encadrements de stages et de thèses sur le sujet et les collaborations mises en place.**

2.2.5.1 Optimisation directe basée sur des métamodèles

L'optimisation basée sur les métamodèles encore appelée SBO pour Surrogate Based Optimization est souvent utilisée dès lors que l'on a des boîtes noires coûteuses à optimiser pour lesquelles aucune information de régularité (convexité, linéarité, ...) n'est disponible. On parle alors de Black Box Optimization (BBO, voir le Chapitre 1 de [11]) pour lesquelles les dérivées ne sont pas accessibles (Derivative Free Optimization - DFO). On remplace la ou les fonctions coûteuses par un métamodèle qui sera appelé par l'algorithme d'optimisation. Cette étape de substitution peut être gérée par l'utilisateur qui choisit le meilleur métamodèle adapté à son problème ou directement par l'algorithme d'optimisation qui construira des directions de recherche (Coordinate Search, Pattern Search, Mesh Adaptive Direct Search). Ces derniers algorithmes possèdent des preuves de convergence intéressantes [11] dès lors que la fonction à minimiser possède des propriétés mathématiques de régularité. Des méthodes basées sur les modèles remplacent les fonctions coûteuses par des approximations linéaires ou quadratiques couplées à des régions de confiance pour la mise à jour des approximations (Chapitre 11 dans [11]). L'utilisation de métamodèles plus généraux comme les fonctions à base radiale [172], les réseaux de neurones [119, 120] ou les processus gaussiens [112] permet de traiter des problèmes plus complexes d'optimisation globale sans aucune connaissance a priori des fonctions à évaluer. La technique des mélanges d'experts décrite précédemment pour approcher des fonctions discontinues ou à

dérivées discontinues a été utilisée dans la thèse de D. Bettebghor pour modéliser les facteurs de flambage (contraintes mécaniques très coûteuses) et ainsi optimiser des panneaux de fuselage avec un temps acceptable [31].

Cette technique originale a également été utilisée afin de prendre en compte des variables discrètes dans un processus d'optimisation dans le cadre du projet interne PRF CREATION. Il s'agissait de concevoir un hélicoptère de transport pour 90 passagers (HO-90) en minimisant 3 objectifs (masse à vide, consommation de carburant et critère acoustique en phase d'approche) en fonction de 4 paramètres liés au rotor principal : nombre de pales, rayon, corde moyenne des pales et vitesse de rotation. L'utilisation des mélanges d'experts permet d'obtenir un modèle réduit global pour chacun des objectifs en fonction des 4 variables de conception. L'optimisation basée sur les modèles réduits peut être alors faite via un algorithme déterministe à base de gradient. En effet, la plupart des modèles réduits sont dérivables analytiquement et ainsi les dérivées analytiques sont directement utilisables par les algorithmes basés sur le gradient. Si ce n'est pas le cas, la méthode des différences finis ou du complex step [137] sur le métamodèle peut être utilisée pour obtenir cette information à un coût raisonnable.

Cette stratégie a été validée via une comparaison avec une approche par algorithmes génétiques et les temps de calcul sont nettement inférieurs, ce qui reste très intéressant en phase de pré-dimensionnement. Elle a été présentée en 2014 à la conférence OPTI2014 [27]. Cette approche est très utile dès lors que la base de données est très riche permettant de construire un métamodèle précis ou bien dans le cas où le métamodèle est déjà existant, par exemple fourni par un industriel pour qui il est plus facile de donner un métamodèle que le vrai code de calcul.

Bilan : 1 chapitre de thèse, 1 article de conférence

2.2.5.2 Optimisation adaptative basée sur des métamodèles

La connaissance de l'incertitude dans un modèle de krigeage est un atout majeur utilisée dans l'optimisation bayésienne. On peut ainsi définir les zones intéressantes comme un compromis entre la valeur du meilleur prédicteur et l'incertitude associée au modèle. L'optimisation bayésienne (BO) est une stratégie d'optimisation adaptative qui repose sur deux principes : un modèle statistique bayésien pour représenter la fonction objectif (un processus gaussien par exemple) et une fonction d'acquisition pour décider du prochain point à échantillonner. Le tutoriel récent proposé dans [90] permet d'avoir une revue des différentes fonc-

tions d'acquisition (ou encore appelées critères d'enrichissement). L'optimisation bayésienne est définie par J. Moćkus [149] comme une technique d'optimisation basée sur la minimisation de l'écart attendu à l'extrémité de la fonction étudiée. Elle a été largement étudiée et utilisée à partir de 1998 avec l'algorithme EGO (Efficient Global Optimization proposé dans [112]). EGO combine les deux informations fournies par le GP (prédiction et incertitude) pour minimiser la fonction objectif. Il est basé sur un critère d'enrichissement (le plus connu est le critère EI pour Expected Improvement ou amélioration espérée), qui permet de faire un compromis entre l'exploration du domaine dans les zones où le modèle a une variance élevée et l'exploitation dans les zones où les prédictions du modèle sont les plus petites. Dans le cas du krigeage avec données déterministes (non bruitées), le critère est nul aux points d'observation et donc permet d'assurer la convergence. De nombreux auteurs modifient le critère EI selon le problème, notamment en grande dimension (critère WB2 par exemple [202] et utilisé dans [178]). Récemment nous avons proposé un nouveau critère notée WB2S qui permet de faire une mise à l'échelle des quantités impliquées. Pour accélérer le temps de calcul de l'algorithme EGO, notamment en grande dimension, nous avons remplacé le modèle de krigeage classique par des modèles KPLS [47] et/ou KPLS-K [46]. Nous avons également adapté l'algorithme à la prise en compte de contraintes en s'inspirant de l'algorithme SuperEGO proposé par M. Sasena dans sa thèse [178]. Les résultats de ce couplage appelé SEGOKPLS ont permis de montrer un gain en terme de temps CPU considérable sur des fonctions académiques et sur une application industrielle aux aubes de turbomachines avec 50 variables de conception et 31 contraintes aérodynamiques et géométriques à respecter [43, 40]. Des comparaisons avec des algorithmes existants ont été faites dans le cadre d'une collaboration avec l'université de St Joseph à Philadelphie (USA). Un article a été publié [48] et est présenté dans la Section 4.4.5. L'application au cas test industriel SNECMA pour l'optimisation des aubes de turbomachines a fait l'objet d'un brevet accepté en 2017 [156].

Suite à ces premières résultats très encourageants fournis par l'algorithme SEGOKPLS, nous avons étendu l'algorithme aux mélanges d'experts. La méthode dénommée SEGOMOE a été testée sur de nombreux cas tests depuis février 2016 au travers de stages et de projets étudiants [169, 191] : un cas test automobile en dimension 30 avec 68 contraintes (cas test MOPTA08 [9, 113]), une optimisation de structure pour une aile d'avion en dimension 12 [21] et une optimisation aérodynamique de l'aile CRM [196, 134] a été réalisée et présentée à l'AIAA/ISSMO [22] en 2017. Ce dernier cas test fait l'objet d'une collaboration avec le MDOLab car leur code CFD ADFlow est utilisé. Chacune de ces nouvelles

applications amène une nouvelle difficulté qu’il faut traiter : le grand nombre de contraintes (750 contraintes dans le cas de l’aile CRM), la grande dimension pour le cas test MOPTA08 (possibilité de considérer 124 variables de conception). La méthode des sous-espaces actifs [63, 200] est une piste regardée dans le cadre de la thèse de R. Priem pour la réduction de dimension.

Récemment, nous avons comparé SEGOMOE avec des méthodes DFO comme COBYLA [168] (interpolation linéaire de la fonction objectif et des contraintes couplée à une région de confiance), NOMAD [10] (construction d’un maillage adaptatif pour trouver des directions de descente), ALBO [166] (méthode du lagrangien augmenté [154] couplée à des variables d’écart pour traiter les contraintes mixtes) ou encore SUR [163] (fonction d’acquisition basée sur le critère EI modifiée pour prendre en compte les contraintes). La méthode SEGOMOE permet d’obtenir le minimum global sur des cas tests analytiques ou plus complexes, contraints ou non (de la dimension 2 à la dimension 17) avec un budget de calcul très réduit [23].

Dans le cadre du projet européen H2020 AGILE, l’algorithme SEGOMOE a été intégré à RCE (outil collaboratif du DLR) pour proposer quelques itérations pour optimiser une aile d’avion en utilisant des codes de calcul distribués un peu partout en Europe (Université de Naples, DLR, TsAGI, PoliTo). Cette application a été présentée à la conférence EASN en octobre 2016 [126]. Toujours dans le cadre de ce projet AGILE, j’ai utilisé SEGOMOE pour optimiser une nacelle d’avion en utilisant un code CFD du TsAGI à partir des 17 variables de conception. Les résultats ont été comparés à l’algorithme EGO de la librairie DAKOTA [6] et SEGOMOE montre de très bonnes performances en minimisant le nombre d’évaluations du code de calcul et en trouvant un meilleur optimum dans un cas non contraint [128, 127] ou contraint [24].

Depuis octobre 2017, la thèse de R. Priem en collaboration ONERA-ISAE a démarré pour améliorer la prise en compte des contraintes. L’idée est considérer l’erreur commise sur le krigeage associé à chacune des contraintes (égalité et inégalité) en adaptant la fonction d’acquisition via un critère de confiance (Upper Trust Bounds). Des comparaisons ont pu être menées sur différents cas analytiques [171, 170]. Cette démarche a été comparée sur le problème bi-dimensionnel de Branin modifié (cf définition dans l’Annexe 12.3) connu pour avoir deux minima locaux et un minimum global. La méthode utilisant le facteur de confiance (SEGO UTB) trouve dans 100% des cas le minimum global contre 65% des cas pour le SEGO classique (SEGO MC) et 60% pour la méthode ALBO [166] basée sur le lagrangien augmenté comme le montre la Figure 2.9. Pour augmenter la complexité des problèmes considérés en terme de nombre de variables et nombre de contraintes, une collaboration est prévue avec Bombardier et le GERAD via le

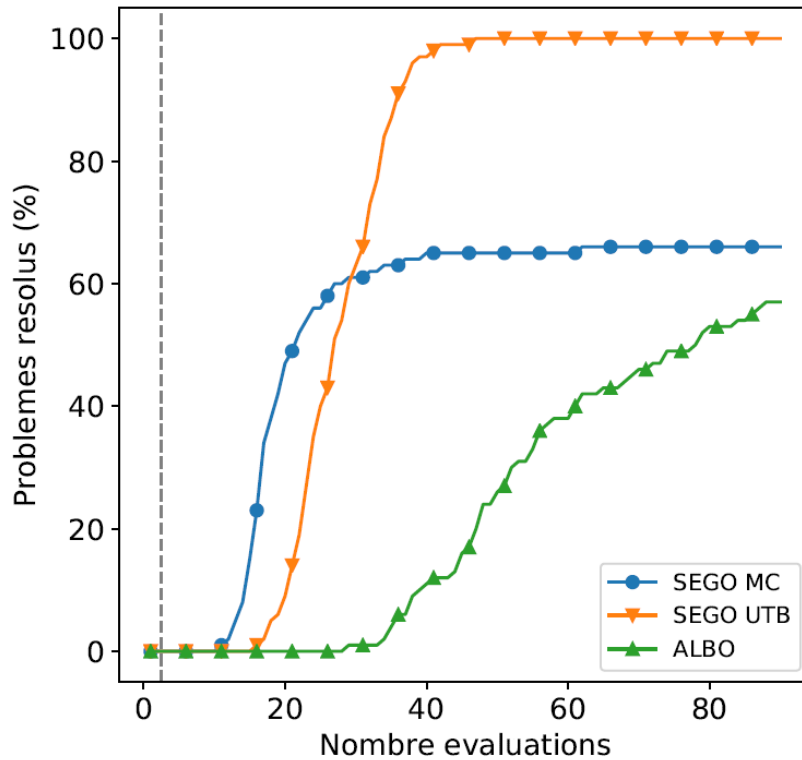


FIGURE 2.9 – Pourcentage de problèmes Branin modifié résolu en fonction du nombre d’évaluations [170].

séjour de R. Priem à Montréal de mai à septembre 2019.

Dans le cadre de la conception avion avec le code FAST (cf Section 2.2.1), une première campagne d’optimisation multiobjectif a été menée en utilisant SEGO-MOE via une fonction composite entre la masse au décollage et la masse de fuel dans le cadre de la thèse de P. Schmollgruber, ce qui a permis de valider les fronts de Pareto avec une approche par algorithmes génétiques [66] et de réduire le temps de calcul par un facteur 5 (1100 appels au lieu de 10000, soit 22h au lieu de 110h). De nouveaux calculs sont en cours dans le cadre de la thèse de A. Sgueglia pour l’optimisation de l’avion à propulsion hybride.

Ce travail sur l’optimisation adaptative basée sur les métamodèles est présenté à partir des deux articles [48, 23] dans les Sections 4.4.5 et 4.4.6.

Bilan : 1 chapitre de thèse, 2 articles acceptés, 3 articles à la conférence AIAA, des collaborations académiques (ISAE-SUPAERO, DLR, Université du Michigan, Université de Philadelphie ...) et industrielles (TsAGi, BOMBARDIER).

2.2.5.3 Amélioration du conditionnement de la matrice du krigeage au cours des enrichissements

Du fait de l'enrichissement de la base de données au cours des itérations, des problèmes de conditionnement de la matrice de covariance du krigeage peuvent apparaître dès lors que deux points sont "trop proches" dans le plan d'expériences (2 colonnes très proches), en particulier en grande dimension. Des travaux plus amont sont donc menés pour essayer de résoudre les problèmes liés au conditionnement de la matrice de covariance du krigeage lorsque deux points sont très proches. Un projet étudiant (été 2016) et le stage de M. Gibert (septembre 2016 - février 2017) ont permis de résoudre cette difficulté en ajoutant un préconditionneur ou bien en proposant des hyperparamètres du krigeage mieux adaptés [93] avec des bornes supérieures et inférieures adaptatives pour évoluer au cours des itérations de SEGOMOE. Ce travail a été mené en collaboration avec Y. Diouane et M. Salaün de l'ISAE-SUPAERO du département mathématiques (DISC).

Bilan : 1 rapport de stage, 1 collaboration avec le département de mathématiques de l'ISAE.

2.2.5.4 Optimisation multi-fidélité

Dès lors que plusieurs fidélités de modèle sont disponibles, il paraît intéressant d'utiliser et de coupler ces différentes sources d'information dans le seul but de réduire le nombre d'évaluations au code de calcul le plus coûteux [167]. Différentes approches existent dans la littérature [124, 160] pour essayer d'accélérer la convergence du processus haute fidélité en utilisant les informations des fidélités inférieures. De fait de notre expérience des algorithmes d'optimisation adaptative basée sur les métamodèles (cf Section 2.2.5.2), nous avons étendu l'algorithme SEGOMOE aux données multi-fidélité avec une approche en deux étapes : i) choix du point à explorer en maximisant le critère Expected Improvement (ou un critère dérivé comme WB2 ou WB2S) et ii) choix du niveau de fidélité associé à ce point. L'idée de l'approche développée et appelée MFEGO est de favoriser l'exploration de zones inconnues avec les modèles peu coûteux et de n'appeler le modèle haute fidélité que lors de la phase d'exploitation pour traiter des problèmes d'optimisation avec ou sans contraintes. Sur l'exemple analytique 1-D issu de [88], nous avons comparé

une approche monofidélité et une approche à deux niveaux de fidélité (LF et HF). L'optimum est trouvé après 7 itérations (cf Figure 2.10b) : 7 points LF ont été ajoutés et 2 points HF à partir d'un plan d'expériences avec 6 points LF et 3 points HF (voir Figure 2.10a). Une approche EGO basée sur des calculs HF seulement aurait nécessité à partir d'un plan d'expériences initial de 4 points HF, l'ajout de 11 points HF pour trouver le minimum global, soit un total de 15 points HF. Au final, le coût de l'approche multifidélité MFEGO est 3 fois inférieur : 5.013 au lieu de 15 en supposant que le ratio de coût est de 1/1000 entre un point LF et un point HF.

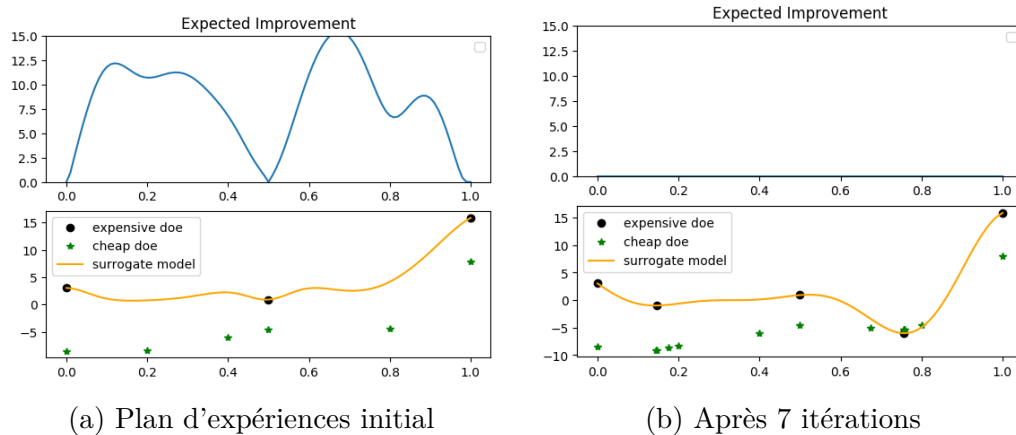


FIGURE 2.10 – Evolution du critère EI et des points d'enrichissement choisis (LF ou HF) au cours des itérations sur un exemple analytique 1-D [88] avec la stratégie MFEGO [146].

Ce travail a donné lieu à un notebook python disponible sous [github-MFEGO](#). Il a été étendu à un cas test aérodynamique via l'utilisation de deux codes de fidélité différente XFOIL [70] et ADflow [135] via une collaboration avec le MDOLab. Durant son stage de fin d'études entre l'ISAE et l'ONERA, M. Meliani a fait un séjour de 3 mois au MDOLab (juillet à septembre 2019) via une bourse de la Fondation ISAE-SUPAERO. Les premiers résultats ont permis de montrer l'intérêt de cette approche MFEGO sur des problèmes aérodynamiques avec ou sans contraintes jusqu'à 15 variables de conception. Le nombre d'appels à la fonction HF (code CFD ADflow) est deux fois moins important que la version EGO mono-fidélité et comparable à une méthode basée sur le gradient qui utilise l'adjoint.

L'extension à plus de variables est en cours via le stage de A. Diaz-Bolivar (mars à juillet 2019, co-encadrement avec T. Lefebvre) sur des exemples aérodynamique et/ou structure en collaboration avec l'ISAE-SUPAERO et l'Université du Michi-

gan.

Bilan : 1 rapport de stage avec une sélection dans le cadre du concours [PEGA-SUS](#) pour représenter l'ISAE-SUPAERO en avril 2019, 1 notebook disponible sous [github](#), 1 collaboration avec l'ISAE-SUPAERO et le MDOLab, 1 article à l'AIAA 2019 et 1 présentation invitée lors de la première édition du [Multifidelity Modeling \(MFM\) Workshop](#) organisé par l'AIAA 2019 à Dallas, 1 article de journal en cours de rédaction.

2.2.6 Formulations MDO adaptées

De nombreuses formulations MDO existent dans la littérature [138] dans le but d'aider l'équipe de conception à formuler de manière adéquate le problème, à rationaliser le processus de conception et d'optimisation de manière à l'organiser de la manière la plus efficace possible (mono ou multi-niveau, respect des couplages à chaque itération ou à convergence...). Mon activité sur les formulations MDO a démarré dans le PRF DOOM et l'ANR OMD via le co-encadrement de la thèse de J. Clément (2005-2009) [56]. Une revue des formulations existantes a été faite avec des adaptations possibles suivant les cas d'application [54, 55]. Dans le cadre de la thèse de D. Bettebghor (2005-2008) [34], une nouvelle formulation de décomposition bi-niveau (Quasi-Separable Decomposition) a été appliquée au calcul de structures [176, 33] et à l'optimisation d'une partie du fuselage dans le cadre du projet FP7 MAAXIMUS [31, 195]. Dans le cadre de l'étude DPAC ARTEMIS (Advanced R&T Enablers for Multidisciplinary Integrated Systems) une approche multi-niveau innovante a été mise en place pour la conception avion. En effet, le processus développé visait à coupler un processus multidisciplinaire de niveau conceptuel (PAG pour Processus Avion Global), un processus bi-discipline aérodynamique-structure haute fidélité (PBD pour Processus Bi Disciplinaire) et un processus détaillé optimisant les panneaux composites de la voilure. Ainsi, le processus PAG s'appuyait sur des outils disciplinaires simplifiés, optimisant les paramètres globaux de l'avion par rapport à une mission de référence. Le processus PBD, quant à lui, appliquait une approche BLISS (formulation MDO pour Bilevel Integrated System Synthesis [189]) à l'optimisation aéro-structure de la voilure en s'appuyant sur des méthodes dites haute fidélité. Les investigations menées entre les différents départements impliqués (DCPS, DAAP et DADS) ont conduit à explorer différentes solutions assurant un couplage entre les deux niveaux. Mon implication dans le projet a été de suivre le couplage PAG-PBD en essayant de proposer un coefficient d'échange, basé sur l'analyse post-optimale et les multipli-

cateurs de Lagrange. Ce travail a donné lieu à deux rapports techniques [144, 145] et de nombreuses présentations.

Une approche similaire est en cours dans le projet MDA-MDO de l'IRT Saint-Exupéry pour la mise en place d'une formulation bi-niveau au sein d'un processus industriel (outils et contraintes AIRBUS) avec la prise en compte des disciplines aérodynamique, structure et avant-projet. Dans le cadre de mon détachement à l'IRT (un jour par semaine entre 2015 et 2018), j'ai pu participer à la mise en place d'une nouvelle formulation [19, 92]. Nous avons mis en place une extension du processus d'optimisation aérodynamique multipoint à une meilleure prise en compte de la mission en couplant les appels au code elSA et au code avant-projet OCCAM. Ce couplage utilise des modèles chaînés avec dérivées pour lesquels nous avons utilisé les splines régularisées [109] développées dans la toolbox SMT [49] couplées à des techniques de bootstrap pour des estimations d'erreur [82]. Le processus complet sera présenté à la conférence AIAA au mois de juin 2019 [91, 65]. Ce travail a fait l'objet de deux sujets de stage M. Meliani (5 mois en 2017) et B. Dabas (6 mois en 2018). Ce travail a été présenté à plusieurs reprises devant les membres du projet IRT (AIRBUS, ISAE, CERFACS, ALTRAN, SOGETI) et devant différents services d'AIRBUS. Les développements ont été faits dans la plateforme python GEMS de l'IRT.

Bilan : 7 conférences dont 3 avec des articles à la conférence AIAA/ISSMO, 5 rapports, 1 collaboration avec l'IRT Saint-Exupéry (détachement 1 jour par semaine entre 2015 et 2018).

2.2.7 Prise en compte des incertitudes

La prise en compte des incertitudes dans la conception, et plus particulièrement en optimisation multidisciplinaire, est primordiale afin de fournir aux instances décisionnaires non seulement une ou des configurations optimales mais aussi les intervalles de confiance associés, permettant ainsi une décision éclairée à risque maîtrisé sur le choix du concept à développer. Les incertitudes peuvent généralement être aléatoires (variabilité inhérente du système physique considéré et/ou de son environnement) ou épistémiques (dues à des méconnaissances et/ou des simplifications de modélisation). La gestion de ces deux types d'incertitudes fait appel à des outils spécifiques car ces incertitudes sont souvent modélisées sous des formalismes différents (probabilités, intervalles, etc.).

2.2.7.1 Gestion des incertitudes de modèles

Depuis octobre 2015 avec l'arrivée de S. Dubreuil en post-doc dans l'unité, puis en CDI depuis 2017, je me suis intéressée à la propagation des incertitudes de modèles dans le processus de conception modélisés par des champs aléatoires. Comme l'idée est de remplacer les codes de calcul coûteux par des métamodèles ou des codes simplifiés, une erreur dite "de modèle" est introduite et peut être quantifiée. En effet, si un métamodèle de type krigeage est utilisé, l'information de la variance fournit une estimation d'erreur intéressante que l'on va pouvoir propager en cas de couplage avec d'autres disciplines (résolution d'une analyse multidisciplinaire MDA). Si un modèle physique simplifié est fourni à partir par exemple d'une base de données, l'expert peut lui associer une information sur l'erreur. Pour quantifier l'incertitude de modèle dans un problème multidisciplinaire couplé (par exemple dans un couplage aéro-structure où des modèles simplifiés sont utilisés), nous avons mis en place une démarche probabiliste qui consiste à modéliser l'incertitude de modèle par des champs aléatoires sur l'espace des variables de couplage et de conception. La résolution du problème couplé est faite uniquement en utilisant ces métamodèles. Pour discrétiser les champs aléatoires modélisant les incertitudes des métamodèles, une représentation spectrale du champ aléatoire est utilisée via la décomposition de Karhunen-Loève [115, 133]. En supposant que chaque variable de couplage (par exemple le champ de déplacement ou les efforts) est une variable aléatoire d'ordre 2, elle peut être approchée par son développement sur le chaos d'Hermite. Nous avons proposé un algorithme de point fixe pour calculer les coefficients des polynômes du chaos et ainsi résoudre le système non linéaire. Après convergence nous obtenons une représentation fonctionnelle de la loi jointe des variables de couplage. Cette représentation peut ensuite être utilisée comme un métamodèle pour d'autres analyses (calcul de la loi de la fonction objectif avec par exemple le rayon d'action ou la consommation de carburant pour chaque valeur de la variable partagée) ou pour le post traitement (calcul des moments, sensibilité, ...). La méthode proposée a été appliquée à la conception d'un avion d'affaires supersonique (cas test SSBJ de Sobieski [7]) où le couplage entre les disciplines Structure, Aérodynamique et Propulsion se traduit par la résolution d'un système avec 6 variables de couplage. L'approche a été validée en comparant les lois jointes (nécessitant 8000 évaluations) à une approche de Monte-Carlo nécessitant 100000 évaluations. Ce travail a fait l'objet d'un article [72] et d'une présentation à la conférence ECCOMAS [71]. L'interpolation par krigeage a également été utilisée pour quantifier l'incertitude lors de l'interpolation de maillage dans un calcul fluide-structure partitionné. Ce travail a été présentée en mai 2017 [80] en lien avec la thèse de J. Mas Colomer.

Dans le but de faire de l'optimisation multidisciplinaire comme le montre la Figure 2.11, nous avons donc remplacé chaque modèle disciplinaire coûteux par un processus gaussien et chercher à caractériser la fonction objectif du problème.

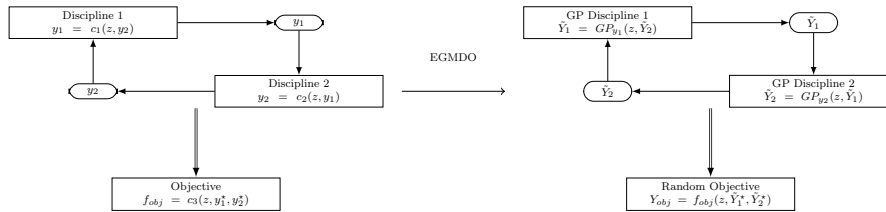


FIGURE 2.11 – Exemple de problème MDO avec deux disciplines couplées et une fonction objectif. Les modèles coûteux (à gauche) sont remplacés par des métamodèles GP (à droite) [72].

L'approche développée a permis de trouver la loi de la fonction objectif pour chaque valeur de la variable partagée en fonction des incertitudes sur les variables de couplage. L'idée étant de minimiser la fonction objectif, il faut trouver la loi de ses valeurs extrêmes. Dans notre cadre d'optimisation par enrichissement, nous sommes intéressés par la dispersion du minimum sur l'espace de conception, c'est à dire trouver les zones intéressantes de l'espace de conception pour enrichir nos métamodèles disciplinaires. Deux approches basées sur les polynômes du chaos, le krigeage et la décomposition de Karhunen-Loève ont été proposées et validées sur des cas tests gaussiens ou non. Ce travail a fait l'objet d'un article [73] et de quatre conférences [78, 79, 75, 77]. Une application dans le cadre du projet H2020 AGILE avec 4 disciplines et 14 variables (7 variables de conception et 7 variables de couplage) a été menée et présentée à l'AIAA 2018 [76] permettant de réduire d'un facteur 30 le nombre d'évaluations par rapport à une approche plus classique (MDF couplée à l'optimiseur SLSQP).

L'extension à la grande dimension en terme de variables de couplage (champ de pression ou de déplacement) est prévue dans la thèse de G. Berthelin (cofinancement ONERA-ISAE-SUPAERO, 2018-2021) où les premiers résultats sur la réduction de dimension ont été obtenus [28].

Bilan : 2 articles parus, 1 chapitre de livre en cours, 7 conférences dont 1 article à l'AIAA/ISSMO, 1 collaboration avec l'Institut Clément Ader et l'ISAE-SUPAERO.

2.2.7.2 Gestion des contraintes de fiabilité

Dès lors que l'on s'intéresse à la conception, il faut pouvoir définir le domaine de sûreté de fonctionnement. En aéronautique, il s'agit d'éviter la rupture d'une aile par exemple. Du fait de la variabilité que l'on peut avoir sur la pièce finalisée due à certaines hypothèses sur les propriétés matériaux par exemple, ou sur les dimensions géométriques de la pièce effectivement construite, il faut pouvoir identifier le domaine de défaillance via une analyse de fiabilité. La méthode classiquement utilisée pour calculer la probabilité de défaillance d'un système reste la méthode de Monte-Carlo qui nécessite la génération d'un grand nombre de points suivant la densité de probabilité associée aux incertitudes (loi gaussienne, loi uniforme,...) qui vont être évalués et classés (défaillant/non défaillant). De manière à réduire les temps de calcul pour des problèmes avec un nombre modéré de variables aléatoires, les méthodes par apprentissage actif [81], connues aussi sous le nom de méthodes par échantillonnage adaptatif, consistent à construire un modèle de substitution (ou métamodèle) de la fonction de performance et enrichir de manière adaptative ce métamodèle. D'autre part, dans le cadre de modèles haute fidélité fournis par exemple par la résolution d'équations aux dérivées partielles par la méthode des éléments finis, la réduction de modèles type base réduite (exemple Proper Orthogonal Decomposition) a également permis des gains importants en temps de calcul. Il existe un potentiel très fort d'obtenir des réductions drastiques des temps de calcul (de plusieurs ordres de grandeur) via une synergie entre les approches d'analyse de fiabilité par échantillonnage adaptatif et la modélisation en base réduite [147]. Ce gain sur le calcul de la contrainte de fiabilité sera d'autant plus important dans un processus d'optimisation robuste [153]. Ce travail est mené en collaboration avec l'Institut Clément Ader dans le cadre de la thèse de M. Menz.

Bilan : 1 article soumis, 3 conférences, 1 collaboration avec l'Institut Clément Ader.

2.2.8 Analyse de sensibilité

Avant de résoudre un problème d'optimisation ou définir un choix de modélisation, il convient de choisir les variables de conception. Ce choix reste une étape importante qui peut être aidée via une analyse de sensibilité de manière à sélectionner les variables les plus pertinentes au regard de la quantité d'intérêt. Cela peut aider à mieux sélectionner les paramètres du modèle ou à réduire la dimension du problème d'optimisation. Les méthodes d'analyse de sensibilité sont divisées en deux grandes classes : les méthodes locales et les méthodes globales [175]. Les

méthodes locales fournissent la pente de la sortie du modèle calculée dans l'espace des paramètres au voisinage d'un point. Les méthodes globales comme celles basées sur la décomposition de la variance de la sortie en fonction de la variance des paramètres d'entrée.

Dans le cadre des études liées à la propulsion hybride, l'introduction de nouvelles technologies comme les batteries avec une entrée en service prévue en 2035 fait apparaître des incertitudes sur la composition et les paramètres de ces batteries par exemple. Une analyse de sensibilité globale est donc intéressante afin de déterminer les paramètres importants qu'il est souhaitable de considérer lors de l'optimisation multidisciplinaire. Parmi toutes les méthodes disponibles [142, 190, 64], la méthode des indices de Sobol est utilisée [190], en raison de son efficacité en matière de calcul et de sa fiabilité dans la prise en compte d'un nombre plus élevé d'interactions entre variables [175]. Pour estimer ces indices de Sobol, nous avons utilisé les polynômes du chaos [192, 38, 74] pour construire une approximation de la quantité d'intérêt à partir d'un plan d'expériences. L'approximation par polynômes du chaos permet de calculer de manière analytique les indices de Sobol. Afin d'éviter la dépendance des résultats au plan d'expériences choisi, nous avons utilisé la technique du bootstrap [74] pour répéter l'opération et ainsi obtenir une valeur moyenne et une variance pour chacun de ces indices de Sobol.

Cette analyse de sensibilité a été menée dans le cadre de la thèse de A. Sgueglia en collaboration avec S. Dubreuil (ONERA/DTIS) et un chapitre de livre a été soumis [185].

2.3 Mon projet de recherche à l'ONERA

Mon projet de recherche s'inscrit directement dans le défi 1 "L'optimisation pluridisciplinaire et évaluation" du plan stratégique scientifique¹ de l'ONERA pour développer de nouveaux outils d'optimisation multidisciplinaire nécessaires à l'évaluation de nouveaux concepts aérospatiaux. La MDO se trouve au carrefour de nombreuses disciplines (modèles de substitution, algorithmique d'optimisation, robustesse aux incertitudes, etc.). L'un des enjeux est, à travers l'utilisation d'outils et formulations adaptés, d'élargir le champ des possibles de manière à trouver des configurations et concepts innovants, tout en essayant de maîtriser l'incertitude associée et le coût calculatoire inhérent aux nombreux appels de modèles au temps d'exécution parfois long. Bien que la MDO ait atteint un niveau de maturité

1. <https://www.onera.fr/fr/pss>

reconnu, de nombreux défis restent à relever. On pourra citer par exemple la gestion des incertitudes dans le processus de conception, la capacité à gérer différents concepts structurels discrets dans un même problème d'optimisation, la gestion des couplages disciplinaires à très grande échelle ainsi que la prise en compte de différents niveaux de fidélité. La grande dimension rajoute une complexité non négligeable à cet ensemble.

Je souhaite donc poursuivre mon activité dans les différents domaines qui ont été présentés et également élargir le spectre en considérant le couple aéronef/opérations au travers de la co-conception. Ce thème porté par la Fédération ONERA-ISA-ENAC va se développer dans les prochaines années de manière à considérer l'aéronef dans son ensemble en prenant en compte les opérations liées aux systèmes de trafic aérien, les exigences environnementales et sociétales. Toutes les composantes autour de l'ATM (Air Traffic Management) avec les opérations sol, l'insertion dans le trafic, le respect des trajectoires 4D,... pourraient être prises en compte si des modèles pour les décrire sont disponibles. Des premiers travaux ont été menés au MDOLab pour la conception de nouveaux aéronefs [107, 108, 106] en prenant l'allocation des missions en fonction des flottes d'avions existantes et de la demande des compagnies aériennes. Je souhaite m'investir dans ce projet autour de la Fédération qui va lier les méthodes autour de la MDO et des techniques d'ingénierie collaborative comme les outils MBSE (Model Based System Engineering) afin de construire une approche globale aéronef/opérations.

Par "co-conception", ou encore conception intégrée, on entend également le couplage de l'avion et des systèmes. En effet, les formulations MDO s'appliquent à la conception d'aéronefs autour des disciplines dites métiers (aérodynamique, structure, propulsion, acoustique,...) mais pourraient être étendues avec la prise en compte des systèmes embarqués. Différents projets de co-conception sont en discussion à échéance plus ou moins longue. Je pense notamment au Projet structurant Drone en cours de finalisation autour du "drone du futur" avec une co-conception intégrée des véhicules et des systèmes avioniques. La co-conception enfin peut se décliner autour du lien avec la fabrication, la maintenance et la certification pour modéliser et optimiser l'avion mais également sa chaîne de production dès les phases amont de conception. Je prévois donc de m'investir autour de ces activités dans le cadre du projet H2020 AGILE 4.0 qui doit démarrer en septembre 2019 pour une durée de 3 ans.

Chapitre 3

Développements de bibliothèques informatiques

Sommaire

3.1	Développement de bibliothèques en interne	46
3.2	Développement de bibliothèques en open source	47

Ce chapitre présente les différents développements informatiques réalisés en interne et en externe pour la communauté scientifique.

3.1 Développement de bibliothèques en interne

Au cours des différents encadrements de stages, thèses et post-docs, des modules informatiques autour des modèles réduits et de l'optimisation ont été développés. Initialement, ces développements ont été effectués sous Scilab dans le cadre de la thèse de J. Clément puis sous Matlab dans la thèse de D. Bettebghor. Le logiciel IMAGE dédié au mélange d'experts [15] a été recodé en Python en 2016 lors de stage du R. Priem [169] à partir de la toolbox scikit-learn couplé au développement de bibliothèques spécifiques. L'extension aux modèles de krigeage en grande dimension a été faite dans le cadre de la thèse de M.-A Bouhrel en 2016. En parallèle, le stage de R. Vauclin [197] a permis la gestion des données multifidélité. Une première version de l'algorithme d'optimisation a été faite en 2016 via le stage de R. Priem [169] puis poursuivie par les stages de V. Stiliz [191], de M. Gibert [93], de R. Olivanti [155] et de M. Meliani [146]. Depuis le départ, une collaboration avec R. Lafage (DTIS/MIDL) a été mise en place sur les aspects informatiques et capitalisation via l'encadrement des stages de C. Leroy (2015) et V. Fievez (2017) pour mettre en place les bibliothèques (gestion sous Redmine et svn). Cela a permis d'appliquer les bonnes pratiques de développement (nécessaires à la capitalisation du logiciel) notamment avec le respect de la charte des développeurs python PEP 8, la présence de tests de non régression et la documentation associée via des jupyter notebooks disponibles pour les utilisateurs. Cela a donné lieu aux bibliothèques internes en python suivantes :

- *doe* dédié aux plans d'expériences,
- *moe* dédié aux mélanges d'experts,
- *kpls* dédié aux modèles de krigeage en grande dimension,
- *mfk* dédié aux modèles de krigeage pour la multi-fidélité,
- *sego* dédié à l'optimisation par enrichissement qui utilise l'ensemble des bibliothèques citées au-dessus.

En interne ONERA, elles ont été utilisées au DCPS puis actuellement au DTIS pour différentes applications (CleanSky 2, SESAR, H2020 AGILE, PR MUFIN) et ont été distribuées au centre du Fauga (collaboration avec S. Mouton en 2016 autour des essais en soufflerie). Une capture du wiki interne mis en place pour la description des outils et algorithmes est donnée sur la Figure 3.1. Aujourd'hui ces développements sont réutilisés au sein de l'application web de conception What-sOpt [123] développée par R. Lafage (DTIS/MIDL). Cette application est pour l'instant disponible uniquement en interne mais des évolutions sont en cours pour la rendre disponible à l'extérieur.

OneraMDAO

Il s'agit des modules dont la source est disponible à l'Onera pour utilisation dans le cadre de l'analyse et l'optimisation multi-disciplinaires.

Listes des plugins Onera

Modèles de Substitution

Name	Description
KPLS	Krigeage combiné à une analyse en composantes principales pour problèmes de grande dimension
MFK	Krigeage multi-fidélité, LHS optimisé (maximinESE), DOE imbriqué
MOE	Mixture of experts based on Gaussian Mixture Model

Optimiseurs

Name	Description
CFSQP	Optimiseur pour problèmes avec contraintes non linéaires
CMAES	Optimiseur utilisant l'algorithme évolutionnaire CMA-ES
SEGOMOE	Optimiseur EGO avec contraintes basé sur les mélanges d'experts MOE

FIGURE 3.1 – Environnement ONERAMDAO sous Redmine.

3.2 Développement de bibliothèques en opensource

En 2013, nous avons contribué au framework OpenMDAO et ainsi collaboré avec la NASA autour des premières versions de la plateforme. Grâce à la collaboration avec R. Lafage, la toolbox *mfk* a été encapsulée à partir de la version 0.12 d'OpenMDAO et depuis fait partie intégrante des évolutions [2]. Plus récemment, en 2017 une collaboration a été mise en place avec l'Université du Michigan, la NASA, l'ISAE-SUPAERO et l'ONERA pour co-développer une toolbox python opensource dédiée aux modèles réduits. Elle contient des méthodes de construction de modèles de substitution, des techniques d'échantillonnage et des fonctions de test. Elle se veut à la fois simple à utiliser, mais aussi simple à enrichir avec de nouveaux algorithmes. Cette bibliothèque se distingue de celles existantes grâce à l'accent mis sur l'apprentissage des dérivées utilisées pour la construction des modèles GEK (Gradient Enhanced Kriging [39]). Elle inclut également des algorithmes innovants : le krigeage combiné aux moindres carrés partiels (modèles KPLS et KPLS-K permettant de traiter des problèmes de grande dimension) et l'interpolation par splines basée sur une minimisation d'énergie (modèles RMTS). SMT est documentée et distribuée sous la licence BSD. Elle peut être téléchargée via [github](#) et la documentation est disponible à l'adresse [doc](#). Un article décrivant les possibilités de SMT a été récemment publié [49]. Depuis la version 0.2, de nouveaux modèles ont été intégrés via le stage de V. Fievez pour *moe* et M. Meliani pour *mfk*. Une capture de la page d'accueil est donnée sur la Figure 3.2.

SMT 0.2 documentation »

AEROSPACE ENGINEERING
UNIVERSITY OF MICHIGAN

NASA

ISAE SUPAERO

ONERA
THE FRENCH AEROSPACE LAB

Table Of Contents

- SMT: Surrogate Modeling Toolbox
- Focus on derivatives
- Documentation contents
 - Indices and tables

Next topic

Getting started

This Page

Show Source

Quick search

SMT: Surrogate Modeling Toolbox

The surrogate modeling toolbox (SMT) is an open-source Python package consisting of libraries of surrogate modeling methods (e.g., radial basis functions, kriging), sampling methods, and benchmarking problems. SMT is designed to make it easy for developers to implement new surrogate models in a well-tested and well-document platform, and for users to have a library of surrogate modeling methods with which to use and compare methods.

The code is available open-source on [GitHub](#).

Focus on derivatives

SMT is meant to be a general library for surrogate modeling (also known as metamodeling, interpolation, and regression), but its distinguishing characteristic is its focus on derivatives, e.g., to be used for gradient-based optimization. A surrogate model can be represented mathematically as

$$y = f(\mathbf{x}, \mathbf{x}_t, \mathbf{y}_t),$$

where $\mathbf{x}_t \in \mathbb{R}^{n_t \times n_x}$ contains the training inputs, $\mathbf{y}_t \in \mathbb{R}^{n_t}$ contains the training outputs, $\mathbf{x} \in \mathbb{R}^{n_x}$ contains the prediction inputs, and $\mathbf{y} \in \mathbb{R}$ contains the prediction outputs. There are three types of derivatives of interest in SMT:

1. Derivatives (dy/dx): derivatives of predicted outputs with respect to the inputs at which the model is evaluated.
2. Training derivatives (dyt/dxt): derivatives of training outputs, given as part of the training data set, e.g., for gradient-enhanced kriging.
3. Output derivatives (dy/dyt): derivatives of predicted outputs with respect to training outputs, representing how the prediction changes if the training outputs change and the surrogate model is re-trained.

Not all surrogate modeling methods support or are required to support all three types of derivatives; all are optional.

FIGURE 3.2 – Toolbox SMT [49].

Cette toolbox SMT a été utilisée pour faire de l’optimisation multifidélité (couplage méthode EGO et SMT-MFK pour la méthode MFEGO, cf Section 2.2.5.4) et un notebook python réalisé dans le cadre du stage de M. Meliani (mars à octobre 2018) est actuellement disponible sous [github-MFEGO](#). Aujourd’hui SMT est utilisée à notre connaissance en interne ONERA dans le cadre du projet interne PR MUFIN sur la multifidélité par exemple, à l’IRT Saint-Exupéry (projet MDA-MDO [65]). Des comparaisons ont été effectuées dans le cadre d’une collaboration avec l’Université de Floride durant la thèse de M. G. Fernández-Godino [84] pour donner lieu à un article de conférence [86] et un article soumis en mars 2019 [85].

Dans le cadre de la formation Machine Learning, des notebooks python (Processus Gaussiens, Optimisation EGO par exemple) sont développées et distribuées aux stagiaires via le lien [github-ML](#).

Dans le cadre du Groupement d’Intérêt Scientifique ”Sizing lab” sur le dimensionnement collaboratif des systèmes technologiques pour la mobilité du futur, une présentation de la toolbox SMT et une formation via un notebook python doit avoir lieu en septembre 2019 afin de pouvoir partager des outils entre les différents partenaires (INSA-ICA-LAAS-ISAE-ICAM-ENAC-ONERA).

Chapitre 4

Contribution de recherche et sélection de publication

Sommaire

4.1	Développement de métamodèles	51
4.2	Optimisation basée sur les métamodèles	52
4.3	Capitalisation dans une librairie informatique open source	55
4.4	Sélection des publications	55
4.4.1	Surrogate modeling approximation using a mixture of experts based on EM joint estimation, Structural and Multidisciplinary Optimization 2011	56
4.4.2	Improving kriging surrogates of high-dimensional design models by Partial Least Squares dimension reduction, Structural and Multidisciplinary Optimization 2016	75
4.4.3	An Improved Approach for Estimating the Hyperparameters of the Kriging Model for High-Dimensional Problems through the Partial Least Squares Method, Mathematical Problems in Engineering 2016	94
4.4.4	Approximation of the critical buckling factor for composite panels, Structural and Multidisciplinary Optimization 2012	106
4.4.5	Efficient global optimization for high-dimensional constrained problems by using the Kriging models combined with the partial least squares method, Engineering Optimization 2018	131

4.4.6	Adaptive modeling strategy for constrained global optimization with application to aerodynamic wing design, Aerospace Science and Technology Journal 2019	158
4.4.7	A Python surrogate modeling framework with derivatives, Advances in Engineering Software 2019	177

Ce chapitre présente ma contribution de recherche autour des métamodèles et de l'optimisation basée sur ces modèles à partir des sept articles sélectionnés.

Chacun des sept articles présentés appartient à l'une des trois catégories suivantes : développement de métamodèles, optimisation basée sur les méta-modèles et développement de librairie informatique. Ces catégories ne sont clairement pas disjointes et permettent de voir le lien entre le traitement et l'optimisation des données jusqu'à la réalisation logicielle.

4.1 Développement de métamodèles

Trois articles traitent des développements autour des métamodèles dans lesquels j'ai contribué. Dans le premier article, l'idée était de développer des métamodèles adaptées aux non-linéarités des fonctions que l'on cherche à approcher, en l'occurrence initialement des changements de mode dans le calcul de flambage. Cette technique d'approximation de type apprentissage statistique consiste à diviser l'espace de design en sous-domaines et à construire dans chacun des sous-domaines le métamodèle (l'expert) le plus adapté. Elle est basée sur l'algorithme Expectation-Maximization qui estime les lois de probabilité pour effectuer le mélange des experts, et l'originalité de la méthode a été de considérer le couple (entrées, sortie) pour estimer ces lois de probabilité et proposer un modèle global pour la régression et/ou l'interpolation suivant le choix des experts. Deux possibilités pour recombinaison des experts ont été proposées suivant le modèle d'approximation qu'on cherche à construire :

- un modèle régulier (C^1) qui pourra être utilisé comme modèle dans un algorithme d'optimisation différentiable, c'est la recombinaison lisse ;
- un modèle discontinu basé sur une probabilité d'appartenance par maximum a posteriori et on parle de recombinaison discontinue.

Une procédure pour choisir le nombre d'experts basée sur différents critères liés à la complexité du modèle ou à la précision a été mise en place. La méthode, baptisée IMAGE (Improved Metamodeling Approximation through Gaussian mixture of Experts) a été mise en œuvre sur plusieurs types de métamodèles locaux (régression linéaire ou quadratique, réseaux de neurones, fonctions à bases radiales, moindre carrés mouvants), et validée à la fois sur des cas test analytiques et sur l'approximation du calcul de flambage d'une structure composite. Aucune limitation sur le type de métamodèles n'a été faite dans cette approche si ce n'est la dimension de l'espace des entrées.

Lorsque le nombre de variables d'entrée augmente, le choix du métamodèle s'avère souvent délicat. C'est la raison pour laquelle, on s'est intéressé au couplage de la méthode du krigeage et des moindres carrés partiels (PLS) en proposant

le deuxième article décrivant le modèle KPLS. L'idée est de réduire le nombre d'hyperparamètres nécessaire à la construction du modèle de krigeage, nombre qui croît linéairement avec la dimension du problème. La méthode de réduction dimensionnelle de PLS permet de réduire le nombre d'hyperparamètres à estimer en les projetant sur un nombre réduit de composantes principales (généralement, on retient entre 1 à 4 composantes principales). Il s'agit d'une variante de l'analyse en composantes principales (ACP). Les variables initiales sont projetées sur un nombre restreint de droites, orthogonales les unes aux autres et dont les directions, contrairement à l'ACP, dépendent non seulement des entrées mais également des sorties du modèle. La méthode KPLS proposée permet de limiter le nombre de paramètres internes du krigeage et d'améliorer de façon globale la capacité de prédiction du métamodèle par une meilleure résolution du problème de maximum de vraisemblance pour l'évaluation des hyperparamètres. Les résultats sur des fonctions analytiques ou sur des cas test issus des turbomachines ont permis de valider la méthode en terme de précision (erreur relative) et de gain en temps de calcul. En étudiant plus précisément les propriétés mathématiques des noyaux KPLS exponentiels, nous avons pu montrer l'équivalence avec des noyaux exponentiels classiques après un changement de paramètres.

Dans le troisième article, un modèle de krigeage KPLS+K est alors proposé avec une seconde maximisation de la vraisemblance qui est cette fois une maximisation locale qui utilise les paramètres issus du noyau KPLS pour initialiser l'optimisation. Ce nouveau modèle appelé KPLS+K revient à un krigeage classique avec un mode d'initialisation des hyperparamètres différent. De nouveaux tests sur les fonctions analytiques et sur un cas test industriel montrent une amélioration des qualités de prédiction du modèle KPLS+K par rapport au modèle KPLS, au prix d'un léger accroissement de temps CPU de la phase d'apprentissage. Les modèles KPLS et KPLS+K ont donc été naturellement rajoutés dans l'outil logiciel IMAGE, initialement codé en Matlab puis ensuite développé en python.

4.2 Optimisation basée sur les métamodèles

L'utilisation des métamodèles pour remplacer des fonctions objectifs et/ou contraintes coûteuses est assez naturelle, c'est ce que l'on appelle Surrogate Based Optimization ou encore SBO. Les trois articles suivants sont liés à cette thématique. Le quatrième article propose une approche en deux étapes pour traiter un problème d'optimisation mixte. L'optimisation de structures composites stratifiés mêle par définition les variables continues (épaisseur, redistribution des

efforts) et les variables discrètes. Pour pallier à la difficulté de ce type de problème MINLP (Mixed Integer Non Linear Programming), il est désormais courant d'utiliser des variables continues : les paramètres de stratification. Le principe général est donc d'effectuer une optimisation continue sur la base de ces variables et ensuite d'identifier les séquences d'empilement les plus proches (dans une métrique adaptée) dans l'espace des séquences d'empilement admissibles. Etant donnée la dimension des problèmes considérés, l'utilisation des modèles réduits pour remplacer les outils métier est une étape inévitable dans la boucle d'optimisation. Un travail théorique a permis de montrer que le facteur de flambage est concave en fonction de certains de ces paramètres de stratification, ce qui permet d'assurer que le problème de maximisation de ce facteur est bien posé. Nous avons étudié numériquement le comportement du facteur de flambage lorsque les paramètres de stratification et les flux varient, on s'aperçoit alors que le comportement typique est de dérivée discontinue par morceaux (fonction définie comme le minimum de plusieurs quantités). On a donc cherché une méthode d'approximation adaptée à ces contraintes de stabilité via la méthodologie de mélanges d'experts pour combiner différents modèles locaux adaptés à chacune des zones de l'espace. Les résultats de ce processus en deux étapes pour l'optimisation des empilements (optimisation des paramètres de stratification et post-identification), ont pu être comparés aux résultats obtenus après une optimisation par algorithme génétique menée sur des variables discrètes (angles d'orientation multiples de 45 degrés avec le logiciel GEOPS d'Airbus-D).

Dans le cinquième article, les modèles de krigeage KPLS sont intégrés à un algorithme d'optimisation avec contraintes. La méthode d'optimisation globale EGO pour Efficient Global Optimization est issue de l'optimisation Bayésienne qui consiste à combiner un modèle statistique Bayésien et une fonction d'acquisition dans un processus itératif. La méthode EGO proposée par Jones et al. [112] basée sur un modèle de krigeage et sur la fonction d'acquisition EI (pour Expected Improvement) qui mélange prédiction et incertitudes de prédiction fournies par le krigeage, ne permettait de résoudre que des problèmes non contraints. Sasena a proposé dans sa thèse [177] une évolution de EGO pour traiter les problèmes contraints appelée SEGO (pour Super Efficient Global Optimization) et a comparé différentes fonctions d'acquisition (EI, WB2, ...). Chaque contrainte du problème est approchée par un modèle de krigeage et le critère d'amélioration espérée EI ou son homologue WB2 qui est le progrès espéré moins le prédicteur de la fonction objectif reste à maximiser en satisfaisant les contraintes prédites. Nous avons étendu l'approche SEGO aux modèles KPLS et KPLS+K afin de traiter des problèmes d'optimisation en plus grande dimension. L'approche résultante appelée SEGOK-

PLS a été comparée à différents algorithmes sur des problèmes analytiques allant de la dimension 4 à 20 avec des contraintes mixtes d'égalité et d'inégalité. Dans un papier précédent de R. Regis datant de 2014, des comparaisons similaires avaient été faites entre l'algorithme COBRA (pour "Constrained Optimization By RAdial basis function interpolation") basé sur les fonctions à bases radiales et différents algorithmes disponibles. Nous avons donc contacté et collaboré avec Rommel Regis de l'Université de St Joseph à Philadelphie (US) afin de considérer les mêmes conditions expérimentales et comparer SEGOKPLS aux meilleurs résultats obtenus dans son papier de 2014 sur différents cas tests analytiques de la dimension 4 à la dimension 20. Le cas test MOPTA issu de l'industrie automobile proposé en 2008 et disponible en ligne via un exécutable permet de considérer une fonction objectif et 68 contraintes avec une dimension pouvant aller jusqu'à la dimension 128. En reprenant ce cas test utilisé dans le papier de R. Regis, nous avons considéré la dimension 12 et la dimension 50 avec 68 contraintes. Nous avons ainsi comparé l'algorithme COBRA, l'algorithme COBYLA et les méthodes SEGOKPLS et SEGOKPLS+K qui permettent d'obtenir le minimum global avec un nombre d'évaluations réduit.

Le sixième article permet d'étendre la méthode SEGO à des fonctions discontinues ou à dérivées discontinues en utilisant des mélanges d'experts basés sur les différentes formes de krigeages disponibles (krigeage, KPLS, KPLS+K). Le mélange de Gaussiennes restant une Gaussienne, les fonctions d'acquisition (EI, WB2) gardent donc une forme semblable et l'extension [23] possède donc les mêmes propriétés que la méthode SEGO. Pour traiter la grande dimension (> 10), un nouveau critère a été proposé en rajoutant un facteur d'échelle dans le critère WB2. En effet, cela permet au cours des itérations d'avoir un équilibre d'échelle entre la valeur du critère EI et la prédiction du modèle. La robustesse du nouveau critère a été validée sur des cas test analytiques connus pour leur multimodalité allant de la dimension 2 à 4. Des problèmes d'optimisation avec ou sans contraintes ont permis de comparer les performances de SEGOMOE à des algorithmes sans dérivée comme COBYLA et NOMAD. Une campagne de simulations sur un cas test aérodynamique en dimension 17 a également été menée en collaboration avec le MDOLab à l'Université du Michigan afin de valider la méthode SEGOMOE sur un problème d'optimisation connu pour être multimodal. L'algorithme SNOPT basé sur le gradient mais qui nécessite un point initial proche de la solution a été utilisé et le nombre d'évaluations comparé à SEGOMOE. Les résultats ont validé l'approche SEGOMOE qui permet de limiter le budget de calcul et de localiser le minimum global.

4.3 Capitalisation dans une librairie informatique open source

Le septième article décrit la toolbox qui a été développée en collaboration entre l'Université du Michigan, la NASA, l'ISAE-SUPAERO et l'ONERA. Elle est le fruit du travail mené depuis plusieurs années sur les métamodèles. Les premiers modèles de mélanges d'experts et de krigeage ont été développés en Matlab, puis recodés en python dans la librairie *moe* en interne ONERA. Ce travail a été mené par des stagiaires et thésards avec l'aide de Rémi Lafage (ONERA/DTIS) qui a une grande expertise dans le développement logiciel. Lors du départ de M.-A. Bouhleb à la fin de sa thèse pour un post-doc au MDOLab, l'idée est venue de capitaliser les développements de *moe*, *kpls* et *kpls+k* dans une toolbox opensource en partenariat entre les différents organismes. SMT for Surrogate Modeling Toolbox a donc démarré début 2017 et petit à petit des modèles et méthodes ont été ajoutés. La version 0.1 a été lancée en août 2017. SMT propose également des méthodes d'échantillonnage (DOE optimisé) et met l'accent sur la connaissance des dérivées. Elle inclut en particulier de nouveaux modèles réduits qui ne sont pas disponibles ailleurs : krigeage combiné aux moindres carrés partiels (KPLS et KPLS+K) et l'interpolation par spline basée sur une minimisation d'énergie (RMTS proposé par la NASA). La toolbox SMT est documentée et distribuée sous la licence New BSD de schéma et peut être téléchargée via <https://github.COM/SMTorg/SMT>. Ce septième article décrit donc les fonctionnalités de cette toolbox et son utilisation sur des exemples. En janvier 2018, nous avons intégré les modèles de mélanges d'experts (*moe*) et en août 2018 les modèles de co-krigeage (*m/k*) pour faire de la multi-fidélité. Les méthodes devraient également évoluer avec une version de SEGO traitant la multi-fidélité.

4.4 Sélection des publications

Les sept articles sélectionnés sont donnés ci-dessous dans leur version éditeur. La liste complète est rappelée ci-dessus :

1. D. Bettebghor, N. Bartoli, S. Grihon, J. Morlier and M. Samuelides. *Surrogate modeling approximation using a mixture of experts based on EM joint estimation*, Structural and Multidisciplinary Optimization, vol 43, n°2, pp 243-259, 2011, 5year-IF 3.269, 2017-IF 2.876, Q1
2. M.-A. Bouhleb, N. Bartoli, A. Otsmane and J. Morlier, *Improving kriging*

surrogates of high-dimensional design models by Partial Least Squares dimension reduction, Structural and Multidisciplinary Optimization, vol 53, n°5, pp 935-952, 2016, 5year-IF 3.269, 2017-IF 2.876, Q1

3. M.-A. Bouhleb, N. Bartoli, A. Otsmane and J. Morlier, *An Improved Approach for Estimating the Hyperparameters of the Kriging Model for High-Dimensional Problems through the Partial Least Squares Method*, Mathematical Problems in Engineering, Volume 2016 (2016), Article ID 6723410, 11 pages, 5year-IF 1.156, 2017-IF 1.145, Q2
4. D. Bettebghor and N. Bartoli. *Approximation of the critical buckling factor for composite panels*, Structural and Multidisciplinary Optimization, vol 46, n°4, pp 561-584, 2012, 5year-IF 3.269, 2017-IF 2.876, Q1
5. M.-A. Bouhleb, N. Bartoli, R. Regis, A. Otsmane and J. Morlier, *Efficient Global Optimization for high-dimensional constrained problems by using the kriging models combined with the Partial Least Squares method*, Engineering optimization, 2018, 5year-IF 1.908, 2017-IF 1.622
6. N. Bartoli, T. Lefebvre, S. Dubreuil, R. Olivanti, R. Priem, N. Bons, J.R.R.A. Martins and J. Morlier, *Adaptive modeling strategy for high-dimensional constrained global optimization*, Aerospace Science and Technology Journal, March 2019, vol 90, p 85-102, 5year-IF 2.439, 2017-IF 2.228
7. M. A. Bouhleb, J. T. Hwang, N. Bartoli, R. Lafage, J. Morlier and J.R.R.A. Martins, *SMT : A Surrogate Modeling Toolbox in Python*, Advances in Engineering Software, Elsevier 2019, 5year-IF 2.948, 2017-IF 3.198

4.4.1 Surrogate modeling approximation using a mixture of experts based on EM joint estimation, Structural and Multidisciplinary Optimization 2011



Open Archive Toulouse Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is a publisher-deposited version published in: <http://oatao.univ-toulouse.fr/>
Eprints ID: 4265

To link to this article: DOI: 10.1007/s00158-010-0554-2

URL: <http://dx.doi.org/10.1007/s00158-010-0554-2>

To cite this version: BETTEBGHOR Dimitri, BARTOLI Nathalie, GRIHON Stéphane, MORLIER Joseph, SAMUELIDES Manuel. Surrogate modeling approximation using a mixture of experts based on EM joint estimation. *Structural and Multidisciplinary Optimization*. 2010. ISSN 1615-1488

Any correspondence concerning this service should be sent to the repository administrator:
staff-oatao@inp-toulouse.fr

Surrogate modeling approximation using a mixture of experts based on EM joint estimation

Dimitri Bettebghor · Nathalie Bartoli ·
Stéphane Grihon · Joseph Morlier ·
Manuel Samuelides

Received: 9 February 2010 / Revised: 15 July 2010 / Accepted: 25 July 2010

Abstract An automatic method to combine several local surrogate models is presented. This method is intended to build accurate and smooth approximation of discontinuous functions that are to be used in structural optimization problems. It strongly relies on the Expectation–Maximization (EM) algorithm for Gaussian mixture models (GMM). To the end of regression, the inputs are clustered together with their output values by means of parameter estimation of the joint distribution. A local expert is then built (linear, quadratic, artificial neural network, moving least squares) on each cluster. Lastly, the local experts are combined using the Gaussian mixture model parameters found by the EM algorithm to obtain a global model. This method is tested over both mathematical test cases and an engineering optimization problem from aeronautics and is found to improve the accuracy of the approximation.

Keywords EM clustering · Gaussian mixture models · Mixture of experts · Surrogate models

1 Introduction

In recent years, design engineers have been provided with many practical tools from mathematical optimization and most of the classical gradient-based optimization techniques are now widely used in all fields (see for instance Haftka and Gurdal 1992 in the field of structural optimization). When using these classical tools, one often faces long simulations times either to compute the optimization constraints, or the objective function, or both. To speed up design, approximation models were developed to tackle the slowness of repetitive code evaluations. When used within a design process, these approximation methods are often called surrogate models. Surrogate models arise from statistics and probability theory and are now widespread tools to approximate complicated functions. They are used inside the optimization process to approximate the objective function or the constraints, or they can directly approximate the results of the optimization process as a function of the optimization problem parameters (materials characteristics, load case in structural optimization for instance). They are also widely used in the multilevel and multidisciplinary optimization framework (Liu et al. 2004; Merval 2008; Merval et al. 2006). Surrogate modeling offers many ways to approximate functions from sample data: artificial neural networks (ANN) (Dreyfus 2005), moving least squares (MLS) (Nealen and Darmstadt 2004), radial basis functions (RBF) (Buhmann 2001), kriging (van Beers and Kleijnen 2004), support vector machines (Smola and Schölkopf 2004), multivariate adaptive regressive splines (MARS) (Friedman et al. 2001). A good

D. Bettebghor · S. Grihon
Airbus France (EDSAZBT), Toulouse, France

S. Grihon
e-mail: stephane.grihon@airbus.com

D. Bettebghor (✉) · N. Bartoli · M. Samuelides
ONERA (DTIM), Toulouse, France
e-mail: dimitri.bettebghor@onera.fr, dimitri.bettebghor@airbus.com

N. Bartoli
e-mail: nathalie.bartoli@onera.fr

J. Morlier · M. Samuelides
ISAE-SupAéro, Toulouse, France

J. Morlier
e-mail: joseph.morlier@isae.fr

M. Samuelides
e-mail: manuel.samuelides@isae.fr

Published online: 29 August 2010

overview of the existing surrogate models can be found in Friedman et al. (2001), Kleijnen et al. (2005), Wang and Shan (2007) and Simpson et al. (2008). In Forrester and Keane (2009), Forrester and Keane provide an intensive review of most of the surrogate models and compare them when used within an optimization process. Nonetheless, one simple surrogate model might not be enough to approximate a complicated function, especially when this function features different behaviors depending on the region of the input space. This situation happens quite often in mechanics when computing critical buckling modes. For instance in Merval (2008), the optimization constraints to be approximated (reserve factors for skin buckling and local web stringer buckling for a composite stiffened panel) happen to be discontinuous and derivative-discontinuous which precludes the training of an accurate surrogate model. Indeed surrogate models usually assume that the function to approximate is smooth and are themselves smooth. This results in a high variance around the discontinuities and that makes the generalization power of the surrogate model poorer. One way to prevent this high variance would be to divide the input space into regions that do not feature discontinuities and then build a surrogate model on each of these regions. This way one could get rid of the discontinuities.

To improve the accuracy of a surrogate model, a common practice is to build several surrogate models on the same common learning basis. One is more likely to find a more accurate surrogate when building many surrogate models. As explained in Viana et al. (2009), several surrogate models prevents from building poorly fitted models. On the basis of the different surrogate models, one can choose the most accurate based on the classical statistical techniques of cross-validation and bootstrap to obtain estimates of the generalization error of the different surrogate models (see for instance Kohavi 1995 and Picard and Cook 1984). As pointed out in Acar and Rais-Rohani (2009), one of the drawbacks of choosing the best predictor is that we do not make use of all the resources used in the construction of the discarded surrogate models. The chosen predictor may be globally precise enough but may lack accuracy in some crucial areas of the input space (boundaries for instance), while one of several of the discarded surrogate models may perform better in these very areas. One could overcome this drawback by combining all the surrogate models by means of weights. This practice of combination relies on the same basis as committees of machines in artificial intelligence. A committee of machines is a collection of intelligent agents that vote and decide all together, hoping that errors would cancel as there are several experts. In the area of machine learning, this practice of combination appears in the bagging and boosting techniques. In the case of an ensemble of surrogate models, the different surrogate models may be simply averaged or weighted. Note that the weighting may be done

globally (constant weights over the input space) as it is done in Viana et al. (2009) and in Acar and Rais-Rohani (2009) or locally (depending on the input space) as it is done in Zepa et al. (2005) and in Sanchez et al. (2008). Even though the idea of combining several surrogate models seems appropriate to approximate functions, there is no evidence that combining is always better than selecting the best surrogate, as it is pointed out in Yang (2003).

Our approach is in the framework of ensembles of locally weighted surrogate models except that it is also based on a partitioning of the learning basis, whereas in most of the described techniques of ensembles, the surrogate models are built on the same common learning basis. Indeed, our concern is about the approximation of functions featuring discontinuities, heterogeneous behaviors and very different landscapes depending on the region of the input space. This is why, in our approach, a surrogate model is built over a specific region of the input space. From an optimization point of view, the global surrogate model needs to be continuous and even smooth (for gradient-based optimization techniques), this is why we combine them in a way that their errors are canceled, notably in the vicinity of discontinuities. As pointed out, this approach differs slightly from the existing ensembles of surrogate models, since each surrogate model, though applied over the whole input space at the very end, is only built over a specific region of the input space. Our approach is based both on the same idea as the committee of machines but also on the ‘Divide and Conquer’ principle. In the literature, this kind of approach is referred to as mixture of experts (MoE’s). A general introduction to mixture of experts can be found in Friedman et al. (2001). One classical application is known as hierarchical mixture of experts (HME) and is described in Jordan and Jacobs (1994). In this study, Jordan and Jacobs present a general architecture of mixture of experts for supervised learning (regression and classification). This architecture is a tree structure where each nonterminal produces a soft split of the input value coming from the upper level. This soft split consists in giving different weights to the lower sub levels by means of a generalized linear model; the authors call it gating network. Each gating network produces soft splits until the terminal leaves, which produce output (real value for regression and binary for classification) by means of a generalized linear model. These terminal leaves are called by the authors expert network. The parameters of the different generalized linear models are estimated using a classical algorithm in statistics: the Expectation-Maximization algorithm (EM algorithm) on the input-output space, which means that in their study, partitioning and learning are based on the same algorithm. We propose a different method where clustering is separated from learning. The gating networks are not generalized linear models (GLIM) but Gaussian mixture models (GMM) still estimated through

EM algorithm. Based on the Gaussian mixture models estimates, the input-output space (or conjoint space) is clustered with respect to the maximum a posteriori (MAP). Once this clustering is done, we have a certain number of sub-bases and a surrogate model is trained over each sub-basis. This surrogate model can be quadratic regression, artificial neural networks and moving least squares regression, while Jordan and Jacobs use a generalized linear model. All the surrogate models are combined on the basis of the Gaussian parameters found by EM algorithm. The proposed technique relies on a certain number of hard tasks that try to answer the following questions

- a) How do we cluster the learning basis? Clustering (or automatic classification) is one of most important areas of unsupervised learning. It aims at finding groups whose individuals are close in some sense into a collection of points. There are many different techniques to cluster data (K-means, K-medoids, quality threshold clustering, density clustering see Berkhin (2002) for instance where most of the classical algorithms for clustering are reviewed). Some of these are hard clustering, where each point of the design space belongs to one and only one cluster. Others are fuzzy clustering, where each point of the design space belongs to several clusters and each point is associated to a random vector defining the probabilities to lie within each cluster.
- b) Which local experts do we build and how do we combine them? As pointed out in Yang (2003), there is no reason that combining several surrogate will perform better than only one, which means that the combination has to be done carefully based on a technique that is expected to cancel errors.
- c) How to choose the number K of clusters? This is a rather difficult question since there might not be perfect number of clusters. Indeed, we just want to find a good number of clusters such that each expert would do well enough to be combined. There might be different choices for K . The question of the number of clusters is central in clustering theory and involves different tools from statistics, probability theory and information theory. In Burnham and Anderson (2004), some of the most common criteria to find the best number of clusters (Akaike Information Criterion and the Bayesian Information Criterion) are thoroughly investigated and compared.

As said earlier, our concern is mostly about the approximation of discontinuous functions to the end of optimization. The clustering should be done such that the boundary between clusters would be as close as possible to the real discontinuities. We assume that these discontinuities may be distributed all over the domain and would naturally divide

the input-output space into bunches of points that are connected but that may be anisotropic. Clustering should also be done in such a way that we have a center and a parameterization of each cluster to combine the local surrogate models. To that end, we assume that a good representation of the conjoint data for discontinuous functions would be Gaussian mixture models. The EM algorithm would allow us to find good estimates of the Gaussian parameters. EM clustering for Gaussian mixture models gives an anisotropic representation of the data and therefore a parameterization of the clusters that makes it possible to combine the different surrogate models built on each cluster. In Bradley et al. (1998), some of the benefits that can be taken from using EM clustering are developed. Next, we describe our method by answering the former questions. We first present the theoretical background of our proposed technique: Gaussian mixtures models, clustering based on the GMM estimates in Section 2 and answer question a). We then focus on the combination of the local experts trained on the basis of the clustering, this will allow us to answer question b) and derive the original algorithm presented in Section 3 to improve the accuracy of surrogate modeling for discontinuous functions. In Section 4, we validate our proposed algorithm on test cases obtained from a discontinuous functions samples generator called Samgen and we also give a practical answer to question c). In Section 5, we finally test our proposed algorithm on a structural optimization problem from aeronautics. The EM algorithm, which is the basis of our method is recalled in Appendix A. We also give an original interpretation of a standard surrogate model the weighted least squares (WLS) in terms of mixture of experts based on a soft clustering and local quadratic experts in Appendix B. The tool called Samgen that we implemented to provide highly discontinuous functions in arbitrary dimension is outlined in Appendix C.

2 Gaussian mixture models and EM clustering

2.1 Gaussian mixture models

We first describe in this section Gaussian mixture models. Suppose we are given $\mathcal{X} = (x_1, \dots, x_n)$ a set of data where $x_i \in \mathbb{R}^d$. Assume that these x_i 's come from $\{X_i\}_{i=1\dots n}$ a set of identical and independently distributed (iid) random variables. An important problem in statistics and in probability theory is the estimation of the probability density function (pdf). In the case of GMM, we assume that the probability law of X is a weighted combination of a given number K of multivariate Gaussian laws

$$X \sim \sum_{k=1}^K \alpha_k \mathcal{N}(\mu_k, \Gamma_k), \quad (1)$$

where the α_k 's are the mixture parameters, i.e. α_k is the proportion of the Gaussian k in the mixture. Note that, if we denote f_k the pdf of Gaussian k , the \sim symbol means that the pdf of X is $\sum_{k=1}^K \alpha_k f_k$. We obviously have $\sum_{k=1}^K \alpha_k = 1$ since it is a probability density function. The other parameters are the Gaussian parameters. Recall that a multivariate Gaussian distribution is defined by its mean $\mu_k \in \mathbb{R}^k$ and its variance-covariance matrix $\Gamma_k \in \mathcal{M}_d(\mathbb{R})$, which is symmetric positive definite. The pdf of Gaussian $\mathcal{N}(\mu_k, \Gamma_k)$ is therefore

$$f_k(x) = \frac{1}{\sqrt{(2\pi)^d \det(\Gamma_k)}} e^{-\frac{1}{2}(x-\mu_k)^T \Gamma_k^{-1} (x-\mu_k)}. \quad (2)$$

To illustrate GMM's, we depicted Fig. 1 two different GMM's, respectively in \mathbb{R} and in \mathbb{R}^2 .

2.2 Expectation-maximization clustering

We recall in Appendix A the EM algorithm, which is a practical tool to determinate the underlying structure of data. More precisely in the Gaussian mixture model case, EM algorithm gives estimates of the means and the variance-covariance matrices. Since we consider regression problems, we assume we have $\mathcal{X} = (x_1, \dots, x_n)$ a set of inputs and $\mathcal{Y} = (y_1, \dots, y_n)$ the corresponding outputs. Note that we consider in the following that the output space is \mathbb{R} , but we could have considered \mathbb{R}^d with no changes. We want to describe the conjoint law of (X, Y) as a GMM.

Consider the conjoint space $\mathcal{X} \times \mathcal{Y} = \mathcal{Z} = (z_1, \dots, z_n)$ where $z_i = (x_i, y_i) \in \mathbb{R}^{d+1}$. Suppose we have set the number of clusters to K . We estimate through EM algorithm the parameters of the K multivariate Gaussian distributions in \mathbb{R}^{d+1} such that

$$Z \sim \sum_{k=1}^K \alpha_k \mathcal{N}(\mu_k, \Gamma_k), \quad (3)$$

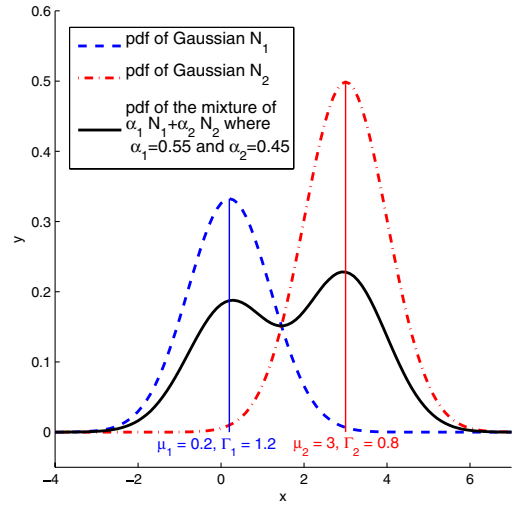
where the α_k 's are the mixture parameters, i.e for all $k \in 1 \dots K$, $\alpha_k \in [0, 1]$ and

$$\sum_{k=1}^K \alpha_k = 1 \quad (4)$$

and $\mu_k \in \mathbb{R}^{d+1}$ is the mean of the Gaussian distribution k and denote

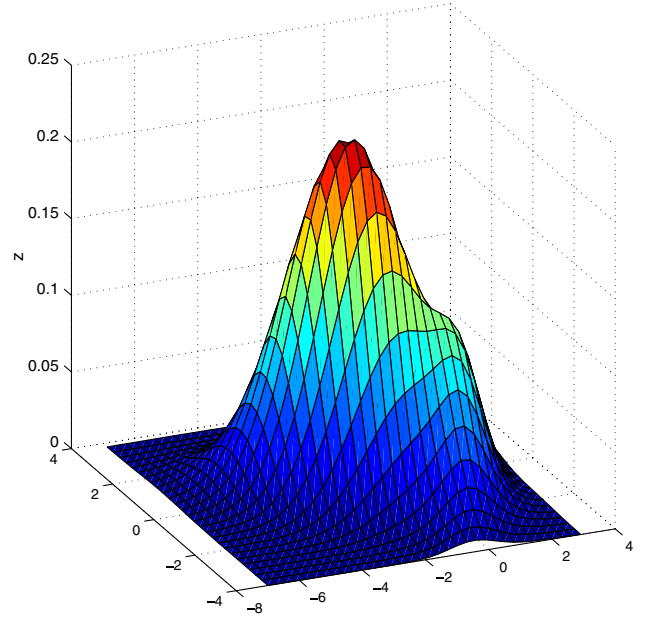
$$\mu_k = \begin{pmatrix} \mu_k^X \\ \mu_k^Y \end{pmatrix}, \quad (5)$$

Probability density function of a Gaussian Mixture in 1D



(a) pdf of a GMM in \mathbb{R}

Probability density function of a Gaussian Mixture in 2D



(b) pdf of a GMM in \mathbb{R}^2

Fig. 1 Examples of GMM pdf's

where $\mu_k^X \in \mathbb{R}^d$ is the X -coordinates of the mean μ_k and $\mu_k^Y \in \mathbb{R}$ is the Y -coordinate of the mean μ_k . $\Gamma_k \in \mathcal{M}_{d+1}(\mathbb{R})$ is the variance-covariance matrix and denote

$$\Gamma_k = \begin{pmatrix} \Gamma_k^X & v_k \\ v_k^T & \xi_k \end{pmatrix}, \quad (6)$$

where $\Gamma_k^X \in \mathcal{M}_d(\mathbb{R})$ is the variance-covariance matrix of X for Gaussian k , $v_k \in \mathbb{R}^d$ is $\text{Cov}(X, Y)$ for Gaussian k and $\xi_k = \text{Var}(Y)$.

Once the GMM parameters are estimated, we can now compute the posterior probabilities, that is to say, the probability for a given $(x, y) \in \mathbb{R}^{d+1}$ to lie within cluster k_i . It is given by Bayes' formula where κ denotes the discrete random variable associated with the clusters

$$\begin{aligned} \mathbb{P}(\kappa = k_i | (X, Y) = (x, y)) \\ = \frac{\mathbb{P}(\kappa = k_i) \mathbb{P}((X, Y) = (x, y) | \kappa = k_i)}{\sum_{k=1}^K \mathbb{P}(\kappa = k) \mathbb{P}((X, Y) = (x, y) | \kappa = k)}. \end{aligned}$$

For the particular case where (X, Y) is assumed to be a Gaussian mixture model: $(X, Y) \sim \sum_{k=1}^K \alpha_k \mathcal{N}(\mu_k, \Gamma_k)$ we have for all $k \in \{1, \dots, K\}$

$$\begin{aligned} \mathbb{P}(\kappa = k) &= \alpha_k, \\ (X, Y) | \kappa = k &\sim \mathcal{N}(\mu_k, \Gamma_k), \end{aligned}$$

which leads with $z = (x, y)$ to

$$\begin{aligned} \mathbb{P}(\kappa = k_i | (X, Y) = (x, y)) \\ = \frac{\det(\Gamma_{k_i})^{-\frac{1}{2}} \alpha_{k_i} e^{-\frac{1}{2}(z - \mu_{k_i})^T \Gamma_{k_i}^{-1} (z - \mu_{k_i})}}{\sum_{k=1}^K \det(\Gamma_k)^{-\frac{1}{2}} \alpha_k e^{-\frac{1}{2}(z - \mu_k)^T \Gamma_k^{-1} (z - \mu_k)}}. \end{aligned} \quad (7)$$

Note that in (7), the 2π factor vanishes, but the determinant of the variance-covariance matrices Γ_k 's remain as the mixture parameters α_k 's. Formula (7) offers two ways of partitioning the data set:

- hard clustering: we simply choose over all the k 's the one that gives the highest probability. Given (x, y) in the conjoint space, (x, y) lies in cluster j where

$$j = \operatorname{argmax}_{j=1, \dots, K} \mathbb{P}(\kappa = j | (X, Y) = (x, y)). \quad (8)$$

- soft clustering: points belong to all clusters, we assign to each point (x_i, y_i) its probability to be generated knowing the cluster j (and thus the mode j). Each point is therefore given a set of K probabilities $p = (p_1, \dots, p_K)$ where

$$p_j = \frac{\mathbb{P}((X, Y) = (x, y) | \kappa = j)}{\sum_{k=1}^K \mathbb{P}((X, Y) = (x, y) | \kappa = k)}. \quad (9)$$

Indeed, we want to train a surrogate model over \mathcal{X} for mode j , therefore each learning point x_i should be weighted by the probability p_j to have x_i knowing that the cluster is j . This probability is normalized such that $\sum_{k=1}^K p_k = 1$.

In this article, we chose the hard clustering. The hard clustering keeps the localization of the examples in the same

cluster in the sense that they are close with respect to the Mahalanobis distance¹ associated with the variance-covariance matrix Γ_k . The other reason is that building surrogate model from weighted examples is not straightforward, apart from linear and polynomial multivariate regressions where it boils down to weighted least squares (see Appendix B). In the next section, we describe the way we combine the different local experts built on this hard clustering.

3 Combining local experts and proposed algorithm

The learning basis is split into K learning sub-bases and an expert f_k is trained over each sub-basis k . Any surrogate model can be used, we give results obtained using the following different local experts:

- linear regression: the most simple expert (apart from the constant expert which leads to radial basis functions regression). It can be computed easily and the multivariate linear regression parameters are directly given by the Gaussian component parameters

$$\begin{aligned} f_k(x) &= \frac{\operatorname{Cov}^k(X, Y)}{\operatorname{Var}^k(X)} (x - \mathbb{E}^k(X)) + \mathbb{E}^k(Y) \\ &= (\Gamma_k^X)^{-1} v_k (x - \mu_k^X) + \mu_k^Y. \end{aligned} \quad (10)$$

- quadratic regression: the original response surfaces, which are quadratic polynomials over \mathbb{R}^d as extensively described in Myers et al. (2009). They are also relatively inexpensive to build however there is no easy

¹The Mahalanobis distance of a random variable $X \in \mathbb{R}^d$ is the distance defined by the inverse the variance-covariance matrix $\Gamma = \operatorname{Var}(X)$: for $\omega_1, \omega_2 \in \mathbb{R}^d$ the Mahalanobis distance is $D_M(\omega_1, \omega_2) = \|\omega_1 - \omega_2\|_{\Gamma^{-1}} = \sqrt{(\omega_1 - \omega_2)^T \Gamma^{-1} (\omega_1 - \omega_2)}$. It does define a distance since the inverse of Γ (sometimes called the precision matrix) is symmetric positive definite.

²All these matrices are symmetric positive definite but they can become nearly-singular especially in case of redundant data (linearity), QR factorization performs better than Gaussian reduction and even Choleski factorization.

formula that can be derived from the Gaussian components parameters. In our case, we computed it in a simple way taking care of the inversion of the system.

- artificial neural networks: we use here the classical Multi Layer Perceptron (MLP) as a local expert; MLP models are thoroughly described in Haykin (2008). We use one-layer networks and the number of hidden neurons is classically determined through a cross-validation procedure and the network is trained using Levenberg–Marquardt algorithm.
- moving least squares: the MLS expert is the most complicated expert to compute for it is an implicit model that needs to be recomputed at each new evaluation point. We implemented a moving least squares method based on the Backus–Gilbert approach that can be found in Fasshauer (2005). We also implemented a golden ratio search to optimize the hyper-parameter σ (width of the Gaussian kernel, see Appendix B). A landmark paper on MLS is Levin (1998) and a brief introduction can be found in Nealen and Darmstadt (2004).

Nonetheless, any surrogate model can be used (kriging, support vector regression, radial basis function, multivariate adapted regressive splines) as a local expert and can be perfectly improved using an ensemble of surrogate models, or boosting (Meir and Ratsch 2003). Moreover, a local expert can be a black-box or itself an MoE and so on.

Once we build our local experts f_k , we want to predict the response y for a new entry $x \in \mathbb{R}^d$. This is done by combining them. We form a linear combination of the local experts f_k . A natural idea is that this linear combination should not be constant over the whole input space and should give more weight to expert f_k when x gets closer to the center of cluster k with respect to the natural Mahalanobis distance inherited by the variance-covariance matrix of Gaussian k . Namely the global model \hat{f} is

$$\hat{f}(x) = \sum_{i=1}^K \beta_i(x) f_i(x), \quad (11)$$

where $\beta = (\beta_1, \dots, \beta_K)$ is an expert that gives the local weights (local in the sense that it does depend on x). A natural gating network would be

$$\beta(x) = (\mathbb{P}(\kappa = 1|X = x), \dots, \mathbb{P}(\kappa = K|X = x)), \quad (12)$$

such that the global model would be

$$\hat{f}(x) = \sum_{i=1}^K \mathbb{P}(\kappa = i|X = x) f_i(x). \quad (13)$$

Equation (13) is the classical probability expression of mixture of experts (as it can be found in Jordan and Jacobs

1994). Note that this expression may represent a lot of different situations and therefore a lot of different MoE's. For instance, as said earlier, the weighted least squares can be interpreted as an Moe using that equation (see Appendix B). To use (13) we need to know what is the law of κ knowing that $X = x$ and without knowing that $Y = y$, this can be easily obtained with the Gaussian parameters found by EM algorithm. Indeed, from the conjoint law $(X, Y) \sim \sum_{k=1}^K \alpha_k \mathcal{N}(\mu_k, \Gamma_k)$, we can derive the law of $X|\kappa = k$ without knowing Y

$$X|\kappa = k \sim \mathcal{N}_d(\mu_k^X, \Gamma_k^X), \quad (14)$$

such that the global GMM law of X is

$$X \sim \sum_{k=1}^K \alpha_k \mathcal{N}(\mu_k^X, \Gamma_k^X). \quad (15)$$

Note that this Gaussian mixture model is different from the one we would have obtained by applying EM only on the inputs X 's for it is the projection on the input space of the conjoint law. Therefore we can derive equivalently the posterior probability from Bayes' formula

$$\begin{aligned} \mathbb{P}(\kappa = k_i|X = x) &= \frac{\det(\Gamma_{k_i}^X)^{-\frac{1}{2}} \alpha_{k_i} e^{-\frac{1}{2}(x-\mu_{k_i}^X)^T \Gamma_{k_i}^{X-1} (x-\mu_{k_i}^X)}}{\sum_{k=1}^K \det(\Gamma_k^X)^{-\frac{1}{2}} \alpha_k e^{-\frac{1}{2}(x-\mu_k^X)^T \Gamma_k^{X-1} (x-\mu_k^X)}}. \end{aligned} \quad (16)$$

Note that the global model defined with (13) and (16) is completely smooth. In the sequel, this mixture of experts will be referred to as smooth mixture. At this point, we can think of another way of combining the local surrogate models that takes more advantage from the clustering made by the EM algorithm. Indeed, based on the Gaussian parameters estimated on the clustering step, we can predict which cluster a new given entry x lies in and simply assign to this entry x the corresponding local surrogate model. This means that we can, at least formally, define a partitionning of the whole input space: $\mathcal{X} = \cup_{k=1}^K \mathcal{X}_k$ and simply define the law of k knowing that $X = x$ as a uniform discrete law such that the global model would be

$$\hat{f}(x) = \sum_{k=1}^K \mathbb{1}_{\mathcal{C}_k^X}(x) f_k(x) \quad (17)$$

where

$$\mathbb{1}_{\mathcal{C}_k^X}(x) = \begin{cases} 1 & \text{if } k = \operatorname{argmax}_{j=1, \dots, K} \mathbb{P}(\kappa = j|X = x) \\ 0 & \text{if } k \neq \operatorname{argmax}_{j=1, \dots, K} \mathbb{P}(\kappa = j|X = x) \end{cases} \quad (18)$$

and (17) defines the most simple mixture of experts where a new entry x is given a cluster k and the predicted value

is simply $f_k(x)$. An important feature of this mixture of experts is that it is not continuous, indeed at the boundary between two adjacent clusters³ the two adjacent local surrogate models need not to match resulting in a globally discontinuous model. This mixture of experts will be referred to as hard mixture of experts. In the case when the functions to approximate is discontinuous the hard mixture version is likely to be more accurate than the smooth version. Besides being discontinuous, the hard mixture version may create artificial discontinuities where the original function does not have ones (see Fig. 2d). In this article, we are mainly concerned with approximating functions that are objective or constraints functions of an optimization problem that is to be solved on the basis of a gradient-based method, this is why we will focus on the smooth mixture of experts. Indeed, in such applications, we are not only concerned with the accuracy of the approximation but also with the regularization of the original approximation. In terms of accuracy the hard version is likely to perform better (see for instance Fig. 2d) but the optimization algorithm may fail to converge due to the non-differentiability of the global approximation model.

Our algorithm is presented here and results of this method are given in the following sections with the help of two test cases and one engineering problem. This method is also illustrated on a one-dimensional test case in Fig. 2.

1. Assemble Z the conjoint learning basis where $z_i \in \mathbb{R}^{d+1}$ contains inputs $x_i \in \mathbb{R}^d$ and output $y_i \in \mathbb{R}$, see Fig. 2a

$$Z = \begin{pmatrix} x_1^{(1)} & \dots & x_n^{(1)} \\ \vdots & & \vdots \\ x_1^{(d)} & \dots & x_n^{(d)} \\ y_1 & \dots & y_n \end{pmatrix}. \quad (19)$$

2. Set the number of clusters K as explained below. In Fig. 2 the number of clusters was set to 3.
3. Apply EM algorithm to Z with K to get $\hat{\alpha}_k$, $\hat{\mu}_k$ and $\hat{\Gamma}_k$, estimates of the real Gaussian parameters.
4. Hard clustering of the data, see Fig. 2b. $z_i = (x_i, y_i)$ belongs to cluster k_i where

$$k_i = \operatorname{argmax}_{j=1, \dots, K} \mathbb{P}(j | (X, Y) = (x_i, y_i)). \quad (20)$$

where $\mathbb{P}(\kappa = j | (X, Y) = (x_i, y_i))$ is computed using (7)

5. Split the conjoint learning basis into K clusters $Z = \cup_{i=1}^K Z^{(i)}$.

6. For $i = 1 \dots K$
 - (a) Remove outliers using for instance Mahalanobis distance.
 - (b) Split randomly $Z^{(i)} = Z_{\text{learn}}^{(i)} \cup Z_{\text{test}}^{(i)}$ into learning basis and test basis.
 - (c) Train expert f_i on $Z_{\text{learn}}^{(i)}$, choose the best expert f_i with $Z_{\text{test}}^{(i)}$, see Fig. 2c.
7. Combine all the experts with

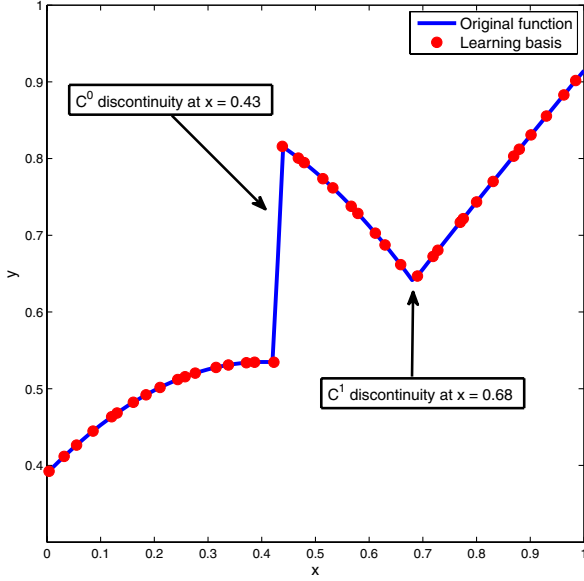
$$\hat{f}(x) = \sum_{i=1}^K \mathbb{P}(\kappa = i | X = x) f_i(x). \quad (21)$$

where $\mathbb{P}(\kappa = i | X = x)$ is computed using (16), see Fig. 2d where we also plotted the hard version of the mixture of experts.

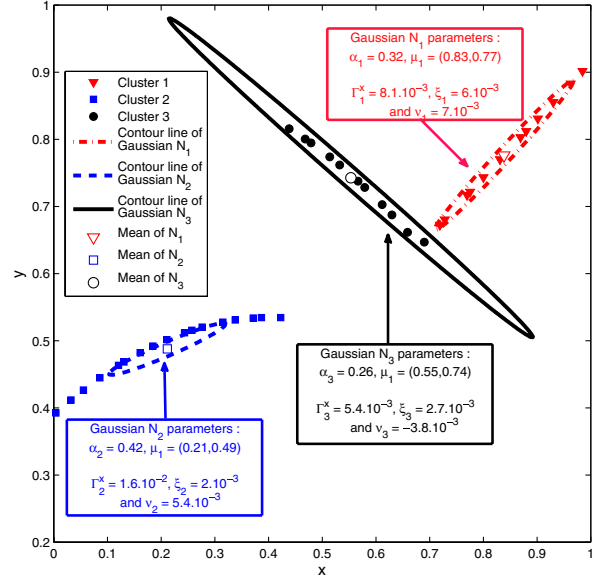
So far we have not answered the crucial question about the number of clusters. We suggest here a practical way to determinate a good number of such clusters. As discussed earlier, there might not be a perfect number of clusters, the function may be continuous and non derivative discontinuous and only one expert may be a good choice to approximate the function. Nonetheless, we usually observed even in that case that several experts can perform better than only one. On the other hand, too many experts may split the learning basis onto clusters with too few examples to train a reasonable surrogate model. There is a trade-off between the complexity of the model (number of free parameters) and the capacity of generalization of the MoE. To set the correct number of clusters (and experts) we suggest to build MoE's with linear experts and quadratic experts, estimate the MSE and the other error measures for different number of clusters, (say for $k = 2 \dots N_A/10$) and then choose the number of clusters that minimizes the MSE and \hat{E}_{mean} . In the case where there are several number of clusters that could do it, we would better choose the lowest to make the clusters as big as possible and thus improve the accuracy of the local experts. Part of the reason for this procedure is that building linear experts and quadratic experts is inexpensive in comparison to building neural experts or moving least squares regression experts.

Note that this algorithm was designed for smoothing discontinuous or derivative-discontinuous functions. It was pointed out that the EM clustering is expected to separate disconnected parts of the conjoint space and therefore to grasp regions where the response to approximate is continuous and where building a surrogate model would be much easier than building one on the whole domain. Nevertheless, combining all the local surrogate models into a global smooth one is likely to make the approximation less accurate around the discontinuities, while in the hard mixture version the discontinuities are preserved. This is

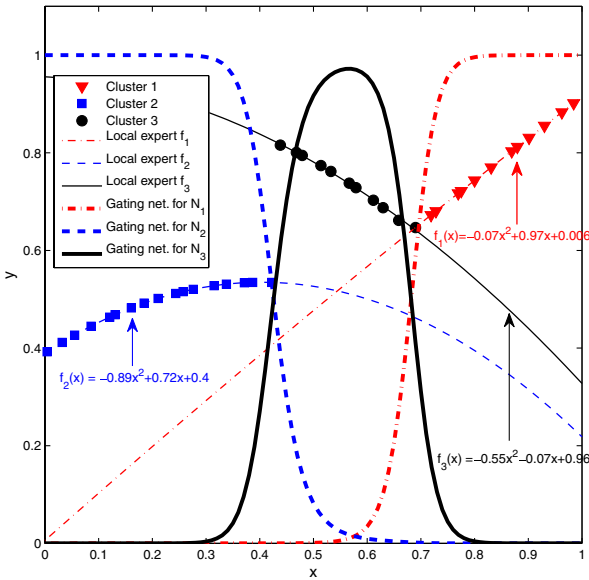
³This boundary is often known in Probability as the Bayes classifier.



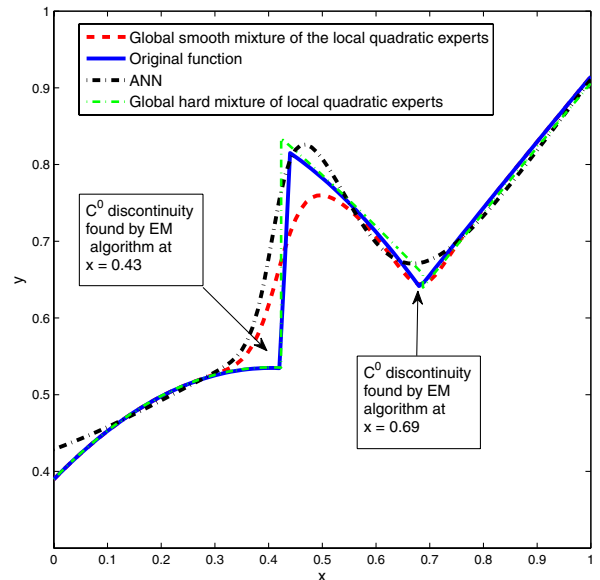
(a) samgen_1D_1 test case and learning basis.



(b) Results of the EM clustering and associated Gaussian laws



(c) Local quadratic experts and gating networks



(d) Global predictor

Fig. 2 Sketch of the proposed algorithm for a 1D test case `samgen_1D_1`. **a** Original function and learning basis. Note that `samgen_1D_1` features two C^0 and C^1 discontinuities. EM clustering is expected to separate them well enough to locally build an accurate surrogate model. **b** EM clustering on the learning basis for $K = 3$. We depict the *contour lines* of the quadratic function associated to the variance-covariance matrix Γ_i at 3×10^{-6} for $i = 1 \dots 3$. Note that these lines are simply the balls of radius 3×10^{-6} centered at μ_i for the Mahalanobis distance associated with the variance-covariance matrices $\begin{pmatrix} \Gamma_i^X & \nu_i \\ \nu_i & \xi_i \end{pmatrix}$. **c** Local quadratic experts and gating networks. See that the gating network associated to cluster 2 is quite steep. This

is due to the relative small size of cluster 2. **d** Global surrogate models obtained through the mixture of local quadratic experts. We plotted the soft and the hard mixture versions. We also depicted a reasonably good artificial neural network to compare. We observe that the smooth predictor is very accurate on clusters 1 and 3 is a little bit less accurate on cluster 2. While the artificial neural network does not generalize very well at the boundaries of the domain, the global smooth predictor performs better at these boundaries due to the local behaviour of the surrogate models. The hard mixture predictor is much more accurate since it does not regularize the discontinuities. It creates though an artificial C^0 discontinuity at $x = 0.69$

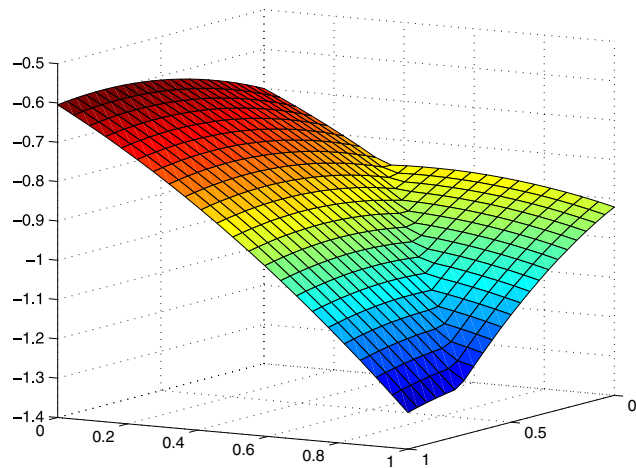
illustrated in Fig. 2d. Besides, such functions may be well approximated using several existing methods. Apart from the methods of ensemble of surrogate model that we briefly describe in the introduction section, the multivariate adaptive regressive splines (MARS) are often said to handle such discontinuous functions especially in high dimensions (see Friedman et al. 2001). Indeed, MARS builds a model using a basis of piecewise linear functions and their products. Nevertheless, MARS does not cluster the learning basis and does not seem to keep the local behavior of the function to be approximated. One of the benefits of our proposed algorithm is to subdivide the whole design space into ‘regions of interest’ where previous knowledge or expertise for a specific surrogate model may help to build an accurate local approximation that will be merged into a global approximation. Indeed, we can use cross-validation when building each local expert and therefore get estimators of the generalization error made by each expert. This information can be used to detect ‘regions’ (that can be a sole cluster or several) where this error is larger than in the rest of the input space. These regions may be enriched (by adding learning points) to get a better global approximation model. Lastly, this local information can be compared to the generalization error of the global approximation model to assess the accuracy of the mixture of experts.

4 Validation on test cases

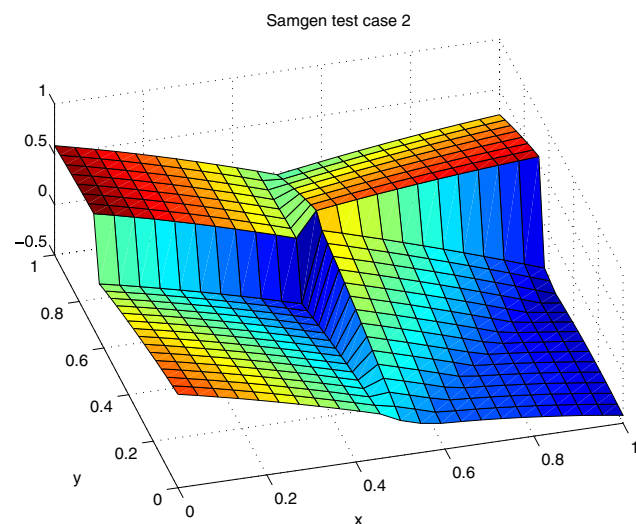
We present the results of our proposed algorithm obtained with three cases. The two first are mathematical examples who were designed to feature C^0 and C^1 discontinuities. These two functions were generated with Samgen (see Appendix C). Part of this study was to carry out new methods to approximate discontinuous functions or functions that feature different global behaviors depending on the region of the design space. These two examples `samgen_2D_1` and `samgen_2D_2` fulfill these requirements.

These two functions are defined through the maximum of non-convex quadratic functions over sub-domains of $[0, 1]^2$. We depicted these two functions Fig. 3. Note that `samgen_2D_1` only features a discontinuity of the derivative, `samgen_2D_2` features two orthogonal discontinuities that subdivide the whole domain into four sub-domains, and it also features a discontinuity of the derivative. These test cases are interesting for several reasons.

- Most surrogate models tend to regularize the discontinuities and therefore hardly handles this test case.
- The outputs are in the same range for all sub-domains. This is why clustering only on the output values does



(a) `samgen_2D_1` test case



(b) `samgen_2D_2` test case

Fig. 3 Mathematical test cases generated with Samgen, outlined in Appendix C. These functions features C^0 discontinuities (b) and C^1 discontinuities (a and b)

not detect the correct four sub-domains. We expect EM clustering to detect these four sub-domains.

To test our proposed algorithm, we used the same procedure for both test cases. The general idea was to compare the ‘best’ surrogate model over the whole domain and the proposed MoE for different types of local experts.

To assess the sensitivity of the clustering to design of experiments, we generated several designs of experiments, ran on each design the proposed algorithm and compared the MoE obtained with the ‘best’ surrogate over the whole domain. At the end, we compared the mean of the different

error measures. More formally, our procedure was the following

1. We first set the number of learning points N_A .
2. We generated N_{doe} designs of experiments.
3. For each design of experiment, we built a reasonably good surrogate model over the whole domain found after a thorough investigation. To that end, we trained ANN, MLS, quadratic and linear models and chose the one that gives the least generalization error estimated through a classical cross-validation procedure.
4. For each design of experiment, we ran the proposed algorithm. We decided to compare five different types of MoE's
 - MOE_LIN: the local experts are linear regression experts
 - MOE_QUAD: the local experts are quadratic regression experts
 - MOE_ANN: the local experts are artificial neural Networks
 - MOE_MLS: the local experts are moving least squares experts.
 - MOE_BEST: once the clustering is done, all the different types of experts are build and the best expert over each cluster is chosen, resulting in a global mixture of experts made of different types of experts.
5. We compared of all the surrogate models built using the following measures of approximation quality
 - RMSE: root mean squared error, most common measure (the surrogate models are usually trained by minimizing this quantity). Note that we could have used the mean squared error (MSE), since, in this very case, we just use the RMSE to compare different surrogate models among them. Owing to the monotonicity of the square root function, both RMSE and MSE would give the same rank among them.

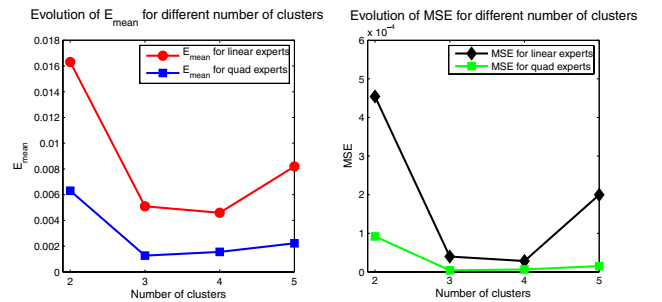
$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n \|f(x_i) - y_i\|^2} \quad (22)$$

- \hat{E}_{max} : maximum of the absolute error
- \hat{E}_{mean} : mean of the absolute error. We did not use the relative error measure to prevent from dividing by zero since `samgen_2D_2` maps to $[-1, 1]$

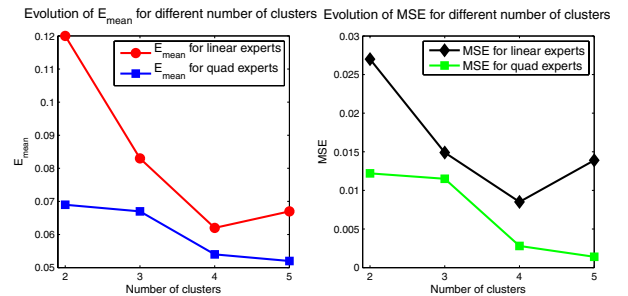
6. Finally, we computed the mean for each error measure to compare all the different surrogate models.

For `samgen_2D_1`, $N_{\text{doe}} = 100$ Latin hyper-cube sampling (LHS) designs of experiments of $N_A = 80$ points

in $[0, 1]^2$ were generated. The global surrogate models and the proposed MoE's local experts were trained over a classical cross-validation partition of $(4/5, 1/5)$ for learning basis and test basis. The errors presented here were computed over a full factorial plan of 35×35 points and the means of the errors are computed over the 100 LHS. The number of clusters was set to 3 based on the practical procedure we describe at the end of Section 3. We plotted the MSE and \hat{E}_{mean} for linear and quadratic expert for different number of clusters in Fig. 4a. We can see that in the case of `samgen_2D_1` that three or four clusters is a good choice. Therefore, we chose three clusters in that case. Results for the different MoE's are given in Table 1. For more than 85 of the different LHS designs, one or several MoE's happened to perform better than the best surrogate model over the whole domain, which happened to be an ANN in most cases. In average, the 'best' surrogate model over the whole domain performs pretty much as well as MOE_QUAD. In this test case, we see that the best surrogate model is MOE_ANN. Note that MOE_MLS is slightly less accurate. Note also that the MOE_BEST had in most cases ANN and MLS experts as best experts and that it performs as well as MOE_ANN. The reasons why the best experts are not quadratic (while the true function is defined as a maximum of quadratic function) is that the EM clustering is not perfect with respect to the true nature of the



(a) `samgen_2D_1` test case



(b) `samgen_2D_2` test case

Fig. 4 \hat{E}_{mean} and MSE for different number of clusters: a `samgen_2D_1` test case, b `samgen_2D_2` test cases

Table 1 Results for samgen_2D_1

Type	RMSE (%)	\hat{E}_{\max}	\hat{E}_{mean}
Best surrogate model over the whole domain			
ANN	0.59	1.2×10^{-2}	7.1×10^{-3}
Proposed MoE			
MOE_LIN	2.66	1.7×10^{-2}	6.6×10^{-3}
MOE_QUAD	0.52	1.1×10^{-2}	2.8×10^{-3}
MOE_ANN	0.73	6×10^{-3}	1.8×10^{-3}
MOE_MLS	0.31	9.7×10^{-3}	3.2×10^{-3}
MOE_BEST	0.29	7.3×10^{-3}	2.9×10^{-3}

Table 2 Results for samgen_2D_2

Type	RMSE (%)	\hat{E}_{\max}	\hat{E}_{mean}
Best surrogate model over the whole domain			
ANN	13	2.14	0.54
Proposed MoE			
MOE_LIN	12.6	0.51	0.5
MOE_QUAD	9.54	0.36	4.5×10^{-2}
MOE_ANN	6.86	0.23	3.7×10^{-2}
MOE_MLS	5.66	0.26	2.4×10^{-2}
MOE_BEST	5.29	0.29	3.1×10^{-2}

function. A few learning points may be misclassified and ANN are more likely to perform better with outliers than quadratic models.

For samgen_2D_2 $N_{\text{doe}} = 100$ LHS designs of experiments of $N_A = 100$ points in $[0, 1]^2$ were generated. More learning points were used to get rid of poorly fitted models that were due to the important discontinuities. The global surrogate models and the proposed MoE's local experts were trained over a classical cross-validation partition of $(4/5, 1/5)$ for learning basis and test basis. The errors presented here were computed over a full factorial plan of 50×50 points (to be sure that we also consider the error near the discontinuities) and the means of the error measures were computed over the 100 plans. The number of clusters was set to 4 on the basis of the procedure described at the end of Section 3. We plotted the MSE

and \hat{E}_{mean} for linear and quadratic expert for different number of clusters in Fig. 4b. We can see that in the case of samgen_2D_2 that four or five clusters is a good choice. Therefore, we chose four clusters in that case. We also depicted in Fig. 5 the gating networks α_i 's we obtained in that case. Results for the different MoE's are drawn in Table 2. We observed that all of the different MoE's performed better than the best surrogate model (which was also an ANN for 90 of the 100 designs, the rest being MLS). Results show that a simple and very inexpensive MOE_LIN performs better than a sophisticated ANN. Regarding the different types MoE's, we clearly see that the best is again MOE_ANN with MOE_MLS being slightly less accurate and MOE_BEST is again very close to MOE_ANN.

In both cases, MOE_BEST does not give the best accuracy, it is however very close to the best MoE. This may

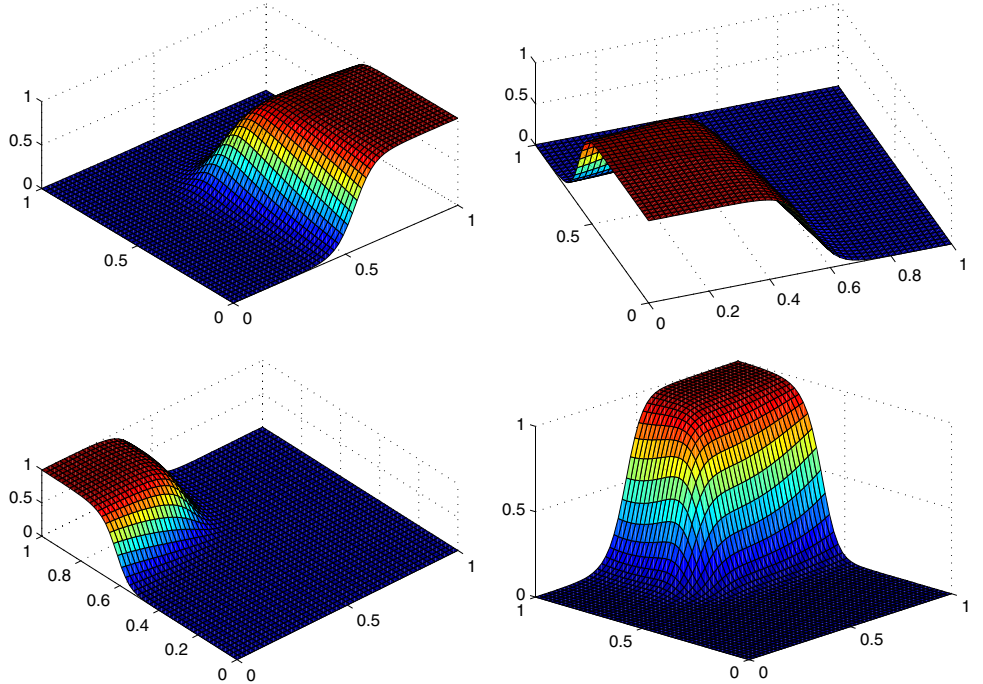
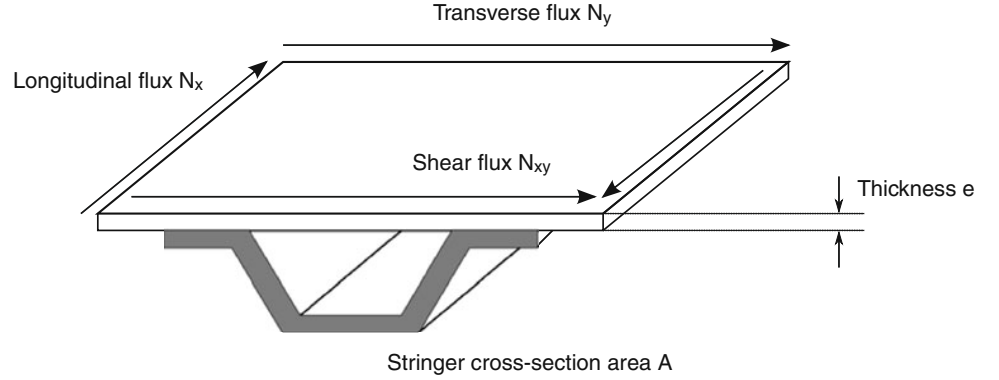
Fig. 5 Gating networks for $K = 4$ clusters in the case of samgen_2D_2. They are very close to the real C^0 discontinuities

Fig. 6 Ω -shaped composite super-stringer under the load case (N_x, N_y, N_{xy})



be due to the fact a very good local expert may not generalize very well outside the cluster where it was built. The best expert is somewhat too local and in these test cases, it seems like combining different types of local experts in the same global MoE does not improve the accuracy. This may also be due to the definition of the test cases, which feature the same kind of quadratic function over each domain (see Appendix C on Samgen). For other test cases, where the function behaves differently depending on the region (polynomial on one region and non-polynomial on another one for instance), MOE_BEST would certainly perform better. It is worth noting that we also observed that for a given type of MoE, the variance of the error measure over the 100 learning bases was quite low, which would indicate that once the type of the expert has been chosen, the accuracy will be nearly independent of the clustering. It should be indicated that in some cases where EM algorithm could not find good estimates and failed to converge (or converged too slowly). Due to the fact there were insufficient example some clusters were almost degenerated which precluded the training of a correct local expert (apart from the linear and quadratic regressions). We must therefore insist in the fact that this procedure needs a certain number of examples and could not be applied in cases where there are very few examples.

5 Stiffened panel optimization test case

We turn now to the third case, which is an engineering problem of structural optimization from aeronautics. It consists of the sizing of a composite panel reinforced by an Ω -shaped stringer, as depicted in Fig. 6. More precisely, we are given N_x, N_y, N_{xy} , the longitudinal flux, the transverse flux and the shear flux. Under this load case, we want to minimize the cross section area of the stringer A and the thickness of the panel e under hard structural constraints (local buckling of the web of the stringer, Euler buckling, skin buckling). These constraints are computed with in-house tools based both on analytical formulas and

a Rayleigh–Ritz approach. We can try to approximate the skill tool, but we are interested here in approximating directly the results of the optimization.⁴ Given N_x, N_y, N_{xy} we want to estimate the optimums A^* and e^* .

To that purpose, the input space, was naturally divided into four sub-domains and 4,000 examples were computed with Boss Quattro (LHS of 1,000 points for each sub-domains)

- $D_1 = \{(N_x, N_y, N_{xy}) | N_x > 0, N_y > 0\}$
- $D_2 = \{(N_x, N_y, N_{xy}) | N_x > 0, N_y < 0\}$
- $D_3 = \{(N_x, N_y, N_{xy}) | N_x < 0, N_y > 0\}$
- $D_4 = \{(N_x, N_y, N_{xy}) | N_x < 0, N_y < 0\}$

This partition of the input space makes sense from physical considerations. Indeed, the shear flux N_{xy} does not influence that much the optimization and we generally observe in mechanics different behaviors depending whether the flux is positive (tension) or negative (compression). We want to find a good approximation of the optimization over each sub-domain. Note that this physical subdivision is in itself a clustering and the resulting global model which consists in assembling all the local experts is a MoE as described in (17). We apply the proposed technique over each sub-domain (and the final global model will be a mixture of mixture of experts) and compare the results with the best surrogate over the whole sub-domain.

Results are depicted in Tables 3, 4, 5, 6 for sub-domains and the overall results are presented in Table 7. Note that we only included results for MOE_ANN since this MoE usually performs slightly better than the other MoE's. We also changed our error measures; we considered the relative error instead of the absolute error and also include the α -quantile for $\alpha = 5$ and $\alpha = 1$, e.g the α -quantile is defined as the percentage of test points that are below an error of $\alpha\%$. By using the relative error, the output values

⁴Such regressions are mainly used to speed up pre-sizing of the aircraft and are known as design curves.

Table 3 Results for D_1

Output	Type	\hat{E}_{\max} (%)	\hat{E}_{mean} (%)	$\alpha = 1\%$ (%)	$\alpha = 5\%$ (%)
Best surrogate model over the whole domain D_1					
A^*	ANN	11	1.2	69	96
e^*	MLS	14.7	1.3	63	95
Proposed MoE with ANN experts ($K = 4$)					
A^*	MOE_ANN	5.6	0.8	80	98
e^*	MOE_ANN	9.5	1.04	65	98

Table 4 Results for D_2

Output	Type	\hat{E}_{\max} (%)	\hat{E}_{mean} (%)	$\alpha = 1\%$ (%)	$\alpha = 5\%$ (%)
Best surrogate model over the whole domain D_2					
A^*	MLS	36.5	2.3	52	88
e^*	MLS	25.7	2.6	41	84
Proposed MoE with ANN experts ($K = 5$)					
A^*	MOE_ANN	17.5	1.3	71	95
e^*	MOE_ANN	19.5	1.8	63	89

Table 5 Results for D_3

Output	Type	\hat{E}_{\max} (%)	\hat{E}_{mean} (%)	$\alpha = 1\%$ (%)	$\alpha = 5\%$ (%)
Best surrogate model over the whole domain D_3					
A^*	MLS	47	3.4	34	83
e^*	ANN	30.5	2.9	38	84
MoE with ANN experts ($K = 6$)					
A^*	MOE_ANN	20.5	2.8	50	81
e^*	MOE_ANN	15.1	1.9	44	89

Table 6 Results for D_4

Output	Type	\hat{E}_{\max} (%)	\hat{E}_{mean} (%)	$\alpha = 1\%$ (%)	$\alpha = 5\%$ (%)
Best surrogate model over the whole domain D_4					
A^*	MLS	43	7	18.5	54
e^*	ANN	22	4.03	22	69
Proposed MoE with ANN experts ($K = 6$)					
A^*	MOE_ANN	29.5	3.7	39	73
e^*	MOE_ANN	17.3	3.1	35	80

Table 7 Overall results

Output	Type	\hat{E}_{\max} (%)	\hat{E}_{mean} (%)	$\alpha = 1\%$ (%)	$\alpha = 5\%$ (%)
Best surrogate model over the whole domain D					
A^*	ANN-MLS	47	3.5	43.4	80.3
e^*	ANN-MLS	30.5	2.68	42.5	83.3
Mixture of MoE's					
A^*	MOE_ANN	29.5	2.3	60	86.2
e^*	MOE_ANN	19.5	2.1	52	89

(thickness and cross-section area) do not vanish and therefore we do not divide by 0. In addition, in a previous work (Merval 2008) this kind of regression (regression of an optimization results) was already tested with respect to that error measure. We observe that the MoE technique always outperforms the best surrogate built on the whole sub-domain for all error measures. For the whole domain, where the global expert is a simple combination of all local models: MoE and ANN, we observe that the error decreases by about 34% for the section area and by 22% for the thickness.

6 Conclusion

We presented an original method to improve accuracy for regression of the objective and constraints functions that arise in structural optimization problems. Despite the functions to approximate are discontinuous, our method allows to build smooth approximations. Instead of assembling a global surrogate model, this method subdivides the global conjoint space by means of EM algorithm for Gaussian mixture models, designs a surrogate model on each of these sub-regions, combines all the surrogate models using the Gaussian parameters estimated through EM algorithm. This clustering can be hard or fuzzy. We implemented and tested the hard clustering version of this method where local surrogate models can be linear and quadratic regression, moving least squares and artificial neural networks. We also derived a classical surrogate model (weighted least squares in Appendix B) from the soft clustering version. This method is tested over mathematical test cases and an industrial case from aeronautics and is found to increase the accuracy of approximation when compared to a sole surrogate model built over the whole domain, especially in the case of a mixture of ANN's. However, as pointed it out in the first section, the literature on ensemble of surrogate models advocates to mix different surrogate models built over the whole learning basis. As far as smooth approximations of discontinuous functions are concerned, the method presented here needs to be compared with the

different methods of ensembles of surrogate models that can be found in the literature, this will be part of our future research. When smoothness is not required (in case for instance of derivative-free optimization, genetic or evolutionary optimization, ant colony optimization...), the hard mixture version that results in a discontinuous global predictor is more appropriate, this will be also part of our future research. Nonetheless, this method offers many advantages, since the training of the local experts can be performed concurrently and drastically reduces the size of the sample data when the number of clusters gets high. It also offers a very good quality even in the linear case (MOE_LIN) which is inexpensive to compute. It also provides local information on the quality of the regression in the hard-clustering case. Indeed, we know the clusters where the error is high. We also carried out a numerical procedure to find the optimum (or at least an appropriate) number of clusters K . As pointed out, one of the main drawbacks of our proposed algorithm is the large number of data needed for clustering. We could improve our technique by using the local information to enrich the clusters where the approximation is the worst. In addition, the accuracy could be further improved by updating the parameters of the local expert within EM algorithm or use a local procedure to get a better local expert (boosting Meir and Ratsch 2003). Lastly, this algorithm can be adapted to mixed input values (discrete and continuous input variables) by centering clusters on each possible discrete value, making an overall continuous surrogate models over mixed variables. Possible applications towards laminated composite material design and optimization are in progress.

Acknowledgments The authors wish to thank the anonymous reviewers for their insightful and constructive comments. They would like to extend their grateful thanks to L. Jones and J. R. Barron, from Airbus and to A. Shahdin, from ISAE-SupAéro for their careful reading of the manuscript.

Appendix A: EM algorithm

The EM algorithm aims at solving maximum likelihood parametric estimation problems. The EM algorithm was first described in Dempster et al. (1977). First, let us recall the well-known maximum-likelihood estimator. Suppose we have a set of observed data $\mathbf{X} = \{x_i\}_{i=1\dots N}$ coming from a random variable X . Actually, we shall consider that \mathbf{X} is a size N iid sample, i.e. that it is a realization of N independent identically distributed random elements with the same law than the random variable X . We assume that the law of X is parametric, governed by parameter θ , e.g for a 1D Gaussian variable X , the parameter θ is (μ, σ^2) the

mean and variance. We can therefore denote the probability density function $g_\theta(x)$. Now define the likelihood

$$\mathcal{L}(\theta; \mathbf{X}) = \prod_{i=1}^N g_\theta(x_i) \quad (23)$$

which is the probability density of the observed data. \mathcal{L} is a function of θ for \mathbf{X} fixed. The sample log-likelihood is

$$l(\theta; \mathbf{X}) = \sum_{i=1}^N \log g_\theta(x_i) \quad (24)$$

Maximum likelihood estimator is the value $\hat{\theta}$ of θ which maximizes $l(\theta; \mathbf{X})$. In simple applications, where the law of X belongs to a classical family as Gaussian, Poisson, Gamma, Student, exponential, the maximum-likelihood estimator is simple to compute and amounts to an analytic expression. But in a lot of modern problems, the law of interest is much more complex and can be readily represented by a mixture $\sum_k \alpha_k g(\theta_k, x)$. Notice that this law is the marginal law of X if we add a latent random finite variable K such that the random couple (K, X) is governed by

$$\forall k, \Omega, \mathbb{P}(K = k, X \in \Omega) = \alpha_k \int_{\Omega} g(\theta_k, x) dx$$

Now suppose that the latent variable K is observed, let us note $(\mathbf{K}, \mathbf{X}) = \{(K_i, X_i)_i\}$ this virtual sample and let \mathbf{X}_k be the sub-sample of \mathbf{X} of the data X_i for which the associate latent variable K_i is equal to k , let N_k be the size of \mathbf{X}_k . Then it is easy to get the maximum likelihood estimation for θ_k and α_k , namely

$$\hat{\theta}_k = \operatorname{argmax}_{\theta} l(\theta; \mathbf{X}_k), \quad \hat{\alpha}_k = \frac{N_k}{N}$$

Actually, we do not know the K_i 's whenever they have a physical existence. If the parameter set $\{\theta_k, \alpha_k\}$ was known, K_i could be recovered through a Bayes posterior estimation

$$\gamma_{i,k}(\theta) = \mathbb{P}(K_i = k | \theta, X = X_i) \quad (25)$$

which is called the responsibility of model k for observation i . Indeed, an estimation of the latent variable may be estimated through the MAP (maximum a posteriori) estimation

$$\hat{K}_i = \operatorname{argmax}_k \gamma_{i,k}(\theta)$$

A similar method is used in the proposed technique for ‘‘hard clustering’’.

We can now derive the EM algorithm as an iterative relaxation of these two steps. Let us describe the $(n + 1)$ -th iteration

1. Take the estimates at the previous step $\{(\hat{\theta}_k^n, \hat{\alpha}_k^n)_k\}$
2. Expectation step. Compute associate responsibilities $\hat{\gamma}_{i,k}^n$ for $i = 1 \dots N$ and $k = 1 \dots m$:

$$\hat{\gamma}_{i,k}^n = \frac{\hat{\alpha}_k^n g_{\hat{\theta}_k^n}(x_i)}{\sum_{j=1}^m \hat{\alpha}_j^n g_{\hat{\theta}_j^n}(x_i)} \quad (26)$$

3. Maximization step. Compute the weighted maximum likelihood estimators for each component of the mixture:

$$\hat{\theta}_k^{n+1} = \frac{\sum_{i=1}^N \hat{\gamma}_{i,k}^n x_i}{\sum_{i=1}^N \hat{\gamma}_{i,k}^n} \quad (27)$$

$$\hat{\alpha}_k^{n+1} = \frac{\sum_{i=1}^N \hat{\gamma}_{i,k}^n}{\sum_{j=1}^m \sum_{i=1}^N \hat{\gamma}_{i,j}^n} \quad (28)$$

The convergence of this algorithm is proven in Wu (1983).

Appendix B: An MoE interpretation of weighted least squares

We give here an interpretation of the weighted least squares regression. Say we have $\mathcal{X} = (x_1, \dots, x_n) \in \mathbb{R}^d$ together with their output values $\mathcal{Y} = (y_1, \dots, y_n)$. A local Weighted Least Squares regression at point $\hat{x} \in \mathbb{R}^d$ consists in finding the best multivariate polynomial approximation $f_{\hat{x}}$:

$$f_{\hat{x}} = \operatorname{argmin}_{p \in \Pi_s^d} \sum_{i=1}^n \theta(\|x_i - \hat{x}\|) \|p(x_i) - y_i\|^2 \quad (29)$$

where Π_s^d is the set of multivariate polynomials of d -variables and degree s and θ is a symmetric positive and decreasing function: maximum value at 0 and 0 when $|x| \rightarrow +\infty$: which can be strictly positive or compactly-supported. θ is called a weight function. A very popular choice for θ is

$$\theta(r) = \frac{1}{\sqrt{(2\pi)^d \sigma^{2d}}} e^{-\frac{1}{2} \frac{r^2}{\sigma^2}} \quad (30)$$

This regression $f_{\hat{x}}$ is a local approximation which is valid only in the neighborhood of \hat{x} . Note that the moving least squares regression consists in a global weighted least squares by making the \hat{x} varying (or moving) over the whole

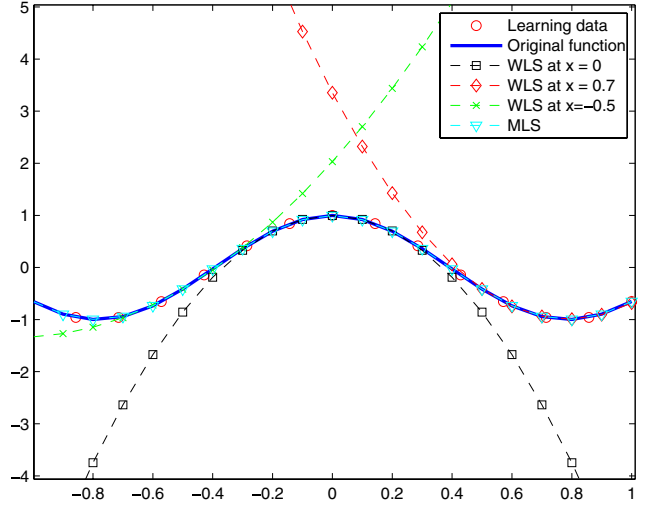


Fig. 7 Weighted least squares regressions (polynomials of degree 2 and $\sigma = 0.1$) of the function $x \rightarrow \cos(4x)$ at different points and moving least squares regression

domain \mathbb{R}^d , this global model happens to be continuous and even smooth whenever the weight function is smooth (Levin 1998). We depicted several WLS regression on different points together with the MLS regression (Fig. 7). Another way to get a global model is by using a partition of unity. Namely, we have K so called support points $\hat{x}_1, \dots, \hat{x}_K$, a local WLS regression $\hat{f}_{\hat{x}_j} = \hat{f}_j$ is built at each support point \hat{x}_j and the global model is

$$F_{WLS}(x) = \sum_{i=1}^K \beta_i(x) \hat{f}_i(x) \quad (31)$$

where

$$\beta_j(x) = \frac{e^{-\frac{1}{2} \frac{\|x - \hat{x}_j\|^2}{\sigma^2}}}{\sum_{i=1}^K e^{-\frac{1}{2} \frac{\|x - \hat{x}_i\|^2}{\sigma^2}}} \quad (32)$$

This popular model has an obvious interpretation in terms of MoE and soft clustering. Assume that the law⁵ of $X \sim \sum_{j=1}^K \frac{1}{K} \mathcal{N}_d(\hat{x}_j, \sigma^2 I_d)$. In that case, we have, keeping the same notations as in Section 2

$$\mathbb{P}(X = x | j) = \frac{1}{\sqrt{(2\pi)^d \sigma^{2d}}} e^{-\frac{1}{2} \frac{\|x - \hat{x}_j\|^2}{\sigma^2}} \quad (33)$$

⁵In this article we focused on Gaussian mixture models that were fully free, i.e. all the parameters of the Gaussian mixture models are not constrained and EM algorithm estimates all the parameters. There are more simple Gaussian mixture models that assume that all the means are the same or all of the variance-covariance matrices are of the form $\sigma^2 I_n$ (this hypothesis is known in statistics as homoscedascity).

and

$$\mathbb{P}(j|X = x) = \frac{e^{-\frac{1}{2} \frac{\|x - \hat{x}_j\|^2}{\sigma^2}}}{\sum_{i=1}^K e^{-\frac{1}{2} \frac{\|x - \hat{x}_i\|^2}{\sigma^2}}} \quad (34)$$

and we see that

$$F_{WLS}(x) = \sum_{i=1}^K \mathbb{P}(\kappa = i|X = x) \hat{f}_i(x) \quad (35)$$

and the WLS model is an MoE for $X \sim \sum_{j=1}^K \frac{1}{K} \mathcal{N}_d(\hat{x}_j, \sigma^2 I_d)$ using a soft partitioning and where local experts are multivariate polynomial regression experts weighted by the probabilities of the soft partitioning.

Appendix C: Samgen: a sample generator of discontinuous functions for surrogate modeling

When studying regression methods, one often builds surrogate models over closed-form functions to assess the accuracy of the surrogate. We wanted to explore the accuracy of surrogate models on discontinuous functions in high dimension (>8). We ended up with the need of such analytical functions, Samgen was created to generate such functions and more precisely functions that mimic the behavior of aeronautical stress tools. Samgen is a simple Matlab code that generates a structure that contains all the data needed to compute a discontinuous function in arbitrary dimension over $[0, 1]^d$ (Fig. 8). It simply separates with hyper-planes the unit hyper-cube $[0, 1]^d$ into the desired number of dis-

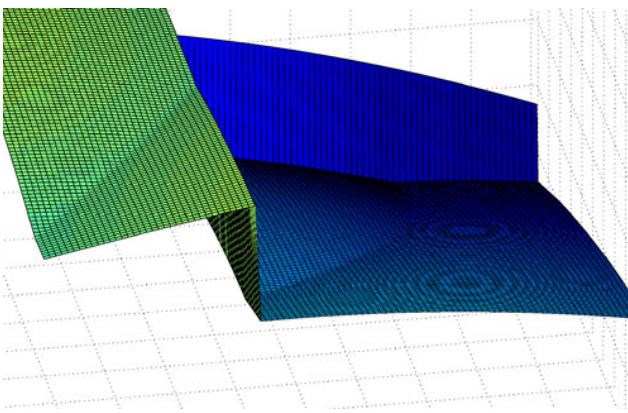


Fig. 8 Example of a C^0 and C^1 discontinuous 2D function created with Samgen

continuities and over each sub-domain generates two non-convex quadratic functions and compute the maximum of them to generate derivative-discontinuities.

References

- Acar E, Rais-Rohani M (2009) Ensemble of metamodels with optimized weight factors. *Struct Multidisc Optim* 37(3):279–294
- Berkhin P (2002) Survey of clustering data mining techniques
- Bradley PS, Fayyad U, Reina C (1998) Scaling EM (expectation-maximization) clustering to large databases. Microsoft Research Report, MSR-TR-98-35
- Buhmann MD (2001) Radial basis functions. *Acta Numer* 9:1–38
- Burnham KP, Anderson DR (2004) Multimodel inference: understanding AIC and BIC in model selection. *Sociol Methods Res* 33(2):261
- Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc B (Methodological)* 39:1–38
- Dreyfus G (2005) *Neural networks: methodology and applications*. Springer, Berlin
- Fasshauer GE (2005) Dual bases and discrete reproducing kernels: a unified framework for RBF and MLS approximation. *Eng Anal Bound Elem* 29(4):313–325
- Forrester AIJ, Keane AJ (2009) Recent advances in surrogate-based optimization. *Prog Aerosp Sci* 45(1–3):50–79
- Friedman J, Hastie T, Tibshirani R (2001) *The elements of statistical learning*
- Haftka RT, Gurdal Z (1992) *Elements of structural optimization*. Kluwer
- Haykin S (2008) *Neural networks: a comprehensive foundation*. Prentice Hall
- Jordan MI, Jacobs RA (1994) Hierarchical mixtures of experts and the EM algorithm. *Neural Comput* 6(2):181–214
- Kleijnen JPC, Sanchez SM, Lucas TW, Cioppa TM (2005) A user's guide to the brave new world of designing simulation experiments. *INFORMS J Comput* 17(3):263–289
- Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *International joint conference on artificial intelligence*, vol 14. Citeseer, pp 1137–1145
- Levin D (1998) The approximation power of moving least-squares. *Math Comput* 67(224):1517–1532
- Liu B, Haftka RT, Watson LT (2004) Global-local structural optimization using response surfaces of local optimization margins. *Struct Multidisc Optim* 27(5):352–359
- Meir R, Ratsch G (2003) An introduction to boosting and leveraging. *Lect Notes Comput Sci* 2600:118–183
- Merval A (2008) *Application des modèles réduits à l'optimisation multi-niveau de structures aéronautiques*. Thèse Supaéro
- Merval A, Samuelides M, Grihon S (2006) Application of response surface methodology to stiffened panel optimization. In: *47th AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics, and materials conference*, 1–4 May 2006, Newport, RI
- Myers RH, Montgomery DC, Anderson-Cook CM (2009) *Response surface methodology, process and product optimization using designed experiments*. Wiley
- Nealen A, Darmstadt TU (2004) An as-short-as-possible introduction to the least squares, weighted least squares and moving least squares methods for scattered data approximation and interpolation. URL: <http://www.nealen.com/projects>
- Picard RR, Cook RD (1984) Cross-validation of regression models. *J Am Stat Assoc* 79(387):575–583

- Sanchez E, Pintos S, Queipo NV (2008) Toward an optimal ensemble of kernel-based approximations with engineering applications. *Struct Multidisc Optim* 36(3):247–261
- Simpson TW, Toropov V, Balabanov V, Viana FAC (2008) Design and analysis of computer experiments in multidisciplinary design optimization: a review of how far we have come or not. In: Proceedings of the 12th AIAA/ISSMO multidisciplinary analysis and optimization conference, 2008 MAO. Victoria, Canada
- Smola AJ, Schölkopf B (2004) A tutorial on support vector regression. *Stat Comput* 14(3):199–222
- van Beers WCM, Kleijnen JPC (2004) Kriging interpolation in simulation: a survey. In: Proceedings of the 36th conference on winter simulation. Winter simulation conference, pp 113–121
- Viana FAC, Haftka RT, Steffen V (2009) Multiple surrogates: how cross-validation errors can help us to obtain the best predictor. *Struct Multidisc Optim* 39(4):439–457
- Wang GG, Shan S (2007) Review of metamodeling techniques in support of engineering design optimization. *J Mech Des* 129:370
- Wu CFJ (1983) On the convergence properties of the EM algorithm. *Ann Stat* 11(1):95–103
- Yang Y (2003) Regression with multiple candidate models: selecting or mixing? *Stat Sin* 13(3):783–810
- Zerpa LE, Queipo NV, Pintos S, Salager JL (2005) An optimization methodology of alkaline–surfactant–polymer flooding processes using field scale numerical simulation and multiple surrogates. *J Pet Sci Eng* 47(3–4):197–208

4.4.2 Improving kriging surrogates of high-dimensional design models by Partial Least Squares dimension reduction, Structural and Multidisciplinary Optimization 2016

Improving kriging surrogates of high-dimensional design models by Partial Least Squares dimension reduction

Mohamed Amine Bouhlel¹ · Nathalie Bartoli² · Abdelkader Otsmane¹ · Joseph Morlier³

Received: 20 June 2014 / Revised: 18 December 2015 / Accepted: 22 December 2015
© Springer-Verlag Berlin Heidelberg 2016

Abstract Engineering computer codes are often computationally expensive. To lighten this load, we exploit new covariance kernels to replace computationally expensive codes with surrogate models. For input spaces with large dimensions, using the kriging model in the standard way is computationally expensive because a large covariance matrix must be inverted several times to estimate the parameters of the model. We address this issue herein by constructing a covariance kernel that depends on only a few parameters. The new kernel is constructed based on information obtained from the Partial Least Squares method. Promising results are obtained for numerical examples with up to 100 dimensions, and significant computational gain is obtained while maintaining sufficient accuracy.

Keywords Kriging · Partial Least Squares · Experiment design · Metamodels

Symbols and notation

Matrices and vectors are in bold type.

Symbol	Meaning
\det	Determinant of a matrix
$ \cdot $	Absolute value
\mathbb{R}	Set of real numbers
\mathbb{R}^+	Set of positive real numbers
n	Number of sampling points
d	Dimensions
h	Number of principal components retained
\mathbf{x}	$1 \times d$ vector
\mathbf{x}_j	j^{th} element of a vector \mathbf{x}
\mathbf{X}	$n \times d$ matrix containing sampling points
\mathbf{y}	$n \times 1$ vector containing simulation of \mathbf{X}
$\mathbf{x}^{(i)}$	i^{th} training point for $i = 1, \dots, n$ (a $1 \times d$ vector)
$\mathbf{w}^{(l)}$	$d \times 1$ vector containing \mathbf{X} weights given by the l^{th} PLS iteration for $l = 1, \dots, h$
$\mathbf{X}^{(0)}$	\mathbf{X}
$\mathbf{X}^{(l-1)}$	Matrix containing residual of inner regression of $(l-1)^{\text{st}}$ PLS iteration for $l = 1, \dots, h$
$k(\cdot, \cdot)$	Covariance function
$\mathcal{N}(\mathbf{0}, k(\cdot, \cdot))$	Distribution of a Gaussian process with mean function $\mathbf{0}$ and covariance function $k(\cdot, \cdot)$
\mathbf{x}^t	Superscript t denotes the transpose operation of the vector \mathbf{x}

✉ Mohamed Amine Bouhlel
mohamed.bouhlel@onera.fr;
mohamed.amine.bouhlel@gmail.com

Nathalie Bartoli
nathalie.bartoli@onera.fr

Abdelkader Otsmane
abdelkader.otsmane@sncma.fr

Joseph Morlier
joseph.morlier@isae.fr

- ¹ SNECMA, Rond-point René Ravaud-Réau, 77550 Moissy-Cramayel, France
- ² ONERA, 2 Avenue Édouard Belin, 31055 Toulouse, France
- ³ Université de Toulouse, CNRS, Institut Clément Ader, ISAE-SUPAERO, 10 Avenue Edouard Belin, 31055 Toulouse Cedex 4, France

1 Introduction and main contribution

In recent decades, because simulation models have striven to more accurately represent the true physics of phenomena, computational tools in engineering have become ever more complex and computationally expensive. To address this new challenge, a large number of input design variables, such as geometric representation, are often considered. Thus, to analyze the sensitivity of input design variables or to search for the best point of a physical objective under certain physical constraints (i.e., global optimization), a large number of computing iterations are required, which is impractical when using simulations in real time. This is the main reason that surrogate modeling techniques have been growing in popularity in recent years. Surrogate models, also called metamodels, are vital in this context and are widely used as substitutes for time-consuming high-fidelity models. They are mathematical tools that approximate coded simulations of a few well-chosen experiments that serve as models for the design of experiments. The main role of surrogate models is to describe the underlying physics of the phenomena in question. Different types of surrogate models can be found in the literature, such as regression, smoothing spline (Wahba and Craven 1978; Wahba 1990), neural networks (Haykin 1998), radial basis functions (Buhmann 2003) and Gaussian-process modeling (Rasmussen and Williams 2006).

In this article, we focus on the kriging model because it estimates the prediction error. This model is also referred to as the Gaussian-process model (Rasmussen and Williams 2006) and was presented first in geostatistics (see, e.g., Cressie 1988 or Goovaerts 1997) before being extended to computer experiments and machine learning (Schonlau 1998; Sasena 2002; Jones et al. 1998; Picheny et al. 2010). The kriging model has become increasingly popular due to its flexibility in accurately imitating the dynamics of computationally expensive simulations and its ability to estimate the error of the predictor. However, it suffers from some well-known drawbacks in high dimension, which may be due to multiple causes. For starters, the size of the covariance matrix of the kriging model may increase dramatically if the model requires a large number of sample points. As a result, inverting the covariance matrix is computationally expensive. The second drawback is the optimization of the subproblem, which involves estimating the hyper-parameters for the covariance matrix. This is a complex problem that requires inverting the covariance matrix several times. Some recent works have addressed the drawbacks of high-dimensional Gaussian processes (Hensman et al. 2013; Damianou and Lawrence 2013; Durrande et al. 2012) or the large-scale sampling of data (Sakata et al. 2004). One way to reduce CPU time when constructing a

kriging model is to reduce the number of hyper-parameters, but this approach assumes that the kriging model exhibits the same dynamics in all directions (Mera 2007).

Thus, because estimating the kriging parameters can be time consuming, especially with dimensions as large as 100, we present herein a new method that combines the kriging model with the Partial Least Squares (PLS) technique to obtain a fast predictor. Like the method of principle components analysis (PCA), the PLS technique reduces dimension and reveals how inputs depend on outputs. PLS is used in this work because PCA only exposes dependencies between inputs. Information given by PLS is integrated in the covariance structure of the kriging model to reduce the number of hyper-parameters. The combination of kriging and PLS is abbreviated KPLS and allows us to build a fast kriging model because it requires fewer hyper-parameters in its covariance function; all without eliminating any input variables from the original problem. In general, the number of kriging parameters is equal to the number of dimension which is reduced to at maximum 4 parameters with our approach.

The KPLS methods is used for many academic and industrial verifications, and promising results have been obtained for problems with up to 100 dimensions. The cases used in this paper do not exceed 100 input variables, which should be quite sufficient for most engineering problems. Problems with more than 100 inputs may lead to memory difficulties with the toolbox Scikit-learn (version 0.14), on which the KPLS method is based.

This paper is organized as follows: Section 2 summarizes the theoretical basis of the universal kriging model, recalling the key equations. The proposed KPLS model is then described in detail in Section 3 by using the kriging equations. Section 4 compares and analyzes the results of the KPLS model with those of the kriging model when applied to classic analytical examples and some complex engineering examples. Finally, Section 5 concludes and gives some perspectives.

2 Universal kriging model

To understand the mathematics of the proposed methods, we first review the kriging equations. The objective is to introduce the notation and to briefly describe the theory behind the kriging model. Assume that we have evaluated a cost deterministic function at n points $\mathbf{x}^{(i)}$ ($i = 1, \dots, n$) with

$$\mathbf{x}^{(i)} = [\mathbf{x}_1^{(i)}, \dots, \mathbf{x}_d^{(i)}] \in B \subset \mathbb{R}^d,$$

and we denote by \mathbf{X} the matrix $[\mathbf{x}^{(1)t}, \dots, \mathbf{x}^{(n)t}]^t$. For simplicity, B is considered to be a hypercube expressed by

the product between intervals of each direction space, i.e., $B = \prod_{j=1}^d [a_j, b_j]$, where $a_j, b_j \in \mathbb{R}$ with $a_j \leq b_j$ for $j = 1, \dots, d$. Simulating these n inputs gives the outputs $\mathbf{y} = [y^{(1)}, \dots, y^{(n)}]^t$ with $y^{(i)} = y(\mathbf{x}^{(i)})$ for $i = 1, \dots, n$. We use $\hat{y}(\mathbf{x})$ to denote the prediction of the true function $y(\mathbf{x})$ which is considered as a realization of a stochastic process $Y(\mathbf{x})$ for all $\mathbf{x} \in B$. For the universal kriging model (Roustant et al. 2012; Picheny et al. 2010), Y is written as

$$Y(\mathbf{x}) = \sum_{j=1}^m \beta_j f_j(\mathbf{x}) + Z(\mathbf{x}), \tag{1}$$

where, for $j = 1, \dots, m$, f_j is a known independent basis function, $\beta_j \in \mathbb{R}$ is an unknown parameter, and Z is a random variable defined by $Z(\mathbf{x}) \sim \mathcal{N}(0, k)$, with k being a stationary covariance function, also called a covariance kernel. The kernel function k can be written as

$$k(\mathbf{x}, \mathbf{x}') = \sigma^2 r(\mathbf{x}, \mathbf{x}') = \sigma^2 r_{\mathbf{xx}'} \forall \mathbf{x}, \mathbf{x}' \in B, \tag{2}$$

where σ^2 is the process variance and $r_{\mathbf{xx}'}$ is the correlation function between \mathbf{x} and \mathbf{x}' . However, the correlation function r depends on some hyper-parameters θ and it is considered to be known. We also denote the $n \times 1$ vector as $\mathbf{r}_{\mathbf{xx}} = [r_{\mathbf{xx}^{(1)}}, \dots, r_{\mathbf{xx}^{(n)}}]^t$ and the $n \times n$ covariance matrix as $\mathbf{R} = [\mathbf{r}_{\mathbf{x}^{(1)}\mathbf{x}}, \dots, \mathbf{r}_{\mathbf{x}^{(n)}\mathbf{x}}]$.

2.1 Derivation of prediction formula

Under the hypothesis above, the best linear unbiased predictor for $y(\mathbf{x})$, given the observations \mathbf{y} , is

$$\hat{y}(\mathbf{x}) = \mathbf{f}(\mathbf{x})^t \hat{\beta} + \mathbf{r}_{\mathbf{xx}'}^t \mathbf{R}^{-1} (\mathbf{y} - \mathbf{F} \hat{\beta}), \tag{3}$$

where $\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}), \dots, f_m(\mathbf{x})]^t$ is the $m \times 1$ vector of basis functions, $\mathbf{F} = [\mathbf{f}(\mathbf{x}^{(1)}), \dots, \mathbf{f}(\mathbf{x}^{(n)})]^t$ is the $n \times m$ matrix, and $\hat{\beta}$ is the vector of generalized least-square estimates of $\beta = [\beta_1, \dots, \beta_m]^t$, which is given by

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_m \end{bmatrix} = (\mathbf{F}^t \mathbf{R}^{-1} \mathbf{F})^{-1} \mathbf{F}^t \mathbf{R}^{-1} \mathbf{y}. \tag{4}$$

Moreover, the universal kriging model provides an estimate of the variance of the prediction, which is given by

$$s^2(\mathbf{x}) = \hat{\sigma}^2 \left(1 - \mathbf{r}_{\mathbf{xx}'}^t \mathbf{R}^{-1} \mathbf{r}_{\mathbf{xx}} \right), \tag{5}$$

with

$$\hat{\sigma}^2 = \frac{1}{n} (\mathbf{y} - \mathbf{F} \hat{\beta})^t \mathbf{R}^{-1} (\mathbf{y} - \mathbf{F} \hat{\beta}). \tag{6}$$

For more details of the derivation of the prediction formula, see, for instance, Sasena (2002) or Schonlau (1998). The theory of the proposed method has been expressed in the same way as for the universal kriging model. The

numerical examples in Section 4 use the ordinary kriging model, which is a special case of the universal model, but with $\mathbf{f}(\mathbf{x}) = \{1\}$ (and $m = 1$). For the ordinary kriging model, (3), (4), and (6) are then replaced by the equations given in Appendix A.

Note that the assumption of known covariance with known hyper-parameters θ is unrealistic in reality. For this reason, the covariance function is typically chosen from among a parametric family of kernels. Table 12 in Appendix B gives some examples of typical stationary kernels. The number of hyper-parameters required for the estimate is typically greater than (or equal to) the number of input variables. In this work, we use in the following a Gaussian exponential kernel:

$$k(\mathbf{x}, \mathbf{x}') = \sigma^2 \prod_{i=1}^d \exp \left(-\theta_i (\mathbf{x}_i - \mathbf{x}'_i)^2 \right) \forall \theta_i \in \mathbb{R}^+.$$

By applying some elementary operations to existing kernels, we can construct new kernels. In this work, we use the property that the tensor product of covariances is a covariance kernel in the product space. More details are available in Rasmussen and Williams (2006), Durrande (2011), Bishop (2007), Liem and Martins (2014).

2.2 Estimation of hyper-parameters θ

The key point of the kriging approximation is how it estimates the hyper-parameters θ , so its main steps are recalled here, along with some mathematical details.

One of the major challenges when building a kriging model is the complexity and difficulty of estimating the hyper-parameters θ , in particular when dealing with problems with many dimensions or with a large number of sampling points. In fact, using (3) to make a kriging prediction requires inverting an $n \times n$ matrix, which typically has cost of $\mathcal{O}(n^3)$, where n is the number of sampling points (Braham et al. 2014). The hyper-parameters are estimated using maximum likelihood (ML) or cross validation (CV), which are based on observations. Bachoc compared the ML and CV techniques (Bachoc 2013) and concluded that, in most cases studied, the CV variance is larger. The ML method is widely used to estimate the hyper-parameters θ ; it is also used in this paper. In practice, the following log-ML estimate is often used:

$$\log \text{-ML}(\theta) = -\frac{1}{2} \left[n \ln(2\pi\sigma^2) + \ln(\det \mathbf{R}(\theta)) + (\mathbf{y} - \mathbf{F}\beta)^t \mathbf{R}(\theta)^{-1} (\mathbf{y} - \mathbf{F}\beta) / \sigma^2 \right]. \tag{7}$$

Inserting $\hat{\beta}$ and $\hat{\sigma}^2$ given by (4) and (6), respectively, into the expression (7), we get the following so-called

concentrated likelihood function, which depends only on the hyper-parameters θ :

$$\begin{aligned} \log\text{-ML}(\theta) &= -\frac{1}{2}[n \ln \hat{\sigma}^2 + \ln \det \mathbf{R}(\theta)] \\ &= -\frac{1}{2} \left[n \ln \left(\frac{1}{n} (\mathbf{y} - \mathbf{F}(\mathbf{F}'\mathbf{R}^{-1}\mathbf{F})^{-1}\mathbf{F}'\mathbf{R}^{-1}\mathbf{y})' \right. \right. \\ &\quad \left. \left. \times \mathbf{R}^{-1}(\mathbf{y} - \mathbf{F}(\mathbf{F}'\mathbf{R}^{-1}\mathbf{F})^{-1}\mathbf{F}'\mathbf{R}^{-1}\mathbf{y}) \right) \right. \\ &\quad \left. + \ln \det \mathbf{R} \right]. \end{aligned} \tag{8}$$

To facilitate reading, $\mathbf{R}(\theta)$ has been replaced by \mathbf{R} in the last line of (8).

Maximizing (8) is very computationally expensive for high dimensions and when using a large number of sample points because the $(n \times n)$ matrix $\mathbf{R}(\theta)$ in (8) must be inverted. The maximization problem is often solved using genetic algorithms (see Forrester and Sobester 2008 for more details). In this work, we use the derivative-free optimization algorithm COBYLA that was developed by Powell (1994). COBYLA is a sequential trust-region algorithm that uses linear approximations for the objective and constraint functions.

Figure 1 recalls the principal stages of building a kriging model, and each step is briefly outlined below:

1. The user must provide the initial design of experiments (\mathbf{X}, \mathbf{y}) and the parametric family of the covariance function k .
2. To derive the prediction formula, the kriging algorithm assumes that all parameters of k are known.

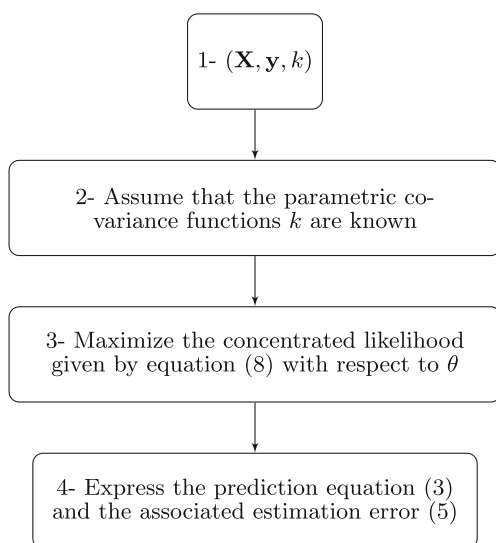


Fig. 1 The main steps for building an ordinary kriging model

3. Under the hypothesis of the kriging algorithm, we estimate hyper-parameters θ from the concentrated likelihood function given by (8) and by using the COBYLA algorithm.
4. Finally, we calculate the prediction (3) and the associated estimation error (5) after estimating all hyper-parameters of the kriging model.

3 Kriging model combined with Partial Least Squares

As explained above, maximizing the concentrated likelihood (8) can be time consuming when the number of covariance parameters is large, which typically occurs in large dimension. Solving this problem can be accelerated by combing the PLS method and the kriging model. The θ parameters from the kriging model represent the range in any spatial direction. Assuming, for instance, that certain values are less significant for the response, then the corresponding θ_i ($i = 1, \dots, d$) will be very small compared to the other θ parameters. The PLS method is a well-known tool for high-dimensional problems and consists of maximizing the variance by projecting onto smaller dimensions while monitoring the correlation between input variables and the output variable. In this way, the PLS method reveals the contribution of all variables—the idea being to use this information to scale the θ parameters.

In this section we propose a new method that can be used to build an efficient kriging model by using the information extracted from the PLS stage. The main steps for this construction are as follows:

1. Use PLS to define weight parameters.
2. To reduce the number of hyper-parameters, define a new covariance kernel by using the PLS weights.
3. Optimize the parameters.

The key mathematical details of this construction are explained in the following.

3.1 Linear transformation of covariance kernels

Let \mathbf{x} be a vector space over the hypercube B . We define a linear map given by

$$\begin{aligned} F : B &\longrightarrow B', \\ \mathbf{x} &\longmapsto [\alpha_1 \mathbf{x}_1, \dots, \alpha_d \mathbf{x}_d], \end{aligned} \tag{9}$$

where $\alpha_1, \dots, \alpha_d \in \mathbb{R}$ and B' is a hypercube included in \mathbb{R}^d (B' can be different from B). Let k be an isotropic covariance kernel with $k : B' \times B' \rightarrow \mathbb{R}$. Since k is isotropic, the covariance kernel $k(F(\cdot), F(\cdot))$

depends on a single parameter, which must be estimated. However, if $\alpha_1, \dots, \alpha_d$ are well chosen, then we can use $k(F(\cdot), F(\cdot))$. In this case, the linear transformation F allows us to approach the isotropic case (Zimmerman and Homer 1991). In the present work, we choose $\alpha_1, \dots, \alpha_d$ based on information extracted from the PLS technique.

3.2 Partial Least Squares

The PLS method is a statistical method that finds a linear relationship between input variables and the output variable by projecting input variables onto a new space formed by newly chosen variables, called principal components (or latent variables), which are linear combinations of the input variables. This approach is particularly useful when the original data are characterized by a large number of highly collinear variables measured on a small number of samples. Below, we briefly describe how the method works. For now, suffice it to say that only the weighting coefficients are central to understanding the new KPLS approach. For more details on the PLS method, please see Helland (1988), Frank and Friedman (1993), Alberto and González (2012).

The PLS method is designed to search out the best multidimensional direction in \mathbf{X} space that explains the characteristics of the output \mathbf{y} . After centering and scaling

the $(n \times d)$ -sample matrix \mathbf{X} and the response vector \mathbf{y} , the first principal component $\mathbf{t}^{(1)}$ is computed by seeking the best direction $\mathbf{w}^{(1)}$ that maximizes the squared covariance between $\mathbf{t}^{(1)} = \mathbf{X}\mathbf{w}^{(1)}$ and \mathbf{y} :

$$\mathbf{w}^{(1)} = \begin{cases} \arg \max_{\mathbf{w}} \mathbf{w}^t \mathbf{X}^t \mathbf{y} \mathbf{y}^t \mathbf{X} \mathbf{w} \\ \text{such that } \mathbf{w}^t \mathbf{w} = 1. \end{cases} \quad (10)$$

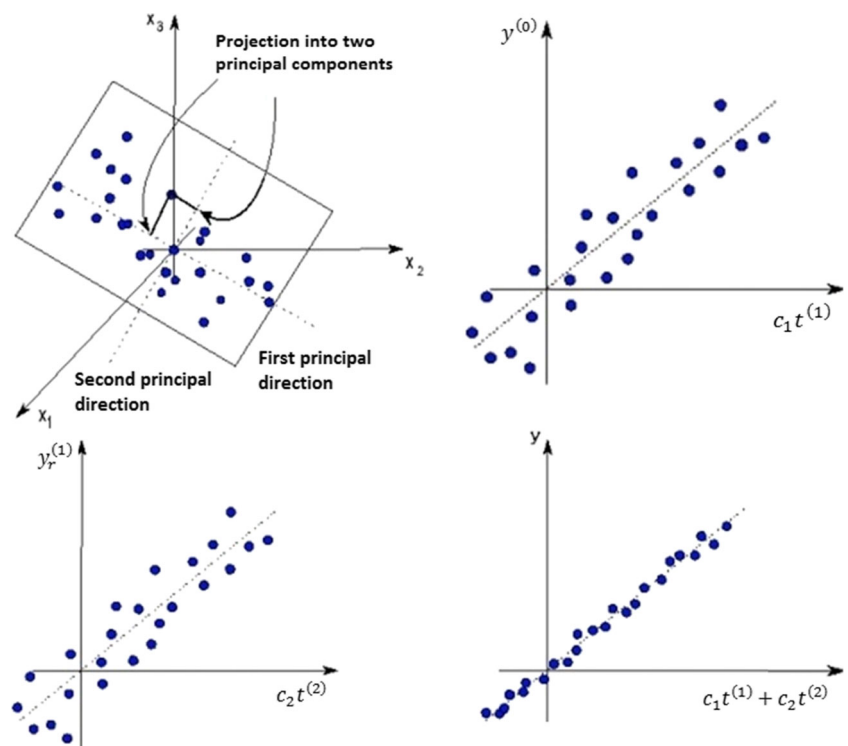
The optimization problem (10) is maximized when $\mathbf{w}^{(1)}$ is the eigenvector of the matrix $\mathbf{X}^t \mathbf{y} \mathbf{y}^t \mathbf{X}$ corresponding to the eigenvalue with the largest absolute value; the vector $\mathbf{w}^{(1)}$ contains the \mathbf{X} weights of the first component. The largest eigenvalue of problem (10) can be estimated by the power iteration method introduced by Lanczos (1950).

Next, the residual matrix from $\mathbf{X} = \mathbf{X}^{(0)}$ space and from $\mathbf{y} = \mathbf{y}^{(0)}$ are calculated; these are denoted $\mathbf{X}^{(1)}$ and $\mathbf{y}^{(1)}$, respectively:

$$\begin{aligned} \mathbf{X}^{(1)} &= \mathbf{X}^{(0)} - \mathbf{t}^{(1)} \mathbf{p}^{(1)}, \\ \mathbf{y}^{(1)} &= \mathbf{y}^{(0)} - c_1 \mathbf{t}^{(1)}, \end{aligned} \quad (11)$$

where $\mathbf{p}^{(1)}$ (a $1 \times d$ vector) contains the regression coefficients of the local regression of \mathbf{X} onto the first principal component $\mathbf{t}^{(1)}$, and c_1 is the regression coefficient of the local regression of \mathbf{y} onto the first principal component $\mathbf{t}^{(1)}$.

Fig. 2 Upper left shows construction of two principal components in \mathbf{X} space. Upper right shows prediction of $y^{(0)}$. Bottom left shows prediction of $y^{(1)}$. Bottom right shows final prediction of y



The system (11) is the local regression of \mathbf{X} and \mathbf{y} onto the first principal component.

Next, the second principal component, which is orthogonal to the first, can be sequentially computed by replacing \mathbf{X} by $\mathbf{X}^{(1)}$ and \mathbf{y} by $\mathbf{y}^{(1)}$ to solve the system (10). The same approach is used to iteratively compute the other principal components. To illustrate this process, a simple three-dimensional (3D) example with two principal components is given in Fig. 2. In the following, we use h to denote the number of principal components retained.

The principal components represent the new coordinate system obtained upon rotating the original system with axes, $\mathbf{x}_1, \dots, \mathbf{x}_d$ (Alberto and González 2012). For $l = 1, \dots, h$, $\mathbf{t}^{(l)}$ can be written as

$$\mathbf{t}^{(l)} = \mathbf{X}^{(l-1)}\mathbf{w}^{(l)} = \mathbf{X}\mathbf{w}_*^{(l)}. \tag{12}$$

This important relationship is used for coding the method. The following matrix $\mathbf{W}_* = [\mathbf{w}_*^{(1)}, \dots, \mathbf{w}_*^{(h)}]$ is obtained by Manne (1987):

$$\mathbf{W}_* = \mathbf{W}(\mathbf{P}'\mathbf{W})^{-1},$$

where $\mathbf{W} = [\mathbf{w}^{(1)}, \dots, \mathbf{w}^{(h)}]$ and $\mathbf{P} = [\mathbf{p}^{(1)t}, \dots, \mathbf{p}^{(h)t}]$.

The vector $\mathbf{w}^{(l)}$ corresponds to the principal direction in X space that maximizes the covariance of $\mathbf{X}^{(l-1)t}\mathbf{y}^{(l-1)}\mathbf{y}^{(l-1)t}\mathbf{X}^{(l-1)}$. If $h = d$, the matrix $\mathbf{W}_* = [\mathbf{w}_*^{(1)}, \dots, \mathbf{w}_*^{(d)}]$ rotates the coordinate space $(\mathbf{x}_1, \dots, \mathbf{x}_d)$ to the new coordinate space $(t^{(1)}, \dots, t^{(d)})$, which follow the principal directions $\mathbf{w}^{(1)}, \dots, \mathbf{w}^{(d)}$.

As mentioned in the introduction, the PLS method is chosen instead of the PCA method because the PLS method shows how the output variable depends on the input variables, whereas the PCA method focuses only on how the input variables depend on each other. In fact, the hyper-parameters θ for the kriging model depend on how each input variable affects the output variable.

3.3 Construction of new kernels for KPLS models

Let B be a hypercube included in \mathbb{R}^d . As seen in the previous section, the vector $\mathbf{w}_*^{(1)}$ is used to build the first principal component $\mathbf{t}^{(1)} = \mathbf{X}\mathbf{w}_*^{(1)}$, where covariance between $\mathbf{t}^{(1)}$ and \mathbf{y} is maximized. The scalars $\mathbf{w}_{*1}^{(1)}, \dots, \mathbf{w}_{*d}^{(1)}$ can then be interpreted as measuring the importance of $\mathbf{x}_1, \dots, \mathbf{x}_d$, respectively, for constructing the first principal component where its correlation with the output y is maximized. However, we know that the hyper-parameters $\theta_1, \dots, \theta_d$ (see Table 12 in Appendix B) can be interpreted as measuring how strongly the variables $\mathbf{x}_1, \dots, \mathbf{x}_d$, respectively, affect the output y . Thus, we define a new kernel $k_{kpls1} : B \times B \rightarrow$

\mathbb{R} given by $k_l(F_1(\cdot), F_1(\cdot))$ with $k_l : B \times B \rightarrow \mathbb{R}$ being an isotropic stationary kernel and

$$F_1 : B \longrightarrow B, \tag{13}$$

$$\mathbf{x} \longmapsto \left[\mathbf{w}_{*1}^{(1)}\mathbf{x}_1, \dots, \mathbf{w}_{*d}^{(1)}\mathbf{x}_d \right].$$

F_1 goes from B to B because it only works for the new coordinate system obtained by rotating the original coordinate axes, $\mathbf{x}_1, \dots, \mathbf{x}_d$. Through the first component $\mathbf{t}^{(1)}$, the elements of the vector $\mathbf{w}_*^{(1)}$ reflect how \mathbf{x} depends on y . However, such information is generally insufficient, so the elements of the vector $\mathbf{w}_*^{(1)}$ are supplemented by the information given by the other principal components $\mathbf{t}^{(2)}, \dots, \mathbf{t}^{(h)}$. Thus, we build a new kernel $k_{kpls1:h}$ sequentially by using the tensor product of all kernels k_{kpls1} , which accounts for all this information in only a single covariance kernel:

$$k_{kpls1:h}(\mathbf{x}, \mathbf{x}') = \prod_{l=1}^h k_l(F_l(\mathbf{x}), F_l(\mathbf{x}')), \tag{14}$$

with $k_l : B \times B \rightarrow \mathbb{R}$ and

$$F_l : B \longrightarrow B \tag{15}$$

$$\mathbf{x} \longmapsto \left[\mathbf{w}_{*1}^{(l)}\mathbf{x}_1, \dots, \mathbf{w}_{*d}^{(l)}\mathbf{x}_d \right].$$

If we consider the Gaussian kernel applied with this proposed approach, we get

$$k(\mathbf{x}, \mathbf{x}') = \sigma^2 \prod_{l=1}^h \prod_{i=1}^d \exp \left[-\theta_l \left(\mathbf{w}_{*i}^{(l)}\mathbf{x}_i - \mathbf{w}_{*i}^{(l)}\mathbf{x}'_i \right)^2 \right],$$

$$\forall \theta_l \in [0, +\infty[.$$

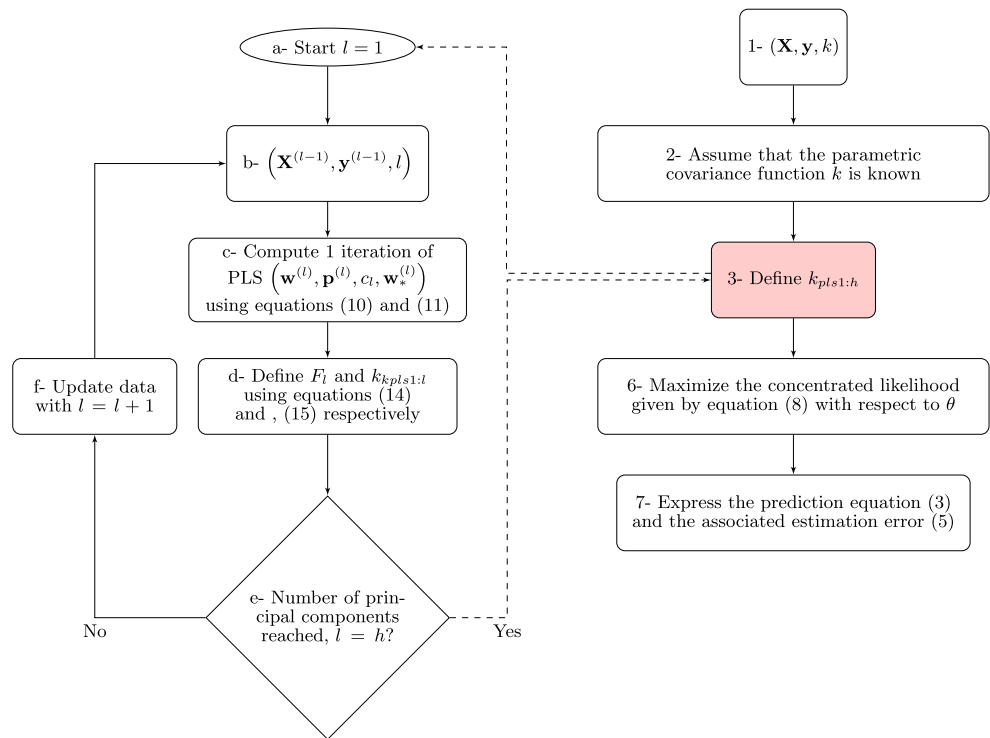
Table 13 in Appendix B presents new KPLS kernels based on examples from Table 12 (also in Appendix B) that contain fewer hyper-parameters because $h \ll d$. The number of principal components is fixed by the following leave-one-out cross-validation method:

- (i) We build KPLS models based on $h = 1$, then $h = 2, \dots$ principal components.
- (ii) We choose the number of components that minimizes the leave-one-out cross-validation error.

The flowchart given in Fig. 3 shows how the information flows through the algorithm, from sample data, PLS algorithm, kriging hyper-parameters, to final predictor. With the same definitions and equations, almost all the steps for constructing the KPLS model are similar to the original steps for constructing the ordinary kriging model. The exception is the third step, which is highlighted in the solid-red box in Fig. 3. This step uses the PLS algorithm to define the new parametric kernel $k_{kpls1:h}$ as follows:

- a. initialize the PLS algorithm with $l = 1$;
- b. if $l \neq 1$, compute the residual of $\mathbf{X}^{(l-1)}$ and $\mathbf{y}^{(l-1)}$ by using system (11);

Fig. 3 Main steps for constructing KPLS model. The solid-red box (step 3 relative to PLS) is what differentiates this approach from the ordinary kriging approach



- c. compute X weights for iteration l ;
- d. define a new kernel $k_{kpls1:h}$ by using (14);
- e. if the number of iterations is reached, return to step 3, otherwise continue;
- f. update data considering $l = l + 1$.

Note that, if kernels k_l are separable at this point, the new kernel given by (14) is also separable. In particular, if all kernels k_l are of the exponential type (e.g., all Gaussian exponentials), the new kernel given by (14) will be the same type as k_l . The proof is given in Appendix C.

4 Numerical examples

We now present a few analytical and engineering examples to verify the proper functioning of the proposed method. The ordinary kriging model with a Gaussian kernel provides the benchmark against which the results of the proposed combined approach are compared. The Python toolbox Scikit-learn v.0.14 (Pedregosa et al. 2011) is used to implement these numerical tests. This toolbox provides hyper-parameters for the ordinary kriging. The computations were done on an Intel® Celeron® CPU 900 2.20 GHz desktop PC. For the proposed method, we combined an ordinary kriging model with a Gaussian kernel with the PLS method with one to three principal components.

4.1 Analytical examples

We use two academic functions and vary the characteristics of these test problems to cover most of the difficulties faced in the field of substitution models. The first function is $g07$ (Michalewicz and Schoenauer 1996) with 10 dimensions,

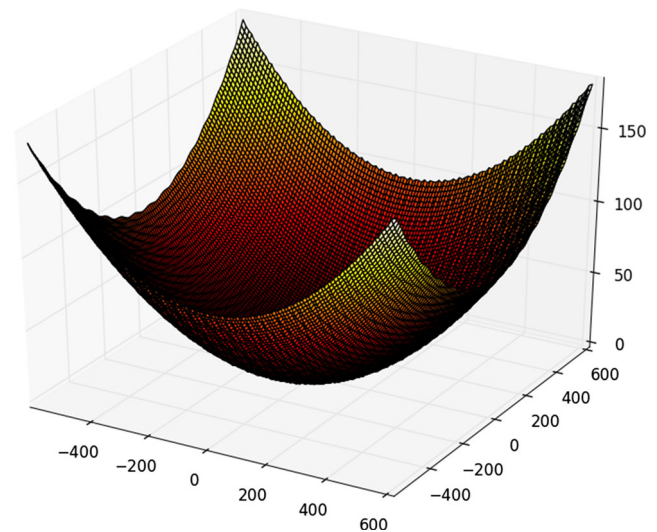


Fig. 4 A two-dimensional Griewank function over the interval $[-600, 600]$

Table 1 Number of data points used for latin hypercube design for the Griewank test function

$d = 2$	$d = 5$	$d = 7$	$d = 10$	$d = 20$	$d = 60$
$n = 70$	$n = 100$	$n = 200$	$n = 300$	$n = 400$	$n = 800$

which is close to what is required by industry in terms of dimensions,

$$\begin{aligned}
 y_{g07}(\mathbf{x}) = & x_1^2 + x_2^2 + x_1x_2 - 14x_1 - 16x_2 \\
 & + (x_3 - 10)^2 + 4(x_4 - 5)^2 + (x_5 - 3)^2 \\
 & + 2(x_6 - 1)^2 + 5x_7^2 + 7(x_8 - 11)^2 \\
 & + 2(x_9 - 10)^2 + (x_{10} - 7)^2 + 45, \\
 & -10 \leq x_i \leq 10, \text{ for } i = 1, \dots, 10.
 \end{aligned}$$

For this function, we use experiments based on a latin hypercube design with 100 data points to fit models.

The second function is the Griewank function (Regis and Shoemaker 2013), which is used because of its complexity, as illustrated in Fig. 4 for the two-dimensional (2D) case. The function is

$$\begin{aligned}
 y_{\text{Griewank}}(\mathbf{x}) = & \sum_{i=1}^d \frac{x_i^2}{4000} - \prod_{i=1}^d \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1, \\
 & -600 \leq x_i \leq 600, \text{ for } i = 1, \dots, d.
 \end{aligned}$$

Two types of experiments are done with this function. The first is defined over the interval $[-600, 600]$ and has varying dimensions (2, 5, 7, 10, 20, 60). This experiment serves to verify the effectiveness of the proposed approach in both low and high dimensions. It is based on the latin hypercube design and uses n data points to fit models, as mentioned in Table 1.

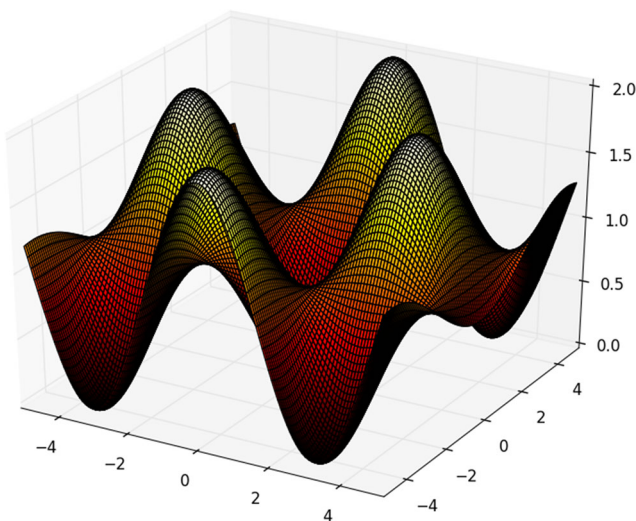


Fig. 5 A two-dimensional Griewank function over the interval $[-5, 5]$

Table 2 Results for $g07$ function in ten dimensions with 100-point latin hypercube

Surrogate	Error (%)	CPU time
Ordinary kriging	0.013	5.14 s
KPLS (1 component)	0.014	0.11 s
KPLS (2 components)	0.0015	0.43 s
KPLS (3 components)	0.0008	0.44 s

The second type of experiment is defined over the interval $[-5, 5]$, where the Griewank function is more complex than for the first type of experiment (cf. Figs. 4 and 5). Over this reduced interval, experiments are done with 20 and 40 dimensions (20D and 40D) and with 50, 100, 200, and 300 sampling points. To analyse the robustness of the method, ten experiments, each with a different latin hypercube design, are used for this case.

To compare the three approaches (i.e., the $g07$ function and the Griewank function of the intervals $[-600, 600]$ and $[-5, 5]$), 5000 random points are computed and the results are stored in a database. The following relative error is used to compare the performance of the ordinary kriging model with the KPLS model:

$$\text{Error} = \frac{\|\hat{\mathbf{Y}} - \mathbf{Y}\|_2}{\|\mathbf{Y}\|_2} 100, \tag{16}$$

where $\|\cdot\|_2$ represents the usual L^2 norm, and $\hat{\mathbf{Y}}$ and \mathbf{Y} are the vectors containing the prediction and the real values of random points, respectively. The CPU time required to fit models is also noted (“h” refers to hours, “min” refers to minutes, and “s” refers to seconds).

4.1.1 Comparison with $g07$ function

The results listed in Table 2 show that the proposed KPLS surrogate model is more accurate than the ordinary kriging model when more than one component is used. Using just one component gives almost the same accuracy as the ordinary kriging model. In this case, only a single θ hyper-parameter from the space correlation needs to be estimated compared to ten θ hyper-parameters for the

Table 3 Griewank function in two dimensions with 70-point latin hypercube over the interval $[-600, 600]$

Surrogate	Error (%)	CPU time
Ordinary kriging	5.50	0.09 s
KPLS (1 component)	7.23	0.04 s
KPLS (2 components)	5.50	0.10 s

Table 4 Griewank function in five dimensions with 100-point latin hypercube over the interval $[-600, 600]$

Surrogate	Error (%)	CPU time
Ordinary kriging	0.605	0.55 s
KPLS (1 component)	0.635	0.12 s
KPLS (2 components)	0.621	0.31 s
KPLS (3 components)	0.623	0.51 s

ordinary kriging model. Increasing the number of components improves the accuracy of the KPLS surrogate model. Whereas the PLS method treats only linearly related input and output variables, this example shows that the KPLS model can treat nonlinear problems. This result is not contradictory because (23) shows that the KPLS model is equivalent to the kriging model with specific hyper-parameters.

4.1.2 Comparison with complex Griewank function over interval $[-600, 600]$

Table 3 compares the ordinary kriging model and the KPLS model in two dimensions.

If two components are used for the KPLS, we expect to obtain the same accuracy and time cost for the two approaches because the difference between the two models consists only of a transformation of the search-space coordinates when a Gaussian kernel is used (the space in which the θ hyper-parameters exist). In this case, the KPLS model with only one component degrades the accuracy of the solution.

Tables 4, 5, and 6, show the results for 5, 7, and 10 dimensions, respectively.

Varying the number of principal components does not significantly affect the accuracy of the model. The gain in computation time does not appear upon increasing the number of principal components: the computation time is reduced when we use the KPLS model. Upon increasing the number of principal components, the CPU time for constructing the KPLS model increases but

Table 5 Griewank function in seven dimensions with 200-point latin hypercube over the interval $[-600, 600]$

Surrogate	Error (%)	CPU time
Ordinary kriging	0.138	3.09 s
KPLS (1 component)	0.141	0.25 s
KPLS (2 components)	0.138	0.52 s
KPLS (3 components)	0.141	0.94 s

Table 6 Griewank function in ten dimensions with 300-point latin hypercube over the interval $[-600, 600]$

Surrogate	Error (%)	CPU time
Ordinary kriging	0.052	21 s
KPLS (1 component)	0.033	0.6 s
KPLS (2 components)	0.035	2.41 s
KPLS (3 components)	0.034	3.58 s

still remains lower than for ordinary kriging. For these three examples, the combined approach with only one PLS component offers sufficient accuracy with a CPU time reduced 35-fold for 10 dimensions (i.e., 21 s for the ordinary kriging model and 0.6 s for the combined model).

In the 20-dimension (20D) example (Table 7), using KPLS with only one principal component leads to a poor relative error (10.15 %) compared with other models. In this case, two principal components are required to build the combined model. The CPU time remains less than that for the ordinary kriging model (11.7 s vs 107 s).

The results in Table 8 for the KPLS model with 60 dimensions (60D) show that this model is faster than the ordinary kriging model. Compared with the kriging model, the CPU time is reduced 42-fold when one principal component is used and over 17-fold when three principal components are used.

Thus, for the Griewank function over the interval $[-600, 600]$ and at the highest dimensions, the majority of the results obtained for the analytical examples are better when using the KPLS model than when using the ordinary kriging model. The proposed method thus appears interesting, particularly in terms of saving CPU time while maintaining good accuracy.

4.1.3 Comparison with complex Griewank function over interval $[-5, 5]$

As shown in Fig. 4, the Griewank function looks like a parabolic function. This is because, over the interval

Table 7 Griewank function in 20 dimensions with 400-point latin hypercube over the interval $[-600, 600]$

Surrogate	Error (%)	CPU time
Ordinary kriging	0.35	107 s
KPLS (1 component)	10.15	1.16 s
KPLS (2 components)	0.003	11.7 s
KPLS (3 components)	0.002	16.23 s

Table 8 Griewank function in 60 dimensions with 800-point latin hypercube over the interval $[-600, 600]$

Surrogate	Error (%)	CPU time
Ordinary kriging	11.47	293 s
KPLS (1 component)	7.4	6.88 s
KPLS (2 components)	6.04	12.57 s
KPLS (3 components)	5.23	16.82 s

$[-600, 600]$, the cosine part of the Griewank function does not contribute significantly compared with the sum of $x_i^2/4000$. This is especially true given that the cosine part is a product of factors each of which is less than unity. If we reduce the interval from $[-600, 600]$ to $[-5, 5]$, we can see why the Griewank function is widely used as a multimodal test function with a very rugged landscape and a large number of local optima (see Fig. 5). Compared with the interval $[-600, 600]$, the opposite happens for the interval $[-5, 5]$: the “cosine part” dominates; at least for moderate dimensions where the product contains few factors. For this case, which seems very difficult, we consider 20 and 60 input variables. For each problem, ten experiments based on the latin hypercube design are built with 50, 100, 200, and 300 sampling points. The mean and the standard error are given in Tables 14 and 15 in Appendix D. To better visualize the results, boxplots are used to show CPU time and the relative error RE given by Figs. 7, 8, 9, 10 and 11 in Appendix D.

For 20 input variables and 50 sampling points, the KPLS model gives a more accurate solution than the ordinary kriging model, as shown in Fig. 7a. The rate of improvement with respect to the number of sampling points is less for the KPLS model than for the kriging model (cf. Fig. 7b–d). Nevertheless, the results shown in Fig. 8 indicate that the

KPLS model leads to an important reduction in CPU time for the various number of sampling points.

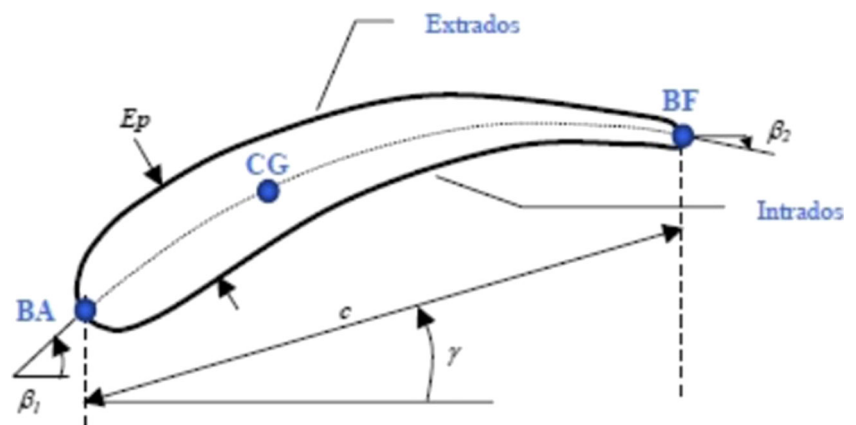
Similar results occur for the 60D Griewank function (Fig. 9). The mean RE obtained with the ordinary kriging model improves from approximately 1.39 to 0.65 % upon increasing the number of sampling points from 50 to 300 (cf. Fig. 9a, d). However, a very important reduction in CPU time is obtained, as shown in Fig. 10. The CPU time required for the KPLS model is hardly visible because it is much, much less than that required by the ordinary kriging model. We thus show in Fig. 11 the CPU time required by the KPLS model alone to show the different CPU times required for the various configurations (KPLS1, KPLS2, and KPLS3). For Griewank function over the interval $[-5, 5]$, the KPLS method seems to perform well when the number of observations is small compared to the dimension d . In this case, the standard separable covariance function for the ordinary kriging model is almost impossible to use because the number of parameters to be estimated is too large compared with the number of observations. Thus, the KPLS method seems more efficient in this case.

4.2 Industrial examples

The following engineering examples are based on results of numerical experiments done at SNECMA on multidisciplinary optimization. The results are stored in tables.

Aerospace turbomachinery consists of numerous blades that transfer energy between air and the rotor. The disks with compressor blades are particularly important because they must satisfy the dual criteria of aerodynamic performance and mechanical stress. Blades are mechanically and aerodynamically optimized by searching parameter space for an aerodynamic shape that ensures the best compromise that

Fig. 6 Example of 2D cut of blade (c is chord; CG is gravity center; β_1 is angle for BA ; β_2 is angle for BF ; E_p is maximum thickness)



satisfies a set of constraints. The blade, which is a 3D entity, is first divided into a number of radial 2D profiles whose thickness is a given percentage of the distance from the hub to the shroud (see Fig. 6).

A new 3D blade is constructed by starting with the 2D profiles and then exporting them to various meshing tools before analyzing them in any specific way. The calculation is integrated into the Optimus platform (Noesis Solutions 2009), which makes it possible to integrate multiple engineering software tools (computational structural mechanics, computational fluid dynamics, ...) into a single automated work flow. Optimus, which is an industrial software package for screening variables, optimizing design, and analyzing the sensitivity and robustness, explores and analyzes the results of the work-flow to optimize product design. Toward this end, it uses high-fidelity codes or a reduced model of these codes. It also exploits a wide range of approximation models, including the ordinary kriging model.

Input variables designate geometric hyper-parameters at different percent height and outputs are related to aerodynamic efficiency, vibration criteria, mechanical stress, geometric constraints, and aerodynamic stress. Three numerical experiments are considered:

- (i) The first experiment is denoted tab_1 and contains 24 input variables and 4 output variables. It has 99 training points and 52 validation points. The outputs are denoted $y_1, y_2, y_3,$ and y_4 .
- (ii) The second experiment is denoted tab_2 and contains 10 input variables and only 1 output variable. It has 1295 training points and 500 validation points.
- (iii) The third experiment is denoted tab_3 and contains 99 input variables and 1 output variable. It has 341 training points and 23 validation points.

Table 10 Results for tab_2 experiment data (10 input variables, 1 output variable y_1) obtained by using 1295 training points, 500 validation points, and error given by (16)

10D	Surrogate	Error (%)	CPU time
tab_2	Kriging	5.37	1 h 30 min
	KPLS1	5.07	11.69 s
	KPLS2	5.02	1 min 22 s
	KPLS3	5.34	7 min 34 s

“Kriging” refers to the ordinary kriging optimus solution and “KPLS h ” refers to the KPLS model with h principal components

Points used in $tab_1, tab_2,$ and tab_3 come from previous computationally expensive computer experiments done at SNECMA, which means that this separation between training points and verification points was imposed by SNECMA. The goal is to compare the ordinary kriging model that is implemented in the Optimus platform with the proposed KPLS model. The relative error given by (16) and the CPU time required to fit the model are reported in Tables 9–11.

The relative errors for the four models are very similar: the KPLS model results in a slightly improved accuracy for the solutions y_1, y_2, y_4 from tab_1, y_1 from $tab_2,$ and y_1 from tab_3 but degrades slightly the solution y_3 from tab_1 . The main improvement offered by the proposed model relates to the time required to fit the model, particularly for a large number of training points. Table 10 shows that, with only one principal component, the CPU time is drastically reduced compared with the Optimus model. More precisely, for $tab_2,$ the ordinary kriging model requires 1 h 30 min whereas the KPLS1 model requires only 11 s and provides better accuracy. In addition, the results for KPLS2 and

Table 9 Results for tab_1 experiment data (24 input variables, 4 output variables y_1, y_2, y_3, y_4) obtained by using 99 training points, 52 validation points, and the error given by (16)

24D	Surrogate	y_1		y_2		y_3		y_4	
		Error (%)	CPU time	Error (%)	CPU time	Error (%)	CPU time	Error (%)	CPU time
tab_1	Kriging	0.082	8 s	4.45	8.4 s	8.97	8.17 s	6.27	8.12 s
	KPLS1	0.079	0.12 s	4.04	0.11 s	10.35	0.18 s	5.67	0.11 s
	KPLS2	0.079	0.43 s	4.06	0.69 s	10.33	0.42 s	5.67	0.19 s
	KPLS3	0.079	0.82 s	4.05	0.5 s	10.41	1.14 s	5.67	0.43 s

“Kriging” refers to the ordinary kriging Optimus solution and “KPLS h ” refers to the KPLS model with h principal components

Table 11 Results for tab_3 experiment data (99 input variables, 1 output variable y_1) obtained by using 341 training points, 23 validation points, and error given by (16)

99D	Surrogate	Error (%)	CPU time
tab_3	Kriging	0.021	20 min 02 s
	KPLS1	0.19	46.6 s
	KPLS2	0.03	2 min 15 s
	KPLS3	0.02	4 min 56 s

“Kriging” refers to the ordinary kriging optimum solution and “KPLS h ” refers to the KPLS model with h principal components

KPLS3 models applied to a 99D problem are very promising (see Table 11).

One other point of major interest for the proposed method is its natural compatibility with sequential enrichment techniques such as the efficient global optimization strategy (see Jones et al. 1998).

4.3 Dimensional limits

This project is financed by SNECMA and most of their design problems do not exceed 100 input variables. In addition, the toolbox Scikit-learn (version 0.14) may have memory problems when a very large number of input variables is considered. Thus, problems with more than 100 input variables are not investigated in this work. However, by optimizing memory access and storage, this limit could easily be increased.

5 Conclusion and future work

Engineering problems that require integrating surrogate models into an optimization process are receiving increasing interest within the multidisciplinary optimization community. Computationally expensive design problems can be solved efficiently by using, for example, a kriging model, which is an interesting method for approximating and replacing high-fidelity codes, largely because these models give estimation errors, which is an interesting way to solve optimization problems. The major drawback involves the construction of the kriging model and in particular the large number of hyper-parameters that must be estimated in high dimensions. In this work, we develop a new covariance kernel for handling this type of higher-dimensional problem (up to 100 dimensions). Although the PLS method requires a very short computation time to estimate θ , the

estimate is often difficult to execute and computationally expensive when the number of input variables is greater than 10. The proposed KPLS model was tested by applying it to two analytic functions and by comparing its results to those tabulated in three industrial databases. The comparison highlights the efficiency of this model for up to 99 dimensions. The advantage of the KPLS models is not only the reduced CPU time, but also in that it reverts to the kriging model when the number of observations is small relative to the dimensions of the problem. Before using the KPLS model, however, the number of principal components should be tested to ensure a good balance between accuracy and CPU time.

An interesting direction for future work is to study how the design of the experiment (e.g., factorial) affects the KPLS model. Furthermore, other verification functions and other types of kernels can be used. In all cases studied herein, the first results with this proposed method reveal significant gains in terms of computation time while still ensuring good accuracy for design problems with up to 100 dimensions. The implementation of the proposed KPLS method requires minimal modifications of the classic kriging algorithm and offers further interesting advantages that can be exploited by methods of optimization by enrichment.

Acknowledgments The authors thank the anonymous reviewers for their insightful and constructive comments. We also extend our grateful thanks to A. Chiplunkar from ISAE SUPAERO, Toulouse and R. G. Regis from Saint Joseph’s University, Philadelphia for their careful correction of the manuscript and to SNECMA for providing the tables of experiment results. Finally, B. Kraabel is gratefully acknowledged for carefully reviewing the paper prior to publication.

Appendix A: Equations for ordinary kriging model

The expression (3) for the ordinary kriging model is transformed into (see Forrester and Sobester 2008)

$$y(\mathbf{x}) = \hat{\beta} + \mathbf{r}'_{\mathbf{x}\mathbf{X}} \mathbf{R}^{-1} (\mathbf{y} - \mathbf{1}\hat{\beta}), \quad (17)$$

where $\mathbf{1}$ denotes an n -vector of ones and

$$\hat{\beta} = (\mathbf{1}'\mathbf{R}^{-1}\mathbf{1})^{-1} \mathbf{1}'\mathbf{R}^{-1}\mathbf{y}. \quad (18)$$

In addition, (6) is written as

$$\hat{\sigma}^2 = \frac{1}{n} (\mathbf{y} - \mathbf{1}\hat{\beta})^t \mathbf{R}^{-1} (\mathbf{y} - \mathbf{1}\hat{\beta}). \tag{19}$$

Appendix B: Examples of kernels

Table 12 presents the most popular examples of stationary kernels. Table 13 presents the new KPLS kernels based on the examples given in Table 12.

Appendix C: Proof of equivalence kernel

For $l = 1, \dots, h$, k_l are separable kernels (or a d -dimensional tensor product) of the same type, so $\exists \phi_{l1}, \dots, \phi_{ld}$ such that

$$k_l(\mathbf{x}, \mathbf{x}') = \prod_{i=1}^d \phi_{li}(F_l(\mathbf{x})_i, F_l(\mathbf{x}')_i), \tag{20}$$

Table 12 Examples of commonly used stationary covariance functions

Covariance functions	Expression	Hyper-parameters θ	Number of hyper-parameters to estimate
Generalized exponential	$\sigma^2 \prod_{i=1}^d \exp(-\theta_i \mathbf{m}_i^{p_i})$	$(\theta_1, \dots, \theta_d, p_1, \dots, p_d)$	$2d$
Gaussian exponential	$\sigma^2 \prod_{i=1}^d \exp(-\theta_i \mathbf{m}_i^2)$	$(\theta_1, \dots, \theta_d)$	d
Matern $\frac{5}{2}$	$\sigma^2 \prod_{i=1}^d \left(1 + \sqrt{5}\theta_i \mathbf{m}_i + \frac{5}{3}\theta_i^2 \mathbf{m}_i^2\right) \exp(-\sqrt{5}\theta_i \mathbf{m}_i)$	$(\theta_1, \dots, \theta_d)$	d
Matern $\frac{3}{2}$	$\sigma^2 \prod_{i=1}^d \left(1 + \sqrt{3}\theta_i \mathbf{m}_i\right) \exp(-\sqrt{3}\theta_i \mathbf{m}_i)$	$(\theta_1, \dots, \theta_d)$	d

The covariance functions are written as functions of the i th component $\mathbf{m}_i = |\mathbf{x}_i - \mathbf{x}'_i|$ with $\theta_i \geq 0$ and $p_i \in [0, 2]$ for $i = 1, \dots, d$

Table 13 Examples of KPLS covariance functions

Covariance functions	Expression	Hyper-parameters θ	Number of hyper-parameters to estimate
Generalized exponential	$\sigma^2 \prod_{l=1}^h \prod_{i=1}^d \exp\left[-\theta_l \left(\mathbf{m}_i^{(l)}\right)^{p_l}\right]$	$(\theta_1, \dots, \theta_h, p_1, \dots, p_h)$	$2h \ll 2d$
Gaussian exponential	$\sigma^2 \prod_{l=1}^h \prod_{i=1}^d \exp\left[-\theta_l \left(\mathbf{m}_i^{(l)}\right)^2\right]$	$(\theta_1, \dots, \theta_h)$	$h \ll d$
Matern $\frac{5}{2}$	$\sigma^2 \prod_{l=1}^h \prod_{i=1}^d \left[1 + \sqrt{5}\theta_l \mathbf{m}_i^{(l)} + \frac{5}{3}\theta_l^2 \left(\mathbf{m}_i^{(l)}\right)^2\right] \exp\left(-\sqrt{5}\theta_l \mathbf{m}_i^{(l)}\right)$	$(\theta_1, \dots, \theta_h)$	$h \ll d$
Matern $\frac{3}{2}$	$\sigma^2 \prod_{l=1}^h \prod_{i=1}^d \left(1 + \sqrt{3}\theta_l \mathbf{m}_i^{(l)}\right) \exp\left(-\sqrt{3}\theta_l \mathbf{m}_i^{(l)}\right)$	$(\theta_1, \dots, \theta_h)$	$h \ll d$

The covariance functions are written as functions of the i th component $\mathbf{m}_i^{(l)} = |\mathbf{w}_{*i}^{(l)}(\mathbf{x}_i - \mathbf{x}'_i)|$ with $\theta_l \geq 0$ and $p_l \in [0, 2]$ for $l = 1, \dots, h$

where $F_l(\mathbf{x}_i)$ is the i th coordinate of $F_l(\mathbf{x})$. If we insert (20) in (14) we get

$$\begin{aligned}
 k_{kpls1:h}(\mathbf{x}, \mathbf{x}') &= \prod_{l=1}^h k_l(F_l(\mathbf{x}), F_l(\mathbf{x}')) \\
 &= \prod_{l=1}^h \prod_{i=1}^d \phi_{li}(F_l(\mathbf{x})_i, F_l(\mathbf{x}')_i) \\
 &= \prod_{i=1}^d \prod_{l=1}^h \phi_{li}(F_l(\mathbf{x})_i, F_l(\mathbf{x}')_i) \\
 &= \prod_{i=1}^d \psi_i(\mathbf{x}_i, \mathbf{x}'_i), \tag{21}
 \end{aligned}$$

with

$$\psi_i(\mathbf{x}_i, \mathbf{x}'_i) = \prod_{l=1}^h \phi_{li}(F_l(\mathbf{x})_i, F_l(\mathbf{x}')_i),$$

corresponding to an one-dimensional kernel. Hence, $k_{kpls1:h}$ is a separable kernel. In particular, if we consider a

generalized exponential kernel with $p_1 = \dots = p_h = p \in [0, 2]$, we obtain

$$\begin{aligned}
 \psi_i(\mathbf{x}_i, \mathbf{x}'_i) &= \sigma^{\frac{2}{d}} \exp\left(-\sum_{l=1}^h \theta_l |\mathbf{w}_{*i}^{(l)}|^p |\mathbf{x}_i - \mathbf{x}'_i|^p\right) \\
 &= \sigma^{\frac{2}{d}} \exp(-\eta_i |\mathbf{x}_i - \mathbf{x}'_i|^p), \tag{22}
 \end{aligned}$$

with

$$\eta_i = \sum_{l=1}^h \theta_l |\mathbf{w}_{*i}^{(l)}|^p.$$

We thus obtain

$$k_l(\mathbf{x}, \mathbf{x}') = \sigma^2 \prod_{i=1}^d \exp(-\eta_i |\mathbf{x}_i - \mathbf{x}'_i|^p). \tag{23}$$

Appendix D: Results of Griewank function in 20D and 60D over interval $[-5, 5]$

In Tables 14 and 15, the mean and standard deviation (std) of the numerical experiments with the Griewank function are given for 20 and 60 dimensions, respectively. To better visualize the results, boxplots are used in Figs. 7–11.

Table 14 Results for Griewank function in 20D over interval $[-5, 5]$

Surrogate	Statistic	50 points		100 points		200 points		300 points	
		Error (%)	CPU time	Error (%)	CPU time	Error (%)	CPU time	Error (%)	CPU time
Kriging	Mean	0.62	30.43 s	0.43	40.09 s	0.15	120.74 s	0.16	94.31 s
	std	0.03	9.03 s	0.04	11.96 s	0.02	27.49 s	0.06	21.92 s
KPLS1	Mean	0.54	0.05 s	0.53	0.12 s	0.48	0.43 s	0.45	0.89 s
	std	0.03	0.007 s	0.03	0.02 s	0.03	0.08 s	0.03	0.02 s
KPLS2	Mean	0.52	0.11 s	0.48	1.04 s	0.42	1.14 s	0.38	2.45 s
	std	0.03	0.05 s	0.04	0.97 s	0.04	0.92 s	0.04	1 s
KPLS3	Mean	0.51	1.27 s	0.46	3.09 s	0.37	3.56 s	0.35	3.52 s
	std	0.03	1.29 s	0.06	3.93 s	0.03	2.75 s	0.06	1.38 s

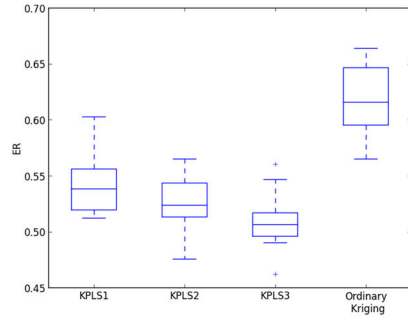
Ten trials are done for each test (50, 100, 200, and 300 training points)

Table 15 Results for Griewank function in 60D over interval $[-5, 5]$

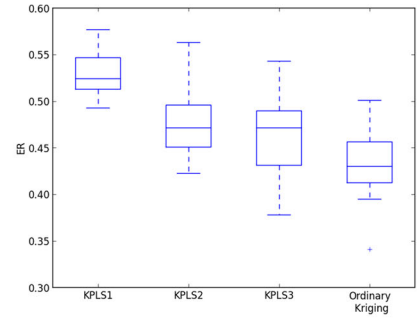
Surrogate	Statistic	50 points		100 points		200 points		300 points	
		Error (%)	CPU time	Error (%)	CPU time	Error (%)	CPU time	Error (%)	CPU time
Kriging	Mean	1.39	560.19 s	1.04	920.41 s	0.83	2015.39 s	0.65	2894.56 s
	std	0.15	200.27 s	0.05	231.34 s	0.04	239.11 s	0.03	728.48 s
KPLS1	Mean	0.92	0.07 s	0.87	0.10 s	0.82	0.37 s	0.79	0.86 s
	std	0.02	0.02 s	0.02	0.007 s	0.02	0.02 s	0.03	0.04 s
KPLS2	Mean	0.91	0.43 s	0.87	0.66 s	0.78	2.92 s	0.74	1.85 s
	std	0.03	0.54s	0.02	1.06 s	0.02	2.57 s	0.03	0.51 s
KPLS3	Mean	0.92	1.57 s	0.86	3.87 s	0.78	6.73 s	0.70	20.01 s
	std	0.04	1.98 s	0.02	5.34 s	0.02	10.94 s	0.03	26.59 s

Ten trials are done for each test (50, 100, 200, and 300 training points)

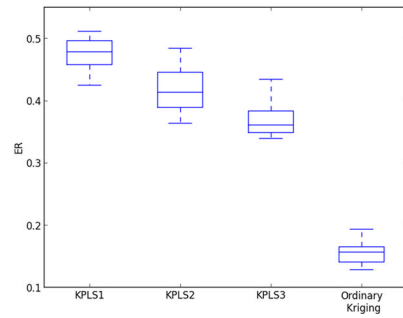
Fig. 7 *RE* for Griewank function in 20D over interval $[-5, 5]$. Experiments are based on the 10 latin hypercube design



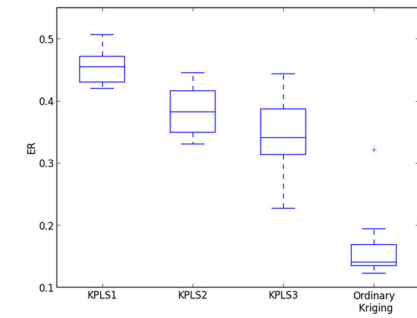
(a) *RE*(%) for 20 input variables and 50 sampling points.



(b) *RE*(%) for 20 input variables and 100 sampling points.

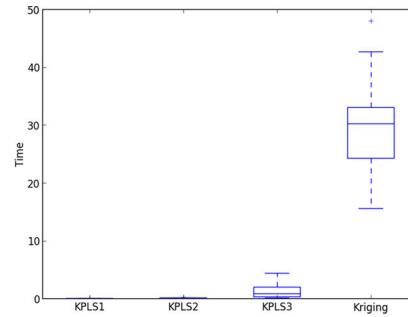


(c) *RE*(%) for 20 input variables and 200 sampling points.

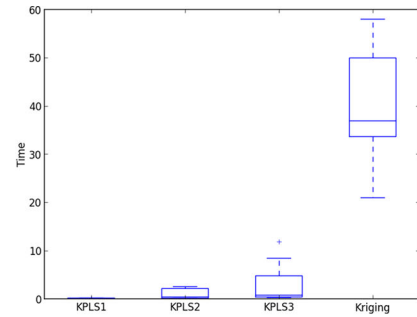


(d) *RE*(%) for 20 input variables and 300 sampling points.

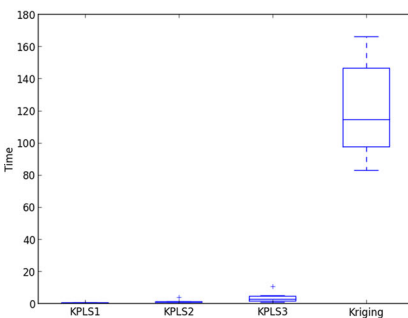
Fig. 8 CPU time for Griewank function in 20D over interval $[-5, 5]$. Experiments are based on the 10 latin hypercube design



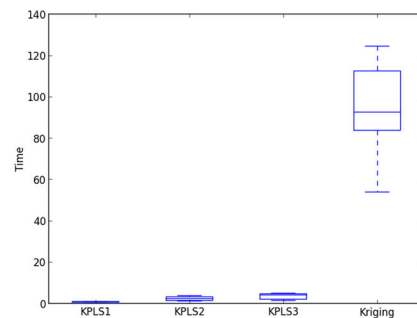
(a) CPU time for 20 input variables and 50 sampling points.



(b) CPU time for 20 input variables and 100 sampling points.

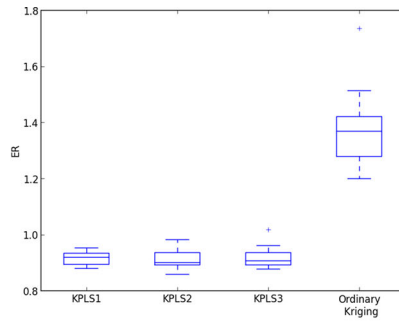


(c) CPU time for 20 input variables and 200 sampling points.

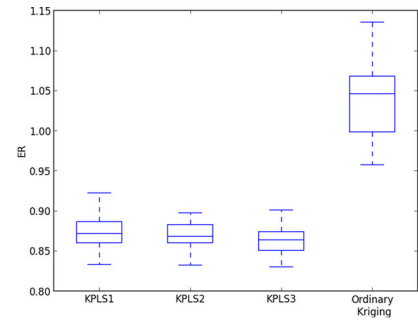


(d) CPU time for 20 input variables and 300 sampling points.

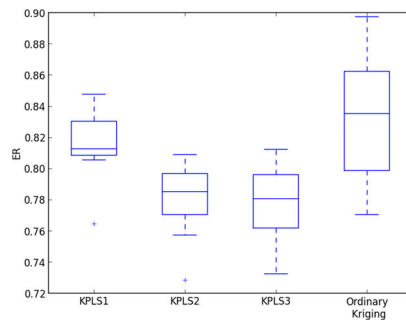
Fig. 9 *RE* for Griewank function in 60D over interval $[-5, 5]$. Experiments are based on the 10 latin hypercube design



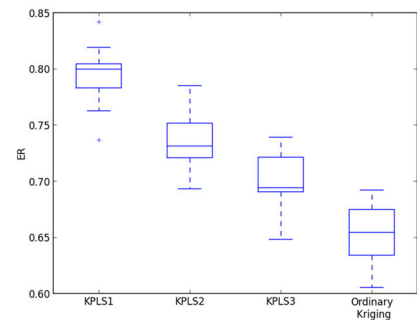
(a) *RE*(%) for 60 input variables and 50 sampling points.



(b) *RE*(%) for 60 input variables and 100 sampling points.

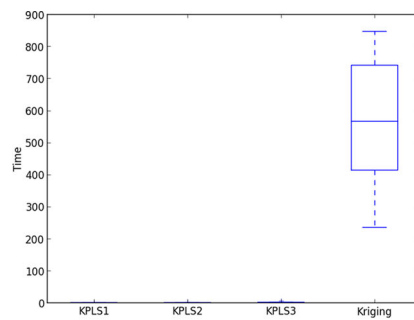


(c) *RE*(%) for 60 input variables and 200 sampling points.

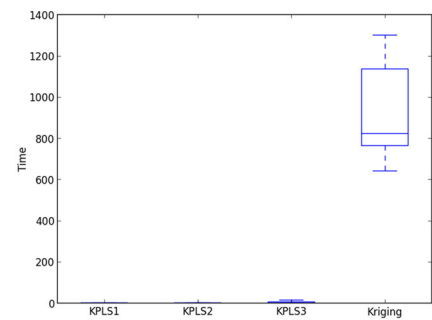


(d) *RE*(%) for 60 input variables and 300 sampling points.

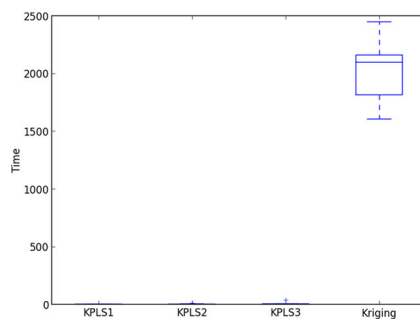
Fig. 10 CPU time for Griewank function in 60D over interval $[-5, 5]$. Experiments are based on the 10 latin hypercube design



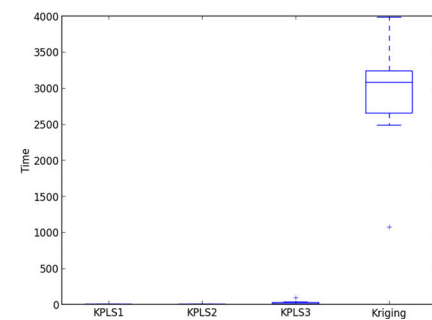
(a) CPU time for 60 input variables and 50 sampling points.



(b) CPU time for 60 input variables and 100 sampling points.

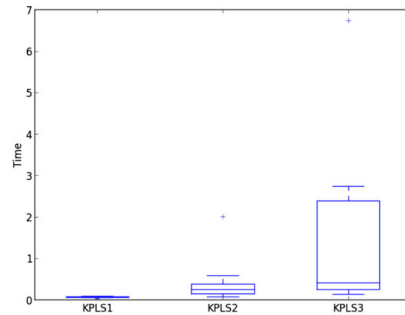


(c) CPU time for 60 input variables and 200 sampling points.

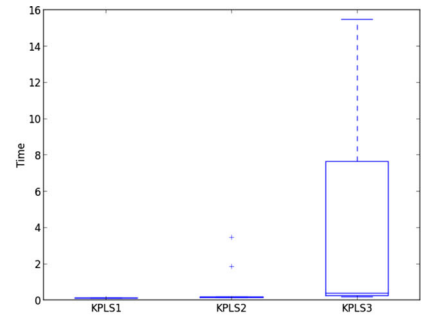


(d) CPU time for 60 input variables and 300 sampling points.

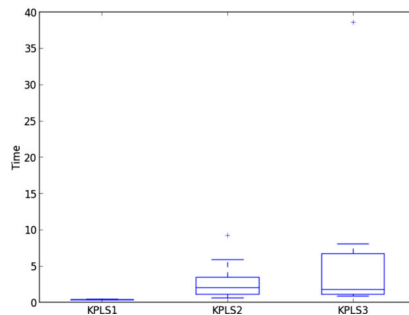
Fig. 11 CPU time for Griewank function in 60D for only KPLS models over interval $[-5, 5]$. Experiments are based on the 10 latin hypercube design



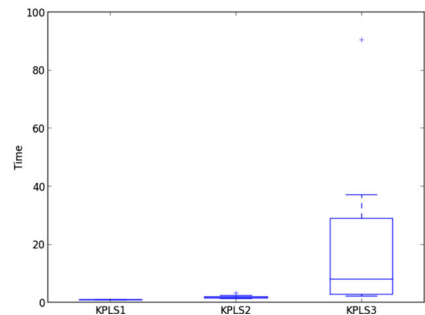
(a) CPU time for 60 input variables and 50 sampling points.



(b) CPU time for 60 input variables and 100 sampling points.



(c) CPU time for 60 input variables and 200 sampling points.



(d) CPU time for 60 input variables and 300 sampling points.

References

- Alberto P, González F (2012) Partial Least Squares regression on symmetric positive-definite matrices. *Rev Col Estad* 36(1):177–192
- Bachoc F (2013) Cross Validation and Maximum Likelihood estimation of hyper-parameters of Gaussian processes with model misspecification. *Comput Stat Data Anal* 66:55–69
- Bishop CM (2007) *Pattern recognition and machine learning (information science and statistics)*. Springer
- Braham H, Ben Jemaa S, Sayrac B, Fort G, Moulines E (2014) Low complexity spatial interpolation for cellular coverage analysis. In: 2014 12th international symposium on modeling and optimization in mobile, ad hoc, and wireless networks (WiOpt). IEEE, pp 188–195
- Buhmann MD (2003) *Radial basis functions: theory and implementations*, vol 12. Cambridge University Press, Cambridge
- Cressie N (1988) *Spatial prediction and ordinary kriging*. *Math Geol* 20(4):405–421
- Damianou A, Lawrence ND (2013) Deep gaussian processes. In: *Proceedings of the sixteenth international conference on artificial intelligence and statistics, AISTATS 2013, Scottsdale*, pp 207–215
- Durrande N (2011) *Covariance kernels for simplified and interpretable modeling. A functional and probabilistic approach*. Theses, Ecole Nationale Supérieure des Mines de saint-Etienne
- Durrande N, Ginsbourger D, Roustant O (2012) Additive covariance kernels for high-dimensional gaussian process modeling. *Ann Fac Sci Toulouse Math* 21(3):481–499
- Forrester A, Sobester A, Keane A (2008) *Engineering design via surrogate modelling: a practical guide*. Wiley, New York
- Frank IE, Friedman JH (1993) A statistical view of some chemometrics regression tools. *Technometrics* 35:109–148
- Goovaerts P (1997) *Geostatistics for natural resources evaluation (applied geostatistics)*. Oxford University Press, New York
- Haykin S (1998) *Neural networks: a comprehensive foundation*, 2nd edn. Prentice Hall PTR, Upper Saddle River
- Helland I (1988) On structure of partial least squares regression. *Commun Stat - Simul Comput* 17:581–607
- Hensman J, Fusi N, Lawrence ND (2013) Gaussian processes for big data. In: *Proceedings of the twenty-ninth conference on uncertainty in artificial intelligence, Bellevue*, p 2013
- Jones DR, Schonlau M, Welch WJ (1998) Efficient global optimization of expensive black-box functions. *J Glob Optim* 13(4):455–492
- Lanczos C (1950) An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *J Res Natl Bur Stand* 45(4):255–282
- Liem RP, Martins JRR (2014) Surrogate models and mixtures of experts in aerodynamic performance prediction for mission analysis. In: 15th AIAA/ISSMO multidisciplinary analysis and optimization conference, Atlanta, GA, AIAA-2014-2301
- Manne R (1987) Analysis of two Partial-Least-Squares algorithms for multivariate calibration. *Chemom Intell Lab Syst* 2(1–3):187–197
- Mera NS (2007) Efficient optimization processes using kriging approximation models in electrical impedance tomography. *Int J Numer Methods Eng* 69(1):202–220
- Michalewicz Z, Schoenauer M (1996) Evolutionary algorithms for constrained parameter optimization problems. *Evol Comput* 4: 1–32
- Noesis Solutions (2009) OPTIMUS. <http://www.noessolutions.com/Noesis/optimus-details/optimus-design-optimization>

- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V et al (2011) Scikit-learn: machine learning in python. *J Mach Learn Res* 12:2825–2830
- Picheny V, Ginsbourger D, Roustant O, Haftka RT, Kim NH (2010) Adaptive designs of experiments for accurate approximation of a target region. *J Mech Des* 132(7):071008
- Powell MJ (1994) A direct search optimization method that models the objective and constraint functions by linear interpolation. In: *Advances in optimization and numerical analysis*. Springer, pp 51–67
- Rasmussen C, Williams C (2006) *Gaussian processes for machine learning*. Adaptive computation and machine learning. MIT Press, Cambridge
- Regis R, Shoemaker C (2013) Combining radial basis function surrogates and dynamic coordinate search in high-dimensional expensive black-box optimization. *Eng Optim* 45(5):529–555
- Roustant O, Ginsbourger D, Deville Y (2012) DiceKriging, DiceOptim: two R packages for the analysis of computer experiments by kriging-based metamodeling and optimization. *J Stat Softw* 51(1):1–55
- Sakata S, Ashida F, Zako M (2004) An efficient algorithm for Kriging approximation and optimization with large-scale sampling data. *Comput Methods Appl Mech Eng* 193(3):385–404
- Sasena M (2002) *Flexibility and efficiency enhancements for constrained global design optimization with Kriging approximations*. PhD thesis, University of Michigan
- Schonlau M (1998) *Computer experiments and global optimization*. PhD thesis, University of Waterloo
- Wahba G (1990) *Spline models for observational data*, CBMS-NSF regional conference series in applied mathematics, vol 59. Society for Industrial and Applied Mathematics (SIAM), Philadelphia
- Wahba G, Craven P (1978) Smoothing noisy data with spline functions. Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numer Math* 31:377–404
- Zimmerman DL, Homer KE (1991) A network design criterion for estimating selected attributes of the semivariogram. *Environmetrics* 2(4):425–441

4.4.3 An Improved Approach for Estimating the Hyperparameters of the Kriging Model for High-Dimensional Problems through the Partial Least Squares Method, Mathematical Problems in Engineering 2016

Research Article

An Improved Approach for Estimating the Hyperparameters of the Kriging Model for High-Dimensional Problems through the Partial Least Squares Method

Mohamed Amine Bouhlel,¹ Nathalie Bartoli,² Abdelkader Otsmane,³ and Joseph Morlier⁴

¹Department of Aerospace Engineering, University of Michigan, 1320 Beal Avenue, Ann Arbor, MI 48109, USA

²ONERA, 2 Avenue Édouard Belin, 31055 Toulouse, France

³SNECMA, Rond-Point René Ravaud-Réau, 77550 Moissy-Cramayel, France

⁴Institut Clément Ader, CNRS, ISAE-SUPAERO, Université de Toulouse, 10 Avenue Edouard Belin, 31055 Toulouse Cedex 4, France

Correspondence should be addressed to Mohamed Amine Bouhlel; bouhlel.mohamed.amine@gmail.com

Received 31 December 2015; Revised 10 May 2016; Accepted 24 May 2016

Academic Editor: Erik Cuevas

Copyright © 2016 Mohamed Amine Bouhlel et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

During the last years, kriging has become one of the most popular methods in computer simulation and machine learning. Kriging models have been successfully used in many engineering applications, to approximate expensive simulation models. When many input variables are used, kriging is inefficient mainly due to an exorbitant computational time required during its construction. To handle high-dimensional problems (100+), one method is recently proposed that combines kriging with the Partial Least Squares technique, the so-called KPLS model. This method has shown interesting results in terms of saving CPU time required to build model while maintaining sufficient accuracy, on both academic and industrial problems. However, KPLS has provided a poor accuracy compared to conventional kriging on multimodal functions. To handle this issue, this paper proposes adding a new step during the construction of KPLS to improve its accuracy for multimodal functions. When the exponential covariance functions are used, this step is based on simple identification between the covariance function of KPLS and kriging. The developed method is validated especially by using a multimodal academic function, known as Griewank function in the literature, and we show the gain in terms of accuracy and computer time by comparing with KPLS and kriging.

1. Introduction

During the last years, the kriging model [1–4], which is referred to as the Gaussian process model [5], has become one of the most popular methods in computer simulation and machine learning. It is used as a substitute of high-fidelity codes representing physical phenomena and aims to reduce the computational time of a particular process. For instance, the kriging model is used successfully in several optimization problems [6–11]. Kriging is not well adapted to high-dimensional problem, principally due to large matrix inversion problems. In fact, the kriging model becomes much time consuming when a large number of input variables are used since a large number of sampling points are required.

Indeed, it is recommended in [12] to use $10d$ sampling points, with d the number of dimensions, for obtaining a good accuracy of the kriging model. As a result, we need to increase the size of the kriging covariance matrix which becomes computationally very expensive to invert. Moreover, this inversion's problem induces difficulty in the classical hyperparameters estimation through the maximization of the likelihood function.

A recent method, called KPLS [13], is developed to reduce computational time which uses, during a construction of the kriging model, the dimensional reduction method “Partial Least Squares” (PLS). This method is able to reduce the number of hyperparameters of a kriging model, such that their number becomes equal to the number of principal

components retained by the PLS method. The KPLS method is thus able to rapidly build a kriging model for high-dimensional problems (100+) while maintaining a good accuracy. However, it has been shown in [13] that the KPLS model is less accurate than the kriging model in many cases, in particular for multimodal functions.

In this paper, we propose an extra step that supplements [13] in order to improve its accuracy. Under hypothesis that kernels used for building the KPLS model are of exponential type with the same form (all Gaussian kernels, e.g.), we choose the hyperparameters found by the KPLS model as an initial point to optimize the likelihood function of a conventional kriging model. In fact, this approach is performed by identifying the covariance function of the KPLS model as a covariance function of a kriging model. The fact of considering the identified kriging model, instead of the KPLS model, leads to extending the search space where the hyperparameters are defined and thus to making the resulting model more flexible than the KPLS model.

This paper is organized in 3 main sections. In Section 2, we present a review of the KPLS model. In Section 3, we discuss our new approach under the hypothesis needed for its applicability. Finally, numerical results are shown to confirm the efficiency of our method followed by a summary of what we have achieved.

2. Construction of KPLS

In this section, we introduce the notation and describe the theory behind the construction of the KPLS model. Assume that we have evaluated a cost deterministic function of n points $\mathbf{x}^{(i)}$ ($i = 1, \dots, n$) with $\mathbf{x}^{(i)} = [x_1^{(i)}, \dots, x_d^{(i)}] \in B \subset \mathbb{R}^d$, and we denote \mathbf{X} by the matrix $[\mathbf{x}^{(1)t}, \dots, \mathbf{x}^{(n)t}]^t$. For simplicity, B is considered to be a hypercube expressed by the product between intervals of each direction space; that is, $B = \prod_{j=1}^d [a_j, b_j]$, where $a_j, b_j \in \mathbb{R}$ with $a_j \leq b_j$ for $j = 1, \dots, d$. Simulating these n inputs gives the outputs $\mathbf{y} = [y^{(1)}, \dots, y^{(n)}]^t$ with $y^{(i)} = y(\mathbf{x}^{(i)})$, for $i = 1, \dots, n$.

2.1. Construction of the Kriging Model. For building the kriging model, we assume that the deterministic response $y(\mathbf{x})$ is realization of a stochastic process [14–17]:

$$Y(\mathbf{x}) = \beta_0 + Z(\mathbf{x}). \quad (1)$$

The presented formula, with β_0 an unknown constant, corresponds to ordinary kriging [8] which is a particular case of universal kriging [15]. The stochastic term $Z(\mathbf{x})$ is considered as realization of a stationary Gaussian process with $E[Z(\mathbf{x})] = 0$ and a covariance function, also called kernel function, given by

$$\text{Cov}(Z(\mathbf{x}), Z(\mathbf{x}')) = k(\mathbf{x}, \mathbf{x}') = \sigma^2 r(\mathbf{x}, \mathbf{x}') = \sigma^2 r_{\mathbf{xx}'}, \quad (2)$$

$$\forall \mathbf{x}, \mathbf{x}' \in B,$$

where σ^2 is the process variance and $r_{\mathbf{xx}'}$ is the correlation function between \mathbf{x} and \mathbf{x}' . However, the correlation function

r depends on hyperparameters θ which are considered to be known. We also denote the $n \times 1$ vector as $\mathbf{r}_{\mathbf{xx}} = [r_{\mathbf{xx}^{(1)}}, \dots, r_{\mathbf{xx}^{(n)}}]^t$ and the $n \times n$ correlation matrix as $\mathbf{R} = [\mathbf{r}_{\mathbf{x}^{(1)}\mathbf{x}}, \dots, \mathbf{r}_{\mathbf{x}^{(n)}\mathbf{x}}]$. We use $\hat{y}(\mathbf{x})$ to denote the prediction of the true function $y(\mathbf{x})$. Under the hypothesis above, the best linear unbiased predictor for $y(\mathbf{x})$, given the observations \mathbf{y} , is

$$\hat{y}(\mathbf{x}) = \hat{\beta}_0 + \mathbf{r}_{\mathbf{xx}}^t \mathbf{R}^{-1} (\mathbf{y} - \hat{\beta}_0 \mathbf{1}), \quad (3)$$

where $\mathbf{1}$ denotes an n -vector of ones and

$$\hat{\beta}_0 = (\mathbf{1}^t \mathbf{R}^{-1} \mathbf{1})^{-1} \mathbf{1}^t \mathbf{R}^{-1} \mathbf{y}. \quad (4)$$

In addition, the estimation of σ^2 is given by

$$\hat{\sigma}^2 = \frac{1}{n} (\mathbf{y} - \mathbf{1} \hat{\beta}_0)^t \mathbf{R}^{-1} (\mathbf{y} - \mathbf{1} \hat{\beta}_0). \quad (5)$$

Moreover, ordinary kriging provides an estimate of the variance of the prediction, which is given by

$$s^2(\mathbf{x}) = \hat{\sigma}^2 (1 - \mathbf{r}_{\mathbf{xx}}^t \mathbf{R}^{-1} \mathbf{r}_{\mathbf{xx}}). \quad (6)$$

Note that the assumption of a known covariance function with known parameters θ is unrealistic in reality and they are often unknown. For this reason, the covariance function is typically chosen from among a parametric family of kernels. In this work, only the covariance functions of exponential type are considered, in particular the Gaussian kernel. Indeed, the Gaussian kernel is the most popular kernel in kriging metamodels of simulation models, which is given by

$$k(\mathbf{x}, \mathbf{x}') = \sigma^2 \prod_{i=1}^d \exp(-\theta_i (x_i - x'_i)^2), \quad \forall \theta_i \in \mathbb{R}^+. \quad (7)$$

We note that the parameters θ_i , for $i = 1, \dots, d$, can be interpreted as measuring how strongly the input variables x_1, \dots, x_d , respectively, affect the output y . If θ_i is very large, the kernel $k(\mathbf{x}, \mathbf{x}')$ given by (7) tends to zero and thus leads to a low correlation. In fact, we see in Figure 1 how the correlation curve rapidly varies from a point to another when $\theta = 10$.

However, the estimator of the kriging parameters ($\hat{\beta}_0, \hat{\sigma}^2$, and $\theta_1, \dots, \theta_d$) makes the kriging predictor, given by (3), nonlinear and makes the estimated predictor variance, given by (6), biased. We note that the vector \mathbf{r} and the matrix \mathbf{R} should get hats above but it is ignored in practice [18].

2.2. Partial Least Squares. The PLS method is a statistical method which searches out the best multidimensional direction \mathbf{X} that explains the characteristics of the output y . It finds a linear relationship between input variables and output variable by projecting input variables onto principal components, also called latent variables. The PLS technique reduces dimension and reveals how inputs depend on output. In the following, we use h to denote the number of principal components retained which are a lot lower than d ($h \ll d$); h does not generally exceed 4, in practice. In addition,

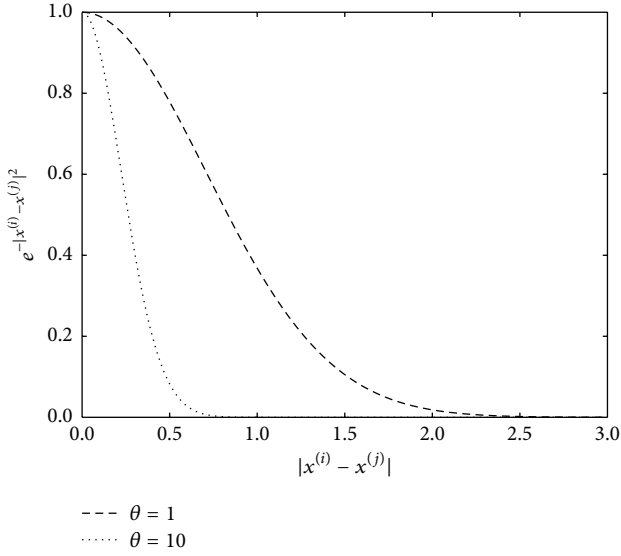


FIGURE 1: Theta smoothness can be tuned to adapt spatial influence to our problem. The magnitude of θ dictates how quickly the squared exponential function varies.

the principal components can be computed sequentially. In fact, the principal component $\mathbf{t}^{(l)}$, for $l = 1, \dots, h$, is computed by seeking the best direction $\mathbf{w}^{(l)}$ which maximizes the squared covariance between $\mathbf{t}^{(l)} = \mathbf{X}^{(l-1)} \mathbf{w}^{(l)}$ and $\mathbf{y}^{(l-1)}$:

$$\mathbf{w}^{(l)} = \begin{cases} \arg \max_{\mathbf{w}^{(l)}} \mathbf{w}^{(l)\top} \mathbf{X}^{(l-1)\top} \mathbf{y}^{(l-1)} \mathbf{y}^{(l-1)\top} \mathbf{X}^{(l-1)} \mathbf{w}^{(l)} \\ \text{such that } \mathbf{w}^{(l)\top} \mathbf{w}^{(l)} = 1, \end{cases} \quad (8)$$

where $\mathbf{X} = \mathbf{X}^{(0)}$, $\mathbf{y} = \mathbf{y}^{(0)}$, and, for $l = 1, \dots, h$, $\mathbf{X}^{(l)}$ and $\mathbf{y}^{(l)}$ are the residual matrix from the local regression of $\mathbf{X}^{(l-1)}$ onto the principal component $\mathbf{t}^{(l)}$ and from the local regression of $\mathbf{y}^{(l-1)}$ onto the principal component $\mathbf{t}^{(l)}$, respectively, such that

$$\begin{aligned} \mathbf{X}^{(l)} &= \mathbf{X}^{(l-1)} - \mathbf{t}^{(l)} \mathbf{p}^{(l)\top}, \\ \mathbf{y}^{(l)} &= \mathbf{y}^{(l-1)} - c_l \mathbf{t}^{(l)}, \end{aligned} \quad (9)$$

where $\mathbf{p}^{(l)}$ (a $1 \times d$ vector) and c_l (a coefficient) contain the regression coefficients. For more details of how PLS method works, please see [19–21].

The principal components represent the new coordinate system obtained upon rotating the original system with axes, x_1, \dots, x_d [21]. For $l = 1, \dots, h$, $\mathbf{t}^{(l)}$ can be written as

$$\mathbf{t}^{(l)} = \mathbf{X}^{(l-1)} \mathbf{w}^{(l)} = \mathbf{X} \mathbf{w}_*^{(l)}. \quad (10)$$

This important relationship is mainly used for developing the KPLS model which is detailed in Section 2.3. The vectors $\mathbf{w}_*^{(l)}$, for $l = 1, \dots, h$, are given by the following matrix $\mathbf{W}_* = [\mathbf{w}_*^{(1)}, \dots, \mathbf{w}_*^{(h)}]$ which is obtained by (for more details, see [22])

$$\mathbf{W}_* = \mathbf{W} (\mathbf{P}^\top \mathbf{W})^{-1}, \quad (11)$$

where $\mathbf{W} = [\mathbf{w}^{(1)}, \dots, \mathbf{w}^{(h)}]$ and $\mathbf{P} = [\mathbf{p}^{(1)\top}, \dots, \mathbf{p}^{(h)\top}]$.

TABLE 1: Results for tab_1 experiment data (24 input variables, output variables y_3) obtained by using 99 training points, 52 validation points, and the error given by (18). ‘‘Kriging’’ refers to the ordinary kriging Optimus solution and ‘‘KPLS h ’’ and ‘‘KPLS h +K’’ refer to KPLS and KPLS+K with h principal components, respectively. Best results of the relative error are highlighted in bold type.

	Surrogate model	RE (%)	CPU time
tab_1	Kriging	8.97	8.17 s
	KPLS1	10.35	0.18 s
	KPLS2	10.33	0.42 s
	KPLS3	10.41	1.14 s
	KPLS1+K	8.77	2.15 s
	KPLS2+K	8.72	4.22 s
	KPLS3+K	8.73	4.53 s

2.3. *Construction of the KPLS Model.* The hyperparameters $\theta = \{\theta_i\}$, for $i = 1, \dots, d$, given by (7) are found by maximum likelihood estimation (MLE) method. Their estimation becomes more and more expensive when d increases. The vector θ can be interpreted as measuring how strongly the variables x_1, \dots, x_d affect the output y , respectively. For building KPLS, coefficients given by vectors $\mathbf{w}_*^{(l)}$ will be considered as measuring of the influence of the input variables x_1, \dots, x_d on the output y . By some elementary operations on the kernel functions, we define the KPLS kernel by

$$k_{\text{KPLS1:}h}(\mathbf{x}, \mathbf{x}') = \prod_{l=1}^h k_l(F_l(\mathbf{x}), F_l(\mathbf{x}')), \quad (12)$$

where $k_l: B \times B \rightarrow \mathbb{R}$ is an isotropic stationary kernel and

$$\begin{aligned} F_l: B &\longrightarrow B \\ \mathbf{x} &\longmapsto [w_{*1}^{(l)} x_1, \dots, w_{*d}^{(l)} x_d]. \end{aligned} \quad (13)$$

More details of such construction are given in [13]. Considering the example of the Gaussian kernel given by (7), we obtain

$$k(\mathbf{x}, \mathbf{x}') = \sigma^2 \prod_{l=1}^h \prod_{i=1}^d \exp \left[-\theta_l (w_{*i}^{(l)} x_i - w_{*i}^{(l)} x'_i)^2 \right], \quad (14)$$

$$\forall \theta_l \in \mathbb{R}^+.$$

Since a small number of principal components are retained, the estimation of the hyperparameters $\theta_1, \dots, \theta_h$ is faster than the hyperparameters $\theta_1, \dots, \theta_d$ given by (7), where d is very high (100+).

3. Transition from the KPLS Model to the Kriging Model Using the Exponential Covariance Functions

In this section, we show that if all kernels k_l , for $l = 1, \dots, h$, used in (12) are of the exponential type with the same form (all Gaussian kernels, e.g.), then the kernel $k_{\text{KPLS1:}h}$ given by (12) will be of the exponential type with the same form as k_l (Gaussian if all k_l are Gaussian).

TABLE 2: Results of the Griewank function in 20D over the interval $[-5, 5]$. Ten trials are done for each test (50, 100, 200, and 300 training points). Best results of the relative error are highlighted in bold type for each case.

Surrogate	Statistic	50 points		100 points	
		Error (%)	CPU time	Error (%)	CPU time
Kriging	Mean	0.62	30.43 s	0.43	40.09 s
	std	0.03	9.03 s	0.04	11.96 s
KPLS1	Mean	0.54	0.05 s	0.53	0.12 s
	std	0.03	0.007 s	0.03	0.02 s
KPLS2	Mean	0.52	0.11 s	0.48	1.04 s
	std	0.03	0.05 s	0.04	0.97 s
KPLS3	Mean	0.51	1.27 s	0.46	3.09 s
	std	0.03	1.29 s	0.06	3.93 s
KPLS1+K	Mean	0.59	1.20 s	0.45	2.42 s
	std	0.04	0.16 s	0.07	0.44 s
KPLS2+K	Mean	0.58	1.28 s	0.42	3.38 s
	std	0.04	0.15 s	0.05	1.06 s
KPLS3+K	Mean	0.58	2.45 s	0.41	5.61 s
	std	0.03	1.32 s	0.05	3.99 s
Surrogate	Statistic	200 points		300 points	
		Error (%)	CPU time	Error (%)	CPU time
Kriging	Mean	0.15	120.74 s	0.16	94.31 s
	std	0.02	27.49 s	0.06	21.92 s
KPLS1	Mean	0.48	0.43 s	0.45	0.89 s
	std	0.03	0.08 s	0.03	0.02 s
KPLS2	Mean	0.42	1.14 s	0.38	2.45 s
	std	0.04	0.92 s	0.04	1 s
KPLS3	Mean	0.37	3.56 s	0.35	3.52 s
	std	0.03	2.75 s	0.06	1.38 s
KPLS1+K	Mean	0.20	8.00 s	0.17	19.07 s
	std	0.04	1.51 s	0.07	3.19 s
KPLS2+K	Mean	0.18	9.71 s	0.16	19.89 s
	std	0.02	1.29 s	0.05	2.67 s
KPLS3+K	Mean	0.16	11.67 s	0.16	20.49 s
	std	0.02	3.88 s	0.05	3.46 s

3.1. *Proof of the Equivalence between the Kernels of the KPLS Model and the Kriging Model.* Let us define, for $i = 1, \dots, d$, $\eta_i = \sum_{l=1}^h \theta_l w_{*i}^{(l)2}$; we have

$$\begin{aligned}
k_{1:h}(\mathbf{x}, \mathbf{x}') &= \prod_{l=1}^h \prod_{i=1}^d \exp\left(-\theta_l w_{*i}^{(l)2} (x_i - x'_i)^2\right) \\
&= \exp\left(\sum_{i=1}^d \sum_{l=1}^h -\theta_l w_{*i}^{(l)2} (x_i - x'_i)^2\right) \\
&= \exp\left(\sum_{i=1}^d -\eta_i (x_i - x'_i)^2\right) \\
&= \prod_{i=1}^d \exp\left(-\eta_i (x_i - x'_i)^2\right).
\end{aligned} \tag{15}$$

In the same way, we can show this equivalence for the other exponential kernels where $p_1 = \dots = p_h$:

$$k_{1:h}(\mathbf{x}, \mathbf{x}') = \sigma^2 \prod_{l=1}^h \prod_{i=1}^d \exp\left(-\theta_l |w_{*i}^{(l)} (x_i - x'_i)|^{p_l}\right). \tag{16}$$

However, we must caution that the above proof shows equivalence between the covariance functions of KPLS and kriging only on a subspace domain. More precisely, the KPLS covariance function is defined in a subspace from \mathbb{R}^{+d} whereas the kriging covariance function is defined in the complete \mathbb{R}^{+d} domain. Thus, our original idea is to extend the space where the KPLS covariance function is defined for the complete space \mathbb{R}^{+d} .

3.2. *A New Step during the Construction of the KPLS Model: KPLS+K.* By considering the equivalence shown in the last

TABLE 3: Results of the Griewank function in 60D over the interval $[-5, 5]$. Ten trials are done for each test (50, 100, 200, and 300 training points). Best results of the relative error are highlighted in bold type for each case.

Surrogate	Statistic	50 points		100 points	
		Error (%)	CPU time	Error (%)	CPU time
Kriging	Mean	1.39	560.19 s	1.04	920.41 s
	std	0.15	200.27 s	0.05	231.34 s
KPLS1	Mean	0.92	0.07 s	0.87	0.10 s
	std	0.02	0.02 s	0.02	0.007 s
KPLS2	Mean	0.91	0.43 s	0.87	0.66 s
	std	0.03	0.54 s	0.02	1.06 s
KPLS3	Mean	0.92	1.57 s	0.86	3.87 s
	std	0.04	1.98 s	0.02	5.34 s
KPLS1+K	Mean	0.99	2.14 s	0.90	2.90 s
	std	0.03	0.72 s	0.03	0.03 s
KPLS2+K	Mean	0.98	2.44 s	0.88	3.44 s
	std	0.04	0.63 s	0.02	1.06 s
KPLS3+K	Mean	0.99	3.82 s	0.88	6.68 s
	std	0.05	2.33 s	0.03	5.34 s
Surrogate	Statistic	200 points		300 points	
		Error (%)	CPU time	Error (%)	CPU time
Kriging	Mean	0.83	2015.39 s	0.65	2894.56 s
	std	0.04	239.11 s	0.03	728.48 s
KPLS1	Mean	0.82	0.37 s	0.79	0.86 s
	std	0.02	0.02 s	0.03	0.04 s
KPLS2	Mean	0.78	2.92 s	0.74	1.85 s
	std	0.02	2.57 s	0.03	0.51 s
KPLS3	Mean	0.78	6.73 s	0.70	20.01 s
	std	0.02	10.94 s	0.03	26.59 s
KPLS1+K	Mean	0.76	9.88 s	0.66	22.00 s
	std	0.03	0.06 s	0.02	0.15 s
KPLS2+K	Mean	0.75	12.38 s	0.60	23.03 s
	std	0.03	2.56 s	0.03	0.50 s
KPLS3+K	Mean	0.74	16.18 s	0.61	41.13 s
	std	0.03	10.95 s	0.03	26.59 s

section, we propose to add a new step during the construction of the KPLS model. This step occurs just after the θ_l -estimation, for $l = 1, \dots, h$. It involves making local optimization of the likelihood function of the kriging model equivalent to the KPLS model. Moreover, we use $\eta_i = \sum_{l=1}^h \theta_l w_{*i}^{(l)2}$, for $i = 1, \dots, d$, as a starting point of the local optimization by considering the solution θ_l , for $l = 1, \dots, h$, found by the KPLS method. Thus, this optimization is done in the complete space, where the vector $\eta = \{\eta_i\} \in \mathbb{R}^{+d}$.

This approach, called KPLS+K, aims to improve the MLE of the kriging model equivalent to the associated KPLS model. In fact, the local optimization of the equivalent kriging offers more possibilities for improving the MLE, by considering a wider search space, and thus it will be able to correct the estimation of many directions. These directions are represented by η_i for the i th direction which is badly estimated by the KPLS method. Because estimating the equivalent kriging hyperparameters can be time consuming,

especially when d is large, we improve the MLE by a local optimization at the cost of a slight increase of computational time.

Figure 2 recalls the principal stages of building a KPLS+K model.

4. Numerical Simulations

We now focus on the performance of KPLS+K by comparing it with the KPLS model and the ordinary kriging model. For this purpose, we use the academic function, named Griewank, over the interval $[-5, 5]$ which is studied in [13]. 20 and 60 dimensions are considered for this function. In addition, an engineering example, done at Snecma for a multidisciplinary optimization, is used. This engineering case is chosen since it was shown in [13] that KPLS is less accurate than ordinary kriging. The Gaussian kernel is used for all surrogate models used herein, that is, ordinary kriging, KPLS, and KPLS+K. For KPLS and KPLS+K using h principal

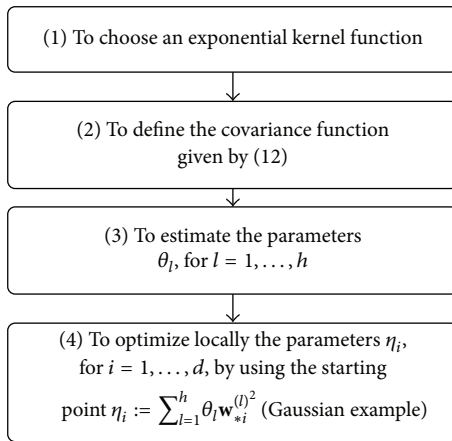


FIGURE 2: Principal stages for building a KPLS+K model.

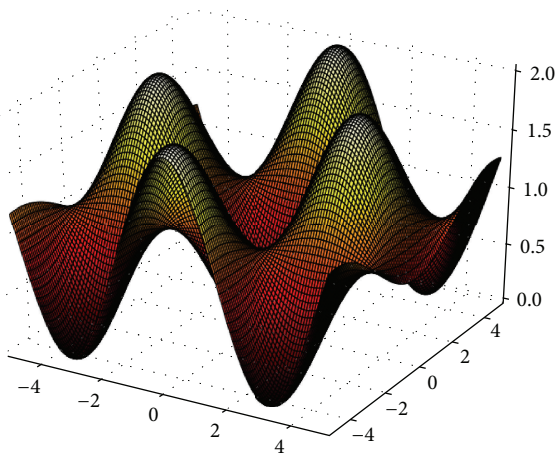


FIGURE 3: A 2D Griewank function over the interval $[-5, 5]$.

components, for $h \leq d$, will be denoted by KPLS h and KPLS h +K, respectively, and this h is varied from 1 to 3. The Python toolbox Scikit-learn v.0.14 [23] is used to achieve the proposed numerical tests, except for ordinary kriging used on the industrial case, where the Optimus version is used. The training and the validation points used in [13] are reused in the following.

4.1. Griewank Function over the Interval $[-5, 5]$. The Griewank function [13, 24] is defined by

$$y_{\text{Griewank}}(\mathbf{x}) = \sum_{i=1}^d \frac{x_i^2}{4000} - \prod_{i=1}^d \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1, \quad (17)$$

$$-5 \leq x_i \leq 5, \text{ for } i = 1, \dots, d.$$

Figure 3 shows the degree of complexity of such function which is very multimodal. As in [13], we consider $d = 20$ and $d = 60$ input variables. For each problem, ten experiments based on the random Latin-hypercube design are built with n (number of sampling points) equal to 50, 100, 200, and 300.

To better visualize the results, boxplots are used to show the CPU time and the relative error RE given by

$$\text{Error} = \frac{\|\hat{\mathbf{Y}} - \mathbf{Y}\|_2}{\|\mathbf{Y}\|_2} 100, \quad (18)$$

where $\|\cdot\|_2$ represents the usual L^2 norm and $\hat{\mathbf{Y}}$ and \mathbf{Y} are the vectors containing the prediction and the real values of 5000 randomly selected validation points for each case. The mean and the standard error are given in Tables 2 and 3, respectively, in Appendix. However, the results of the ordinary kriging model and the KPLS model are reported from [13].

For 20 input variables and 50 sampling points, the KPLS models always give a more accurate solution than ordinary kriging and KPLS+K, as shown in Figure 4(a). Indeed, the best result is given by KPLS3 with a mean of RE equal to 0.51%. However, the KPLS+K models give more accurate models than ordinary kriging in this case (0.58% for KPLS2+K and KPLS3+K versus 0.62% for ordinary kriging). For the KPLS model, the rate of improvement with respect to the number of sampling points is less than for ordinary kriging and KPLS+K (see Figures 4(b)–4(d)). As a result, KPLS h +K, for $h = 1, \dots, 3$, and ordinary kriging give almost the same accuracy ($\approx 0.16\%$) when 300 sampling points are used (as shown in Figure 4(d)), whereas the KPLS models give a RE of 0.35% as a best result, when $h = 3$.

Nevertheless, the results shown in Figure 5 indicate that the KPLS+K models lead to an important reduction in CPU time for the various number of sampling points compared to ordinary kriging. For instance, 20.49 s are required for building KPLS3+K when 300 training points are used, whereas ordinary kriging is built in 94.31 s; in this case, KPLS3+K is thus approximately 4 times cheaper than the ordinary kriging model. Moreover, the computational time required for building KPLS+K is more stable than the computational time for building ordinary kriging; standard deviations of approximately 3 s for KPLS+K and 22 s for the ordinary kriging model are observed.

For 60 input variables and 50 sampling points, a slight difference of the results occurs compared to the 20 input variables case (Figure 6(a)). Indeed, the KPLS models remain always better, with a mean of RE approximately equal to 0.92%, than KPLS+K and ordinary kriging. However, the KPLS+K models give more accurate results than ordinary kriging with an accuracy close to that of KPLS ($\approx 0.99\%$ versus 1.39%). Increasing the number of sampling points, the accuracy of ordinary kriging becomes better than the accuracy given by the KPLS models, but it remains less accurate than for the KPLS h +K models, for $h = 2$ or 3. For instance, we obtain a mean of RE with 0.60% for KPLS2+K against 0.65% for ordinary kriging (see Figure 6(d)), when 300 sampling points are used.

As we can observe from Figure 7(d), a very important reduction in terms of computational time is obtained. Indeed, a mean time of 2894.56 s is required for building ordinary kriging, whereas KPLS2+K is built in 23.03 s; KPLS2+K is approximately 125 times cheaper than ordinary kriging in this case. In addition, the computational time for building

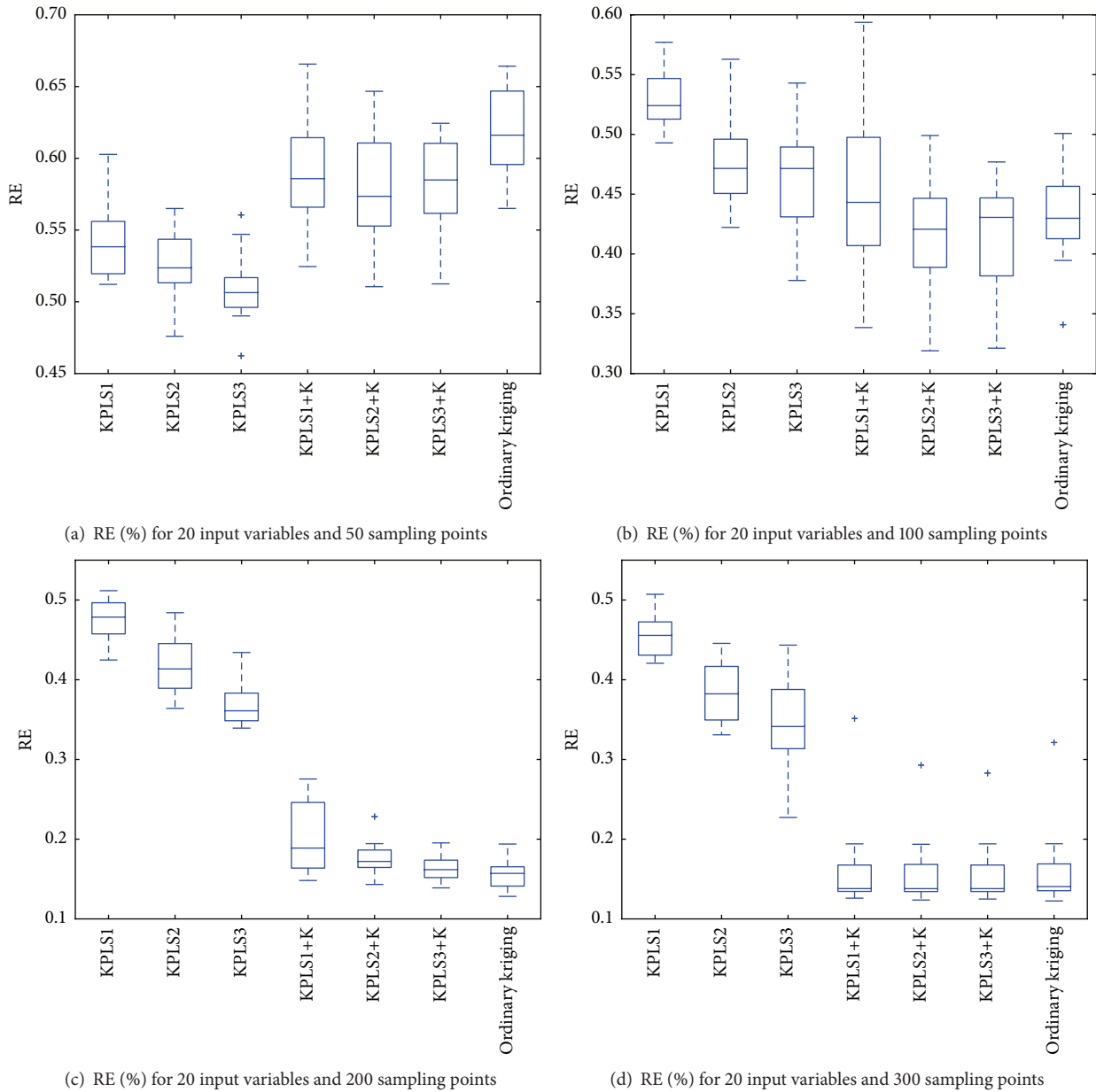


FIGURE 4: RE of the Griewank function in 20D over the interval $[-5, 5]$. The experiments are based on the 10-Latin-hypercube design.

KPLS+K is more stable than ordinary kriging, except the KPLS3+K case; a standard deviation of approximately 0.30 s for KPLS1+K and KPLS2+K is observed, against 728.48 s for ordinary kriging. However, the relatively large standard of deviation of KPLS3+K (26.59 s) is probably due to the dispersion caused by KPLS3 (26.59 s). But, it remains too lower than the standard deviation of the ordinary kriging model.

For the Griewank function over the interval $[-5, 5]$, the KPLS+K models are slightly more time consuming than the KPLS models, but they are more accurate, in particular when the number of observations is greater than the dimension d , as is implied by the rule-of-thumb $n = 10d$ in [12]. They seem

to perform well compared to the ordinary kriging model with an important gain in terms of saving CPU time.

4.2. Engineering Case. In this section, let us consider the third output, y_3 , from tab_1 problem studied in [13]. This test case is chosen because the KPLS models, from 1 to 3 principal components, do not perform well (see Table 1). We recall that this problem contains 24 input variables. 99 training points and 52 validation points are used and the relative error (RE) given by (18) is considered.

As we see in Table 1, we improve the accuracy of KPLS by adding the step for building KPLS+K. This improvement is verified whatever the number of principal components

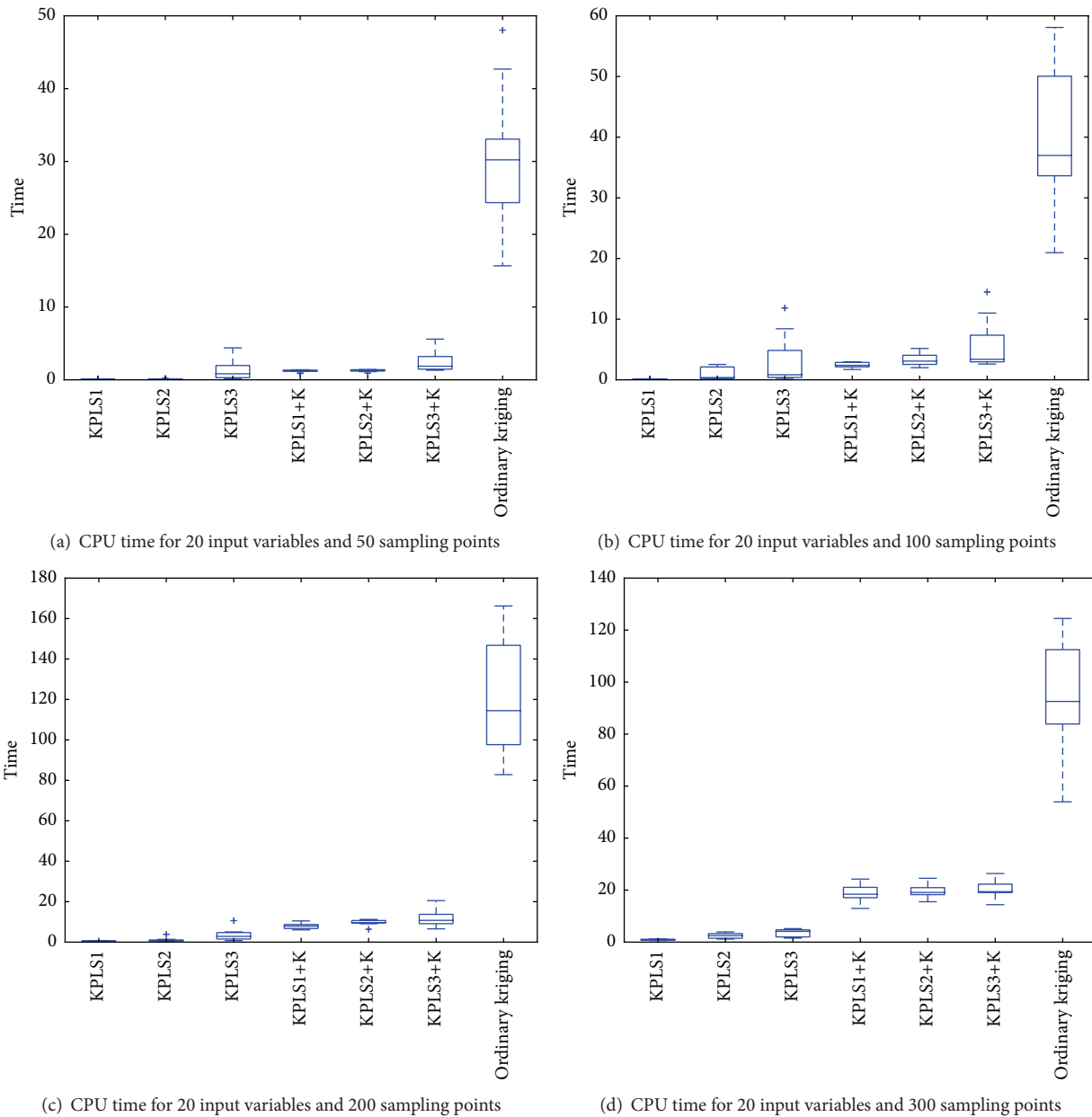


FIGURE 5: CPU time of the Griewank function in 20D over the interval $[-5, 5]$. The experiments are based on the 10-Latin-hypercube design.

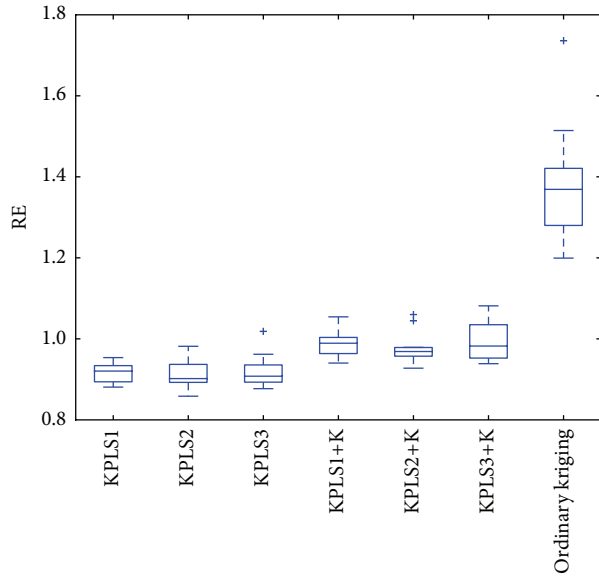
used (1, 2, and 3 principal components). For these three models, a better accuracy is even found than the ordinary kriging model. The computational time required to build KPLS+k is approximately twice lower than the time required for ordinary kriging.

5. Conclusions

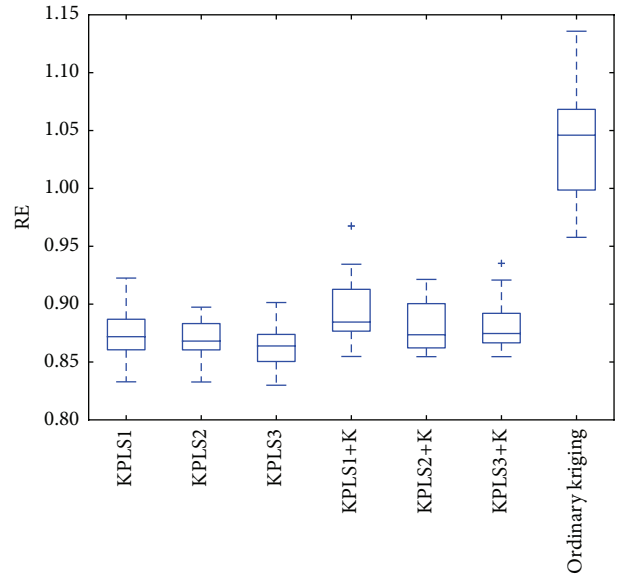
Motivated by the need to accurately approximate high-fidelity codes rapidly, we develop a new technique for building the kriging model faster than classical techniques used in literature. The key idea for such construction relies on the choice of the start point for optimizing the likelihood function of

the kriging model. For this purpose, we firstly prove equivalence between KPLS and kriging when an exponential covariance function is used. After optimizing hyperparameters of KPLS, we then choose the solution obtained as an initial point to find the MLE of the equivalent kriging model. This approach will be applicable only if the kernels used for building KPLS are of the exponential type with the same form.

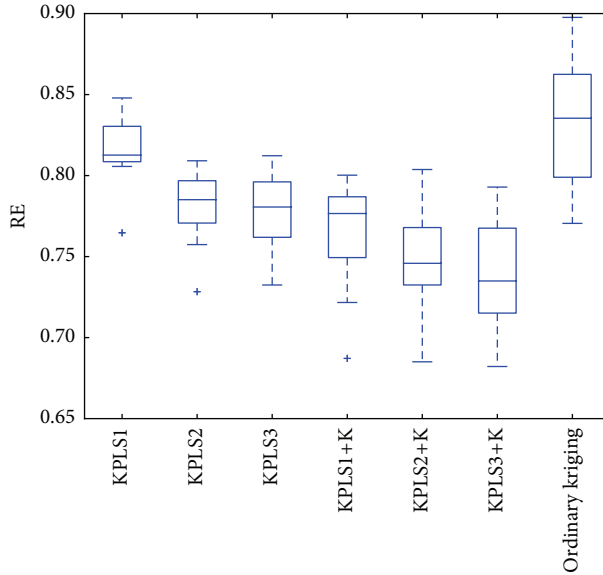
The performance of KPLS+K is verified in the Griewank function over the interval $[-5, 5]$ with 20 and 60 dimensions and an industrial case from Snecma, where the KPLS models do not perform well in terms of accuracy. The results of KPLS+K have shown a significant improvement in terms of



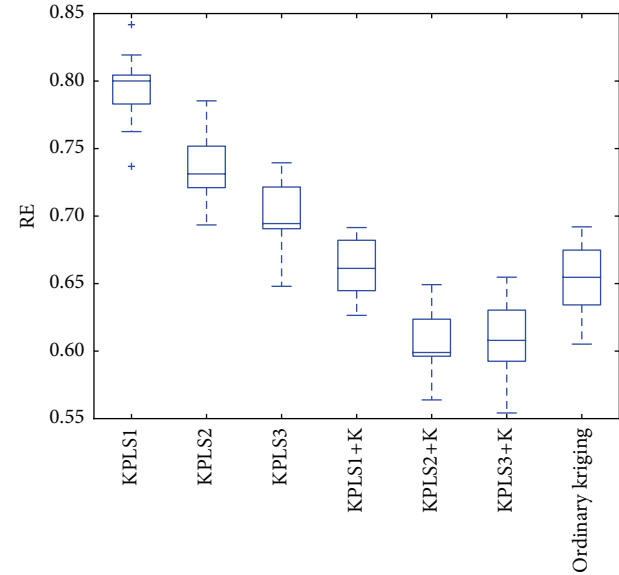
(a) RE (%) for 60 input variables and 50 sampling points



(b) RE (%) for 60 input variables and 100 sampling points



(c) RE (%) for 60 input variables and 200 sampling points



(d) RE (%) for 60 input variables and 300 sampling points

FIGURE 6: RE of the Griewank function in 60D over the interval $[-5, 5]$. The experiments are based on the 10-Latin-hypercube design.

accuracy compared to the results of KPLS, at the cost of a slight increase in computational time. We have also seen, in some cases, that accuracy of KPLS+K is even better than accuracy given by the ordinary kriging model.

Appendix

Results of Griewank Function in 20D and 60D over the Interval $[-5, 5]$

In Tables 2 and 3, the mean and the standard deviation (std) of the numerical experiments with the Griewank function are given for 20 and 60 dimensions, respectively.

Symbols and Notation (Matrices and Vectors Are in Bold Type)

- $|\cdot|$: Absolute value
- \mathbb{R} : Set of real numbers
- \mathbb{R}^+ : Set of positive real numbers
- n : Number of sampling points
- d : Dimension
- h : Number of principal components retained
- \mathbf{x} : A $1 \times d$ vector
- x_i : The i th element of a vector \mathbf{x}
- \mathbf{X} : A $n \times d$ matrix containing sampling points

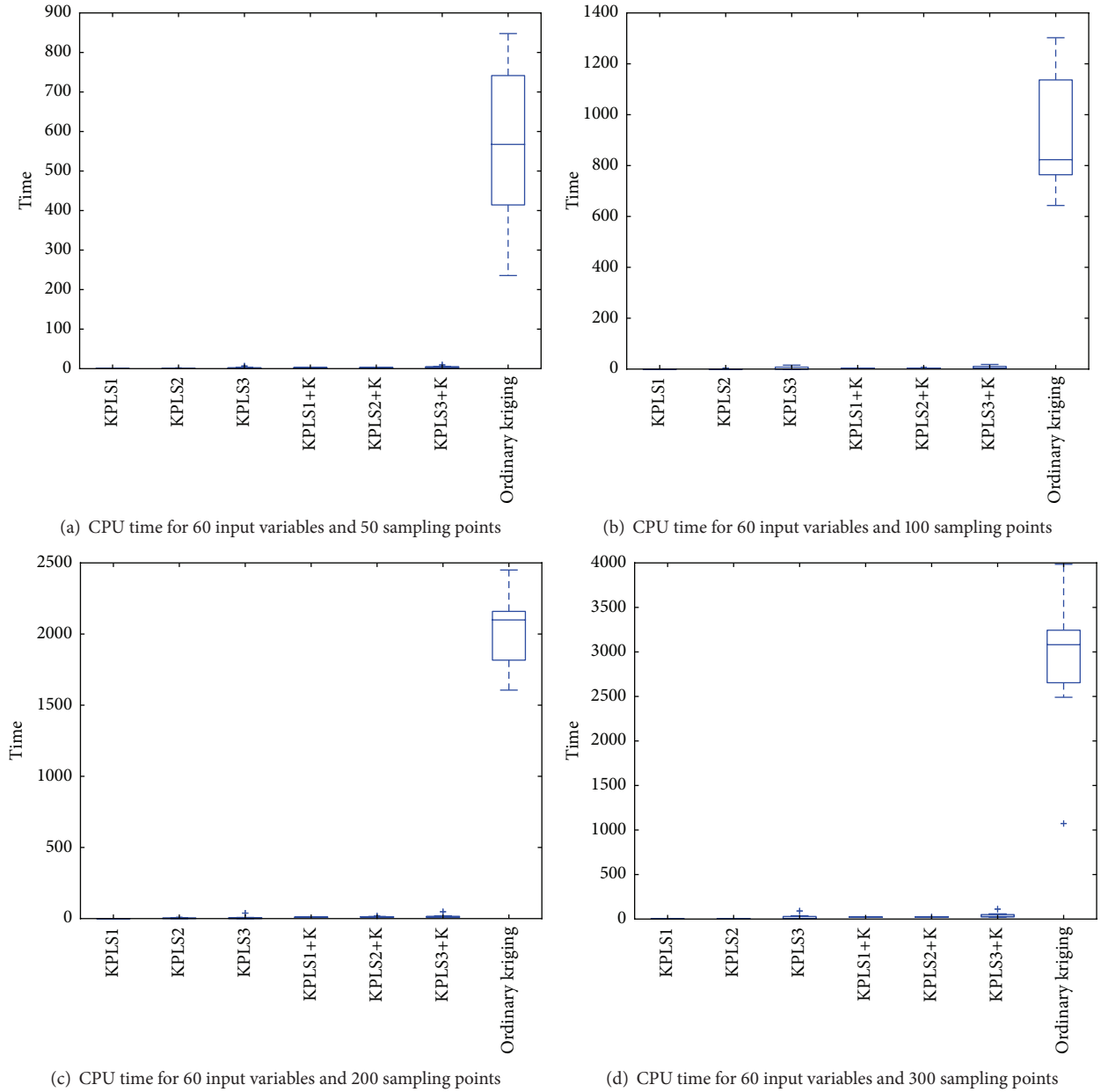


FIGURE 7: CPU time of the Griewank function in 60D over the interval $[-5, 5]$. The experiments are based on the 10-Latin-hypercube design.

$y(\mathbf{x})$: The true function y performed on the vector \mathbf{x}

\mathbf{y} : $n \times 1$ vector containing simulation of \mathbf{X}

$\hat{y}(\mathbf{x})$: The prediction of the true function $y(\mathbf{x})$

$Y(\mathbf{x})$: A stochastic process

$\mathbf{x}^{(i)}$: The i th training point for $i = 1, \dots, n$ (a $1 \times d$ vector)

$\mathbf{w}^{(l)}$: A $d \times 1$ vector containing X -weights given by the l th PLS-iteration for $l = 1, \dots, h$

$\mathbf{X}^{(0)}$: \mathbf{X}

$\mathbf{X}^{(l-1)}$: Matrix containing residual of the inner regression of the $(l-1)$ th PLS-iteration for $l = 1, \dots, h$

$k(\cdot, \cdot)$: A covariance function

\mathbf{x}^t : Superscript t denoting the transpose operation of the vector \mathbf{x}

\approx : Approximately sign.

Competing Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors extend their grateful thanks to A. Chiplunkar from ISAE-SUPAERO, Toulouse, for his careful correction of the paper.

References

- [1] D. Krige, "A statistical approach to some basic mine valuation problems on the Witwatersrand," *Journal of the Chemical, Metallurgical and Mining Society*, vol. 52, pp. 119–139, 1951.
- [2] G. Matheron, "Principles of geostatistics," *Economic Geology*, vol. 58, no. 8, pp. 1246–1266, 1963.
- [3] N. Cressie, "Spatial prediction and ordinary kriging," *Mathematical Geology*, vol. 20, no. 4, pp. 405–421, 1988.
- [4] J. Sacks, S. B. Schiller, and W. J. Welch, "Designs for computer experiments," *Technometrics*, vol. 31, no. 1, pp. 41–47, 1989.
- [5] C. E. Rasmussen and C. K. Williams, *Gaussian Processes for Machine Learning, Adaptive Computation and Machine Learning*, MIT Press, Cambridge, Mass, USA, 2006.
- [6] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box functions," *Journal of Global Optimization*, vol. 13, no. 4, pp. 455–492, 1998.
- [7] S. Sakata, F. Ashida, and M. Zako, "Structural optimization using Kriging approximation," *Computer Methods in Applied Mechanics and Engineering*, vol. 192, no. 7–8, pp. 923–939, 2003.
- [8] A. Forrester, A. Sobester, and A. Keane, *Engineering Design via Surrogate Modelling: A Practical Guide*, John Wiley & Sons, New York, NY, USA, 2008.
- [9] J. Laurenceau, *Kriging based response surfaces for aerodynamic shape optimization [Thesis]*, Institut National Polytechnique de Toulouse (INPT), Toulouse, France, 2008.
- [10] J. P. C. Kleijnen, W. van Beers, and I. van Nieuwenhuysse, "Constrained optimization in expensive simulation: novel approach," *European Journal of Operational Research*, vol. 202, no. 1, pp. 164–174, 2010.
- [11] J. P. Kleijnen, W. van Beers, and I. van Nieuwenhuysse, "Expected improvement in efficient global optimization through bootstrapped kriging," *Journal of Global Optimization*, vol. 54, no. 1, pp. 59–73, 2012.
- [12] J. L. Loepky, J. Sacks, and W. J. Welch, "Choosing the sample size of a computer experiment: a practical guide," *Technometrics*, vol. 51, no. 4, pp. 366–376, 2009.
- [13] M. A. Bouhlel, N. Bartoli, A. Otsmane, and J. Morlier, "Improving kriging surrogates of high-dimensional design models by partial least squares dimension reduction," *Structural and Multidisciplinary Optimization*, vol. 53, no. 5, pp. 935–952, 2016.
- [14] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn, "Design and analysis of computer experiments," *Statistical Science*, vol. 4, no. 4, pp. 409–435, 1989.
- [15] M. Sasena, *Flexibility and efficiency enhancements for constrained global design optimization with Kriging approximations [Ph.D. thesis]*, University of Michigan, 2002.
- [16] V. Picheny, D. Ginsbourger, Y. Richet, and G. Caplin, "Quantile-based optimization of noisy computer experiments with tunable precision," *Technometrics*, vol. 55, no. 1, pp. 2–13, 2013.
- [17] O. Roustant, D. Ginsbourger, and Y. Deville, "DiceKriging, DiceOptim: two R packages for the analysis of computer experiments by kriging-based metamodeling and optimization," *Journal of Statistical Software*, vol. 51, no. 1, pp. 1–55, 2012.
- [18] J. P. Kleijnen, *Design and Analysis of Simulation Experiments*, vol. 230, Springer, Berlin, Germany, 2015.
- [19] I. S. Helland, "On the structure of partial least squares regression," *Communications in Statistics. Simulation and Computation*, vol. 17, no. 2, pp. 581–607, 1988.
- [20] I. E. Frank and J. H. Friedman, "A statistical view of some chemometrics regression tools," *Technometrics*, vol. 35, no. 2, pp. 109–135, 1993.
- [21] R. A. Pérez and G. González-Farías, "Partial least squares regression on symmetric positive-definite matrices," *Revista Colombiana de Estadística*, vol. 36, no. 1, pp. 177–192, 2013.
- [22] R. Manne, "Analysis of two partial-least-squares algorithms for multivariate calibration," *Chemometrics and Intelligent Laboratory Systems*, vol. 2, no. 1–3, pp. 187–197, 1987.
- [23] F. Pedregosa, G. Varoquaux, A. Gramfort et al., "Scikit-learn: machine learning in Python," *Journal of Machine Learning Research (JMLR)*, vol. 12, pp. 2825–2830, 2011.
- [24] R. G. Regis and C. A. Shoemaker, "Combining radial basis function surrogates and dynamic coordinate search in high-dimensional expensive black-box optimization," *Engineering Optimization*, vol. 45, no. 5, pp. 529–555, 2013.

4.4.4 Approximation of the critical buckling factor for composite panels, Structural and Multidisciplinary Optimization 2012

Approximation of the critical buckling factor for composite panels

Dimitri Bettebghor · Nathalie Bartoli

Received: 10 August 2011 / Revised: 11 January 2012 / Accepted: 15 February 2012 / Published online: 15 March 2012
© Springer-Verlag 2012

Abstract This article is concerned with the approximation of the critical buckling factor for thin composite plates. A new method to improve the approximation of this critical factor is applied based on its behavior with respect to lamination parameters and loading conditions. This method allows accurate approximation of the critical buckling factor for non-orthotropic laminates under complex combined loadings (including shear loading). The influence of the stacking sequence and loading conditions is extensively studied as well as properties of the critical buckling factor behavior (e.g concavity over tensor D or out-of-plane lamination parameters). Moreover, the critical buckling factor is numerically shown to be piecewise linear for orthotropic laminates under combined loading whenever shear remains low and it is also shown to be piecewise continuous in the general case. Based on the numerically observed behavior, a new scheme for the approximation is applied that separates each buckling mode and builds linear, polynomial or rational regressions for each mode. Results of this approach and applications to structural optimization are presented.

Keywords Buckling · Lamination parameters · Piecewise polynomial regression

1 Introduction

Structural optimization for composite thin-walled structures often exhibits large computational times due to repetitive stability analysis. For composite structure made of thin plates or thin shallow shells (such as aircraft fuselage), buckling computation is of primary importance since it is one of the critical sizing constraints when minimizing the weight. This computational burden becomes even critical when we address laminated composites and when the stacking sequence is optimized. The prescribed orientations and the feasibility rules for stacking sequences make this kind of problem NP-hard. To solve such problems, one often goes to population-based heuristics (evolutionary algorithms, particle swarm optimization...) to address the discreteness of the design variables. Such methods require many evaluations to ensure that a reasonably good optimum has been found. To circumvent such issues, one classical strategy uses a continuous representation of the in-plane and out-of-plane behaviors of the stacking sequence by means of either lamination parameters, polar invariant representation or directly A and D tensors terms. Lamination parameters are dimensionless quantities that only depend on the stacking sequence. Unlike A and D tensors, their use allows to treat exclusively the dependence with respect to stacking sequences and not the material dependence. We distinguish in-plane lamination parameters (ξ_A), bending-stretching lamination parameters (ξ_B) and out-of-plane (or bending or flexural) lamination parameters (ξ_D), leading to 12 variables for arbitrary stacking sequences. This continuous relaxation offers a way to treat the optimization (for instance the maximization of the critical buckling factor of a laminate plate under a given loading) as a continuous one, reducing the number of evaluations needed to converge whenever the optimization problem is well posed (e.g.

D. Bettebghor (✉)
The French Aerospace Lab, Onera, 92260, Chatillon, France
e-mail: dimitri.bettebghor@onera.fr

N. Bartoli
The French Aerospace Lab, Onera, 31055, Toulouse, France
e-mail: nathalie.bartoli@onera.fr

convex). A good property of lamination parameters is that most classical objective functions have interesting optimization properties when evaluated over the lamination space, which happens to be convex. However, this continuous relaxation has two major drawbacks. First when a continuous optimum is found, it usually does not match a discrete solution, there is no practical stacking sequence that gives these optimal lamination parameters. A new discrete optimization step is to be performed to identify a correct discrete solution and there is no evidence that the discrete optimum found by this step is the real discrete optimum. The other important drawback is that not all sizing constraints can be computed on the basis of lamination parameters. Some optimization constraints do need the real stacking sequence to be computed: mechanical constraints (first-ply failure...), feasibility constraints (contiguity of same orientations to avoid matrix cracking for instance). However, lamination parameters do provide a practical continuous representation of in-plane and out-of-plane behavior making possible to build approximation models of the expensive optimization constraints, such as buckling computations. Even though continuous optimization over lamination is not performed, such an approximation model allows to perform discrete optimizations in a much more efficient way.

Since their first application to the design of composite in the seminal work of Miki (1983) and Miki and Sugiyama (1991), where the author derives a graphical procedure to optimize stiffness properties, lamination parameters have been extensively studied and covered as a practical and convenient tool to optimize laminated composites. Early work by Miki allowed to bound the lamination space. Much work was also done in describing the dependence of the critical buckling factor in the lamination space: Grenestedt (1989, 1990, 1991) or Walker et al. (1996) investigated the influence of lamination parameters for the critical buckling factor in different cases (pure shear loading, non-orthotropic laminates, ...) through different approximation methods: orthotropic closed-form expression, finite strip method, Rayleigh–Ritz method... Grenestedt found out that non-orthotropy decreases the critical buckling factor for uni-axial compression loading and in many situations the optimum in the lamination space belongs to the boundary previously described by Miki, corresponding to angle-ply laminate $[+\theta/-\theta]$ with continuous angle θ . However practical stacking sequences that satisfy manufacturability constraints are not orthotropic nor angle-ply stacking sequences. A lot of work was also done in applying directly the optimization problem in the lamination space and then get from the continuous optimum a discrete solution usually by means of genetic algorithms (Fukunaga et al. 1995; Harrison et al. 1995). Lamination parameters associated to real practical stacking sequences of orientations $[0^\circ/45^\circ/-$

$45^\circ/90^\circ]$ exhibit a fractal structure that was exploited in Terada et al. (2001) to derive a fractal branch-and-bound method. The precise description of the feasible set of lamination parameters and more specifically the coupling between in-plane lamination parameters and out-of-plane lamination parameters was solved in Diaconu et al. (2002) and Bloomfield et al. (2009), resulting into so-called compatibility equations that constrain the feasible space for out-of-plane lamination parameters based on the values of in-plane lamination parameters. These compatibility equations describe a convex feasible set and can be directly integrated into an optimization (Herencia et al. 2008b). Regarding the area of building approximations of the critical buckling factor over lamination parameters, much work was done with the help of response surfaces (polynomial regression of degree 2; Todoroki et al. 2003; Abu-Odeh and Jones 1998; Harrison et al. 1995), Taylor-based first-order linear approximation (Herencia et al. 2008b), adaptive response surface (Todoroki and Sasai 2002). The non-differentiability of the critical buckling factor with respect to lamination parameters, which was already noticed in a more general framework (see Rousselet and Chenaïs (1990) or De Gournay (2005) for general results on critical buckling factor differentiability not specifically devoted to composite) does not seem to have been considered when building approximation models. Regarding potential property of the critical buckling factor well suited for optimization, concavity was cited in several references (Autio 2001), but to the authors' knowledge, it has never been proved on a sound mathematical basis for the general case (arbitrary geometry, non-orthotropic laminate).

This paper aims at studying numerically some properties of the critical buckling factor that will ease the construction of accurate approximations to be used within an optimization process. The approximation of the critical buckling factor over lamination parameters is not restricted to continuous optimization over lamination parameters, since such an approximation can be used when the stacking is directly addressed as an optimization variable (genetic algorithms, simulated annealing, particle swarm optimization). In particular, we show that the critical buckling factor is concave over lamination parameters. We also give numerical evidence of the piecewise behavior of the critical buckling factor depending on the region of the input space. Based on this knowledge we apply a recent strategy to build piecewise approximation models developed in Bettebghor et al. (2011). The behavior of the critical buckling factor both over lamination parameters and varying loading conditions is covered and the suggested scheme seems to be well suited to approximate such derivative-discontinuous functions, applications to optimization (maximization of the critical buckling factor) are also given.

2 Lamination parameters and feasible stacking sequences

We briefly recall here the definition of out-of-plane lamination parameters and the feasibility constraints for laminated composites that we use in this paper.

2.1 Out-of-plane lamination parameters

Consider a composite of given material characteristics E_1, E_2, G_{12} and ν_{12} . Note Q_{11}, Q_{22}, Q_{12} and Q_{66} the reduced stiffness's. The Tsai–Pagano parameters $(U_i)_{i=1...5}$ (or material invariants) are defined as

$$\begin{pmatrix} U_1 \\ U_2 \\ U_3 \\ U_4 \\ U_5 \end{pmatrix} = \begin{pmatrix} 3/8 & 3/8 & 1/4 & 1/2 \\ 1/2 & -1/2 & 0 & 0 \\ 1/8 & 1/8 & -1/4 & -1/2 \\ 1/8 & 1/8 & 3/4 & -1/2 \\ 1/8 & 1/8 & -1/4 & 1/2 \end{pmatrix} \begin{pmatrix} Q_{11} \\ Q_{22} \\ Q_{12} \\ Q_{66} \end{pmatrix} \quad (1)$$

the out-of-plane (or bending) lamination parameters are defined as

$$\xi_D^{\{1,2,3\}} = \frac{12}{h^3} \int_{-h/2}^{h/2} \{\cos(2\theta(z)), \cos(4\theta(z)), \sin(2\theta(z))\} dz \quad (2)$$

Note that we do not consider ξ_D^4 since it vanishes when computed over stacking sequences made of $[0^\circ/45^\circ/90^\circ]$ orientations. Now recall that the D tensor is expressed as an affine function of the lamination parameters as follows.

$$D_{11} = \frac{h^3}{12} (U_1 + U_2 \xi_1^D + U_3 \xi_2^D) \quad (3)$$

$$D_{22} = \frac{h^3}{12} (U_1 - U_2 \xi_1^D + U_3 \xi_2^D) \quad (4)$$

$$D_{12} = \frac{h^3}{12} (U_4 - U_3 \xi_2^D) \quad (5)$$

$$D_{66} = \frac{h^3}{12} (U_5 - U_3 \xi_2^D) \quad (6)$$

$$D_{16} = \frac{h^3}{24} U_2 \xi_3^D \quad (7)$$

$$D_{26} = \frac{h^3}{24} U_2 \xi_3^D \quad (8)$$

2.2 Feasible stacking sequences

We are interested here in conventional laminates of fiber orientations $[0^\circ/45^\circ/-45^\circ/90^\circ]$. We used the following

common feasibility rules that both come from mechanical properties and manufacturability constraints:

- Symmetric laminates. This ensures that the B tensor vanishes. Note that there exist laminates with $B = 0$ that are not symmetric.
- Balanced laminates. The number of 45° layers is equal to the number of -45° layers. This makes the A_{16} and A_{26} terms of the A tensor vanish.
- Contiguity rule. No more than 4 successive layers of the same orientation.
- Disorientation rule. No more than 45° of angle difference between 2 successive layers. In particular the sequence $[45^\circ/-45^\circ]$ is not allowed although many references in the literature on composite design allow it.
- Percentages rule. Each orientation should be represented in the laminate. We took 8% as a minimum percentage of each orientation.

Based on these rules, the number of feasible stacking sequences remains relatively low for thin laminates. The number of feasible stacking sequences based on these rules for a 2×10 layers symmetric laminate is for instance 7,032. We found numerically that we could very easily exhaust all feasible stacking sequences based on vector-oriented programming language (such as Matlab) up to symmetric laminates of 34 layers (2×17). Our concern here is not in finding such feasible stackings but it should be noted that for thicker laminates this exhaustive search might take excessive memory resources unless proportions are given and a discrete optimization is preferable. However, for relatively

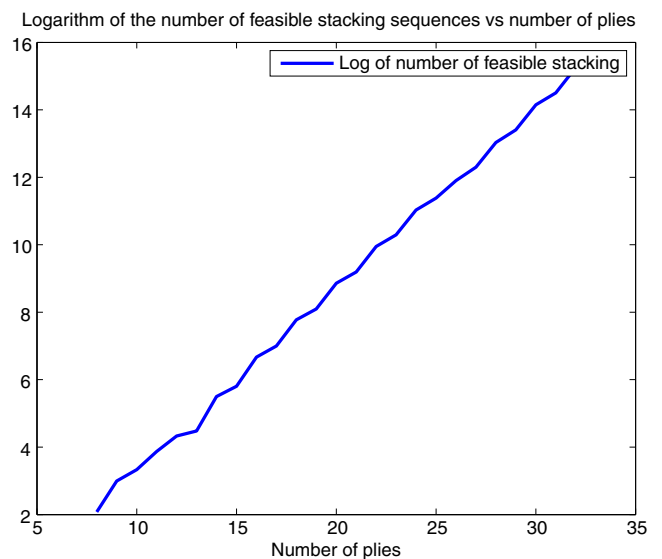


Fig. 1 Logarithm of the number of feasible stacking sequences from 8 to 34 plies

thin laminates (such as the ones used for aircraft fuselage or wing) we consider that we have all feasible stacking sequences and that an exact optimization can be performed by listing all the feasible laminates. We depicted Fig. 1 the logarithm of the number of feasible stacking sequences from 8 to 34 plies. This number $N_{\text{feas}}(n)$ where n stands for the number of plies can be approximated by

$$N_{\text{feas}}(n) = K 1.72^n \quad (9)$$

where $K \simeq 0.12$.

3 Behavior of the critical buckling factor

In this section, we first recall some basic facts on buckling, including the definition of the critical buckling factor for thin composite plates through variational formulation. Based on this definition we show that this critical buckling factor is concave over bending lamination parameters. We then depict two interesting aspects: firstly, the typical behavior of the critical buckling factor when the stacking sequence varies and secondly the typical behavior of its reciprocal when the loading conditions vary.

3.1 Generalities on buckling

The buckling phenomenon for thin plates and shells has been extensively studied since the beginning of the 20th century. The first rigorous equations for large deflections of thin plates were written down by von Kármán in 1910. In their first derivation for isotropic homogeneous structures, these equations were a system of two nonlinear partial differential equations of order four. As outlined in Antman (2006), the mathematical analysis used to solve these equations involved development of bifurcation theory, as it can be found in Bloom and Coffin (2000), Lebedev and Vorovich (2003) and Antman and Renardy (2005). A very brief introduction to bifurcation theory not specifically devoted to buckling can be found in Stakgold (1971). The major idea is that bifurcation points of the nonlinear partial differential equation's (or stability points where the trivial solution $w = 0$ ceases to be unique) are to be found amongst the points where the linearized operator is not invertible thanks to implicit functions theorem. The topological Leray–Schauder degree theory offers the other way around. Namely, the points where the linearized operator is not invertible (precisely the eigenelements of the linearized operator) are actually bifurcation points of the original nonlinear pde system. We will not delve into the mathematical theory but this brief explanation means that buckling is a nonlinear phenomenon that is often partially solved through

a linear eigenvalue problem. It is worth noting that the linearized buckling can not describe the behavior of the material after buckling, it only predicts the onset of buckling. Being an eigenfunction, the buckled transverse displacement is known up to a multiplicative constant, which means that no maximal transverse displacement can be computed through linear buckling. To study the behavior of plates or shells after buckling one should go to nonlinear analysis, the first work on post-buckling can be found in the seminal paper (Koiter 1962). Our concern here is on the prediction of buckling, we will therefore study buckling of laminated composite plates into the framework of linear buckling. We will only consider symmetric laminates, which means that the B tensor vanishes and hence that there is no coupling between in-plane and out-of-plane behaviors. Under Kirchhoff hypothesis, one can show (see Turvey 1995) that the equation for buckling of a simply supported composite plate leads to the following eigenvalue problem. The *critical buckling factor* is defined as the smallest positive eigenvalue λ_1 of

$$\begin{aligned} D_{11} \frac{\partial^4 w}{\partial x^4} + D_{22} \frac{\partial^4 w}{\partial y^4} + 2(D_{12} + 2D_{66}) \frac{\partial^4 w}{\partial x^2 \partial y^2} \\ + 4D_{16} \frac{\partial^4 w}{\partial x^3 \partial y} + 4D_{26} \frac{\partial^4 w}{\partial y^3 \partial x} \\ = \lambda_1 \left(N_x \frac{\partial^2 w}{\partial x^2} + N_y \frac{\partial^2 w}{\partial y^2} + N_{xy} \frac{\partial^2 w}{\partial x \partial y} \right) \end{aligned} \quad (10)$$

where w is the transverse displacement, with $w = 0$ on $\partial\Omega$ and $\Delta w = 0$ on $\partial\Omega$. N_x , N_y and N_{xy} are the loads per unit length respectively for longitudinal, lateral and shear loading, and equivalently the *critical buckling load* N_{cr} is defined as $N_{\text{cr}} = \lambda_{\text{cr}}(N_x, N_y, N_{xy})$. Note that (10) is valid for any geometry Ω . For the numerical experiments in this article, we only focus on rectangular plates but, in general, we can already see that buckling essentially depends on three different types of parameters:

- Ω : the domain where the buckling (10) is defined. Dependence of eigenvalue problems to the geometry is quite complicated. For rectangular plates $\Omega = [0, a] \times [0, b]$, the eigenvalues depend only on the aspect ratio $\frac{a}{b}$.
- D tensor: the idealized buckling problem studied in shape optimization (in Ashbaugh 1999 for instance) only defines bi-Laplacian and Laplacian operators. This corresponds to buckling for isotropic homogeneous material under bi-axial compression of the same magnitude ($N_x = N_y < 0$). For composite, the D tensor makes the buckling equation somewhat more complex since it weights the different derivatives and we clearly see that the D_{16} and D_{26} (bending-twisting terms) are not part of the bi-laplacian.

- $N = (N_x, N_y, N_{xy})$: the loading conditions. When $N_{xy} = 0$ and $N_x = N_y < 0$, note that the right-hand differential operator boils down to a Laplacian.

Buckling problems of this type can be solved exactly in a few cases with simple geometry (rectangles, disk, ...) and simple operators (bi-Laplacian that corresponds to orthotropy, Laplacian that corresponds to uniform bi-axial compression with no shear), many closed-form expressions can be found in Lekhnitskii (1968). In the orthotropic case, where $D_{16} = D_{26} = 0$, buckling of a rectangular composite plate of dimension $[0, a] \times [0, b]$, under bi-axial loading (no shear), the critical buckling factor is

$$\lambda_1 = \min_{m,n \in \mathbb{N}^2} \frac{\pi^2 (D_{11}m^4 + 2(D_{12} + 2D_{66})m^2n^2R^2 + D_{22}n^4R^4)}{a^2 (N_x m^2 + N_y n^2 R^2)} \tag{11}$$

where $R = a/b$ is the aspect ratio and n (resp. m) is the number of half-waves of the transverse displacement along x direction (resp. y direction). When no closed-form expression can be derived, more complex geometry, non-orthotropic laminates, shear loading, approximation methods are to be used. Amongst the classical methods, we have

- Rayleigh–Ritz method: historically, this was the first method developed to tackle with eigenvalue problem arising in vibration and buckling. This yields to solve a generalized eigenvalue problem over a finite-dimensional space which is the span of natural eigenfunctions for a simple operator, e.g $\sin(\frac{i\pi x}{a}) \sin(\frac{j\pi y}{b})$ for a rectangular plate. The idea is therefore to write the variational characterization of eigenvalue as an optimization problem solved in this finite-dimensional space. The Rayleigh–Ritz method can be thought as a perturbation method (see Appendix B). General introduction can be found in Courant (1994), Akulenko and Nesterov (2005) and Turvey (1995) for composite buckling,
- Finite strip method (see Turvey 1995),
- Finite elements method.

In general, the approximation space of finite dimension \mathcal{V}_h is included in the \mathcal{V} Hilbert space where the solution w belongs to (conformal approximation). Under this usual property, the exact eigenvalues can be shown to be approximated by above through these different approximation methods. In this paper, most of the numerical computations were done with a Rayleigh–Ritz method.

A few basic facts can be noted in (11) regarding the dependence of λ_1 with respect to stacking sequence and loading conditions.

- For fixed loading conditions, critical buckling factor is concave over the D tensor variables, being the minimum of concave functions (linear). We will see in the sequel that this can be established rigorously in the general non-orthotropic case, for any loading and any geometry Ω .
- For fixed stacking sequence, hence fixed D tensor, the reciprocal of the critical buckling factor is piecewise linear. We will see that numerically it seems to be a good approximation for non-orthotropic laminates under bi-axial loading.
- For fixed loading conditions, the critical buckling factor is piecewise linear on D hence on lamination parameters. In the general case (combined loading conditions) this does not seem to be still valid, however, simple piecewise quadratic models over lamination parameters are shown to be very accurate to predict the critical buckling factor.
- Define formally a critical buckling factor function

$$\lambda_{cr} : S_3(\mathbb{R})^{++} \times \mathbb{R}^3 \mapsto \mathbb{R}^+ \\ (D, N_x, N_y, N_{xy}) \mapsto \lambda_1 = \lambda_{cr}(D, N_x, N_y, N_{xy}) \tag{12}$$

where D is the out-of-plane (symmetric positive definite) tensor in $S_3(\mathbb{R})^{++}$ (convex cone of symmetric definite positive tensors) and $N = (N_x, N_y, N_{xy})$ the loading conditions. We see that in the orthotropic case, λ_{cr} does have some partial homogeneity properties (e.g $\lambda_{cr}(\kappa N) = \kappa^{-1} \lambda_{cr}(N)$). Most of these properties are still valid in the general case (non-orthotropic and combined loading), this is showed in the sequel.

3.2 Characterization of the critical buckling factor and properties

In most cases, complex geometry, shear loading, non-orthotropic laminate, no closed-form expression can be derived. One way to derive some properties of the critical buckling factor λ_1 is through variational formulation. Let us define some notations for sake of clarity

- $\mathcal{V} = W_0^{2,2}(\Omega)$, the Hilbert space of feasible displacement satisfying the boundary conditions (clamped).¹

¹For simply supported boundary conditions, the variational formulation remains the same. The main difference is that the boundary condition $\Delta u = 0$ can not be satisfied in the Hilbert space but is imposed in the variational formulation. However, a posteriori, solution u is in $W_0^{2,4}(\Omega)$ and the trace of the Laplacian can be defined.

- $\mathcal{H} = W_0^{2,1}(\Omega)$. Note that a key property is that $\mathcal{V} \hookrightarrow \mathcal{H}$ and this embedding is continuous and compact. This ensures that the spectrum is countable. Moreover \mathcal{V} is dense in \mathcal{H} .
- $a_D : \mathcal{V} \times \mathcal{V} \mapsto \mathbb{R}^+$ a continuous symmetric bilinear coercive form where D is an element of $S_3(\mathbb{R})^{++}$
- $b_N : \mathcal{H} \times \mathcal{H} \mapsto \mathbb{R}$ a continuous symmetric bilinear form that need not be positive with $N = (N_x, N_y, N_{xy})$. A basic assumption to ensure existence of (realistic) positive buckling eigenvalues is that there exists $w \in \mathcal{H}$ such that $b_N(w, w) \geq 0$ (see Appendix C).

In the case of buckling of composite laminate these bilinear forms are

$$a_D(u, v) = \iint_{\Omega} D_{11} \frac{\partial^2 u}{\partial x^2} \frac{\partial^2 v}{\partial x^2} + D_{22} \frac{\partial^2 u}{\partial y^2} \frac{\partial^2 v}{\partial y^2} + 2(D_{12} + D_{66}) \frac{\partial^2 u}{\partial x \partial y} \frac{\partial^2 v}{\partial x \partial y} + 4D_{16} \frac{\partial^2 u}{\partial x^2} \frac{\partial^2 v}{\partial x \partial y} + 4D_{26} \frac{\partial^2 u}{\partial y^2} \frac{\partial^2 v}{\partial x \partial y} dx dy \tag{13}$$

and

$$b_N(u, v) = - \iint_{\Omega} N_x \frac{\partial u}{\partial x} \frac{\partial v}{\partial x} + N_y \frac{\partial u}{\partial y} \frac{\partial v}{\partial y} + \frac{N_{xy}}{2} \left(\frac{\partial u}{\partial x} \frac{\partial v}{\partial y} + \frac{\partial u}{\partial y} \frac{\partial v}{\partial x} \right) dx dy \tag{14}$$

The coercitivity of a_D simply follows from Poincaré’s inequality applied to each component of the gradient of u and from the positivity of the D tensor. We now have the following characterization of λ_1 , known as the Courant–Fisher characterization²

$$\lambda_{cr}(D, N_x, N_y, N_{xy}) = \lambda_1 = \min_{w \in \mathcal{V} \setminus \{0\}} \frac{a_D(w, w)}{|b_N(w, w)|} \tag{15}$$

Let’s first recall a result from Rousselet and Chenais (1990) that states that λ_{cr} is continuous over D, N_x, N_y, N_{xy} . However differentiability can not be established, indeed, what is showed in Rousselet and Chenais (1990) is that λ_{cr} is differentiable whenever it is a simple eigenvalue, in case it is a multiple eigenvalue only directional differentiability can be established.

We can use (15) to derive some homogeneity properties. Let $\kappa > 0$ and $D \in S_3(\mathbb{R})^{++}, N = (N_x, N_y, N_{xy})$, for any $w \in \mathcal{V}$

$$a_{\kappa D}(w, w) = \kappa a_D(w, w) \tag{16}$$

²Note that the original Courant–Fisher characterization does not use absolute value over the bilinear form b_N . This convention, used also in Rousselet and Chenais (1990) does not change the value of λ_{cr} for loading conditions that ‘really’ give rise to buckling, since in that case λ_{cr} is attained over a positive eigenfunction a_D being coercive (see Appendix C). It simply allows to define λ_{cr} everywhere in \mathbb{R}^3 .

and

$$|b_{\kappa N}| = \kappa |b_N(w, w)| \tag{17}$$

Taking the infimum over \mathcal{V} , we get

$$\lambda_{cr}(\kappa D, N) = \kappa \lambda_{cr}(D, N) \tag{18}$$

and

$$\lambda_{cr}(D, \kappa N) = \frac{1}{\kappa} \lambda_{cr}(D, N) \tag{19}$$

Equations (18) and (19) mean that λ_{cr} is *homogeneous of degree 1* over D and *homogeneous of degree -1* over N . Applying Euler identity theorem (see Appendix A) to (18) and (19) whenever λ_{cr} is differentiable gives

$$\sum_{i=1}^6 D_i \frac{\partial \lambda_{cr}}{\partial D_i} = \lambda_{cr} \tag{20}$$

$$\sum_{i=1}^3 N_i \frac{\partial \lambda_{cr}}{\partial N_i} = -\lambda_{cr}. \tag{21}$$

These equations can be used within an optimization process to save sensitivity analysis computations. Note that by simply rewriting (19) in terms of the reciprocal of the critical buckling factor

$$\frac{1}{\lambda_{cr}(D, \kappa N)} = \frac{\kappa}{\lambda_{cr}(D, N)} \tag{22}$$

and we see that the reciprocal of λ_{cr} is homogeneous of degree 1, which yields

$$N_x \frac{\partial \frac{1}{\lambda_{cr}}}{\partial N_x} + N_y \frac{\partial \frac{1}{\lambda_{cr}}}{\partial N_y} + N_{xy} \frac{\partial \frac{1}{\lambda_{cr}}}{\partial N_{xy}} = \frac{1}{\lambda_{cr}}. \tag{23}$$

Note any of (20), (21) and (23) can be used to directly derive an exact expression of the reciprocal of the critical buckling factor whenever the partial derivatives are simple (e.g piecewise constant). It should be emphasized however that these equations are not globally valid in the whole domain since the derivatives of λ_{cr} do not exist everywhere. We will see in the sequel that these equations are valid piecewise, since λ_{cr} is differentiable over connected regions that can be easily described.

The homogeneity of the critical buckling factor, which was already noticed, for instance with respect to the loading conditions in Grenestedt (1991), can of course be used to reduce the size of the regression problem. One way to use it would be to build an approximation model over the sphere of radius $\{||N|| = K\}$ and use for instance the spherical coordinates (θ, ϕ) as approximation variables. However

doing so we would definitely reduce the dimension but we are likely to miss some simple dependence of λ_{cr} over the natural variables N . We can also observe that if we consider the critical buckling factor defined over the panel thickness h and the lamination parameters ξ_D instead of being defined over the D tensor, we get

$$\lambda_{cr}(\kappa h) = \kappa^3 \lambda_{cr}(h) \tag{24}$$

This equation means that we can get rid of panel thickness h and only focus on lamination parameters. Finally, we can also observe that for the specific case of rectangular plates, we have by applying the change of coordinates $x' = \kappa x$ and $y' = \kappa y$

$$\lambda_{cr}(\kappa a, \kappa b) = \frac{1}{\kappa^2} \lambda_{cr}(a, b) \tag{25}$$

and the corresponding Euler identity

$$a \frac{\partial \lambda_{cr}}{\partial a} + b \frac{\partial \lambda_{cr}}{\partial b} = 2 \lambda_{cr} \tag{26}$$

Note that (25) in reality can be derived for any domain Ω . Following the derivation³ in Henrot (2006) where this formula is stated for the eigenvalues of the Laplacian, we have for any domain Ω and $\kappa > 0$

$$\lambda_{cr}(\kappa \Omega) = \frac{\lambda_{cr}(\Omega)}{\kappa^2}. \tag{27}$$

3.3 Concavity of the buckling factor over lamination parameters

We now turn on the concavity of λ_{cr} . To show that it is concave over the ξ_D , we first show that it is concave over D , concavity being preserved by pre-composition by an affine function (with h fixed), concavity over ξ_D will be ensured.

In this paragraph, the loading conditions remain fixed, we therefore skip them when writing the dependence of λ_{cr} . We want to show that λ_{cr} is concave over $S_3(\mathbb{R})^{++}$. Let $\alpha > 0$, consider the upper level set $D_\alpha = \{D \in S_3(\mathbb{R})^{++} \text{ s.t. } \lambda_{cr} \geq \alpha\}$. Let $D^{(1)}$ and $D^{(2)}$ in D_α and $\theta \in (0, 1)$, let denote $D^{(\theta)} = \theta D^{(1)} + (1 - \theta) D^{(2)}$. $S_3(\mathbb{R})^{++}$ being convex, $D^{(\theta)}$ is clearly a symmetric positive definite tensor and consider now the bilinear form $a_{D^{(\theta)}}$. It is clearly continuous, symmetric and coercive. Let $w \in \mathcal{V}$ we then have

$$a_{D^{(\theta)}}(w, w) = \theta a_{D^{(1)}}(w, w) + (1 - \theta) a_{D^{(2)}}(w, w) \tag{28}$$

³By noting that $H_\kappa \circ \Delta^2 = \kappa^4 \Delta^2 \circ H_\kappa$ and $H_\kappa \circ \Delta = \kappa^2 \Delta \circ H_\kappa$ where $H_\kappa v$ is the function defined over $\mathcal{V}(\kappa \Omega)$ such that $H_\kappa v(x) = v(x/\kappa)$. Note that (25) is homogeneous of order 2 because buckling equation is essentially the diagonalization of a fourth-order operator with respect to a second-order operator. For pure fourth-order problems (vibrations), we would get homogeneity of order 4.

hence

$$\frac{a_{D^{(\theta)}}(w, w)}{b_N(w, w)} = \theta \frac{a_{D^{(1)}}(w, w)}{b_N(w, w)} + (1 - \theta) \frac{a_{D^{(2)}}(w, w)}{b_N(w, w)} \tag{29}$$

Since $D^{(1)}$ and $D^{(2)}$ are in D_α , for all $w \in \mathcal{V}$

$$\frac{a_{D^{(1)}}(w, w)}{b_N(w, w)} \geq \alpha \tag{30}$$

and

$$\frac{a_{D^{(2)}}(w, w)}{b_N(w, w)} \geq \alpha \tag{31}$$

hence for all $w \in \mathcal{V}$

$$\frac{a_{D^{(\theta)}}(w, w)}{b_N(w, w)} \geq \alpha \tag{32}$$

and $D^{(\theta)}$ is in D_α . This implies that λ_{cr} is quasiconcave, since its upper level sets D_α are convex. Being also homogeneous of degree 1, it is concave (see Appendix A). Note that strict concavity cannot be ensured, this follows obviously from (11) where we see that in the vicinity of $(\xi_D^1, \xi_D^2, 0)$, λ_{cr} is a minimum of linear functions, hence at a point where λ_{cr} is differentiable, there exists a neighborhood where λ_{cr} is linear over (ξ_D^1, ξ_D^2) and cannot be strictly concave. Another quick way to see that is to note that, in case λ_{cr} is twice continuously differentiable over D , Euler identity for the homogeneity of order 0 of each partial derivative of λ_{cr} w.r.t to D_j yields

$$\forall j = 1 \dots 6 \quad \sum_{i=1}^6 D_i \frac{\partial^2 \lambda_{cr}}{\partial D_j \partial D_i} = 0 \tag{33}$$

Equation (33) simply means that D belongs to the kernel of the Hessian $H\lambda_{cr}(D)$ of λ_{cr} at D and hence $H\lambda_{cr}(D)$ is not definite.

Using the following definition of the reciprocal $1/\lambda_{cr}(N)$

$$1/\lambda_{cr}(N) = \max_{w \in \mathcal{V}} \frac{|b_N(w, w)|}{a_D(w, w)} \tag{34}$$

together with the following lower level sets $N_\alpha = \{N \in \mathbb{R}^3 \text{ s.t. } 1/\lambda_{cr}(N) \leq \alpha\}$ and the triangular inequality leads to the quasiconvexity of $1/\lambda_{cr}(N)$ over N . Being homogeneous of degree 1 it is convex. Strict convexity of $1/\lambda_{cr}(N)$ cannot be ensured the same way as the strict concavity of $\lambda_{cr}(\xi_D)$.

3.4 Influence of the stacking sequence

We focus in this section on the behavior of the critical buckling factor λ_{cr} when the loading N is fixed and when the

stacking sequence of the plate laminate varies. We use lamination parameters ξ_D to assess this behavior. We focused here on lamination parameters of conventional laminates $[0^\circ/45^\circ/-45^\circ/90^\circ]$, in that case $\xi_D^4 = 0$. We then restrict our study to the influence of $(\xi_D^1, \xi_D^2, \xi_D^3)$ over the feasible domain described in Bloomfield et al. (2009) and Herencia et al. (2008a, b)

$$2|\xi_D^1| - \xi_D^2 - 1 \leq 0 \tag{35}$$

and

$$2|\xi_D^3| + \xi_D^2 - 1 \leq 0 \tag{36}$$

Note that it is a convex domain. The triangle in $[-1, 1]^2$ that (35) will be referred to as *Miki's triangle* (with $\xi_D^3 = 0$) and the polyhedron in $[-1, 1]^3$ defined both by (35) and (36) will be referred to as *Miki's tetrahedron*. The main interest of the compatibility equations described in Herencia et al. (2008a) is that they provide a convex feasible set for $\xi_D \in [-1, 1]^3$ for given in-plane lamination parameters ξ_A (or equivalently proportions of fiber orientation). In an optimization process, this offers a way to ensure compatibility in a continuous way between in-plane properties (A tensor) and out-of-plane properties (D tensor). In the sequel we study the behavior of λ_{cr} all over its definition domain with no regards to proportions, but we should keep in mind that in practice, the out-of-plane lamination parameters are related to proportions (or equivalently ξ_A). Note that most of the computations done in this article were based on a Rayleigh–Ritz method (see Appendix B) with $N = 20 \times 20$ basis functions over a rectangular plate of dimensions $a = 650$ mm and $b = 250$ mm, we chose a classical Graphite/Epoxy whose characteristics are given Table 1.

3.4.1 Behavior of λ_{cr} for orthotropic laminates under combined loadings and pure shear

For uni-axial and bi-axial compressive loadings, a closed-form equation is given by (11). For fixed loading, it follows that λ_{cr} is piecewise linear over tensor D terms and therefore piecewise affine over ξ_D . For combined loadings (including shear) or pure shear loading, no closed form expression

Table 1 Graphite/epoxy material characteristics

Characteristics	Graphite/epoxy
E_1 longitudinal modulus	112 GPa
E_2 transverse modulus	8.2 GPa
G_{12} shear modulus	4.5 GPa
t elementary ply thickness	0.26 mm

can be derived. The main reason for this is that the $\frac{\partial^2 w}{\partial x \partial y}$ term prevents from using the classical techniques as a double sine series (see Reddy 2004). Correction formulas were given (see for instance Whitney 1969), but all of them were approximation expressions based on simplified assumption (infinite length plate for instance) and there is no evidence at the time being that

- for arbitrary loading (combined loading, pure shear), λ_{cr} is still piecewise affine over ξ_D for orthotropic laminate
- for arbitrary loading, λ_{cr} can be defined in a piecewise manner for non-orthotropic laminates.

In the sequel, we show numerically that λ_{cr} seems to be piecewise affine over ξ_D for combined loading where shear is not very high. However, for pure shear loading, we observed that λ_{cr} can still be defined in a piecewise manner but each part of the function seems to be more complex than affine functions. The most important fact shown here is that λ_{cr} behaves in a piecewise manner and a good approximation technique should be able to detect these different regions or so-called modes and build an approximator for each region.

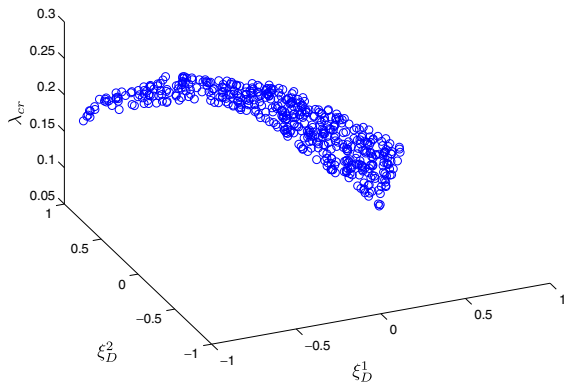
To assess the evolution of λ_{cr} , we sample uniformly Miki's triangle for ξ_D^1, ξ_D^2 and compute for each couple of lamination parameters its associated λ_{cr} . The loading is taken as combined loading with shear and we also compute the associated sensitivities $\frac{\partial \lambda_{cr}}{\partial \xi_D^1}$ and $\frac{\partial \lambda_{cr}}{\partial \xi_D^2}$ and plot the phase portrait: for each couple we associate a point of the gradient space.

The different plots depicted Fig. 2 were obtained for two different loading conditions

- Combined loading: $N_x = -50$ N.mm⁻¹, $N_y = -20$ N.mm⁻¹, $N_{xy} = 20$ N.mm⁻¹
- Pure shear loading: $N_x = 0$ N.mm⁻¹, $N_y = 0$ N.mm⁻¹, $N_{xy} = 20$ N.mm⁻¹.

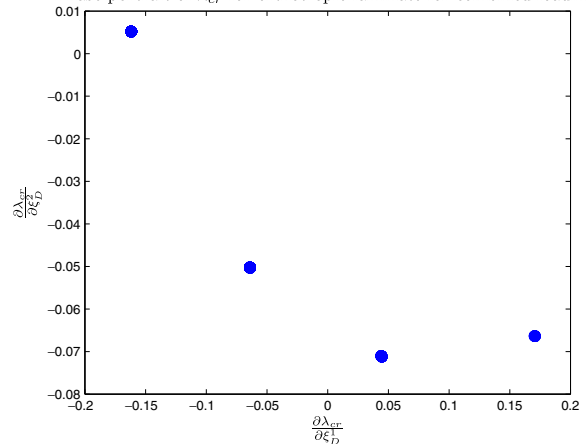
For each loading case, we generated 800 couples (ξ_D^1, ξ_D^2) uniformly distributed over Miki's triangle. We depicted the behavior of λ_{cr} Fig. 2a and c. We can observe that for the first loading case, we clearly see the piecewise dependence of λ_{cr} over (ξ_D^1, ξ_D^2) , as in the bi-axial compression case of (11), λ_{cr} seems still defined as the minimum of hyperplanes, making the overall function simply defined over a sole hyperplane but not differentiable at the boundaries between hyperplanes. It is worth noting that directional differentiability seems still valid as it is proved in Rousselet and Chenais (1990). For this loading case, λ_{cr} seems to be still defined as piecewise linear, this is numerically ensured by the phase portrait Fig. 2b, where we see all the 800 points shrink down to only 4 different points, which means over

λ_{cr} over Miki's triangle for orthotropic laminates for combined loading



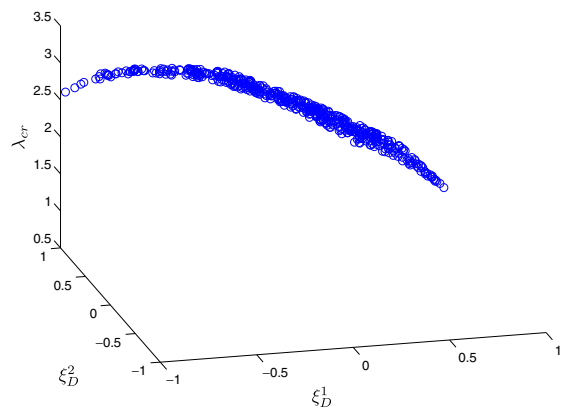
(a) Piecewise dependence of λ_{cr} over ξ_D^1 and ξ_D^2 on Miki's triangle for combined loading $N_x = -50N.mm^{-1}$, $N_y = -20N.mm^{-1}$ and $N_{xy} = 20N.mm^{-1}$

Phase portrait of λ_{cr} for orthotropic laminate for combined loading



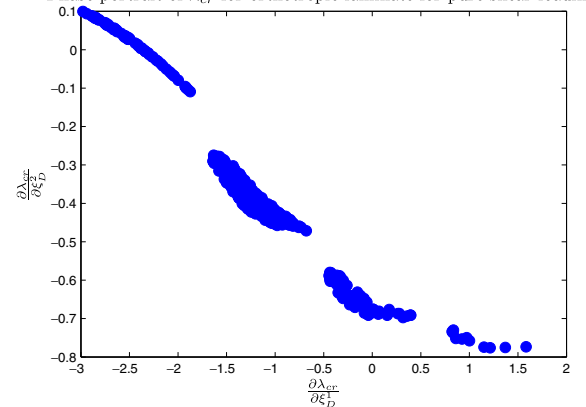
(b) Phase portrait associated to the couple ξ_D^1, ξ_D^2 for combined loading. Each point is the associated partial derivatives

λ_{cr} over Miki's triangle for orthotropic laminates for pure shear



(c) Piecewise dependence of λ_{cr} over ξ_D^1 and ξ_D^2 on Miki's triangle for pure shear loading $N_x = 0N.mm^{-1}$, $N_y = 0N.mm^{-1}$ and $N_{xy} = 20N.mm^{-1}$

Phase portrait of λ_{cr} for orthotropic laminate for pure shear loading



(d) Phase portrait associated to the couple ξ_D^1, ξ_D^2 for pure shear loading. Each point is the associated partial derivatives

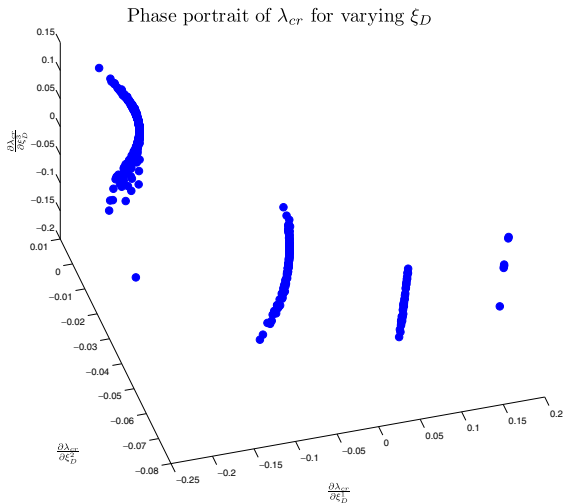
Fig. 2 Results for orthotropic laminates under combined loading and pure shear loading

each region, λ_{cr} is linear or very close to be linear. For the pure shear loading case, this dependence does not seem to hold, indeed, the regions are not as clearly defined as in the combined loading. The phase portrait shows 4 different regions disconnected, which again means that λ_{cr} can be defined in a piecewise manner, though its behavior over each region seems more complicated than linear.

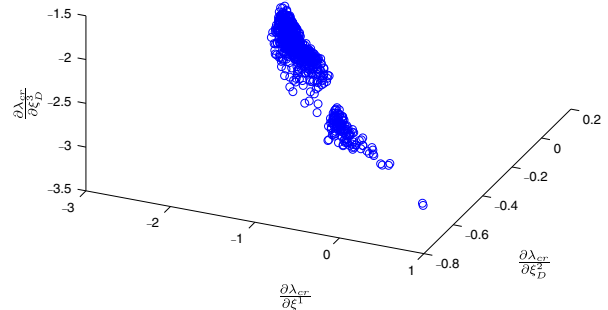
3.4.2 Behavior of λ_{cr} for non-orthotropic laminates under combined loadings and pure shear

We also depicted the phase portrait for non-orthotropic laminates by making the ξ_D varying in the Miki's tetrahedron for the same loadings. We observe in both cases that the phase portrait has the same structure that is to say 4

disconnected components. However, the dispersion of these clusters is quite different for the two loadings (cf. Fig. 3). In the first case, for combined loading, we observe a low variance over the two first dimensions ($\frac{\partial \lambda_{cr}}{\partial \xi_D^1}, \frac{\partial \lambda_{cr}}{\partial \xi_D^2}$) and bigger variations for the last dimension ($\frac{\partial \lambda_{cr}}{\partial \xi_D^3}$), this would naturally indicate that λ_{cr} is close to depend linearly over the two first dimensions and that it depends in a more complex manner over the last dimension. In case of pure shear loading, the dependence is much more complex, it does not seem that there is a dimension that has a bigger influence than the others. However, the main observation that we can make is that in all cases, λ_{cr} seems to depend in a piecewise manner and that the function is likely to be well approximated when taking advantage of this structure.



Phase portrait of λ_{cr} for non-orthotropic laminates for pure shear loading



(a) Phase portrait associated to the couple ξ_D^1, ξ_D^2 for combined loading $N_x = -50 N.mm^{-1}$, $N_y = -20 N.mm^{-1}$ and $N_{xy} = 20 N.mm^{-1}$. Each point is the associated partial derivatives

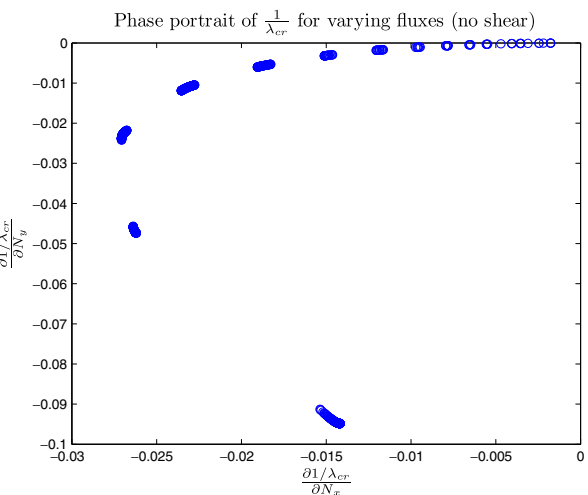
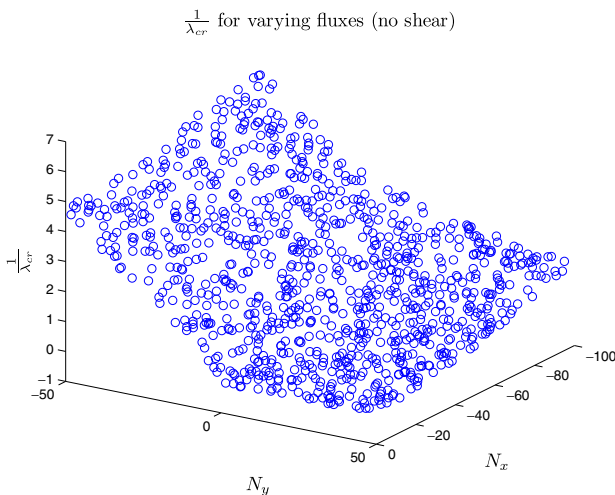
(b) Phase portrait associated to the couple ξ_D^1, ξ_D^2 for pure shear loading $N_x = 0 N.mm^{-1}$, $N_y = 0 N.mm^{-1}$ and $N_{xy} = 20 N.mm^{-1}$. Each point is the associated partial derivatives

Fig. 3 Results for non-orthotropic laminates under combined loading and pure shear loading

3.5 Influence of the loading conditions

In this section, we are interested in the behavior of the λ_{cr} for fixed stacking sequence ($[-45/-45/0/45/90/-45/0]_S$) when the loading conditions vary. In most cases, when loading conditions vary, the reciprocal of λ_{cr} was found to be more easier to depict and interpret. This can

be easily seen again in the general formula (15) where all the stacking sequence dependence appear in the numerator and all the loading conditions appear in the denominator. Furthermore, it also follows from both theoretical and computational considerations. Indeed, for instance in Rousselet and Chenais (1990), the authors do not study directly λ_{cr} but first use its reciprocal to show it is continuous and



(a) $\frac{1}{\lambda_{cr}}$ over (N_x, N_y) for non-orthotropic laminate

(b) Phase portrait associated to the couple (N_x, N_y) for non-orthotropic laminate. Each point is the associated partial derivatives

Fig. 4 Results for reciprocal of the buckling critical load for varying loading conditions $N_{xy} = 0 N.mm^{-1}$ for a fixed non-orthotropic stacking sequence

differentiable whenever it is a simple eigenvalue. Numerically, it is preferable to compute the reciprocal of λ_{cr} that is to find w and the largest κ such that

$$Lw = \kappa K w \tag{37}$$

where K is the matrix associated to the a_D bilinear form (stiffness material matrix) and L matrix associated to the b_N bilinear form (geometry stiffness matrix) and eventually $\lambda_{cr} = 1/\kappa$. Next, as noted before, in case of the orthotropic formula, we can observe that $1/\lambda_{cr}$ is piecewise linear over $N = (N_x, N_y)$. Regarding computational issues when making the loading conditions vary, it can be noted that for general loading (including shear) the feasible set (i.e the fluxes that gives rise to buckling, this is related to the assumption of positivity of b_N) are not easy to describe. When the loading conditions approach near the boundary, the eigensolver might not converge or exhibit numerical ill-conditioning behavior and λ_{cr} increases very quickly and behaves like an hyperbola (cf. Fig. 4).

4 Original strategy to approximate the critical buckling factor

In a recent paper (Bettebghor et al. 2011), the authors developed a strategy to approximate discontinuous or derivative discontinuous functions by means of clustering of the learning basis into regions where the function to be approximated is expected to be continuous or at least more simple. It strongly relies on the EM algorithm (Expectation–Maximization), which is a classical algorithm in statistics to solve maximum-likelihood problem, such as density estimation. In this case, the learning data (input and output) is modeled as a finite mixture of multivariate Gaussians. The EM algorithm is first used to estimate the mixtures parameters (weights of each mixture component, means and variance-covariance matrices of each Gaussian component), a clustering is performed on the data (by means of Maximum A Posteriori), a local expert is built over each region, marginal densities are then derived from the finite mixture parameters and a final approximation model is built as

$$f(x) = \sum_{i=1}^K \mathbb{P}(\kappa = i | X = x) f_i(x) \tag{38}$$

where K is the number of Gaussian components, $\mathbb{P}(\kappa = i | X = x)$ is the probability to lie in cluster i knowing that $X = x$ and f_i is the local expert built on cluster i . Equation (38) leads to two different approximation models depending on the computation of $\mathbb{P}(\kappa = i | X = x)$. When choosing the Gaussian laws to compute this quan-

tity, (38) leads to a smooth model that smoothly recombine different local experts. If $\mathcal{P}(i | X = x)$ is computed as characteristic functions of clusters (being equal to 0 or 1) this leads to a discontinuous approximation model. Depending on the application, the smooth recombination can be used for instance in a gradient-based method, while if derivative free optimization algorithms are to be used, the discontinuous (or hard) recombination is preferable whenever it gives a better accuracy. Based on former numerical experiments where this strategy showed to be adequate to approximate discontinuous or derivative discontinuous functions and on the behavior of λ_{cr} as described in the former sections, we want to apply this strategy to the approximation of λ_{cr} with a specified type of local experts f_i (linear, polynomial and rational).

4.1 Piecewise polynomial approximation in lamination space

The behavior of $\lambda_{cr}(\xi_D^1, \xi_D^2)$ for orthotropic laminates for combined loading (with low shear) seems to be piecewise linear (or close to be linear). We then want to approximate λ_{cr} based on a mixture of linear experts (referred to as MOL). The number of clusters is chosen through a range of potential clusters by minimizing the generalization error for the mixture of linear experts. We depicted Fig. 5 the result of the EM clustering for 4 clusters on the orthotropic data for combined loading. We see that the hyperplanes are correctly detected through the clustering.

For pure shear and non-orthotropic laminates, the phase portraits showed that the linear dependence does not hold, however λ_{cr} does not seem to behave in a very complicated way over each cluster. We then build an approximation of λ_{cr} on the basis of a mixture of quadratic models (MOQ).

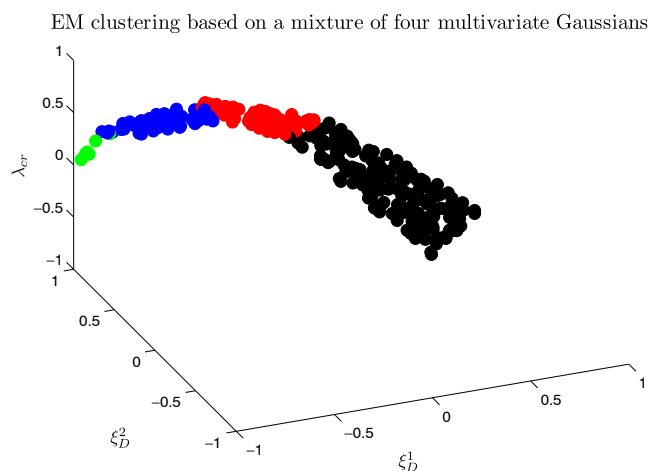


Fig. 5 Automatic clustering (normalized output) for an orthotropic laminate and combined loading

4.2 Piecewise polynomial approximation in forces space

When the loading conditions vary, the dependence of $1/\lambda_{cr}$ seems to get less simple as linear as the plate endures less compression and more complex loading with tension for instance. However for pure compression loading cases, $1/\lambda_{cr}$ is close to be piecewise linear, we then build an approximation model as a mixture of quadratic models since it is expected to be able to handle different situations. It should be noted that if ordinary least squares are used, a linear model can not totally be recovered from a polynomial regression due to the \mathcal{L}_2 minimization that regularizes polynomial coefficients. So more advanced techniques should be used to build purely linear experts and quadratic experts in different regions (best-subset selection, lasso, least angles regression, compressed sensing).

4.3 Piecewise rational approximation in the general case

Based on the behavior of λ_{cr} with respect to lamination parameters and of $1/\lambda_{cr}$ with respect to loads per unit length N , we want to build an approximation model as a mixture of (quadratic) rational experts

$$f_i(x) = \frac{a_0 + a_1 \xi_D^1 + \dots + a_{n-1} (\xi_D^2)^2 + a_n (\xi_D^3)^2}{b_0 + b_1 N_x + \dots + b_{n-1} N_y^2 + b_n N_{xy}^2} \quad (39)$$

such models will be referred to as Mixture of Rationals (MoR).

4.4 Using directly Rayleigh–Ritz active buckling modes

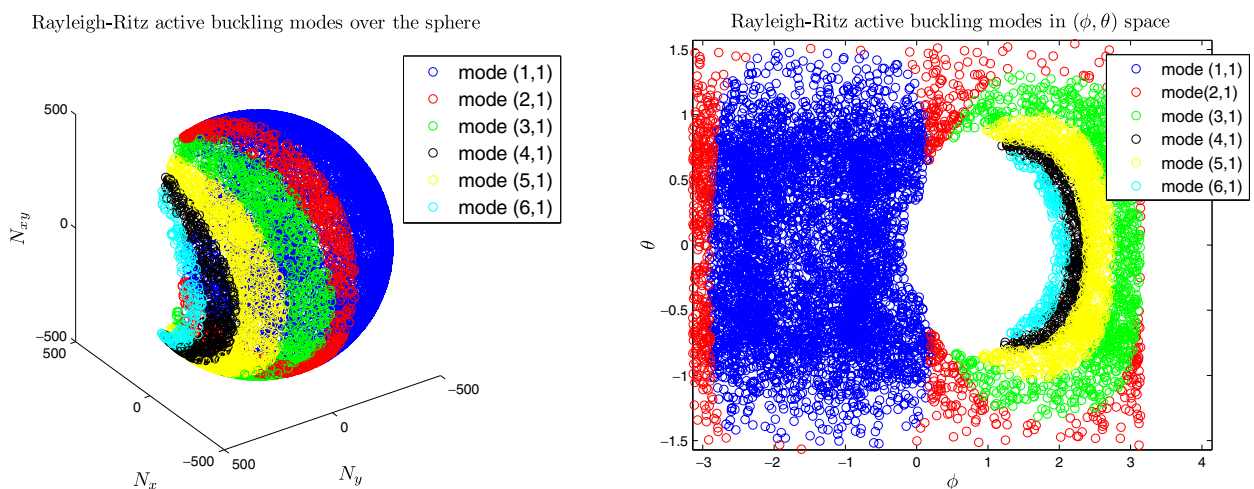
It is worth noting when using Rayleigh–Ritz method, we usually get an extra information by looking at the eigenvector w_{cr} associated to λ_{cr} . Indeed, it can be written

$$w_{cr} = \sum_{i=1}^N \alpha_i u_i \quad (40)$$

where the u_i 's are the basis functions. These basis functions can be associated to 'natural' modes (that is eigenfunctions of the unperturbed operator or over a more simple geometry). Suppose we take functions of the type

$$\sin(i\pi x/a) \sin(j\pi y/b). \quad (41)$$

We can associate each eigenvector w_{cr} to the 'closest natural mode' simply by finding the largest (in magnitude) component w_{cr}^{ij} of w_{cr} . However, this process can lead to wrong associated modes, in the case where we have two or more large components in w_{cr} , the numerical approximation is likely to invert the corresponding natural modes. Next, finite element buckling methods do not provide directly such information, since the critical buckled shape (w_{cr}) is defined over local, mesh-dependent, non natural basis functions (\mathcal{P}_1 basis functions...). Post-treatment of the transverse displacement is needed when identifying the corresponding natural mode. This includes tools from mode shape identification well known in vibrations or structural health monitoring like *Modal Assurance Criteria* (MAC). In case modal information is directly provided and can be easily retrieved from the analysis, this offers a natural way to approximate the



(a) Associated natural buckling modes in the flux spaces (b) Associated natural buckling modes in (θ, ϕ) space

Fig. 6 Active buckling modes found by Rayleigh–Ritz method for different representations

critical buckling load by approximating mode per mode and building classifiers (linear or more advanced like Support Vector Machine) that give the corresponding natural mode to a new input and hence apply the corresponding approximation model. This can be done for few degrees-of-freedom problems since the regions that separate each natural buckling mode are usually quite simple in the input space. We depicted for instance Fig. 6, the different natural modes when the loading conditions vary both over the loading conditions sphere ($\|N\| = K$) and over spherical angles space (θ, ϕ) , since the active mode does not depend on the norm $\|N\|$. We clearly see Fig. 6a that the associated regions are quite simple over the sphere but for the (θ, ϕ) space Fig. 6b the regions are not easily separated, which prevents from building accurate classifiers. Even though, this approach seems at first glance easy to apply, there are several drawbacks that prevent from detecting the correct mode. This is the difficult problem of mode tracking well known in structural dynamics and vibrations. For low-dimensional problem (<3), several phenomena make it difficult to follow one mode when state parameters change: mode crossing, where two modes simply cross each other, mode veering where two nearly-simultaneous modes start to interact and the associated eigenvalues abruptly diverge and finally mode coalescence where two modes shrink down in only one mode. For high-dimensional problems (more than 3 parameters), such our problem, this mode tracking becomes very challenging and even impossible. Furthermore, one has to fix the number of modes to approximate and ensure that all these modes are available at learning points. Last but not least, the accuracy of the eigenvalues decreases for higher modes, this means that each mode is likely to have a different level of noise all over the input space.

5 Numerical results

We describe the different results obtained with the suggested strategies to approximate λ_{cr} in the different situations.

- For fixed loading: we approximate λ_{cr} first for orthotropic laminates (over ξ_D^1 and ξ_D^2) and then for [0/45/90] laminates (over ξ_D^1 , ξ_D^2 and ξ_D^3) for both combined loading and pure shear loading.
- For a fixed non-orthotropic stacking sequence: we approximate λ_{cr} first for bi-axial loading (over N_x and N_y) and then for combined loading (over N_x , N_y and N_{xy}).
- In the general case when stacking and loading vary: we approximate λ_{cr} over ξ_D^1 , ξ_D^2 , ξ_D^3 , N_x , N_y and N_{xy} .

In all situations, the learning of the approximation models is based on three different bases: learning basis B_A , test basis B_T and validation basis B_V . The learning in itself is directly based on B_A , to compare several approximation model built over B_A , we use B_T to rank them and chose the best one. Finally at the end of this selection process, the generalization error and the other accuracy metrics are computed over the validation basis B_V . Following for instance Friedman et al. (2001), we have a classical partition of 60% for B_A , 20% for B_T and 20% for B_V . To assess the goodness of the fit, we use several classical metrics, amongst them, we use the *Mean Squared Error* (MSE) and *Lack Of Fit* (LOF) defined as

$$MSE = \frac{1}{N_V} \sum_{i=1}^{N_V} |\hat{f}(x_i) - y_i|^2 \tag{42}$$

and

$$LOF = \frac{MSE}{\text{Var}(\{y_i\}_{i=1\dots N_V})} \tag{43}$$

where N_V is the number of points in the validation basis, y is the real output function and \hat{f} is the approximation model. The LOF simply divides the MSE by the variance of the output to avoid scaling effect. However the knowledge of only the MSE might not be enough for accuracy evaluation. Indeed, first it does not provide any clue regarding the relative error and more important it is an average over a whole basis and does not give any insight on the distribution of the error. Suppose we have two different approximation models, one giving a very good approximation for a majority of points and being very bad for a small amount of points, it will be equivalent in terms of MSE to an approximation model that has a medium error for all points, while the first approximation model is likely to be preferable when

Table 2 Results for the approximation of λ_{cr} for orthotropic laminates for combined loading and pure shear loading

Loading type	Combined loading		Pure shear loading	
	MOL	ANN	MOQ	ANN
MSE	$8.5 \cdot 10^{-8}$	$7.8 \cdot 10^{-7}$	$2.1 \cdot 10^{-6}$	$1.7 \cdot 10^{-6}$
LOF(%)	$9.1 \cdot 10^{-4}$	$8.5 \cdot 10^{-3}$	$8.5 \cdot 10^{-4}$	$6.7 \cdot 10^{-4}$
$\hat{E}_{rel}(\%)$	7.10^{-3}	$1.6 \cdot 10^{-1}$	$8.8 \cdot 10^{-3}$	$3.4 \cdot 10^{-3}$
\hat{E}_{abs}	$3.1 \cdot 10^{-5}$	$6.3 \cdot 10^{-4}$	$2.4 \cdot 10^{-4}$	$8.1 \cdot 10^{-4}$
$E_{rel}^{max}(\%)$	1.16	1.6	0.9	0.5
E_{abs}^{max}	$4.6 \cdot 10^{-3}$	$5.4 \cdot 10^{-3}$	$2.5 \cdot 10^{-2}$	$7.5 \cdot 10^{-3}$
$Q_{1\%}(\%)$	99.7	99.7	100	99.7
$Q_{0.1\%}(\%)$	98.8	43.6	97.9	92.9

used within a population-based algorithm (such as genetic algorithms). This is why, we add several metrics

- \mathcal{L}_1 and \mathcal{L}_∞ errors:
- Mean relative error:

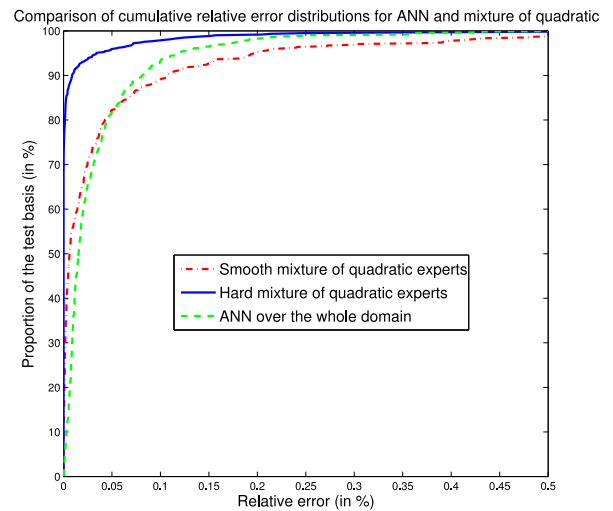
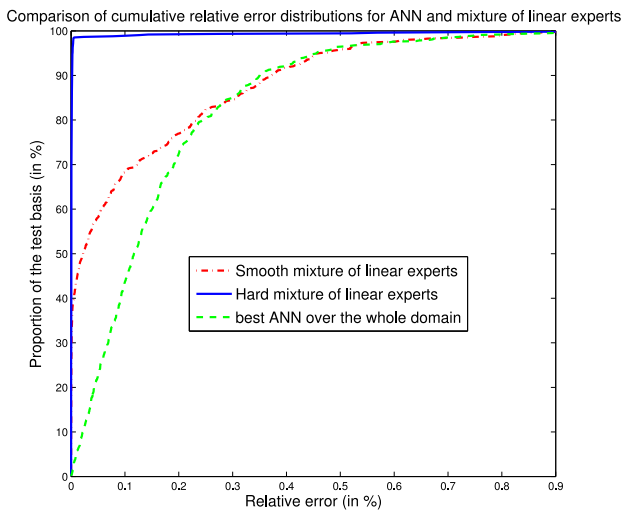
$$\hat{E}_{rel} = \frac{1}{N_V} \sum_{i=1}^{N_V} \frac{|\hat{f}(x_i) - y_i|}{|y_i|}$$

- Mean absolute error:

$$\hat{E}_{abs} = \frac{1}{N_V} \sum_{i=1}^{N_V} |\hat{f}(x_i) - y_i|$$

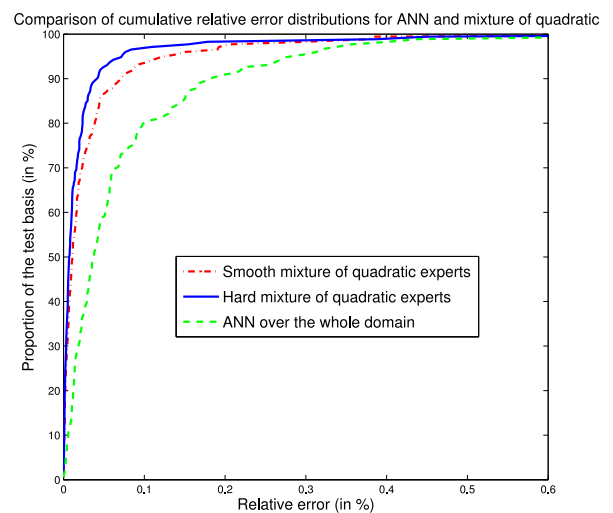
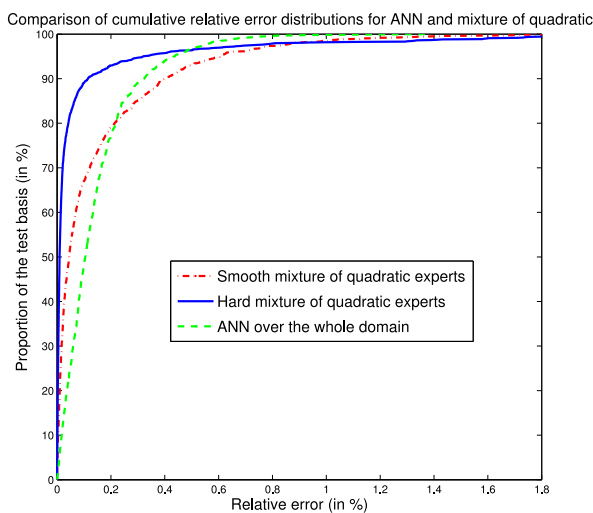
- Max relative error:

$$E_{rel}^{max} = \max_{i=1 \dots N_V} \frac{|\hat{f}(x_i) - y_i|}{|y_i|}$$



(a) Comparison of approximations of λ_{cr} over ξ_D^1 and ξ_D^2 on Miki's triangle for combined loading $N_x = -50N.mm^{-1}$, $N_y = 20N.mm^{-1}$ and $N_{xy} = 20N.mm^{-1}$

(b) Comparison of approximations of λ_{cr} over ξ_D^1 and ξ_D^2 on Miki's triangle for pure shear loading $N_x = 0N.mm^{-1}$, $N_y = 0N.mm^{-1}$ and $N_{xy} = 20N.mm^{-1}$



(c) Comparison of approximations of λ_{cr} over ξ_D^1 , ξ_D^2 and ξ_D^3 for combined loading $N_x = -50N.mm^{-1}$, $N_y = 20N.mm^{-1}$ and $N_{xy} = 20N.mm^{-1}$

(d) Comparison of approximations of λ_{cr} over ξ_D^1 , ξ_D^2 and ξ_D^3 for pure shear loading $N_x = 0N.mm^{-1}$, $N_y = 0N.mm^{-1}$ and $N_{xy} = 20N.mm^{-1}$

Fig. 7 Quantile curves for orthotropic and non-orthotropic laminates under combined loading and pure shear loading

- Max absolute error:

$$E_{\text{abs}}^{\text{max}} = \max_{i=1 \dots N_V} |\hat{f}(x_i) - y_i|$$

- Cumulative error distribution :

- α - quantiles:

$$Q_\alpha = \frac{\text{Card}\{x_i \in B_V \text{ s.t. } \frac{|\hat{f}(x_i) - y_i|}{\hat{f}(x_i)} \leq \alpha\% \}}{N_V}$$

the proportion of the validation points that are below a relative error of $\alpha\%$.

- Quantile curve: curve obtained by making α varying from 0% to $E_{\text{rel}}^{\text{max}}\%$.

Note also that the number of points used in each case corresponds to the total points used for all bases. To the end of comparison, in each case we compared with a reasonably good artificial neural network built over the whole domain with classical techniques such as cross-validation for the choice of the number of hidden units or backpropagation error for the weight update.

5.1 Approximation models for varying stacking

For the approximation of λ_{cr} for orthotropic laminates, we used $N = 420$ points for the first loading case (combined, $N_x = -50 \text{ N.mm}^{-1}$, $N_y = 20 \text{ N.mm}^{-1}$ and $N_{xy} = 20 \text{ N.mm}^{-1}$) and $N = 430$ points for the second loading case (pure shear). For the combined loading case, a mixture of linear experts was built and for the pure shear case, a mixture of quadratic models. We see Table 2 that the mixture model performs equally as the artificial neural network (ANN) for the pure shear and performs better for the combined loading. The quantile curves as described on Fig. 7a and b and the quantiles $Q_{1\%}$ and $Q_{0.1\%}$ (cf. Table 2) show that the hard recombination of mixture models even when they have the same accuracy in terms of MSE is much more accurate for a large amount of validation points. For combined loading for instance, the quantile curve is very close the y-axis, which means the approximation model is close to be exact except for a few points for which the mixture model gives bad accuracy. This might be due to misclassification, the wrong local expert is called for these few points, while the ANN regularizes the error over the whole domain. The ANN is of a reasonably good accuracy over all validation points but can not achieve very high accuracy even for a small amount of validation points. Regarding the relative error measures, the mixture models always perform better than ANN.

Table 3 Results for the approximation of λ_{cr} for non-orthotropic laminates for combined loading and pure shear loading

Loading type	Combined loading		Pure shear loading	
	MOQ	ANN	MOQ	ANN
MSE	$6.9 \cdot 10^{-4}$	$2.7 \cdot 10^{-4}$	$1.7 \cdot 10^{-4}$	$9.3 \cdot 10^{-4}$
LOF(%)	$3 \cdot 10^{-2}$	$1.2 \cdot 10^{-2}$	$3.4 \cdot 10^{-4}$	$1.9 \cdot 10^{-3}$
$\hat{E}_{\text{rel}}(\%)$	$7.7 \cdot 10^{-2}$	$1.4 \cdot 10^{-1}$	$3 \cdot 10^{-2}$	$7.6 \cdot 10^{-2}$
\hat{E}_{abs}	$6.4 \cdot 10^{-3}$	$1.2 \cdot 10^{-2}$	$7 \cdot 10^{-3}$	$1.8 \cdot 10^{-2}$
$E_{\text{rel}}^{\text{max}}(\%)$	5.1	1.3	0.47	0.9
$E_{\text{abs}}^{\text{max}}$	$3.8 \cdot 10^{-1}$	$8.8 \cdot 10^{-2}$	$8.9 \cdot 10^{-2}$	0.15
$Q_{1\%}(\%)$	98.2	99.8	100	100
$Q_{0.1\%}(\%)$	89	48.1	93.1	79.9

In the case of non-orthotropic laminates, $N = 853$ points were used for combined loading and $N = 768$ points in the case of pure shear loading. Results are in Table 3, again we observe the same behavior (see Fig. 7c and d), the mixture models are by far better in terms of relative error while the MSE show the same tendency. The quantile curves clearly show that the hard recombination of mixture models is better for a majority of points. Regarding the smooth recombination, no clear conclusion can be drawn since it performs better than ANN for pure shear loading and it is less accurate for combined loading.

5.2 Approximation models for varying loading conditions

Based on the observations previously made, λ_{cr} is approximated through its reciprocal, the result of the approximation is then inverted and the accuracy is measured on λ_{cr} . For the first case (no shear), $N = 468$ points were used with $N_x \in [-100, 0]$ and $N_y = [-50, 50]$ and for the second

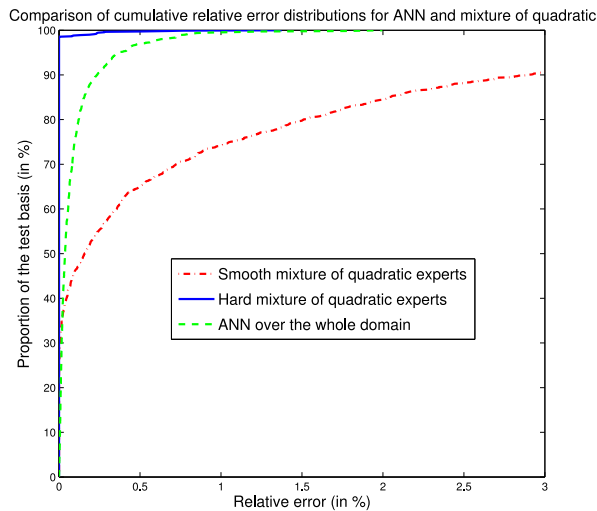
Table 4 Results for the approximation of λ_{cr} for varying loadings (bi-axial loading and combined loading) and fixed SS

Loading type	Bi-axial loading		Combined loading	
	MOQ	ANN	MOQ	ANN
MSE	$1.93 \cdot 10^{-6}$	$4.8 \cdot 10^{-5}$	$3.6 \cdot 10^{-3}$	$6.6 \cdot 10^{-3}$
LOF(%)	$2.1 \cdot 10^{-4}$	$5 \cdot 10^{-3}$	$2.6 \cdot 10^{-1}$	$4.7 \cdot 10^{-1}$
$\hat{E}_{\text{rel}}(\%)$	$5.1 \cdot 10^{-3}$	$9.3 \cdot 10^{-2}$	$4 \cdot 10^{-1}$	$7.1 \cdot 10^{-1}$
\hat{E}_{abs}	$1 \cdot 10^{-4}$	$2 \cdot 10^{-3}$	$1.1 \cdot 10^{-2}$	$1.7 \cdot 10^{-2}$
$E_{\text{rel}}^{\text{max}}(\%)$	1.4	2	11.1	29.4
$E_{\text{abs}}^{\text{max}}$	$3.9 \cdot 10^{-2}$	$9.9 \cdot 10^{-2}$	1.1	1.1
$Q_{1\%}(\%)$	99.9	99.4	90.4	81.8
$Q_{0.1\%}(\%)$	98.8	75.6	65.9	28.9

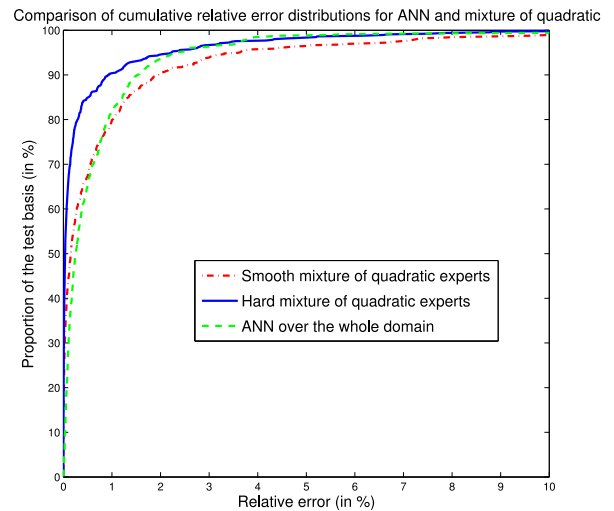
case 940 points were used with $N_x \in [-100, 0]$, $N_y = [-50, 50]$ and $N_{xy} = [0, 50]$. In both cases, the stacking sequence that was fixed was a non-orthotropic one $[-45/-45/0/45/90/-45/0]_S$ (2×7 plies). Results are in Table 4. In that case, mixture of quadratic models performs better for all accuracy metrics. Quantiles measures and quantile curves (see Fig. 8a and b) show the same tendency as in the former test cases. Mixture models are much more accurate for a large number of points of the validation basis.

5.3 Approximation models in the general case

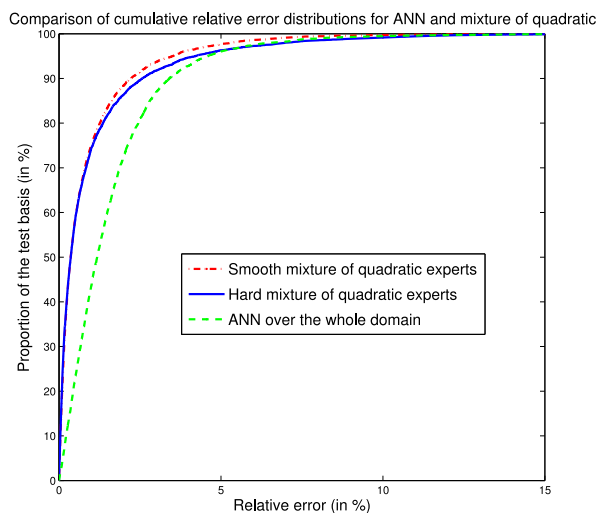
In that case, we make both stacking sequence and loading vary. We distinguish two cases. In the first one, we consider non-orthotropic laminates and no shear loading (biaxial loading including tension/compression). Dimension of regression in that case is 5: $\xi_D^1, \xi_D^2, \xi_D^3, N_x$ and N_y . We used 5,400 points ranging in Miki's tetrahedron for lamination parameters and $N_x \in [-100, 0]$ and $N_y \in [-50, 50]$. In the second, we added shear loading, dimension of



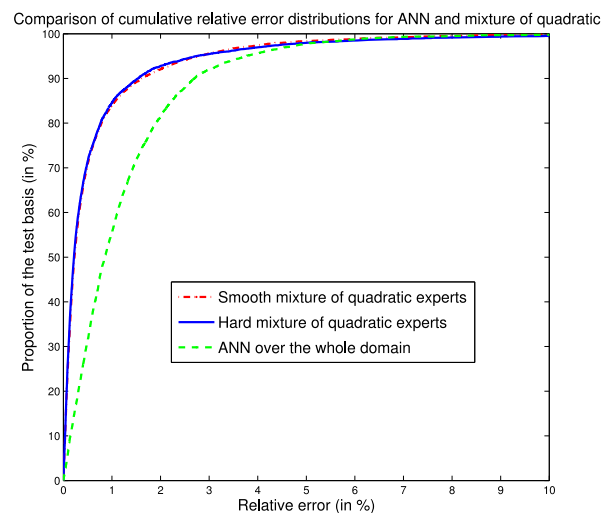
(a) Comparison of approximations of λ_{cr} over N_x and N_y for a non-orthotropic laminate



(b) Comparison of approximations of λ_{cr} over N_x, N_y and N_{xy} for a non-orthotropic laminate



(c) Comparison of approximations of λ_{cr} over $\xi_D^1, \xi_D^2, \xi_D^3, N_x$ and N_y



(d) Comparison of approximations of λ_{cr} over $\xi_D^1, \xi_D^2, \xi_D^3, N_x, N_y$ and N_{xy}

Fig. 8 Quantile curves for the approximation of λ_{cr} **a** over N_x and N_y **b** over N_x, N_y and N_{xy} **c** over $\xi_D^1, \xi_D^2, \xi_D^3, N_x$ and N_y **d** $\xi_D^1, \xi_D^2, \xi_D^3, N_x, N_y$ and N_{xy}

Table 5 Results for the approximation of λ_{cr} for varying non-orthotropic laminates and for varying flux

Loading type	Bi-axial loading		Combined loading	
	MOR	ANN	MOR	ANN
MSE	$7.6 \cdot 10^{-3}$	$3.4 \cdot 10^{-3}$	$1.9 \cdot 10^{-2}$	$7.1 \cdot 10^{-3}$
LOF(%)	$3.8 \cdot 10^{-1}$	$1.7 \cdot 10^{-1}$	$3.8 \cdot 10^{-1}$	$1.3 \cdot 10^{-1}$
$\hat{E}_{rel}(\%)$	0.9	1.6	0.67	1.25
\hat{E}_{abs}	$3.2 \cdot 10^{-2}$	$3.6 \cdot 10^{-2}$	$3.3 \cdot 10^{-2}$	$3.7 \cdot 10^{-2}$
$E_{rel}^{max}(\%)$	20.9	40.5	31	22.4
E_{abs}^{max}	1.23	0.7	3.64	1.56
$Q_{1\%}(\%)$	73.7	43.6	84.6	55.6
$Q_{0.1\%}(\%)$	19.5	4.1	28.6	6.6

regression was then 6 and we used 8,400 points within the same domain with $N_{xy} \in [0, 50]$. Results are in Table 5 and in Fig. 8c and d. For both regressions, results show that ANN's give lower MSE, but for all the other accuracy metrics, the mixture of rational experts gives a better accuracy. Regarding quantile errors, we have the same tendency as in the other cases, that is mixture models gives a better accuracy for most points than ANN's.

5.4 Conclusion on the accuracy of the approximation

We see that in many cases, mixture models offers quite a good performance in terms of the accuracy of the approximation. For most cases, the hard recombination was found to be better than the smooth recombination and a reasonably good artificial neural network. In case one type of variables, amongst stacking or loading, is fixed, the accuracy is very high allowing to use the approximation models within an optimization process (see Section 6). When material and loading vary the quality of the approximation is degraded, this is of course due to the dimension but also of the growing complexity of λ_{cr} ; much modes are indeed covered when making material and loading vary. What suggests these results is that approximation models should take advantage of this mode-based structure to improve the accuracy of the approximation. However, doing so requires a fairly good number of learning points within each mode, which is not an easy task to obtain without prior knowledge.

6 Applications to optimization

6.1 Optimization problem

In order to illustrate the interest of the approximation of λ_{cr} , we consider a stacking sequence optimization to maximize

critical buckling factor of a composite plate under combined loading. The optimization problem can be stated as follows

$$\begin{cases} \max \lambda_{cr}(\xi_D^1, \xi_D^2, \xi_D^3) \\ \text{s.t. } 2|\xi_D^1| - \xi_D^2 - 1 \leq 0 \\ \quad 2|\xi_D^3| + \xi_D^2 - 1 \leq 0 \end{cases}$$

where some parameters are fixed: the material, the dimensions of the plate a and b , the total number of plies n , the loading with N_x , N_y and N_{xy} .

6.2 Description of the two distinct approaches

Two approaches will be compared to solve this problem:

- An explicit approach: it consists in listing all the feasible stacking sequences for a fixed number of plies n and computing the buckling load with a Rayleigh–Ritz method. The optimization problem is written as follows

$$\max_{\text{feasible SS}} \lambda_{cr}(\text{SS})$$

where the acceptable stacking sequence space is generated by listing the whole of the possible combinations of angles $[0^\circ, \pm 45^\circ, 90^\circ]$ which comply with the manufacturing rules (see Section 2.2). The critical buckling factor λ_{cr} is computed with a Rayleigh–Ritz (RR) method on the subset of feasible stacking sequences. The optimal stacking sequence SS^* is referred to as the reference solution for the optimization problem.

- A two-step approach composed of a continuous optimization in the out-of-plane lamination parameters space followed by a post-identification to find the associated optimal stacking sequence.

1. Solve the continuous optimization problem

$$\begin{cases} \max \lambda_{cr}(\xi_D^1, \xi_D^2, \xi_D^3) \\ \text{s.t. } 2|\xi_D^1| - \xi_D^2 - 1 \leq 0 \\ \quad 2|\xi_D^3| + \xi_D^2 - 1 \leq 0 \end{cases}$$

and determine the optimal out-of-plane lamination parameters $(\xi_D)^* = (\xi_D^1, \xi_D^2, \xi_D^3)^*$. The concavity

Table 6 Optimization result with the combined loading $N_x = -50 \text{ N.mm}^{-1}$, $N_y = 20 \text{ N.mm}^{-1}$ and $N_{xy} = 20 \text{ N.mm}^{-1}$

	Explicit	2-step with RR	2-step with MOQ
Optimal SS		$[-45/-45/-45/0/45/45/0/0/45/90]_s$	
Error at optimum (%)	Exact	Exact	0.03%
Number of λ_{cr} evaluations	7,032	269 + 20	39 + 20

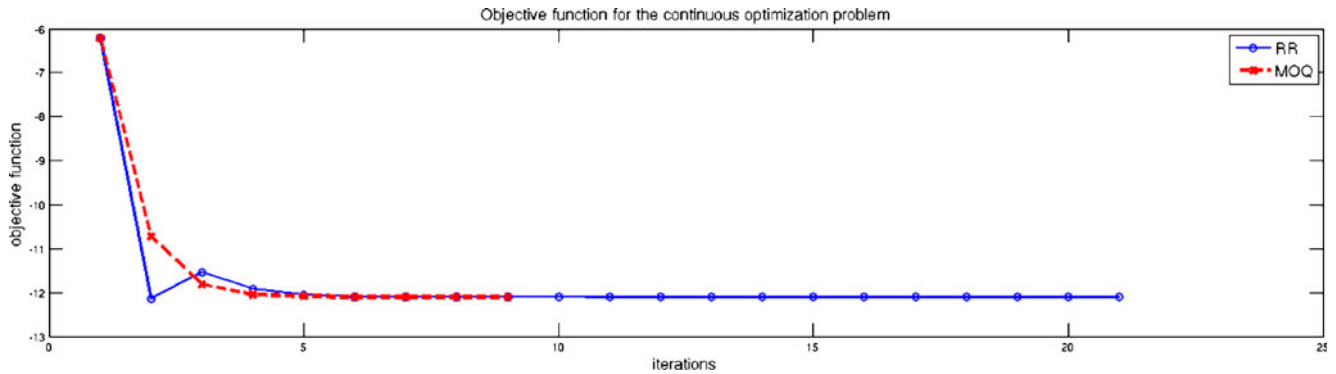


Fig. 9 Objective function during the optimization process with the combined load

of λ_{cr} ensures that any local maximum is a global maximum and that a gradient-based methods (such as SQP,...) will converge to a global maximum.⁴

- From the optimal vector $(\xi_D)^*$ find the associated SS^* by solving

$$\begin{cases} \max \lambda_{cr}(SS) \\ \text{s.t. } \xi_D(SS_{feasible}) \in \mathcal{N}((\xi_D)^*) \end{cases}$$

where $\mathcal{N}((\xi_D)^*)$ denotes the neighborhood of the optimal vector $(\xi_D)^*$. To define this neighborhood, we consider the ball $\mathcal{B}((\xi_D)^*, r)$ of center $(\xi_D)^*$ and of radius r chosen such that the cardinal of points inside \mathcal{B} is equal to 20 (fixed number). The reason for choosing the best stacking within a neighborhood is again related on the concavity of λ_{cr} . Indeed, this ensures that the discrete optimum can not be too far away from the continuous optimum. There is no reason however that the discrete optimum corresponds to the closest discrete stacking sequence. This local research could definitely be improved by constructing a new distance that will weight the influence of each ξ_D^i (based on the Hessian of λ_{cr} at the optimum for instance).

6.3 Comparison of the optimization results

We consider the Graphite/Epoxy material given in Table 1, a plate of 650×250 mm with 10×2 plies according to

⁴Convergence cannot be simply ensured because of the non-differentiability of the objective function. In reality, this problem is a typical non smooth optimization problem (being formally a maximization problem of an objective function computed through a minimization problem). Optimality conditions need a more general definition for the gradient (such as Clarke’s subdifferential, see Kiwiel (1985) and Lewis (2006)) and well-posedness and convergence can be shown for this particular problem where the objective function is a piecewise differentiable concave function. We did not use here such optimality conditions nor specific algorithms for non smooth optimizations (e.g Shor’s algorithm).

the symmetry. A piecewise polynomial regression of degree 2 (MOQ) is built to approximate the critical buckling factor λ_{cr} and is used in the optimization process to reduce computational time (compared with the RR approach).

6.3.1 A combined loading example

First the combined loading $N_x = -50 \text{ N.mm}^{-1}$, $N_y = 20 \text{ N.mm}^{-1}$ and $N_{xy} = 20 \text{ N.mm}^{-1}$ is considered. We report in Table 6 the results obtained with the two approaches: in all cases we obtained the same optimal stacking sequence, the only difference results in the number of objective function evaluations: 7,032 by listing all the feasible subset of stacking sequences with 10×2 plies, 269 function evaluations for the continuous optimization by using the two-step approach with the RR method to compute λ_{cr} instead of 39 evaluations with the MOQ approximation. For the SS-identification process, 20 more evaluations of λ_{cr} are required to build the neighborhood $\mathcal{B}((\xi_D)^*, r)$ of the optimal vector and then choose the optimal stacking sequence.

The convergence of the optimization process is compared on Fig. 9. Due to the smoothness of the quadratic approximate model (MOQ), the number of iterations is drastically reduced with 9 iterations instead of 21 with the RR method. Figure 10 shows the feasible out-of-plane lamination parameters in Miki’s tetrahedron. The considered neighborhood of the continuous solution is also represented and contains the feasible optimal solution. As expected, we can see that these optimal parameters are on the boundary of the domain. The use of a piecewise polynomial approximation of the λ_{cr} offers many advantages in term of convergence, computing time and makes it possible in this case to find the reference solution.

6.3.2 A pure shear loading example

To confirm these first very promising results, a second loading case is considered with $N_x = N_y = 0 \text{ N.mm}^{-1}$ and

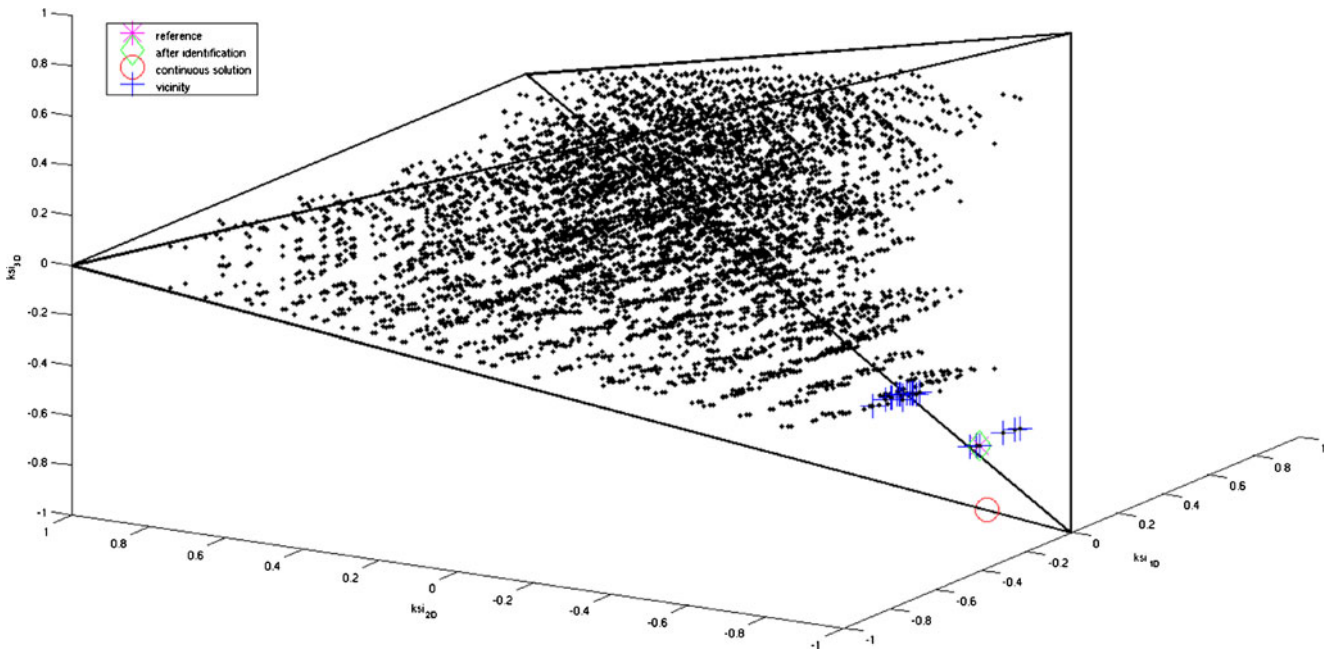


Fig. 10 Optimal solution in the lamination parameters space (the 7,032 feasible ξ_D are illustrated by a black dot market) under the combined loading

Table 7 Optimization result with the shear loading $N_x = N_y = 0 \text{ N.mm}^{-1}$ and $N_{xy} = 20 \text{ N.mm}^{-1}$

	Explicit 2-step with RR		2-step with MOQ
Optimal SS	[-45/ -45/ -45/ -45/90/45/45/45/0] _s		
Error at optimum (%)	Exact	Exact	0.1%
Number of λ_{cr} evaluations	7,032	38 + 20	32 + 20

$N_{xy} = 20 \text{ N.mm}^{-1}$. The results are presented in Table 7. As previously, the reference stacking sequence is found by the two-step approach in both cases: with the RR method or the MOQ method to compute λ_{cr} . The number of iterations is very close and Fig. 11 confirms that the objective function converges in a similar way (8 or 9 iterations). One more time, the optimal solution belongs to the boundary of the Miki' tetrahedron as shown in Fig. 12.

6.4 Conclusion of the optimization applications

The use of our approximation model of λ_{cr} is particularly useful for the optimization strategy developed here. The

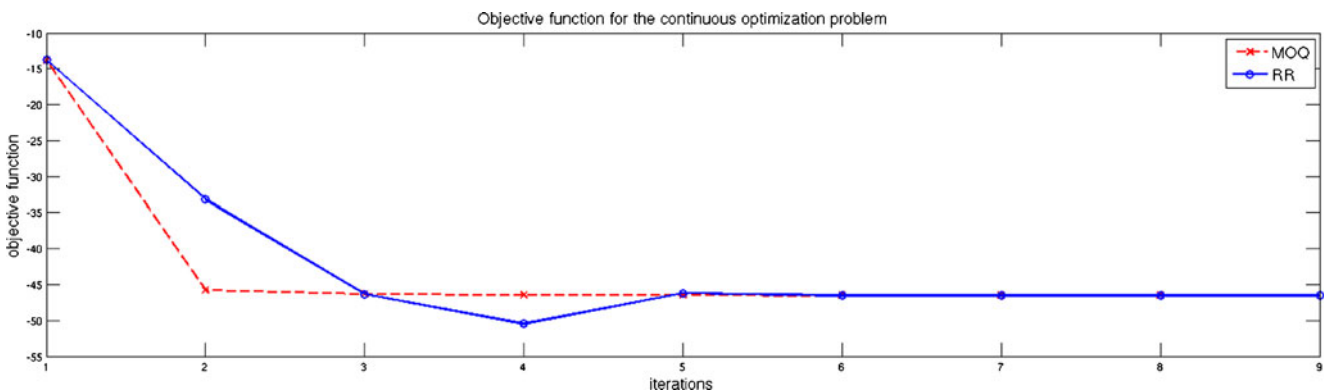


Fig. 11 Objective function during the optimization process with the shear loading

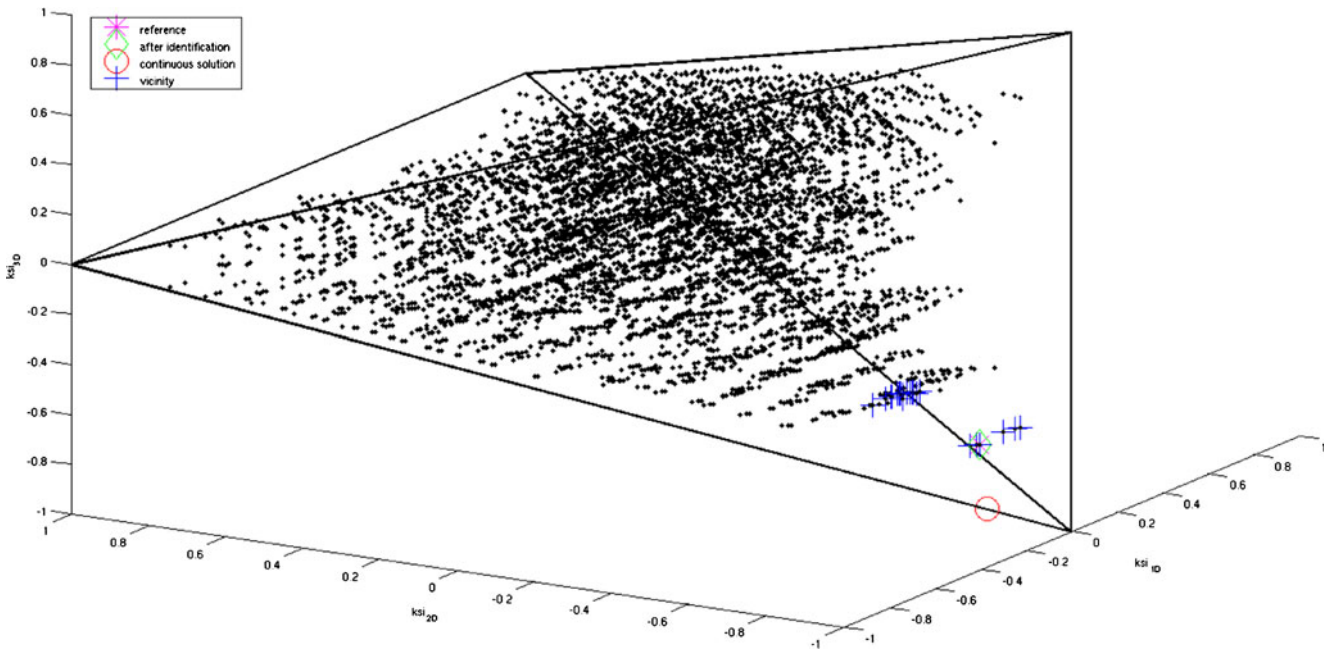


Fig. 12 Optimal solution in the lamination parameters space (the 7,032 feasible ξ_D are illustrated by a black dot market) under the shear loading

approach in 2 stages with MOQ makes it possible to find the reference solution with few evaluations of the objective function, and in a negligible computational time. If the number of plies n would be higher, acceptable space would be larger and thus one could clarify all the combinations with difficulty. One will prefer in this case to use an evolutionary algorithm (genetic algorithm, ...) coupled with the MOQ approximation for the objective function to perform the stacking sequence identification.

7 Conclusions and perspectives

The behavior of the critical buckling load λ_{cr} for laminate composite plate was investigated. The dependence of λ_{cr} with respect to the stacking dependent lamination parameters was studied and shown to be concave. When considering internal load redistribution, the loading conditions vary it is therefore important to have a good approximation of λ_{cr} . It is shown that the reciprocal of λ_{cr} is easier to describe than λ_{cr} and then an approximation model of λ_{cr} building through the use of the reciprocal can achieve a reasonable accuracy. The basis of the approximation strategy used here takes advantage of the piecewise behavior of λ_{cr} and it is shown that piecewise approximation models based on simple model (linear, quadratic, rational) often performs better than a sophisticated artificial neural network. More precisely, this piecewise approximation model seems to increase the approximation power for a large part of the validation bases, however a few points are misclassified

leading to a large error for a small part of the validation bases. The accuracy of the approximation is often on the same range as an artificial neural network for global accuracy metrics (MSE) but for relative error distribution based metrics, piecewise models are much more accurate. When used within an optimization process, these piecewise models show very good performance, since it gives the same stacking sequence as the discrete optimization process and it seems to converge faster than the optimization based on the real buckling computations. As noticed in Bettebghor et al. (2011) one of the drawbacks of that piecewise modeling is that it may require quite a lot of points to correctly detect and grasp regions where λ_{cr} behaves simply.

One way of improvement for that technique would be to consider the gradient of the function to approximate directly in the clustering since the regions of λ_{cr} are totally disconnected which would make the clustering even easier. Another way would be to build directly the more appropriate local model, for instance, for low shear combined loadings, linear model gives very good accuracy but for high shear, the dependence seems more complicated and an adaptive strategy that automatically selects the local model (e.g amongst linear, sparse or full polynomial model) would definitely improve the method. Regarding potential applications, this technique is likely to be appropriate to build approximation model of optimization results. For instance, in a bilevel optimization framework, one often consider maximizing the critical buckling factor with respect to thickness and in-plane lamination parameters (or equivalently relative thicknesses). Such functions are known to be

derivative-discontinuous and even discontinuous whenever discrete local optimization variables are addressed. This piecewise approximation model strategy would definitely be of great interest when building such approximation models. Another interesting area of application of such an approximation strategy would be structural optimization problems including aeroelasticity constraints and structural dynamics constraints where some of the quantities of interest are often computed through an eigenvalue problem similar to the buckling equation. Regarding the specific area of the approximation of buckling computations, this technique would also be of great interest for any homogenized anisotropic materials and even inhomogeneous anisotropic material such as curved-fibers composites.

Acknowledgments The authors are deeply grateful to the reviewers and Pr. Nagendra Somanath for their meticulous reading about the manuscript making several useful remarks. They would like to also thank Pr. A. Henrot and Pr. Z. Gürdal for their useful comments and discussions regarding buckling issues.

Appendix A: Quasiconcave functions and homogeneous functions

We recall here some basic definitions and properties of quasiconcave functions and homogeneous functions. In particular, we also give the Euler identity formula for homogeneous functions. Let $f : V \mapsto \mathbb{R}$ defined over a convex subset V of a vector space, f is *quasiconcave* if

$$\forall x, y \in V \quad \forall \theta \in (0, 1) f(\theta x + (1 - \theta)y) \geq \min(f(x), f(y)) \tag{44}$$

or equivalently, a quasiconcave function can be defined in terms of *upper level sets* $S_\alpha = \{x \in V | f(x) \geq \alpha\}$ that is, f is quasiconcave if for any $\alpha \in \mathbb{R}$, S_α is a convex set. This is the definition we used in this article. Quasiconcavity (and quasiconvexity) are a generalization of the well known concavity, a concave function is quasiconcave. Quasiconcave functions often appear in economics (e.g Cobb–Douglas function), game theory and optimization. These are assumptions for the minmax theorem.

Let $f : U \mapsto V$ a function defined a vector space U that maps to another vector space V . f is said to be *positive homogeneous* of degree α whenever

$$\forall x \in U \quad \forall \kappa \in \mathbb{R}^+ \quad f(\kappa x) = \kappa^\alpha f(x) \tag{45}$$

Linear functions are homogeneous functions of degree 1, the determinant function over $n \times n$ matrices is homogeneous of degree n . One important feature of homogeneous functions whenever they are continuously differentiable is

that they admit the Euler identity characterization. $f : \mathbb{R}^n \mapsto \mathbb{R}$ is a positive homogeneous functions if and only if

$$\forall x \in \mathbb{R}^n \quad \sum_{i=1}^n x_i \frac{\partial f}{\partial x_i} = \alpha f(x) \tag{46}$$

To obtain (46), we simply differentiate (45) at κx with respect to κ to get

$$\sum_{i=1}^n x_i \frac{\partial f(\kappa x)}{\partial x_i} = \alpha \kappa^{\alpha-1} f(x). \tag{47}$$

It is worth noting that, if F is homogeneous of degree α and continuously differentiable, then its derivative is homogeneous of degree $\alpha - 1$. We end up with the following useful property, whose proof together with a lot of results and examples on quasiconcave and homogeneous functions can be found in Simon and Blume (1994). Let f be a homogeneous function of degree 1 and quasiconcave then f is concave.

Appendix B: Rayleigh–Ritz approximation for buckling and sensitivity analysis

We briefly recall here the Rayleigh–Ritz method for buckling computations. Note that, Rayleigh–Ritz methods are also widely used in structural dynamics, vibrations, quantum chemistry,... The main idea is to write the solution of a partial differential equation system or eigenproblem as a linear combination of test functions. Unlike finite element methods, the test functions are defined over the whole domain and satisfy boundary conditions. These basis functions are usually based on a closed-form solutions of a close unperturbed problem. This means that Rayleigh–Ritz methods are used when solving a pde or eigenproblem close in some sense (simple geometry, simple differential operator) to a simple one and are not as general as finite element methods. They are also close to spectral methods in the sense the test functions are defined over the whole domain and their convergence properties are usually quite good whenever the problem remains close to an original simple problem. In addition to that, as in spectral methods, the discretization step leads to dense matrices and high accuracy can be achieved with much less degrees-of-freedom than in finite elements methods.

Consider the following eigenproblem under its variational form: find $\lambda \in \mathbb{R}$ and $u \in \mathcal{V} - \{0\}$ (a Hilbert space) such that

$$a(u, u) - \lambda b(u, u) = 0 \tag{48}$$

with prescribed boundary conditions (e.g $u = 0$ on the boundary of the domain), with a and b being bilinear symmetric forms defined over \mathcal{V} , for sake of simplicity we will

assume that they are both coercive and restrict ourselves to eigenproblems. The minmax theorem (or variational characterization) for eigenvalues are local minima of the Rayleigh ratio

$$R(u) = \frac{a(u, u)}{b(u, u)} \tag{49}$$

To ensure unicity of eigenmodes, it is convenient to add a normalization condition⁵ over u (see Haftka and Gurdal 1992, e.g $b(u, u) = 1, \|b\|_\infty = 1$). Let's take $b(u, u) = 1$. The process of computing eigenvalues and eigenmodes of (48) boils down to solving the following problem

$$\begin{aligned} \min_{u \in \mathcal{V}} \quad & R(u) \\ \text{s.t} \quad & b(u, u) = 1 \end{aligned} \tag{50}$$

Instead of solving problem (50), we solve it over a finite-dimensional space \mathcal{V}_N . Consider (u_1, \dots, u_N) the basis of $\mathcal{V}_N \subset \mathcal{V}$ and let's denote for $v \in \mathcal{V}_N$

$$v(x, c) = \sum_{i=1}^N c_i u_i(x) \tag{51}$$

where $c \in \mathbb{R}^N$, and we can write finite-dimensional approximation of a and b

$$a(v, v) = a_N(c) = \sum_{i,j=1}^N a_{ij} c_i c_j \tag{52}$$

with

$$a_{ij} = a(u_i, u_j) \tag{53}$$

and

$$b(v, v) = b_N(c) = \sum_{i,j=1}^N b_{ij} c_i c_j \tag{54}$$

with

$$b_{ij} = b(u_i, u_j) \tag{55}$$

and denote by A and B the (symmetric positive definite) matrices of elements $(a_{ij})_{1 \leq i, j \leq N}$ and $(b_{ij})_{1 \leq i, j \leq N}$. We can write the KKT optimality conditions of problem (50) over \mathbb{R}^N

$$\min_{c \in \mathbb{R}^N | b_N(c)=1} a_N(c) \tag{56}$$

⁵This is indeed necessary to ensure unicity of the eigenmodes, since they are defined up to a multiplicative constant. This is even more critical when we have to compute the derivatives of the eigenmodes.

and so $c \in \mathbb{R}^N$ a local minimum of problem (56) necessary satisfies the following conditions (any point satisfying these conditions is also a local minimum since a_N and b_N are convex)

$$\frac{1}{2} \nabla a_N(c) - \lambda \nabla b_N(c) = (A - \lambda B)c = 0 \tag{57}$$

with

$$b_N(c) = c^T B c = 1 \tag{58}$$

And the linear equation defined by (57) has a nontrivial solution if and only if

$$\det(A - \lambda B) = 0 \tag{59}$$

that the Lagrange multiplier λ is an eigenvalue of A relatively to B .

To make this method more clear, let's take the example of composite plate buckling. As outlined in this paper, for simple geometry like plate, there exists a closed-form solution whenever the laminate is orthotropic and the loading is biaxial. In that case, buckling equation boils down to

$$\begin{aligned} D_{11} \frac{\partial^4 w}{\partial x^4} + D_{22} \frac{\partial^4 w}{\partial y^4} + 2(D_{12} + 2D_{66}) \frac{\partial^4 w}{\partial x^2 \partial y^2} \\ = N_x \frac{\partial^2 w}{\partial x^2} + N_y \frac{\partial^2 w}{\partial y^2} \end{aligned} \tag{60}$$

Closed form solution of (60) are obtained considering the following test functions

$$w_{ij}(x, y) = \sin(i\pi/x) \sin(j\pi/y) \tag{61}$$

and writing (60) we obtain the corresponding eigenvalues λ_{ij} as defined in (11). We therefore use $(w_{ij})_{i,j \in [1,N]}$ as test functions, leading to matrices A and B of size $N^2 \times N^2$ defined as

$$A_{(i-1)N+j, (k-1)N+l} = a_D(w_{ij}, w_{kl}) \tag{62}$$

and

$$B_{(i-1)N+j, (k-1)N+l} = b_N(w_{ij}, w_{kl}) \tag{63}$$

and λ_{cr} is found as the smallest positive eigenvalue of A relatively to B . It is worth noting here that when making the parameters of that buckling computation vary, there is no need to reassemble each time the full matrices A and B . One

easy way to save computation is to assemble elementary matrices once for all such that for instance

$$A_D = D_{11}A_{11} + \dots + D_{16}A_{16} \tag{64}$$

where A_{ij} are matrices of size $N^2 \times N^2$ computed while performing the first analysis. This allows to perform new computations for new values of D faster, since the assembly becomes negligible and the main part of the computations lies in the eigensolver.

Suppose now we want to compute the sensitivity of λ_{cr} w.r.t to any design variable x , we have

$$\frac{\partial \lambda_{cr}}{\partial x} = w_{cr}^T \left(\frac{\partial A}{\partial x} + \lambda_{cr} \frac{\partial B}{\partial x} \right) w_{cr} \tag{65}$$

with w_{cr} the associated eigenvector whenever λ_{cr} is simple.

Appendix C: Buckling space for loading conditions

In Fig. 6 we see that some values of (N_x, N_y, N_{xy}) do not give rise to buckling. In both plots we observe a 'hole' for which buckling analysis gives either a negative critical buckling factor or an infinite value. This is related to the assumption that the bilinear form $b_N(\cdot, \cdot)$ has at least one positive direction, i.e a displacement w such that $b_N(w, w) > 0$. This assumption ensures that there exists at least one eigenvalue $\lambda > 0$, the positive part of the spectrum is not empty. In the literature concerned with linear buckling in the frame of three-dimensional linear elasticity (Dauge and Suri 2006; Szabó and Királyfalvi 1999), this condition is written over the stress tensor and is referred to as *non-negativity property*. The aim of that section is to show that this assumption naturally explains the 'non-buckling' area in the loading conditions space.

The bilinear form $b_N(\cdot, \cdot)$ is

$$b_N(u, u) = - \iint_{\Omega} N_x \frac{\partial u}{\partial x} + N_y \frac{\partial u}{\partial y} + N_{xy} \frac{\partial u}{\partial x} \frac{\partial u}{\partial y} dx dy \tag{66}$$

Equation (66) can be written in a matrix form

$$b_N(u, u) = - \iint_{\Omega} \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \end{pmatrix} \begin{pmatrix} N_x & \frac{N_{xy}}{2} \\ \frac{N_{xy}}{2} & N_y \end{pmatrix} \begin{pmatrix} \frac{\partial u}{\partial x} \\ \frac{\partial u}{\partial y} \end{pmatrix}^T dx dy \tag{67}$$

and the non-negativity property is satisfied whenever the forces tensor N

$$N = \begin{pmatrix} N_x & \frac{N_{xy}}{2} \\ \frac{N_{xy}}{2} & N_y \end{pmatrix} \tag{68}$$

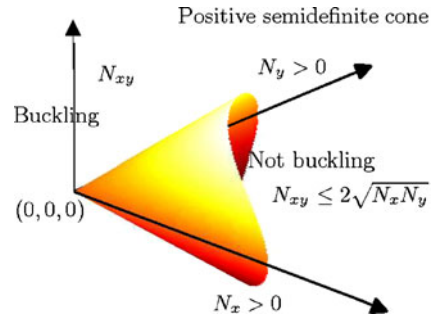


Fig. 13 Buckling and non-buckling zones in loading space

has at least one negative eigenvalue. This ensures that there exists a positive eigenvalue and that $\lambda_1 \geq 0$. Equivalently, loading conditions do not give rise to buckling whenever tensor N is semidefinite positive. The buckling zone in (N_x, N_y, N_{xy}) is then the complement in \mathbb{R}^3 of the positive semidefinite cone. This then leads to the condition that buckling occurs if and only if

$$|N_{xy}| > 2\sqrt{N_x N_y}. \tag{69}$$

We depicted the buckling and non buckling zones and the positive semidefinite cone in Figs. 13 and 14. Now we can comment and explain the main part of the 'hole' in Fig. 6 when compared to the buckling/non-buckling zones in (θ, ϕ) space depicted Fig. 14. The rest of the hole can be explained by the fact as a discrete numerical approximation scheme was used, we could not explore all the different modes. Typically in our case, since 20×20 basis functions were used, modes with more than 21 half-waves along one direction can not be detected.

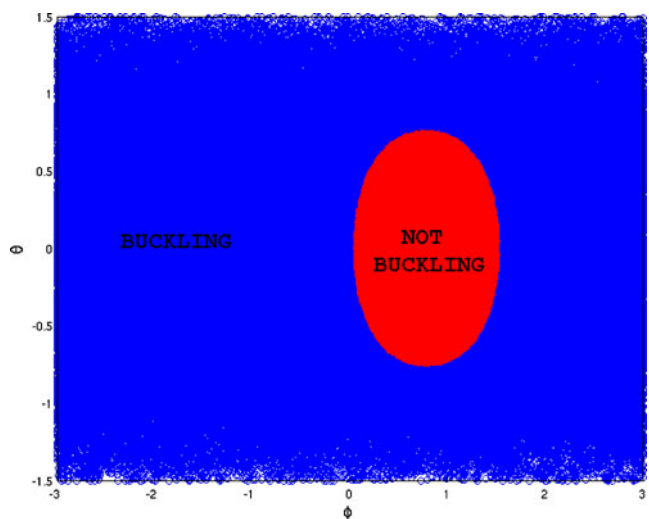


Fig. 14 Buckling and non-buckling zones in (θ, ϕ) space

References

- Abu-Odeh AY, Jones HL (1998) Optimum design of composite plates using response surface method. *Compos Struct* 43(3):233–242
- Akulenko LD, Nesterov SV (2005) High precision methods in eigenvalue problems and their applications. CRC Press, Boca Raton, FL
- Antman SS (2006) Theodore von Kármán. In: A panorama of Hungarian mathematics in the twentieth century I, pp 373–382
- Antman SS, Renardy M (2005) Nonlinear problems of elasticity. Springer
- Ashbaugh MS (1999) Isoperimetric and universal inequalities for eigenvalues. In: Spectral theory and geometry: ICMS instructional conference, Edinburgh 1998. Cambridge University Press, Cambridge, UK, p 95
- Autio M (2001) Optimization of coupled thermal-structural problems of laminated plates with lamination parameters. *Struct Multidisc Optim* 21(1):40–51
- Bettebghor D, Bartoli N, Grihon S, Morlier J, Samuelides M (2011) Surrogate modeling approximation using a mixture of experts based on EM joint estimation. *Struct Multidisc Optim* 43:1–17
- Bloom F, Coffin D (2000) Handbook of thin plate buckling and postbuckling. CRC Press, Boca Raton, FL
- Bloomfield MW, Diaconu CG, Weaver PM (2009) On feasible regions of lamination parameters for lay-up optimization of laminated composites. *Proc R Soc A, Math Phys Eng Sci* 465(2104):1123
- Courant R (1994) Variational methods for the solution of problems of equilibrium and vibrations. *Lect Notes Pure Appl Math* 1–1
- Dauge M, Suri M (2006) On the asymptotic behaviour of the discrete spectrum in buckling problems for thin plates. *Math Methods Appl Sci* 29(7):789–817
- De Gournay F (2005) Optimisation de formes par la méthode des courbes de niveaux
- Diaconu CG, Sato M, Sekine H (2002) Feasible region in general design space of lamination parameters for laminated composites. *AIAA J* 40(3):559–565
- Friedman J, Hastie T, Tibshirani R (2001) The elements of statistical learning
- Fukunaga H, Sekine H, Sato M, Iino A (1995) Buckling design of symmetrically laminated plates using lamination parameters. *Comput Struct* 57(4):643–649
- Grenestedt JL (1989) A study on the effect of bending-twisting coupling on buckling strength. *Compos Struct* 12(4):271–290
- Grenestedt JL (1990) Composite plate optimization only requires one parameter. *Struct Multidisc Optim* 2(1):29–37
- Grenestedt JL (1991) Layup optimization against buckling of shear panels. *Struct Multidisc Optim* 3(2):115–120
- Haftka RT, Gurdal Z (1992) Elements of structural optimization. Kluwer
- Harrison PN, Le Riche R, Haftka RT (1995) Design of stiffened composite panels by genetic algorithm and response surface approximations. In: AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics and materials conference, 36th and AIAA/ASME adaptive structures forum, New Orleans, LA, pp 58–68
- Henrot A (2006) Extremum problems for eigenvalues of elliptic operators: Antoine Henrot. Birkhauser, Cambridge, MA
- Herencia JE, Haftka RT, Weaver PM, Friswell MI (2008a) Lay-up optimization of composite stiffened panels using linear approximations in lamination space. *AIAA J* 46(9):2387–2391
- Herencia JE, Weaver PM, Friswell MI (2008b) Optimization of anisotropic composite panels with T-shaped stiffeners including transverse shear effects and out-of-plane loading. *Struct Multidisc Optim* 37(2):165–184
- Kiwiel KC (1985) Methods of descent for nondifferentiable optimization. Springer, Berlin
- Koiter WT (1962) Elastic stability and post-buckling behavior. *Nonlinear Probl* 257–275
- Lebedev LP, Vorovich II (2003) Functional analysis in mechanics. Springer
- Lekhnitskii SG (1968) Anisotropic plates. Technical report, DTIC document
- Lewis A (2006) Eigenvalues and nonsmooth optimization. *Lond Math Soc Lect Note Ser* 331:208
- Miki M (1983) A graphical method for designing fibrous laminated composites with required in-plane stiffness. *Trans Jpn Soc Compos Mater* 9:51–55
- Miki M, Sugiyama Y (1991) Optimum design of laminated composite plates using lamination parameters. In: AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics, and materials conference, 32nd, Baltimore, MD, technical papers, Pt. 1, vol 8
- Reddy JN (2004) Mechanics of laminated composite plates and shells, vol 2. CRC Press, Boca Raton, FL
- Rousselet B, Chenais D (1990) Continuité et différentiabilité d'éléments propres: application à l'optimisation de structures. *Appl Math Optim* 22(1):27–59
- Simon CP, Blume L (1994) Mathematics for economists, vol 7. Norton, New York
- Stakgold I (1971) Branching of solutions of nonlinear equations. *SIAM Rev* 13(3):289–332
- Szabó B, Királyfalvi G (1999) Linear models of buckling and stress-stiffening. *Comput Methods Appl Mech Eng* 171(1):43–59
- Terada Y, Todoroki A, Shimamura Y (2001) Stacking sequence optimizations using fractal branch and bound method for laminated composites. *JSME Int J Ser A* 44(4):490–498
- Todoroki A, Sasai M (2002) Stacking sequence optimizations using GA with zoomed response surface on lamination parameters. *Adv Compos Mater* 11(3):299–318
- Todoroki A, Suenaga K, Shimamura Y (2003) Stacking sequence optimizations using modified global response surface in lamination parameters. *Adv Compos Mater* 12(1):35–55
- Turvey GJ (1995) Buckling and postbuckling of composite plates. Springer
- Walker M, Adali S, Verijenko V (1996) Optimization of symmetric laminates for maximum buckling load including the effects of bending-twisting coupling. *Comput Struct* 58(2):313–319
- Whitney JM (1969) The effect of transverse shear deformation on the bending of laminated plates. *J Compos Mater* 3(3):534

4.4.5 Efficient global optimization for high-dimensional constrained problems by using the Kriging models combined with the partial least squares method, Engineering Optimization 2018




Efficient global optimization for high-dimensional constrained problems by using the Kriging models combined with the partial least squares method

Mohamed Amine Bouhlej, Nathalie Bartoli, Rommel G. Regis, Abdelkader Otsmane & Joseph Morlier


To cite this article: Mohamed Amine Bouhlej, Nathalie Bartoli, Rommel G. Regis, Abdelkader Otsmane & Joseph Morlier (2018): Efficient global optimization for high-dimensional constrained problems by using the Kriging models combined with the partial least squares method, Engineering Optimization, DOI: [10.1080/0305215X.2017.1419344](https://doi.org/10.1080/0305215X.2017.1419344)

To link to this article: <https://doi.org/10.1080/0305215X.2017.1419344>

 View supplementary material [↗](#)

 Published online: 01 Feb 2018.

 Submit your article to this journal [↗](#)

 View related articles [↗](#)

 View Crossmark data [↗](#)



Efficient global optimization for high-dimensional constrained problems by using the Kriging models combined with the partial least squares method

Mohamed Amine Bouhlef , Nathalie Bartoli , Rommel G. Regis, Abdelkader Otsmane and Joseph Morlier

ABSTRACT

In many engineering optimization problems, the number of function evaluations is often very limited because of the computational cost to run one high-fidelity numerical simulation. Using a classic optimization algorithm, such as a derivative-based algorithm or an evolutionary algorithm, directly on a computational model is not suitable in this case. A common approach to addressing this challenge is to use black-box surrogate modelling techniques. The most popular surrogate-based optimization algorithm is the efficient global optimization (EGO) algorithm, which is an iterative sampling algorithm that adds one (or many) point(s) per iteration. This algorithm is often based on an infill sampling criterion, called expected improvement, which represents a trade-off between promising and uncertain areas. Many studies have shown the efficiency of EGO, particularly when the number of input variables is relatively low. However, its performance on high-dimensional problems is still poor since the Kriging models used are time-consuming to build. To deal with this issue, this article introduces a surrogate-based optimization method that is suited to high-dimensional problems. The method first uses the 'locating the regional extreme' criterion, which incorporates minimizing the surrogate model while also maximizing the expected improvement criterion. Then, it replaces the Kriging models by the KPLS(+K) models (Kriging combined with the partial least squares method), which are more suitable for high-dimensional problems. Finally, the proposed approach is validated by a comparison with alternative methods existing in the literature on some analytical functions and on 12-dimensional and 50-dimensional instances of the benchmark automotive problem 'MOPTA08'.

ARTICLE HISTORY

Received 6 July 2016
Accepted 3 December 2017


KEYWORDS

Kriging; KPLS; Partial Least Squares; Optimization; Expected Improvement

1. Introduction

The field of optimization has become significantly important in engineering and industrial design. In the literature, many approaches have been used for finding the global optimum. One of the most popular methods is to run a gradient-based optimization algorithm with a multi-start procedure (Hickernell and Yuan 1997; Sendin, Banga, and Csendes 2009). For this type of method, the derivative of the objective function is needed. When the derivative of the objective function is not available, it can be approximated by finite-difference methods. However, finite-difference methods are unreliable when the function to be optimized is non-smooth. For this reason, derivative-free optimization methods (Conn, Scheinberg, and Vicente 2009) and heuristic methods such as simulated

CONTACT Mohamed Amine Bouhlef  mbouhlef@umich.edu

 Supplemental data for this article can be accessed at <https://doi.org/10.1080/0305215X.2017.1419344>

annealing (Laarhoven and Aarts 1987), evolutionary algorithms (Simon 2013), and particle swarm optimization (Kennedy and Eberhart 1995) became popular in the last few decades.

In many engineering problems, numerical simulations require several hours, sometimes more, to run one simulation representative of the physics. Thus, direct optimization methods for such cases are too expensive and infeasible in practice. In addition, many iterations are often required when the number of input variables is large. Likewise, to find the global optimum using a relatively small number of evaluations of the true function is almost impossible for high-dimensional problems. Hopefully, it is possible to obtain a reasonably good feasible solution through a good management of the number of function evaluations.

A suitable optimization approach is to use a surrogate model, also called metamodel, instead of the true function. The evaluation of new points through a surrogate model, called predictions, is very cheap since it consists of computing a simple analytical function. Moreover, the surrogate model is a good tool for indicating promising areas where new high-fidelity simulations should be run. Several approaches for constrained black-box optimization have been developed recently. For instance, Li *et al.* (2016) developed a Kriging-based algorithm for constrained black-box optimization that involves two phases. The first phase finds a feasible point while the second phase obtains a better feasible point. This approach is similar to what is done in the COBRA algorithm (Regis 2014, briefly described in Sec. 4.2.2 in this article). Liu *et al.* (2016) proposed a two-phase method that uses a constraint-handling technique based on the DIRECT algorithm and that uses an adaptive metamodeling strategy for the objective and constraints. Gramacy *et al.* (2016) developed a hybrid approach that uses Kriging with the expected improvement (EI) criterion combined with the augmented Lagrangian framework for constrained optimization. Moreover, Regis and Wild (2017) developed a model-based trust-region algorithm for constrained optimization that uses radial basis functions (RBF) models. In addition, Kriging-based approaches have been proposed for constrained multi-objective optimization (*e.g.* Singh *et al.* (2014); Feliot, Bect, and Vazquez (2016)). An important application of Kriging in an optimization context is the efficient global optimization (EGO) algorithm (Jones, Schonlau, and Welch 1998). It uses both the prediction and error estimations provided by Kriging to guide an infill sampling criterion towards promising areas. To the best of the authors' knowledge, EGO is used for problems with a relatively small number of input variables because of the Kriging limitations in high dimension (Haftka, Villanueva, and Chaudhuri 2016). Sasena (2002) has developed a constrained version of EGO, called SuperEGO (denoted SEGO in this article), where the optimization of the infill sampling criterion takes into account the satisfaction of constraint functions.

In this article, two new methods for solving high-dimensional constrained optimization problems are developed, both based on SEGO approach: (i) SEGOKPLS uses SEGO with the KPLS model (Bouhleb *et al.* 2016b) and (ii) SEGOKPLS+K uses SEGO with the KPLS+K model (Bouhleb *et al.* 2016a). The SEGOKPLS(+K) algorithms build KPLS(+K) for each output function at each iteration of optimization. Either KPLS or KPLS+K are used since they are faster to build than the classical Kriging models while maintaining a good accuracy for high-dimensional problems. Once the KPLS(+K) models are built, a specific infill sampling criterion, that adds promising points into the training points in order to improve the accuracy of the metamodels in relevant areas, is maximized. For this purpose, the 'locating the regional extreme' (WB2) criterion proposed by Watson and Barnes (1995) is used. Sasena, Papalambros, and Goovaerts (2002) implemented and validated this criterion on several analytical functions. This approach proposes a balance between exploitation and exploration of the metamodel—herein, exploitation means that the metamodel is minimized and exploration means that points are added where the uncertainty of the metamodel is high. Moreover, WB2 is more local than the popular EI criterion. This latter characteristic is important for expensive high-dimensional problems, since the number of true function evaluations is limited and the uncertainty of the metamodel is high.

There have been other methods that use principal components for high-dimensional surrogate-based optimization. For example, Kyriacou, Asouti, and Giannakoglou (2014) used principal

components analysis (PCA) in a metamodel-assisted evolutionary algorithm to guide the application of evolution operators and the training of the metamodel. Moreover, Chen *et al.* (2015) used Karhunen-Loève expansion, which is similar to PCA, to reduce design-space dimensionality in metamodel-based shape optimization. The proposed approach, which is based on the EGO algorithm, incorporates a Kriging model that uses principal components in the context of partial least squares to make it suitable for high-dimensional constrained optimization. Several analytical problems and an automotive problem (MOPTA08) are tested with the proposed approaches. In addition, a comparison between SEGOKPLS(+K) and several optimization methods existing in the literature is provided. This comparison has shown that the SEGOKPLS(+K) methods outperform the alternative algorithms on most of the test problems.

The article is organized as follows. Section 2 details the proposed optimization algorithms, SEGOKPLS(+K). Sections 3 and 4 apply and validate the SEGOKPLS(+K) algorithms on the analytical and MOPTA08 problems, respectively. Finally, Section 5 concludes the article.

2. SEGOKPLS(+K) for high-dimensional constrained optimization problems

Assume that a deterministic cost function has been evaluated at n points $\mathbf{x}^{(i)}$ ($i = 1, \dots, n$) with $\mathbf{x}^{(i)} = [x_1^{(i)}, \dots, x_d^{(i)}] \in B$, with $B \subset \mathbb{R}^d$. For simplicity, consider B to be a hypercube expressed by the product of intervals of each direction space, *i.e.* $B = \prod_{j=1}^d [a_j, b_j]$, where $a_j, b_j \in \mathbb{R}$ with $a_j \leq b_j$ for $j = 1, \dots, d$. Also, denote by \mathbf{X} the matrix $[(\mathbf{x}^{(1)})^t, \dots, (\mathbf{x}^{(n)})^t]^t$. Evaluating these n inputs gives the outputs $\mathbf{y} = [y^{(1)}, \dots, y^{(n)}]^t$ with $y^{(i)} = y(\mathbf{x}^{(i)})$, for $i = 1, \dots, n$.

2.1. The KPLS(+K) models

The KPLS and KPLS+K models are derived by combining the Kriging model with the partial least squares (PLS) technique. Therefore, the theory of Kriging and PLS are briefly depicted before describing the KPLS(+K) models.

2.1.1. The Kriging model

The Kriging model has been developed first in geostatistics (Krige 1951; Cressie 1988; Goovaerts 1997) before being extended to computer experiments (Sacks *et al.* 1989; Schonlau 1998; Jones, Schonlau, and Welch 1998; Sasena, Papalambros, and Goovaerts 2002; Forrester, Sobester, and Keane 2006; Picheny *et al.* 2010; Liem, Mader, and Martins 2015). In order to develop the Kriging model, assume that the deterministic response $y(\mathbf{x})$ is a realization of a stochastic process $Y(\mathbf{x})$ (Koehler and Owen 1996; Schonlau 1998; Sasena 2002)

$$Y(\mathbf{x}) = \beta_0 + Z(\mathbf{x}), \quad (1)$$

where β_0 is an unknown constant and $Z(\mathbf{x})$ is a stochastic term considered as a realization of a stationary Gaussian process with $\mathbb{E}[Z(\mathbf{x})] = 0$ and a covariance function, also called a kernel function, given by

$$\text{Cov}(Z(\mathbf{x}), Z(\mathbf{x}')) = k(\mathbf{x}, \mathbf{x}') = \sigma^2 r_{\mathbf{x}\mathbf{x}'}, \quad \forall \mathbf{x}, \mathbf{x}' \in B, \quad (2)$$

where σ^2 is the process variance and $r_{\mathbf{x}\mathbf{x}'}$ is the correlation function between \mathbf{x} and \mathbf{x}' . In the following, the exponential covariance function type is only considered, in particular the squared Gaussian

kernel, given by

$$k(\mathbf{x}, \mathbf{x}') = \sigma^2 \prod_{i=1}^d \exp\left(-\theta_i (x_i - x'_i)^2\right), \quad \forall \theta_i \in \mathbb{R}^+. \quad (3)$$

Equation (1) corresponds to a particular case of the Kriging model: the ordinary Kriging model (Forrester, Sóbester, and Keane 2008). The function k , given by Equation (3), depends on some hyperparameters θ . To construct the Kriging model, these hyperparameters are assumed to be known. Also, denote the $n \times 1$ vector as $\mathbf{r}_{\mathbf{xx}} = [r_{\mathbf{xx}(1)}, \dots, r_{\mathbf{xx}(n)}]^t$ and the $n \times n$ covariance matrix as $\mathbf{R} = [\mathbf{r}_{\mathbf{x}(1)\mathbf{X}}, \dots, \mathbf{r}_{\mathbf{x}(n)\mathbf{X}}]$. Moreover, $\hat{y}(\mathbf{x})$ denotes the prediction of the true function $y(\mathbf{x})$. Under the hypothesis considered above, the best linear unbiased predictor for $y(\mathbf{x})$, given the observations \mathbf{y} , is

$$\hat{y}(\mathbf{x}) = \hat{\beta}_0 + \mathbf{r}_{\mathbf{xx}}^t \mathbf{R}^{-1} (\mathbf{y} - \hat{\beta}_0 \mathbf{1}), \quad (4)$$

where $\mathbf{1}$ denotes an n -vector of ones and

$$\hat{\beta}_0 = (\mathbf{1}^t \mathbf{R}^{-1} \mathbf{1})^{-1} \mathbf{1}^t \mathbf{R}^{-1} \mathbf{y}. \quad (5)$$

In addition, the estimate of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{n} (\mathbf{y} - \mathbf{1} \hat{\beta}_0)^t \mathbf{R}^{-1} (\mathbf{y} - \mathbf{1} \hat{\beta}_0). \quad (6)$$

Moreover, the ordinary Kriging model provides an estimate of the variance of the prediction given by

$$s^2(\mathbf{x}) = \hat{\sigma}^2 (1 - \mathbf{r}_{\mathbf{xx}}^t \mathbf{R}^{-1} \mathbf{r}_{\mathbf{xx}}). \quad (7)$$

Finally, the vector of hyperparameters $\theta = \{\theta_i\}$, for $i = 1, \dots, d$, is estimated by the maximum likelihood estimation method.

2.1.2. The partial least squares technique

The PLS method is a statistical method that searches out the best multidimensional direction \mathbf{X} that explains the characteristics of the output \mathbf{y} . Moreover, it finds a linear relationship between input variables and output variables by projecting input variables onto principal components (Wold 1966). The PLS technique reveals how inputs depend on the output variable. In the following, h denotes the number of principal components retained that is much lower than d ($h \ll d$). The PLS components are computed sequentially. In fact, the principal component \mathbf{t}_l is computed by seeking the best direction $\mathbf{w}^{(l)}$ that maximizes the squared covariance between $\mathbf{t}_l = \mathbf{X}^{(l-1)} \mathbf{w}^{(l)}$ and $\mathbf{y}^{(l-1)}$

$$\mathbf{w}^{(l)} = \begin{cases} \operatorname{argmax}_{\mathbf{w}^{(l)}} \mathbf{w}^{(l)t} \mathbf{X}^{(l-1)t} \mathbf{y}^{(l-1)} \mathbf{y}^{(l-1)t} \mathbf{X}^{(l-1)} \mathbf{w}^{(l)} \\ \text{s.t. } \mathbf{w}^{(l)t} \mathbf{w}^{(l)} = 1. \end{cases} \quad (8)$$

where $\mathbf{X} = \mathbf{X}^{(0)}$, $\mathbf{y} = \mathbf{y}^{(0)}$, and, for $l = 1, \dots, h$, $\mathbf{X}^{(l)}$ and $\mathbf{y}^{(l)}$ are the residual matrix from the local regression of $\mathbf{X}^{(l-1)}$ onto the principal component \mathbf{t}_l and from the local regression of $\mathbf{y}^{(l-1)}$ onto the principal component \mathbf{t}_l , respectively, such that

$$\begin{aligned} \mathbf{X}^{(l-1)} &= \mathbf{t}_l \mathbf{p}^{(l)} + \mathbf{X}^{(l)}, \\ \mathbf{y}^{(l-1)} &= c_l \mathbf{t}_l + \mathbf{y}^{(l)}, \end{aligned} \quad (9)$$

where $\mathbf{p}^{(l)}$ (a $1 \times d$ vector) and c_l (a coefficient) contain the regression coefficients. For more details of how the PLS method works, please see the work by Helland (1988); Frank and Friedman (1993); Alberto and González (2012).

The principal components represent the new coordinate system obtained upon rotating the original system with axes, x_1, \dots, x_d (Alberto and González 2012). For $l = 1, \dots, h$, \mathbf{t}_l is written as

$$\mathbf{t}_l = \mathbf{X}^{(l-1)} \mathbf{w}^{(l)} = \mathbf{X} \mathbf{w}_*^{(l)}. \quad (10)$$

This important relationship is mainly used for developing the KPLS model that is described in the following section.

2.1.3. Construction of the KPLS and KPLS+K models

The hyperparameters $\boldsymbol{\theta} = \{\theta_i\}$, for $i = 1, \dots, d$, given by Equation (3) can be interpreted as measuring how strongly the variables x_1, \dots, x_d , respectively, affect the output y . For building KPLS, consider the coefficients given by the vectors $\mathbf{w}_*^{(l)}$, for $l = 1, \dots, h$, as a measure of the influence of input variables x_1, \dots, x_d on the output y . By some elementary operations applied on the kernel functions, define the KPLS kernel by

$$k_{1:h}(\mathbf{x}, \mathbf{x}') = \prod_{l=1}^h k_l(F_l(\mathbf{x}), F_l(\mathbf{x}')), \quad (11)$$

where $k_l : B \times B \rightarrow \mathbb{R}$ is an isotropic stationary kernel and

$$\begin{aligned} F_l : B &\longrightarrow B \\ \mathbf{x} &\longmapsto \left[w_{*1}^{(l)} x_1, \dots, w_{*d}^{(l)} x_d \right]. \end{aligned} \quad (12)$$

More details of such construction are given in Bouhlel *et al.* (2016b). Considering the example of the squared Gaussian kernel given by Equation (3) yields

$$k(\mathbf{x}, \mathbf{x}') = \sigma^2 \prod_{l=1}^h \prod_{i=1}^d \exp \left[-\theta_l \left(w_{*i}^{(l)} x_i - w_{*i}^{(l)} x'_i \right)^2 \right], \forall \theta_l \in \mathbb{R}^+. \quad (13)$$

This equation is the new kernel associated to the KPLS model. Since a small number of principal components is retained, the estimation of the hyperparameters $\theta_1, \dots, \theta_h$ is very fast compared to the hyperparameters $\theta_1, \dots, \theta_d$ given by Equation (3), where $h \ll d$. Moreover, the solution of the maximum likelihood of the KPLS covariance function can be improved. To do this, add a new step, right after the estimation of the θ_l -parameters, that is based on the following transition

$$\begin{aligned} k_{1:h}(\mathbf{x}, \mathbf{x}') &= \sigma^2 \prod_{l=1}^h \prod_{i=1}^d \exp \left(-\theta_l w_{*i}^{(l)2} (x_i - x'_i)^2 \right) \\ &= \sigma^2 \exp \left(\sum_{i=1}^d \sum_{l=1}^h -\theta_l w_{*i}^{(l)2} (x_i - x'_i)^2 \right) \\ &= \sigma^2 \exp \left(\sum_{i=1}^d -\eta_i (x_i - x'_i)^2 \right) \\ &= \sigma^2 \prod_{i=1}^d \exp \left(-\eta_i (x_i - x'_i)^2 \right), \end{aligned} \quad (14)$$

with $\eta_i = \sum_{l=1}^h \theta_l w_{*i}^{(l)2}$, for $i = 1, \dots, d$. Equation (14) makes it possible to express the hyperparameters' solution, provided by the KPLS kernel, from the reduced space (with h dimensions) into the

original space (with d dimensions). Thus, $\eta_i = \sum_{l=1}^h \theta_l w_{*i}^{(l)2}$, for $i = 1, \dots, d$, is used as a starting point for a gradient-based maximization of the likelihood function for a standard Kriging model. This optimization is done in the complete space where the vector $\eta = \{\eta_i\} \in \mathbb{R}^{+d}$. Such construction is similar to the method developed by Ollar *et al.* (2016), where a gradient-free optimization algorithm is used with an isotropic Kriging model followed by a gradient-based optimization starting from the solution given by the first optimization. More details of this model can be found in Bouhlel *et al.* (2016a). The two models KPLS and KPLS+K developed for high-dimensional problems are denoted by KPLS(+K) below.

2.2. The SEGOKPLS(+K) algorithms

During the last decade, the EGO algorithm (Jones, Schonlau, and Welch 1998) has become one of the most popular methods for surrogate-based optimization. It is an adaptive sampling algorithm that adds one (or more) point per cycle (Chevalier and Ginsbourger 2013). This algorithm uses the well-known expected improvement (EI) criterion based on the prediction value and the uncertainty provided by the Kriging model. In this section, this algorithm is adapted to constrained black-box optimization problems in high dimension.

2.2.1. Definition of the constrained optimization problem

The optimization problem with constraint functions is formalized as follows

$$\min_{\mathbf{x} \in B} \begin{cases} f(\mathbf{x}) \\ \text{s.t.} \\ g_k(\mathbf{x}) \leq 0, k = 1, \dots, m. \end{cases} \quad (15)$$

From this formulation, note that equality constraints are not considered in this article. The functions f, g_1, \dots, g_m are deterministic black-box functions that are computationally expensive. Moreover, assume that the derivatives of the objective function f and the constraint functions g_k ($k = 1, \dots, m$) are unavailable. These assumptions are typical in many engineering applications.

2.2.2. EGO for constrained optimization problems: SEGO

To construct a surrogate model in an optimization context is a complex task. Many classifications exist in the literature, mainly based on the type of metamodel and the method used to select the search points. For more details, Jones (2001) gives a taxonomy of surrogate-based optimization methods. In this article, the focus is on the so-called ‘two-stage’ method. The first step consists of fitting the metamodel and estimating the associated hyperparameters. The second step consists of using this metamodel instead of the true function and searching promising points following chosen infill sampling points. The details of this approach for bound constrained optimization problems are given as follows:

1. Evaluate the initial design of experiments: a set of initial points $(\mathbf{X}, \mathbf{y}) = (\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}, y^{(1)}, \dots, y^{(n)})$ is constructed. In this article, the Latin hypercube design (Jin, Chen, and Sudjianto 2005) is used. Since the proposed approach is designed to optimize time-consuming engineering design problems, only a small number of initial points n is used. In addition, set $n = d+1$ for comparison conveniences with Regis (2014).
2. Construct (or update) the metamodel involved in the optimization problem: the hyperparameters of the metamodel are fitted with the initial (or the enriched) design of experiments.
3. Optimize the infill sampling criterion.
4. Evaluate the new point $(\mathbf{x}^{(n+1)}, y^{(n+1)})$.

5. Check if the number of iterations is reached: if the number of points permitted is reached (stopping criterion), the algorithm stops, otherwise, the new point is added in the design of experiments and the algorithm returns to step 2.

The EI sampling criterion is one of the most popular infill criteria and is used to balance local and global search (Jones, Schonlau, and Welch 1998; Forrester, Sóbester, and Keane 2008). This algorithm performs well when the number of dimensions is relatively low. However, it cannot effectively deal with high-dimensional problems, this is mainly due to the uncertainty of the model, which is often high when the number of input variables is large. Indeed, the EI criterion puts more emphasis on metamodel exploration rather than on its exploitation, because it is almost impossible to fill all regions of the domain in this case. For this reason, the criterion used by Sasena (WB2) 'to locate the regional extreme' is chosen in this article (for more details, see Sasena 2002). The WB2 criterion is essentially the predicted function value added to the EI function and is, therefore, a slightly more local searching criterion. The expression of the WB2 criterion is given by

$$\text{WB2}(\mathbf{x}) = \begin{cases} -\hat{y}(\mathbf{x}) + \overbrace{(f_{\min} - \hat{y}(\mathbf{x}))\Phi\left(\frac{f_{\min} - \hat{y}(\mathbf{x})}{s(\mathbf{x})}\right) + s(\mathbf{x})\phi\left(\frac{f_{\min} - \hat{y}(\mathbf{x})}{s(\mathbf{x})}\right)}^{\text{EI}(\mathbf{x})}, & \text{if } s < 0 \\ -\hat{y}(\mathbf{x}), & \text{if } s = 0 \end{cases} \quad (16)$$

The WB2 is quite similar to the EI and needs to be maximized. The only difference is in the additional first term ($-\hat{y}(\mathbf{x})$) on the right-hand side of Equation (16). The main advantage of this criterion is that it is smoother than the EI function since it does not return to zero at the sampled points. After comparing the WB2 and the EI criteria on several analytical functions, Sasena (2002) recommends using the WB2 criterion instead of the EI criterion. In addition, the WB2 criterion exploits the surrogate model more than the EI criterion. This is more suitable for high-dimensional problems since the budget allocated is usually low and not proportional to the number of dimensions. Therefore, an

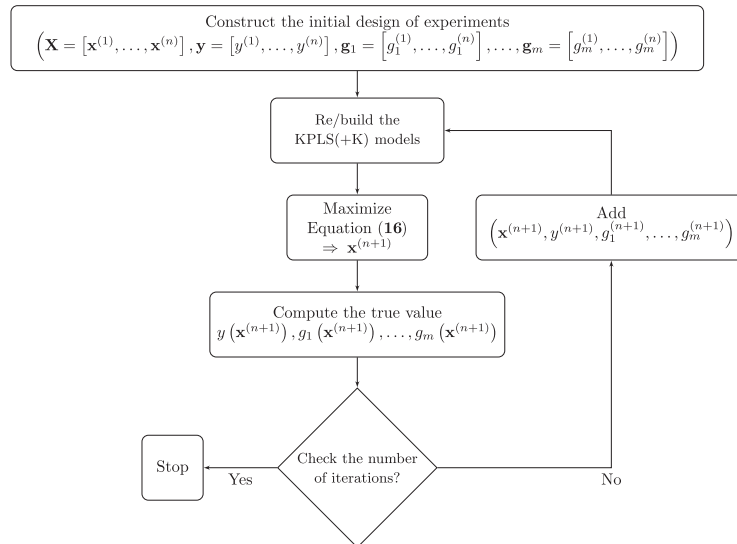


Figure 1. Steps used for the SEGOKPLS(+K) methods.

algorithm needs to rapidly converge towards a promising solution while maintaining its exploration behaviour.

To extend the algorithm for constrained optimization problems, a KPLS(+K) model for each constraint function is constructed, and the constrained infill sampling criterion is defined by

$$\min_{\mathbf{x} \in B} \begin{cases} \text{WB2}(\mathbf{x}) \\ \text{s.t.} \\ \hat{g}_k(\mathbf{x}) \leq 0, \quad k = 1, \dots, m, \end{cases} \quad (17)$$

where $\hat{g}_k(\mathbf{x})$, for $k = 1, \dots, m$ are the predictions of $g_k(\mathbf{x})$ given by Equation (15).

This approach will be called SEGOKPLS if the KPLS models are used, otherwise SEGOKPLS+K when the KPLS+K models are used. The final procedure is summarized on Figure 1:

1. Evaluate of the initial design of experiments:
 $(\mathbf{X}, \mathbf{y}, \mathbf{g}_1, \dots, \mathbf{g}_m) = (\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}, y^{(1)}, \dots, y^{(n)}, g_1^{(1)}, \dots, g_1^{(n)}, \dots, g_m^{(1)}, \dots, g_m^{(n)})$.
2. Construct or update of the KPLS(+K) models: the parameters of the metamodel are fitted with the initial or the enriched design of experiments.
3. Maximize Equation (16).
4. Evaluate the new point: the point solution is evaluated

$$\left(\mathbf{x}^{(n+1)}, y^{(n+1)}, g_1^{(n+1)}, \dots, g_m^{(n+1)} \right).$$

5. Check if the number of iterations is reached: if the number of points permitted is reached (stopping criterion), the algorithm stops, otherwise, the new point is added in the design of experiments and the algorithm returns to step 2.

3. Analytical functions

To demonstrate the performance of the SEGOKPLS algorithm, 12 well known analytical functions, defined in Table A.1 in Appendix A (included in an online supplement), are used. The SEGOKPLS algorithm is compared to several optimization methods that have been applied to these analytical functions in Regis (2014).

It should be noted that the original g_{03} and g_{05} functions have equality constraints in their definition. This article considers the same modified formulations used by Regis (2014) (noted by G3MOD and G5MOD in Regis 2014) where each constraint function is replaced by an inequality (\leq) constraint.

3.1. Alternative optimization methods

A comparison is made of SEGOKPLS with COBRA (constrained optimization by radial basis function interpolation) (Regis 2014) and ConstrLMSRBF (constrained local metric stochastic radial basis function) (Regis 2011). Both of these methods are compared with several alternative optimization methods existing in the literature by Regis (2014):

- SDPEN for sequential penalty derivative-free method for nonlinear constrained optimization (Liuzzi, Lucidi, and Sciandrone. 2010),
- NOMAD for nonlinear optimization by mesh adaptive direct search that uses DACE surrogates (NOMADm-DACE) (Abramson and Audet 2006; Audet and Jr. 2006),
- GLOBALm algorithm (Sendin, Banga, and Csendes 2009) which is a multistart clustering algorithm using two types of local solvers,

- OQNLP (Ugray *et al.* 2007), which is implemented as the GlobalSearch solver in the Matlab Global Optimization toolbox (the Mathworks 2010).

In this article, only the best results, given by Regis (2014), are used for the comparison with the SEGOKPLS algorithm as detailed in the following section.

3.2. Experimental setup

The SEGOKPLS computation is performed using the Python toolbox ‘Scikit-learn v.014’ (Pedregosa *et al.* 2011) on an Intel(R) Core(TM) i7-4500U CPU @ 1.80 Hz 2.40 GHz. The results of alternative optimization methods are transcribed from the work by Regis (2014) and they are obtained using an Intel(R) Core(TM) i7 CPU 860 2.8 GHZ. To increase the reliability of the comparison between the results of the SEGOKPLS algorithm and the alternative optimization methods, the conditions in which alternative optimization methods are performed are closely approximated. To achieve these conditions, a new Latin hypercube design of size $d+1$ is used (since the initial design of experiments used by Regis (2014) are not available but the same size is used) in which all points are infeasible. For each experiment, the SEGOKPLS algorithm is run 30 times and the computation budget is fixed at 100 objective and constraint function evaluations. These several runs reduce the influence of the initial design of experiments on the final results and on the performance of each optimization method involved in the comparison. The number of principal components used in the KPLS models for the SEGOKPLS algorithm is equal to 3. The constraint tolerance used in these tests is 10^{-5} .

To achieve the comparison, five criteria are used: the best solution, the worst solution, the median, the mean, and the standard deviation error (std) of the 30 trials computed. To facilitate the comparison of the results for each test case, only the best result of each statistic (*e.g.* the worst solution, ...) are transcribed from Table B3 in the work by Regis (2014). These results used at least one feasible point in the initial design of experiments because this is required by the alternative methods. However, the SEGOKPLS algorithm does not require an initial feasible point. In addition, searching for an initial feasible point is out of the scope of this article. Therefore, as mentioned above, the initial design of experiments was intentionally chosen to be free from feasible points, which is a common situation in real engineering design.

As mentioned in Bouhlef *et al.* (2016a), the KPLS+K model is costlier than the KPLS model; *e.g.* KPLS is over two times faster than KPLS+K for a function with 60 input variables and 300 sampling points. For the analytical functions, 30 trials are used, so SEGOKPLS is compared only to alternative methods to reduce the cost of the experimental tests. A comparison is done between SEGOKPLS and SEGOKPLS+K on the MOPTA08-12D using five runs in Section 4.1.

3.3. Results on the analytical functions

Table C.2 in Appendix C (included in an online supplement) provides statistics for both results of the SEGOKPLS algorithm (first value written in brackets) and the best results of the alternative algorithms (second value written in brackets) given by Regis (2014). The SEGOKPLS algorithm yields better results compared with the best algorithm by Regis (2014) for many test cases. In particular, SEGOKPLS is better than the alternative algorithms for all functions in terms of the worst, the median, and the mean, except for the median of functions g_{03} , g_{10} , and GTCD4. This result indicates that the SEGOKPLS algorithm is not very sensitive to the quality of the initial Latin hypercube design. Moreover, the SEGOKPLS algorithm finds a better result than alternative algorithms, in terms of the best solution, for both g_{03} and g_{07} functions. In addition, the SEGOKPLS algorithm reaches the best-known solution over all statistics used in the comparison for five functions (SR7, Hesse, g_{04} , g_{05} and g_{07}).

For all the test functions, SEGOKPLS returns the best possible optimum in the worst-case scenario of those 30 trials considered. Since the initial distribution of the sampling points plays an important role into an optimization design process that can affect the final solution, the last result (best

of worst cases) shows that SEGOKPLS' solution is less deteriorated than the alternative methods in the worst cases, which is an important characteristic in real-world engineering design optimization. Indeed, it is almost impossible to know in advance the best type and distribution of an initial design of experiments for certain engineering problems.

These results show the capability of SEGOKPLS to compete with optimization algorithms from the literature. The application of this algorithm to a more realistic problem is considered in the following section with the MOPTA08 test case.

4. MOPTA08 test problems

MOPTA08 is an automotive problem (Jones 2008) available as a Fortran code at '<http://www.miguelanjos.com/jones-benchmark>'. It is an optimization problem for minimizing the mass of a vehicle (one objective function) under 68 inequality constraints which are well normalized. One run of a real simulation of the MOPTA08 problem takes about 1–3 days. However, one simulation with the Fortran code is immediate since it is a 'good' approximation of the real version using Kriging models. Nevertheless, the type of Kriging used is not available. The MOPTA08 problem contains 124 input variables normalized to $[0, 1]$. Thus, this problem is considered large-scale in the area of a surrogate-based expensive black-box optimization.

For this test case, the SEGOKPLS+K algorithm is first applied on the complete optimization problem, *i.e.* MOPTA08 problem with 124 input variables. Despite the good performance of SEGOKPLS+K during the first optimization iterations, serious numerical problems, where the most important of them are explained in detail in Section 4.1, are encountered. Next, the SEGOKPLS+K algorithm is applied to two smaller problems: MOPTA08 with 12 and 50 input variables.

4.1. Why reduce the complexity of the MOPTA08 problem?

When applying the SEGOKPLS+K algorithm on the complete MOPTA08 problem, an initial Latin hypercube design of size $d+1$ (125 points) and 221 iterations are considered. The number of iterations is limited because there is a system crash after the 221th iteration. In order to understand the reasons of such a system crash, consider the mean evolution of both the objective function and the sum of violated constraints occurring during 10 successive iterations, which are shown in Figure 2. The objective function value decreases during the first 100 iterations as shown in Figure 2a. However, it starts to oscillate at the 165th iteration. Indeed, the SEGOKPLS+K algorithm focuses its search in a

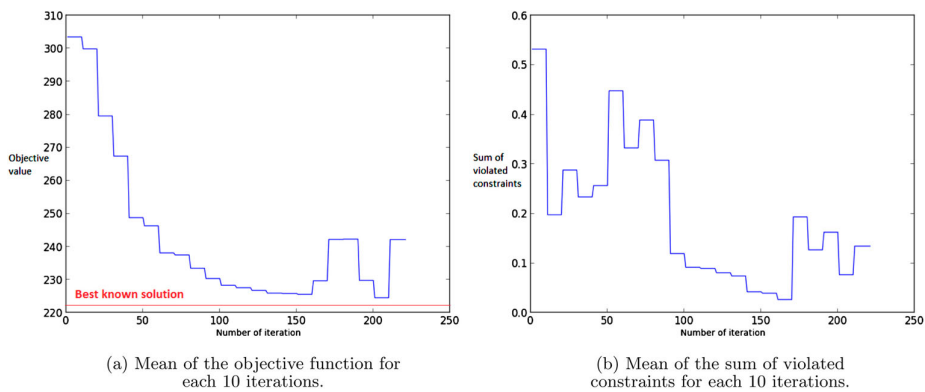


Figure 2. MOPTA08 test case with 124 input variables. Left panel: Mean of the best feasible objectives occurring for 10 successive iterations. Right panel: Mean of the sum of violated constraints for 10 successive iterations.

Table 1. MOPTA08-12D – Total number of calls (value for each of the five runs) of the black box function to reach the optimum with the SEGOKPLS and SEGOKPLS+K approaches.

MOPTA08-12D	SEGOKPLS	SEGOKPLS+K
Trial 1	53	25
Trial 2	52	34
Trial 3	52	28
Trial 4	36	29
Trial 5	38	32

local region. Moreover, it adds several similar points (*i.e.* points that are nearly collinear) to the training points. As a consequence, this set of new points leads to an ill-conditioned correlation matrix. Therefore, the algorithm crashes after a certain number of iterations, in particular when the number of dimensions d is very high (> 50). In order to reduce the complexity of the problem, the number of input variables involved in the optimization problem is reduced. To achieve this, certain input variables are fixed to inputs' values of the best-known solution. The best-known solution is found by Regis (2014) and its objective function equals to 222.22.

In this article, two reduced problems are treated: 12 and 50 input variables given in Sections 4.4.1 and 4.4.2, respectively. The indexes of the variables used during the optimization are given in Appendix D (included in an online supplement). The same ill-conditioning problem that occurs with 124 input variables are encountered beyond 50 input variables.

In addition, the SEGOKPLS algorithm is not included in the comparison. Indeed, Table 1 shows the results of SEGOKPLS and SEGOKPLS+K on the MOPTA08-12D case for five trials with a constraint tolerance of 10^{-5} . The SEGOKPLS+K algorithm outperforms SEGOKPLS, and the best-known solution of the problem is reached with less iterations using SEGOKPLS+K. Therefore, only SEGOKPLS+K is used in the following MOPTA08 study.

4.2. Alternative optimization methods

4.2.1. Description of the COBYLA algorithm

In this article, the baseline method used is the so-called COBYLA method (Powell 1994). COBYLA is a derivative-free trust region method that uses an approximating linear interpolation model of the objective and constraint functions. The algorithm is written in Fortran and iteratively finds a candidate for the optimal solution of an approximating linear programming problem. Using the original objective and constraint functions, the candidate solution is evaluated at each iteration of optimization. Afterwards, it is used to improve the approximating linear programming problem while maintaining a regular shaped simplex over iterations. When the solution can no longer be improved, the step size is reduced until it becomes sufficiently small and then the algorithm terminates.

4.2.2. Description of the COBRA algorithm

SEGOKPLS+K will be compared with an algorithm called COBRA (Regis 2014) that is designed for constrained black-box optimization when the objective and constraint functions are computationally expensive. Like SEGOKPLS+K, COBRA treats each inequality constraint individually instead of combining them into one penalty function. However, instead of using Kriging, it uses RBF surrogates to approximate the objective and constraint functions. Moreover, COBRA implements a two-phase approach where Phase I finds a feasible point while Phase II searches for a better feasible point.

In each iteration of Phase II of COBRA, RBF surrogates for the objective function and for each of the constraint functions are fit. Then, the iterate is chosen to be a minimizer of the RBF surrogate of the objective function that satisfies the RBF surrogates of the constraints within some small margin and that also satisfies a distance requirement from previous iterates. As explained in the work by Regis (2014), the margin is useful because it helps maintain feasibility of the iterates, especially on

problems with many inequality constraints. More precisely, the next sample point $\mathbf{x}^{(n+1)}$ as a solution to the optimization sub-problem:

$$\begin{aligned} & \min \hat{y}(\mathbf{x}) \\ & \text{s.t. } \mathbf{x} \in B \\ & \hat{g}_i(\mathbf{x}) + \epsilon_n^{(i)} \leq 0, \quad i = 1, 2, \dots, m \\ & \|\mathbf{x} - \mathbf{x}^{(j)}\| \geq \rho_n, \quad j = 1, \dots, n \end{aligned}$$

where $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}$ are the previous sample points, $\hat{y}, \hat{g}_1, \dots, \hat{g}_m$ are the RBF surrogates for the objective function $f(x)$ and constraint functions $g_1(x), \dots, g_m(x)$, respectively. Moreover, $\epsilon_n^{(i)} > 0$ is the margin for the inequality constraints, which vary depending on performance starting with a value of $0.005\ell(B)$, where $\ell(B)$ is the length of one side of the hypercube B . Also, ρ_n is the distance requirement from previous sample points and it is allowed to cycle to balance local and global search. That is, $\rho_n = \gamma_n \ell(B)$, where γ_n is an element of the cycle of distance requirements Ξ . Here, SEGOKPLS+K is compared with an implementation of COBRA that emphasizes local search since it performed better on the benchmark test problems used in the work by Regis (2014). In this implementation of COBRA that focuses on local search, $\Xi = (0.01, 0.001, 0.0005)$.

4.3. Experimental setup

The SEGOKPLS+K computations are performed with the Python toolbox ‘Scikit-learn v.014’ (Pedregosa *et al.* 2011) using an Intel(R) Core(TM) i7-4500U CPU @ 1.80 Hz 2.40 GHz desktop machine. The alternative optimization methods are performed in Matlab 7.11.0 using an Intel(R) Core(TM) i7 CPU 860 2.8 GHz desktop machine. This is one of the reasons why the computational cost was not investigated, in addition to the non-practical environment for getting such a consistent study. Bouhlel *et al.* (2016a) provide a comparison of CPU time between KPLS and KPLS+K using a different number of sampling points and input variables, that gives an idea about the total cost of the optimization design. The optimization of Equation (17) requires between 10 and 20 minutes per iteration.

For each set of experiments, each algorithm is run five times and each run corresponds to a different Latin hypercube design of size $d+1$. In addition, these latin hypercube designs do not include any feasible point. For the SEGOKPLS+K algorithm, a computational budget of 39 and 63 evaluations is fixed for MOPTA08 with 12 and 50 input variables, respectively. Each evaluation corresponds to one objective and constraint functions evaluation. On the other hand, a computational budget of 500 evaluations is fixed for both the COBYLA and COBRA algorithms. Then, three principal components are fixed to build the KPLS+K models for all outputs. Finally, a constraint tolerance of 10^{-5} is used as for the analytical problems.

4.4. Results of the MOPTA08 problem

4.4.1. Comparison on the MOPTA08 benchmark with 12 dimensions

The objective in this section is to optimize the MOPTA08 problem in 12 dimensions and compare the three algorithms (COBYLA, COBRA and SEGOKPLS+K), the best-known solution of this weight minimization being 222.22.

Table 2 gives the total number of calls and the objective value of the best feasible point from the five runs performed. In this previous example, only 36 (59–13) iterations of optimization are sufficient for SEGOKPLS+K to find the best-known solution of the problem for all runs. Unlike the SEGOKPLS+K algorithm, the COBYLA and COBRA algorithms require more than 100 iterations of optimization to get close to the best-known solution. However, COBRA performs reasonably well on MOPTA08-12D. In fact, the COBRA algorithm requires 293 iterations in the worst case for finding the best-known solution, whereas, COBYLA finds the best-known solution for only two runs after 411 iterations in

Table 2. MOPTA08-12D – Total number of calls, the initial design of experiments (13 points) + number of iterations, for the five runs to reach the optimum with the COBYLA, COBRA, and SEGOKPLS+K methods. The objective value of the best feasible point is given in brackets.

	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5
COBYLA	362 (222.22)	513 (222.26)	513 (223.06)	424 (222.22)	513 (236.96)
COBRA	137 (222.22)	176 (222.22)	194 (222.22)	143 (222.22)	206 (222.22)
SEGOKPLS+K	28 (222.22)	42 (222.22)	51 (222.22)	36 (222.22)	52 (222.22)

the worst case. Moreover, two runs converge to solutions with objective values 223.06 and 222.26. In addition, the COBYLA algorithm fails to find the best-known solution for one run and remains on the objective value 236.96 with a sum of violated functions equals to 6.77.

Hence, for this relatively low-dimensional case, the SEGOKPLS+K algorithm shows a very good performance and seems to be a very competitive method among optimization methods existing in the literature.

4.4.2. Comparison on the MOPTA08 benchmark with 50 dimensions

To assess the performance of the SEGOKPLS+K optimizer on a higher dimensional problem, the same test case with 50 design variables is considered. All optimization algorithms search to reach the best-known solution starting from different Latin hypercube designs of size 51.

As for the 12D test case, Table 3 gives the total number of calls and the objective value of the best feasible point from the five runs performed. In this case, 63 (114–51) iterations of optimization for SEGOKPLS+K are used whereas 500 iterations are used for both COBYLA and COBRA. In fact, four runs of SEGOKPLS+K rapidly converge near the best-known solution before the 23th iteration. Only one run remains stable on the objective function value 261.33 during 49 iterations of optimization before converging to the best-known solution, which costs more iterations before converging. This leads to fixing 63 iterations of optimization for finding the best-known solution that is 222.22. Indeed, the total number of calls to the expensive black-box function (size of the initial design of experiments + number of iterations of optimization) is equal to 114, which is greater than $2d$. To optimize a function of 50 dimensions and 68 constraint functions using that total budget is an important improvement in the area of surrogate-based optimization, especially for high-dimensional and expensive black-box optimization problems.

The COBYLA algorithm does not perform well on the MOPTA08-50D problem. In fact, the solutions found by COBYLA after the 500th iteration from all runs do not satisfy all the constraints. The sums of constraint violations of the solutions from all runs are equal to 0.007, 0.35, 3.76, 6.76, and 8.72 and the corresponding objective values are equal to 238.15, 247.42, 271.68, 286.67, and 274.74, respectively. Thus, COBYLA is not really designed for high-dimensional expensive black-box functions.

Table 3. MOPTA08-50D – Total number of calls, the initial design of experiments (51 points) + number of iterations, for the five runs to reach the optimum with the COBYLA, COBRA, and SEGOKPLS+K methods. The objective value of the best feasible point is given in brackets.

	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5
COBYLA	551 –	551 –	551 –	551 –	551 –
COBRA	551 (223.89)	449 (222.22)	551 (222.27)	551 (222.52)	551 (222.96)
SEGOKPLS+K	101 (222.22)	91 (222.22)	114 (222.22)	93 (222.22)	109 (222.22)

For all runs, COBRA reaches feasible solutions detailed as follows: after the 500th iteration, the objective values of four runs are equal to 222.52, 222.89, 222.96, and 223.27, and the remaining run finds the best-known solution after 398 iterations.

To summarize the comparisons of this test case, the optimization with the SEGOKPLS+K algorithm can reach the best-known solution for all runs. This means that SEGOKPLS+K is less sensitive to the distribution of sampling points than the alternative methods. Furthermore, SEGOKPLS+K rapidly finds the solution using very few calls to the expensive black-box functions. Therefore, SEGOKPLS+K is a very competitive optimization algorithm among existing optimization algorithms.

5. Conclusions

This article introduced the SEGOKPLS and SEGOKPLS+K algorithms adapted to high-dimensional constrained surrogate-based optimization methods using expensive black-box functions. Both combine the iterative optimization algorithm SEGO and the KPLS(+K) models. Moreover, SEGOKPLS(+K) aim to rapidly handle high-dimensional optimization problems through a good management of the budget available, which is represented by a number of calls of the expensive true functions. The key improvement of this method is the use of the KPLS(+K) models and the local infill sampling criterion WB2, which are suitable for high-dimensional problems (up to 50 dimensions). Furthermore, note that the SEGOKPLS(+K) algorithms treat each constraint individually instead of gathering them into one penalty function.

Analytical test cases and an automotive test problem are used for comparing the SEGOKPLS(+K) algorithms with many existing optimization methods. SEGOKPLS(+K) outperform the alternative methods including COBRA and COBYLA, in particular on the MOPTA-12D and MOPTA-50D problems. Moreover, SEGOKPLS+K is able to reach the best-known solution for the MOPTA-12D and MOPTA-50D problems in much less expensive calls than either COBYLA or COBRA.

An interesting direction for future work is to use classical preconditioning methods to handle problems with more than 50 dimensions, or to require SEGOKPLS(+K) to respect some conditions for adding new points into the design of experiments to avoid a bad conditioning of the covariance matrix.

ORCID

Mohamed Amine Bouhleb  <http://orcid.org/0000-0003-2182-3340>

Nathalie Bartoli  <http://orcid.org/0000-0002-6451-2203>

References

- Abramson, M. A., and C. Audet. 2006. "Convergence of Mesh Adaptive Direct Search to Second-Order Stationary Points." *SIAM Journal on Optimization* 17 (2): 606–619. doi:10.1137/050638382.
- Alberto, P. R., and F. G. González. 2012. "Partial Least Squares Regression on Symmetric Positive-Definite Matrices." *Revista Colombiana de Estadística* 36 (1): 177–192.
- Audet, C., and J.E. Jr. Dennis. 2006. "Mesh Adaptive Direct Search Algorithms for Constrained Optimization." *SIAM Journal on Optimization* 17 (1): 188–217. doi:10.1137/040603371.
- Bouhleb, M. A., N. Bartoli, J. Morlier, and A. Otsmane. 2016a. "An Improved Approach for Estimating the Hyperparameters of the Kriging Model for High-Dimensional Problems through the Partial Least Squares Method." *Mathematical Problems in Engineering* 2016: Article ID 6723410. doi:10.1155/2016/6723410
- Bouhleb, M. A., N. Bartoli, A. Otsmane, and J. Morlier. 2016b. "Improving Kriging Surrogates of High-Dimensional Design Models by Partial Least Squares Dimension Reduction." *Structural and Multidisciplinary Optimization* 53 (5): 935–952.
- Chen, X., M. Diez, M. Kandasamy, Z. Zhang, E. F. Campana, and F. Stern. 2015. "High-Fidelity Global Optimization of Shape Design by Dimensionality Reduction, Metamodels and Deterministic Particle Swarm." *Engineering Optimization* 47 (4): 473–494. doi:10.1080/0305215X.2014.895340.
- Chevalier, C., and D. Ginsbourger. 2013. *Fast Computation of the Multi-Points Expected Improvement with Applications in Batch Selection*. Berlin, Heidelberg: Springer.

- Conn, Andrew R., Katya Scheinberg, and Luis N Vicente. 2009. *Introduction to Derivative-Free Optimization*. Philadelphia, PA: Society for Industrial and Applied Mathematics.
- Cressie, N. 1988. "Spatial Prediction and Ordinary Kriging." *Mathematical Geology* 20 (4): 405–421.
- Feliot, P., J. Bect, and E. Vazquez. 2016. "A Bayesian Approach to Constrained Single- and Multi-Objective Optimization." *Journal of Global Optimization* 67 (1–2): 97–133. doi:10.1007/s10898-016-0427-3.
- Forrester, A. I. J., A. Söbester, and A. J. Keane. 2006. "Optimization with Missing Data." *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 462 (2067): 935–945.
- Forrester, A. I., A. Söbester, and A. J. Keane. 2008. *Engineering Design via Surrogate Modeling: A Practical Guide*. Hoboken, NJ: Wiley.
- Frank, I. E., and J. H. Friedman. 1993. "A Statistical View of Some Chemometrics Regression Tools." *Technometrics* 35: 109–148.
- Goovaerts, P. 1997. *Geostatistics for Natural Resources Evaluation (Applied Geostatistics)*. New York: Oxford University Press.
- Gramacy, R. B., G. A. Gray, S. LeDigabel, H. K. H. Lee, P. Ranjan, G. Wells, and S. M. Wild. 2016. "Modeling an Augmented Lagrangian for Blackbox Constrained Optimization." *Technometrics* 58 (1): 1–11. doi:10.1080/00401706.2015.1014065.
- Haftka, R. T., D. Villanueva, and A. Chaudhuri. 2016. "Parallel Surrogate-Assisted Global Optimization with Expensive Functions—A Survey." *Structural and Multidisciplinary Optimization* 54 (1): 3–13.
- Helland, I. S. 1988. "On Structure of Partial Least Squares Regression." *Communication in Statistics – Simulation and Computation* 17: 581–607.
- Hickernell, F. J., and Y. X. Yuan. 1997. "A Simple Multistart Algorithm for Global Optimization." *Operations Research Transactions* 1 (2): 1–11.
- Jin, R., W. Chen, and A. Sudjianto. 2005. "An Efficient Algorithm for Constructing Optimal Design of Computer Experiments." *Journal of Statistical Planning and Inference* 134 (1): 268–287.
- Jones, D. R. 2001. "A Taxonomy of Global Optimization Methods Based on Response Surfaces." *Journal of Global Optimization* 21 (4): 345–383.
- Jones, D. R. 2008. "Large-Scale Multi-Disciplinary Mass Optimization in the Auto-Industry." Paper presented at the *Modeling and Optimization: Theory and Application (MOPTA) Conference*, Ontario, Canada, August 20.
- Jones, D. R., M. Schonlau, and W. J. Welch. 1998. "Efficient Global Optimization of Expensive Black-Box Functions." *Journal of Global Optimization* 13 (4): 455–492.
- Kennedy, J., and R. C. Eberhart. 1995. "Particle Swarm Optimization." In *Proceedings of IEEE International Conference on Neural Networks, 1942–1948*. Piscataway, NJ: IEEE.
- Koehler, J. R., and A. B. Owen. 1996. "Computer Experiments." *Handbook of Statistics* 13: 261–308.
- Krige, D. G. 1951. "A Statistical Approach to Some Basic Mine Valuation Problems on the Witwatersrand." *Journal of the Chemical, Metallurgical and Mining Society* 52: 119–139.
- Kyriacou, S. A., V. G. Asouti, and K. C. Giannakoglou. 2014. "Efficient PCA-Driven EAs and Metamodel-Assisted EAs, with Applications in Turbomachinery." *Engineering Optimization* 46 (7): 895–911. doi:10.1080/0305215X.2013.812726.
- Laarhoven, P. J., and E. H. L. Aarts. 1987. *Simulated Annealing: Theory and Applications*. Norwell, MA, USA: Kluwer Academic Publishers.
- Li, Y., Y. Wu, J. Zhao, and L. Chen. 2016. "A Kriging-Based Constrained Global Optimization Algorithm for Expensive Black-Box Functions with Infeasible Initial Points." *Journal of Global Optimization* 67 (1–2): 343–366.
- Liem, R. P., C. A. Mader, and J. R. R. A. Martins. 2015. "Surrogate Models and Mixtures of Experts in Aerodynamic Performance Prediction for Aircraft Mission Analysis." *Aerospace Science and Technology* 43: 126–151. doi:10.1016/j.ast.2015.02.019.
- Liu, H., S. Xu, X. Chen, X. Wang, and Q. Ma. 2016. "Constrained Global Optimization via a DIRECT-Type Constraint-Handling Technique and an Adaptive Metamodeling Strategy." *Structural and Multidisciplinary Optimization* 55 (1): 155–177. doi:10.1007/s00158-016-1482-6.
- Liuzzi, G., S. Lucidi, and M. Sciandrone. 2010. "Sequential Penalty Derivative-Free Methods for Nonlinear Constrained Optimization." *SIAM Journal on Optimization* 20 (5): 2614–2635.
- Ollar, J., C. Mortished, R. Jones, J. Siens, and V. Toropov. 2016. "Gradient Based Hyper-Parameter Optimisation for Well Conditioned Kriging Metamodels." *Structural and Multidisciplinary Optimization* 55 (6): 2029–2044.
- Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, and M. Blondel, et al. 2011. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research* 12: 2825–2830.
- Picheny V., D. Ginsbourger, O. Roustant, R. T. Haftka, and N. Kim. 2010. "Adaptive Designs of Experiments for Accurate Approximation of a Target Region." *Journal of Mechanical Design* 132 (7): 071008. doi:10.1115/1.4001873.
- Powell, M. 1994. "A Direct Search Optimization Method that Models the Objective and Constraint Functions by Linear Interpolation." In *Advances in Optimization and Numerical Analysis*, edited by S. Gomez and J. P. Hennart, 51–67. Dordrecht: Springer.
- Regis, R. G. 2011. "Stochastic Radial Basis Function Algorithms for Large-Scale Optimization Involving Expensive Black-Box Objective and Constraint Functions." *Computers and Operations Research* 38 (5): 837–853.

- Regis, R. G. 2014. "Constrained Optimization by Radial Basis Function Interpolation for High-Dimensional Expensive Black-Box Problems with Infeasible Initial Points." *Engineering Optimization* 46 (2): 218–243.
- Regis, R. G., and S. M. Wild. 2017. "CONORBIT: Constrained Optimization by Radial Basis Function Interpolation in Trust Regions." *Optimization Methods and Software* 32 (3): 552–580.
- Sacks, J., W. J. Welch, W. J. Mitchell, and H. P. Wynn. 1989. "Design and Analysis of Computer Experiments." *Statistical Science* 4 (4): 409–435.
- Sasena, M., P. Papalambros, and P. Goovaerts. 2002. "Exploration of Metamodeling Sampling Criteria for Constrained Global Optimization." *Engineering optimization* 34 (3): 263–278.
- Sasena, M. J. 2002. "Flexibility and Efficiency Enhancements for Constrained Global Design Optimization with Kriging Approximations." Ph.D. diss., University of Michigan.
- Schonlau, M. 1998. "Computer Experiments and Global Optimization." Ph.D. diss., University of Waterloo, Canada.
- Sendín, J. O. H., J. R. Banga, and T. Csendes. 2009. "Extensions of a Multistart Clustering Algorithm for Constrained Global Optimization Problems." *Industrial and Engineering Chemistry Research* 6 (48): 3014–3023.
- Simon, D. 2013. *Evolutionary Optimization Algorithms*. Hoboken, NJ: Wiley.
- Singh, P., I. Couckuyt, F. Ferranti, and T. Dhaene. 2014. "A Constrained Multi-Objective Surrogate-Based Optimization Algorithm." *2014 IEEE Congress on Evolutionary Computation (CEC)* 175: 3080–3087.
- Ugray, Z., L. Lasdon, J. Plummer, F. Glover, J. Kelly, and R. Martí. 2007. "Scatter Search and Local NLP Solvers: A Multistart Framework for Global Optimization." *INFORMS J. on Computing* 19 (3): 328–340. doi:10.1287/ijoc.1060.0175.
- Watson, A. G., and R. J. Barnes. 1995. "Infill Sampling Criteria to Locate Extremes." *Mathematical Geology* 27 (5): 589–608.
- Wold, H. 1966. *Estimation of Principal Components and Related Models by Iterative Least squares*. 391–420. New York: Academic Press.

Online Supplement for: Efficient Global Optimization for High-Dimensional Constrained Problems by Using the Kriging Models Combined with the Partial Least Squares Method

Mohamed Amine Bouhlel^a, Nathalie Bartoli^b, Rommel G. Regis^c,
Abdelkader Otsmane^d, Joseph Morlier^e

^a*University of Michigan, Department of Aerospace Engineering, Ann Arbor, MI, United States; bouhlel.mohamed.amine@gmail.com*

^b*ONERA, 2 avenue Édouard Belin, 31055 Toulouse, France; nathalie.bartoli@onera.fr*

^c*Saint Joseph's University, 5600 City Avenue, Philadelphia, PA 19131; USA
rregis@sju.edu*

^d*SNECMA, Rond-point René Ravaut-Réau, 77550 Moissy-Cramayel, France;
abdelkader.otsmane@sneema.fr*

^e*Université de Toulouse, Institut Clément Ader, CNRS, ISAE-SUPAERO, 10 Avenue
Edouard Belin, 31055 Toulouse Cedex 4, France; joseph.morlier@isae.fr*

Appendix A. Definition of the analytical problems

Table A.1: Constrained optimization test problems. d is the number of input variables and m is the number of inequality constraints. The formulations and references of these functions are given in the appendix Appendix B.

Name of the function	d	m	Bound constraints	True or best known minimum
g_{02}	10	2	$[0, 10]^{10}$	-0.4
g_{03}	20	1	$[0, 1]^{20}$	-0.69
g_{04}	5	6	$[78, 102] \times [33, 45] \times [27, 45]^3$	-30655.539
g_{05}	4	5	$[0, 1200]^2 \times [-0.55, 0.55]^2$	5126.5
g_{07}	10	8	$[-10, 10]^{10}$	24.3062
g_{09}	7	4	$[-10, 10]^7$	680.6301
g_{10}	8	6	$[10^2, 10^4] \times [10^3, 10^4] \times [10, 10^3]^5$	7049.3307
WB4	4	6	$[0.125, 10] \times [0.1, 10]^3$	1.725
GTCD4	4	1	$[20, 50] \times [1, 10] \times [20, 50] \times [0.1, 60]$	2964893.85
PVD4	4	3	$[0.1]^2 \times [0, 50] \times [0, 240]$	5804.45
Hesse	6	6	$[0, 5] \times [0, 4] \times [1, 5] \times [0, 6] \times [1, 5] \times [0, 10]$	-310
SR7	7	11	$[2.6, 3.6] \times [0.7, 0.8] \times [17, 28] \times [7.3, 8.3]^2 \times [2.9, 3.9] \times [5, 5.5]$	2994.42

Appendix B. Formulation of the analytical test functions

This section describes the analytical test problems used in this study. Some of the constraint functions are modified by either dividing by a positive constant or applying a logarithmic transformation without changing the feasible space. The $plog$ transformation was introduced in [8] and it is given by

$$plog(x) = \begin{cases} \log(1+x) & \text{if } 0 \leq x \\ -\log(1-x) & \text{if } x \leq 0 \end{cases} \quad (\text{B.1})$$

The function g_{02} [6]

$$\begin{aligned}
\min \quad & - \left| \frac{\sum_{i=1}^d \cos^4(x_i) - 2 \prod_{i=1}^d \cos^2(x_i)}{\sqrt{\sum_{i=1}^d i x_i^2}} \right| \\
\text{s.t.} \quad & \\
g_1 = & \frac{p \log(-\prod_{i=1}^d x_i + 0.75)}{p \log(10^d)} \leq 0 \\
g_2 = & \frac{\sum_{i=1}^d x_i - 7.5d}{2.5d} \leq 0 \\
\text{for } i = & 1, \dots, d, \quad 0 \leq x_i \leq 10.
\end{aligned} \tag{B.2}$$

The function g_{03} [6]

$$\begin{aligned}
\min \quad & -p \log \left[(\sqrt{d})^d \prod_{i=1}^d x_i \right] \\
\text{s.t.} \quad & \\
g_1 = & \sum_{i=1}^d x_i^2 - 1 \leq 0 \\
\text{for } i = & 1, \dots, d, \quad 0 \leq x_i \leq 1.
\end{aligned} \tag{B.3}$$

The function g_{04} [6]

For $u = 85.334407 + 0.0056858x_2x_5 + 0.0006262x_1x_4 - 0.0022053x_3x_5$,
 $v = 80.51249 + 0.0071317x_2x_5 + 0.0029955x_1x_2 + 0.0021813x_3^2$, and
 $w = 9.300961 + 0.0047026x_3x_5 + 0.0012547x_1x_3 + 0.0019085x_3x_4$, we have

$$\begin{aligned}
\min \quad & 5.3578547x_3^2 + 0.8356891x_1x_5 + 37.293239x_1 - 40792.141 \\
\text{s.t.} \quad & \\
g_1 = & -u \leq 0 \\
g_2 = & u - 92 \leq 0 \\
g_3 = & -v + 90 \leq 0 \\
g_4 = & v - 110 \leq 0 \\
g_5 = & -w + 20 \leq 0 \\
g_6 = & w - 25 \leq 0 \\
78 \leq & x_1 \leq 102, \quad 33 \leq x_2 \leq 45, \quad \text{for } i = 3, 4, 5, \quad 27 \leq x_i \leq 45.
\end{aligned} \tag{B.4}$$

The function g_{05} [6]

$$\begin{aligned}
\min \quad & 3x_1 + e^{-6}x_1^3 + 2x_2 + \frac{2e^{-6}}{3}x_2^3 \\
\text{s.t.} \quad & \\
& g_1 = x_3 - x_4 - 0.55 \leq 0 \\
& g_2 = x_4 - x_3 - 0.55 \leq 0 \\
& g_3 = 1000 \sin(-x_3 - 0.25) + 1000 \sin(-x_4 - 0.25) + 894.8 - x_1 \leq 0 \\
& g_4 = 1000 \sin(x_3 - 0.25) + 1000 \sin(x_3 - x_4 - 0.25) + 894.8 - x_2 \leq 0 \\
& g_5 = 1000 \sin(x_4 - 0.25) + 1000 \sin(x_4 - x_3 - 0.25) + 1294.8 \leq 0 \\
& 0 \leq x_1, x_2 \leq 1200, \quad -0.55 \leq x_3, x_4 \leq 0.55.
\end{aligned} \tag{B.5}$$

The function g_{07} [6]

$$\begin{aligned}
\min \quad & x_1^2 + x_2^2 + x_1x_2 - 14x_1 - 16x_2 + (x_3 - 10)^2 + 4(x_4 - 5)^2 + (x_5 - 3)^2 \\
& + 2(x_6 - 1)^2 + 5x_7^2 + 7(x_8 - 11)^2 + 2(x_9 - 10)^2 + (x_{10} - 7)^2 + 45 \\
\text{s.t.} \quad & \\
& g_1 = \frac{4x_1 + 5x_2 - 3x_7 + 9x_8 - 105}{105} \leq 0 \\
& g_2 = \frac{10x_1 - 8x_2 - 17x_7 + 2x_8}{370} \leq 0 \\
& g_3 = \frac{-8x_1 + 2x_2 + 5x_9 - 2x_{10} - 12}{158} \leq 0 \\
& g_4 = \frac{3(x_1 - 2)^2 + 4(x_2 - 3)^2 + 2x_3^2 - 7x_4 - 120}{1258} \leq 0 \\
& g_5 = \frac{5x_1^2 + 8x_2 + (x_3 - 6)^2 - 2x_4 - 40}{1258} \leq 0 \\
& g_6 = \frac{0.5(x_1 - 8)^2 + 2(x_2 - 4)^2 + 3x_5^2 - x_6 - 30}{816} \leq 0 \\
& g_7 = \frac{x_1^2 + 2(x_2 - 2)^2 - 2x_1x_2 + 14x_5 - 6x_6}{834} \leq 0 \\
& g_8 = \frac{-3x_1 + 6x_2 + 12(x_9 - 8)^2 - 7x_{10}}{788} \leq 0 \\
& g_9 = \frac{-3x_1 + 6x_2 + 12(x_9 - 8)^2 - 7x_{10}}{4048} \leq 0 \\
& \text{for } i = 1, \dots, 10, \quad -10 \leq x_i \leq 10.
\end{aligned} \tag{B.6}$$

The function g_{09} [6]

$$\begin{aligned}
\min \quad & (x_1 - 10)^2 + 5(x_2 - 12)^2 + x_3^4 + 3(x_4 - 11)^2 + 10x_5^6 + 7x_6^2 + x_7^4 - 4x_6x_7 \\
& - 10x_6 - 8x_7 \\
\text{s.t.} \quad & \\
& g_1 = \frac{2x_1^2 + 3x_2^4 + x_3 + 4x_4^2 + 5x_5 - 127}{127} \leq 0 \\
& g_2 = \frac{7x_1 + 3x_2 + 10x_3^2 + x_4 - x_5 - 282}{282} \leq 0 \\
& g_3 = \frac{23x_1 + x_2^2 + 6x_6^2 - 8x_7 - 196}{196} \leq 0 \\
& g_4 = 4x_1^2 + x_2^2 - 3x_1x_2 + 2x_3^2 + 5x_6 - 11x_7 \leq 0 \\
& \text{for } i = 1, \dots, 7, \quad -10 \leq x_i \leq 10.
\end{aligned} \tag{B.7}$$

The function g_{10} [6]

$$\begin{aligned}
\min \quad & x_1 + x_2 + x_3 \\
\text{s.t.} \quad & \\
& g_1 = -1 + 0.0025(x_4 + x_6) \leq 0 \\
& g_2 = -1 + 0.0025(-x_4 + x_5 + x_7) \leq 0 \\
& g_3 = -1 + 0.01(-x_5 + x_8) \leq 0 \\
& g_4 = \text{plog}(100x_1 - x_1x_6 + 833.33252x_4 - 83333.333) \leq 0 \\
& g_5 = \text{plog}(x_2x_4 - x_2x_7 - 1250x_4 + 1250x_5) \leq 0 \\
& g_6 = \text{plog}(x_3x_5 - x_3x_8 - 2500x_5 + 1250000) \leq 0 \\
& 10^2 \leq x_1 \leq 10^4, \quad 10^3 \leq x_2, x_3 \leq 10^4, \quad \text{for } i = 4, 5, 6, 7, 8 \quad 10 \leq x_i \leq 10^3.
\end{aligned} \tag{B.8}$$

The welded beam function (WB4) [5]

For $P = 6000$, $L = 14$, $E = 30e^6$, $G = 12e^6$, $t_{max} = 13600$, $s_{max} = 30000$, $x_{max} = 10$, $d_{max} = 0.25$, $M = P(L + \frac{x_2}{2})$, $R = \sqrt{0.25(x_2^2 + (x_1 + x_3)^2)}$, $J = \sqrt{2}x_1x_2(\frac{x_2^2}{12} + 0.25(x_1 + x_3)^2)$, $Pc = \frac{4.013E}{6L^2}x_3x_4^3(1 - 0.25x_3\frac{\sqrt{E}}{L})$, $t_1 = \frac{P}{\sqrt{2}x_1x_2}$, $t_2 = M\frac{R}{J}$, $t = \sqrt{t_1^2 + t_1t_2\frac{x_2}{R} + t_2^2}$, $s = 6P\frac{L}{x_4x_3^3}$ et $d = 4P\frac{L^3}{Ex_4x_3^3}$, we have

$$\begin{aligned}
\min \quad & 1.10471x_1^2x_2 + 0.04811x_3x_4(14 + x_2) \\
\text{s.t.} \quad & \\
& g_1 = \frac{t - t_{max}}{t_{max}} \leq 0 \\
& g_2 = \frac{s - s_{max}}{s_{max}} \leq 0 \\
& g_3 = \frac{x_1 - x_4}{x_{max}} \leq 0 \\
& g_4 = \frac{0.10471x_1^2 + 0.04811x_3x_4(14 + x_2) - 5}{5} \leq 0 \\
& g_5 = \frac{d - d_{max}}{d_{max}} \leq 0 \\
& g_6 = \frac{P - Pc}{P} \leq 0 \\
& 0.125 \leq x_1 \leq 10, \quad \text{for } i = 2, 3, 4, \quad 0.1 \leq x_i \leq 10.
\end{aligned} \tag{B.9}$$

The gas transmission compressor design (GTCD) [1]

$$\begin{aligned}
\min \quad & 8.61e^5x_1^{\frac{1}{2}}x_2x_3^{-\frac{2}{3}}x_4^{-\frac{1}{2}} + 3.69e^4x_3 + 7.72e^8x_1^{-1}x_2^{0.219} - 765.43e^6x_1^{-1} \\
\text{s.t.} \quad & \\
& g_1 = x_4x_2^{-2} + x_2^{-2} - 1 \leq 0 \\
& 20 \leq x_1 \leq 50, \quad 1 \leq x_2 \leq 10, \quad 20 \leq x_3 \leq 50, \quad 0.1 \leq x_4 \leq 60.
\end{aligned} \tag{B.10}$$

The pressure vessel design function (PVD4) [2]

$$\begin{aligned}
\min \quad & 0.6224x_1x_3x_4 + 1.7781x_2x_3^2 + 3.1661x_1^2x_4 + 19.84x_1^2x_3 \\
\text{s.t.} \quad & \\
& g_1 = -x_1 + 0.0193x_3 \leq 0 \\
& g_2 = -x_2 + 0.00954x_3 \leq 0 \\
& g_3 = p \log(-\pi x_3^2 x_4 - \frac{4}{3}\pi x_3^3 + 1296000) \leq 0 \\
& 0 \leq x_1, x_2 \leq 1, \quad 0 \leq x_3 \leq 50, \quad 0 \leq x_4 \leq 240.
\end{aligned} \tag{B.11}$$

The Hesse function [4]

$$\begin{aligned}
\min \quad & -25(x_1 - 2)^2 - (x_2 - 2)^2 - (x_3 - 1)^2 - (x_4 - 4)^2 - (x_5 - 1)^2 - (x_6 - 4)^2 \\
\text{s.t.} \quad & \\
& g_1 = \frac{2-x_1-x_2}{2} \leq 0 \\
& g_2 = \frac{x_1+x_2-6}{6} \leq 0 \\
& g_3 = \frac{-x_1+x_2-2}{2} \leq 0 \\
& g_4 = \frac{x_1-3x_2-2}{2} \leq 0 \\
& g_5 = \frac{4-(x_3-3)^2-x_4}{4} \leq 0 \\
& g_6 = \frac{4-(x_5-3)^2-x_6}{4} \leq 0 \\
& 0 \leq x_1 \leq 5, \quad 0 \leq x_2 \leq 4, \quad 1 \leq x_3 \leq 5, \quad 0 \leq x_4 \leq 6, \quad 1 \leq x_5 \leq 5, \\
& 0 \leq x_6 \leq 10.
\end{aligned} \tag{B.12}$$

The SR7 function [3]

$$\text{For } A = 3.3333x_3^2 + 14.9334x_3 - 43.0934, \quad B = x_6^2 + x_7^2, \quad C = x_6^3 + x_7^3, \quad D = x_4x_6^2 + x_5x_7^2, \\
A1 = [(745\frac{x_4}{x_2x_3})^2 + 16.91e^6]^{0.5}, \quad B1 = 0.1x_6^3, \quad A2 = [(745\frac{x_5}{x_2x_3})^2 +$$

$157.5e^{6]^{0.5}}$ et $B2 = 0.1x_7^3$, we have

$$\begin{aligned}
& \min \quad 0.7854x_1x_2^2A - 1.508x_1B + 7.477C + 0.7854D \\
& \text{s.t.} \\
& g_1 = \frac{27-x_1x_2^2x_3}{27} \leq 0 \\
& g_2 = \frac{397.5-x_1x_2^2x_3^2}{397.5} \leq 0 \\
& g_3 = \frac{1.93-\frac{x_2x_6^4x_3}{x_3^4}}{1.93} \leq 0 \\
& g_4 = \frac{1.93-\frac{x_2x_7^4x_3}{x_3^5}}{1.93} \leq 0 \\
& g_5 = \frac{\frac{A_1}{B_1}-1100}{1100} \leq 0 \\
& g_6 = \frac{\frac{A_2}{B_2}-850}{850} \leq 0 \\
& g_7 = \frac{x_2x_3-40}{40} \leq 0 \\
& g_8 = \frac{5-\frac{x_1}{x_2}}{5} \leq 0 \\
& g_9 = \frac{\frac{x_1}{x_2}-12}{x_2} \leq 0 \\
& g_{10} = \frac{1.9+1.5x_6-x_4}{1.9} \leq 0 \\
& g_{11} = \frac{1.9+1.1x_7-x_5}{1.9} \leq 0 \\
& 2.6 \leq x_1 \leq 3.6, \quad 0.7 \leq x_2 \leq 0.8, \quad 17 \leq x_3 \leq 28, \quad 7.3 \leq x_4, x_5 \leq 8.3, \\
& 2.9 \leq x_6 \leq 3.9, \quad 5 \leq x_7 \leq 5.5.
\end{aligned} \tag{B.13}$$

Appendix C. Results on the analytical functions

Table C.2: Statistics on the best feasible objective value for 30 repetitions obtained by optimization algorithms after 100 evaluations. Results in brackets represent: on the left, the SEGOKPLS results, on the right, best results given by [7]. The best results are written in bold.

	Name of the function (best known value)	
	Hesse (-310)	PVD4 (5804)
best	(-310.00, -310.00) (GlobalSearch))	(5848.66, 5807.79 (GlobalSearch))
worst	(-310.00, -261.82) (COBRA-Global))	(6179.71, 7669.36) (ConstrLMSRBF))
median	(-310.00, -297.87) (COBRA-Global))	(5959.71, 6303.98) (Cobra-Local))
mean	(-310.00, -296.25) (COBRA-Global))	(5960.54, 6492.45) (Cobra-Local))
std error	$4.90e^{-5}$	75.66
	SR7 (2994.42)	GTCD4 (2964894)
best	(2994.42, 2994.42) (NOMADm-DACE))	(2977393, 2966625 (GlobalSearch))
worst	(2994.42, 2994.69) (COBRA-Local))	(3189857, 3336814) (ConstrLMSRBF))
median	(2994.42, 2994.66) (COBRA-Global))	(3078724, 3033081 (ConstrLMSRBF))
mean	(2994.42, 2994.67) (COBRA-Local))	(3084897, 3087779) (ConstrLMSRBF))
std error	$2.69e^{-4}$	55865
	WB4 (1.725)	g_{02} (-0.40)
best	(2.22, 2.22) (NOMADm-DACE))	(-0.30, -0.33 (ConstrLMSRBF))
worst	(2.88, 3.25) (ConstrLMSRBF))	(-0.19, -0.16) (GLOBALm-SOLNP))
median	(2.22, 2.56) (COBRA-Local))	(-0.23, -0.20) (ConstrLMSRBF))
mean	(2.35, 2.61) (ConstrLMSRBF))	(-0.23, -0.21) (NOMADm-DACE))
std error	$2.08e^{-6}$	0.04
	g_{03} (-0.69)	g_{04} (-30665.54)
best	(-0.69, -0.64) (COBRA-Local))	(-30665.54, -30665.54) (NOMADm-DACE))
worst	(0.00, 0.00) (COBRA-Local))	(-30665.54, -30664.58) (COBRA-Local))
median	(0.00, -0.01 (COBRA-Global))	(-30665.54, -30665.15) (COBRA-Local))
mean	(-0.15, -0.12) (COBRA-Global))	(-30665.54, -30665.07) (COBRA-Local))
std error	0.23	$5.94e^{-4}$
	g_{05} (5126.50)	g_{07} (24.30)
best	(5126.50, 5126.50) (COBRA-Local))	(24.30, 24.48) (COBRA-Local))
worst	(5126.50, 5126.53) (COBRA-Local))	(24.30, 28.60) (COBRA-Global))
median	(5126.50, 5126.50) (GlobalSearch))	(24.30, 25.28) (COBRA-Global))
mean	(5126.50, 5126.51) (COBRA-Local))	(24.30, 25.35) (COBRA-Global))
std error	$1.11e^{-5}$	$5.51e^{-6}$
	g_{09} (680.63)	g_{10} (7049.33)
best	(920.51, 702.04 (ConstrLMSRBF))	(8210.08, 7818.53 (ConstrLMSRBF))
worst	(1861.88, 3131.82) (ConstrLMSRBF))	(11285, 13965.60) (ConstrLMSRBF))
median	(1198.47, 1308.49) (ConstrLMSRBF))	(9957.84, 9494.06 (ConstrLMSRBF))
mean	(1305.15, 1413.85) (ConstrLMSRBF))	(9702.63, 10018.33) (ConstrLMSRBF))
std error	302.2	1647

Appendix D. The indexes of the retained variables for the MOPTA08 problem

The choice of the fixed variables is done in a random way. The indexes of the used variables during the optimization are:

- 12 dimensions: 2, 42, 45, 46, 49, 51, 60, 66, 72, 78, 82 and 95.
 - 50 dimensions: 2, 3, 6, 11, 13, 14, 15, 19, 21, 22, 24, 29, 30, 31, 33, 34, 36, 40, 42, 45, 46, 49, 51, 53, 54, 56, 57, 60, 62, 63, 65, 66, 67, 71, 72, 76, 77, 78, 81, 82, 85, 86, 92, 93, 95, 96, 101, 103, 121 and 124.
- [1] Beightler C, Phillips D (1976) Applied geometric programming. Wiley, URL <http://books.google.fr/books?id=6W5RAAAAMAAJ>
 - [2] Carlos AC, Efrén MM (2002) Constraint-handling in genetic algorithms through the use of dominance-based tournament selection. *Advanced Engineering Informatics* 16:2002
 - [3] Floudas C, Pardalos P (1990) A Collection of Test Problems for Constrained Global Optimization Algorithms. *Lecture Notes in Computer Science*, Springer, URL <http://books.google.fr/books?id=zRIfcNQ0V04C>
 - [4] Hesse R (1973) A heuristic search procedure for estimating a global solution of nonconvex programming problems. *Operations Research* 21:1267–1280, DOI 10.1287/opre.21.6.1267
 - [5] Kalyanmoy D (1998) An efficient constraint handling method for genetic algorithms. In: *Computer Methods in Applied Mechanics and Engineering*, pp 311–338
 - [6] Michalewicz Z, Schoenauer M (1996) Evolutionary Algorithms for Constrained Parameter Optimization Problems. *Evolutionary Computation* 4:1–32
 - [7] Regis RG (2014) Constrained optimization by radial basis function interpolation for high-dimensional expensive black-box problems with infeasible initial points. *Engineering Optimization* 46(2):218–243
 - [8] Regis RG, Shoemaker CA (2013) A quasi-multistart framework for global optimization of expensive functions using response surface models. *Journal of Global Optimization* 56(4):1719–1753

4.4.6 Adaptive modeling strategy for constrained global optimization with application to aerodynamic wing design, Aerospace Science and Technology Journal 2019



Adaptive modeling strategy for constrained global optimization with application to aerodynamic wing design



N. Bartoli^{a,*}, T. Lefebvre^a, S. Dubreuil^a, R. Olivanti^b, R. Priem^a, N. Bons^c, J.R.R.A. Martins^c, J. Morlier^d

^a ONERA/DTIS, Université de Toulouse, Toulouse, France

^b Université de Toulouse, ISAE-SUPAERO, Toulouse, France

^c University of Michigan, Ann Arbor, MI, 48109, United States

^d Université de Toulouse, ISAE-SUPAERO-INSA-Mines Albi-UPS, CNRS UMR5312, Institut Clément Ader, Toulouse, France

ARTICLE INFO

Article history:

Received 27 March 2018

Received in revised form 19 March 2019

Accepted 24 March 2019

Available online 9 April 2019

Keywords:

Surrogate modeling

Global optimization

Multimodal optimization

Mixture of experts

Aerodynamic shape optimization

Wing design

ABSTRACT

Surrogate models are often used to reduce the cost of design optimization problems that involve computationally costly models, such as computational fluid dynamics simulations. However, the number of evaluations required by surrogate models usually scales poorly with the number of design variables, and there is a need for both better constraint formulations and multimodal function handling. To address this issue, we developed a surrogate-based gradient-free optimization algorithm that can handle cases where the function evaluations are expensive, the computational budget is limited, the functions are multimodal, and the optimization problem includes nonlinear equality or inequality constraints. The proposed algorithm—super efficient global optimization coupled with mixture of experts (SEGOMOE)—can tackle complex constrained design optimization problems through the use of an enrichment strategy based on a mixture of experts coupled with adaptive surrogate models. The performance of this approach was evaluated for analytic constrained and unconstrained problems, as well as for a multimodal aerodynamic shape optimization problem with 17 design variables and an equality constraint. Our results showed that the method is efficient and that the optimum is much less dependent on the starting point than the conventional gradient-based optimization.

© 2019 Elsevier Masson SAS. All rights reserved.

1. Introduction

In aerodynamic shape optimization, realistic wing design problems require both a large number of design variables and constraints. A typical problem involves a few tens of design variables and objective and constraint functions that require significant CPU time. Because the computational budget is usually restricted, only a limited number of objective and constraint function evaluations (a few hundreds) can be performed to find the best design. The emergence of adjoint methods [1–3] was a considerable breakthrough in the field of aerodynamic shape optimization because it renders the cost of computing objective and constraint gradients independent of the number of design variables. In conjunction with gradient-based optimizers, this provides an efficient way to converge to a local minimum in a high-dimensional problem. Although airfoil and wing design optimization problems are unimodal [4,5], once the wing planform is allowed to change, the

design space becomes multimodal [6]. More generally, in many engineering design optimization problems, we do not know *a priori* whether the design space is multimodal or not and performing a multistart is often intractable in terms of CPU time. When the gradient information is not available, an interesting option is to use gradient-free optimizers known as derivative-free optimization (DFO) methods [7–10]. Stochastic DFO methods mostly fall into the category of evolutionary algorithms and are efficient to solve constrained problems [11]. One of the most promising evolutionary algorithms that can handle constraints is the covariance matrix adaptation evolution strategy (CMA-ES) [12] and its variants [13–15]. They have gained popularity in the solution of high-dimensional problems. Their only drawback is still the large number of function evaluations required to find the optimum [16]. Deterministic DFO methods, such as mesh adaptive direct search (MADS), coordinate search (CS), and generalized pattern search (GPS), have been proposed [10]. These methods have been implemented in different packages such as NOMAD [17], HOPSPACK [18], and the DFL Library [19]. Deterministic DFO methods perform local searches and therefore depend on the initial starting point. For

* Corresponding author.

E-mail address: nathalie.bartoli@onera.fr (N. Bartoli).

solving a multimodal problem, a multistart approach is generally used and the number of function evaluations increases, often with more than a thousand calls. In addition, some DFO algorithms, such as NOMAD, artificially increase the number of inequality constraints to manage equality constraints, making convergence even more difficult. Surrogate-based optimization (SBO) or Bayesian optimization (BO) approaches build an inexpensive approximation of the original function that can be used to rapidly find an approximate optimum [20–23]. In BO, an acquisition function is built from Gaussian process approximations of the objective and constraint functions [24]. For some BO frameworks [17,25], inequality constraints are an issue. Most of the cited algorithms consider equality constraints as two inequality constraints, which becomes intractable for large numbers of constraints. The augmented Lagrangian framework ALBO [26] handles mixed-constrained problems; however, it is still not suitable for solving large-scale problems in a reasonable time. Various reviews on SBO can be found in the literature [27–30].

Another BO alternative is based on sequential enrichment applied to efficient global optimization (EGO) [21] using an adaptive surrogate model. EGO uses a kriging model (also called Gaussian process in the machine learning community [31]) as a substitute for high-fidelity models, taking advantage of the prediction of the variance that is built into these models to inform the adaptive sampling [32–34]. Wessing and Preuss [16] performed a comparison between EGO and CMA-ES for multimodal unconstrained problems. They concluded that EGO is effective in finding multiple optima when the budget of function evaluations is limited. To handle constrained problems, Sasena et al. [35] proposed an extension of EGO called SuperEGO. SBO has been applied to aerodynamic shape optimization problems in previous research efforts [36,37], some of which have used adaptive sampling [38–40]. Despite the works cited above, the number of design variables that have been handled so far is insufficient for wing design optimization and multimodality with a restricted budget is still an issue.

For realistic aircraft wing shape optimization problems, the required number of design variables exceeds 200 [4], and, therefore, trying to directly solve the problems using EGO with a conventional kriging approach is not feasible. We recently proposed to use EGO with a new type of kriging model adapted to high-dimensional problems (namely, KPLS and KPLS+K) [41,42]. The resulting optimization approach, which we call SEGOKPLS(+K), was demonstrated in problems with up to 50 design variables that were solved using approximately a hundred function evaluations [43]. To handle nonlinear functions that vary significantly within a wide domain, researchers have proposed to cluster the data and construct an assembly of local surrogates, known as *mixture of experts* (MOE), which facilitate global optimization [44–48].

The first contribution of this study is the extension of the SEGOKPLS algorithm to handle highly nonlinear and non-smooth functions by using MOE with KPLS or KPLS+K models as the local experts. We have previously presented an approach combining SuperEGO and MOE (SEGOMOE), with preliminary results for analytic functions [49]. We start by constructing surrogate models for the objective and constraint functions by combining automatic clustering and best expert selection. Then, we approach the solution iteratively by balancing the exploration and exploitation phases with a new proposed criterion for the acquisition function. Analytic test cases are presented to compare SEGOMOE with other DFO algorithms such as COBYLA, NOMAD, or ALBO.

The second contribution is the comparison of the proposed approach with a gradient-based algorithm for an aerodynamic shape optimization problem. This problem is based on a benchmark developed by the AIAA Aerodynamic Design and Optimization Discussion Group (ADODG) [50]. The case that we solve is a simplified version of ADODG Case 6, which is a subsonic wing de-

sign problem with a multimodal objective. The aerodynamic model is expensive, requiring high-performance parallel computing resources to solve the Reynolds-averaged Navier–Stokes equations using computational fluid dynamics (CFD). The objective is to minimize the drag coefficient at a given lift coefficient. The design variables are the angle of attack, twist distribution, and dihedral distribution of the wing, for a total of 17 design variables. Multiple optima have been identified by using a gradient-based optimizer starting from different design points. The goal of this study was to compare the effectiveness of SEGOMOE versus an established gradient-based approach in converging to the global optimum.

We start the rest of this paper by summarizing the previously developed methods that are relevant to the present work and by introducing a new enrichment criterion (i.e., WB2S). We demonstrate the ability of the proposed approach to find the global optima for five analytic test cases that are multimodal. Then, we demonstrate the effectiveness of SEGOMOE in finding the global optimum of the aerodynamic shape optimization problem and discuss how it compares to a gradient-based method.

2. The SEGOMOE approach

In this section, we describe the proposed approach (SEGOMOE), which solves the constrained optimization problem

$$\begin{cases} \min_{\mathbf{x} \in \Omega} & y(\mathbf{x}) \\ \text{s.t.} & c_1(\mathbf{x}) \leq 0, \dots, c_m(\mathbf{x}) \leq 0, \end{cases} \quad (1)$$

where $\Omega \subset \mathbb{R}^d$ defines the design space. Equality constraints can be considered a particular case where $c_i(\mathbf{x}) = 0$.

We start this section with a description of previously developed techniques (Section 2.1) and follow with a description of the novel contributions in the present work (Section 2.2).

2.1. Background on SEGO

The approach proposed in this paper builds upon SEGO, which, in turn, is based on EGO. Therefore, we start with an overview of EGO and follow with an explanation of how constraints are handled in SEGO. We also include a description of a more local infill criterion (WB2). Finally, we describe the techniques that we use to handle high-dimensional design spaces efficiently (namely, KPLS and KPLS+K).

2.1.1. Original EGO

As mentioned in the Introduction, the proposed algorithm was inspired by EGO [21]. The main idea of the EGO approach is to assume that the unknown objective function is a realization of a Gaussian process (GP) (also known as kriging [31,51]). This requires an initial design of experiments (DOE) of n_{DOE} points $\mathbf{X} = \{\mathbf{x}^{(i)}, i = 1, \dots, n_{\text{DOE}}\}$ with $\mathbf{x}^{(i)} \in \mathbb{R}^d$, where the objective function is evaluated to obtain the samples $\mathbf{y} = \{y(\mathbf{x}^{(i)}), i = 1, \dots, n_{\text{DOE}}\}$ with $y(\mathbf{x}^{(i)}) \in \mathbb{R}$. Then, we can build a conditioned GP that approximates the objective function at any point \mathbf{x} by a Gaussian random variable $\hat{Y}(\mathbf{x})$ with a mean of

$$\hat{y}(\mathbf{x}) = \hat{\mu} + \mathbf{r}_{\mathbf{x}\mathbf{X}}^t \mathbf{R}^{-1} (\mathbf{y} - \mathbf{1}\hat{\mu}) \quad (2)$$

and a standard deviation of

$$\hat{\sigma}^2(\mathbf{x}) = \hat{\sigma}^2 (\mathbf{1} - \mathbf{r}_{\mathbf{x}\mathbf{X}}^t \mathbf{R}^{-1} \mathbf{r}_{\mathbf{x}\mathbf{X}}), \quad (3)$$

where $\mathbf{1}$ is an $n_{\text{DOE}} \times 1$ column vector of 1's, $\mathbf{r}_{\mathbf{x}\mathbf{X}} = \{k(\mathbf{x}, \mathbf{x}^{(i)}), i = 1, \dots, n_{\text{DOE}}\}$, \mathbf{R} is the covariance matrix of components $\mathbf{R}_{ij} = k(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$, and $k(\cdot, \cdot)$ is a given covariance function. In this work, we use the squared exponential covariance function

$$k(\mathbf{x}, \mathbf{x}') = \hat{\sigma}^2 \prod_{i=1}^d \exp\left(-\theta_i (\mathbf{x}_i - \mathbf{x}'_i)^2\right) \\ \forall \theta_i \in \mathbb{R}^+, \forall i \in [1, \dots, d]. \quad (4)$$

The scalar parameters $\hat{\mu}$, $\hat{\sigma}$, and θ_i , $i = 1, \dots, d$, can be found using a number of methods; here, we use likelihood maximization [22,31]. Using this Gaussian approximation, the EGO algorithm initially proposed by Jones et al. [21] iteratively adds points to the DOE to increase the accuracy of the GP. To find new points \mathbf{x}^{new} that help the overall optimization, the EGO algorithm uses the *expected improvement* (EI) criterion,

$$\text{EI}(\mathbf{x}) = \mathbb{E} \left[\max(0, y_{\min} - \hat{Y}(\mathbf{x})) \right], \quad (5)$$

where y_{\min} is the minimum value of the objective function over the DOE, i.e., $y_{\min} = \min_{i \in [1, \dots, n_{\text{DOE}}]} y(\mathbf{x}^{(i)})$. Because $\hat{Y}(\mathbf{x})$ is a Gaussian random variable defined by its mean (2) and its variance (3), the EI criterion can be written as

$$\text{EI}(\mathbf{x}) = \begin{cases} (y_{\min} - \hat{y}(\mathbf{x})) \Phi\left(\frac{y_{\min} - \hat{y}(\mathbf{x})}{\hat{s}(\mathbf{x})}\right) + \hat{s}(\mathbf{x}) \phi\left(\frac{y_{\min} - \hat{y}(\mathbf{x})}{\hat{s}(\mathbf{x})}\right), & \text{if } \hat{s} > 0 \\ 0, & \text{if } \hat{s} = 0, \end{cases} \quad (6)$$

where $\phi(\cdot)$ is the probability density function and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. This expression highlights the trade-off between *exploitation* of the Gaussian surrogate model and *exploration* of the design space. If $\Phi(\cdot)$ is large when $\hat{y}(\mathbf{x})$ is small compared to y_{\min} , then the EI criterion promotes exploitation. On the other hand, if $\phi(\cdot)$ is large when $\hat{s}(\mathbf{x})$ is large, then it promotes exploration.

Each iteration of the EGO algorithm consists of three main steps:

1. Construct a conditioned GP defined by a mean (2) and the variance (3) based on a DOE of size n_{DOE} .
2. Maximize the expected improvement (6).
3. Add a new point to the DOE (\mathbf{x}^{new}) and evaluate it (y^{new}).

This iterative process is repeated from an initial DOE until convergence. The usual stopping criterion is the maximum number of function evaluations corresponding to a computational budget.

2.1.2. Handling constraints

The original EGO algorithm outlined above was designed to minimize unconstrained functions. Because most engineering design problems are subject to constraints, it is crucial to have a sound strategy to handle design constraints. Although we can always add constraint functions to the objective as penalties, this approach is inefficient and inaccurate. To address this issue, Sasena et al. [52] proposed an approach they called super-efficient global optimization (SEGO), which can solve general nonlinearly mixed constrained problems. As in the EGO approach, the objective function y is approximated by a GP. To consider the m nonlinear constraints c_i , $i = 1, \dots, m$, we construct a surrogate model for each constraint c_i , denoted by \hat{c}_i , usually based on the same DOE as that used for the objective function. This results in the following constrained optimization problem:

$$\max_{\mathbf{x} \in \Omega_f} \text{EI}(\mathbf{x}), \quad (7)$$

where the feasible domain Ω_f is defined by the nonlinear constraints, i.e.,

$$\Omega_f := \{\mathbf{x} \in \mathbb{R}^d : \hat{c}_1(\mathbf{x}) \leq 0, \dots, \hat{c}_m(\mathbf{x}) \leq 0\}, \quad (8)$$

where equality constraints, $\hat{c}_i(\mathbf{x}) = 0$, could replace or be added to the set of inequality constraints above. At each iteration, the point \mathbf{x}^{new} that solves this optimization problem is added to the DOE and the process is repeated until the convergence. Even if a point \mathbf{x}^{new} is not feasible, evaluating the true functions adds information to the DOE.

In this procedure, the iterative construction of the various surrogate models (one for the objective function and m for the constraints) is driven only by the EI of the objective function. Therefore, if the surrogate models of the constraints are not accurate enough, the accuracy of the optimum is compromised.

In some examples, the optimal solution of Eq. (7) cannot satisfy the “true” constraint functions $c_i(\mathbf{x})$ because only the mean value of the GP $\hat{c}_i(\mathbf{x})$ is used to approximate the constraints during the optimization process. To consider the associated error estimation $\hat{s}_{c_i}^2(\mathbf{x})$, we should implement different strategies. For handling constraints in a Bayesian optimization framework considering the error estimation, various approaches are possible: probabilistic [53], expected violation [54], predictive entropy search with constraints [55], or slack variables with Lagrangian formulation to handle mixed-constraint problems [26]. Work is still in progress when it comes to addressing this challenge [56].

2.1.3. Infill criterion

The optimization problem (7) is multimodal, and it is challenging to find the global maximum without incurring a large computational cost. To improve the efficiency of this optimization, Sasena et al. [52] recommended to use the following criterion named “locating the regional extreme” proposed by Watson and Barnes [57]:

$$\text{WB2}(\mathbf{x}) = -\hat{y}(\mathbf{x}) + \text{EI}(\mathbf{x}), \quad (9)$$

where the mean value of the Gaussian surrogate is subtracted from the EI. The result is that this criterion (WB2) is less multimodal than the EI, which eases the solution of the global optimization, as demonstrated by Sasena [58].

Nonetheless, the WB2 criterion can, in certain cases, prevent the algorithm from finding the global minimum, yielding a local minimum instead. In fact, the magnitude of the term $\text{EI}(\mathbf{x})$ is expected to decrease during the iterative process, as the GP surrogate model becomes more accurate in the promising areas of the design space. Thus, after a few iterations, the WB2 criterion is only driven by the mean value of the GP surrogate model, which makes the algorithm focus only on the *exploitation* of the surrogate model rather than on the *exploration* of the design space. To address these issues, we developed a new criterion that builds on WB2, which is described in Section 2.2.1.

2.1.4. Handling a large number of design variables

One of the issues with SEGO is the scalability with the number of design variables d . To construct the GP of the objective and the constraint functions, we have to estimate the hyperparameters θ_i , $i = 1, \dots, d$ of the covariance function (4). The estimation of these hyperparameters by maximization of the likelihood function can be time-consuming, especially when the number of design variables is high ($d > 10$). This is because the likelihood function is multimodal, requiring a large number of inversions of the correlation matrix to maximize it.

A recently developed surrogate technique, kriging with partial least squares (KPLS), was proposed by Bouhrel et al. [41] to handle the large number of variables for Gaussian processes (up to 100 variables). This technique uses the partial least squares (PLS) method to reduce the number of hyperparameters and the size of the estimation problem. Bouhrel et al. [42] subsequently improved KPLS by adding a new step in the construction of the surrogate

model, which improves the accuracy for high-dimensional problems (KPLS+K). These approaches were used within the SEGO algorithm, and they were demonstrated to be efficient for problems with up to 50 design variables [43].

In this study, we use KPLS and KPLS+K to model both the objective function and the constraints. The use of the KPLS(+K) method in the SEGO algorithm is what enables us to solve nonlinearly constrained optimization problems with a large number of design variables ($d > 10$).

2.2. SEGOMOE and its improvements

In this section, we describe the improvements we made to SEGOMOE that constitute the contributions of the present study. These contributions include a new enrichment technique that improves the overall performance of the optimization, and an MOE approach that improves the accuracy of the surrogate model over a wide range in the design space.

2.2.1. New infill criterion

As discussed in Section 2.1.3, the WB2 criterion (9) improves some of the issues of EI; however, the lack of scaling between the EI term and the prediction of the model can compromise the exploration properties of the method. Therefore, we add the scaling

$$\text{WB2S}(\mathbf{x}) = s \text{EI}(\mathbf{x}) - \hat{y}(\mathbf{x}), \quad (10)$$

where s is a non-negative scale, EI is given by Eq. (9), and $\hat{y}(\mathbf{x})$ is given by Eq. (2).

This new criterion (WB2S) has two objectives:

1. Keep the exploration property of the expected improvement metric over the feasible design space Ω_f .
2. Keep the original properties of the WB2 metric by smoothing it to facilitate optimization.

The second condition is fulfilled as in the original WB2 criterion by penalizing the expected improvement with the term $-\hat{y}(\mathbf{x})$ in Eq. (10). The first condition is more difficult to satisfy and we use the following heuristic. Intuitively, this condition can be translated to

$$\arg \max_{\mathbf{x} \in \Omega_f} \text{EI}(\mathbf{x}) \approx \arg \max_{\mathbf{x} \in \Omega_f} \text{WB2S}(\mathbf{x}). \quad (11)$$

To enforce this condition, we need to at least ensure that, for $\mathbf{x}^* = \arg \max_{\mathbf{x} \in \Omega_f} \text{EI}(\mathbf{x})$, the following inequality is verified: $s \text{EI}(\mathbf{x}^*) > \hat{y}(\mathbf{x}^*)$. As a consequence, we define $s = \beta |\hat{y}(\mathbf{x}^*)| / \text{EI}(\mathbf{x}^*)$, where $\beta > 1$. This heuristic approach does not guarantee that Eq. (11) is true because this depends on the variation of $y(\mathbf{x}^*)$ over Ω_f . However, by setting a relatively large value for the parameter β , we expect this heuristic to give an acceptable approximation.

Finally, a new approximation is performed to improve the computational efficiency of the proposed criterion. Finding \mathbf{x}^* over Ω_f is a difficult task, and we prefer to redefine \mathbf{x}^* with the following approximation: $\mathbf{x}^* = \arg \max_{\mathbf{x} \in \mathbf{X}_0} \text{EI}(\mathbf{x})$, where the finite subset $\mathbf{X}_0 \subset \Omega_f$ contains the starting points used in the multistart optimization of the criterion. The following procedure is used to compute $s > 0$:

1. Compute EI for each starting point in the optimization (in a multistart approach).
2. Evaluate the prediction of the surrogate model for the point with the highest EI.
3. Compute the scale s such that

$$s = \begin{cases} \beta \frac{|\hat{y}(x_{\text{start}, \text{EI}_{\text{max}}})|}{\text{EI}(x_{\text{start}, \text{EI}_{\text{max}}})} & \text{if } \text{EI}(x_{\text{start}, \text{EI}_{\text{max}}}) \neq 0 \\ 1 & \text{if } \text{EI}(x_{\text{start}, \text{EI}_{\text{max}}}) = 0, \end{cases} \quad (12)$$

where $\beta > 1$ is a scaling factor. In our experiments, we have found that a fixed value of $\beta = 100$ works well; however, this could be adjusted during optimization if the EI values are too small compared to the $\hat{y}(\mathbf{x})$ values. Some analytic experiments will be presented in Section 3.3 to study the sensitivity analysis of the β parameter.

To illustrate the different behaviors associated with each criterion, we provide a one-dimensional example in Fig. 1. The objective function, its associated GP mean, and the GP uncertainty (with a confidence interval of 99%) are plotted in Fig. 1(a). In Fig. 1(c), the exploration phase is less obvious with the WB2 criterion compared to that with the EI or the WB2S criterion in Figs. 1(b) and 1(d), respectively. The maximum value of WB2 is located close to the minimum of the kriging prediction (see Fig. 1 at $x \approx 0.4$), which illustrates that, in that case, WB2 focuses more on the exploitation of the surrogate model rather than on the exploration. In contrast, the EI and WB2S criteria reach their maximum at $x = 0$, which allows to explore a promising area of the design space. The two advantages of WB2S are illustrated in Fig. 1(d). Similarly to the EI criterion, the WB2S criterion facilitates exploration of the three local maxima. On the other hand, the WB2 criterion is more unimodal, which prevents the optimizer from getting stuck in flat areas in the interval $x \in [0.8, 1]$, where the gradient is zero.

2.2.2. Mixture of experts

One of the main contributions of this study is the combination of MOE with EGO. The motivation for using MOE comes from industrial optimization problems for which both the objective function and the constraints might be strongly nonlinear, discontinuous, or both. In such cases, a good approximation over the whole design space by a kriging model might be inaccurate or require a large-size DOE.

To deal with the approximation of highly nonlinear functions, some researchers have proposed the MOE technique [59,60]. The key idea is to construct different local approximations (experts) for different domains in the design space. These local approximations can be tailored to deal with disparate local trends in the function, including flat regions, discontinuities, and strong nonlinearities.

We use MOE with KPLS(+K) surrogate models as the experts and implemented these methods in the Surrogate Modeling Toolbox [61]. MOE relies on the expectation-maximization algorithm for Gaussian mixture models [62]. The input space is clustered together with its output value by means of a parameter estimation of the joint distribution. A local expert (e.g., polynomial fit, radial basis functions, and kriging) is then built on each cluster, and all the local experts are then combined using the Gaussian mixture model parameters found by the expectation-maximization algorithm to obtain a global model.

In this approach, the Gaussian mixture model is used to combine the data to both partition the input space and derive the mixing proportion. To perform the clustering, we need n inputs, $\mathbf{X} = \{\mathbf{x}^{(i)}, i = 1, \dots, n\}$, and the corresponding outputs, $\mathbf{y} = \{y(\mathbf{x}^{(i)}), i = 1, \dots, n\}$. Therefore, we can only know the cluster posterior probabilities of vectors like $(\mathbf{x}^{(i)}, y(\mathbf{x}^{(i)})) \in \mathbb{R}^{d+1}$. To predict the cluster posterior probabilities of a sample knowing only its inputs, we must project each multivariate Gaussian function k in the Gaussian mixture model (trained in $d + 1$ dimensions) onto the input space, which has d dimensions.

Thus, for each cluster k , we create a multivariate Gaussian function in $(d + 1)$ -dimensional space with the covariance matrix,

$$\Gamma_k = \begin{pmatrix} \mathbf{\Gamma}_k^X & \nu_k \\ \nu_k^T & \xi_k \end{pmatrix}, \quad (13)$$

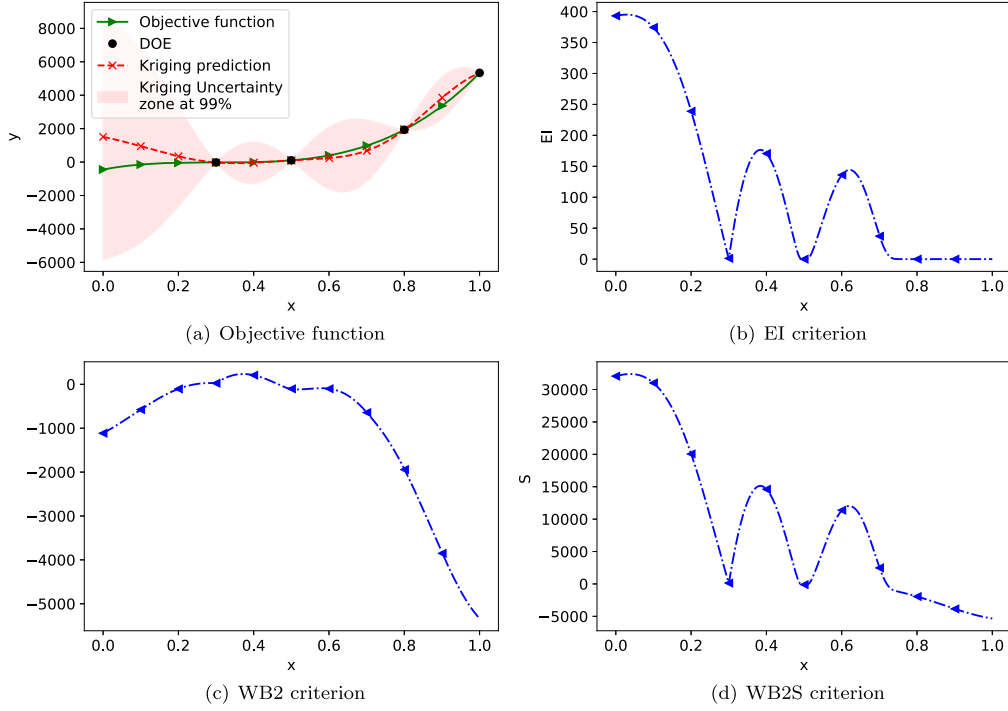


Fig. 1. Explicit infill criteria for a one-dimensional example. The objective function in green is approximated by the kriging mean (red curve) from a four-point DOE. The confidence interval of 99% is represented by the pink shaded area. The criterion (EI/WB2/WB2S) is given by the blue curve. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

where Γ_k is the covariance matrix of (\mathbf{X}, \mathbf{y}) , $\Gamma_k^X \in \mathbb{R}^d$ is the covariance matrix of \mathbf{X} , $\nu_k \in \mathbb{R}$ is $\text{Cov}(\mathbf{X}, \mathbf{y})$, and $\xi_k \in \mathbb{R}$ is $\text{Var}(\mathbf{y}, \mathbf{y})$. The mean vector of the multivariate Gaussian function is

$$\boldsymbol{\mu}_k = \begin{pmatrix} \mu_k^X \\ \mu_k^y \end{pmatrix}, \quad (14)$$

where μ_k^X represents the \mathbf{x} -coordinates of the mean $\boldsymbol{\mu}_k$ and μ_k^y is the \mathbf{y} -coordinates of the mean.

The clustering is based on the spatial location and function values (supervised approach between the inputs \mathbf{X} and the output \mathbf{y}) as in the initial version of Bettebghor et al. [45]. If available, some other clustering criteria could be used, such as a derivative value (supervised approach between the inputs \mathbf{X} and the outputs $\partial \mathbf{y} / \partial \mathbf{x}$) to obtain a better indicator for the heterogeneity in the function, as proposed by Liem et al. [60]. In our MOE toolbox, the user can provide the full or partial Jacobian as a criterion for the MOE clustering.

Thanks to hyperplane projection and linear recombination, the posterior probabilities of each cluster can be predicted and local models can be performed. When local models \hat{f}_i are known, the global model is

$$\hat{f}(\mathbf{x}) = \sum_{i=1}^K \mathbb{P}(\kappa = i | X = \mathbf{x}) \hat{f}_i(\mathbf{x}), \quad (15)$$

which is the classical probability expression for MOE. In the equation above, K is the number of Gaussian components, $\mathbb{P}(\kappa = i | X = \mathbf{x})$ is the probability that X lies in cluster i knowing that $X = \mathbf{x}$, and \hat{f}_i is the local expert built on cluster i . This probability is called the *gating network*.

2.2.3. Recombination of the infill criterion with the mixture of experts

Another key contribution of the proposed approach is the strategy used to recombine the local experts and compute the infill criterion. In MOE, the Gaussian mixture model is used to both partition the input space and derive the mixing proportion.

Based on Eq. (15), the global model can be rewritten as

$$\hat{f}(\mathbf{x}) = \sum_{i=1}^K \alpha_i \hat{f}_i(\mathbf{x}), \quad (16)$$

where K is the number of Gaussian components. This number is chosen automatically to minimize the generalization error on a validation data set [49]. The coefficients α_i are given by

$$\alpha_i = \mathbb{P}(\kappa = i | X = \mathbf{x}) \quad \forall i \in 1 \dots K$$

for the smooth recombination

or

$$\alpha_i = \begin{cases} = 1 & \text{if } i = \arg \max_{k \in [1, \dots, K]} \mathbb{P}(\kappa = k | X = \mathbf{x}) \\ = 0 & \text{otherwise} \end{cases} \quad (17)$$

for the hard recombination

Their computation leads to two different approximation models: smooth recombination and discontinuous recombination.

Smooth recombination considers all the cluster posterior probabilities when choosing the Gaussian laws to compute these quantities. The output is given by the sum of the different model outputs weighted by the cluster posterior probabilities. The resulting mixture of experts is also normally distributed, assuming the local experts are independent random variables that are normally distributed (and therefore also jointly so). The uncertainties of these models follow the Gaussian functions. The Jacobian of the surrogate model can be predicted if the analytic Jacobians of the local models are known, as in the case of kriging-based models. Thus, we can perform a gradient-based optimization, thanks to the smoothness of the surrogate model.

Discontinuous recombination—also known as hard recombination—considers the maximum of the cluster posterior probabilities; in which case, $\mathbb{P}(\kappa = i | X = \mathbf{x})$ is computed using characteristic functions for the clusters that are equal to 0 or 1. Thus, only

outputs given by a local model corresponding to the cluster are allowed. This method can be efficient for discontinuous functions. Moreover, the mixture of experts can also predict the uncertainty: for one sample, we choose the uncertainty corresponding to its cluster.

For both recombinations, the mixture of experts based on kriging models approximates functions with a heterogeneous behavior, providing a global model with a prediction of both the Jacobian and the uncertainty. The proposed algorithm automatically chooses the best type of recombination, according to a cross-validation procedure. The MOE can be written as the sum

$$\sum_{i=1}^K \alpha_i \mathcal{N}(\hat{y}_i, \hat{s}_i^2) = \mathcal{N}\left(\sum_{i=1}^K \alpha_i \hat{y}_i, \sum_{i=1}^K \alpha_i^2 \hat{s}_i^2\right) = \mathcal{N}(\hat{y}, \hat{s}^2), \quad (18)$$

where α_i is defined by Eq. (17), and $\hat{y}_i(\mathbf{x})$ and $\hat{s}_i^2(\mathbf{x})$ are given by Eqs. (2) and (3), respectively. Thus, the MOE can also predict the uncertainty. This information is useful to define the infill criterion used in the EGO or SEGO algorithm. The EI, WB2, and WB2S criteria defined by Eqs. (6), (9), and (10), respectively, are adapted for the MOE models using

$$EI_{MOE}(\mathbf{x}) = \begin{cases} (y_{\min} - \hat{y}(\mathbf{x})) \Phi\left(\frac{y_{\min} - \hat{y}(\mathbf{x})}{\hat{s}(\mathbf{x})}\right) + \hat{s}(\mathbf{x}) \phi\left(\frac{y_{\min} - \hat{y}(\mathbf{x})}{\hat{s}(\mathbf{x})}\right), & \text{if } \hat{s} > 0 \\ 0, & \text{if } \hat{s} = 0, \end{cases} \quad (19)$$

where $\phi(\cdot)$ is the probability density function and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution $\mathcal{N}(0, 1)$. The only difference between Eq. (6) and Eq. (19) is the nature of \hat{y} . Because MOE has multiple models, \hat{y} is derived from Eq. (18) and expressed as a combination of kriging-based local experts, i.e.,

$$\hat{y}(\mathbf{x}) = \sum_{i=1}^K \alpha_i \hat{y}_i(\mathbf{x}), \quad (20)$$

where $\hat{y}_i(\mathbf{x})$ is given by Eq. (2). The associated variance \hat{s}^2 is expressed in the same way, i.e.,

$$\hat{s}^2(\mathbf{x}) = \sum_{i=1}^K \alpha_i^2 \hat{s}_i^2(\mathbf{x}), \quad (21)$$

where \hat{s}_i^2 is given by Eq. (3). Because the infill criterion considers only the objective function with the knowledge of the uncertainty estimation (see Eq. (3)), we have a wider choice for the constraint surrogate models. These could be a single surrogate or a mixture of local experts. Analytic derivatives of the infill criterion (EI, WB2, or WB2S) are also available and computed for gradient-based optimization.

2.2.4. The SEGOMOE algorithm

Once recombination is achieved, MOE is used in combination with the SEGO algorithm to solve global optimization problems subject to nonlinear constraints involving a large number of design variables ($d > 10$) and potentially highly nonlinear objective and constraint functions. Figure 2 shows a flowchart of the SEGOMOE algorithm used in this study. The main steps in the algorithm are as follows:

1. Construct the initial DOE and build the associated MOE models for the objective and constraint functions.
2. Maximize the infill criterion (EI, WB2, or the new criterion defined in Section 2.2.1) subject to the design constraints and variable bounds, and propose the new enrichment point.

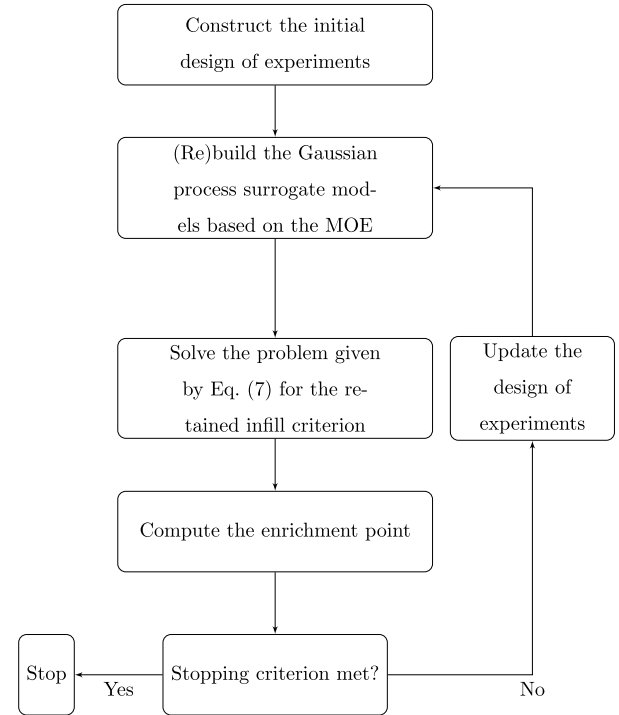


Fig. 2. Overview of the SEGOMOE algorithm.

3. Compute the values of the objective and constraint functions at the new enrichment point.
4. Check if the new enrichment point is in the feasible domain or not, and identify inactive, active, and violated constraints.
5. Return to Step 2 and update the DOE until the stopping criterion is met. A common criterion is the maximum number of function calls corresponding to the available computational budget.

To solve the constrained optimization problem (with inequality and/or equality constraints) in Step 2, we can use a gradient-based algorithm or a derivative-free optimizer. Because we use local optimizers, we perform a multistart optimization with different starting points (10 points by default). The gradient-based optimizers use the analytic Jacobian of the MOE (as described in Section 2.2.3) to compute the gradients of both the objective function and the constraints of the global optimization problem (7).

Regarding the number of clusters involved in the MOE, Bartoli et al. [49] compared the efficiency of SEGOMOE with K clusters versus one cluster on the MOPTA test case [63]. It was found that the number of function evaluations could be decreased by automatically choosing the number of clusters K (see [49] for more details on the proposed strategy) and that, in every test case studied, the use of MOE in the EGO approach never increases the number of function evaluations required for convergence. As a consequence, in the following, the SEGOMOE framework is used with the automatic cluster number approach already presented by Bartoli et al. [49]. The algorithm is implemented within the NASA OpenMDAO framework [64] and, additionally, can use surrogates available within the Surrogate Modeling Toolbox [61].

3. Analytic benchmark problems

This section presents numerical results that demonstrate the ability of SEGOMOE to solve multimodal analytic optimization problems. A more realistic application is presented in Section 4. Here, five multimodal analytic optimization problems (three unconstrained problems and two constrained problems) are con-

sidered. The three unconstrained problems are well known two-dimensional benchmark analytic functions: the six-hump camel-back function, the Michalewicz function, and the Ackley function. The first constrained problem is a two-dimensional analytic problem proposed by Parr et al. [65] that has two design variables and one inequality constraint. The solution consists of one global minimum and two local minima. The second constrained problem is a four-dimensional analytic problem proposed by [26] that has a (known) linear objective and two constraints: one inequality and one equality. Before dealing with the numerical results, we review first the performance criteria based either on the objective value at the optimal point or on the global minimum location and the choice of the β parameter for the WB2S criterion (see Section 2.2.1). We also introduce derivative-free optimizers to compare their performance with that of SEGOMOE. We choose to consider COBYLA, NOMAD, and ALBO, which are three derivative-free optimizers available as an open-source toolbox that can handle nonlinear constraints.

3.1. Performance criteria

We use three performance criteria to compare the results for SEGOMOE with those for the other optimization algorithms (COBYLA, NOMAD, and ALBO).

1. Percentage of converged runs
2. Mean number of function evaluations for converged runs
3. Standard deviation of the number of function evaluations for converged runs.

The convergence is assessed by measuring the error in the objective value or the proximity of the design variables. When the objective value is used, the relative error is

$$RE = \frac{|f_{\text{opt}}^* - f_{\text{ref}}^*|}{|f_{\text{ref}}^*|}, \quad (22)$$

where f_{opt}^* is the best point given by each algorithm and f_{ref}^* is the known reference solution given in Appendix A. When using the proximity of the design variables, we measure the difference between two solutions \mathbf{x}_1 and \mathbf{x}_2 using

$$\mu_{\text{prox}}(\mathbf{x}_1, \mathbf{x}_2) = 1 - \frac{1}{d} \sum_{i=1}^d \frac{|x_{2i} - x_{1i}|}{\text{ub}_i - \text{lb}_i}, \quad (23)$$

where ub_i and lb_i denote the upper bound and the lower bound, respectively, of the i th design variable. The distances are scaled to $[0, 1]$ to confer the same weight on the design variables.

In the test cases presented here, one of the solutions is the reference and the other one is the optimal point provided by the algorithm. To ensure that an optimization is converged, we check that $1 - \mu_{\text{prox}}(\mathbf{x}_1, \mathbf{x}_2) \leq \epsilon$, where ϵ is a given threshold value (10^{-3} by default). The type of validation and the convergence threshold are specified for each problem. It is crucial to look at this criterion because the other two are only computed for converged runs.

3.2. Derivative-free optimizers: COBYLA, NOMAD, and ALBO

To evaluate the surrogate-based strategies, we consider the other derivative-free algorithms (COBYLA, NOMAD, and ALBO) for the constrained and unconstrained analytic test cases. COBYLA (Constrained Optimization BY Linear Approximation) is a trust region optimization method that uses an approximating linear interpolation model of the objective and constraint functions [66]. NOMAD is a derivative-free algorithm based on a C++ implementation of the mesh adaptive direct search (MADS) algorithm [17],

which is designed to solve difficult blackbox optimization problems. NOMAD discretizes the design space into a mesh that refines adaptively to find better successive solutions. According to Audet et al. [67], NOMAD is intended for time-consuming blackbox simulation with a small number of variables. Another BO algorithm, ALBO [26], is used here to handle mixed-constrained problems. ALBO combines an unconstrained BO framework with the classical augmented Lagrangian (AL) framework [68]. Originally designed for the equality constraint problems [69], it has been extended to inequality constraints by means of the slack variables. The ALBO procedure is the same as the AL framework except that the minimization of the AL function is replaced by the maximization of an acquisition function. The new acquisition function is not given explicitly but only through an estimation method. We found that ALBO is not suitable for the solution of large-scale problems in a reasonable time.

We tested COBYLA and NOMAD on three well-known two-dimensional multimodal unconstrained problems and tested ALBO on a four-dimensional mixed-constrained problem to compare the convergence to the global minimum and the number of function evaluations.

Both COBYLA and NOMAD require the user to specify the maximum number of function evaluations. For COBYLA, this is set to 500 evaluations, which is much higher than the number of evaluations required to reach the convergence in the analytic test cases we solve. For NOMAD, we try different values for the maximum number of evaluations keeping in mind that our budget is limited to 1000 calls. If no stopping criterion is specified, NOMAD stops as soon as the mesh size reaches a given tolerance. Because COBYLA and NOMAD require a starting point, it may be necessary to use multistart to find the global optimum of these multimodal functions. To use the same set of starting points between the two algorithms, we generate an initial set of 100 points using Latin hypercube sampling (LHS) [70].

As explained earlier, SEGOMOE uses an LHS strategy to generate the initial DOE to build the GP approximations of the objective and the constraints. Thus, to compare the three infill criteria (EI, WB2, and WB2S), we use the same set of initial LHS-generated points for each criterion with different numbers of sampling points (5, 10, 20, and 30 points). This set of initial LHS-generated points is also used for ALBO, which, like SEGOMOE, requires an initial sample of points to build a GP approximation. The size of the sampling set is also a studied parameter.

We used the Python library Scipy [71] for COBYLA, the Python NOMAD framework [72], and DiceOptim [73] with the default settings for ALBO.

3.3. Sensitivity analysis of the WB2S infill criterion

To find the relevant values for the β parameter in the WB2S infill criterion (defined in Section 2.2.1), we performed a sensitivity analysis study for two analytic functions (Michalewicz and Ackley) defined in Appendices A.2 and A.1. The sensitivity analysis study consisted of 100 runs for 32 values of $\beta \in [10^{-6}, 10^9]$, resulting in a total of 3200 runs for each function. An initial DOE of 5 points was used for the two test cases. The percentage of solutions obtained for varying β is plotted in Fig. 3.

The success rate for the Michalewicz problem was 100% for $\beta > 1$, as shown in Fig. 3(a). This was expected because s in Eq. (12) increases with β and the WB2S criterion tends to the EI criterion when β tends to infinity, as can be seen in Eq. (10). On the other hand, for $\beta < 1$, the WB2S criterion tends to the surrogate-based criterion (SBO approach), which minimizes the kriging-based surrogate model without any exploration. Therefore, small values of β lead to poor performance in both problems.

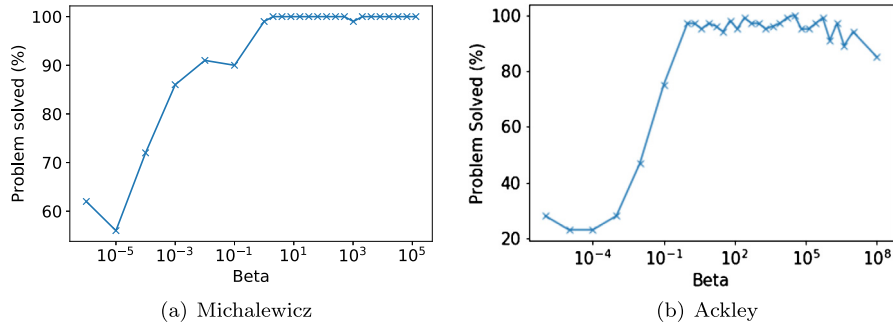


Fig. 3. Effect of the value of β in the WB2S infill criterion. One hundred DOE points (initial set with 5 points) were tested with a total of 3200 runs for each function.

For the Ackley function, the percentage of problems solved with WB2S initially increased with increasing β and then tended to decrease for $\beta > 10^5$ (see Fig. 3(b)). As a consequence, one can see that the range of β values leading to a higher success rate is quite large $\beta \in [10, 10^5]$. Thus, even if this range is problem dependent (especially the upper bound, as illustrated by the comparison of Fig. 3(a) and Fig. 3(b)), one might be confident in choosing an intermediate value for β . In the following, $\beta = 100$ is retained.

3.4. Analytic unconstrained multimodal problems

As mentioned above, the main objective of this section is to demonstrate the performance of SEGOMOE for multimodal problems. In particular, we want to compare the performance of the different infill criteria (EI, WB2, and WB2S). We first consider three analytic unconstrained multimodal functions:

- The six-hump camel-back function (defined in Appendix A.3) is a two-dimensional function with six minima—two of which are global—that is smooth and easy to optimize. The convergence is assessed based on the objective value and a threshold of 10^{-3} for the associated relative error given by Eq. (22).
- The Michalewicz function (defined in Appendix A.2) exhibits valleys and ridges with a tunable steepness. The convergence is assessed based on the objective value and a threshold of 10^{-3} for the associated relative error (22).
- The Ackley function (defined in Appendix A.1 for an arbitrary number of dimensions d) is characterized by many local minima, which makes it difficult to optimize. The value of the global minimum is independent of the number of dimensions and is located in an area of steep gradient. We consider the $d = 2$ case. The convergence is assessed based on the proximity of the design variables with the reference solution, as the relative error on the objective function value is not defined in that case ($f(x^*) = 0$). The proximity is computed with Eq. (23) and a threshold is set to 10^{-3} .

For SEGOMOE, as stated above, the same sets of DOE (one set per initial size) are used for all the criteria for a fair comparison. Universal kriging models (with squared exponential correlation functions and linear regression functions) are used as local experts to build the surrogate model of the objective $f(x)$. The iteration budget is set to 300 evaluations. Each computation can be stopped early, either because convergence has been reached or because of a singularity arising when enrichment points are too close. The optimization of the infill SEGOMOE criterion is performed using the SLSQP algorithm [74] with a multistart approach. The gradients for SLSQP are computed analytically for speed and accuracy. The convergence is assessed with a threshold of 10^{-3} for the relative error (see Eq. (22)) or the proximity index (see Eq. (23)), with an iteration budget of 300 function evaluations. The results are summarized in Table 1.

Table 1

SEGOMOE results for 3600 runs (100 per combination of DOE size, criterion, and function) for the six-hump camel-back, Michalewicz, and Ackley problems, showing the percentage of runs that converged to the analytic solution (with a relative error threshold or a proximity index of 10^{-3}). The best results are shown in bold.

Criterion	Results	DOE 5	DOE 10	DOE 20	DOE 30
Six-hump camel-back function					
WB2	Converged	94%	100%	100%	100%
	Mean (evaluations)	23	29	40	43
	σ (evaluations)	6	7	7	4
EI	Converged	100%	100%	100%	100%
	Mean (evaluations)	39	40	42	44
	σ (evaluations)	4	4	4	3
WB2S	Converged	100%	99%	100%	100%
	Mean (evaluations)	39	39	42	44
	σ (evaluations)	3	3	4	3
Michalewicz function					
WB2	Converged	71%	76%	83%	87%
	Mean (evaluations)	25	31	36	46
	σ (evaluations)	14	24	15	13
EI	Converged	100%	100%	100%	100%
	Mean (evaluations)	47	45	51	59
	σ (evaluations)	36	43	31	21
WB2S	Converged	100%	99%	100%	100%
	Mean (evaluations)	48	46	51	55
	σ (evaluations)	36	33	27	15
Ackley function					
WB2	Converged	32%	58%	90%	91%
	Mean (evaluations)	41	55	45	50
	σ (evaluations)	30	38	27	17
EI	Converged	71%	98%	98%	98%
	Mean (evaluations)	95	73	56	80
	σ (evaluations)	96	68	30	56
WB2S	Converged	97%	100%	100%	100%
	Mean (evaluations)	107	60	58	68
	σ (evaluations)	90	45	26	23

In this table, higher success rates were obtained in decreasing order by WB2S (10 times), EI (8 times), and WB2 (3 times). WB2 seems more efficient when judged on the basis of the mean and standard deviation of the evaluations. However, it is also important to consider its rate of convergence. For the Michalewicz and the Ackley function, a significant number of runs did not converge (between 70% and 40% for the initial DOEs with fewer than 20 points), although they converged with the EI and the WB2S criterion. This can be explained by the multimodality of the considered functions, as some runs performed with the WB2 criterion ended up converging near the local minima. This issue is mitigated as the size of the initial DOE is increased. The use of more points improves exploitation and reduces the likelihood of being trapped near the local optima. The EI and the WB2S criterion performed similarly across all sizes of the DOE in terms of both rate of con-

Table 2

Six-hump, Michalewicz, and Ackley results using COBYLA or NOMAD (100 runs). Different stopping criteria relative to the maximum number of blackbox evaluations were tested. The best results are shown in bold.

Algorithm	COBYLA		NOMAD	
	500	100	300	without
MAX_BB_EVAL				
Six-hump camel-back				
Converged	77%	93%	95%	95%
Mean (evaluations)	51	100	294	308
σ (evaluations)	5	0	10	25
Michalewicz				
Converged	29%	87%	89%	89%
Mean (evaluations)	60	100	290	300
σ (evaluations)	27	0	13	26
Ackley				
Converged	38%	98%	100%	100%
Mean (evaluations)	59	100	300	499
σ (evaluations)	5	0	0	12

vergence and mean number of evaluations. The standard deviation of the WB2S was slightly better for DOE sizes ranging from 10 to 30. This behavior was expected, as explained in Section 2.2.1.

The comparison results between COBYLA and NOMAD are listed in Table 2. For COBYLA, the convergence to the global minima was obtained with a significantly lower success rate compared to that for SEGOMOE. However, when COBYLA converges, it requires roughly the same number of function evaluations (between 51 and 60) as that of SEGOMOE. Table 2 shows also the NOMAD results for the three performance criteria as a function of the stopping criterion chosen (MAX_BB_EVAL for a maximum number of blackbox evaluations of 100, 300, and no limit). Thus, the number of evaluations with NOMAD was much higher than that with the SEGOMOE approach with a similar success rate.

The efficiency of SEGOMOE was validated for three multimodal unconstrained problems compared to the two other derivative-free algorithms (COBYLA and NOMAD). In terms of criteria, EI and WB2S yielded similar results in two problems, whereas WB2S performed better in the third one.

3.5. Analytic constrained multimodal problems

To evaluate the ability of SEGOMOE to tackle multimodal and constrained problems, we used the following two test cases:

- The modified Branin problem, which is a two-dimensional nonlinear problem with a single nonlinear constraint [65]. This modification, which was proposed by Parr et al. [65], adapts the original Branin function to have a single global optimum and two local ones, rather than three optima of equal value.

$$\begin{cases} \min_{\mathbf{x} \in [-5, 10] \times [0, 15]} f(\mathbf{x}) \\ g(\mathbf{x}) \geq 0 \end{cases}, \quad (24)$$

where $f(\mathbf{x})$ and $g(\mathbf{x})$ are as detailed in Appendix A.4. The feasible space for this problem consists of isolated feasible regions within their design space, which often occurs for severely constrained practical cases. The problem features three distinct feasible regions delimited by green curves in Fig. 4, making it an excellent test for constraint strategies [25, 65].

- The Linear-Ackley-Hartman (LAH) mixed-constrained problem with four input dimensions, a linear objective, and two constraints like the one described by Picheny et al. [26]

$$\begin{cases} \min_{\mathbf{x} \in [0, 1]^4} f(\mathbf{x}) \\ g(\mathbf{x}) \leq 0 \\ h(\mathbf{x}) = 0 \end{cases} \quad (25)$$

The first inequality constraint is the Ackley function in four dimensions, and the second one is an equality constraint following the Hartman four-dimensional function. The problem is denoted by LAH in the following, and details on the function expressions and the reference solution are given in Appendix A.5.

For these two constrained test cases, we compared SEGOMOE, COBYLA, and NOMAD with the ALBO framework. As the LAH problem is a mixed-constrained problem, the equality constraint $h(\mathbf{x})$ in Eq. (25) has been transformed into two inequality constraints for COBYLA and NOMAD. SEGOMOE and ALBO are both able to consider mixed-constraint problems.

In our tests, we sought to quantify the impact of the size of the initial DOE and the choice of infill criterion for SEGOMOE. As for the unconstrained case, the results were obtained by considering universal kriging models (with squared exponential correlation functions and linear regression functions) as local experts to build the surrogate models of both the objective $f(\mathbf{x})$ and the constraints $g(\mathbf{x})$ and $h(\mathbf{x})$. The iteration budget was still set to 300 evaluations. Optimization under constraints of the infill SEGOMOE criterion was still performed using the SLSQP algorithm [74] with a multistart approach. Because the initial DOE was computed using an enhanced stochastic evolutionary (ESE) algorithm [70] for building the LHS, we ran the same case 100 times to obtain statistically significant results. Furthermore, to reduce the bias that could exist between the results obtained with different infill criteria should different batches of initial DOEs be used, we kept the same initial batches for the criteria (size-wise) by means of a hot start feature implemented in SEGOMOE. The same DOEs were used to initialize the ALBO framework.

To compare the performances of the different algorithms (SEGOMOE, ALBO, COBYLA, and NOMAD), we used the criteria defined in Section 3.1. For the modified Branin problem, the convergence was assessed for each feasible point found using the relative error of the objective with respect to the known solution (see Eq. (22)), and a convergence threshold of 10^{-3} . A point is only a priori feasible when the relative violation of the constraint is less than 10^{-4} . For the LAH problem, a convergence threshold of 10^{-3} (see Eq. (23)) on the proximity index was used. The results of the optimizations are summarized in Table 3.

From Table 3, we can see that the convergence success rate increases with the size of the initial DOE for all the criteria considered, reaching nearly 100% for SEGOMOE with WB2 or WB2S when more than 20 points were used. When comparing the three criteria we used, we found that WB2S was significantly better than the other two for sparse initial DOEs (sizes of 5 and 10), for which it had a 69% success rate for modified Branin and 10% for the LAH function. This can be attributed to the fact that, in WB2S, the scaling eases exploration. For all criteria, the mean of the SEGOMOE cost increased with the size of the initial DOE; however, it is important to note that the variance decreased. The lower success rates for the sparser initial DOEs can be explained by the inaccuracy of the surrogate of the constraint, which prevented the enrichment optimizer from searching the promising areas.

If the WB2 and WB2S criteria both gave a convergence success rate close to 100% for all the DOE sizes (10, 20, and 30) with a similar number of function evaluations (mean and variance) for the LAH function, the EI criterion was less efficient on this example. Concerning the other BO algorithm, ALBO converged with a success rate of less than 45% with a large number of function evaluations (greater than 50). The ALBO results on the LAH func-

Table 3

Results for 3200 runs (100 per combination of DOE size and criterion) of the modified Branin and the LAH problems. DOEs range from 5 to 30 points, and the number of function evaluations to reach the minimum value is reported both for SEGOMOE depending on the infill criterion used (EI, WB2, or WB2S) and for ALBO. The number of converged runs to the analytic solution (with a relative error threshold or a proximity index of 10^{-3}) is given in percent. The best results are shown in bold.

Criterion	Results	DOE 5	DOE 10	DOE 20	DOE 30
Modified Branin function					
WB2	Converged	62%	78%	93%	99%
	Mean (evaluations)	38	39	53	51
	σ (evaluations)	28	22	20	9
EI	Converged	50%	75%	96%	100%
	Mean (evaluations)	27	45	49	45
	σ (evaluations)	22	29	21	8
WB2S	Converged	69%	76%	90%	100%
	Mean (evaluations)	34	39	49	47
	σ (evaluations)	20	17	19	9
ALBO	Converged	41%	32%	35%	19%
	Mean (evaluations)	58	57	70	70
	σ (evaluations)	15	13	8	7
LAH function					
WB2	Converged	100%	100%	100%	100%
	Mean (evaluations)	17	19	28	37
	σ (evaluations)	4	3	3	3
EI	Converged	96%	86%	80%	68%
	Mean (evaluations)	45	33	43	43
	σ (evaluations)	31	15	15	11
WB2S	Converged	100%	100%	100%	100%
	Mean (evaluations)	18	20	29	37
	σ (evaluations)	3	2	3	3
ALBO	Converged	5%	8%	13%	27%
	Mean (evaluations)	29	50	65	84
	σ (evaluations)	11	7	8	8

tion are quite different from those presented by Picheny et al. [26]; this could be explained by the difference in the relative violation of the constraint, i.e., 10^{-4} instead of 10^{-2} .

To visualize the enrichment process in two dimension, Fig. 4 shows the modified Branin function, where the three feasible regions are shown in green. The global optimum of the modified Branin case is at the border of the feasible region at the bottom right-hand corner of the contour plots. When the WB2S or WB2 infill criterion was used, the solution successfully converged to the global minimum. However, when the conventional EI criterion was used, the solution remained trapped in the feasible region discovered by the initial DOE. This shows that the numerical improvement of WB2S relative to EI eases the enrichment optimization, allowing a better exploration. Moreover, the increased contribution of EI within WB2S, compared to WB2, enabled the optimizer to identify the second feasible region and, therefore, to converge for this particular configuration.

As for the unconstrained test cases, comparisons were performed with the DFO algorithms (COBYLA and NOMAD), as reported in Table 4, where 100 runs (corresponding to 100 different starting points from an LHS) were performed and the same threshold criterion of 10^{-3} for the relative error or the proximity index was used. For the modified Branin problem, the convergence success rate was very low for COBYLA (12%). For NOMAD, the constraint was treated with the Progressive Barrier option [75], which needs to be satisfied only at the solution and not necessarily at the intermediate points (relaxable constraint). The convergence success rate was very low (less than 27%), and the number of function evaluations was still significant (more than 300). These values were compared with fewer than 51 evaluations for the SEGOMOE results (see Table 3). For the LAH problem, COBYLA with 53% con-

Table 4

Modified Branin and LAH problems with COBYLA or NOMAD (100 runs). Different stopping criteria relative to the maximum number of evaluations were tested. The best results are shown in bold.

Algorithms	COBYLA		NOMAD	
	500	100	300	Without
Modified Branin				
Converged	12%	20%	27%	27%
Mean (evaluations)	60	100	300	641
σ (evaluations)	6	0	0	123
LAH				
Converged	53%	0%	0%	–
Mean (evaluations)	42	–	–	–
σ (evaluations)	11	–	–	–

vergence success rate had a number of function evaluations quite comparable to the SEGOMOE-EI results. For NOMAD, the equality constraint was treated as two inequality constraints with the Progressive Barrier option [75]. The convergence success rate was 0% for the 10^{-3} threshold and was very low (2%) if we increased the threshold value to 10^{-1} (requiring 500 function evaluations in this case). In this LAH test case, a maximum number of function evaluations `MAX_BB_EVAL` is required to obtain a reasonable a CPU time. The obtained values were compared with fewer than 43 evaluations for the SEGOMOE results (see Table 3). To sum up on these two constrained test cases, SEGOMOE with the EI, WB2, or WB2S criterion was still more efficient than ALBO, NOMAD, or COBYLA in terms of both convergence success rate and number of function evaluations.

In the five analytic multimodal optimization problems considered in this section, EI and WB2S had a similar behavior in some cases, which justifies the use of a scalar multiplier $s > 1$ in the WB2S formula to promote the exploration phase. In other cases, WB2S gave better results than those of EI; especially when the size of the initial DOE was small, its convergence success rate was close to 100%. When EI found a global optimum, WB2S managed to find it also, although the reciprocity was not always satisfied.

4. Application to wing aerodynamic shape optimization

In the previous sections, we demonstrated the performance of SEGOMOE for analytic cases. We now demonstrate it on a design optimization problem that is more representative of a real-world problem: a nonlinearly constrained wing aerodynamic shape optimization problem where the objective function is multimodal. We compare SEGOMOE with a gradient-based optimizer (SNOPT) that requires a multistart approach to reach the global optimum. These comparisons are based on the performance criteria described in Section 3.1.

The final problem is a simplified version of an aerodynamic shape optimization problem defined by the ADODG, which has one equality constraint and exhibits multimodality.

4.1. High-fidelity aerodynamic shape optimization framework

To perform the aerodynamic shape optimization, we use a framework that combines a Reynolds-averaged Navier–Stokes (RANS) CFD solver, a geometry parametrization engine, and a mesh perturbation algorithm. The CFD solver is ADflow, which uses a second-order finite-volume scheme to solve the compressible Euler equations, laminar Navier–Stokes, and RANS equations (steady, unsteady, and time periodic) on overset meshes [76,77]. The Spalart–Allmaras turbulence model [78] is used to complete the RANS equations. The solver combines a Runge–Kutta method

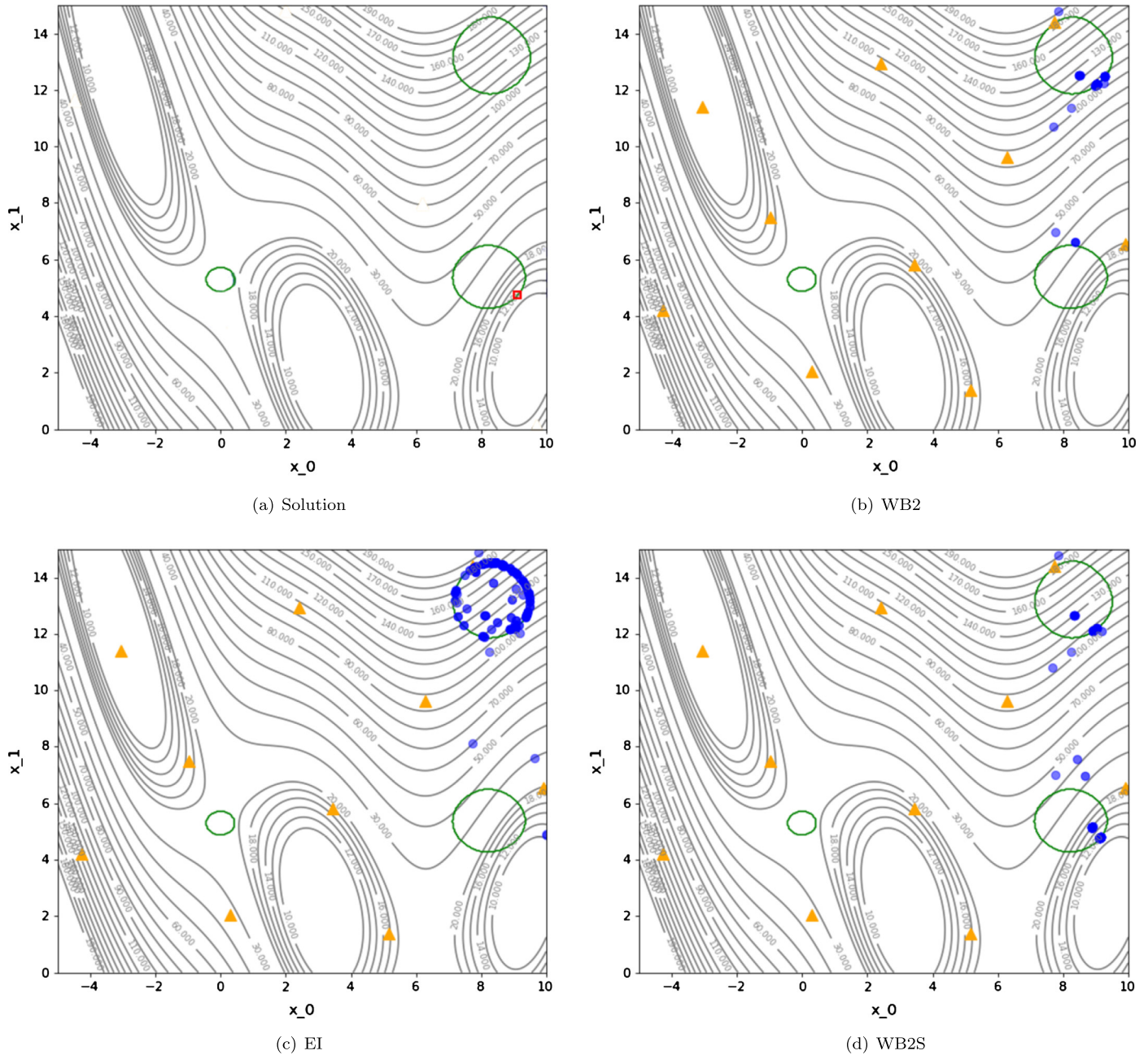


Fig. 4. Contour plots of the modified Branin function showing an example of criterion sensitivity for an initial DOE of 10 points. The color of the enrichment points (blue) fades according to how old they are in the optimization history. The feasible areas are inside the three green circles, and the optimal exact solution is shown by the red square in (a). The enrichment process with WB2, EI, and WB2S is given in (b), (c), and (d), respectively. EI is trapped in a local optimal point, whereas WB2 and WB2S find the global optimum.

for the initial iterations with a Newton–Krylov algorithm that increases the convergence success rate in the later iterations.

ADflow is especially effective when used in conjunction with gradient-based optimization because it efficiently computes accurate gradients with respect to large numbers of design variables using an adjoint method. The adjoint method is implemented using a hybrid approach that selectively uses automatic differentiation to generate the code that computes the partial derivatives in the adjoint equations [77]. The geometry is parametrized using an implementation of free-form deformation (FFD) [79], and the mesh deformation is performed with an efficient analytic inverse distance method [80].

The integration of ADflow into the optimization algorithms is achieved through pyOpt, a common Python interface that facil-

itates the use of different optimization algorithms [81,82]. The gradient-based optimizer that we use here is SNOPT [83], a sequential quadratic programming optimizer that can handle large-scale nonlinear constrained problems. This aerodynamic shape optimization framework has been used to solve various problems [4, 84–87] and has also been extended to perform aerostructural design optimization [3,88–90]. For this specific study, SEGOMOE was integrated into pyOpt, enabling us to reuse the same Python scripts when benchmarking SEGOMOE against SNOPT.

4.2. Problem definition

The problem is a simplification of the ADODG Case 6 benchmark (“Multimodal Subsonic Inviscid Optimization”) [6,50], where

Table 5
Definition of the simplified ADODG Case 6 optimization problem.

	Function/ variable	Description	Quantity	Range
Minimize	C_D	Drag coefficient	1	
with respect to	α	Angle of attack	1	$[-3.0, 6.0]$ ($^\circ$)
	θ	Twist	8	$[-3.12, 3.12]$ ($^\circ$)
	δ	Dihedral	8	$[-0.25, 0.25]$ (δ/c)
		Total variables	17	
subject to	$C_L = 0.2625$	Lift coefficient	1	
		Total constraints	1	

we reduce the number of variables. This case was devised to explore the existence of multiple local minima in aerodynamic wing design. The baseline geometry is a rectangular wing with a chord of 1.0 m and a NACA 0012 airfoil cross section with a sharp trailing edge. The semi-span is 3.06 m and the wing is fitted with a rounded wingtip cap. In the full benchmark problem, the optimizer is given freedom to change the twist, chord, dihedral, sweep, span, and sectional shape variables. In the modified version used for this study, we reduced the variables to twist and dihedral, for a total of 17 design variables. The geometry is parameterized using the FFD approach implemented in pyGeo [79], which allows the definition of global design variables with control of the sections of the B-spline control points. Nine sections are defined along the span of the wing with heavier clustering toward the wing tip. The twist variables rotate eight of these sections (excluding the root section) about the quarter-chord. Likewise, each of the eight dihedral variables controls the vertical displacement of one of the spanwise sections, excluding the root section. The angle of attack can be varied to allow the optimizer to satisfy the lift constraint. The objective of the problem is to perform a lift-constrained drag minimization of a simple wing under subsonic flow ($M_\infty = 0.5$), using the Euler equations. The main characteristics of the optimization problem are summarized in Table 5.

4.3. Gradient-based optimization results

The SEGOMOE approach is compared to SNOPT on the same test case. These gradient-based results, as well as those of other cases related to ADODG Case 6, are presented in more detail by Bons et al. [6]. Here, we just compare the SEGOMOE results for the Euler-based twist and dihedral optimization case, and summarize the corresponding results for completeness. The optimization problem is solved using SNOPT [83] starting from the 15 random shapes shown in Fig. 5(a), where we use a coarse mesh grid (L3 mesh with 180K cells). The average number of iterations required to converge is approximately 100. Nevertheless, in addition

Table 6
Results obtained with SNOPT and ADflow using the adjoint method. The global optimum is written in bold.

$C_D \times 10^4$ (drag counts)	Number of runs	Mean constraint violation
39.9091	5	7×10^{-10}
40.1971	3	1×10^{-9}
40.1972	3	1×10^{-8}
40.1975	2	6×10^{-10}
40.1986	1	8×10^{-10}
40.1995	1	5×10^{-10}

to the standard evaluation, ADflow also needs to compute the gradients for each aerodynamic-related function (drag and lift), each of which requires an adjoint solution. Therefore, for this problem, the mean evaluation budget is equivalent to approximately 300 standard ADflow calls. The optimal shapes are shown in Fig. 5(b), and the optimal results are summarized in Table 6. This problem has multiple local optima and a single global one, although the designs are very close in terms of drag value: 39.9091 drag counts for the global optimum and around 40.1980 drag counts for the local ones. Of the 15 runs, 5 converged toward the global optimum, which is characterized by an upward winglet. For the other 10 runs, multiple local minima were found for shapes with a downward winglet.

The radar chart in Fig. 6 shows the values of the 17 design variables for the various optima. We can see that all local minima differ only slightly on the twist parameters and on the dihedral parameters in the inner part of the wing. The multimodality of this problem, together with the complexity of the equality constraint, makes this a challenging proving ground for SEGOMOE.

4.4. SEGOMOE results

We performed a series of studies similar to those performed for the analytic functions in Sections 3.4 and 3.5. The runs were performed using a fixed evaluation budget of 500 evaluations, including the initial DOE and iterations. Six initial DOE sizes were considered, ranging from $1 \times d = 17$ to $6 \times d = 102$. Given the dimension of the problem, KPLS+K surrogates were selected as local experts of the MOE because we have found in previous studies that they offer the best trade-off between efficiency and accuracy [42]. The tolerance on the C_L was set to 10^{-5} , and the enrichment step was performed using the SLSQP optimizer [74] with 50 randomly picked starting points at each step.

Both the WB2 and the WB2S criterion were tested using the same batches of initial DOEs (18 DOEs per batch, computed using an optimized LHS algorithm) for each size of initial DOE, leading to a total of 216 runs. As previously mentioned, this allows us to

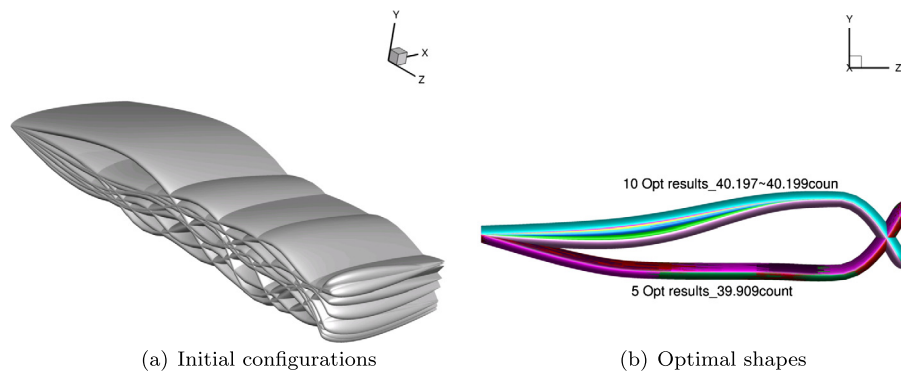


Fig. 5. SNOPT results for wing optimization with respect to the twist and dihedral variables. Multiple local minima were found when starting from 15 randomly generated initial configurations.

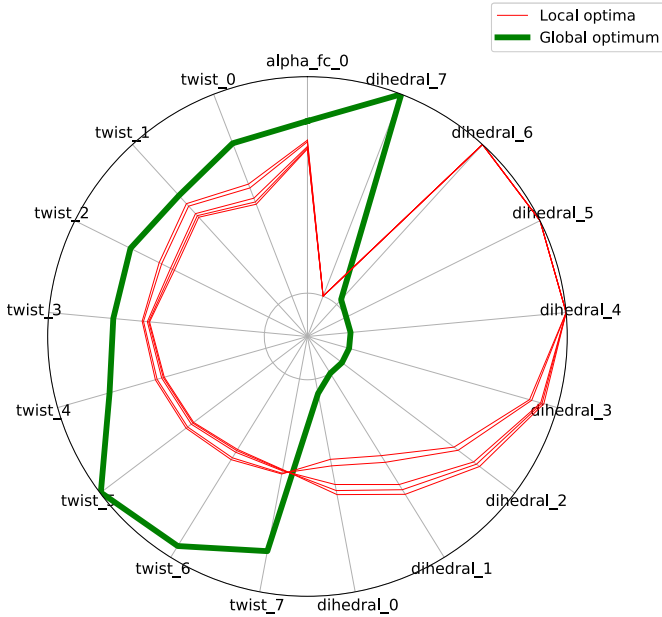


Fig. 6. Radar plot of the optima found by gradient-based optimization: SNOPT global minimum (green) and SNOPT local minima (red).

Table 7

Proximity indices between the local optima and the global one ($C_D = 39.9091$ drag counts) computed with Eq. (23). The best value in bold was obtained for the global optimum.

C_D (drag counts)	39.9091	40.1971	40.1972	40.1975	40.1986	40.1995
$\mu_{prox-glob}$	1	0.458	0.449	0.441	0.494	0.505

study the sensitivity with respect to the initial sampling without introducing bias into the results.

The best point obtained for each run is shown in a radar plot that compares the results for different initial DOE sizes and the infill criteria. Only the four most significant plots (out of 12) are shown in Figs. 7 and 8, where we compare the WB2 and WB2S criteria. The optima found by the gradient-based optimizer (SNOPT) in Section 4.3 are also shown for comparison, where the global optimum is shown in green and the five local optima are shown in red.

Because not all runs converge exactly to one of the optimum of the problem within the given evaluation budget, we use the proximity index to quantify the difference between two solutions \mathbf{x}_1 and \mathbf{x}_2 defined by Eq. (23). The proximity indices of the local min-

ima computed with the global optimum as reference ($\mu_{prox-glob}$) are listed in Table 7. When the point is the global optimum, then $\mu_{prox} = 1$.

Because two distinct optimal regions exist, proximity indices can be used to assess whether a point belongs more to the “local region” or to the “global one,” by comparing its proximity to the global and local optima.

The cost of the optimization is evaluated using the mean, maximum, and minimum number of evaluations corresponding to the best points found. Owing to the relatively small number of runs performed, the minimum and maximum are deemed more relevant than the standard deviation when analyzing the sensitivity to the initial DOE. The accuracy is assessed by computing the mean proximity index of the best points of a batch with the known global optimum ($\mu_{prox-glob}$), and the feasibility is assessed by computing the mean constraint violation. All the results are compiled in Table 8.

Both the radar plots and the statistics in Table 8 show that the WB2S criterion is more robust than the WB2 criterion in finding the “global region” for all sizes of the initial DOE within the specified budget. For instance, Figs. 7 and 8 compare the behavior of both criteria for $1 \times d$ and $4 \times d$ DOE points. We can see that, for $1 \times d$, only WB2S managed to find the global optimum shape (higher proximity index). For $4 \times d$, all WB2S optimizations converged to the global optimum area, which was not the case with WB2. We can see this by looking at the success rate in finding the global optimum (“Global” in Table 8) and the closely related $\mu_{prox-glob}$ index defined by Eq. (23).

The robustness of WB2S comes at an additional cost of 40 to 120 iterations compared to WB2 in terms of mean evaluations. The sensitivity to the initial DOE can be noticed by looking at the sharp differences between the maximum and the minimum number of evaluations for the WB2S criterion (≈ 200 – 250). It is nonetheless important to keep in mind that these results are linked to the evaluation budget as the SEGOMOE approach performs exploration without using a stopping criterion other than the budget. For instance, several runs yield their best point on the last iteration, which might be noticeably improved with few additional iterations.

The mean evaluation budget, irrespective of the infill criterion selected, is around 50% higher than that required for one gradient-based optimization with SNOPT and the adjoint method. Likewise and logically, the mean constraint violation obtained with SEGOMOE was also higher than what was achieved by the gradient-based optimization using SNOPT, although it is totally in line with the chosen feasibility tolerance (10^{-5}).

The gradient-based optimization only converged a third of a time to the global optimum. Although SEGOMOE rarely exactly

Table 8

SEGOMOE statistics for the simplified ADODG Case 6 optimization problem for different infill criteria (WB2 and WB2S) and initial DOE sizes. The best results in terms of proximity index or mean evaluations are shown in bold.

Criterion	Results	DOE 17 $= d$	DOE 34 $= 2d$	DOE 51 $= 3d$	DOE 68 $= 4d$	DOE 85 $= 5d$	DOE 102 $= 6d$
WB2	Global	61.1%	77.8%	77.8%	88.9%	83.3%	77.8%
	$\mu_{prox-glob}$	0.714	0.785	0.801	0.892	0.844	0.801
	Mean evaluations	382	426	414	409	428	368
	Maximum evaluations	500	500	500	497	498	489
	Minimum evaluations	225	273	189	270	254	248
	Mean violation	3.9×10^{-6}	4.0×10^{-6}	3.2×10^{-6}	2.97×10^{-6}	2.64×10^{-6}	3.22×10^{-6}
WB2S	Global	88.9%	94.5%	100%	100%	100%	100%
	$\mu_{prox-glob}$	0.808	0.856	0.888	0.954	0.934	0.949
	Mean evaluations	399	458	468	477	478	468
	Maximum evaluations	497	500	500	500	498	500
	Minimum evaluations	256	312	335	406	370	408
	Mean violation	3.5×10^{-6}	3.5×10^{-6}	4.5×10^{-6}	3.24×10^{-6}	2.52×10^{-6}	2.76×10^{-6}

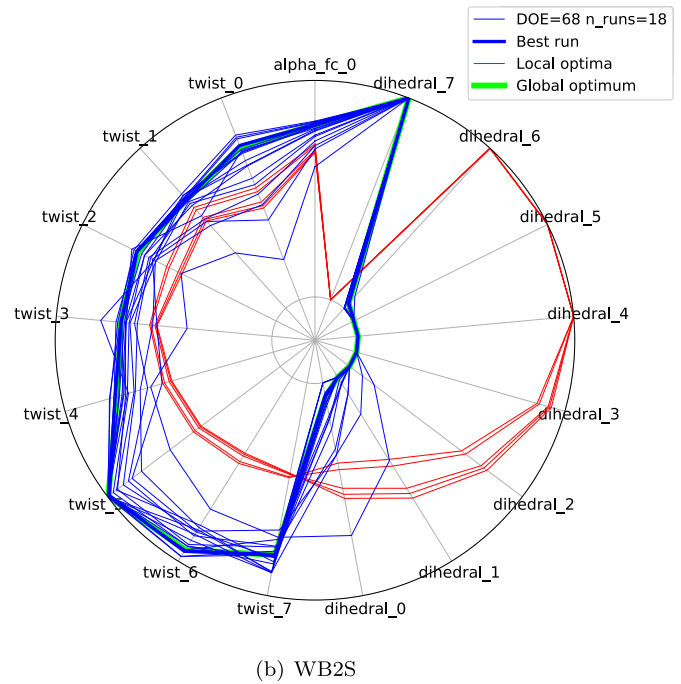
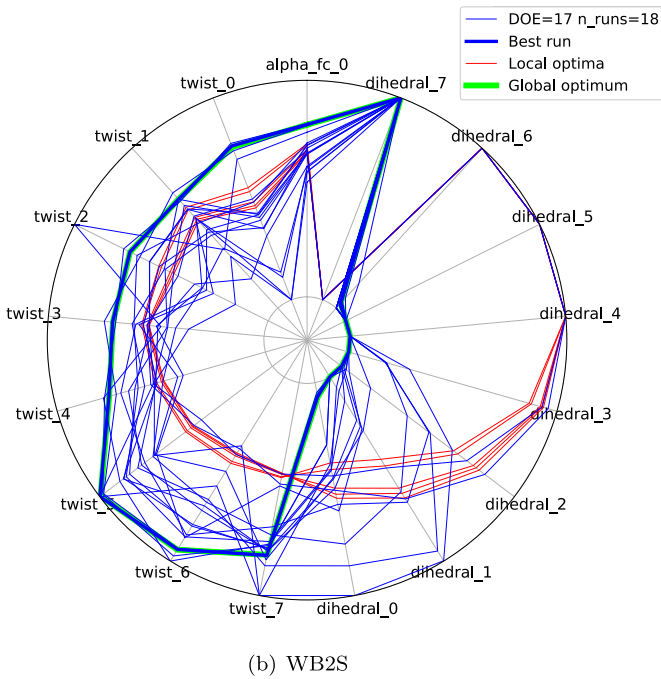
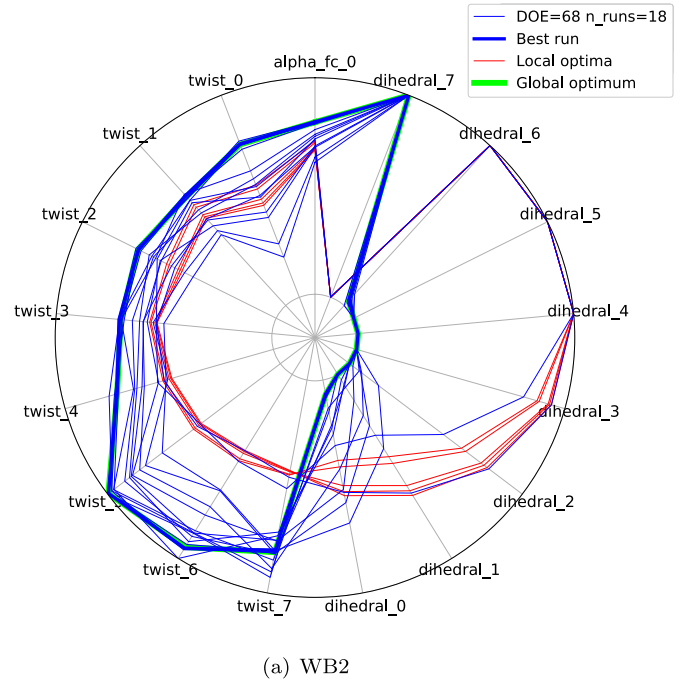
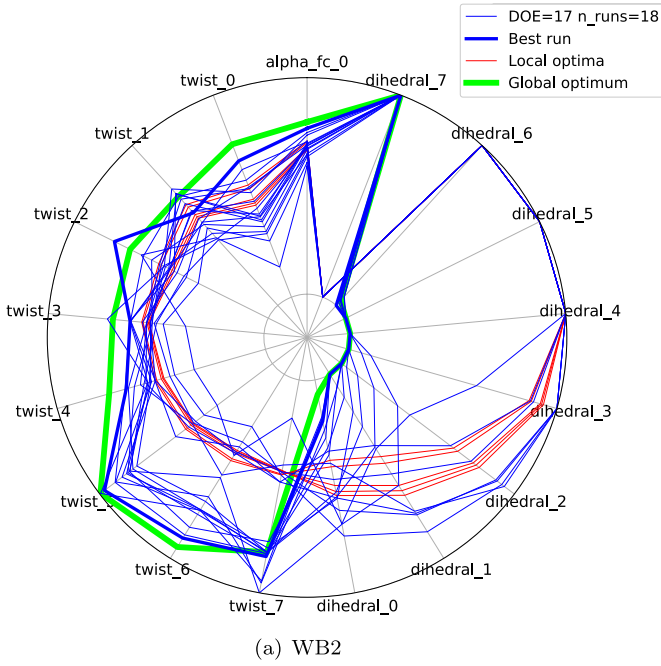


Fig. 7. Best points obtained with SEGOMOE (blue) for the initial DOE of 17 points, as well as for the SNOPT global minimum (green) and SNOPT local minima (red).

Fig. 8. Best points obtained with SEGOMOE (blue) for the initial DOE of 68 points, as well as for the SNOPT global minimum (green) and SNOPT local minima (red).

converges toward an optimum within the budget, it reliably identifies the neighborhood of the global optimum (189 out of 216 times) with acceptable statistical accuracy. This is especially true when we used the WB2S criterion (105 out of 108 times, or 97%). Table 8 shows that the best size of the initial DOE is $4 \times d$, both in terms of finding the global optimum and in terms of accuracy.

5. Conclusions

We presented a new constrained global optimization methodology, namely, super-efficient global optimization coupled with

mixture of experts (SEGOMOE), that can handle nonlinearly constrained multimodal problems at a reasonable computational cost, even when the function evaluations are costly. We built on previous algorithms by developing and implementing a new infill criterion (WB2S) and by implementing an MOE with a new recombination strategy. The use of MOE based on kriging models allows us to have a global surrogate model that approximates heterogeneous functions for the objective and the constraints.

The proposed approach can handle complex design optimization problems through the use of an adaptive surrogate modeling approach based on kriging surrogate models. The ability

of SEGOMOE to tackle multimodal optimization problems (both constrained and unconstrained) was evaluated using five analytic benchmark problems. For the first three problems, we compared SEGOMOE with two other derivative-free algorithms (NOMAD and COBYLA). The number of function evaluations was drastically reduced with the proposed approach, and the convergence success rate was higher: 99% for WB2S, compared to 89% for NOMAD and 29% for COBYLA in the worst cases. In the best cases, WB2S achieved 100%, compared to 100% for NOMAD and 77% for COBYLA. The fourth analytic problem (a modified Branin function) enabled us to obtain statistically significant results for different infill criteria and different initial DOE sizes for SEGOMOE. The results show the overall efficiency of the SEGOMOE algorithm and demonstrate the advantages of WB2S, which performs a better design space exploration, especially for sparse initial DOEs. On this constrained multimodal analytic problem, COBYLA and NOMAD had a small convergence success rate (12% and 27%, respectively) with a significant number of function evaluations. For the mixed-constrained problem considered in four dimensions, NOMAD had a poor convergence success rate (2% to be closed to the reference solution). The ALBO algorithm was compared to the different variants of SEGOMOE in this fifth analytic example and managed to converge to the reference solution with a 27% success rate and a high number of function evaluations (84 calls compared to 37 with SEGOMOE).

We also presented the results for an aerodynamic shape constrained optimization problem based on the ADODG Case 6 benchmark, which exhibits multimodality. An exhaustive study of the infill criteria with various DOE sizes was conducted, and SEGOMOE converged to the vicinity of the global minimum area under a fixed computational budget. Among the criteria, WB2S appeared more robust than WB2 for all DOE sizes, confirming its advantages. Compared to a gradient-based optimizer (SNOPT) that takes advantage of adjoint gradients, and assuming a DOE size of $4 \times d$, SEGOMOE always identified the global minimum for an average budget that is only 50% higher than that of the gradient-based one. The gradient-based approach, on the other hand, only reached the global optimum 33% of the time.

Conflict of interest statement

None declared.

Acknowledgements

This work is part of the activities of the ONERA–ISAE–ENAC joint research group. We would like to thank the following internship students for their work on the SEGOMOE optimizer: Vivien Stilz, Marie Gibert, and Valentin Fievez. We are also grateful to Rémi Lafage for his support with the SEGOMOE framework and the Adflow installation at ONERA, Régine Leconte for her support on the ISAE Cluster, and Youssef Diouane for his suggestions. This work was supported by the AGILE project (Aircraft 3rd Generation MDO for Innovative Collaboration of Heterogeneous Teams of Experts), as well as by the European Union Horizon 2020 Program (H2020-MG-2014–2015) under grant agreement number 636202. Additional support was provided by the Air Force Office of Scientific Research (AFOSR) MURI on “Managing multiple information sources of multi-physics systems,” program officer Jean–Luc Cambier, award number FA9550-15-1-0038. Joseph Morlier acknowledges the ISAE–SUPAERO foundation, which supported his visiting scholar position at the University of Michigan. This work was also partially supported by the EU project 658570–NextGen Airliners–funded by Marie Skłodowska–Curie Actions (MSCA-IF-2014-EF).

Appendix A

A.1. Ackley function

Objective

$$f(x) = -a \exp \left(-b \sqrt{\frac{1}{d} \sum_{i=1}^d x_i^2} \right) - \exp \left(\frac{1}{d} \sum_{i=1}^d \cos(cx_i) \right) + a + \exp(1)$$

Parameters (usual)

$$a = 20 \quad b = 0.2 \quad c = 2\pi$$

Search domain

$$x_i \in [-32.768, 32.768] \quad \forall i \in [1, \dots, d]$$

Global minimum

$$f(x^*) = 0 \quad x^* = [0, \dots, 0]$$

A.2. Michalewicz function

Objective

$$f(x) = - \sum_{i=1}^d \sin(x_i) \sin^{2m} \left(\frac{ix_i^2}{\pi} \right)$$

Search domain

$$x_i \in [0, \pi] \quad \forall i \in [1, \dots, d]$$

Global minimum according to the dimension d

$$d = 2 \quad f(x^*) = -1.8013 \quad x^* = [2.20, 1.57]$$

$$d = 5 \quad f(x^*) = -4.687658$$

$$d = 10 \quad f(x^*) = -9.66015$$

A.3. Six-hump camel-back function

Objective

$$f(x) = \left(4 - 2.1x_1^2 + \frac{x_1^4}{3} \right) x_1^2 + x_1 x_2 + \left(-4 + 4x_2^2 \right) x_2^2$$

Search domain

$$x_1 \in [-3, 3] \quad x_2 \in [-2, 2]$$

Global minimum

$$f(x^*) = -1.0316 \quad x^* = [0.0898, -0.7126] \quad \text{and}$$

$$x^* = [-0.0898, 0.7126]$$

A.4. Modified Branin problem

The modified Branin function [65] is a version of the Branin function normalized to $[0, 1]$ [22]. The constraint function is a normalized version of the Gomez#3 function [47], with an additional sine wave to increase multimodality:

Objective

$$f(x_1, x_2) = \left(x_2 - \frac{5.1}{4\pi^2} x_1^2 + \frac{5}{\pi} x_1 - 6 \right)^2 + 10 \left[\left(1 - \frac{1}{8\pi} \right) \cos(x_1) + 1 \right] + \frac{5x_1 + 25}{15}$$

Constraint

$$g(x_1, x_2) = \left(4 - 2.1y^2 + \frac{y^4}{3}\right)y^2 + yz + 4(y^2 - 1)z^2 + 3 \sin[6(1 - y)] + 3 \sin[6(1 - z)] - 6$$

$$\text{where } y = \frac{x_1 - 2.5}{7.5}, z = \frac{x_2 - 7.5}{7.5}$$

$$g(x_1, x_2) \geq 0$$

Search domain

$$x_1 \in [-5, 10] \quad x_2 \in [0, 15]$$

Global minimum

$$f(x^*) = 12.005$$

A.5. Linear Ackley–Hartman problem

The objective of the linear Ackley–Hartman function is a simple linear function, the inequality constraint is the Ackley function (centered) and the equality constraint is the “Hartman” function (centered, rescaled) as described by Picheny et al. [26].

Objective

$$f(x_1, x_2, x_3, x_4) = \sum_{i=1}^4 x_i$$

Inequality constraint

$$g(x_1, x_2, x_3, x_4) = 3 + 20 \exp\left(-0.2 \sqrt{\frac{1}{4} \sum_{i=1}^4 (3x_i - 1)^2}\right) + \exp\left(\frac{1}{4} \sum_{i=1}^4 \cos(2\pi(3x_i - 1))\right) - 20 - \exp(1)$$

$$g(x_1, x_2, x_3, x_4) \leq 0$$

Equality constraint

$$h(x_1, x_2, x_3, x_4) = \frac{1}{0.8387} \left[-1.1 + \sum_{i=1}^4 C_i \exp\left(-\sum_{j=1}^4 a_{ji}(x_j - p_{ji})^2\right) \right]$$

with

$$\mathbf{C} = \begin{bmatrix} 1.0 \\ 1.2 \\ 3.0 \\ 3.2 \end{bmatrix}, \quad \mathbf{a} = \begin{bmatrix} 10.0 & 0.05 & 3.0 & 17.0 \\ 3.0 & 10.0 & 3.5 & 8.0 \\ 17.0 & 17.0 & 1.70 & 0.05 \\ 3.5 & 0.1 & 10.0 & 10.0 \end{bmatrix},$$

$$\mathbf{p} = \begin{bmatrix} 0.131 & 0.232 & 0.234 & 0.404 \\ 0.169 & 0.413 & 0.145 & 0.882 \\ 0.556 & 0.830 & 0.352 & 0.873 \\ 0.012 & 0.373 & 0.288 & 0.574 \end{bmatrix}.$$

$$h(x_1, x_2, x_3, x_4) = 0$$

Search domain

$$x_i \in [0, 1] \quad \forall i \in [1, \dots, 4]$$

Global minimum

$$f(x^*) = 0.0516605 \quad \text{and} \quad x^* = [0.0, 0.0, 0.0, 0.0516605]$$

References

- [1] C.A. Mader, J.R.R.A. Martins, J.J. Alonso, E. van der Weide, ADjoint: an approach for the rapid development of discrete adjoint solvers, *AIAA J.* 46 (4) (2008) 863–873, <https://doi.org/10.2514/1.29123>.
- [2] J.R.R.A. Martins, J.T. Hwang, Review and unification of methods for computing derivatives of multidisciplinary computational models, *AIAA J.* 51 (11) (2013) 2582–2599, <https://doi.org/10.2514/1.J052184>.
- [3] G.K.W. Kenway, J.R.R.A. Martins, Multipoint high-fidelity aerostructural optimization of a transport aircraft configuration, *J. Aircr.* 51 (1) (2014) 144–160, <https://doi.org/10.2514/1.C032150>.
- [4] Z. Lyu, G.K.W. Kenway, J.R.R.A. Martins, Aerodynamic shape optimization investigations of the common research model wing benchmark, *AIAA J.* 53 (4) (2015) 968–985, <https://doi.org/10.2514/1.J053318>.
- [5] Y. Yu, Z. Lyu, Z. Xu, J.R.R.A. Martins, On the influence of optimization algorithm and starting design on wing aerodynamic shape optimization, *Aerosp. Sci. Technol.* 75 (2018) 183–199, <https://doi.org/10.1016/j.ast.2018.01.016>.
- [6] N. Bons, X. He, C.A. Mader, J.R.R.A. Martins, Multimodality in aerodynamic wing design optimization, *AIAA J.* 57 (3) (2019) 1004–1018, <https://doi.org/10.2514/1.J057294>.
- [7] A.R. Conn, K. Scheinberg, L.N. Vicente, *Introduction to Derivative-Free Optimization*, vol. 8, SIAM, 2009.
- [8] L.M. Rios, N.V. Sahinidis, Derivative-free optimization: a review of algorithms and comparison of software implementations, *J. Glob. Optim.* 56 (3) (2013) 1247–1293.
- [9] F. Boukouvava, R. Misener, C.A. Floudas, *Global optimization advances in mixed-integer nonlinear programming, MINLP, and constrained derivative-free optimization*, CDFO, *Eur. J. Oper. Res.* 252 (3) (2016) 701–727.
- [10] C. Audet, W. Hare, *Derivative-Free and Blackbox Optimization*, Springer, 2017.
- [11] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, *IEEE Trans. Evol. Comput.* 6 (2) (2002) 182–197.
- [12] N. Hansen, S.D. Müller, P. Koumoutsakos, Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES), *Evol. Comput.* 11 (1) (2003) 1–18.
- [13] N. Hansen, S. Kern, Evaluating the CMA Evolution Strategy on Multimodal Test Functions, *PPSN*, vol. 8, Springer, 2004, pp. 282–291.
- [14] G.A. Jastrebski, D.V. Arnold, Improving evolution strategies through active covariance matrix adaptation, in: 2006 IEEE Congress on Evolutionary Computation, CEC 2006, IEEE, 2006, pp. 2814–2821.
- [15] Y. Diouane, S. Gratton, L.N. Vicente, Globally convergent evolution strategies for constrained optimization, *Comput. Optim. Appl.* 62 (2) (2015) 323–346.
- [16] S. Wessing, M. Preuss, The true destination of EGO is multi-local optimization, in: 2007 IEEE Latin American Conference on Computational Intelligence, LACCI, IEEE, 2017, pp. 1–6.
- [17] S. Le Digabel, Algorithm 909: NOMAD: nonlinear optimization with the MADS algorithm, *ACM Trans. Math. Softw.* 37 (4) (2011) 44.
- [18] T.D. Plantenga, *Hopsack 2.0 user manual*, Sandia National Laboratories, Albuquerque, NM and Livermore, CA, SAND2009–SAND6265.
- [19] G. Fasano, G. Liuzzi, S. Lucidi, F. Rinaldi, A linesearch-based derivative-free approach for nonsmooth constrained optimization, *SIAM J. Optim.* 24 (3) (2014) 959–992.
- [20] J. Moćkus, On Bayesian methods for seeking the extremum, in: *Optimization Techniques IFIP Technical Conference*, 1975, pp. 400–404.
- [21] D.R. Jones, M. Schonlau, W.J. Welch, Efficient global optimization of expensive black-box functions, *J. Glob. Optim.* 13 (4) (1998) 455–492.
- [22] A. Forrester, A. Sobester, A. Keane, *Engineering Design via Surrogate Modelling: A Practical Guide*, John Wiley & Sons, 2008.
- [23] Z. Wang, S. Jegelka, Max-value entropy search for efficient Bayesian optimization, in: 34th International Conference on Machine Learning, vol. 70, 2017, pp. 3627–3635.
- [24] M.A. Gelbart, *Constrained Bayesian Optimization and Applications*, Ph.D. thesis, Harvard, 2015.
- [25] V. Picheny, A stepwise uncertainty reduction approach to constrained global optimization, in: *AISTATS*, 2014, pp. 787–795.
- [26] V. Picheny, R.B. Gramacy, S. Wild, S. Le Digabel, Bayesian optimization under mixed constraints with a slack-variable augmented Lagrangian, in: *Advances in Neural Information Processing Systems*, 2016, pp. 1435–1443.
- [27] S. Shan, G.G. Wang, Survey of modeling and optimization strategies to solve high-dimensional design problems with computationally-expensive black-box functions, *Struct. Multidiscip. Optim.* 41 (2) (2010) 219–241.
- [28] A.I. Forrester, A.J. Keane, Recent advances in surrogate-based optimization, *Prog. Aerosp. Sci.* 45 (1) (2009) 50–79.
- [29] G.G. Wang, S. Shan, Review of metamodeling techniques in support of engineering design optimization, *J. Mech. Des.* 129 (4) (2007) 370–380.
- [30] N.V. Queipo, R.T. Haftka, W. Shyy, T. Goel, R. Vaidyanathan, P.K. Tucker, Surrogate-based analysis and optimization, *Prog. Aerosp. Sci.* 41 (1) (2005) 1–28.
- [31] C.E. Rasmussen, C.K. Williams, *Gaussian Processes for Machine Learning*, vol. 1, MIT Press, Cambridge, 2006.
- [32] G. Matheron, Principles of geostatistics, *Econ. Geol.* 58 (8) (1963) 1246–1266.

- [33] J. Sacks, W.J. Welch, T.J. Mitchell, H.P. Wynn, Design and analysis of computer experiments, *Stat. Sci.* (1989) 409–423.
- [34] J.P. Kleijnen, Kriging metamodeling in simulation: a review, *Eur. J. Oper. Res.* 192 (3) (2009) 707–716.
- [35] M.J. Sasena, P.Y. Papalambros, P. Goovaerts, The use of surrogate modeling algorithms to exploit disparities in function computation time within simulation-based optimization, in: *The Fourth World Congress of Structural and Multidisciplinary Optimization*, 2001, pp. 1–6.
- [36] F. Palacios, J.J. Alonso, M. Colonno, J. Hicken, T. Lukaczyk, Adjoint-based method for supersonic aircraft design using equivalent area distributions, *AIAA Pap.* 269 (2012) 2012.
- [37] Z.-H. Han, R. Zimmermann, S. Görtz, A New Cokriging Method for Variable-Fidelity Surrogate Modeling of Aerodynamic Data, *AIAA Paper* 1225, 2010, p. 2010.
- [38] T. Benamara, P. Breitkopf, I. Lepot, C. Sainvitu, Adaptive infill sampling criterion for multi-fidelity optimization based on Gappy-POD, *Struct. Multidiscip. Optim.* 54 (4) (2016) 843–855.
- [39] J. Liu, W.-P. Song, Z.-H. Han, Y. Zhang, Efficient aerodynamic shape optimization of transonic wings using a parallel infilling strategy and surrogate models, *Struct. Multidiscip. Optim.* 55 (3) (2017) 925–943.
- [40] Y. Ma, W. Zhou, Q. Han, Research of multi-point infill criteria based on multi-objective optimization front and its application on aerodynamic shape optimization, *Adv. Mech. Eng.* 9 (6) (2017) 1687814017703340.
- [41] M.A. Bouhlef, N. Bartoli, A. Otsmane, J. Morlier, Improving kriging surrogates of high-dimensional design models by Partial Least Squares dimension reduction, *Struct. Multidiscip. Optim.* (ISSN 1615-1488) 53 (5) (2016) 935–952, <https://doi.org/10.1007/s00158-015-1395-9>.
- [42] M.A. Bouhlef, N. Bartoli, A. Otsmane, J. Morlier, An improved approach for estimating the hyperparameters of the Kriging model for high-dimensional problems through the partial least squares method, *Math. Probl. Eng.* (2016).
- [43] M.A. Bouhlef, N. Bartoli, R.G. Regis, A. Otsmane, J. Morlier, Efficient global optimization for high-dimensional constrained problems by using the Kriging models combined with the partial least squares method, *Eng. Optim.* (2018) 1–16, <https://doi.org/10.1080/0305215X.2017.1419344>.
- [44] T. Hastie, R. Tibshirani, J. Friedman, J. Franklin, The elements of statistical learning: data mining, inference and prediction, *Math. Intell.* 27 (2) (2005) 83–85.
- [45] D. Bettebghor, N. Bartoli, S. Grihon, J. Morlier, M. Samuelides, Surrogate modeling approximation using a mixture of experts based on EM joint estimation, *Struct. Multidiscip. Optim.* (ISSN 1615-147X) 43 (2) (2011) 243–259, <https://doi.org/10.1007/s00158-010-0554-2>.
- [46] Y. Tenne, An optimization algorithm employing multiple metamodels and optimizers, *Int. J. Autom. Comput.* 10 (3) (2013) 227–241.
- [47] M. Sasena, P. Papalambros, P. Goovaerts, Global optimization of problems with disconnected feasible regions via surrogate modeling, in: *9th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, Atlanta, GA, 2002, Paper No. AIAA-2002-5573.
- [48] S. Chen, Z. Jiang, S. Yang, W. Chen, Multimodel fusion based sequential optimization, *AIAA J.* 55 (1) (2016) 241–254.
- [49] N. Bartoli, M.-A. Bouhlef, I. Kurek, R. Lafage, T. Lefebvre, J. Morlier, R. Priem, V. Stütz, R. Regis, Improvement of efficient global optimization with application to aircraft wing design, in: *17th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, Washington DC, USA, 2016, p. 4001.
- [50] AIAA Aerodynamic Design Optimization Discussion Group, <http://mdolab.engin.umich.edu/content/aerodynamic-design-optimization-workshop>, 2019.
- [51] D.G. Krige, A Statistical Approach to Some Mine Evaluations and Allied Problems at the Witwatersrand, Master's thesis, University of Witwatersrand, 1951.
- [52] M.J. Sasena, P. Papalambros, P. Goovaerts, Exploration of metamodeling sampling criteria for constrained global optimization, *Eng. Optim.* 34 (3) (2002) 263–278, <https://doi.org/10.1080/03052150211751>.
- [53] M. Schonlau, W.J. Welch, D.R. Jones, Global Versus Local Search in Constrained Optimization of Computer Models, *Lecture Notes-Monograph Series*, 1998, pp. 11–25.
- [54] C. Audet, J. Denni, D. Moore, A. Booker, P. Frank, A surrogate-model-based method for constrained optimization, in: *8th Symposium on Multidisciplinary Analysis and Optimization*, 2000, p. 4891.
- [55] J.M. Hernández-Lobato, M.A. Gelbart, R.P. Adams, M.W. Hoffman, Z. Ghahramani, A general framework for constrained Bayesian optimization using information-based search, *J. Mach. Learn. Res.* 17 (160) (2016) 1–53.
- [56] R. Priem, N. Bartoli, Y. Diouane, S. Dubreuil, An adaptive feasibility approach for constrained bayesian optimization with application in aircraft design, in: *Engopt: 6th International Conference on Engineering Optimization*, Lisboa, Portugal, 2018.
- [57] A.G. Watson, R.J. Barnes, Infill sampling criteria to locate extremes, *Math. Geol.* 27 (5) (1995) 589–608.
- [58] M. Sasena, Flexibility and Efficiency Enhancements for Constrained Global Design Optimization with Kriging Approximations, Ph.D. thesis, University of Michigan, 2002.
- [59] D. Bettebghor, N. Bartoli, Approximation of the critical buckling factor for composite panels, *Struct. Multidiscip. Optim.* 46 (4) (2012) 561–584, <http://www.springerlink.com/content/e2855877v3x8557/>.
- [60] R.P. Liem, C.A. Mader, J.R.R.A. Martins, Surrogate models and mixtures of experts in aerodynamic performance prediction for mission analysis, *Aerosp. Sci. Technol.* 43 (2015) 126–151, <https://doi.org/10.1016/j.ast.2015.02.019>.
- [61] M.A. Bouhlef, J.T. Hwang, N. Bartoli, R. Lafage, J. Morlier, J.R.R.A. Martins, A Python surrogate modeling framework with derivatives, *Adv. Eng. Softw.*, <https://doi.org/10.1016/j.advengsoft.2019.03.005>.
- [62] P.S. Bradley, U.M. Fayyad, C. Reina, et al., Scaling clustering algorithms to large databases, in: *KDD*, 1998, pp. 9–15.
- [63] M.F. Anjos, D.R. Jones, MOPTA 2008 benchmark, <http://www.miguelanjos.com/jones-benchmark>, 2009.
- [64] J.S. Gray, J.T. Hwang, J.R.R.A. Martins, K.T. Moore, B.A. Naylor, OpenMDAO: an open-source framework for multidisciplinary design, in: *analysis, and optimization*, *Struct. Multidiscip. Optim.* (2019), <https://doi.org/10.1007/s00158-019-02211-z>.
- [65] J. Parr, A. Keane, A.I. Forrester, C. Holden, Infill sampling criteria for surrogate-based optimization with constraint handling, *Eng. Optim.* 44 (10) (2012) 1147–1166.
- [66] M.J. Powell, A direct search optimization method that models the objective and constraint functions by linear interpolation, in: *Advances in Optimization and Numerical Analysis*, Springer, 1994, pp. 51–67.
- [67] C. Audet, S. Le Digabel, C. Tribes, NOMAD User Guide, Tech. Rep. G-2009-37, Les cahiers du GERAD, 2009, https://www.gerad.ca/nomad/Downloads/user_guide.pdf.
- [68] J. Nocedal, S.J. Wright, *Numerical Optimization*, Springer, ISBN 978-0-387-30303-1, 2006.
- [69] R.B. Gramacy, G.A. Gray, S. Le Digabel, H.K. Lee, P. Ranjan, G. Wells, S.M. Wild, Modeling an augmented Lagrangian for blackbox constrained optimization, *Technometrics* 58 (2016) 1–11.
- [70] R. Jin, W. Chen, A. Sudjianto, An efficient algorithm for constructing optimal design of computer experiments, *J. Stat. Plan. Inference* 134 (1) (2005) 268–287.
- [71] E. Jones, T. Oliphant, P. Peterson, et al., SciPy: open source scientific tools for Python, <http://www.scipy.org/>, 2001. (Accessed 27 April 2016) [Online].
- [72] M. Abramson, C. Audet, G. Couture, J. Dennis Jr., S. Le Digabel, C. Tribes, The NOMAD project, Software available at <https://www.gerad.ca/nomad/>, 2009.
- [73] V. Picheny, D. Ginsbourger, O. Roustant, DiceOptim: Kriging-based optimization for computer experiments, <https://github.com/cran/DiceOptim>, 2014.
- [74] D. Kraft, et al., A Software Package for Sequential Quadratic Programming, DFLR Obersfaffeuhofen, Germany, 1988.
- [75] C. Audet, J.E. Dennis Jr., A progressive barrier for derivative-free nonlinear programming, *SIAM J. Optim.* 20 (1) (2009) 445–472.
- [76] G.K.W. Kenway, N. Secco, J.R.R.A. Martins, A. Mishra, K. Duraisamy, An efficient parallel overset method for aerodynamic shape optimization, in: *Proceedings of the 58th AIAA/ASCE/AHS/JSC Structures, Structural Dynamics, and Materials Conference*, AIAA SciTech Forum, Grapevine, TX, 2017.
- [77] Z. Lyu, G.K. Kenway, C. Paige, J.R.R.A. Martins, Automatic differentiation adjoint of the Reynolds-averaged Navier–Stokes equations with a turbulence model, in: *21st AIAA Computational Fluid Dynamics Conference*, San Diego, CA, 2013.
- [78] P. Spalart, S. Allmaras, A one-equation turbulence model for aerodynamic flows, in: *30th Aerospace Sciences Meeting and Exhibit*, Aerospace Sciences Meetings, American Institute of Aeronautics and Astronautics, 1992.
- [79] G.K. Kenway, G.J. Kennedy, J.R.R.A. Martins, A CAD-free approach to high-fidelity aerostuctural optimization, in: *Proceedings of the 13th AIAA/ISSMO Multidisciplinary Analysis Optimization Conference*, Fort Worth, TX, AIAA 2010-9231, 2010.
- [80] E. Luke, E. Collins, E. Blades, A fast mesh deformation method using explicit interpolation, *J. Comput. Phys.* (ISSN 0021-9991) 231 (2) (2012) 586–601, <https://doi.org/10.1016/j.jcp.2011.09.021>.
- [81] R.E. Perez, P.W. Jansen, J.R.R.A. Martins, pyOpt: a python-based object-oriented framework for nonlinear constrained optimization, *Struct. Multidiscip. Optim.* 45 (1) (2012) 101–118, <https://doi.org/10.1007/s00158-011-0666-3>.
- [82] G.K.W. Kenway, J.R.R.A. Martins, pyOptSparse – Python OPTimization (sparse) framework, <https://github.com/mdolab/pyoptsparse>, 2018.
- [83] P.E. Gill, W. Murray, M.A. Saunders, SNOPT: an SQP algorithm for large-scale constrained optimization, *SIAM Rev.* 47 (1) (2005) 99–131.
- [84] G.K.W. Kenway, J.R.R.A. Martins, Buffet onset constraint formulation for aerodynamic shape optimization, *AIAA J.* 55 (6) (2017) 1930–1947, <https://doi.org/10.2514/1.J055172>.
- [85] R.P. Liem, J.R.R.A. Martins, G.K. Kenway, Expected drag minimization for aerodynamic design optimization based on aircraft operational data, *Aerosp. Sci. Technol.* 63 (2017) 344–362, <https://doi.org/10.1016/j.ast.2017.01.006>.
- [86] S. Chen, Z. Lyu, G.K.W. Kenway, J.R.R.A. Martins, Aerodynamic shape optimization of the common research model wing-body-tail configuration, *J. Aircr.* 53 (1) (2016) 276–293, <https://doi.org/10.2514/1.C033328>.
- [87] G.K.W. Kenway, J.R.R.A. Martins, Multipoint aerodynamic shape optimization investigations of the common research model wing, *AIAA J.* 54 (1) (2016) 113–128, <https://doi.org/10.2514/1.J054154>.
- [88] G.K.W. Kenway, G.J. Kennedy, J.R.R.A. Martins, Scalable parallel approach for high-fidelity steady-state aeroelastic analysis and derivative computations, *AIAA J.* 52 (5) (2014) 935–951, <https://doi.org/10.2514/1.J052255>.

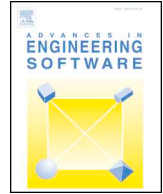
- [89] T.R. Brooks, G.K.W. Kenway, J.R.R.A. Martins, Benchmark aerostructural models for the study of transonic aircraft wings, *AIAA J.* 56 (7) (2018) 2840–2855, <https://doi.org/10.2514/1.J056603>.
- [90] D.A. Burdette, J.R.R.A. Martins, Design of a transonic wing with an adaptive morphing trailing edge via aerostructural optimization, *Aerosp. Sci. Technol.* 81 (2018) 192–203, <https://doi.org/10.1016/j.ast.2018.08.004>.

4.4.7 A Python surrogate modeling framework with derivatives, Advances in Engineering Software 2019



Contents lists available at ScienceDirect

Advances in Engineering Software

journal homepage: www.elsevier.com/locate/advengsoft

A Python surrogate modeling framework with derivatives

Mohamed Amine Bouhlel^{*,a}, John T. Hwang^b, Nathalie Bartoli^c, Rémi Lafage^c, Joseph Morlier^d, Joaquim R.R.A. Martins^a^a Department of Aerospace Engineering, University of Michigan, Ann Arbor, MI, USA^b Department of Mechanical and Aerospace Engineering, University of California San Diego, La Jolla, CA, USA^c ONERA/DTIS, Université de Toulouse, Toulouse, France^d ICA, Université de Toulouse, ISAE-SUPAERO, INSA, CNRS, MINES ALBI, UPS, Toulouse, France

ARTICLE INFO

Keywords:

Surrogate modeling
Gradient-enhanced surrogate modeling
Derivatives

ABSTRACT

The surrogate modeling toolbox (SMT) is an open-source Python package that contains a collection of surrogate modeling methods, sampling techniques, and benchmarking functions. This package provides a library of surrogate models that is simple to use and facilitates the implementation of additional methods. SMT is different from existing surrogate modeling libraries because of its emphasis on derivatives, including training derivatives used for gradient-enhanced modeling, prediction derivatives, and derivatives with respect to training data. It also includes unique surrogate models: kriging by partial least-squares reduction, which scales well with the number of inputs; and energy-minimizing spline interpolation, which scales well with the number of training points. The efficiency and effectiveness of SMT are demonstrated through a series of examples. SMT is documented using custom tools for embedding automatically tested code and dynamically generated plots to produce high-quality user guides with minimal effort from contributors. SMT is maintained in a public version control repository.¹

1. Motivation and significance

In the last few decades, numerical models for engineering have become more complex and accurate, but the computational time for running these models has not necessarily decreased. This makes it difficult to complete engineering tasks that rely on these models, such as design space exploration and optimization. Surrogate modeling is often used to reduce the computational time of these tasks by replacing expensive numerical simulations with approximate functions that are much faster to evaluate. Surrogate models are constructed by evaluating the original model at a set of points, called training points, and using the corresponding evaluations to construct an approximate model based on mathematical functions.

Surrogate modeling is often used in the context of design optimization because of the repeated model evaluations that are required. Derivatives play an important role in optimization, because problems with a large number of optimization variables require gradient-based algorithms for efficient scalability. Therefore, situations frequently arise where there are requirements for surrogate models associated with the computation or use of derivatives.

There are three types of derivatives in surrogate modeling: prediction derivatives, training derivatives, and output derivatives. Prediction derivatives are the derivatives of the surrogate model outputs with respect to the inputs, and they are the derivatives required when using a surrogate model in gradient-based optimization. Training derivatives are derivatives of the training outputs with respect to the training inputs that provide additional training data that increase the accuracy of the surrogate model. Output derivatives are derivatives of the prediction outputs with respect to the training outputs, which are required if a surrogate model is reconstructed within a gradient-based optimization process. We describe these derivatives in more detail in Section 3.3.

Various packages that build surrogate models have been developed using different programming languages, such as Scikit-learn in Python [1], SUMO in MATLAB [2], and GPML in MATLAB and Octave [3]. However, these packages do not handle the derivatives.

In this paper, we introduce a new Python package called the surrogate modeling toolbox (SMT). SMT is different from existing surrogate modeling libraries because of its emphasis on derivatives, including all three types of derivatives described above. SMT also includes newly developed surrogate models that handle derivatives and

* Corresponding author.

E-mail addresses: mbouhlel@umich.edu (M.A. Bouhlel), jhwang@eng.ucsd.edu (J.T. Hwang), nathalie.bartoli@onera.fr (N. Bartoli), remi.lafage@onera.fr (R. Lafage), joseph.morlier@isae-supaero.fr (J. Morlier), jrram@umich.edu (J.R.R.A. Martins).¹ <https://github.com/SMTorg/SMT>.<https://doi.org/10.1016/j.advengsoft.2019.03.005>Received 10 September 2018; Received in revised form 5 March 2019; Accepted 11 March 2019
0965-9978/ Published by Elsevier Ltd.

do not exist elsewhere: partial least-squares (PLS)-based surrogate models [4–6], which are suitable for high-dimensional problems, and the regularized minimal-energy tensor-product spline (RMTS), which is suitable for low-dimensional problems with up to hundreds of thousands of sampling points [7]. To use SMT, the user should first provide a set of training points. This could be done either by using the sampling techniques and benchmarking functions implemented in SMT or by directly importing the data. Then, the user can build a surrogate model based on the training points and make predictions for the function values and derivatives.

The main goal of this work is to provide a simple Python toolbox that contains a set of surrogate modeling functions and supports different kinds of derivatives that are useful for many applications in engineering. SMT includes sampling techniques, which are necessary for the construction of a surrogate model. Various sample functions are also included to facilitate the benchmarking of different techniques and for reproducing results. SMT is suitable for both novice and advanced users and is supported by detailed documentation available online with examples of each implemented surrogate modeling approach.² It is hosted publicly in a version-controlled repository,³ and is easily imported and used within Python scripts. It is released under the New BSD License and runs on Linux, macOS, and Windows operating systems. Regression tests are run automatically on each operating system whenever a change is committed to the repository.

The remainder of this paper is organized as follows. First, we present the architecture and the main implementation features of SMT in Section 2 and then we describe the methods implemented in SMT in Section 3. Section 4 gives an example of SMT usage and presents a set of benchmarking functions implemented within SMT. We apply SMT to two engineering problems in Section 5 and present the conclusions in Section 6.

2. Software architecture, documentation, and automatic testing

SMT is composed of three main modules (*sampling methods*, *problems*, and *surrogate models*) that implement a set of sampling techniques, benchmarking functions, and surrogate modeling techniques, respectively. Each module contains a common interface inherited by the corresponding methods, and each method implements the functions required by the interface, as shown in Fig. 1.

SMT's documentation is written using reStructuredText and is generated using the Sphinx package for documentation in Python, along with custom extensions.⁴ The documentation pages include embedded code snippets that demonstrate the usage. These code snippets are extracted dynamically from actual tests in the source code, ensuring that the code snippets are always up to date. The print output and plots from the code snippets are also generated dynamically by custom Sphinx extensions and embedded in the documentation page. This leads to high-quality documentation with low effort, requiring only the addition of a custom directive and the path to locating the test code. Similarly, another custom directive embeds a table of options, default values, valid types, valid values, and descriptions in the documentation for the surrogate model, sampling method, or benchmarking problem. The documentation also uses existing Sphinx directives to embed descriptions of the user-callable methods in the classes.

In addition to the user documentation, we also provide developer documentation that explains how to contribute code to SMT. The developer documentation includes a different list of application programming interface (API) methods for the *SurrogateModel*, *SamplingMethod*, and *Problem* classes, which are classes that must be implemented to create a new surrogate modeling method, sampling

technique, or benchmarking problem, respectively.

When a developer issues a pull request, the request is merged once the automatically triggered tests run successfully and at least one reviewer approves it. The repository on GitHub⁵ is linked to two continuous integration testing services, Travis CI (for Linux and macOS) and AppVeyor (for Windows), which trigger test suite execution whenever code is committed and prevent changes from being merged if any of the tests fail.

3. Surrogate modeling methods

To build a surrogate model, two main steps are necessary. First, we generate a set of training points from the input space where the quantity of interest is computed. This step can be done by using one of the sampling techniques implemented in SMT (Section 3.1), or by uploading an existing training dataset. Second, we train the desired surrogate model on those points and make a prediction of the output values and derivatives (Section 3.2).

3.1. Sampling methods

SMT contains a library of sampling methods used to generate sets of points in the input space, either for training or for prediction.

Random sampling This class creates random samples from a uniform distribution over the design space.

Latin hypercube sampling This is a statistical method for generating a quasi-random sampling distribution. It is among the most popular sampling techniques in computer experiments thanks to its simplicity and projection properties in high-dimensional problems. Five construction criteria are available in SMT: four criteria defined in the *pyDOE* package⁶ and the enhanced stochastic evolutionary criterion [8].

Full-factorial sampling This is a common sampling method where all input variables are set at two levels each. These levels are usually denoted by +1 and -1 for *high* and *low* levels, respectively. Full-factorial sampling computes all possible combinations of these levels across all such input variables.

Figure 2 shows the implementation of the *Random* class, which inherits from the *SamplingMethod* class.

3.2. Surrogate models

Table 1 lists the surrogate modeling methods currently available in SMT and summarizes the advantages and disadvantages of each method. The methods include both well-established methods and methods recently developed by the authors that use derivative information. Among the well-established methods, we implement kriging [9], radial basis functions (RBF) [10], inverse distance weighting (IDW) [11], least squares (LS) [12, ch. 3], and quadratic polynomials (QP) [12, ch. 3]. In Fig. 3, we show the implementation of the *RBF* class, which inherits from the *SurrogateModel* class.

The methods that we recently developed are kriging combined with partial least squares (KPLS) [4], KPLSK for the construction of a standard kriging model in high-dimensional problems [5], gradient-enhanced KPLS (GE-KPLS) [6], and RMTS [7].

Surrogate models provide a vector of prediction outputs \mathbf{y} for a given vector of prediction inputs $\mathbf{x} \in \mathbb{R}^{n_x}$, where n_x is the number of dimensions, and can be expressed as

$$\mathbf{y} = f(\mathbf{x}, \mathbf{x}_t, \mathbf{y}_t), \quad (1)$$

² <http://smt.readthedocs.io/en/latest>.

³ <https://github.com/SMTorg/smt>.

⁴ <https://smt.readthedocs.org>.

⁵ <https://github.com/SMTorg/smt>.

⁶ <https://pythonhosted.org/pyDOE/randomized.html>

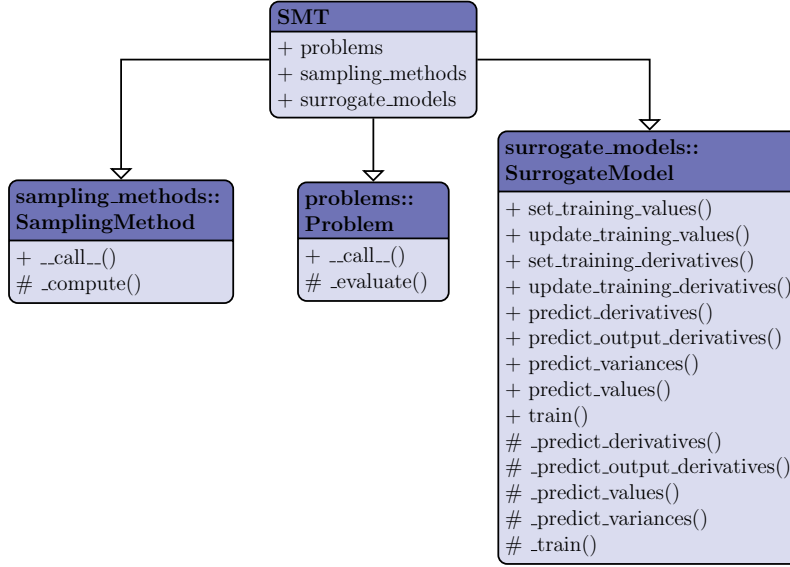


Fig. 1. Architecture of SMT.

```

import numpy as np
from six.moves import range
from smt.sampling_methods.sampling_method
    import SamplingMethod

class Random(SamplingMethod):

    def _compute(self, n):
        """
        Compute the requested number of sampling points.
        Arguments
        -----
        n : int
            Number of points requested.
        Returns
        -----
        ndarray[n, nx]
            The sampling locations in the input space.
        """
        xlimits = self.options['xlimits']
        nx = xlimits.shape[0]
        return np.random.rand(n, nx)
  
```

Fig. 2. Implementation of the *Random* class, which inherits from the *SamplingMethod* class. The *compute* function evaluates the requested number of sampling points uniformly over the design space.

where $\mathbf{x}_t \in \mathbb{R}^{n_x}$ and \mathbf{y}_t are the vectors of training inputs and outputs, respectively, which are used to build the surrogate model a priori, and \mathbf{x} is the unknown point to predict with the surrogate model. We now describe the surrogate modeling methods available in SMT.

3.2.1. LS and QP

The LS method [12] fits a linear model with coefficients $\beta = (\beta_0, \beta_1, \dots, \beta_{n_x})$, to minimize the residual sum of the squares between the observed responses in the dataset, and the responses predicted by the linear approximation, i.e.,

$$\min_{\beta} \|\mathbf{X}\beta - \mathbf{y}\|_2^2, \quad (2)$$

where $\mathbf{X} = (1, \mathbf{x}^{(1)T}, \dots, \mathbf{x}^{(n_t)T})^T$. The vectors in \mathbf{X} have $(n_t \times n_x + 1)$ dimensions and n_t is the number of training points. The square polynomial model is given by Hastie et al. [12]

$$\mathbf{y} = \mathbf{X}\beta + \epsilon, \quad (3)$$

where ϵ is a vector of random errors and

$$\mathbf{X} = \begin{bmatrix} 1 & x_1^{(1)} & \dots & x_{n_x}^{(1)} & x_1^{(1)}x_2^{(1)} & \dots & x_{n_x-1}^{(1)}x_{n_x}^{(1)} & x_1^{(1)2} & \dots & x_{n_x}^{(1)2} \\ \vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots \\ 1 & x_1^{(n_t)} & \dots & x_{n_x}^{(n_t)} & x_1^{(n_t)}x_2^{(n_t)} & \dots & x_{n_x-1}^{(n_t)}x_{n_x}^{(n_t)} & x_1^{(n_t)2} & \dots & x_{n_x}^{(n_t)2} \end{bmatrix}. \quad (4)$$

The vector of estimated polynomial regression coefficients using ordinary LS estimation is

$$\beta = \mathbf{X}^T\mathbf{X}^{-1}\mathbf{X}^T\mathbf{y}. \quad (5)$$

3.2.2. IDW

The IDW model is an interpolating method where the unknown points are calculated with a weighted average of the sampling points using [11]

$$y = \begin{cases} \frac{\sum_i^{n_t} \beta(\mathbf{x}, \mathbf{x}_i^{(i)})y_i^{(i)}}{\sum_i^{n_t} \beta(\mathbf{x}, \mathbf{x}_i^{(i)})}, & \text{if } \mathbf{x} \neq \mathbf{x}_i^{(i)} \quad \forall i \\ y_i^{(i)}, & \text{if } \mathbf{x} = \mathbf{x}_i^{(i)} \text{ for some } i \end{cases}, \quad (6)$$

Table 1
Surrogate modeling methods provided by SMT.

Method	Advantages (+) and disadvantages (-)	Derivatives			References
		Train.	Pred.	Out.	
Kriging	+ Prediction variance, flexible - Costly if number of inputs or training points is large - Numerical issues when points are too close to each other	No	Yes	No	Sacks et al. [9]
KPLS	+ Prediction variance, fast construction + Suitable for high-dimensional problems - Numerical issues when points are too close to each other	No	Yes	No	Bouhleh et al. [4]
KPLSK	+ Prediction variance, fast construction + Suitable for high-dimensional problems - Numerical issues when points are too close to each other	No	Yes	No	Bouhleh et al. [5]
GE-KPLS	+ Prediction variance, fast construction + Suitable for high-dimensional problems + Control of the correlation matrix size - Numerical issues when points are too close to each other - Choice of step parameter is not intuitive	Yes	Yes	No	Bouhleh and Martins [6]
RMTS	+ Fast prediction + Training scales well up to 10^5 training points + No issues with points that are too close to each other - Poor scaling with number of inputs above 4 - Slow training overall	Yes	Yes	Yes	Hwang and Martins [7]
RBF	+ Simple, only a single tuning parameter + Fast training for small number of training points - Susceptible to oscillations - Numerical issues when points are too close to each other	No	Yes	Yes	Powell [10]
IDW	+ Simple, no training required - Derivatives are zero at training points - Poor overall accuracy	No	Yes	Yes	Shepard [11]
LS	+ Simple, fast construction - Accurate only for linear problems	No	Yes	No	Hastie et al. [12]
QP	+ Simple, fast construction - Large number of points required for large number of inputs	No	Yes	No	Hastie et al. [12]

where $\mathbf{x} \in \mathbb{R}^{n_x}$ is the prediction input vector, $y \in \mathbb{R}$ is the prediction output, $\mathbf{x}_i^{(i)} \in \mathbb{R}^{n_x}$ is the input vector for the i th training point, and $y_i^{(i)} \in \mathbb{R}$ is the output value for the i th training point. The weighting function β is defined as

$$\beta(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = \|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\|_2^{-p}, \quad (7)$$

where p is a positive real number called the power parameter, which must be strictly greater than one for the derivatives to be continuous.

3.2.3. RBF

The RBF surrogate model [10] represents the interpolating function as a linear combination of basis functions, one for each training point. RBF are named as such because the basis functions depend only on the distance from the prediction point to the training point for the basis function. The coefficients of the basis functions are computed during the training stage. RBF are frequently augmented to global polynomials to capture the general trends. The prediction equation for RBF is given by

$$y = \mathbf{p}(\mathbf{x})\mathbf{w}_p + \sum_i^{n_t} \phi(\mathbf{x}, \mathbf{x}_i^{(i)})\mathbf{w}_r, \quad (8)$$

where $\mathbf{x} \in \mathbb{R}^{n_x}$ is the prediction input vector, $y \in \mathbb{R}$ is the prediction output, $\mathbf{x}_i^{(i)} \in \mathbb{R}^{n_x}$ is the input vector for the i th training point, $\mathbf{p}(\mathbf{x}) \in \mathbb{R}^{n_p}$ is the vector mapping the polynomial coefficients to the prediction output, $\phi(\mathbf{x}, \mathbf{x}_i^{(i)}) \in \mathbb{R}^{n_t}$ is the vector mapping the RBF coefficients to the prediction output, $\mathbf{w}_p \in \mathbb{R}^{n_p}$ is the vector of polynomial

coefficients, and $\mathbf{w}_r \in \mathbb{R}^{n_t}$ is the vector of RBF coefficients. The coefficients, \mathbf{w}_p and \mathbf{w}_r , are computed by solving the following linear augmented Gram system:

$$\begin{bmatrix} \phi(\mathbf{x}_1^{(1)}, \mathbf{x}_1^{(1)}) & \dots & \phi(\mathbf{x}_1^{(1)}, \mathbf{x}_1^{(n_t)}) & \mathbf{p}(\mathbf{x}_1^{(1)})^T \\ \vdots & \ddots & \vdots & \vdots \\ \phi(\mathbf{x}_1^{(n_t)}, \mathbf{x}_1^{(1)}) & \dots & \phi(\mathbf{x}_1^{(n_t)}, \mathbf{x}_1^{(n_t)}) & \mathbf{p}(\mathbf{x}_1^{(n_t)})^T \\ \mathbf{p}(\mathbf{x}_1^{(1)}) & \dots & \mathbf{p}(\mathbf{x}_1^{(n_t)}) & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{w}_{r1} \\ \vdots \\ \mathbf{w}_{rnt} \\ \mathbf{w}_p \end{bmatrix} = \begin{bmatrix} y_1^{(1)} \\ \vdots \\ y_1^{(n_t)} \\ 0 \end{bmatrix}. \quad (9)$$

Only Gaussian basis functions are implemented currently. These can be written as follows

$$\phi(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = \exp\left(-\frac{\|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\|_2^2}{d_0^2}\right), \quad (10)$$

where d_0 is a scaling parameter.

3.2.4. Kriging-based models

Kriging [9] is an interpolating model that is a linear combination of a known function $f_i(\mathbf{x})$ added to a realization of a stochastic process, $Z(\mathbf{x})$, to obtain

$$\hat{y} = \sum_{i=1}^k \beta_i f_i(\mathbf{x}) + Z(\mathbf{x}), \quad (11)$$

where β_i is the i th linear regression coefficients and k is the number of linear regression coefficients to be determined. Here, $Z(\mathbf{x})$ has a mean of zero and a spatial covariance function given by

```

from smt.surrogate_models.surrogate_model
    import SurrogateModel

class RBF(SurrogateModel):
    def _initialize(self):
        super(RBF, self)._initialize()
        ...

    def _setup(self):
        options = self.options
        ...

    def _train(self):
        self._setup()
        ...

    def _predict_values(self, x):
        """
        Evaluates the model at a set of points.
        x : np.ndarray [n_evals, dim]
            Evaluation point input values
        y (output): Evaluation point output values
        """
        n = x.shape[0]
        ...

    def _predict_derivatives(self, x, kx):
        """
        Evaluates the derivatives at a set of points.
        x : np.ndarray [n_evals, dim]
            Evaluation point input variable values
        kx : int
            The 0-based index of the input variable with
            respect to which derivatives are desired.
        dy_dx (output): Derivative values.
        """
        n = x.shape[0]
        ...

    def _predict_output_derivatives(self, x):
        """
        Evaluates the output derivatives at a set of points.
        x : np.ndarray [n_evals, dim]
            Evaluation point input variable values
        dy_dyt (output): Output derivative values.
        """
        n = x.shape[0]
        ...

```

Fig. 3. Implementation of the RBF class, which inherits from the *SurrogateModel* class.

$$\text{cov}[Z(\mathbf{x}^{(i)}), Z(\mathbf{x}^{(j)})] = \sigma^2 R(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}), \quad (12)$$

where σ^2 is the process variance and R is the correlation. Two types of correlation functions are available in SMT: the exponential (Ornstein-Uhlenbeck process)

$$R(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = \prod_{l=1}^{n_x} \exp(-\theta_l |x_l^{(i)} - x_l^{(j)}|),$$

and the Gaussian correlation function

$$R(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = \prod_{l=1}^{n_x} \exp(-\theta_l (x_l^{(i)} - x_l^{(j)})^2),$$

where $\theta_l \in \mathbb{R}^+$. The number of hyperparameters θ is equal to the number of variables n_x . They are estimated by maximizing the likelihood function using the gradient-free optimization algorithm COBYLA [13]. These two correlation functions are called *abs_exp* (exponential) and *squar_exp* (Gaussian) in SMT. The deterministic term, the sum in the kriging prediction (11), is replaced by a constant, a linear model, or a quadratic model, all of which are available in SMT.

KPLS [4] is a kriging model that uses the PLS method. KPLS is faster than kriging because fewer hyperparameters need to be estimated to achieve high accuracy. This model is suitable for high-dimensional problems owing to the kernel constructed through the PLS method. The PLS method is a well-known tool for high-dimensional problems that searches the direction that maximizes the variance between the input and output variables. This is done by a projection in a smaller space spanned by the *principal components*. The PLS information is integrated into the kriging correlation matrix to scale the number of inputs by reducing the number of hyperparameters. The number of principal components, h , which corresponds to the number of hyperparameters in KPLS is much lower than n_x . For example, the PLS-Gaussian correlation function is

$$R(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = \prod_{l=1}^{n_x} \prod_{k=1}^h \exp(-\theta_k w_l^{(k)^2} (x_l^{(i)} - x_l^{(j)})^2). \quad (13)$$

We provide more details on the KPLS method in previous work [4].

The KPLSK model is built in two steps [5]. The first step is to run

KPLS and estimate the hyperparameters expressed in the reduced space with h dimensions. The second step is to express the estimated hyperparameters in the original space with n_x dimensions, and then use this as a starting point to optimize the likelihood function of a standard kriging model. The idea here is to guess a “good” initial hyperparameter and apply a gradient-based optimization using a classic kriging kernel. This guess is provided by the KPLS construction: the solutions $(\theta_1^*, \dots, \theta_h^*)$ and the PLS-coefficients $(w_1^{(k)}, \dots, w_{n_x}^{(k)})$ for $k = 1, \dots, h$. Using the change of variables

$$\eta_l = \sum_{k=1}^h \theta_k^* w_l^{(k)2} \quad (14)$$

for $l = 1, \dots, n_x$, we express the initial hyperparameters point in the original space. We demonstrate this using a KPLS-Gaussian correlation function R_{KPLS} in the following example [5]:

$$\begin{aligned} R_{\text{KPLS}}(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) &= \prod_{k=1}^h \prod_{l=1}^{n_x} \exp(-\theta_k w_l^{(k)2} (x_l^{(i)} - x_l^{(j)})^2) \\ &= \exp(\sum_{l=1}^{n_x} \sum_{k=1}^h -\theta_k w_l^{(k)2} (x_l^{(i)} - x_l^{(j)})^2) \\ &= \exp(\sum_{l=1}^{n_x} -\eta_l (x_l^{(i)} - x_l^{(j)})^2) \\ &= \prod_{l=1}^{n_x} \exp(-\eta_l (x_l^{(i)} - x_l^{(j)})^2), \end{aligned} \quad (15)$$

where the change of variables is performed on the third line, and the last line is a standard Gaussian correlation function. The hyperparameters point (14) provides a starting point for a gradient-based optimization applied on a standard kriging method.

3.2.5. Gradient-enhanced KPLS models

GE-KPLS [6] is a gradient-enhanced kriging (GEK) model with the PLS approach. GEK is an extension of kriging that exploits gradient information. GEK is usually more accurate than kriging; however, it is not computationally efficient when the number of inputs, the number of sampling points, or both, are high. This is primarily owing to the size of the corresponding correlation matrix that increases in proportion to both the number of inputs and the number of sampling points.

To address these issues, GE-KPLS exploits the gradient information with a slight increase in the size of the correlation matrix and reduces the number of hyperparameters. The key idea of GE-KPLS is to generate a set of approximating points around each sampling point using a first-order Taylor approximation. Then, the PLS method is applied several times, each time on a different number of sampling points with the associated sampling points. Each PLS provides a set of coefficients that contribute to each variable nearby the associated sampling point to the output. Finally, an average of all PLS coefficients is computed to estimate the global influence to the output. Denoting these coefficients by $(w_1^{(k)}, \dots, w_{n_x}^{(k)})$, the GE-KPLS Gaussian correlation function is given by

$$R_{\text{GE-KPLS}}(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = \sigma \prod_{l=1}^{n_x} \prod_{k=1}^h \exp(-\theta_k (w_l^{(k)} x_l^{(i)} - w_l^{(k)} x_l^{(j)})^2). \quad (16)$$

This approach reduces the number of hyperparameters (reduced dimension) from n_x to h , where $h \ll n_x$.

As mentioned previously, PLS is applied several times with respect to each sampling point, which provides the influence of each input variable around that point. The idea here is to add only m approximating points ($m \in [1, n_x]$) around each sampling point. Only the m highest coefficients given by the first principal component are considered, which usually capture the most useful information. Bouhlel and Martins [6] provide more details on this approach.

3.2.6. RMTS Method

The RMTS model is a type of surrogate model for low-dimensional problems with large datasets that has fast prediction capability [7]. The underlying mathematical functions are tensor-product splines, which limits RMTS to up to four-dimensional problems, or five-dimensional problems in certain cases. On the other hand, tensor-product splines

enable a fast prediction time that does not increase with the number of training points. Unlike other methods, such as kriging and RBF, RMTS is not susceptible to numerical issues when there is a large number of training points or when points are too close together. The prediction equation for RMTS is given by

$$y = \mathbf{F}(\mathbf{x})\mathbf{w}, \quad (17)$$

where $\mathbf{x} \in \mathbb{R}^{n_x}$ is the prediction input vector, $y \in \mathbb{R}$ is the prediction output, $\mathbf{w} \in \mathbb{R}^{n_w}$ is the vector of spline coefficients, and $\mathbf{F}(\mathbf{x}) \in \mathbb{R}^{n_w}$ is the vector mapping the spline coefficients to the prediction output.

RMTS computes the coefficients of the splines, \mathbf{w} , by solving an energy minimization problem subject to the conditions that the splines pass through the training points. This is formulated as an unconstrained optimization problem where the objective function consists of a term with the second derivatives of the splines, a term that represents the approximation error for the training points, and another term for regularization. Thus, this optimization problem can be written as

$$\min_{\mathbf{w}} \frac{1}{2} \mathbf{w}^T \mathbf{H} \mathbf{w} + \frac{1}{2} \beta \mathbf{w}^T \mathbf{w} + \frac{1}{2} \frac{1}{\alpha} \sum_i^{n_t} [\mathbf{F}(\mathbf{x}_i^{(i)}) \mathbf{w} - y_i^{(i)}]^2, \quad (18)$$

where $\mathbf{x}_i^{(i)} \in \mathbb{R}^{n_x}$ is the input vector for the i th training point, $y_i^{(i)} \in \mathbb{R}$ is the output value for the i th training point, $\mathbf{H} \in \mathbb{R}^{n_w \times n_w}$ is the matrix of second derivatives, $\mathbf{F}(\mathbf{x}_i^{(i)}) \in \mathbb{R}^{n_w}$ is the vector mapping the spline coefficients to the i th training output, and α and β are the regularization coefficients.

In problems with a large number of training points relative to the number of spline coefficients, the energy minimization term is not necessary and can be set to zero by setting the *reg_cons* option to zero. In problems with a small dataset, the energy minimization is necessary. When the true function has high curvature, the energy minimization can be counterproductive. This can be addressed by increasing the quadratic approximation term to one of higher order and using Newton’s method to solve the resulting nonlinear system. The nonlinear formulation is given by

$$\min_{\mathbf{w}} \frac{1}{2} \mathbf{w}^T \mathbf{H} \mathbf{w} + \frac{1}{2} \beta \mathbf{w}^T \mathbf{w} + \frac{1}{2} \frac{1}{\alpha} \sum_i^{n_t} [\mathbf{F}(\mathbf{x}_i^{(i)}) \mathbf{w} - y_i^{(i)}]^p, \quad (19)$$

where p is the order set by the *approx_order* option. The number of Newton iterations is specified in the *nonlinear_maxiter* option.

RMTS is implemented in SMT with two choices of splines: B-splines and cubic Hermite splines. RMTB uses B-splines with a uniform knot vector in each dimension. The number of B-spline control points and the B-spline order in each dimension are options that trade off efficiency and precision of the interpolant. For the cubic Hermite splines, RMTS divides the domain into tensor-product cubic elements. For adjacent elements, the values and derivatives are continuous. The number of elements in each dimension is an option that trades off efficiency and precision. B-splines are usually the better choice when training time is the most important factor, whereas cubic Hermite splines are the better choice when the accuracy of the interpolant is most important [7].

3.3. Derivatives

There are three types of derivatives in SMT.

Prediction derivatives (dy/dx) are derivatives of predicted outputs with respect to the inputs at which the model is evaluated. These are computed together with the prediction outputs when the surrogate model is evaluated [14]. These are required for gradient-based optimization algorithms based on surrogate models [15].

Training derivatives (dy_i/dx_i) are derivatives of the training outputs with respect to the corresponding inputs. These are provided by the user and are used to improve the model accuracy in GE-KPLS.

When the adjoint method is used to compute training derivatives, a high-quality surrogate model can be constructed with a low relative

cost, because the adjoint method computes these derivatives at a cost independent of the number of inputs.

Output derivatives (dy/dy_i) are derivatives of predicted outputs with respect to training outputs, which is a measure of how the prediction changes with a change in training outputs, accounting for the re-training of the surrogate model. These post-training derivatives are used when the surrogate model is trained within an optimization iteration. This feature is not commonly available in other frameworks; however, it is required when the training of the surrogate model is embedded in a gradient-based optimization. In this case, derivatives of the prediction outputs with respect to the training outputs must be computed and combined with derivatives from other parts of the model using, for example, the chain rule.

Given its focus on derivatives, SMT is synergistic with the OpenMDAO framework [16], which is a software framework for gradient-based multidisciplinary analysis and optimization [17–19]. An SMT surrogate model can be a component that is part of a larger model developed in OpenMDAO and can provide the derivatives that OpenMDAO requires from its components to compute the coupled derivatives of the multidisciplinary model.

3.4. Additional surrogate modeling methods

To extend surrogate modeling to higher-level methods, we implemented a component within SMT named *extensions*. These methods require additional and sometimes different steps than the usual surrogate modeling. For example, multi-fidelity methods combine data generated from different sources, such as coarse and fine mesh solutions. Another example is the mixture of experts (MoE) class of methods, which linearly combines several surrogates. Three such methods are available within SMT.

MoE modeling performs a weighted sum of local surrogate models (experts) instead of one single model. It is based on splitting the input space into several subspaces via clustering algorithms and training a surrogate model within each subspace. Hastie et al. [12] provide a general introduction to the MoE method. The implementation of MoE within SMT is based on Gaussian mixture models and expectation maximization [20].

Variable-fidelity modeling (VFM) samples points using both low-fidelity function evaluations that are cheap to evaluate and more costly high-fidelity evaluations. It then uses the differences between the high- and low-fidelity evaluations to construct a bridge function that corrects the low-fidelity model [21]. SMT implements additive and multiplicative bridge functions.

Multi-fidelity kriging (MFK) uses a correlation factor $\rho(x)$ (constant, linear, or quadratic) and a discrepancy function $\delta(x)$. The high-fidelity model is given by

$$y_{\text{high}}(x) = \rho(x)y_{\text{low}}(x) + \delta(x).$$

SMT follows the formulation proposed by Le Gratiet [22], which is recursive and can be easily extended to multiple levels of fidelity.

4. Example and benchmark functions

SMT contains a library of analytical and engineering problems that can be used for instructional or benchmarking purposes. The analytical functions currently implemented in SMT are the sphere, Branin, L_p norm, Rosenbrock, and tensor-product functions, all of which are detailed in Appendix A. The engineering functions currently implemented in SMT are the cantilever beam, robot arm, torsion vibration, water flow, welded beam, and wing weight problems (described in Appendix B).

Figure 4 shows the implementation of the *Sphere* class, which

inherits from the *Problem* class. This function is a simple example that serves as a good first test for newly developed surrogate modeling and surrogate-based optimization methods. This function is a continuous and convex function where the global minimum is at the origin.

Figure 5 shows an example of the use of RMTC within SMT for the robot arm function [23]. This function is scalable; however, the number of dimensions must be an even number (we use two dimensions in this example). This function gives the position of a robot arm, which is made by multiple segments, in a two-dimensional space where the shoulder of the arm is fixed at the origin. This function is highly non-linear, and the use of training derivatives samples is particularly beneficial in this case [6].

5. Applications

In this section, we describe two applications that highlight the unique features of SMT. We also provide an overview of the main previous applications realized using SMT. The first application is the computation and validation of the prediction derivatives for an airfoil analysis tool, where we apply GE-KPLS. The second application describes a practical use for output derivatives, where we use RMTCs to compute the outputs derivatives for a surrogate model that is dynamically trained with an optimization iteration.

5.1. Airfoil analysis and shape optimization tool

Li et al. [14] used the GE-KPLS implementation in SMT with an MoE technique to develop a data-driven approach to airfoil and shape optimization based on computational fluid dynamics (CFD) simulations.⁷ They used a database of 1100 existing airfoils⁸ and enriched the database with 100,000 more generated airfoils. The data was used to create a surrogate model of force coefficients (lift, drag, and moment) with respect to flight speed (Mach number), the angle of attack, and airfoil shape variables. The shape design variables consisted of shape modes obtained by singular-value decomposition of the existing airfoil shapes. The approach proved successful, enabling shape optimization in 2 sec using a personal computer with an error of less than 0.25% for subsonic flight conditions. This is thousands of times faster than optimization using direct calls to CFD.

To validate the prediction derivatives provided by SMT, we built a surrogate model of the airfoil drag coefficient similar to the one above using the following four inputs: first thickness mode t_1 , first camber mode c_1 , Mach number M , and angle of attack α . We compute the partial derivatives of the drag coefficient (C_d) with respect to these four variables. The range of the first thickness and camber modes are such that we cover 55% of the airfoil database. The range of M is [0.3,0.4] and the range of α is [0,4]. To train the surrogate model, we use 20 training points with their respective training derivatives. We use the GE-KPLS model, where we provide training derivatives using an adjoint approach [24,25]. To validate the prediction derivatives of the constructed surrogate model, we generate 100 new points different from the training points and compute the relative error on those points using

$$\varepsilon = \frac{\|\frac{\partial C_d}{\partial x} - \hat{\frac{\partial C_d}{\partial x}}\|_2}{\|\frac{\partial C_d}{\partial x}\|_2}, \quad (20)$$

where $\hat{\frac{\partial C_d}{\partial x}}$ is the vector of 100 surrogate model prediction derivatives with respect to one of the four variables ($x = t_1, c_1, M, \alpha$), $\frac{\partial C_d}{\partial x}$ is the vector of corresponding reference derivative values with respect to x , and $\|\cdot\|_2$ is the L_2 norm. The resulting relative errors corresponding to each variable (listed in Table 2) show that the surrogate model provides an accurate estimate of the derivatives with a relative

⁷ <https://github.com/mdolab/adflow>

⁸ <http://webfoil.engin.umich.edu>

```

import numpy as np
from smt.problems.problem import Problem

class Sphere(Problem):
    # Class Sphere inherits class Problem

    # Set up methods of class
    def _initialize(self):
        self.options.declare('name', 'Sphere', types=str)

    def _setup(self):
        # Bornes of the sphere problem
        self.xlimits[:, 0] = -10.
        self.xlimits[:, 1] = 10.

    def _evaluate(self, x, kx):
        """
        Arguments
        -----
        x : ndarray[ne, nx]
            Evaluation points.
        kx : int or None
            Index of derivative (0-based) to return
            values with respect to. None means return
            function value rather than derivative.
        Returns
        -----
        ndarray[ne, 1]
            Functions values if kx=None or derivative
            values if kx is an int.
        """
        ne, nx = x.shape
        y = np.zeros((ne, 1), complex)
        if kx is None:
            y[:, 0] = np.sum(x**2, 1).T
        else:
            y[:, 0] = 2 * x[:, kx]
        return y

```

Fig. 4. Implementation of the *Sphere* class, which inherits from the *Problem* class. The main function is `_evaluate`, which computes either the output values or the derivatives depending on the variable `kx`.

error less or equal to 10^{-2} .

Using the constructed surrogate, we compare two gradient-based optimization algorithms:

Optimization 1 uses the surrogate model derivatives provided by SMT;

Optimization 2 uses finite-difference derivatives.

The optimization problem is to minimize the drag coefficient (C_d) subject to a lift constraint ($C_l = 0.5$). Optimization 1 achieved a slightly lower drag (0.03%) than Optimization 2 using 21 times fewer evaluations (Fig. 6).

5.2. Aircraft design optimization considering dynamics

The design optimization of commercial aircraft using CFD is typically done by modeling the aircraft performance at a small number of representative flight conditions [26,27]. This simplification is made because it would be prohibitively expensive to simulate the full, discretized mission profile using CFD at all points. However, full-mission simulation is sometimes necessary, such as when the flight is short or when considering morphing aircraft designs [28,29]. This requires a surrogate model of the aerodynamic performance as a function of a small number of parameters such as flight speed and altitude. In a design optimization context, the surrogate model must be retrained each

optimization iteration because the aircraft design changes from iteration to iteration as the shape is being optimized.

This is the situation in an aircraft allocation-mission-design optimization problem recently solved using RMTS [30]. In this work, the optimization problem maximized airline profit by optimizing the twist, span, sweep, and shape of the wing of a next-generation commercial aircraft. The profit was computed by simulating the fuel efficiency and flight time for the aircraft on a set of routes operated by the hypothetical airline. To do this, a surrogate model was generated for the wing lift and drag coefficients (C_L and C_D) as a function of the angle of attack (α), Mach number (M), and altitude (h). In each optimization iteration, the training C_L and C_D values were computed at a series of points in M - α - h space using CFD. The training outputs were used to retrain the RMTS surrogate model, and the mission simulations for all the airline routes were performed using inexpensive evaluations of the trained surrogate model, as shown in Fig. 7. However, because the overall optimization problem was solved using a gradient-based algorithm, we required derivatives of the prediction outputs (C_L and C_D values at the M - α - h points at which we evaluated the surrogate) with respect to the training outputs (C_L and C_D values at the fixed M - α - h points where the training points were located).

Figure 8 shows the surrogate model with the altitude axis eliminated by projecting onto the other two axes, where C_L replaces α as one of the inputs. The horizontal and vertical axes represent the inputs. The red points are the training points and the black points are the prediction

```

"""
Example solving an SMT benchmark problem using the
Regularized Minimal-energy Tensor-product
Cubic hermite spline surrogate model
"""
from smt.problems import RobotArm
from smt.sampling_methods import LHS
from smt.surrogate_models import RMTC
import numpy as np

# Sample the training points
fun = RobotArm(ndim=2)
sampling = LHS(xlimits=fun.xlimits)
xt = sampling(750)
yt = fun(xt)
for i in range(fun.options['ndim']):
    # Derivative with respect to the i-th variable
    yd = fun(xt,kx=i)
    yt = np.concatenate((yt,yd),axis=1)

# Construct the RMTC model
t = RMTC(xlimits=fun.xlimits, min_energy=True, \
         nonlinear_maxiter=20)
t.set_training_values(xt,yt[:,0])
for i in range(fun.options['ndim']):
    # Add the gradient information to the
    # sampling points
    t.set_training_derivatives(xt,yt[:,1+i],i)
t.train()

# Predict function values
xtest = sampling(5000)
yp = t.predict_values(xtest)

# Predict the derivative values
for i in range(fun.options['ndim']):
    ydp = t.predict_derivatives(xtest,i)

# Predict the derivatives with respect to the
# training points
ytdp = t.predict_output_derivatives(xtest)

```

Fig. 5. Example of use of RMTC within SMT on the robot arm function [23].

Table 2

Relative error of the prediction derivative of C_d with respect to t_1 , c_1 , M , and α . The surrogate model yields an accurate prediction derivative with an error less than or equal to 10^{-2} .

	$\partial C_d / \partial t_1$	$\partial C_d / \partial c_1$	$\partial C_d / \partial M$	$\partial C_d / \partial \alpha$
ϵ	0.010	0.007	0.009	0.002

points. Therefore, the output derivatives are the derivatives of the outputs at the black points with respect to the training outputs at the red points.

5.3. Other applications

In addition to the two applications we just described, SMT has been used to solve other engineering problems. We summarize these applications in Table 3. The number of input variables ranges from 2 to 99. This demonstrates SMT's ability to solve different engineering problems of various complexities.

6. Conclusions

SMT is unique compared with existing surrogate modeling libraries in that it is designed from the ground up to handle derivative

information effectively. The derivative information includes training derivatives used for gradient-enhanced modeling, prediction derivatives used for surrogate-model-based optimization, and derivatives with respect to the training data used when the optimization loop includes reconstructing the surrogate model. However, SMT does not need to involve derivatives and provides a simple general interface to various surrogate modeling techniques. Together with the provided problems, SMT is a convenient framework to develop new methods for sampling and surrogate modeling, and then benchmarking them against existing methods.

SMT arose from our own research needs, namely the development of the GE-KPLS and RMTS surrogate models, which require the handling of derivatives. GE-KPLS is a method we recently developed that scales much better with the number of inputs than the other methods, thanks in part to the use of the derivative information. RMTS is another method that we recently developed that is best suited to low-dimensional problems and handles both structured and unstructured training points.

SMT is distributed through an open-source license and is freely available online⁹. We provide documentation that caters to both users and potential developers.¹⁰ SMT allows users collaborating on the same

⁹ <https://github.com/SMTorg/SMT>

¹⁰ <http://smt.readthedocs.io/en/latest/>

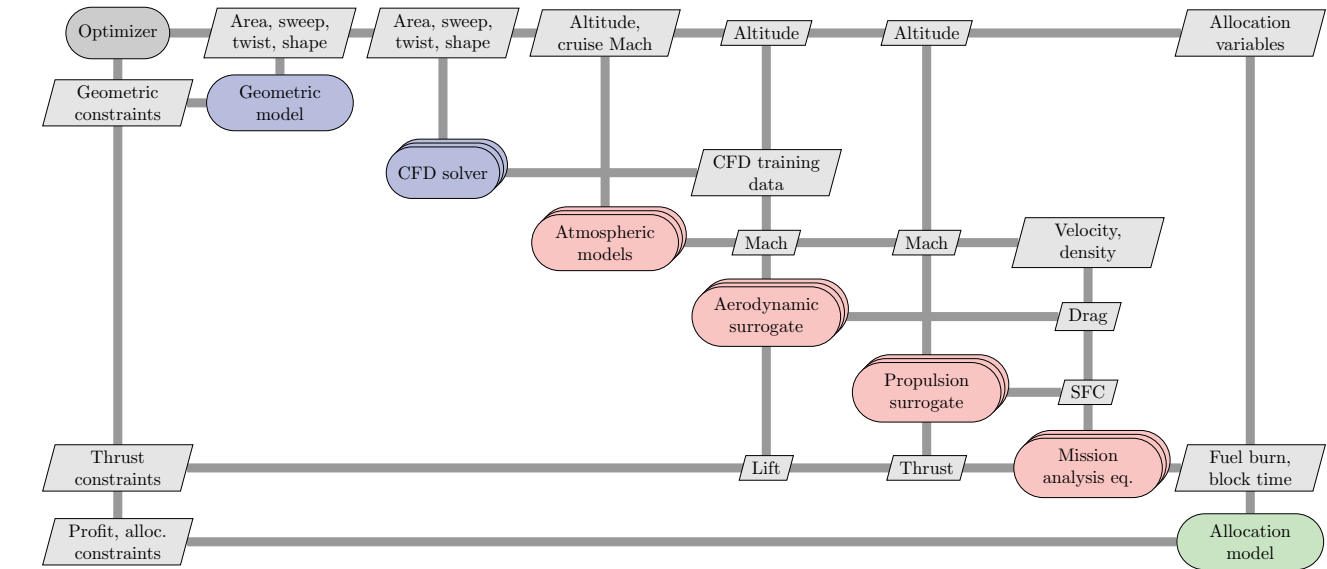
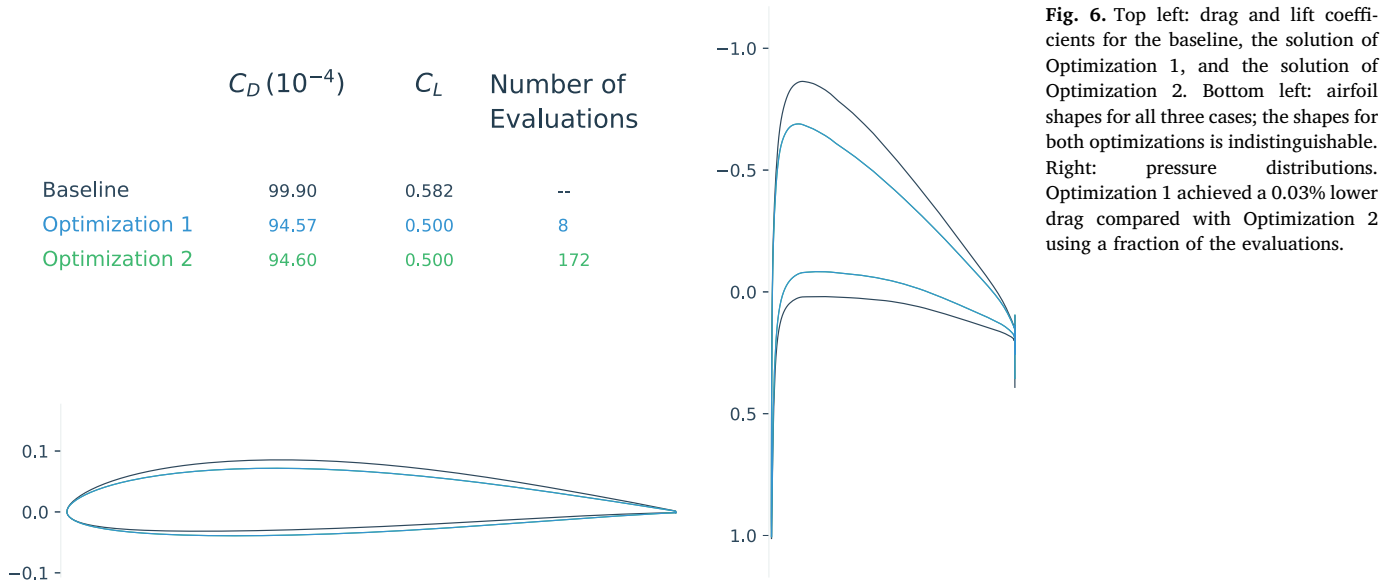


Fig. 7. Extended design structure matrix [31] for the aircraft allocation-mission-design optimization problem that used RMTS [30]. The aerodynamic surrogate model predicts wing lift and drag coefficient as a function of the angle of attack, Mach number, and altitude. This surrogate model is retrained each iteration because the lift and drag training data change each optimization iteration as the aircraft wing design changes. Since a gradient-based algorithm is used for this large-scale optimization problem, we require output derivatives: derivatives of the lift and drag prediction outputs with respect to the lift and drag training data. These are provided by RMTS.

project to have a common surrogate modeling tool that facilitates the exchange of methods and reproducibility of results.

SMT is a recently developed tool and, so far, it has primarily been used in aerospace engineering applications. However, this tool is useful to anyone needing to use or develop surrogate modeling techniques, regardless of the application. In particular, given the advantages of the newer surrogate modeling techniques (GE-KPLS and RMTS), using SMT is particularly attractive.

Acknowledgments

This work was partially supported by the Air Force Office of Scientific Research (AFOSR) MURI on “Managing multiple information sources of multi-physics systems,” Program Officer Jean-Luc Cambier, Award Number FA9550-15-1-0038. This work is part of the activities of ONERA–ISAE–ENAC joint research group and was partially supported by an ONERA internal project on multi-fidelity methods (MUFIN). The authors would like to thank Rémi Vauclin and Mostafa Meliani for the implementation of MFK, and Rémy Priem for the implementation of MoE.

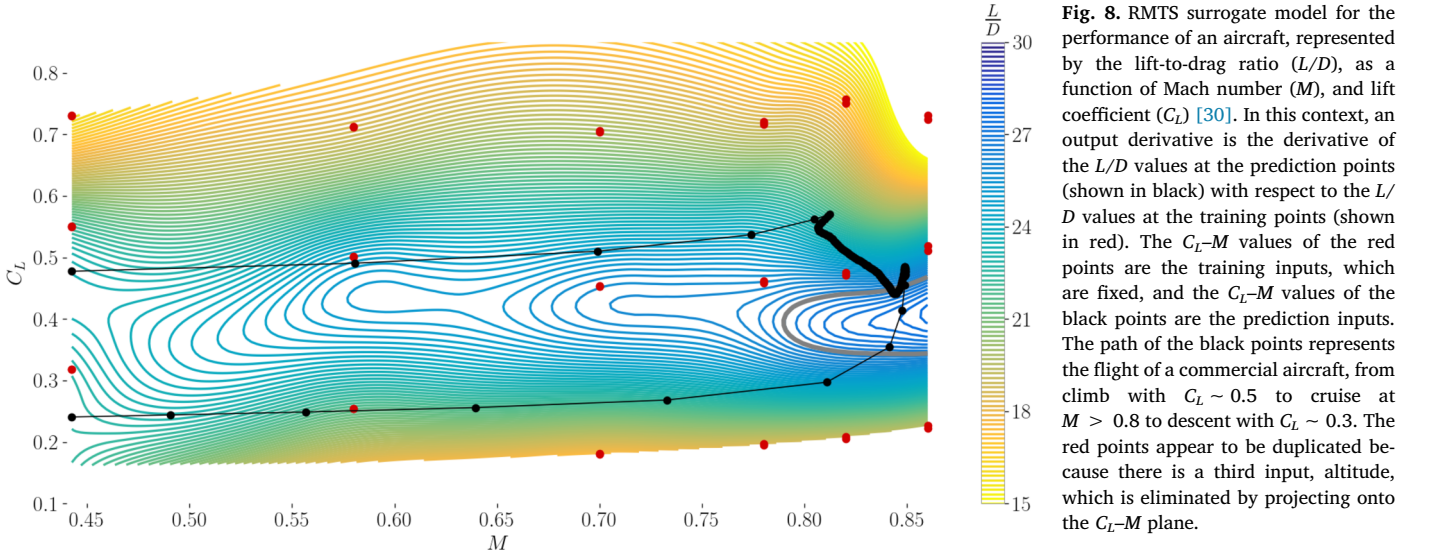


Fig. 8. RMTS surrogate model for the performance of an aircraft, represented by the lift-to-drag ratio (L/D), as a function of Mach number (M), and lift coefficient (C_L) [30]. In this context, an output derivative is the derivative of the L/D values at the prediction points (shown in black) with respect to the L/D values at the training points (shown in red). The C_L - M values of the red points are the training inputs, which are fixed, and the C_L - M values of the black points are the prediction inputs. The path of the black points represents the flight of a commercial aircraft, from climb with $C_L \sim 0.5$ to cruise at $M > 0.8$ to descent with $C_L \sim 0.3$. The red points appear to be duplicated because there is a third input, altitude, which is eliminated by projecting onto the C_L - M plane.

Table 3
Summary of SMT applications to engineering design problems.

Problem	Surrogate used	Objective of the study	Number of design variables	Reference
3D blade	KPLS	Compare the accuracy and efficiency of the KPLS model and the Optimus [32] implementation of the kriging model	Up to 99	Bouhleh et al. [4]
3D blade	KPLSK	Compare the accuracy and efficiency of the KPLSK model and the KPLS model	99	Bouhleh et al. [5]
Automotive	KPLS, KPLSK	Minimize the mass of a vehicle subject to 68 constraints	50	Bouhleh et al. [33]
Eight different engineering problems	GE-KPLS	Compare the accuracy and efficiency of the GE-KPLS model and the indirect GEK	Up to 15	Bouhleh and Martins [6]
Aircraft wing	MoE	Minimize the drag of aircraft wings subject to lift constraints	Up to 17	Bartoli et al. [15]
2D airfoils	GE-KPLS	Build a fast interactive airfoil analysis and design optimization tool	Up to 16	Li et al. [14]
Aircraft performance	RMTS	Optimize an aircraft design while analyzing the full mission using a CFD surrogate	3	Hwang and Munster [30]
Rotor analysis	RMTS	Optimize a rotor using a surrogate model for the lift and drag of blade airfoils	2	Hwang and Ning [34]

Appendix A. Analytical functions implemented in SMT

In the following descriptions, n_x is the number of dimensions.

Sphere. The sphere function is quadratic, continuous, convex, and unimodal. It is given by

$$\sum_{i=1}^{n_x} x_i^2, \quad -10 \leq x_i \leq 10, \quad \text{for } i = 1, \dots, n_x.$$

Branin [35]. The Branin function is commonly used in optimization and has three global minima. It is given by

$$f(x) = \left(x_2 - \frac{5.1}{4\pi^2} x_1^2 + \frac{5}{\pi} x_1 - 6 \right)^2 + 10 \left(1 - \frac{1}{8\pi} \right) \cos(x_1) + 10,$$

where $x = (x_1, x_2)$ with $-5 \leq x_1 \leq 10$ and $0 \leq x_2 \leq 15$. Rosenbrock [36]. The Rosenbrock function is a continuous, nonlinear, and non-convex function and usually used in optimization. The minimum of this function is situated in a parabolic-shaped flat valley and is given by

$$\sum_{i=1}^{n_x-1} [(x_{i+1} - x_i^2)^2 + (x_i - 1)^2], \quad -2 \leq x_i \leq 2, \quad \text{for } i = 1, \dots, n_x.$$

Tensor-product. The tensor-product function approximates a step function, which causes oscillations with some surrogate models [7], and is given by

$$\prod_{i=1}^{n_x} \cos(a\pi x_i), \quad -1 \leq x_i \leq 1, \quad \text{for } i = 1, \dots, n_x,$$

or

$$\prod_{i=1}^{n_x} \exp(x_i), \quad -1 \leq x_i \leq 1, \quad \text{for } i = 1, \dots, n_x,$$

or

$$\prod_{i=1}^{n_x} \tanh(x_i), \quad -1 \leq x_i \leq 1, \quad \text{for } i = 1, \dots, n_x,$$

or

$$\prod_{i=1}^{n_x} \exp(-2x_i^2), \quad -1 \leq x_i \leq 1, \quad \text{for } i = 1, \dots, n_x.$$

Appendix B. Engineering problems implemented in SMT

Cantilever beam [37]

This function models a simple uniform cantilever beam with vertical and horizontal loads

$$\frac{50}{600} \sum_{i=1}^{17} \left[\frac{12}{b_i h_i^3} \left(\left(\sum_{j=i}^{17} l_j \right)^3 - \left(\sum_{j=i+1}^{17} l_j \right)^3 \right) \right],$$

where $b_i \in [0.01, 0.05]$, $h_i \in [0.3, 0.65]$, $l_i \in [0.5, 1]$.

Robot arm [23]

This function gives the position of a robot arm

$$\sqrt{\left(\sum_{i=1}^4 L_i \cos \left(\sum_{j=1}^i \theta_j \right) \right)^2 + \left(\sum_{i=1}^4 L_i \sin \left(\sum_{j=1}^i \theta_j \right) \right)^2},$$

where $L_i \in [0, 1]$ for $i = 1, \dots, 4$ and $\theta_j \in [0, 2\pi]$ for $j = 1, \dots, 4$.

Torsion vibration [38]

This function gives the low natural frequency of a torsion problem

$$\frac{1}{2\pi} \sqrt{\frac{-b - \sqrt{b^2 - 4ac}}{2a}},$$

where $K_i = \frac{\pi G_i d_i}{32 L_i}$, $M_j = \frac{\rho_j \pi l_j D_j}{4g}$, $J_j = 0.5 M_j \frac{D_j^4}{2}$, $a = 1$, $b = -\left(\frac{K_1 + K_2}{J_1} + \frac{K_2 + K_3}{J_2} \right)$, $c = \frac{K_1 K_2 + K_2 K_3 + K_3 K_1}{J_1 J_2}$, for $d_1 \in [1.8, 2.2]$, $L_1 \in [9, 11]$, $G_1 \in [105300000, 128700000]$, $d_2 \in [1.638, 2.002]$, $L_2 \in [10.8, 13.2]$, $G_2 \in [5580000, 6820000]$, $d_3 \in [2.025, 2.475]$, $L_3 \in [7.2, 8.8]$, $G_3 \in [3510000, 4290000]$, $D_1 \in [10.8, 13.2]$, $t_1 \in [2.7, 3.3]$, $\rho_1 \in [0.252, 0.308]$, $D_2 \in [12.6, 15.4]$, $t_2 \in [3.6, 4.4]$, and $\rho_1 \in [0.09, 0.11]$.

Water flow [39]

This function characterizes the flow of water through a borehole that is drilled from the ground surface through two aquifers

$$\frac{2\pi T_u (H_u - H_l)}{\ln \left(\frac{r}{r_w} \right) \left[1 + \frac{2LT_u}{\ln \left(\frac{r}{r_w} \right)^2 K_w} + \frac{T_u}{T_l} \right]},$$

where $0.05 \leq r_w \leq 0.15$, $100 \leq r \leq 50000$, $63070 \leq T_u \leq 115600$, $990 \leq H_u \leq 1110$, $63.1 \leq T_l \leq 116$, $700 \leq H_l \leq 820$, $1120 \leq L \leq 1680$, and $9855 \leq K_w \leq 12045$.

Welded beam [40]

The shear stress of a welded beam problem is given by

$$\frac{\sqrt{\tau'^2 + \tau''^2 + l\tau'\tau''}}{\sqrt{0.25(l^2 + (h+t)^2)}}$$

where $\tau' = \frac{6000}{\sqrt{2}hl}$, $\tau'' = \frac{6000(14+0.5l)\sqrt{0.25(l^2+(h+t)^2)}}{2\left[0.707hl\left(\frac{l^2}{12}+0.25(h+t)^2\right)\right]}$, for $h \in [0.125, 1]$, and $l, t \in [5, 10]$.

Wing weight [35]

The estimate of the weight of a light aircraft wing is given by

$$0.036S_w^{0.758}W_{fw}^{0.0035}\left(\frac{A}{\cos^2\Lambda}\right)q^{0.006}\lambda^{0.04}\left(\frac{100t_c}{\cos\Lambda}\right)^{-0.3}(N_zW_{dg})^{0.49}+S_wW_p,$$

where $150 \leq S_w \leq 200$, $220 \leq W_{fw} \leq 300$, $6 \leq A \leq 10$, $-10 \leq \Lambda \leq 10$, $16 \leq q \leq 45$, $0.5 \leq \lambda \leq 1$, $0.08 \leq t_c \leq 0.18$, $2.5 \leq N_z \leq 6$, $1700 \leq W_{dg} \leq 25000$, and $0.025 \leq W_p \leq 0.08$.

References

- [1] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. *J Mach Learn Res* 2011;12:2825–30.
- [2] Gorissen D, Crombecq K, Couckuyt I, Dhaene T, Demeester P. A surrogate modeling and adaptive sampling toolbox for computer based design. *J Mach Learn Res* 2010;11:2051–5.
- [3] Rasmussen CE, Williams CKI. *Gaussian processes for machine learning*. Adaptive Computation and Machine Learning Cambridge, MA, USA: MIT Press; 2006.
- [4] Bouhleh MA, Bartoli N, Otsmane A, Morlier J. Improving kriging surrogates of high-dimensional design models by partial least squares dimension reduction. *Struct Multidiscip Optim* 2016;53(5):935–52. <https://doi.org/10.1007/s00158-015-1395-9>.
- [5] Bouhleh MA, Bartoli N, Morlier J, Otsmane A. An improved approach for estimating the hyperparameters of the kriging model for high-dimensional problems through the partial least squares method. *Math Probl Eng* 2016. <https://doi.org/10.1155/2016/6723410>. Article ID 6723410
- [6] Bouhleh MA, Martins JRRA. Gradient-enhanced kriging for high-dimensional problems. *Eng Comput* 2019;1(35):157–73. <https://doi.org/10.1007/s00366-018-0590-x>.
- [7] Hwang JT, Martins JRRA. A fast-prediction surrogate model for large datasets. *Aerosp Sci Technol* 2018;75:74–87. <https://doi.org/10.1016/j.ast.2017.12.030>.
- [8] Jin R, Chen W, Sudjianto A. An efficient algorithm for constructing optimal design of computer experiments. *J Stat Plan Inference* 2005;134(1):268–87. <https://doi.org/10.1016/j.jspi.2004.02.014>.
- [9] Sacks J, Schiller SB, Welch WJ. Designs for computer experiments. *Technometrics* 1989;31(1):41–7.
- [10] Powell MJD. *The theory of radial basis function approximation in 1990*. Oxford University Press; 1992. p. 105–210.
- [11] Shepard D. A two-dimensional interpolation function for irregularly-spaced data. Proceedings of the twenty-third ACM national conference, ACM '68. New York, NY, USA: ACM; 1968. p. 517–24. <https://doi.org/10.1145/800186.810616>.
- [12] Hastie T, Tibshirani R, Friedman J. *The elements of statistical learning*. Springer Series in Statistics New York, NY, USA: Springer New York Inc.; 2001.
- [13] Powell MJD. Dordrecht: Springer Netherlands; 1994. p. 51–67. ISBN 978-94-015-8330-5
- [14] Li J, Bouhleh MA, Martins JRRA. Data-based approach for fast airfoil analysis and optimization. *J Aircr* 2019;57(2):581–96. <https://doi.org/10.2514/1.J057129>.
- [15] Bartoli N, Lefebvre T, Dubreuil S, Olivanti R, Priem R, Bons N, Martins JRRA, Morlier J. Adaptive modeling strategy for constrained global optimization with application to aerodynamic wing design. *Aerosp Sci Technol* 2019. (In press).
- [16] Gray JS, Hwang JT, Martins JRRA, Moore KT, Naylor BA. OpenMDAO: an open-source framework for multidisciplinary design, analysis, and optimization. *Structural and Multidisciplinary Optimization* 2019. <https://doi.org/10.1007/s00158-019-02211-z>.
- [17] Martins JRRA, Lambe AB. Multidisciplinary design optimization: a survey of architectures. *AIAA J* 2013;51(9):2049–75. <https://doi.org/10.2514/1.J051895>.
- [18] Martins JRRA, Hwang JT. Review and unification of methods for computing derivatives of multidisciplinary computational models. *AIAA J* 2013;51(11):2582–99. <https://doi.org/10.2514/1.J052184>.
- [19] Hwang JT, Martins JRRA. A computational architecture for coupling heterogeneous numerical models and computing coupled derivatives. *ACM Trans Math Softw* 2018;44(4):Article37. <https://doi.org/10.1145/3182393>.
- [20] Bettebghor D, Bartoli N, Grihon S, Morlier J, Samuelides M. Surrogate modeling approximation using a mixture of experts based on em joint estimation. *Struct Multidiscip Optim* 2011;43(2):243–59. <https://doi.org/10.1007/s00158-010-0554-2>.
- [21] Han Z-H, Görtz S, Zimmermann R. Improving variable-fidelity surrogate modeling via gradient-enhanced kriging and a generalized hybrid bridge function. *Aerosp Sci Technol* 2013;25(1):177–89. <https://doi.org/10.1016/j.ast.2012.01.006>.
- [22] Le Gratiet L. Multi-fidelity gaussian process regression for computer experiments. Université Paris-Diderot - Paris VII; 2013. Theses. <https://tel.archives-ouvertes.fr/tel-00866770>
- [23] An J, Owen A. Quasi-regression. *J Complex* 2001;17(4):588–607. <https://doi.org/10.1006/jcom.2001.0588>.
- [24] Mader CA, Martins JRRA, Alonso JJ, van der Weide E. Adjoint: an approach for the rapid development of discrete adjoint solvers. *AIAA J* 2008;46(4):863–73. <https://doi.org/10.2514/1.29123>.
- [25] Lyu Z, Kenway GK, Paige C, Martins JRRA. Automatic differentiation adjoint of the Reynolds-averaged Navier–Stokes equations with a turbulence model. In: Proceedings of the twenty-first AIAA computational fluid dynamics conference. San Diego, CA; doi:10.2514/6.2013-2581.
- [26] Kenway GW, Martins JRRA. Multipoint aerodynamic shape optimization investigations of the common research model wing. *AIAA J* 2016;54(1):113–28. <https://doi.org/10.2514/1.J054154>.
- [27] Kenway GW, Martins JRRA. Multipoint high-fidelity aerostructural optimization of a transport aircraft configuration. *J Aircr* 2014;51(1):144–60. <https://doi.org/10.2514/1.C032150>.
- [28] Martins JRRA. Fuel burn reduction through wing morphing. *Encyclopedia of aerospace engineering*. Green Aviation. Wiley; 2016. p. 75–9. <https://doi.org/10.1002/9780470686652.eae1007>. ISBN 978-1-118-86635-1
- [29] Burdette DA, Martins JRRA. Impact of morphing trailing edge on mission performance for the common research model. *J Aircr* 2019;56(1):369–84. <https://doi.org/10.2514/1.C034967>.
- [30] Hwang JT, Jasa J, Martins JRRA. High-fidelity design-allocation optimization of a commercial aircraft maximizing airline profit. *J Aircr* 2019. <https://doi.org/10.2514/1.C035082>.
- [31] Lambe AB, Martins JRRA. Extensions to the design structure matrix for the de-structuring of multidisciplinary design, analysis, and optimization processes. *Struct Multidiscip Optim* 2012;46:273–84. <https://doi.org/10.1007/s00158-012-0763-y>.
- [32] Noesis Solutions. OPTIMUS. 2009. <http://www.noesisolutions.com/Noesis/optimus-details/optimus-design-optimization>.
- [33] Bouhleh MA, Bartoli N, Regis RG, Otsmane A, Morlier J. Efficient global optimization for high-dimensional constrained problems by using the kriging models combined with the partial least squares method. *Eng Optim* 2018;50(12):2038–53. <https://doi.org/10.1080/0305215X.2017.1419344>.
- [34] Hwang JT, Ning A. Large-scale multidisciplinary optimization of an electric aircraft for on-demand mobility. Proceedings of the 2018 AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference. Kissimmee, FL. 2018. <https://doi.org/10.2514/6.2018-1384>.
- [35] Forrester AJJ, Sobester A, Keane AJ. *Engineering design via surrogate modelling-A practical guide..* Wiley; 2008. ISBN 978-0-470-06068-1
- [36] Shang Y, Qiu Y. A note on the extended rOsenbrock function. *Evol Comput* 2006;14(1):119–26. <https://doi.org/10.1162/106365606776022733>.
- [37] Cheng GH, Younis A, Hajikolaei KH, Wang GG. Trust region based mode pursuing sampling method for global optimization of high dimensional design problems. *J Mech Des* 2015;137(2):021407.
- [38] Liping W, Don B, Gene W, Mahidhar R. A comparison of metamodeling methods using practical industry requirements. Proceedings of the 47th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, Newport, RI. 2006.
- [39] Morris MD, Mitchell TJ, Ylvisaker D. Bayesian design and analysis of computer experiments: use of derivatives in surface prediction. *Technometrics* 1993;35(3):243–55.
- [40] Deb K. An efficient constraint handling method for genetic algorithms. *Comput Methods Appl Mech Eng* 1998;186(2–4):311–38. [https://doi.org/10.1016/S0045-7825\(99\)00389-8](https://doi.org/10.1016/S0045-7825(99)00389-8).

Chapitre 5

Activités d'encadrement

Sommaire

5.1	Activités d'encadrement d'étudiants	192
5.1.1	Encadrements de projets élèves	192
5.1.2	Encadrements de stage	192
5.2	Encadrements de thèse	195
5.3	Encadrements de post-doc	196

Ce chapitre présente la liste des encadrements que j'ai effectués : projets étudiants, stages, thèses et post-docs.

5.1 Activités d'encadrement d'étudiants

Dans le cadre de mes activités de recherche depuis 1997, j'ai été amenée à participer à l'encadrement ou au co-encadrement de projets étudiants, stages, thèses et post-docs. La liste n'est pas exhaustive et j'ai essayé de rassembler dans ce document les activités d'encadrement liées à mon projet de recherche autour de l'optimisation et des modèles réduits.

5.1.1 Encadrements de projets élèves

L'ensemble des encadrements porte principalement sur l'optimisation multidisciplinaire. Un projet a été proposé autour de l'optimisation multi-critères.

- Optimisation multidisciplinaire, ISAE-SUPAERO (filière Systèmes Complexes et MASTER MAE) et INSA (Génie Mathématique et Modélisation).

Depuis 2013, j'ai participé à l'encadrement de 10 projets étudiants (1 à 2 groupes d'élèves par an) pour mettre en place des algorithmes d'optimisation mono ou multi-objectifs (MGDA, CMAES, multiEGO, jeux de Nash) ou tester des stratégies multi-agents. Les développements sont faits en python pour être intégrés à la plate-forme OpenMDAO. Ces encadrements sont faits en collaboration avec T. Lefebvre (ONERA/DTIS), V. Picheny (INRA), C. Gogu (UPS-ICA), Y. Diouane, M. Salaun et J. Morlier de l'ISAE-SUPAERO. Certains de ces sujets ont été l'occasion d'échanger avec J.-A. Désidéri de l'INRIA Sophia Antipolis et F. Poirion (ONERA/DADS).

- Optimisation multicritère, ISAE-SUPAERO, filière Systèmes Complexes

Un sujet sur l'optimisation multicritères dans le cadre du projet interne PRF VESTA (Porteur anciennement DCPS) a été réalisé en utilisant le logiciel PROMETHEE en collaboration avec A. Joulia (ONERA/DTIS).

5.1.2 Encadrements de stage

Depuis 1998, je n'ai quasiment jamais cessé d'encadrer des étudiants en stage pour des durées allant de deux mois à six mois. Ces encadrements ont pour la plupart été des co-encadrements et ont fait l'objet de collaboration avec l'ISAE-SUPAERO, l'IRT Saint-Exupéry, l'Institut Clément Ader ou l'Université du Michigan.

2019

- Co-encadrement du stage de 2ème année de Alejandro Díaz Bolívar (ISAE-SUPAERO) avec T. Lefebvre (ONERA/DTIS), J. Morlier (ISAE-SUPAERO) et J.R.R.A. Martins (Univ. Michigan) sur de nouvelles applications de l'optimisation multi-fidélité basée sur des modèles réduits en aérodynamique (mars à juillet 2019).

2018

- Co-encadrement du stage ouvrier de T. Lahire (ISAE-SUPAERO) avec R. Priem (doctorant ONERA/DTIS) pour la mise en place d'un plan d'expérience avec enrichissement adaptatif (juin et juillet 2018).
- Co-encadrement du stage de fin d'études de B. Dabas (Ecole Polytechnique et ISAE-SUPAERO) avec T. Lefebvre et F. Gallard à l'IRT Saint-Exupéry pour la mise en place d'une formulation d'optimisation multidisciplinaire pour la conception avion (avril à septembre 2018). Ce stage a fait l'objet d'un abstract de 23 pages accepté pour la conférence AIAA/AVIATION de juin 2019 à Dallas [65].
- Co-encadrement du stage de fin d'études de M. Meliani (ISAE-ENSICA) avec J. Morlier (ISAE-SUPAERO) et J.R.R.A. Martins (Univ. Michigan) sur l'optimisation multi-fidélité basée sur des modèles réduits pour l'aérodynamique. Un séjour de 3 mois (1er juillet au 30 septembre 2018) a été réalisé à l'Université du Michigan au sein du MDOLab via le soutien de la Fondation ISAE-SUPAERO. Le stage a été pré-sélectionné pour représenter l'ISAE au concours Pegasus-AIAA [student Conference](#) en avril 2019 et un abstract de 18 pages [146] a été accepté à la conférence AIAA/AVIATION en juin 2019 à Dallas.
- Participation à l'encadrement du stage de fin d'études de G. Berthelin (ISAE-SUPAERO) avec S. Dubreuil, T. Lefebvre (ONERA/DTIS), M. Salaun (ISAE-SUPAERO) et C. Gogu (ICA) sur la propagation des incertitudes de modèles dans un couplage aéro-structure (avril à septembre 2018). Un abstract a été soumis à la conférence CSMA à Giens en juin 2019 [28].

2017

- Co-encadrement du stage de fin d'études de R. Olivanti (ISAE-SUPAERO) sur l'optimisation en grande dimension pour une aile d'avion avec T. Lefebvre (ONERA/DCPS) et J. Morlier (ISAE-SUPAERO), avril 2017 - oct 2017. Ce stage a fait l'objet d'un papier à la conférence AIAA/AVIATION à Atlanta en juin 2018 [22].
- Co-encadrement du stage de 2A de G. Dematté (ISAE-SUPAERO) sur la prise en compte de contraintes dans l'algorithme CMAES avec Y. Diouane (ISAE-SUPAERO), mai 2017 - juin 2017.
- Co-encadrement du stage de césure de M. Meliani (ISAE-ENSICA) sur la mise

en place d'un couplage entre la chaîne aérodynamique haute-fidélité et le processus avant-projet avec T. Lefebvre (ONERA/DCPS) et T. Druot (AIRBUS), IRT Saint-Exupéry, Projet MDA-MDO, avril 2017 - aout 2017.

- Co-encadrement du stage de césure de V. Fievez (ISAE-SUPAERO) sur la mise en place de cas tests de non régression dans une toolbox python de mélanges d'experts en partenariat avec R. Lafage (ONERA/DCPS), fev 2017 - juin 2017.

2016

- Co-encadrement du stage de fin d'études de M. Gibert (UTBM) sur l'adaptation des matrices de covariance pour le krigeage en partenariat avec Y. Diouane et M. Salaun (ISAE-SUPAERO), sept 2016 - fev 2017.

- Co-encadrement du stage de fin d'études de V. Stiliz (ISAE-ENSICA) sur l'optimisation en grande dimension en partenariat avec T. Lefebvre (ONERA/DCPS) et J. Morlier (ISAE-SUPAERO), fev 2016 - aout 2016. Ce stage a fait l'objet d'un papier à la conférence AIAA de juin 2016 à Denver [21].

- Co-encadrement du stage de 2A de H. Akmouch (ISAE-SUPAERO) sur l'Amélioration du conditionnement des matrices de krigeage avec Y. Diouane et M. Salaun (ISAE-SUPAERO), juin 2016 - aout 2016.

2015

- Co-encadrement du stage de césure de R. Priem (ISAE-ENSICA) sur le développement d'une toolbox python sur les mélanges d'experts en partenariat avec T. Lefebvre (ONERA/DCPS), sept 2015 - fev 2016.

- Co-encadrement du stage de fin d'études de I. Kurek (ISAE-ENSICA) sur l'optimisation structure pour la conception avion en partenariat avec T. Lefebvre (ONERA/DCPS) et J. Morlier (ISAE-SUPAERO), fev 2015 - aout 2015.

- Co-encadrement du stage de 1ere année de C. Leroy (ISAE-ENSICA) sur l'adaptation du module python de co-krigeage en partenariat avec R. Lafage (ONERA/DCPS), mars 2015 - juillet 2015.

- Co-encadrement du stage de 1ere année de F. Vidal (ISAE-SUPAERO) sur l'algorithme EGO dans OpenMDAO, juin-juillet 2015.

2014

- Co-encadrement du stage de fin d'études de C. Avilla Swistunowicz (Univ Politecnica de Madrid) sur l'optimisation multidisciplinaire pour la conception avion en partenariat avec T. Lefebvre (ONERA/DCPS), fev 2014 - aout 2014.

- Co-encadrement du stage de fin d'études de R. Vauclin (Mines St Etienne) sur le développement d'un module de co-krigeage pour la multi-fidélité en partenariat avec J. Morlier (ISAE-SUPAERO), avril 2014 - septembre 2014.

2013

- Co-encadrement du stage de fin d'études de M. Fielder (Univ Hambourg) sur le couplage multi-fidélité pour la conception préliminaire avion en partenariat avec T. Lefebvre (ONERA/DCPS) oct 2012 - fev 2013.

Avant 2010

- Co-encadrement du stage de fin d'études - DEA de L. Boussof (INSA Toulouse) sur les modèles réduits robustes pour l'optimisation multi-objectifs, 2008.
- Co-encadrement du stage ingénieur de H. Busson (INSA Toulouse) sur l'étude des approches space-mapping pour la construction de surfaces de réponses, été 2006.
- Co-encadrement de 7 stagiaires (ingénieur-DEA) au CERFACS (étés 1998 à 2004).

5.2 Encadrements de thèse

Dès mon arrivée à l'ONERA en 2005, j'ai été amenée à participer à l'encadrement de doctorants. La liste de ces co-encadrements est donnée ci-dessous et les activités de recherche associées à ces thèses sont décrites dans le chapitre 6.11. A ce jour, six thèses ont été soutenues, trois sont en cours et deux nouvelles thèses vont démarrer courant 2019. Le chapitre 6.11 détaille chacun des sujets de thèse.

Thèses en cours

- Co-directrice (25%) de la thèse CIFRE avec ATR de V. Palladino avec P. Schmollgruber (ONERA), V. Budinger et E. Benard (ISAE-SUPAERO), *Multi-fidelity design process applied to a hybrid electric regional aircraft*, 2019-2022.
- Co-directrice (33%) de la thèse de M. Saporito avec P. Schmollgruber (ONERA) et A. Da Ronch (Université de Southampton), *Robust Analysis and Optimization Process for Virtual Flight Simulation of Very-flexible Aircraft*, 2018-2021.
- Participation à l'encadrement de la thèse de G. Berthelin avec M. Salaün (ISAE-SUPAERO), S. Dubreuil (ONERA) et C. Gogu (ICA-UPS Toulouse), 2018-2021
- Co-directrice (25%) de la thèse de M. Menz avec C. Gogu (ICA-UPS Toulouse), S. Dubreuil et J. Morio (ONERA), *Vers une synergie entre analyses de fiabilité par échantillonnage adaptatif et réduction de modèles*, 2017-2020.
- Directrice (50%) de la thèse de R. Priem avec Y. Diouane (ISAE-SUPAERO), *Optimisation multidisciplinaire sous contraintes en grande dimension pour la concep-*

tion avion avant-projet, 2017-2020.

- Co-directrice (25%) de la thèse de A. Sgueglia avec J. Morlier, E. Benard (ISAE-SUPAERO) et P. Schmollgruber (ONERA), *Exploration of sizing and optimization priorities in Aircraft Design with application to a Blended Wing Body with Distributed Electric Propulsion*, 2017-2019. La soutenance est prévue en décembre 2019.

Thèses soutenues

- Co-directrice (50%) de la thèse de J. Mas Colomer avec J. Morlier (ISAE-SUPAERO), *Aeroelastic Similarity of a Flight Demonstrator via Multidisciplinary Optimization*, 2015-2018. La thèse a été soutenue en décembre 2018 [59].
- Co-directrice (50%) de la thèse de P. Schmollgruber avec Y. Gourinat (ISAE-SUPAERO), *Enhancement of the aircraft design process through certification constraints management and full mission simulations*, 2012-2018. La thèse a été soutenue en décembre 2018 [180].
- Co-encadrement (33%) de la thèse de M. Gruet avec A. Sicard (ONERA/DPHY), S. Rochel (ONERA/DPHY) et G. Rolland (CNES), *Intelligence Artificielle et prévision de l'impact de l'activité solaire sur l'environnement magnétique terrestre*, 2015-2018. La thèse a été soutenue en septembre 2018 [96].
- Co-directrice (50%) de la thèse de M-A. Bouhrel avec J. Morlier (ISAE-SUPAERO) et A. Otsmane (SNECMA), 2012-2016, *Optimisation auto-adaptative en environnement d'analyse multidisciplinaire via les modèles de krigeage combinés à la méthode PLS*, ISAE janvier 2016 [40].
- Co-encadrement (33%) de la thèse de D. Bettebghor avec M. Samuelides (ONERA-ISAE), S. Grihon (AIRBUS) et J. Morlier (ISAE-SUPAERO), 2008-2011, *Optimisation biniveau de structures aéronautiques composites*, ISAE 2011 [34].
- Co-encadrement (50%) de la thèse de J. Clément avec M. Masmoudi (MIP-UPS Toulouse) et J. Hermetz (DPRS), 2005-2009, *Optimisation multi-disciplinaire : étude théorique et appliquée*, ISAE 2009 [56].

5.3 Encadrements de post-doc

Depuis mon arrivée à l'ONERA, j'ai eu la possibilité de co-encadrer deux post-doctorants. Plus de détails sur les réalisations faites sont donnés dans le chapitre 6.11.

- Co-encadrement du post-doc de S. Dubreuil avec T. Lefebvre (ONERA/DTIS) sur la prise en compte des incertitudes de modèles dans un processus d'optimisation multidisciplinaire, 2015-2017.

- Participation à l'encadrement du post-doc de F. Luraghi avec M. Samuelides (ONERA/DTIM) et J. Morlier (ISAE-SUPAERO) sur l'optimisation des fibres composites courbes, 2011.

Chapitre 6

Activités d'encadrement de thèses et post-docs

Sommaire

6.1	Thèse de R. Priem	200
6.2	Thèse de M. Menz	201
6.3	Thèse de A. Sgueglia	203
6.4	Thèse de J. Mas Colomer	204
6.5	Thèse de P. Schmollgruber	206
6.6	Thèse de M. Gruet	207
6.7	Thèse de M.-A. Bouhlel	209
6.8	Thèse de D. Bettebghor	211
6.9	Thèse de J. Clément	213
6.10	Post-Doc de S. Dubreuil	214
6.11	Post-Doc de F. Luraghi	216

Ce chapitre détaille les thèses et post-docs dans lesquels je me suis impliquée depuis 2005.

Depuis mon arrivée à l'ONERA en 2005, j'ai eu la possibilité de participer à l'encadrement de dix doctorants et de deux post-doctorants. A ce jour, 6 thèses ont été soutenues, 4 thèses sont en cours et 2 doivent démarrer courant 2019. Dans les paragraphes qui suivent, je vais donc décrire rapidement pour chacun de ces encadrements, le contexte, les objectifs scientifiques et les principaux résultats obtenus ou attendus et je terminerai par la production scientifique associée.

Dans un souci d'amélioration continue, j'ai effectué la formation Encadrement de doctorants proposée par l'Association Bernard Grégory en décembre 2018.

6.1 Thèse de Rémy Priem : *Optimisation multidisciplinaire sous contraintes en grande dimension pour la conception avion avant-projet, 2017-2020*

Je dirige la thèse de Rémy Priem via une dérogation de l'école doctorale EDAA en collaboration avec Y. Diouane (ISAE-SUPAERO). La thèse de Rémy a démarré en novembre 2017 via une bourse en cofinancement ONERA - ISAE-SUPAERO.

Les méthodes d'optimisation bayésienne, comme la méthode Efficient Global Optimisation (EGO) et son extension aux contraintes présentée dans la Section 4.4.6 permettent d'améliorer la qualité du modèle de substitution dans les zones potentiellement intéressantes et de réaliser l'optimisation en ajoutant des points à la base de données. L'utilisation de la méthode est efficace pour des problèmes avec une cinquantaine de variables et une dizaine de contraintes. Dans leur version actuelle, ces approches ne peuvent cependant pas gérer le grand nombre de variables et de contraintes (une centaine). La thèse va donc tenter de répondre à ces deux verrous.

Les premiers 18 mois de la thèse ont permis de répondre à la problématique du traitement des contraintes. Dans la version initiale de SEGOMOE, seule la prédiction du modèle du krigeage est utilisée, l'information de la variance n'étant pas considérée. L'idée a été de relaxer les contraintes en utilisant un facteur de confiance et l'estimation d'erreur fournie par la variance et ainsi transformer la contrainte en Upper Trust Bound (UTB). Cette modification permet en outre de détecter les zones faisables avec une probabilité facilement calculable grâce aux propriétés des distributions gaussiennes. On a pu montrer, à partir de nombreux tests sur des problèmes de la littérature (Branin modifié, G07, PVD4, ...), que le

critère utilisant le facteur de confiance (SEGO-UTB) était le plus prometteur. Les performances de SEGO-UTB ont également été améliorées en faisant décroître le facteur de confiance (de 3 vers 0) au cours de l'optimisation. Cette tendance autorise l'exploration de zones faisables selon UTB mais qui ne le sont pas au regard de la version SEGO classique basée uniquement sur la prédiction. Les tests ont montré l'efficacité de la méthode proposée par rapport à la version SEGO classique et à la méthode ALBO [166] basée sur le Lagrangien Augmenté.

Une mobilité est prévue entre mai 2019 et septembre 2019 chez Bombardier au CANADA afin de tester l'algorithme développé sur des problèmes industriels de conception avion préliminaire avec une cinquantaine de variables et de contraintes.

Le deuxième volet du travail de thèse porte sur la prise en compte d'un grand nombre de variables de conception dans un problème d'optimisation bayésienne. Pour cela, la théorie des sous espace actifs [200] va être étudiée afin de l'adapter dans un cas plus général.

Communications :

- Use of Upper Trust Bounds in Constrained Bayesian Optimization Infill Criteria, In 20th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, Dallas June 2019
- An upper trust bound feasibility criterion for constrained Bayesian optimization, Optimizations Days 2019, HEC Montréal, Canada, May 2019
- Adaptive feasibility function for constrained Bayesian optimization with application in aircraft design, ENGOPT, Septembre 2018
- SEGOMOE : Super Efficient Global Optimization with Mixture of Experts, CIMI, Septembre 2018
- 1er prix du challenge [AGILE academy 2018](#)

6.2 Thèse de Morgane Menz : *Vers une synergie entre analyses de fiabilité par échantillonnage adaptatif et réduction de modèles*, 2017-2020

La thèse de Morgane Menz a démarré en octobre 2017 via une bourse en cofinancement Université Paul Sabatier et ONERA. Le directeur de thèse est Christian Gogu (ICA-UPS Toulouse), je codirige cette thèse avec Jérôme Morio et Sylvain Dubreuil de l'ONERA.

Les problèmes d'analyse de fiabilité sur des structures complexes sont souvent

numériquement coûteux à résoudre, ce qui en limite l'utilisation dans un processus d'optimisation. La première grande étape de la thèse est donc de réduire le coût numérique en couplant deux approches disjointes. D'une part des développements récents dans le domaine des approches d'analyse de fiabilité par apprentissage actif [37, 81, 13], ont permis des réductions importantes des temps de calcul pour des problèmes avec un nombre modéré de variables aléatoires. Les méthodes par apprentissage actif (AKMCS), connues aussi sous le nom de méthodes par échantillonnage adaptatif, consistent à construire un modèle de substitution (ou métamodèle) de krigeage (processus gaussien conditionné) pour le calcul de contraintes fiabilistes et enrichir de manière adaptative ce métamodèle en fonction de la structure d'incertitude du krigeage. D'autre part, l'utilisation de la réduction de modèles a également permis des gains importants en temps de calcul. Nous nous intéressons ici à la réduction de modèles appelée "en base réduite", qui consiste à rechercher une solution dans un sous-espace judicieusement construit, qui est de dimension notablement réduite par rapport à l'espace initial. Il existe un potentiel très fort d'obtenir des réductions drastiques des temps de calcul (de plusieurs ordres de grandeur) via une synergie entre les approches d'analyse de fiabilité par échantillonnage adaptatif et la modélisation en base réduite. Un des premiers objectifs de la thèse a été de développer le couplage entre ces deux approches afin de rendre possible cette synergie. La méthode a été validée sur deux cas tests simples (un modèle thermique et un modèle de plaque trouée stratifiée) et les gains en temps de calcul sont importants : un facteur 7 a été obtenu sur l'accélération entre la méthode AKMCS et la méthode couplant AKMCS et la base réduite. Le deuxième objectif sera d'utiliser l'approche dans une méthode d'optimisation adaptative basée sur les processus gaussiens et ainsi se comparer aux approches plus classiques d'optimisation robuste [153].

Communications :

- Couplage adaptatif des méthodes d'analyses de fiabilité par apprentissage actif basées sur les métamodèles de krigeage et de la modélisation en base réduite, CFM 24ème Congrès Français de Mécanique, Brest France, Août 2019
- Towards reliability analysis by adaptive sampling with multiple fidelity levels by the use of reduced basis methods, 3rd International Conference on Uncertainty Quantification in Computational Sciences and Engineering, 24-26 June 2019, Crete, Greece
- Vers des analyses de fiabilité à fidélité adaptative par l'utilisation de modèles en base réduite, aux 10èmes Journées Fiabilité des Matériaux et des Structures, Bordeaux, Mars 2018.

Article en révision :

- Adaptive coupling of reduced basis modeling and kriging based active learning methods for reliability analyses, submitted to Reliability Engineering and System Safety en janvier 2019

6.3 Thèse de Alessandro Sgueglia : *Sizing and optimization priorities applied to a Blended Wing-Body with distributed electric ducted fans*, 2017-2019

La thèse de Alessandro Sgueglia est une collaboration entre l'ISAE-SUPAERO et l'ONERA avec un co-encadrement entre J. Morlier (ISAE), E. Bénard (ISAE), P. Schmollgruber (ONERA) et moi-même. La soutenance est prévue en décembre 2019.

Au cours des prochaines décennies, l'objectif principal de l'aviation civile est de réduire considérablement les émissions, la consommation de bruit, de carburant et d'énergie à l'horizon 2050. Pour atteindre ces objectifs, les changements au niveau des aéronefs doivent être faits en explorant de nouvelles configurations. L'aile volante est une piste intéressante du fait de sa conception intégrée qui améliore les performances et réduit la consommation de carburant. Une nouvelle configuration n'est pas le seul moyen pour atteindre les objectifs demandés, le système de propulsion est également important. Ces dernières années, beaucoup d'efforts ont été consacrés à l'étude de la propulsion distribuée et de la propulsion électrique. L'idée de la thèse est donc l'étude de l'aile volante à propulsion électrique distribuée. Pour réaliser cet objectif, il faut :

- Mettre en place un processus complet de conception avec l'identification des priorités pour le dimensionnement des différents composants ;
- Développer d'un outil d'analyse de conception multidisciplinaire (MDA) pour mieux représenter des phénomènes physiques spécifiques ;
- Explorer et sélectionner la formulation MDO la plus appropriée pour ce problème, compte tenu des deux points précédents.

Dans une première partie de la thèse, le processus de dimensionnement a concerné un avion classique (tube and wing) avec une propulsion électrique distribuée afin de mettre en place les briques logicielles et les boucles de conception en comparant les performances à une propulsion classique. Dans une deuxième étape, le concept d'aile volante a été dimensionné avec une propulsion classique. La dernière étape sera de coupler l'aile volante et la propulsion distribuée.

Alessandro a bénéficié du soutien de la Fondation ISAE-SUPAERO pour effectuer un séjour de 3 mois entre janvier 2018 et avril 2018 au MDOLab à l'Université du Michigan pour collaborer sur la mise en place du processus sur OpenMDAO.

Communications :

- Exploration of a Blended Wing-Body concept featuring distributed electric propulsion with gradient optimization techniques, EASN 2019
- Preliminary Sizing of a Medium Range Blended Wing-Body Using a Multidisciplinary Design Analysis Approach, MATEC Web of Conferences, Volume 23, n. 14, EASN 2018
- Exploration and Sizing of a Large Passenger Aircraft with Distributed Ducted Electric Fans, AIAA SciTech Conference, Kissimee, Florida, 8-12 January 2018

Chapitre de livre en révision :

- Technologies sensitivity analysis of an hybrid aircraft, chapitre 10, Aerospace System Analysis and Optimization in Uncertainty, Springer, Optimization and Its Applications, New York, 2019

Article en révision :

- Development of a multidisciplinary design optimization framework with gradient calculation applied to hybrid aircraft, submitted to Journal of Aircraft, 2019

6.4 Thèse de Joan Mas Colomer : *Similitude Aéroélastique d'un Démonstrateur en Vol via l'Optimisation Multidisciplinaire*, 2015-2018

La thèse de Joan Mas Colomer a démarré en octobre 2015 avec un financement ONERA-ISAE. j'en ai assuré la co-direction avec J. Morlier (ISAE-SUPAERO) en collaboration avec P. Schmollgruber et T. Lefebvre de l'ONERA. La soutenance a eu lieu en décembre 2018. Joan est aujourd'hui en post-doc à l'ISAE-SUPAERO depuis novembre 2018.

La recherche de configurations d'aéronefs plus efficaces mène les ingénieurs à explorer de nouveaux concepts tels que l'aile volante, l'aile haubanée ou l'aile jointive. Contrairement à la configuration classique aile-fuselage, qui est bien connue et étudiée, le comportement en vol de ces nouveaux concepts d'avion est peu connu. Dans ce contexte, la conception, la construction et les essais de modèles à l'échelle aéroélastiquement semblables se présentent comme un moyen peu risqué

d'acquérir des connaissances expérimentales sur ces nouveaux concepts. Un modèle aéroélastiquement semblable présente le même comportement aéroélastique (mis à l'échelle) que l'avion de référence à échelle réelle. En général, le même comportement aéroélastique implique de reproduire les mêmes déplacements pour des conditions du flux d'air données, ainsi que les mêmes vitesses de flottement ou de divergence statique mises à l'échelle. Pour résoudre le problème de similitude, l'approche est divisée en trois parties. Dans le premier cas, nous traitons le problème de similitude aéroélastique lorsque les paramètres de similitude du flux aérodynamique peuvent être complètement préservés. Dans cette situation, le problème consiste simplement à reproduire la réponse dynamique modale de l'aile mise à l'échelle en optimisant les propriétés de la structure et de la masse. Dans la deuxième partie, nous nous concentrons sur l'optimisation du design de la forme de l'aile pour reproduire la réponse du flottement lorsque les paramètres de remise à l'échelle du flux aérodynamique ne peuvent pas être atteints. Finalement, nous abordons la similitude des déflexions aéroélastiques statiques, même lorsque la similitude du flux d'air ne peut pas être atteinte, grâce à l'optimisation simultanée des propriétés de la structure et de la forme aérodynamique.

Joan a bénéficié du soutien de l'Université Fédérale pour effectuer un séjour de 3 mois entre septembre 2017 et décembre 2017 au MDOLab à l'Université du Michigan.

Communications :

- Similarity Maximization of a Scaled Aeroelastic Flight Demonstrator via Multidisciplinary Optimization, 58th AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, Grapewine, United States, 2017
- Static Aeroelastic Scaling with Non-Similar Flow through Multidisciplinary Optimization, Engopt : 6th International Conference on Engineering Optimization, 2018
- Static and Dynamic Aeroelastic Scaling of the CRM Wing via Multidisciplinary Optimization, WCSMO12 12th World Congress of Structural and Multidisciplinary Optimisation, Braunschweig, Germany, 2017

Article en révision :

- Optimization Approach for Aeroelastic Scaling of Wings under Non-Similar Flow, submitted to Journal of Aircraft, 2018
- Aeroelastic Scaling using the Modal Assurance Criterion, submitted to Aerospace Science and Technology, 2019

6.5 Thèse de Peter Schmollgruber : *Enhancement of the conceptual aircraft design process through certification constraints management and full mission simulations*, 2012-2018

J'ai accompagné Peter Schmollgruber, ingénieur de recherche à l'ONERA dans sa démarche de mener un doctorat en parallèle de ses études contractuelles et de ses activités de chef de projet Clean Sky 2. La soutenance a eu lieu en décembre 2018.

Pour les prochaines générations d'avions de transport civil, les contraintes environnementales fixent des objectifs très ambitieux au niveau de la consommation de carburant. Avec une configuration classique optimisée pendant des décennies offrant peu de marges d'amélioration, de nouveaux concepts basés sur des configurations disruptives et/ou des technologies innovantes sont étudiés. Cependant, peu d'informations sur ces architectures sont disponibles dans les premières phases du processus de conception. Ainsi, les travaux de recherche en conception avion visent à ajouter des connaissances supplémentaires au sein de l'analyse multidisciplinaire. Actuellement, différentes approches sont utilisées : implémentation de boucles d'optimisation multidisciplinaire, ajout de précision grâce à des analyses haute-fidélité, introduction de nouvelles disciplines ou systèmes et gestion de l'incertitude. Suite à l'analyse des résultats obtenus lors d'une première optimisation d'un avion de transport utilisant un système d'assistance au sol basée sur une formulation monolithique et des modèles d'analyse structure avancés, des besoins spécifiques pour le dimensionnement d'un véhicule viable ont été identifiés. Ainsi, cette thèse propose d'ajouter de la connaissance au sein du processus de conception et d'optimisation via l'implémentation d'un nouveau module associé aux contraintes de certification et de capacités de simulations de mission.

Développé entièrement dans le cadre de ces travaux, le module des contraintes de certification (CCM) a été utilisé pour résoudre quatre problèmes d'optimisation associés à un avion de transport civil classique basés sur l'outil de dimensionnement ONERA / ISAE-SUPAERO appelé FAST. Grâce aux fonctionnalités du CCM, le temps d'implémentation de ces optimisations a été réduit et les résultats ont clairement confirmé la nécessité de prendre en compte les contraintes de certification très tôt dans le processus de conception. Afin de mettre en place des capacités de simulation complète, l'analyse multidisciplinaire au sein de FAST a dû être revue. Tout d'abord, l'outil d'analyse aérodynamique a été modifié afin de générer les

coefficients nécessaires pour un modèle à 6 degrés de liberté. Ensuite, un module de calcul des inerties a été ajouté. Enfin, le simulateur open source JSBSim a été utilisé avec différentes boucles de contrôle agissant sur la stabilité et la navigation. Les comparaisons entre les trajectoires de vol obtenues avec FAST et les données réelles sur les avions enregistrées avec une antenne ADS-B ont confirmé la validité de l'approche.

Communications :

- Improvement of the Aircraft Design process for Air Traffic Management evaluations, AIAA Aerospace Sciences Meeting, Kissimmee, United States, 2018
- Use of a Certification Constraints Module for Aircraft Design Activities, 17th AIAA Aviation Technology, Integration, and Operations Conference, 2017
- Development of a certification module tailored to Aircraft Multi Disciplinary Optimization, CEAS Air & Space, Delft 2015
- Virtual flight testing in an aircraft sizing and optimization process, 15th AIAA Aviation Technology, Integration, and Operations Conference, Dallas, June 2015
- Impact of Certification constraints in a Multi-Disciplinary Design Process, Workshop EASN on Aircraft Design, Aachen, 2014

6.6 Thèse de Marina Gruet : *Intelligence artificielle et prévision de l'impact de l'activité solaire sur l'environnement magnétique terrestre*, 2015-2018

En collaboration avec le département DPHY de l'ONERA, j'ai participé à l'encadrement de la thèse de Marina Gruet (financement CNES) sous la direction de A. Sicard et S. Rochel (ONERA). La soutenance a eu lieu en septembre 2018. Marina est aujourd'hui Data Scientist en CDI chez Airbus DS à Toulouse dans une équipe de Machine Learning pour l'imagerie satellite.

Dans cette thèse, nous avons présenté des modèles appartenant au domaine de l'intelligence artificielle afin de prédire l'indice magnétique global am à partir des paramètres du vent solaire. Ceci est fait dans l'optique de fournir des modèles opérationnels basés sur les données enregistrées par le satellite ACE situé au point de Lagrange L1. L'indice am ne possède pas à l'heure actuelle de modèles de prédiction. Pour prédire cet indice, nous avons fait appel à des modèles non-

linéaires que sont les réseaux de neurones, permettant de modéliser le comportement complexe et non-linéaire de la magnétosphère terrestre. Nous avons dans un premier temps travaillé sur le développement et l'optimisation des modèles de réseaux classiques comme le perceptron multi-couche. Ces modèles ont fait leurs preuves en météorologie de l'espace pour prédire aussi bien des indices magnétiques spécifiques à des systèmes de courant comme l'indice Dst, caractéristique du courant annulaire, que des indices globaux comme l'indice Kp. Nous avons en particulier étudié un réseau temporel appelé Time Delay Neural Network (TDNN) et évalué sa capacité à prédire l'indice magnétique am à une heure, uniquement à partir des paramètres du vent solaire. Nous avons analysé la sensibilité des performances des réseaux de neurones en considérant d'une part les données fournies par la base OMNI au niveau de l'onde de choc, et d'autre part des données obtenues par le satellite ACE en L1.

Après avoir étudié la capacité de ces réseaux à prédire am, nous avons développé un réseau de neurones encore jamais utilisé en météorologie de l'espace, le réseau Long Short Term Memory ou LSTM. Ce réseau possède une mémoire à court et à long terme, et comme le TDNN, fournit des prédictions de l'indice am uniquement à partir des paramètres du vent solaire. Nous l'avons optimisé afin de modéliser au mieux le comportement de la magnétosphère et avons ainsi obtenu de meilleures performances de prédiction de l'indice am par rapport à celles obtenues avec le TDNN. Nous avons souhaité continuer le développement et l'optimisation du LSTM en travaillant sur l'utilisation de fonctions de couplage en entrée de ce réseau de neurones, et sur le développement de réseaux multisorties pour prédire les indices magnétiques am sectoriels, spécifiques à chaque secteur Temps Magnétique Local. Enfin, nous avons développé une nouvelle technique combinant réseau LSTM et processus gaussiens, afin de fournir une prédiction probabiliste jusqu'à six heures des indices magnétiques Dst et am. Cette méthode a été dans un premier temps développée pour l'indice magnétique Dst afin de pouvoir comparer les performances du modèle hybride à des modèles de référence, puis appliquée à l'indice magnétique am. Marina a effectué un séjour de 3 mois (janv 2018 à avril 2018) au CWI (centrum voor wiskunde und informatica), laboratoire de recherche appliquée en informatique et mathématique d'Amsterdam) sous la direction d'Enrico Camporeale, chercheur dans l'équipe MultiScale Dynamics.

Communications :

- Prediction of the magnetic index am based on development and comparisons of performances of static and dynamic neural networks, IAGA (International Association of Geomagnetism and Aeronomy), 2017, Cape Town, South Africa

- Prediction of the magnetic index am based on development and the performance comparisons of performances of static and dynamic neural networks, Space Weather, a multidisciplinary approach, Leiden, 2017
- From Artificial Intelligence to Space Weather, Séminaire à l’Institut royal d’Aéronomie Spatiale de Belgique (IASB), Bruxelles Septembre 2017
- Neural networks and Am predictions, a preliminary study, 13th European Space Weather Week, Oostende, Belgium, 2016

Article en révision :

- Prediction of the geomagnetic index am based on the development and performance comparisons of static and dynamic Neural Networks, submitted to Space Weather Space Climate, 2018

6.7 Thèse de Mohamed-Amine Bouhlef : *Optimisation auto-adaptative en environnement d’analyse multidisciplinaire via les modèles de krigeage combinés à la méthode PLS*, 2012-2016

La thèse de Mohamed-Amine Bouhlef était une thèse CIFRE avec la SNECMA que j’ai co-dirigée avec J. Morlier (ISAE) et A. Otsmane (SNECMA). La soutenance a eu lieu en janvier 2016. Mohamed-Amine est en post-doc au MDOLab à l’Université du Michigan depuis 2016.

Les turbomachines aéronautiques sont composées de plusieurs roues aubagées dont la fonction est de transférer l’énergie de l’air au rotor. Les roues aubagées des modules compresseur et turbine sont des pièces particulièrement sensibles car elles doivent répondre à des impératifs de performance aérodynamique, de tenue mécanique, de tenue thermique et de performance acoustique. L’optimisation aéroméca-acoustique ou aéro-thermo-mécanique des aubages consiste à chercher, pour un ensemble de formes aérodynamiques paramétrées (par plusieurs dizaines de variables), celle assurant le meilleur compromis entre la performance aérodynamique du moteur et la satisfaction de plusieurs dizaines de contraintes souvent contradictoires. Cette thèse introduit une méthode d’optimisation basée sur les métamodèles et adaptée à la grande dimension pour répondre à la problématique industrielle des aubages. Les contributions de cette thèse portent sur deux aspects : le développement de modèles de krigeage, et l’adaptation d’une stratégie d’optimisation pour la gestion du grand nombre de variables et de contraintes. La première partie de ce travail traite des modèles de krigeage. Nous avons proposé une nouvelle formula-

tion du noyau de covariance permettant de réduire le nombre de paramètres du modèle afin d'accélérer sa construction. Une des limitations connues du modèle de krigeage concerne l'estimation de ses paramètres. Cette estimation devient de plus en plus difficile lorsque nous augmentons la dimension du phénomène à approcher. En particulier, la base de données nécessite davantage de points et par conséquent la matrice de covariance du modèle du krigeage est de plus en plus coûteuse à inverser. Notre approche consiste à réduire le nombre de paramètres à estimer en utilisant la méthode de régression des moindres carrés partiels (PLS pour Partial Least Squares). Cette méthode de réduction dimensionnelle fournit des informations sur la relation linéaire entre les variables d'entrée et la variable de sortie. Ces informations ont été intégrées dans les noyaux du modèle de krigeage tout en conservant les propriétés de symétrie et de positivité des noyaux. Grâce à cette approche, la construction de ces nouveaux modèles appelés KPLS est très rapide étant donné le faible nombre de paramètres nécessaires à estimer. La validation de ces modèles KPLS sur des cas test académiques ou industriels a démontré leur qualité de prédiction équivalente voire même meilleure que celle des modèles de krigeage classiques. Dans le cas de noyaux de covariance de type exponentiel, la méthode KPLS peut être utilisée pour initialiser les paramètres du krigeage classique, afin d'accélérer la convergence de l'estimation des paramètres du modèle. La méthode résultante, notée KPLS+K, a permis d'améliorer la qualité des modèles dans le cas de fonctions fortement multimodales.

La deuxième contribution de la thèse a consisté à développer une stratégie d'optimisation globale sous contraintes pour la grande dimension, en s'appuyant sur les modèles KPLS ou les modèles KPLS+K. En effet, nous avons étendu la méthode d'optimisation auto-adaptative connue dans la littérature sous le nom "Efficient Global Optimisation, EGO" pour gérer les problèmes d'optimisation sous contraintes en grande dimension. Différents critères d'enrichissement adaptatifs ont pu être explorés. Cette stratégie a permis de retrouver l'optimum global sur des problèmes académiques jusqu'à la dimension 50. La méthode proposée a été confrontée à deux types de problèmes industriels, le cas test MOPTA issu de l'industrie automobile (124 variables d'entrée et 68 fonctions contraintes [9]) et le cas test Snecma des aubes de turbomachines (50 variables d'entrée et 31 fonctions contraintes). Les résultats ont permis de montrer la validité de la démarche ainsi que les limites de la méthode pour une application dans un cadre industriel.

Communications :

- Improvement of efficient global optimization with application to aircraft wing design, 17th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, Washington D.C., USA, 2016

- Kriging combining Partial Least Squares for optimization in high dimension, EngOpt 2014, 4th International Conference on Engineering Optimization, Lisbon, 2014
- Surrogate models for optimization in high dimension using a mixed PLS/Kriging method, OPT-i 14, Greece, 2014
- Surrogate models for optimization in high dimension using a mixed Kriging/PLS method, PLS2014, 8th International Conference on Partial Least Squares and Related Method, Paris, 2014

Articles parus :

- Efficient global optimization for high-dimensional constrained problems by using the Kriging models combined with the partial least squares method, Engineering Optimization, Taylor & Francis Online, 2018
- An Improved Approach for Estimating the Hyperparameters of the Kriging Model for High-Dimensional Problems through the Partial Least Squares Method, Mathematical Problems in Engineering, Hindawi Publishing Corporation, 2016
- Improving kriging surrogates of high-dimensional design models by Partial Least Squares dimension reduction, Structural and Multidisciplinary Optimization, Springer, vol 53 (5), 2016

Brevet :

- Procédé de conception de pièces mécaniques, notamment d’aubes de turbomachines, Brevet FR.15.57301 - cas 2625, 2017

6.8 Thèse de Dimitri Bettebghor : *Optimisation biniveau de structures aéronautiques composites*, 2008-2011

J’ai participé à l’encadrement de la thèse CIFRE de Dimitri Bettebghor avec M. Samuelides (ONERA & ISAE), J. Morlier (ISAE) et S. Grihon (AIRBUS). La soutenance a eu lieu en décembre 2011. Dimitri est ingénieur de recherche à l’ONERA au département structure (DADS puis DMAS) depuis décembre 2011.

Ce travail de thèse s’inscrit dans le domaine de l’optimisation de structures aéronautiques composites. On cherche à rendre possible le traitement de problèmes de dimensionnement de telles structures, telles que celles rencontrées dans l’industrie aéronautique. Ce type de problèmes présente deux aspects bloquants. En premier lieu, la taille des structures et le type de matériaux rendent le problème

d'optimisation à la fois de très grande taille et de variables mixtes (continues, discrètes). D'autre part, le très grand nombre d'analyses de stabilité locale (flambage) nécessaires rend le problème d'optimisation très difficile à traiter en terme de coût de calculs.

On cherche donc à résoudre le premier aspect au travers de schémas d'optimisation dits de décomposition qui permettent de décomposer le problème d'optimisation initial en une multitude de sous-problèmes d'optimisation pouvant être résolus en parallèle et dont le couplage est résolu par un problème d'optimisation sur un ensemble de variables réduit. L'équivalence théorique entre les différents problèmes d'optimisation (en termes de minima locaux) est prouvée et on présente et développe un schéma adapté à la fois aux spécificités des composites et aux contraintes industrielles.

Le second point est résolu de manière originale par le développement d'une stratégie d'approximation des contraintes de stabilité. Cette stratégie de mélanges d'experts se base sur des outils statistiques avancés et se révèle adaptée au comportement des composites. Les deux principales avancées de ce travail sont validées sur des cas test académiques et sur une structure aéronautique réaliste.

Le fil directeur de ce travail est la mécanique des structures composites, néanmoins le caractère pluridisciplinaire du sujet nous a conduit à des incursions vers les domaines des statistiques (apprentissage), de l'analyse numérique (étude de l'équation aux dérivées partielles relative au flambage) et enfin de l'optimisation théorique.

Communications :

- Enhancing approximation of critical buckling factor for laminated composites, ODAS 2012, 12th Onera-DLR Aerospace Symposium, Braunschweig, Germany, 2012
- Approche en paramètres de stratification pour l'optimisation biniveau de structures composites, CSMA, 10e Colloque national en calcul des structures, 2011
- Using mixture of experts surrogate models and EM learning for multilevel structural optimization, IV European Conference on Computational Mechanics, 2010,

Articles parus :

- Approximation of the critical buckling factor for composite panels, Structural and Multidisciplinary Optimization, Springer, 46 (4), 2012
- Surrogate modeling approximation using a mixture of experts based on EM joint estimation, Structural and Multidisciplinary Optimization, Springer, 43 (2), 2011

6.9 Thèse de Joël Clément : *Optimisation multi-disciplinaire : étude théorique et application à la conception des avions en phase d'avant projet*, 2005-2009

En mon arrivée à l'ONERA, j'ai participé à l'encadrement de la thèse de Joël Clément avec M. Masmoudi (MIP-UPS Toulouse) et J. Hermetz (ONERA). La soutenance a eu lieu en juin 2009. Joël a effectué un post-doc au CEA à Tahiti entre 2011 et 2013.

L'optimisation multi-disciplinaire propose des solutions aux problèmes de conception de systèmes complexes. Le terme "optimisation multi-disciplinaire" laisse sous-entendre à tort qu'il ne s'agit que d'un problème d'optimisation. Nous lui préférons ici le terme de "conception collaborative". En effet, l'optimisation ne représente qu'un aspect, qui ne peut être séparée du reste du problème de conception. Le but n'est pas de créer un processus automatique, mais de faciliter les échanges entre les équipes des différentes disciplines. De nombreuses méthodes, appelées communément formulations MDO (de Multi-Disciplinary Optimization), apparaissent dans la littérature (MDF, IDF, AAO, BLISS, CO). Elles proposent des stratégies permettant, d'une part, d'assurer la cohérence de la description du système complexe et, d'autre part, d'effectuer la recherche de la configuration optimale. Dans un premier temps, nous dressons un état de l'art des formulations MDO. Nous mettons en avant leurs points communs et leurs différences, afin de proposer une implémentation de la manière la plus générale qui soit. Nous proposons, avec la méthode DIVE (Discipline Interaction Variable Elimination), un cadre d'utilisation de méta-modèles au sein des formulations MDO. Le méta-modèle peut se limiter à une approximation linéaire ou quadratique. Il peut s'appuyer sur des méthodes classiques d'apprentissage, telles que les réseaux neuronaux, le krigeage ou la SVM. Chaque méta-modèle est accompagné d'une région de confiance qui en détermine la validité. Cette approche par approximations locales et successives permet d'aborder les problèmes de grande dimension. Nous présentons des résultats obtenus avec deux cas-tests d'avions d'affaires supersoniques obtenus sous deux environnements différents (Scilab et ModelCenter).

Communications :

- DIVE : une méthode d'utilisation de méta-modèles dans un processus d'optimisation multi-disciplinaire, 9th Colloque national en calcul des structures, Giens, 2009

- Multi-disciplinary optimization methods studies, 2nd European Conference for Aero-Space Sciences EUCASS, Brussels, Belgium, 2007

6.10 Post-doctorat de Sylvain Dubreuil, 2015-2017

En 2015, j'ai proposé avec Thierry Lefebvre (ONERA/DCPS) un sujet de post-doc sur financement ONERA intitulé "MultiDisciplinary Optimization for aircraft conceptual design : development of formulations adapted to mixed fidelity approaches" qui a été confié en octobre 2015 à Sylvain Dubreuil pour une durée de 24 mois. Sylvain a été recruté en CDI dans l'équipe avant la fin de son post-doc, il est ingénieur de recherche depuis juillet 2017.

L'évaluation et l'optimisation d'un nouveau concept d'aéronef nécessite la mise en place de processus couplant les différentes disciplines physiques (aérodynamique, structure, propulsion, système etc.) modélisant son comportement. Chaque discipline est représentée par un modèle numérique (modèle empirique, éléments finis, CFD, etc.). L'analyse multidisciplinaire consiste à résoudre le système non linéaire formé par les différents solveurs disciplinaires. Lorsque ces solveurs disciplinaires sont numériquement coûteux à évaluer, la résolution de ce problème d'optimisation devient prohibitive voir impossible.

Afin de palier à ce problème nous proposons de remplacer chaque solveur disciplinaire par un méta-modèle de krigeage (processus gaussien). La solution du système multidisciplinaire est donc un vecteur aléatoire dont la loi est la nouvelle inconnue du problème. Dans [72], nous avons proposé d'approcher ce vecteur aléatoire par son développement sur le chaos polynomial d'Hermite. Ainsi, en prenant en compte l'incertitude introduite par l'utilisation des méta-modèles de krigeage dans l'analyse multidisciplinaire, la fonction objectif peut être vue comme un champ aléatoire sur l'espace de conception. Nous sommes en particulier intéressés par les valeurs extrêmes de ce champ *i.e.* la variable aléatoire qui modélise la valeur aléatoire du minimum et celle qui modélise sa position. Afin d'approximer ces 2 variables aléatoires nous avons proposé [73] une méthode originale de discrétisation et de représentation du champ aléatoire basée sur un couplage entre chaos polynomial et interpolation par krigeage. Cette approche permet une discrétisation adaptative du champ aléatoire (sélection itérative des points de conception où la propagation d'incertitude par chaos polynomial doit être effectuée) dans les zones de l'espace de conception où la probabilité d'atteindre le minimum est importante

(un critère de type *amélioration espérée* est proposé dans [73]).

Enfin, un processus d'optimisation par enrichissement des méta-modèles disciplinaires est mis en place à l'aide du critère de discrétisation. L'objectif de cet enrichissement est de *réduire* itérativement la dispersion des variables aléatoires (valeur minimale et position). Lorsque celle-ci est *suffisamment* faible, la solution du problème d'optimisation déterministe est approchée par les espérances de ces variables aléatoires.

Six mois avant la fin de son post-doc, Sylvain Dubreuil a été embauché en CDI dans l'équipe ce qui a permis de mener à bout ce travail et de poursuivre encore aujourd'hui cette thématique de recherche avec l'extension aux variables de couplage vectorielles.

Communications :

- Propagation d'incertitude de modèle pour l'optimisation multidisciplinaire, CANUM 2018, 44e Congrès National d'Analyse Numérique, 2018
- Multidisciplinary optimization based on surrogate models, OLA 2018, Workshop on Optimization and Learning : Challenges and Applications, Alicante Spain, 2018
- Towards an efficient global MDO based on surrogate models, 1st European Workshop on MDO for Industrial Applications in Aeronautics – Challenges and Expectations, Braunschweig Germany, 2017
- Adaptive random field discretization for optimization under uncertainties, WCSMO12 12th World Congress of Structural and Multidisciplinary Optimisation, Braunschweig, Germany, 2017
- An adaptive kriging based approach to robust optimization, UNCECOMP 2017, 2nd International Conference on Uncertainty Quantification in Computational Sciences and Engineering, Greece, 2017
- Évaluation de l'incertitude associée à l'interpolation de maillage dans un calcul couplé partitionné, 13e Colloque National en calcul des structures, Giens (Var), 2017
- Multidisciplinary optimization by surrogate models : handling epistemic uncertainties by polynomial chaos expansion, ECCOMAS Congress, European Congress on Computational Methods in Applied Sciences and Engineering, Crete Island, Greece, 2016

Articles parus :

- Extreme value oriented random field discretization based on a hybrid polynomial chaos expansion — Kriging approach, Computer Methods in Applied Mechanics and Engineering, 332, Elsevier, 2018
- Propagation of modeling uncertainty by polynomial chaos expansion in mul-

tidisciplinary analysis, Journal of Mechanical Design, 138 (11), American Society of Mechanical Engineers, 2016

6.11 Post-doctorat de Fabio Luraghi, 2011

Dans le cadre du projet régional RNTL OSYCAF, j'ai participé à l'encadrement de Fabio Luraghi avec M. Samuelides (ONERA & ISAE) et J. Morlier (ISAE). Fabio travaille aujourd'hui à Londres dans une société d'analyse de risques financiers.

Dans le but final d'optimiser un caisson composite d'aile d'avion, une des étapes originales concernait la conception de panneaux composites à fibres courbes. L'objectif était de concevoir un panneau composite avec une orientation des fibres courbes optimale de manière à maximiser une charge critique de flambage. Cela se traduit par un problème d'optimisation sous contraintes où la fonction objectif à maximiser est par exemple la charge critique de flambage ou encore la charge de rupture, les variables de conception sont définies, par exemple les paramètres de représentation des fibres courbes, et des contraintes de fabrication comme par exemple une courbure maximale imposée pour l'orientation des fibres, ou bien une contrainte de rupture du stratifié (critère de Tsai-Wu). La modélisation des fibres courbes qui a été choisie est basée sur le fait que pour chacun des plis, les fibres sont définies à partir d'une fibre de référence, qui elle est décrite via les courbes NURBS (Non Uniform Rational B-Splines). Pour obtenir des panneaux qui maximisent la charge critique de flambage, le problème de conception est traité comme un problème d'optimisation de forme où les variables d'optimisation sont les paramètres des courbes NURBS (appelés points de contrôle). Les propriétés du stratifié sont déterminées via un code élément finis appelé SASCREP (Static Analysis of Symmetrically-Laminated Composite REctangular Panels) concernant le calcul du flambage des panneaux composites, développé sous Matlab. Le code a été couplé avec un algorithme évolutionnaire (CMA-ES pour Covariance Matrix Adaptation Evolution Strategy) pour essayer de trouver ces points de contrôle optimaux. Les algorithmes évolutionnaires sont bien connus pour permettre une exploration large du domaine de conception mais pour cela nécessitent un grand nombre d'appels de la fonction objectif (le code SASCREP) et donc des temps de calcul qui peuvent s'avérer coûteux. Le couplage SASCREP-CMAES peut atteindre plusieurs jours de calcul pour certaines optimisations. Un axe du projet a donc été de remplacer le code éléments finis exact par un modèle analytique approché peu coûteux à évaluer. Cette technique d'utilisation de méta-modèles (ou surrogate models) est bien connue depuis quelques années et donc la littérature est

très riche. L'idée ici a été d'utiliser des bibliothèques Open Sources développées sur Matlab (SED, DACE, ARESLab,...) permettant l'appel de différents types de modèles réduits (fonction polynomiale linéaire ou quadratique PRS, fonctions à base radiale RBF, réseaux de neurones, kriging, splines adaptatives MARS, les moindres carrés mobiles MLS,...). Ce projet a accompagné la participation de l'ONERA/DTIM aux projets européens VIVACE et MAAXIMUS où des modèles réduits ont été développés pour être utilisés lors d'algorithmes d'optimisation multi-niveau de structures aéronautiques. Dans cette étude, une stratégie a été mise place pour choisir de manière séquentielle les points d'échantillonnage basée sur un critère exploitation-exploration selon le principe LOLA-Voronoi. Il s'agit d'enrichir les zones pour lesquelles une approximation locale linéaire n'est pas suffisante (critère d'exploitation : LOLA pour LOcal Linear Approximation) tout en s'assurant de couvrir de manière "uniforme" l'espace de design (critère d'exploration basé sur le pavage de Voronoi). Concernant le deuxième point, l'algorithme développé permet de choisir le meilleur modèle réduit suivant un critère d'erreur à minimiser (erreur moyenne, erreur maximale, ...). Donc au final, le code développé en MATLAB (appelé AMEDEO pour Adaptive Metamodeling for Efficient DEsign Optimization) définit un plan d'expériences initial avec peu de points qui va être enrichi de manière adaptative via la méthode LOLA-Voronoi de manière à minimiser le critère d'erreur choisi par l'utilisateur. La précision est évaluée soit sur une base de validation (si fournie par l'utilisateur) soit par validation croisée. Deux cas d'applications issues de la littérature ont été traitées, il s'agissait de concevoir un stratifié constitué de 16 plis qui maximise la charge critique de flambage. Du fait des symétries du stratifié, seuls 4 plis ont été considérés avec 5 points de contrôle pour la courbe NURBS de référence, soit un total de 20 variables de conception. Les résultats montrent un gain important, de plus de 47% sur la charge critique de flambage, entre un composite à fibres courbes et un composite à fibres droites. Outre la construction du modèle réduit qui nécessite entre 5 et 7 heures de calcul (pour 200 à 300 points de calcul éléments finis), l'étape d'optimisation est réduite à quelques minutes. Les résultats obtenus qui montrent le gain d'utiliser des composites courbes sont très prometteurs quand à l'utilisation de ces matériaux.

Communication :

- On some applications of surrogate model-based optimization techniques to the design of composite panels for aeronautical applications, AFG'11, 15th Austrian-French-German conference on Optimization, Toulouse, 2011

Chapitre 7

Participation à des projets de recherche

Sommaire

7.1 Projets internes ONERA	220
7.2 Projets régionaux et nationaux	221
7.3 Projets européens	222
7.4 Projets en cours de montage	223

Ce chapitre détaille les projets internes et externes dans lesquelles j'ai été impliquée depuis mon arrivée à l'ONERA.

Au CERFACS, entre 1997 et 2004, j'ai participé à des projets internes et des projets en lien avec les organismes de recherche (CEA-CESTA) et des industriels (EADS, CNES, ALCATEL). J'ai fait des études théoriques et numériques avec des livrables associés (rapports et code de calcul) autour des Equations Intégrales et de la méthode Multipôle pour accélérer les calculs de diffraction d'ondes. Depuis mon arrivée à l'ONERA en 2005, j'ai participé à de nombreux projets internes et externes dont la liste est détaillée ci-dessous.

7.1 Projets internes ONERA

Dès mon arrivée à l'ONERA j'ai été impliquée dans des projets internes dont la durée peut varier : PRF pour Projet de Recherche Fédérateur (3 à 4 ans), PR pour Projet de Recherche (2 à 3 ans) et ARF pour Axe de recherche Fédérateur (1 an).

- PRF DOOM (Démarche Outillée pour Optimisation Multidisciplinaire), Projet de Recherche Fédérateur ONERA, 01/2004-12/2007, coordinateur J. Hermetz, Participation au recensement des algorithmes d'optimisation, des formulations multidisciplinaires et développements de métamodèles (réseaux de neurones) dans un contexte MDO
- PRF HEPHAIISTOS, Projet de Recherche Fédérateur ONERA, 01/2008-12/2011, coordinateur C. Guin, Méthodologie pour l'optimisation des émissions de polluants du système d'injection et de la chambre de combustion
- PRF MAHPSO (Modèles d'Approximation de Haute Précision pour les Systèmes d'équations de Propagations d'Ondes), Projet de Recherche Fédérateur ONERA 01/2006 - 12/2010, coordinateur F. Rogier, Développement d'un logiciel "d'étude" sur la méthode Galerkin Discontinu pour intégrer le raffinement en espace et le raffinement en ordre d'approximation dans les équations aéroacoustiques (à partir des équations de Euler Linéarisées) de manière à capturer les tourbillons de Kelvin-Helmholtz
- ARF DMORPH (Développement de Méthodes pour Optimiser la Résolution numérique des modèles PHysiques), Axe de recherche Fédérateur 01/2011-12/2013, coordinateur C. Marmignon, Développements autour des méthodes variationnelles de type Galerkin Discontinu et des raffinements adaptatifs hp pour les équations d'Euler linéarisées
- PR VESTA (Méthodologie d'évaluation et de Validation de Systèmes de Transport Aérien futurs - Application au ces des contrats 4D), Projet de Recherche, 01/2014-12/2018, coordinateur T. Chaboud, Activités sur l'ana-

- lyse multi-critères appliquée à l'ATM
- PRF CREATION (Concepts of Rotorcraft Enhanced Assessment Through Integrated Optimization Network), Projet de Recherche Fédérateur ONERA, 01/2005-12/2009, coordinateur P.M. Basset, Optimisation multiobjectif d'un hélicoptère de transport pour 90 passagers à l'aide des mélanges d'experts pour traiter les variables discrètes et continues
 - PRF ACADIA (Atelier de Conception et d'Aide à la Décision pour l'Innovation Aérospatiale), Projet de Recherche Fédérateur ONERA, 01/2014-12/2018, coordinateur S. Defoort, Participation au développement de méthodes dans le domaine de la génération de modèles de substitution, des méthodes d'optimisation, et de la gestion des incertitudes
 - ARF MERITE (Météorologie de l'Espace et Risques Induits sur la Terre et l'Environnement), Axe de Recherche, 2016-2018, coordinateur D. Boshier. J'ai participé au projet via l'encadrement de thèse de Marina Gruet pour la prédiction des indices magnétiques.
 - PR MUFIN (MULTidisciplinarité et MULTI Fidélité sous Incertitudes pour l'étude de Nouveaux concepts), Projet de recherche, 01/2019-12/2021, coordinateur T. Lefebvre, Responsable des tâches dédiées aux modèles de substitution pour la multifidélité et aux méthodes d'optimisation adaptative en multifidélité.

7.2 Projets régionaux et nationaux

Au niveau national, j'ai été impliquée dans des projets nationaux financés par l'ANR ou la DGAC, et des projets régionaux financés par la fondation STAE ou l'IRT Siant-Exupéry.

- ANR OMD (Optimisation MultiDisciplinaire) pour optimiser les systèmes complexes avec Scilab, Juin 2006 – Novembre 2009, coordinateur : R. Leriche (Ecole des Mines de St-Etienne), Participation à l'Optimisation Collaborative dans le cadre de la thèse de J. Clément
- Projet DPAC NECST-2 ARTEMIS (Advanced R&T Enablers for Multidisciplinary Integrated Systems) dont le but était de développer un processus d'optimisation multidisciplinaire dédié à la conception préliminaire d'un avion de transport civil, 2009-2014, coordinateur P. Schmollgruber, Participation à la mise en place de la formulation bi-niveau entre le processus bi-discipline (aéro-structure haute fidélité) et le processus avion global
- RTRA [OSYCAF](#) sur l'Optimisation d'un système couplé fluide-structure

représentant une aile flexible dans un écoulement transsonique, 04/2010-06/2013, coordinateur CERFACS, Participation à la formulation multidisciplinaire et à l'optimisation de panneaux composites à fibres courbes dans le cadre du post-doc de F. Luraghi

- Projet [MDA-MDO](#) (MultiDisciplinary Analysis - MultiDisciplianry Optimization) de l'IRT Saint-Exupéry dédié au développement de méthodes, processus et outils d'analyse et d'optimisation multidisciplinaire. Mise à disposition dans le projet à 20% (1 jour par semaine environ) entre décembre 2014 et décembre 2018. Participation à la mise en place du processus d'optimisation bi-niveau pour prendre en compte les contraintes industrielles, extension du processus d'optimisation aérodynamique multipoint à une meilleure prise en compte de la mission, encadrement de deux stagiaires (Mars 2017 à sept 2017 et Mars 2018 à sept 2018), participation aux réunions devant les membres du projet et aux revues de projet devant AIRBUS
- Projet CONCORDE - DGA, DGAC dans le cadre de la Fédération de recherche entre l'ONERA, ISAE et l'ENAC, Analyse et conception intégrée vers la certification des futurs systèmes de drones et de leurs opérations, démarrage prévue en septembre 2019 pour une durée de 6 ans, Encadrement d'une thèse prévue autour de la co-conception multidisciplinaire intégrant les modèles aéro-structure et les opérations

7.3 Projets européens

Depuis 2005, je me suis investie dans plusieurs projets européens dont la liste est donnée ci-dessous.

- FP6 [INTELLECT D.M.](#) (Integrated lean low emission combustor design methodology) 01/2004-12/2009, coordinateur ROLLS-ROYCE DEUTSCHLAND, Participation à l'optimisation d'un tube de pré-mélange LPP via une approche adaptative basée sur les réseaux de neurones et les techniques de bootstrap
- FP7 [MAAXIMUS](#) (More Affordable Aircraft structure through eXtended, Integrated, and Mature nUmerical Sizing), 04/2008 - 09/2016, coordinateur AIRBUS OPERATIONS GMBH, participation à la tâche sur l'optimisation des panneaux composites en utilisant des mélanges d'experts dans le cadre de la thèse de D. Bettebghor
- H2020 [AGILE](#) (Aircraft 3rd Generation MDO for Innovative Collaboration of Heterogeneous Teams of Experts) 05/2015-11/2018, coordinateur DLR,

responsable de la tâche dédiée aux formulations MDO et aux méthodes d'approximation

- Projet CleanSky 2 [Airframe/ITD](#), NACOR (New Innovative Aircraft COnfigurations and Related Issues), Projet autour de la comparaison de différentes configurations d'avion, 10/2015- 2021, coordinateur ONERA, Participation à la mise en place du processus d'optimisation
- H2020 [AGILE 4.0](#), Extension des développements de processus de dimensionnement et d'optimisation aux aspects certification et fabrication, 09/2019-08/2022, coordinateur DLR, Responsable d'une tâche dédiée à l'optimisation et à la validation des processus

7.4 Projets en cours de montage

A ce jour, quatre projets sont en cours de montage.

- CRP (Common Reserch Projet) avec le DLR Hamburg, eMDO dédié aux processus MDO intégrés pour la conception des avions électriques
- PR IAHPHI (Intelligence Artificielle Hybride Pour l'Industrie 4.0), Projet de Recherche, coordinateur T. Polacsek, en cours de soumission pour un démarrage 01/2020-12/2023
- ANR internationale avec ENAC et Hong-Kong University of Science and Technology pour développer une méthodologie qui lie les données réelles opérationnelles de l'avion et les données de conception pour dessiner des trajectoires de vol plus efficaces, coordinateur T. Klein (ENAC) et R. Liem (HKUST)
- GIS Sizing Lab, structure locale avec INSA, LAAS, ICA, ICAM, ISAE, ENAC pour le dimensionnement de composants technologiques (par ex. actionneurs, chaînes électriques), coordinateur INSA, Participation aux aspects développements méthodologiques autour des formulations MDO et des métamodèles.

Chapitre 8

Collaborations

Sommaire

8.1	Collaborations régionales	226
8.2	Collaborations européennes	228
8.3	Collaborations internationales	230

Ce chapitre est dédié aux collaborations académiques et industrielles dans lesquelles je me suis impliquée depuis 2005.

Depuis quelques années, j'ai mis en place de nombreuses collaborations au niveau régional, européen et international.

8.1 Collaborations régionales

Au niveau régional, la liste de mes collaborations est présentée ci-dessous.

- La collaboration avec l'ISAE-SUPAERO a débuté par des enseignements dès 2006 puis depuis 2008 par une activité recherche via des co-encadrements de thèses, de stages et la participation à des projets communs (Projet RNTL OSYCAF, FP7 MAAXIMUS et projet ANR OMD). J'ai travaillé avec J. Morlier (Professeur à l'ISAE-SUPAERO, département mécanique des structures et matériaux, membre de l'Institut Clément Ader) autour de l'optimisation multidisciplinaire dans le cadre de deux thèses CIFRE (AIRBUS et SNECMA). Le démarrage des équipes de recherche communes (ECR MOIS et ECR CSDV) entre l'ONERA et l'ISAE est venu renforcer cette collaboration en accueillant de nombreux stagiaires (convention d'accueil facilitée entre les deux organismes) en co-encadrement. Des bourses de thèses entre les deux organismes ont été demandées et obtenues (2015 et 2017). Nous organisons environ 2 fois par an des séminaires communs afin de présenter le travail des stagiaires, doctorants et post-docs et d'organiser les activités de recherche des prochains mois. Six articles [36, 47, 46, 48, 49, 23] ont été publiés, un brevet [156] accepté et deux articles sont en cours de révision [140, 187] dans le cadre de cette collaboration. Une thèse en co-encadrement a également démarré en novembre 2017 avec le département de mathématiques (Y. Diouane et M. Salaun ISAE/DISC) autour de l'optimisation bayésienne sous contraintes en grande dimension.
- La collaboration avec l'Institut Clément Ader est plus récente. Depuis 2015, nous travaillons avec C. Gogu (Enseignant-Chercheur à l'université Toulouse III, membre de l'Institut Clément Ader) autour de l'optimisation de structures et les incertitudes dans un cadre multidisciplinaire. Des réunions régulières sont organisées à l'ONERA et deux articles ont été rédigés en collaboration [72, 73] et un troisième article vient d'être soumis [151]. C. Gogu a obtenu une ANR jeune chercheur en octobre 2016 (projet ReBRed) pour développer une nouvelle méthodologie pour la conception fiabiliste par modélisation en base réduite. Dans le cadre de ce projet ReBRed et en partenariat avec l'ONERA, je co-encadre avec C. Gogu, J. Morio et S. Dubreuil (ONERA) la thèse de Morgane Menz autour de l'analyse de fiabi-

lité par échantillonnage adaptatif et réduction de modèles. Un autre projet est en cours de montage autour des stratégies d'optimisation à mettre en place dans la conception préliminaire en modélisation de manière explicite les parties prenantes influençant le processus de conception (compagnies aériennes, autorités de régulation...). Une étude préliminaire est en cours avec des étudiants de Master de l'ISAE-SUPAERO.

- Ma participation au projet MDA-MDO de l'IRT Saint-Exupéry pour mettre en place une optimisation multidisciplinaire pour redimensionner un mat moteur avec une mise à disposition un jour par semaine a facilité les échanges avec les entreprises partenaires du projet : AIRBUS, CERFACS, ALTRAN, SOGETI. Je suis particulièrement en lien avec les personnes impliquées dans les avant-projets AIRBUS. Cette mise à disposition (2014-2018) s'est achevée en décembre 2018 avec l'extension du processus d'optimisation aérodynamique multipoint à une meilleure prise en compte de la mission (avion global) et une soumission à la conférence AIAA/AVIATION de Dallas 2019 [92, 91, 65].
- Une collaboration avec l'ENAC (département Mathématiques appliquées) a été initiée en octobre 2010. Entre 2010 et 2014, j'ai donc participé à différents montages de projets (FP7 CORINA et CARAT) et en particulier au montage de l'ANR STARGATE acceptée en 2014. Mon changement d'équipe en 2012 avec une réorientation de mes activités vers l'optimisation multidisciplinaire m'a contraint en avril 2015 à abandonner ma contribution à ce projet. Depuis je suis en lien avec M. Mongeau (participation au jury de thèse de M.-A. Bouhleb en janvier 2016 et de P. Schmollgruber en décembre 2018) et N. Peteilh qui effectue une thèse autour de l'optimisation avion avec la prise en compte des contraintes aéroportuaires. Une ANR entre la France et Hong-Kong a été proposée en 2018 entre l'ENAC (porteur), l'ONERA et l'University of Science and Technology mais refusée lors de la première sélection. En, revanche, une bourse de voyages a été acceptée en février 2019 ce qui va permettre à Rhea Liem de l'HKUST de venir à Toulouse durant l'été 2019 et de pouvoir avancer sur le montage de l'ANR internationale.
- Une collaboration a été initiée avec V. Picheny de l'INRIA sur les critères d'enrichissement de la méthode EGO [163, 164]. Une journée a été organisée en novembre 2016 pour la présentation de nos activités respectives et un co-encadrement de projet étudiants a été initié avec les étudiants de l'ISAE-SUPAERO (octobre 2017 et mars 2018) autour de l'optimisation multiobjectif. La mobilité de V. Picheny en Angleterre pour un an a mis un frein à cette collaboration pour le moment.

- La mise en place de la Fédération ONERA-ISAE-ENAC lancée en juin 2018 est venue renforcer ce lien. Dans les 8 thèmes retenus, je suis particulièrement impliquée dans le thème eXtended-Overall Aircraft Design (X-OAD) qui traite de l'Analyse et la co-conception. Les deux objectifs de ce thème sont de développer des stratégies et processus de co-conception d'aéronefs/opérations innovants, en faisant appel aux méthodes de l'ingénierie système et de la conception multidisciplinaire, et de prescrire, développer et valider des méthodes et outils communs dans le domaine de la conception d'aéronefs (plateforme partagée). Le thème compte actuellement 32 membres des trois organismes. J'assure la représentation de ce thème au sein du comité de pilotage de la Fédération depuis novembre 2018. A court terme, une des premières tâches sera la mise en place et la présentation du projet X-OAD en lien avec le projet de la Fédération pour l'évaluation HCERES prévue fin 2019.
- Plus récemment, le démarrage de la thèse CIFRE de V. Palladino en juin 2019 a permis de nouer les premiers contacts avec ATR, ce qui devrait se renforcer au cours de trois années à venir.

8.2 Collaborations européennes

Depuis mon arrivée à l'ONERA en 2005, j'ai été impliquée dans plusieurs projets européens : projet INTELLECT D.M. (2004-2007), projet MAAXIMUS (2008-2011), projet AGILE (2015-2018) et plus récemment le projet AGILE 4.0 (2019-2022) qui devrait démarrer en septembre 2019. De part ces opportunités, j'ai pu nouer des contacts avec le NLR, TUDelft, et le DLR qui se sont renforcés avec le temps. La Figure 8.1 permet de visualiser les différentes collaborations à l'échelle européenne.

- Mon implication dans le projet H2020 AGILE (responsabilité d'une tâche méthodologique avec 15 partenaires) m'a permis de collaborer étroitement avec le DLR, le NLR, l'Université de Naples, le TUDelft, NOESIS (Belgique), Airinnova (Suède) et le TsAGI (Russie). Lors des réunions d'avancement du projet AGILE sur le site de Toulouse (organisation pour 40 personnes en décembre 2015, 25 personnes en septembre 2016, 40 personnes en décembre 2017) ces contacts ont pu être renforcés. Entre juin 2016 et juin 2017, j'ai organisé des réunions régulières et participé au moins une fois par semaine à des téléconférences pour discuter de l'avancement des travaux avec ces différents organismes. Pendant 18 mois environ, j'ai par-



FIGURE 8.1 – Collaborations européennes.

icipé et présenté à toutes les réunions d’avancement du projet (environ tous les 3 mois) et ma participation à la revue de mi-parcours du projet AGILE devant les industriels (décembre 2016 et novembre 2018) lors des Open Day AGILE et devant la commission européenne (en janvier 2017 à Bruxelles et novembre 2018 à Hambourg) a permis d’affirmer ma position au sein du projet. J’ai pu également mettre en place plusieurs cas tests avec le TsAGI en utilisant leur code CFD (pour Computational Fluid Dynamics) pour l’optimisation d’une nacelle d’avion avec ou sans contraintes. Les résultats très encourageants laissent envisager une collaboration sur du long terme (présentation lors des journées Tsagi-ONERA en novembre 2018). Dans ce cadre, deux présentations ont été faites à l’AIAA/ISSMO à Denver en juin 2017. En juin 2018, j’ai organisé une session dédiée au projet AGILE à l’AIAA/ISSMO à Atlanta avec 6 présentations sélectionnées du projet. Rappelons que le projet AGILE a connu un immense succès auprès de la commission européenne mais également auprès de la communauté aéronautique et qu’il a été récompensé lors de la conférence ICAS 2018 au Brésil en recevant le prix ”ICAS Award for Innovation in Aeronautics” ([EU news](#)).

- Je suis également impliquée dans le Challenge AGILE ACADEMY qui regroupe des étudiants dans le monde entier pour résoudre un processus d'optimisation multidisciplinaire [Agile Academy](#). J'ai participé au co-encadrement d'une équipe de 10 étudiants (ONERA, Concordia University, ISAE-SUPAERO, University of Michigan, University of Pisa) entre janvier 2018 et juin 2018, qui a remporté le challenge
- J'ai pu également inviter Imco van Gent, étudiant en thèse du TUDelft, pendant deux mois à l'ONERA (décembre 2017 à février 2018) pour mettre en place une collaboration autour des outils collaboratifs KADMOS et CMDOWS développés par le TUDdelft [Open MDO suite](#) avec les outils internes ONERA (WhatsOpt [123]). Des cours autour de ces outils ont d'ailleurs été mis en place lors de la session MDO à l'ISAE-SUPAERO en février 2019 et mai 2019 en collaboration avec T. Lefebvre.
- Dans le cadre de mes activités autour de l'optimisation multidisciplinaire, ma participation à l'organisation du 1er workshop européen [MDO-workshop](#) en octobre 2017 m'a permis de nouer des contacts dans le milieu académique (University of Southampton) et industriel (AIRBUS UK et Allemagne) et de participer au Steering Group meeting en février 2018 chez AIRBUS. La deuxième édition devrait avoir lieu à l'automne 2019 à Toulouse (IRT Saint-Exupéry).
- Le démarrage de la thèse de M. Saporito en 2019 entre l'Université de Southampton et l'ONERA va permettre de renforcer cette collaboration académique au cours des trois prochaines années.

8.3 Collaborations internationales

Au niveau international comme le montre la Figure 8.2, j'ai pu participer à des collaborations avec différents organismes américains, un industriel canadien et l'université de Hong-Kong.

- Dans le cadre de la thèse de Mohamed-Amine Bouhleb, j'ai pris contact avec Rommel Regis de l'Université St Joseph, Philadelphie aux Etats-Unis. Il s'agissait de comparer différents optimiseurs sur des fonctions analytiques. Nous avons donc échangé des fonctions et des résultats par mail et organisé plusieurs conférences téléphoniques. R. Regis a participé à la rédaction d'un article de conférence [21] et d'un article de journal [48].
- Dans le cadre de la venue de Joaquim R.R.A. Martins à l'ISAE-SUPAERO via une bourse Pierre et Marie Curie (aout 2015-juillet 2016), de nombreux



FIGURE 8.2 – Collaborations internationales.

échanges ont été organisés avec le [MDOLab](#) de l'Université du Michigan aux Etats-Unis. De nombreuses discussions ont eu lieu concernant les sujets de thèse, les outils et les modèles que l'on peut partager notamment pour la mise en place d'un processus de conception d'une aile volante. Depuis, les méthodes que nous développons, sont évaluées sur des cas tests communs (cas test OpenAeroStruct [110]).

- Cette collaboration solide a permis de mettre en place des échanges d'étudiant. J. Mas Colomer, thésard ONERA-ISAE (2015-2018) a obtenu une aide à la mobilité de l'Université fédérale Toulouse et l'Ecole des Docteurs pour un séjour de 3 mois au MDOLab effectué entre septembre et décembre 2017. Cela a permis de faire le lien entre les outils PANAIR & NASTRAN utilisés à l'ONERA et le code CFD ADFlow du MDOLab (codes de calcul haute-fidélité). Un article pour l'AIAA/Scitech en janvier 2017 [139] a d'ailleurs été rédigé dans le cadre de cette collaboration et un article a été soumis dans la revue Journal of Aircraft [140].

En 2018, Alessandro Sgueglia, étudiant en thèse, est parti au MDOLab (janvier à avril 2018) pour un séjour de 3 mois durant lequel il a intégré dans la plateforme OpenMDAO le code avant-projet FAST de l'ONERA-ISAE.

Un article a été soumis dans la revue Journal of Aircraft [187].

Plus récemment, un étudiant de l'ISAE-SUPAERO, Mostafa Meliani, a effectué 3 mois durant son stage de fin d'études au MDOLab (Juillet à Septembre 2018) grâce à un financement de la Fondation ISAE-SUPAERO. Mostafa a travaillé sur l'optimisation multi-fidélité et cela a donné lieu à un papier soumis à la conférence AIAA 2019 [146] et a présenté son travail de recherche au challenge Pegasus en avril 2019. Ce travail sera poursuivi en 2019 dans le cadre du projet interne MUFIN dédié à la multifidélité.

- Concernant les modèles réduits basés sur les mélanges d'experts (Mixture of Experts), une collaboration existe entre l'ONERA, l'ISAE, le MDOLab, et l'Université de Hong-Kong. Des comparaisons entre le code *moe* (développé en interne ONERA/DCPS, cf Section 3.1) et le code du MDOLab développé par Rhea Liem durant sa thèse [130, 129] ont été menées durant le stage de R. Priem (sept 2015 à fev 2016). J'ai organisé plusieurs téléconférences durant l'année 2016 pour mettre en place un travail collaboratif entre les différents organismes. R. Liem, actuellement en poste à l'Université de Hong-Kong, travaille avec un étudiant de master sur cette thématique. A la fin de sa thèse à l'ONERA, M.-A. Bouhlel est parti en post-doc au MDOLab (juin 2016 pour 2 à 3 ans) pour travailler en partie sur les modèles réduits et notamment sur la prise en compte de l'information des gradients. L'objectif était de développer une toolbox opensource dédiée aux métamodèles dont les mélanges d'experts basée sur la toolbox scikit-learn [159] et qui intégrerait les méthodes venant des codes ONERA, ISAE et MDOLab. Une première version de la toolbox SMT (Surrogate Models Toolbox) est disponible sous GitHub [SMT Git](#), elle est actuellement utilisée par plusieurs universités américaines. Une version actualisée avec les mélanges d'experts a été mise en ligne en février 2018. Un article décrivant les fonctionnalités de cette toolbox a été accepté en février 2019 dans le Journal of Advances in Engineering Software [49]. Une évolution récente (aout 2018) de cette toolbox a permis d'intégrer les travaux de stage de Mostafa Meliani sur les modèles réduits multifidélité.
- Concernant la capitalisation des outils, nous avons choisi d'intégrer nos développements dans la plateforme OpenMDAO, plateforme open-source développée par la NASA-Glenn. Dans ce cadre, une collaboration a été initiée avec le chef de projet d'OpenMDAO et certains de nos modules python sont à présent intégrés et proposés dans la plateforme (module de co-krigeage en 2014, cas test SSBJ en 2016). Cette collaboration a fait l'objet de deux faits marquants à l'ONERA (mars 2015 et décembre 2016). J'ai pu organiser en octobre 2017 le premier workshop européen des uti-

lisateurs d'OpenMDAO [agenda](#) qui a rassemblé plus de 60 participants (académiques, instituts de recherche européens et partenaires industriels).

- Depuis l'été 2017, une collaboration a été mise en place avec l'Université de Floride et le professeur R. Haftka autour des modèles réduits en multi-fidélité. Une de ses doctorantes, G. Fernández-Godino est venue passé trois mois pour travailler avec C. Gogu à l'Institut Clément Ader. Cela a été le début d'un travail commun autour des modèles de co-krigeage. Cela a donné lieu à plusieurs réunions téléphoniques entre 2017 et 2019, un papier de conférence AIAA [86] et plus récemment un papier soumis dans le journal SMO [85].
- Dans le cadre du projet AGILE, des contacts ont eu lieu avec l'industriel BOMBARDIER concernant l'algorithme d'optimisation SEGOMOE. Des comparaisons ont été effectuées entre SEGOMOE et leur optimiseur interne pour traiter des problèmes académiques. Suite aux résultats encourageants de SEGOMOE, un NDA a été signé et Rémy Priem est parti à Montréal pour 4 mois (mai à septembre 2019) afin de tester la méthode sur un cas test de conception avion en phase préliminaire et en phase conceptuelle ce qui va permettre d'augmenter considérablement le nombre de variables de design et le nombre de contraintes.

Chapitre 9

Activités de dissémination

Sommaire

9.1	Organisation de workshops	236
9.2	Participation à des formations	238
9.2.1	Formation Matlab	238
9.2.2	Formation métamodèles - Machine Learning	238
9.3	Publications	239
9.3.1	Revue internationale avec comité de lecture	239
9.3.2	Ouvrage	242
9.3.3	Chapitres de livre	243
9.3.4	Mémoires de DEA et de thèse	243
9.4	Conférences	244
9.4.1	Conférences internationales avec actes	244
9.4.2	Autres conférences et workshops sans acte	250
9.4.3	Conférences invitées et séminaires	253
9.5	Brevets	254
9.6	Administratif et Scientifique	255
9.6.1	Participation à des activités de relecture	255
9.6.2	Participation à des jurys	255

Ce chapitre détaille les workshops pour lesquels je me suis investie dans l'organisation, les formations auxquelles je participe et les différentes publications et communications pour lesquelles j'ai collaboré.

9.1 Organisation de workshops

Je détaille dans la liste ci-dessous les workshops pour lesquels je me suis impliquée dans l'organisation.

- Participation à l'organisation du premier workshop européen pour la plate-forme OpenMDAO en collaboration avec l'ISAE-SUPAERO et la NASA Glenn, ONERA Toulouse 12 et 13 octobre 2017, [agenda](#). Cet événement, bénéficiant de la présence de l'équipe de la NASA Glenn en charge du développement du logiciel, a réuni plus de 60 personnes en provenance de centres de recherche et d'universités européennes ainsi que de l'industrie aéronautique. Pendant 2 jours, les participants ont pu échanger sur évolutions de la plateforme ainsi que sur les principales applications en optimisation multi-disciplinaire (MDO) actuellement développées en Europe. Ce workshop affiche un bilan très positif tant sur le plan technique que sur les aspects liés au rayonnement scientifique. Il a permis de resserrer les échanges entre la communauté OpenMDAO européenne et la NASA, d'affirmer le positionnement de l'ONERA et sa capacité d'initiative sur la thématique MDO, et de nouer ou renforcer des collaborations externes.

- Participation à l'organisation du premier workshop européen sur la MDO en collaboration avec l'IRT Saint-Exupéry et le DLR, Braunschweig Germany, 24 et 25 Octobre 2017. Le thème général était : "MDO Overview & Challenges at Organizing Research Institutes and at Industry" [Workshop MDO](#). Le workshop a compté plus de 110 participants (50% de centres de recherche, 30 % d'industriels, 20% d'universitaires). Les sujets proposés autour de la MDO étaient :

- Gradients in MDO
- MDO Concepts, Methods & Algorithms
- Airbus Challenge Test Case
- Collaborative MDO, Architectures, Data Management and IT Aspect
- CAD and Parameterization for MDO
- Loads and Structure Sizing for MDO
- Uncertainty in MDO

avec une Panel Session : "MDO organizational and technical bottlenecks, what industry expects from research". Deux personnes ont été invitées pour présenter leurs expériences P. Piperni (Bombardier puis Univ. Clarkson) and J.R.R.A Martins (Univ. of Michigan).

- Une nouvelle édition du workshop européen sur la MDO en collaboration avec

l'IRT Saint-Exupéry et le DLR est prévue courant Novembre 2019 (19 et 20) à Toulouse et des discussions sont en cours pour le choix des sujets à proposer et des interlocuteurs. En parallèle, une troisième journée (21 novembre 2019) sera proposée dans le cadre de la Fédération ONERA-ENAC-ISAE dédiée aux activités de recherche MDO dans le monde académique.

- J'ai proposé un minisymposium autour de la multifidélité dans le cadre de la conférence [WCCM/ECCOMAS](#). Le programme du minisymposium dont la référence est [MS442 Multifidelity Optimization](#) est en cours de finalisation.
- De manière plus ponctuelle, j'ai organisé des séminaires à l'auditorium ONERA avec Joaquim Martins de l'Université du Michigan en 2015 et plus récemment avec Karen Willcox de l'Oden Institute (Austin, Texas) en juillet 2019 avec la présence d'académiques et d'industriels locaux.

9.2 Participation à des formations

A ce jour, j'ai été impliquée dans des formations autour du logiciel Matlab dans le cadre de ma thèse et autour des métamodèles liés à mes activités de recherche.

9.2.1 Formation Matlab

Dans le cadre de certaines formations proposées par le département de mathématiques de l'INSA Toulouse, j'ai participé à deux formations du logiciel Matlab en février et mai 2000 (Initiation à Matlab, Interpolation, Intégration numérique, EDO). Cette formation était adressée à des enseignants de l'INSA et à des ingénieurs de Motorola Toulouse.

- Personnel INSA : 9 heures en février 2000
- Ingénieurs Motorola : 6 heures en juin 2000

9.2.2 Formation métamodèles - Machine Learning

Dans le cadre de mes activités de recherche à l'ONERA, j'ai mis en place, en collaboration avec D. Bettebghor (ONERA/DADS) et S. Lefebvre (ONERA/DOTA), une formation autour des métamodèles (ou modèles réduits) qui a été proposée au personnel de l'ONERA. La formation interne ONERA "Panorama des métamodèles pour l'approximation : principes généraux et applications" qui a été proposée a permis de présenter ces différentes techniques aux ingénieurs et chercheurs de l'Office et de nouer des collaborations entre les départements sur cette thématique.

- une formation de 8h a été réalisée en décembre 2014 à l'ONERA - Palaiseau (30 personnes présentes)
- une formation de 8h a été réalisée en mars 2015 à l'ONERA - Toulouse (25 personnes présentes)

Une nouvelle édition devrait avoir lieu fin 2019.

En externe, cette formation a également été proposée à EUROSABE (théorie et pratique en python) et le stage intitulé *Machine learning dans un contexte aéronautique : motivation, présentation et applications* [MTS-003](#) a été réalisé en 2018 et 2019 pour former des ingénieurs de l'industrie et des organismes d'état. Cette formation devrait évoluer vers une version plus longue (3 jours) en 2020.

9.3 Liste détaillée des publications

Je liste ici mes différentes publications et chapitres d'ouvrage. En utilisant les résultats de Web of Knowledge [5], on obtient la distribution suivante de mes publications:

- Engineering Multidisciplinary 57%
- Mathematics Applied 35%

ce qui est représenté sur la Figure 9.1 par les rectangles bleu et orange.

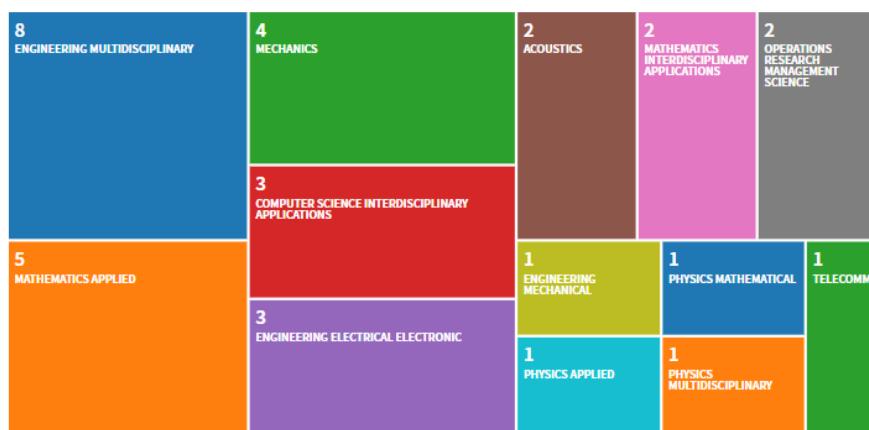


FIGURE 9.1 – Analyse des domaines de publication faite par Web of Knowledge [5].

9.3.1 Revues internationales avec comité de lecture

Dans la liste des articles, j'ai choisi de différencier les articles parus, les articles en révision et les articles soumis récemment. Les impacts facteurs donnés par Journal Citation Reports et les quartiles fournis par SCImago Journal Rank Indicator [4] sont donnés pour information:

- Moyenne de l'Impact Factor sur 5 ans (5year-IF)
- Journal Impact Factor de 2017 (2017-IF) donnés par Web of Knowledge
- Quartile (Q1 à Q4) disponibles jusqu'en 2018 sur SCImago

9.3.1.1 Articles soumis

Deux articles soumis récemment sont à ce jour en attente d'un premier retour.

- R. Priem, N. Bartoli, Y. Diouane, A. Sgueglia *Upper Trust Bound Feasibility Criterion for Mixed Constrained Bayesian Optimization with Application to Aircraft Design*, submitted to Aerospace Science and Technology Journal, October 2019
- J. Mas Colomer, N. Bartoli, T. Lefebvre, J. Morlier, *Aeroelastic Scaling using the Modal Assurance Criterion*, submitted to Aerospace Science and Technology, May 2019

9.3.1.2 Articles en révision

A ce jour, huit articles sont en cours de première ou seconde révision.

- S. Dubreuil, N. Bartoli, C. Gogu and T. Lefebvre, *Towards an Efficient Global Multidisciplinary Design Optimization algorithm*, in review to Structural and Multidisciplinary Optimization Journal, July 2019
- M. Menz, C. Gogu, S. Dubreuil, N. Bartoli and J. Morio, *Adaptive coupling of reduced basis modeling and Kriging based active learning methods for reliability analyses*, in revision to Reliability Engineering and System Safety journal, September 2019
- A. Sgueglia, P. Schmollgruber, N. Bartoli, E. Benard, J. Morlier, J. Jasa, J. R.R.A. Martins, J. T. Hwang, J. S. Gray, *Development of a multidisciplinary design optimization framework with gradient calculation applied to hybrid aircraft*, in revision to Journal of Aircraft, August 2019
- M. Zhang, N. Bartoli and A. Jungo, *Disciplinary Data Fusion for Multi-Fidelity Aerodynamic Application*, in review to CEAS Journal, 2018, paper no.953
- M. Zhang, N. Bartoli, A. Jungo, W. Lammen, E. Baalbergen and M. Voskuijl, *Disciplinary Data Fusion and Aerodynamic Surrogate Modeling for Flight Simulation*, submitted to Progress in Aerospace Sciences, Elsevier 2018
- T. Lefebvre, N. Bartoli, S. Dubreuil, M. Panzeri, R. Lombardi, P. Della Vecchia, L. Stingo, F. Nicolosi, K. Anisimov, A. Savelyev, A. Mirzoyan and A. Isyanov, *Optimization enhancements in a collaborative framework for nacelle and wing applications*, submitted to Progress in Aerospace Sciences, Elsevier 2018
- M. Gruet, N. Bartoli, S. Rochel, T. Pellegrini, A. Sicard, R. Benacquista and G. Rolland, *Prediction of the magnetic index α_m based on development and performance comparisons of static and dynamic neural networks*, under revision to

Space Weather Space Climate, February 2018

- J. Mas Colomer, N. Bartoli, T. Lefebvre, J. R.R.A. Martins, J. Morlier, *Optimization Approach for Aeroelastic Scaling of Wings under Non-Similar Flow*, submitted to Journal of Aircraft, September 2018

9.3.1.3 Articles parus

Aujourd’hui, dix articles parus sont liés à mes activités de recherche autour de l’optimisation et des métamodèles.

- M. Giselle Fernández-Godino, S. Dubreuil, N. Bartoli, C. Gogu, S. Balachandar, R. T. Haftka, *Linear Regression Multi-fidelity Surrogate for Disturbance Amplification in Multi-phase Explosion*, Structural and Multidisciplinary Optimization Journal, 2019
- N. Bartoli, T. Lefebvre, S. Dubreuil, R. Olivanti, R. Priem, N. Bons, J.R.R.A. Martins and J. Morlier, *Adaptive modeling strategy for high-dimensional constrained global optimization*, Aerospace Science and Technology Journal, March 2019, vol 90, p 85-102, 5year-IF 2.439, 2017-IF 2.228
- M. A. Bouhlel, J. T. Hwang, N. Bartoli, R. Lafage, J. Morlier and J.R.R.A. Martins, *A Python surrogate modeling framework with derivatives*, Advances in Engineering Software, Elsevier 2019, 5year-IF 2.948, 2017-IF 3.198
- S. Dubreuil, N. Bartoli, C. Gogu and T. Lefebvre, *Extreme value oriented random field discretization based on polynomial chaos expansion and kriging*, Computer Methods in Applied Mechanics and Engineering, 332, p 540-57, 2018, 5year-IF 4.499, 2017-IF 4.441
- M.-A. Bouhlel, N. Bartoli, R. Regis, A. Otsmane and J. Morlier, *Efficient Global Optimization for high-dimensional constrained problems by using the kriging models combined with the Partial Least Squares method*, Engineering optimization, 2018, 5year-IF 1.908, 2017-IF 1.622
- S. Dubreuil, N. Bartoli, C. Gogu and T. Lefebvre, *Propagation of modeling uncertainty by polynomial chaos expansion in multidisciplinary analysis*, ASME. J. Mech. Des., doi:10.1115/1.4034110, 2016, 5year-IF 2.698, 2017-IF 2.786, Q1
- M.-A. Bouhlel, N. Bartoli, A. Otsmane and J. Morlier, *An Improved Approach for Estimating the Hyperparameters of the Kriging Model for High-Dimensional Problems through the Partial Least Squares Method*, Mathematical Problems in Engineering, Volume 2016 (2016), Article ID 6723410, 11 pages, 5year-IF 1.156, 2017-IF 1.145, Q2
- M.-A. Bouhlel, N. Bartoli, A. Otsmane and J. Morlier, *Improving kriging sur-*

rogates of high-dimensional design models by Partial Least Squares dimension reduction, Structural and Multidisciplinary Optimization, vol 53, n°5, pp 935-952, 2016, 5year-IF 3.269, 2017-IF 2.876, Q1

• D. Bettebghor and N. Bartoli. *Approximation of the critical buckling factor for composite panels*, Structural and Multidisciplinary Optimization, vol 46, n°4, pp 561-584, 2012, 5year-IF 3.269, 2017-IF 2.876, Q1

• D. Bettebghor, N. Bartoli, S. Grihon, J. Morlier and M. Samuelides. *Surrogate modeling approximation using a mixture of experts based on EM joint estimation*, Structural and Multidisciplinary Optimization, vol 43, n°2, pp 243-259, 2011, 5year-IF 3.269, 2017-IF 2.876, Q1

Les publications qui suivent ont été effectuées dans le cadre de mes activités eu CERFACS.

• N. Bartoli, F. Collino, F. Dodu and T. Koleck. *A far-near field Transformation using the Fast Multipole Techniques*, IEEE Transactions on Antennas and Propagation, vol 52, n°12, pp 3329-3336, Décembre 2004, 5year-IF 4.120, 2017-IF 4.130, Q1

• N. Bartoli and A. Bendali. *Robust and High-Order Effective Boundary Conditions for Perfectly Conducting Scatterers Coated by a thin Dielectric Layer*, IMA Journal of Applied Mathematics, vol 67, n°5, pp 479-508, Octobre 2002, 5year-IF 1.228, 2017-IF 1.369, Q3

• N. Bartoli and F. Collino. *Integral equations via Saddle Point Problems for Acoustic Problems*, Mathematical Modelling and Numerical Analysis, vol 34, n°5, pp 1023-1049, Sept/Oct 2000, 5year-IF 0.648, 2017-IF 0.716, Q1

9.3.2 Ouvrage

Durant ma thèse, j'ai pu participer à l'écriture d'un ouvrage lié à mes enseignements avec P. Del Moral.

• N. Bartoli et P. Del Moral. *Simulation et algorithmes stochastiques : Une introduction avec applications*, [Cépaduès éditions](#), 2001.

9.3.3 Chapitres de livre

Un chapitre de livre a été soumis en janvier 2019 et un deuxième chapitre vient d’être soumis en septembre 2019.

- A. Sgueglia, S. Dubreuil, N. Bartoli, chapter 10: *Technologies sensitivity analysis of an hybrid aircraft*, Aerospace System Analysis and Optimization in Uncertainty (Springer), In press 2019
- S. Dubreuil, N. Bartoli, C. Gogu, T. Lefebvre, *Réduction d’incertitudes en analyse multidisciplinaire basée sur une étude de sensibilité par chaos polynomial*, Ouvrage ”Traitement des incertitudes en ingénierie mécanique”, Collection Sciences, Editions ISTE, soumis en septembre 2019

9.3.4 Mémoires de DEA et de thèse

Les références liés à mon doctorat et mon mémoire de DEA sont données ci-dessous.

- Nathalie Bartoli. *Modèles pour la diffraction d’ondes par des obstacles revêtus de couches minces. Résolution de problèmes de diffraction d’ondes via une formulation intégrale de type point selle*, Thèse de doctorat en mathématiques appliquées, INSA-MIP et CERFACS, soutenue le 21 décembre 2000. Directeur de Thèse : Abderrahmane Bendali. Rapporteurs : Bruno Després et Jean-Claude Nédélec. Jury : Alain Bachelot, Abderrahmane Bendali, Pierre Degond, Bruno Després, Mohammed Masmoudi, Jean-Claude Nédélec, Jean Ovidia et Jean-Paul Vila. Document disponible à partir de la page <http://www.cerfacs.fr/emc/>.
- Nathalie Bartoli. *Association de l’information polarimétrique et de l’étude des résonances pour l’identification d’objets enfouis à partir de la diffraction d’impulsions électromagnétiques très courtes*, Responsable de stage : Pierre Borderies (CERT-ONERA), DEA de mathématiques appliquées, INSA-Université Paul Sabatier de Toulouse, Septembre 1997.

9.4 Présentations à des conférences nationales et internationales

La liste des conférences est donnée ci-dessous en distinguant les conférences avec acte ou sans acte.

9.4.1 Conférences internationales avec actes

Les conférences auxquelles j'ai participé et pour lesquelles les proceedings sont disponibles sont listées ci-dessous.

2019

- A. Sgueglia, P. Schmollgruber, E. Benard, N. Bartoli, J. Morlier, *Exploration of a Blended Wing-Body concept featuring distributed electric propulsion with gradient optimization techniques*, 9th EASN Conference on Innovation in Aviation & Space, September 2019, Greece
- M. Menz, C. Gogu, S. Dubreuil, N. Bartoli, J. Morio, *Couplage adaptatif des méthodes d'analyses de fiabilité par apprentissage actif basées sur les métamodèles de krigeage et de la modélisation en base réduite*, CFM 24ème Congrès Français de Mécanique, Brest, France, Aout 2019
- M. Menz, C. Gogu, S. Dubreuil, N. Bartoli, J. Morio, *Towards reliability analysis by adaptive sampling with multiple fidelity levels by the use of reduced basis methods*, UNCECOMP 2019 3rd International Conference on Uncertainty Quantification in Computational Sciences and Engineering, Crete, Greece, June 2019
- M. Meliani, N. Bartoli, T. Lefebvre, M.-A. Bouhlel, J. R.R.A Martins, J. Morlier, *Multi-fidelity efficient global optimization: Methodology and application to airfoil shape design*, 20th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, ID 3130197, June 2019, Dallas, United States
- R. Priem, N. Bartoli, Y. Diouane, *Use of Upper Trust Bounds in Constrained Bayesian Optimization Infill Criteria*, 20th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, ID 3130342 , June 2019, Dallas, United States
- A. Gazaix, F. Gallard, V. Amber, D. Guénot, M. Hamadi, P. Sarrouille, V. Gachelin, J. Plakoo, N. Desfachelles, N. Bartoli, T. Lefebvre, S. Grihon, *MDA-MDO Project: Advanced MDO Processes, Achievements and Results*, 20th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, ID 3130467 , June 2019, Dallas, United States
- B. Dabas, N. Bartoli, T. Lefebvre, F. Gallard, A. Gazaix, T. Druot, D. Guénot, *Error-Based Adaptive Coupling Process Between Multipoint High-Fidelity Aero-*

dynamics and Mission Performance for Shape Optimization in the MDA-MDO Project, 20th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, ID 3130886 , June 2019, Dallas, United States

2018

- R. Priem, N. Bartoli, Y. Diouane, S. Dubreuil, T. Lefebvre, *An adaptive feasibility approach for constrained bayesian optimization with application in aircraft design*, EngOpt 2018, 6th International Conference on Engineering Optimization, Lisbon, September 2018
- J. Mas Colomer, N. Bartoli, T. Lefebvre, J. Morlier, *Static Aeroelastic Scaling with Non-Similar Flow through Multidisciplinary Optimization*, EngOpt 2018, 6th International Conference on Engineering Optimization, Lisbon, September 2018
- A. Sgueglia, P. Schmollgruber, E. Benard, N. Bartoli, J. Morlier, *Preliminary Sizing of a Medium Range Blended Wing-Body using a Multidisciplinary Design Analysis Approach*, 8th EASN-CEAS International Workshop on Manufacturing for Growth & Innovation, vol 233, <https://doi.org/10.1051/mateconf/201823300014>, September 2018, Glasgow, United Kingdom
- N. Bartoli, T. Lefebvre, S. Dubreuil, M. Panzeri, R. Lombardi, R. D'Ippolito, K. Anisomov, A. Savelyev, *Robust Nacelle Optimization Design investigated in the AGILE European project*, 19th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, 10.2514/6.2018-3250, June 2018, Atlanta, United States
- S. Dubreuil, N. Bartoli, T. Lefebvre, C. Gogu, *Efficient global multidisciplinary optimization based on surrogate models*, 19th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, 10.2514/6.2018-3745, June 2018, Atlanta, United States
- T. Lefebvre, N. Bartoli, S. Dubreuil, M. Panzeri, R. Lombardi, W. Lammen, M. Zhang, I. van Gent, P. D. Ciampa, *A clustered and surrogate-based MDA use case for MDO scenarios in AGILE project*, 19th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference 10.2514/6.2018-3252, June 2018, Atlanta, United States
- M. G. Fernández-Godino, R. T. Haftka, S. Balachandar, C. Gogu, S. Dubreuil, N. Bartoli, *Multi-fidelity Surrogate based Optimization as a Tool to Explore Explosive Dispersal of Solid Particles*, 2018 AIAA Non-Deterministic Approaches Conference, AIAA SciTech Forum, (AIAA 2018-2176)
- P. Schmollgruber, J. Bedouet, N. Bartoli, E. Bénard, Y. Gourinat, *Improvement of the Aircraft Design process towards Air Traffic Management evaluations*, 2018 AIAA Aerospace Sciences Meeting, AIAA SciTech Forum, (AIAA 2018-0283)
- A. Sgueglia, P. Schmollgruber, N. Bartoli, O. Atinault, E. Bénard and J. Morlier *Exploration and Sizing of a Large Passenger Aircraft with Distributed Ducted*

2017

- M. Gruet, S. Rochel, N. Bartoli, R. Benacquista, A. Sicard, G. Rolland, T. Pelligrini, *Prediction of the magnetic index am based on development and the performance comparisons of performances of static and dynamic neural networks*, Space Weather, a multidisciplinary approach, Leiden, The Netherlands, November 2017
- M. Zhang, A. Jungo and N. Bartoli, *Disciplinary Data Fusion for Multi-Fidelity Aerodynamic Application*, 6th CEAS Aerospace Europe Conference, paper no.953, Bucharest, Romania, October 2017
- T. Lefebvre, N. Bartoli, S. Dubreuil, M. Panzeri, R. Lombardi, W. Lammen, B. de Wit, M. Zhang, I. van Gent and P. D. Ciampa, *Overview of MDO enhancement in AGILE project: a clustered and surrogate-based MDA use case*, 6th CEAS Aerospace Europe Conference, paper no.956, Bucharest, Romania, October 2017
- T. Lefebvre, N. Bartoli, S. Dubreuil, M. Panzeri, R. Lombardi, R. D'Ippolito, P. Della Vecchia, F. Nicolosi, P. D. Ciampa, *Methodological enhancements in MDO process investigated in the AGILE European project*, 18th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, AIAA AVIATION Forum, (AIAA 2017-4140)
- P. Schmollgruber, N. Bartoli, J. Bedouet, Sébastien Defoort, Y. Gourinat, E. Bénard, *Use of a Certification Constraints Module for Aircraft Design Activities*, 17th AIAA Aviation Technology, Integration, and Operations Conference, AIAA AVIATION Forum, (AIAA 2017-3762)
- N. Bartoli, T. Lefebvre, N. Bons, J.R.R.A. Martins, J. Morlier, *An adaptive optimization strategy based on mixture of experts for wing aerodynamic design optimization*, 18th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, AIAA AVIATION Forum, (AIAA 2017-4433)
- A. Gazaix, F. Gallard, V. Gachelin, T. Druot, S. Grihon, V. Ambert, D. Guénot, R. Lafage, C. Vanaret, B. Pauwels, N. Bartoli, T. Lefebvre, P. Sarouille, N. Desfachelles, J. Brézillon, M. Hamadi, S. Gurol, *Towards the industrialization of new MDO methodologies and tools for aircraft design*. In 18th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference (p. 3149), 2017
- J. Mas Colomer, N. Bartoli, T. Lefebvre, S. Dubreuil, P. Schmollgruber, J. Morlier, *Static and Dynamic Aeroelastic Scaling of the CRM Wing via Multidisciplinary Optimization*, WCSMO12 12th World Congress of Structural and Multidisciplinary Optimisation, June 2017, Braunschweig, Germany
- S. Dubreuil, C. Gogu, N. Bartoli, T. Lefebvre, *Adaptive random field discretization for optimization under uncertainties*, WCSMO12 12th World Congress of Structural and Multidisciplinary Optimisation, June 2017, Braunschweig, Ger-

many

- S. Dubreuil, C. Gogu, N. Bartoli, T. Lefebvre, *An adaptive kriging based approach to robust optimization*, UNCECOMP 2017, 2nd International Conference on Uncertainty Quantification in Computational Sciences and Engineering, June 2017, Greece
- S. Dubreuil, J. Mas Colomer, M. Salaiün, N. Bartoli, T. Lefebvre, *Évaluation de l'incertitude associée à l'interpolation de maillage dans un calcul couplé partitionné*, CSMA 13e Colloque National en calcul des structures, Giens (Var), 2017
- J. Mas Colomer, N. Bartoli, T. Lefebvre, S. Dubreuil, J. R.R.A Martins, E. Bénard, J. Morlier, *Similarity Maximization of a Scaled Aeroelastic Flight Demonstrator via Multidisciplinary Optimization*, 58th AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, AIAA SciTech Forum, (AIAA 2017-0573)

2016

- T. Lefebvre, N. Bartoli, R. Lafage, P.D. Ciampa, *AGILE DC-1 MDO process using an efficient global optimization approach*, 6th EASN International Conference on Innovation in European Aeronautics Research, October 2016, Porto, Portugal
- N. Bartoli, I. Kurek, R. Lafage, T. Lefebvre, R. Priem, M.-A. Bouhlef, J. Morlier, V. Stiliz, R. Regis, *Improvement of efficient global optimization with mixture of experts: methodology developments and preliminary results in aircraft wing design*, 17th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, AIAA AVIATION Forum, (AIAA 2016-4001)
- S. Dubreuil, N. Bartoli, C. Gogu, T. Lefebvre, *Multidisciplinary optimization by surrogate models: handling epistemic uncertainties by polynomial chaos expansion*, ECCOMAS, European Congress on Computational Methods in Applied Sciences and Engineering, Crete Island, Greece, June 2016

2015

- P. Schmollgruber, J. Bedouet, N. Bartoli, Yves Gourinat, *Development of a certification module tailored to Aircraft Multi Disciplinary Optimization*, CEAS Air & Space, Delft, September 2015
- P. Schmollgruber, N. Bartoli, Yves Gourinat, *Virtual flight testing in an aircraft sizing and optimization process*, 15th AIAA Aviation Technology, Integration, and Operations Conference, AIAA AVIATION Forum, (AIAA 2015-2546)

2014

- P. Schmollgruber, N. Bartoli, Yves Gourinat, *Impact of Certification constraints*

in a Multi-Disciplinary Design Process, Workshop EASN on Aircraft Design (Aachen), October 2014

- M.-A. Bouhlef, N. Bartoli, A. Otsmane, J. Morlier, *Kriging combining Partial Least Squares for optimization in high dimension*, EngOpt 2014, 4th International Conference on Engineering Optimization, Lisbon, 8-11 September 2014
- P.-M. Basset, A. Tremolet, N. Bartoli, T. Lefebvre, *Helicopter presizing by multidisciplinary-multiobjective optimization*, 1st International Conference on Engineering and Applied Sciences Optimization, Greece, June 2014
- M.-A. Bouhlef, N. Bartoli, A. Otsmane, J. Morlier, *Surrogate models for optimization in high dimension using a mixed PLS/Kriging method*, OPT-i 14, Greece, 4-6 June 2014
- M.-A. Bouhlef, N. Bartoli, A. Otsmane, J. Morlier, *Surrogate models for optimization in high dimension using a mixed Kriging/PLS method*, PLS2014, 8th International Conference on Partial Least Squares and Related Method, Paris, 26-28 May 2014

2011

- N. Bartoli, D. Bettebghor, F. Luraghi, M. Samuelides, *On some applications of surrogate model-based optimization techniques to the design of composite panels for aeronautical applications*, AFG'11, 15th Austrian-French-German conference on Optimization, Toulouse, September 19-23, 2011
- N. Bartoli, P. Delorme, *An Aeroacoustic Benchmark for Discontinuous Galerkin Method*, WAVES 2011, 10th International Conference on Mathematical and Numerical Aspects of Waves, July 25-29, Vancouver, 2011
- D. Bettebghor, N. Bartoli, M. Samuelides, J. Morlier, S. Grihon, CSMA, 10e colloque national en calcul des structures, Giens (var), 9-13 mai 2011

Avant 2011

- N. Bartoli, P. Mazet, V. Mouysset, F. Rogier, *Highly Accurate Schemes for Wave Propagation Systems: Application in Aeroacoustics*, AIP Conference Proceedings, september 2010
- M. Samuelides, D. Bettebghor, S. Grihon, N. Bartoli, J. Morlier, A. Merval, *Using mixture of experts surrogate models and EM learning for multilevel structural optimization*, IV European Conference on Computational Mechanics, May 16-21, 2010
- N. Bartoli, P. Mazet, V. Mouysset, F. Rogier, *Highly accurate schemes for wave propagation systems: application in aeroacoustic*, WAVES 2009, les aspects mathématiques et numériques de la propagation des ondes, 15-19 Juin 2009
- J. Clément, M. Masmoudi, J. Hermetz, N. Bartoli, Y. Parte, *DIVE: une méthode*

d'utilisation de méta-modèles dans un processus d'optimisation multi-disciplinaire, Giens (Var), 25-29 Mai 2009, 9e Colloque National en calcul des structures

- J. Clément, M. Masmoudi, N. Bartoli, J. Hermetz, *Multi-disciplinary optimization methods studies*, 2nd European Conference for Aero-Space Sciences EUCASS, Brussels, Belgium, July 1–6, 2007
- P. Klotz, N. Bartoli, L. Cornez, R. Lecourt, N. Savary, *Database enrichment for multi-objective optimization*, AAAF, 42ème Colloque d'Aérodynamique Appliquée : Couplages et optimisation multidisciplinaires, Nice, March 19-21, 2007
- P. Klotz, N. Bartoli, L. Cornez, M. Samuelides, P. Villedieu, *Database Enrichment for Optimization with Artificial Neural Networks*, Second International Workshop on Surrogate Modelling and Space Mapping for Engineering Optimization (SMSMEO-06), Technical University of Denmark, November 9-11, 2006
- N. Bartoli, P.-A. Mazet, F. Rogier, Y. Ventribout, *Highly Accurate Schemes for Wave Propagation Systems: Application to Noise Reduction in Aeroacoustic*, ODAS'2006, 7th ONERA-DLR Aerospace symposium, October, 2006
- N. Bartoli, F. Collino. *Numerical method applied to the antenna-satellite interactions*, Progress in Electromagnetism Research Symposium (PIERS Pisa, Italy, March 2004), pp 469
- N. Bartoli, F. Collino, F. Dodu. *Using the Multipole Techniques for the Far to Near Field Transformation*, WAVES 2003, Mathematical and Numerical Aspects of Wave Propagation (Jyvaskyla, Finland, July 2003), ouvrage de proceedings Springer, pp 662-668
- N. Bartoli, F. Collino. *A Far to Near Interpolation*, Progress in Electromagnetism Research Symposium (PIERS Boston, July 2002), pp 297
- N. Bartoli, F. Collino. *Integral Equations and Iterative Methods*, ICEAA01, International Conference on Electromagnetics In Advanced Applications (Torino, Italy, September 2001), pp 19
- N. Bartoli and F. Collino. *Integral equations via saddle point problem for acoustic problems*, IMA 2000, Second international Conference on Boundary Integral Methods : Theory and Applications (University of Bath, UK, September 2000), pp 19-22
- N. Bartoli, F. Collino. *Integral equations via saddle point problem for acoustic problems*, WAVES 2000, Fifth international conference on mathematical and numerical aspects of waves propagation (Santiago de Compostela, Spain, July 2000), Siam-INRIA, pp 730-734
- N. Bartoli, A. Bendali, M'B. Fares. *Using fictitious surface impedance to remove interior resonances in BEM solutions of electromagnetic boundary value-problem*, Progress in Electromagnetism Research Symposium (PIERS Taipei, Taiwan, March 1999), pp 391

- N. Bartoli, A. Bendali, F. Collino. *High order effective boundary conditions for perfectly conducting scatterers coated by a thin dielectric layer and their boundary element solution*, Progress in Electromagnetism Research Symposium (PIERS Taipei, Taiwan, March 1999), pp 579
- E. Bachelier, N. Bartoli, P. Borderies, J. Grando. *Ultra wide band subsurface scattering : extraction of discriminating features*, Progress in Electromagnetism Research Symposium (PIERS USA, July 1997)
- E. Bachelier, N. Bartoli and P. Borderies. *Identification of 3D subsurface scatterers based on numerical simulation of UWB electromagnetic scattering*, International Geoscience and Remote Sensing Symposium (IGARSS, Singapore, August 1997)
- V. Babovic, N. Bartoli. *An application of artificial neural network in computational hydraulics*, Congress of the International Association for Hydraulic Research (USA 1997), ASCE publications, pp 865-870.

9.4.2 Autres conférences et workshops sans acte

La liste ci-dessous présente les conférences ou workshops pour lesquels aucun acte n'a été fourni.

2019

- G. Berthelin, S. Dubreuil, M. Salaun, C. Gogu, N. Bartoli, *Couplage Optimisation et Réduction de Modèles par Projection: Application à la conception avion avant-projet*, 14e Colloque National en calcul des structures, Giens (Var), Mai 2019
- R. Priem, N. Bartoli, Y. Diouane, *An upper trust bound feasibility criterion for constrained Bayesian optimization*, Optimizations Days 2019, HEC Montréal, Canada, May 2019

2018

- A. Savelyev, K. Anisomov, N. Bartoli, T. Lefebvre, S. Dubreuil, M. Panzeri, R. D'Ippolito, *Robust design optimization applied to turbofan nacelle*, 17th TsAGI-ONERA Seminar, November 2018, Moscow, Russia
- N. Bartoli, T. Lefebvre, S. Dubreuil, R. Priem, J. R.R.A. Martins, J. Morlier, *Adaptive modeling strategy for high-dimensional constrained global optimization*, ISMP: 23rd International Symposium on Mathematical Programming, July 2018, Bordeaux, France
- S. Dubreuil, N. Bartoli, T. Lefebvre, C. Gogu, *Propagation d'incertitude de modèle pour l'optimisation multidisciplinaire*, CANUM 2018, 44e Congrès Natio-

nal d'Analyse Numérique, Mai 2018, Cap d'Agde, France

- M. Menz, C. Gogu, S. Dubreuil, N. Bartoli, J. Morio, *Vers des analyses de fiabilité à fidélité adaptative par l'utilisation de modèles en base réduite*, 10e Journées de Fiabilité des Matériaux et des Structures JFMS, Bordeaux, France, Mars 2018
- S. Dubreuil, N. Bartoli, T. Lefebvre, C. Gogu, *Multidisciplinary optimization based on Kriging surrogate models*, OLA 2018, Workshop on Optimization and Learning: Challenges and Applications, Alicante, Spain, February 2018

2017

- S. Dubreuil, N. Bartoli, T. Lefebvre, C. Gogu, *Towards an efficient global MDO based on surrogate models*, 1st European Workshop on MDO for Industrial Applications in Aeronautics – Challenges and Expectations, Braunschweig, Germany, October 2017
- M. Gruet, S. Rochel, N. Bartoli, A. Sicard, T. Pelligrini, G. Rolland, *Prediction of the magnetic index am based on development and comparisons of performances of static and dynamic neural networks*, IAGA (International Association of Geomagnetism and Aeronomy), November 2017, Cape Town, South Africa
- M. Gruet, N. Bartoli, S. Rochel, T. Pelligrini A. Sicard, R. Benacquista, G. Rolland, *From Artificial Intelligence to Space Weather*, Séminaire à l'Institut royal d'Aéronomie Spatiale de Belgique (IASB), Bruxelles, Belgium, September 2017

2016

- M. Gruet, N. Bartoli, S. Rochel, A. Marchaudon, P.-L. Blelly, Guy Rolland, *Neural networks and Am predictions, a preliminary study*, 13th European Space Weather Week, November 2016, Oostende, Belgium

2014

- T. Lefebvre, M. Balesdent, N. Bartoli, L. Brevault, S. Defoort, R. Lafage, P. Schmollgruber, *MDO advances for aircraft design in ONERA*, 4th SCAD Symposium, Toulouse, November 2014
- S. Defoort, M. Balesdent, L. Brevault, N. Bartoli, T. Lefebvre, R. Vauclin, D. Bailly, P. Klotz, *Modèles de substitution pour la conception et l'optimisation multidisciplinaires*, ONERA JSO Optimisation sous incertitude - Modèles de Substitution et Optimisation, Novembre 2014

2012

- D. Bettebghor, N. Bartoli, M. Samuelides, *Enhancing approximation of critical buckling factor for laminated composites*, ODAS 2012, 12th Onera-DLR Aerospace Symposium, Braunschweig, Germany, 4th to 6th June 2012

- D. Bettebghor, N. Bartoli, *Application des mélanges d'experts à l'approximation des contraintes de flambage en calcul de structures*, ONERA JSO Modèles de Substitution & Optimisation Multi-Disciplinaire, Janvier 2012

Avant 2010

- P. Klotz, N. Bartoli, R. Lecourt, *LPP optimization using response surface method*, 2nd Colloque INCA, 2008
- J. Clément, M. Masmoudi, J. Hermetz, N. Bartoli. *A Scilab tool for MDO*, EU-CASS, 2nd European Conference for Aero-Space Sciences, Brussels, Belgique, July 1-6, 2007
- J. Clément, M. Masmoudi, J. Hermetz, N. Bartoli. *MDO formulation comparison*, AAAF, 42ème Colloque d'Aérodynamique Appliquée : Couplages et optimisation multidisciplinaires, Nice, March 19-21, 2007
- N. Bartoli, P. Mazet, F. Rogier, Y. Ventribout. *Highly accurate schemes for wave propagation systems*, ODAS'2006, 7th ONERA-DLR Aerospace Symposium, ONERA, Toulouse, 4-6 October 2006
- T. Koleck, N. Mailhat, N. Bartoli, F. Millot. *Antenna/Structure Interaction Computation by using Fast Multipole Methods*, JINA2004, 13èmes Journées Internationales de Nice sur les Antennes (Nice, Novembre 2004)
- T. Koleck, N. Bartoli, F. Millot. *Calcul d'interaction antenne/structure par les méthodes multipoles rapides*, Journées scientifiques du CNFRS, Paris, France, 2005
- N. Bartoli, F. Collino, F. Dodu, T. Koleck. *Transformation champ lointain - champ proche par l'approche multipolaire*, GDR ondes "interférences d'ondes" (Marseille, Décembre 2003)
- N. Bartoli, F. Collino. *Numerical methods applied to the antenna-satellite interactions*, Workshop INRIA-GOES/KTH, Hybrid and Multiscale Methods for Electromagnetism Solvers, PSCI scientific meeting (Stockholm, Sweden, October 2003), [agenda](#)
- N. Bartoli. *Scattering of a perfectly conducting obstacle coated by a thin dielectric layer - Some appropriate impedance boundary conditions*, European Optical Society Topical Meetings Digest series, EOS2, Electromagnetic Optics (Paris, août 2001)
- N. Bartoli, F. Collino. *Equations intégrales et problèmes de point selle pour la diffraction des ondes bidimensionnelles*, Journées Micro-ondes et Electromagnétisme de Toulouse JMETS (Janvier 2001)
- N. Bartoli, F. Collino. *Méthode des équations intégrales de Després pour la diffraction des ondes bidimensionnelles*, Congrès d'analyse numérique CANUM (Port d'Albret, juin 2000)

- N. Bartoli. *Conditions d'impédance de surface d'ordre élevé couplées à une méthode itérative*, Journées Micro-ondes et Electromagnétisme de Toulouse JMET (Toulouse, Janvier 2000)
- N. Bartoli. *Conditions de surface d'ordre élevé*, Congrès d'analyse numérique CANUM (Ax-Bonascre, Mai 1999)

9.4.3 Conférences invitées et séminaires

J'ai listé ci-dessous les conférences ou séminaires pour lesquels j'ai été invitée à participer.

2019

- N. Bartoli, AIAA Multifidelity Modeling In Support Of Design And Uncertainty Quantification [workshop](#), June 2019, Dallas, United States

2018

- N. Bartoli, séminaire interne DTIS, *Utilisation de méta-modèles à base de processus gaussiens pour la conception et l'optimisation*, octobre 2018

2016

- N. Bartoli, T. Lefebvre, R. Lafage, J. Morlier, *Improvement of efficient global optimization with mixture of experts: methodology developments and preliminary results in aircraft wing design*, Séminaire de la thématique Mécanique Probabiliste des matériaux et des Structures (Laboratoire Institut Pascal UMR 6602 - CNRS-UBP/SIGMA), IFMA, décembre 2016,
- N. Bartoli, T. Lefebvre *MDO for novel configurations*, AIRBUS Flight Physics Capability EGD-ONERA collaboration exchange, MDO-UQM webex, June 2016

Avant 2010

- Invitation à un Workshop en Suède dans le cadre du programme GOES (INRIA-KTH) en novembre 2003
- Invitation au CEA-CESTA dans le cadre du séminaire "Les activités Scientifiques du CESTA" (février 2001).

9.5 Liste des brevets

Dans le cadre de ma collaboration avec SNECMA, un brevet a été déposé.

- Brevet FR.15.57301 - cas 2625 : A. Otsmane, M.-A. Bouhlef, N. Bartoli and J. Morlier, *Procédé de conception de pièces mécaniques notamment d'aubes de turbomachine*, requested by SNECMA, déposée le 30 juillet 2015, acceptée le 02 Mars 2017.

9.6 Activités administratives et rayonnement scientifique

9.6.1 Participation à des activités de relecture

Dans le cadre de mes activités autour de l'Optimisation multidisciplinaire et métamodèles, j'ai été sollicitée à plusieurs reprises en interne ONERA pour relire et sélectionner des projets internes.

Au niveau des revues internationales, je suis régulièrement sollicitée pour relire des papiers (1 à 3 par an environ)

- Structural and Multidisciplinary Optimization
- Transactions on Mathematical Software

Depuis août 2019, je suis Review Editor du journal Structural and Multidisciplinary Optimization.

Dans le cadre de mes activités au MDO TC de l'AIAA, je relis

- entre 10 et 15 abstracts par an (pour la conférence annuelle AIAA/SCITECH et la conférence annuelle AIAA/AVIATION)
- ainsi que des papiers étendus (autour de 5 à 6 papiers par an) pour la "student competition" de l'AIAA/AVIATION chaque année.

9.6.2 Participation à des jurys

J'ai été sollicitée pour participer à des jurys de thèse et à des jurys de sélection.

• Jury de thèse

- A. Chiplunkar, *Incorporating Prior Information from Engineering Design into Gaussian Process Regression : with applications to Aeronautical Engineering*, 2017, Toulouse ISAE-SUPAERO [51]
- S. Delbecq, *Knowledge-Based Multidisciplinary Sizing and Optimization of Embedded Mechatronic Systems - Application to Aerospace Electro-Mechanical Actuation Systems*, 2018, INSA Toulouse [67]
- S. Coniglio, *Optimisation topologique à formalisme Eulérien et Lagrangien appliquée à la conception d'un ensemble propulsif*, 2019, Toulouse ISAE-SUPAERO [62]

• Comité de sélection

- J'ai fait partie du jury pour la campagne de recrutement d'un enseignant/chercheur en optimisation à l'ISAE-SUPAERO en 2015.
- Depuis 2016, je fais partie des jurys pour faire passer les entretiens de recrutement des candidats 1ère, 2ème, 3ème et 4ème année de l'INSA Toulouse.

Chapitre 10

Activités d'enseignement

Sommaire

10.1 Introduction	258
10.2 Optimisation mono et multidisciplinaire	258
10.2.1 Optimisation multidisciplinaire, ISAE-SUPAERO	258
10.2.2 Optimisation stochastique, ISAE-SUPAERO	259
10.2.3 Fiabilité, ISAE-SUPAERO	259
10.2.4 Optimisation de structure multidisciplinaire, ISAE-SUPAERO	260
10.2.5 Optimisation filière Masters spécialisés, ISAE-SUPAERO	260
10.2.6 Cours d'optimisation non linéaire, ISAE-SUPAERO	260
10.2.7 Projet en optimisation, ISAE-SUPAERO	260
10.2.8 Travaux dirigés de probabilités et statistiques, INSA	260
10.3 Calcul scientifique	261
10.3.1 Travaux dirigés d'introduction à la simulation numérique, ISAE-SUPAERO	261
10.3.2 Travaux dirigés de calcul scientifique, ISAE-SUPAERO	261
10.3.3 Cours pour les EDP, ISAE-ENSICA	261
10.3.4 Travaux pratiques Matlab - DEUG MASS	261
10.3.5 Travaux pratiques Internet - DEUG MASS	262
10.3.6 Travaux dirigés et projets en Analyse Numérique - première année ENSICA	262
10.3.7 Travaux pratiques de Fortran 90 - INSA, option Génie Mathématique et Modélisation	262

10.1 Introduction

Je mène une activité d'enseignement depuis ma première année de thèse en 1997 avec environ 60 heures de cours/an. De par mon parcours, j'ai enseigné dans deux domaines des mathématiques appliquées assez distincts :

- le calcul scientifique, l'analyse numérique, les équations aux dérivées partielles et les schémas numériques associés,
- les probabilités et statistiques, l'optimisation (linéaire, non-linéaire, multidisciplinaire, ...) et les modèles réduits (plans d'expériences, processus gaussiens, ...).

Dans ce cadre, j'ai été amenée à rédiger des énoncés et des corrigés de TD, des sujets et des corrections d'examens. J'ai également mis en place des BE sous Matlab et plus récemment sous python via l'utilisation de Jupyter Notebooks. Les transparents des cours en amphi sont distribués aux élèves avant chacun des cours ou mis sur le site intranet de l'école. La liste présentée ci-dessous reprend mes activités d'enseignements sur les quinze dernières années (depuis 2002 environ). Entre 1997 et 2002, j'ai essentiellement enseigné à l'INSA Toulouse (département Génie mécanique, Génie civil et Génie mathématique) un cours de probabilité et un cours de statistiques et à l'université des Sciences Sociales (DEUG MASS). Le cours mis en place à l'INSA Génie Mathématique et Modélisation sur les algorithmes stochastiques avec P. Delmoral (mesure invariante des chaînes de Markov et images fractales, fonctions itérées stochastiques et algorithme de Robbins-Monro, recuit simulé avec le problème du voyageur de commerce, filtrage non-linéaire et algorithmes génétiques) a donné lieu à un ouvrage basé sur les notes de cours paru aux Editions Cépaduès en 2001.

Depuis 2005, j'ai essentiellement enseigné à l'ENSICA et SUPAERO initialement le calcul scientifique et les équations aux dérivées partielles. Depuis 2011, j'enseigne l'optimisation mono-objectif non linéaire et depuis 2016 nous avons mis en place un cours autour de l'Optimisation MultiDisciplinaire et des métamodèles. en français et en anglais.

10.2 Optimisation mono et multidisciplinaire

10.2.1 Optimisation multidisciplinaire, ISAE-SUPAERO

Cours pour les élèves en 1ere année MAE (Master of Science in Aerospace engineering) et en 3eme année de la filière Structure, enseignement réalisé en anglais

ou en français : cours sur les formulations MDO et mise en place d'un BE d'application en python sous OpenMDAO avec un problème analytique (Sellar problem) et un problème d'optimisation aéro-structure avec [Openaerostruct](#). Ce cours a été mis en place en coordination avec mes collègues de l'ONERA S. Dubreuil et T. Lefebvre, J. R.R.A Martins (Univ du Michigan) et J. Morlier (ISAE-SUPAERO). En février 2019, nous avons organisé un cours de 3h avec Imco Van Gent du TU-Delft pour coupler OpenMDAO aux technologies développées dans le cadre du projet H2020 AGILE (CMDOWS, OpenLego, VISTOMS, [disponibles en ligne via la Open MDO suite](#))

- 8 heures/an pour les masters et 7 heures/an pour les 3A
- Années scolaires : 2017, 2018, 2019

Ce cours étant nouveau, il va sûrement encore évoluer dans les prochaines éditions.. En particulier, on envisage l'utilisation de l'application web WhatsOpt [123] pour générer facilement des modules OpenMDAO pour 2020.

10.2.2 Optimisation stochastique, ISAE-SUPAERO

Ce cours sous la forme de Notebooks python est destiné aux élèves de 3ème année, option science de la décision. Les méthodes de recuit simulé, d'algorithmes évolutionnaires (algorithmes génétiques, CMA-ES) sont vues et les élèves choisissent ensuite une méthode à expliquer aux autres (Particle Swarm Optimization, Ant colony, ...)

- 16 heures/an
- Années scolaires : 2018, 2019

10.2.3 Fiabilité, ISAE-SUPAERO

Participation aux Notebooks python pour les élèves de 3ème année option SXS (Systèmes Complexes) avec l'application de la méthode de Monte-Carlo ou de la méthode FORM (First Order Reliability Method) pour le calcul de probabilité de défaillance dans un problème d'optimisation sous contrainte.

- 4 heures/an
- Années scolaires : 2018, 2019

10.2.4 Optimisation de structure multidisciplinaire, ISAE-SUPAERO

Cours sur le krigeage et les méthodes d'optimisation par enrichissement et mise en place d'un BE sous Matlab

- 5 heures/an
- Années scolaires : 2016, 2017, 2018, 2019

10.2.5 Optimisation filière Masters spécialisés, ISAE-SUPAERO

Notebook python pour les élèves en 1ere année MAE (Master of Science in Aerospace engineering), enseignement réalisé en anglais : optimisation linéaire, programmation linéaire en nombres entiers, optimisation non linéaire, programmation par contraintes, optimisation dans les graphes, correction de l'examen

- 20 heures/an
- Années scolaires : 2016, 2017, 2018, 2019

10.2.6 Cours d'optimisation non linéaire, ISAE-SUPAERO

Co-responsable (avec G. Haine de l'ISAE) du module d'optimisation non linéaire à l'ISAE pour les 1ères années avec 3 cours magistraux (200 élèves en amphithéâtre), 4 cours en petites classes, la mise en place des BE d'optimisation sur machine (avec MATLAB), la rédaction et la correction du sujet d'examen.

- 12 heures/an
- Années scolaires : 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019

10.2.7 Projet en optimisation, ISAE-SUPAERO

Il s'agissait de suivre plusieurs projets en optimisation pour les élèves de 1ère année de SUPAERO :

- 15 heures/an
- Année scolaire : 2007

10.2.8 Travaux dirigés de probabilités et statistiques, INSA

Entre 1997 et 2001, j'ai effectué des travaux dirigés pour sensibiliser les élèves ingénieurs aux probabilités et aux statistiques avec les tests d'hypothèses. Cet

enseignement s'adressait aux élèves de l'INSA en option génie mécanique avec un volume de 28h par an.

10.3 Calcul scientifique

10.3.1 Travaux dirigés d'introduction à la simulation numérique, ISAE-SUPAERO

Ce module d'introduction à la simulation numérique comprenait une initiation à Matlab et des travaux dirigés pour la mise en place de schémas numériques (équations aux dérivées ordinaires, différences finies, éléments finis 1D) aux élèves de 1ère année de ISAE-SUPAERO :

- 12 heures/an
- Années scolaires : 2010, 2011, 2012, 2013, 2014, 2015

10.3.2 Travaux dirigés de calcul scientifique, ISAE-SUPAERO

Il s'agissait de donner un enseignement en calcul scientifique (différences finies, éléments finis, volumes finis) aux élèves de 2ème année de ISAE-SUPAERO :

- 12 heures/an
- Années scolaires : 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015

10.3.3 Cours pour les EDP, ISAE-ENSICA

Il s'agissait de donner un enseignement théorique sur les Equations aux dérivées partielles aux étudiants de 2ème année de l'ENSICA.

- 25 heures de cours en petite classe (20 étudiants)
- Années scolaires : 2006, 2008, 2009, 2010, 2011, 2013, 2014, 2015

10.3.4 Travaux pratiques Matlab - DEUG MASS

Il s'agissait d'illustrer les notions d'algèbres linéaires avec le logiciel Matlab.

- 2×30 étudiants
- 15 heures
- Année scolaire : 2002-03, 2003-04

10.3.5 Travaux pratiques Internet - DEUG MASS

Il s'agissait de mettre en place une page internet en utilisant le logiciel Composer de Netscape.

- 2×30 étudiants
- 15 heures
- Année scolaire : 2002-03, 2003-04

10.3.6 Travaux dirigés et projets en Analyse Numérique - première année ENSICA

Des notions générales en analyse numérique étaient enseignées sous la forme de bureau d'étude (optimisation, différences finies, éléments finis, intégration numérique, splines, équations différentielles). Des projets (environ 20 heures par binôme) sur des problématiques physiques sont proposés aux élèves pour mettre en application ces notions avec le logiciel Matlab.

- 2×30 étudiants
- 30 heures
- Année scolaire : 2002-03, 2003-04, 2004-05

10.3.7 Travaux pratiques de Fortran 90 - INSA, option Génie Mathématique et Modélisation

Des travaux pratiques de Fortran 90 étaient proposés aux élèves de première année de la filière Mathématique et Modélisation pour illustrer des notions algorithmiques (boucles, récursivité, manipulation de chaînes, tableaux, listes...) et des notions de programmation (modules, debugueur, compilation et makefile).

- 2×30 étudiants
- 30 heures
- Année scolaire : 2003-04

Chapitre 11

Conclusion et perspectives de recherche

Un autre verrou concerne la gestion des contraintes (contraintes physiques connues et/ou contraintes inconnues liées à la robustesse du code de calcul) avec en particulier la gestion des succès ou échecs des calculs. Différentes méthodes existent dans la littérature [89] mais devront être étudiées sur des problèmes de grande taille. Le couplage avec des techniques d'apprentissage pour le partitionnement pourrait s'avérer efficace et permettre ainsi de traiter des problèmes plus proches des problématiques industrielles.

L'hybridation de SEGOMOE avec des algorithmes existants pourrait être intéressante en fonction de l'application. Je pense notamment aux algorithmes évolutionnaires comme CMAES [103]. Des projets étudiants en collaboration avec l'ISAE ont été lancés depuis 2017 pour proposer une hybridation de EGO et de CMAES afin d'accélérer le processus de convergence [52, 150, 69]. Les développements n'ont pas encore assez avancé pour obtenir les premiers résultats mais la piste semble prometteuse. L'enrichissement multipoint qui permet de choisir plusieurs points intéressants à évaluer [50] et ainsi paralléliser les calculs coûteux a été implémenté mais des tests restent à mener pour montrer l'accélération de la convergence en fonction des différents critères de la méthode (nombre de points, valeur estimée, ...).

Aujourd'hui SEGOMOE ne traite que les variables continues, ce qui est une limite de la méthode pour traiter l'optimisation de rotors ou l'optimisation d'architectures distribuées ; dans ces cas d'application le nombre de pales ou le nombre de batteries doivent être considérés. L'extension aux variables discrètes est donc prévue au travers des différentes approches proposées récemment [174, 161]. Le caractère mono-objectif de SEGOMOE est aujourd'hui une limite pour traiter les compromis et l'évolution vers le multi-objectif semble nécessaire. Lors de projets étudiants précédents, nous avons mené différentes campagnes de tests pour utiliser l'algorithme multiEGO [89] dans la plateforme OpenMDAO. Le changement radical de version d'OpenMDAO en juillet 2015 avec la non compatibilité avec les codes précédemment développés a considérablement nui à la capitalisation de ces investigations. Des discussions autour de l'algorithme MGDA par J-A Désidéri [68] ont eu lieu à plusieurs reprises. Lors d'un projet étudiant, nous avons utilisé les modèles de krigeage et leurs dérivées analytiques comme entrées de l'algorithme MGDA et les résultats permettaient d'obtenir des points du front de Pareto en limitant l'appel aux codes de calcul coûteux. Les derniers travaux sur **MGDA** pour la prise en compte de contraintes pourraient permettre d'étendre l'approche aux problèmes contraints. Dans la revue sur les algorithmes d'optimisation faite par R. Haftka en 2016 [102], la méthode EGO est largement citée. Cependant, il existe des alternatives aux modèles de krigeage afin d'obtenir une estimation d'er-

reur. J'ai mené en 2010 et 2018 des activités autour des techniques de bootstrap qui pourraient être appliquées pour calculer cette estimation à partir de modèles type réseaux de neurones. De même, la construction et la comparaison de modèles de nature très diverses est une piste qui pourrait être étudiée dans notre cadre d'application.

En conclusion, l'algorithme SEGOMOE pourrait donc d'une part être étendu aux variables mixtes [174, 161], et d'autre part être amélioré via des approches hybrides (avec CMAES ou NOMAD par exemple), avec de nouveaux modèles disponibles (multiple surrogate techniques de Viana [199, 198]), via de nouveaux critères d'enrichissement en mono-objectif [163, 90] ou en multiobjectif [89, 164, 105]. L'algorithme résultant devra faciliter le calcul parallèle et ainsi minimiser davantage les temps de restitution.

La gestion des données multi-fidélité a été initiée avec une évolution de SEGOMOE via un modèle de co-kriging [146]. Cela consiste à construire des métamodèles en utilisant les différentes sources d'information et en choisissant à chaque étape le nouveau point à ajouter et le niveau de fidélité adapté pour la simulation. La première application a été menée sur une seule discipline aérodynamique avec deux niveaux de fidélité. Une extension à deux disciplines (aérodynamique et structure) est en cours en collaboration avec le MDOLab et l'ISAE-SUPAERO. Dans les premières expérimentations, les points proposés par la méthode SEGOMOE sont rajoutés dans les deux disciplines avec le niveau de fidélité choisi mais identique pour chacune des deux disciplines. La prochaine étape serait de pouvoir différencier le niveau de fidélité adapté à chacun des deux solveurs. Cette problématique sur le choix des modèles à enrichir est également présente dans le travail que nous avons mené sur la propagation d'incertitudes de modèles dans un processus d'optimisation multidisciplinaire. Cette source d'incertitude est directement liée à la fidélité du modèle que l'on considère : basse fidélité, modèle analytique simplifié, utilisation d'un métamodèle. Dès lors que l'on arrive à la quantifier via une distribution de probabilité, nous sommes capables de propager cette source d'incertitude jusqu'à l'optimum et ainsi identifier les modèles à raffiner pour trouver le design optimal. La loi des valeurs extrêmes de la fonction objectif a été modélisée et les premiers résultats d'intégration dans un processus d'optimisation avec enrichissement sont très encourageants. Le passage à l'échelle industrielle risque d'amener des difficultés supplémentaires pour la prise en compte des incertitudes surtout si l'on ajoute à cela plusieurs modèles de fidélités différentes. Des méthodes plus récentes et moins coûteuses que les polynômes du chaos basées sur des approximations de rang faible [121, 122] pourraient être étudiées. Quelle stratégie adopter, faut-il enrichir tous les modèles à chaque itération, les moins coûteux, les plus

influent? L'analyse basée sur les indices de Sobol que nous avons menée sur le cas test analytique de l'avion supersonique [72] est une première réponse pour le choix des modèles à enrichir mais cela nécessiterait une étude plus approfondie pour prendre en compte les temps de retour et le coût associé aux différents modèles. Des travaux de recherche sont en cours avec l'ISAE-SUPAERO et une collaboration avec l'Université de Southampton viendra compléter cette étude en fournissant des erreurs quantifiées entre les modules de différentes fidélités.

Concernant l'optimisation des temps de calcul, des améliorations pourraient être menées au niveau de la formulation MDO : par exemple la formulation ASO pour Asymmetric Subspace Optimization [138] permet de tenir compte des disparités importantes dans les temps de simulation entre les disciplines utilisées dans les processus MDO avec un exemple pour un couplage aéro-structure [116]. Dès lors que de la multi-fidélité est introduite dans le processus, cette formulation pourrait avoir un intérêt non négligeable. Le démarrage de la thèse CIFRE avec ATR et l'ISAE-SUPAERO autour d'un processus de conception multi-fidélité pourrait être le cadre applicatif de cette formulation.

Un dernier point plus prospectif concerne les métamodèles dynamiques. Mon expérience sur les réseaux de neurones au travers de la thèse de M. Gruet m'a permis de mieux comprendre les nouveaux réseaux de type LSTM (Long Short Term Memory) issus des développements en intelligence artificielle qui ont la capacité à gérer automatiquement la fenêtre temporelle sur laquelle les entrées doivent être fournies. Dans le domaine de l'analyse sismique, des travaux récents [136] ont couplé le modèle dynamique NARX (Non linear AutoRegressive with exogenous inputs) avec les polynômes du chaos et ainsi obtenir une réponse dynamique prenant en compte les incertitudes des paramètres d'entrée. Un autre couplage entre les réseaux de neurones et les processus gaussiens a eu lieu dans le cadre de la thèse de Marina [101] appliqué à la prévision météorologie spatiale. Le modèle couplé permet d'une part d'obtenir une prédiction en temps de meilleure qualité que les réseaux classiques mais également de disposer des intervalles de confiance et ainsi d'évaluer le risque maximal pour la prévention des orages magnétiques. La modélisation et l'optimisation de systèmes dynamiques par des méthodes basées sur les processus gaussiens seraient donc un nouveau champ d'application et d'évolution de l'algorithme SEGOMOE.

Tous les développements de méthodes et d'outils sont capitalisés dans la plateforme interne de l'ONERA (WhatsOpt [123]) qui peut proposer un web service accessible via l'extérieur. Certains des développements viendront compléter la toolbox SMT. L'objectif étant de pouvoir appliquer les différents développements méthodologiques aux configurations avion classiques et dérivées (exemple : nou-

velle motorisation) mais aussi aux drones à voilure fixe, à l'aile volante avec propulsion distribuée,... La mise à jour de cette plate-forme avec la capitalisation de ces différents développements permettront de résoudre la problématique de conception des aéronefs du futur présentant des architectures innovantes.

Concernant les enseignements menés autour des métamodèles, de l'optimisation mono et multidisciplinaire, je souhaite les poursuivre en intégrant dans la formation l'utilisation des toolbox que l'on co-développe. Une première étape a été de mettre en place la formation basée sur OpenMDAO et la méthode EGO depuis 2017. Depuis février 2019, nous avons intégré les outils opensource développés dans le projet européen AGILE [Open MDO Suite framework](#). La prochaine étape sera de mettre en place des notebooks autour de la toolbox [SMT](#). A plus long terme, je souhaiterais mettre en place un cours liant la MDO et les outils MBSE dans le cadre de la fédération en particulier.

Par ailleurs, la poursuite des collaborations avec l'ISAE-SUPAERO, l'ICA, l'IRT, le MDOLab, le DLR, le NLR, le TUDelft, la NASA, le Tsagi, l'université HKUST de Hong-Kong, l'université de Mc-Gill de Montréal, l'université de Clarkson,... vont se poursuivre et s'amplifier de par notamment la participation à des programmes d'échanges pour les étudiants en thèse et l'organisation de workshop.

Chapitre 12

Appendix

Sommaire

12.1	Fonction G07	270
12.2	Fonction de Griewank	270
12.3	Fonction de Branin Modifiée	270

Cette annexe décrit les fonctions analytiques utilisées dans les exemples de ce manuscrit.

12.1 Fonction G07

Objectif et contraintes

$$\left\{ \begin{array}{l} f(x) = x_1^2 + x_2^2 + x_1x_2 - 14x_1 - 16x_2 + (x_3 - 10)^2 + 4(x_4 - 5)^2 \\ + (x_5 - 3)^2 + 2(x_6 - 1)^2 + 5x_7^2 + 7(x_8 - 11)^2 + 2(x_9 - 10)^2 + (x_{10} - 7)^2 + 45 \\ g_1 = \frac{4x_1+5x_2-3x_7+9x_8-105}{105} \leq 0 \\ g_2 = \frac{10x_1-8x_2-17x_7+2x_8}{370} \leq 0 \\ g_3 = \frac{-8x_1-2x_2+5x_9-2x_{10}-12}{158} \leq 0 \\ g_4 = \frac{3(x_1-2)^2+4(x_2-3)^2+2x_3^2-7x_4-120}{1258} \leq 0 \\ g_5 = \frac{5x_1^2+8x_2+(x_3-6)^2-2x_4-40}{816} \leq 0 \\ g_6 = \frac{0.5(x_1-8)^2+2(x_2-4)^2+3x_5^2-x_6-30}{834} \leq 0 \\ g_7 = \frac{x_1^2+2(x_2-2)^2-2x_1x_2+14x_5-6x_6}{788} \leq 0 \\ g_8 = \frac{-3x_1+6x_2+12(x_9-8)^2-7x_{10}}{4048} \leq 0 \end{array} \right.$$

Domaine de recherche

$$x_i \in [-10, 10] \quad \forall i \in [1, \dots, 10]$$

12.2 Fonction de Griewank

Objectif

$$f(x) = \sum_{i=1}^d \frac{x_i^2}{4000} - \prod_{i=1}^d \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1$$

Domaine de recherche

$$x_i \in [-600, 600] \quad \forall i \in [1, \dots, d]$$

12.3 Fonction de Branin Modifiée

La fonction de Branin modifiée [158] est issue de la fonction de Branin normalisée sur $[0, 1]$ [89]. La fonction contrainte est une fonction normalisée issue

de la fonction Gomez#3 [177], avec l'ajout d'un sinus pour accroître le caractère multimodale.

Objectif

$$f(x_1, x_2) = \left(x_2 - \frac{5.1}{4\pi^2}x_1^2 + \frac{5}{\pi}x_1 - 6\right)^2 + 10 \left[\left(1 - \frac{1}{8\pi}\right) \cos(x_1) + 1\right] + \frac{5x_1 + 25}{15}$$

Contrainte

$$g(x_1, x_2) = \left(4 - 2.1y^2 + \frac{y^4}{3}\right)y^2 + yz + 4(y^2 - 1)z^2 + 3 \sin [6(1 - y)] + 3 \sin [6(1 - z)] - 6$$

where $y = \frac{x_1 - 2.5}{7.5}$, $z = \frac{x_2 - 7.5}{7.5}$

$$g(x_1, x_2) \geq 0$$

Domaine de recherche

$$x_1 \in [-5, 10] \quad x_2 \in [0, 15]$$

Minimum global

$$f(x^*) = 12.005$$

Bibliographie

- [1] Aviation renaissance : NASA advances concepts for next-gen aircraft. <https://www.nasa.gov/feature/aviation-renaissance-nasa-advances-concepts-for-next-genaircraft>.
- [2] MultiFi Cokriging in OpenMDAO. http://openmdao.org/twodocs/versions/2.0.0/_srcdocs/packages/surrogate_models/multifi_cokriging.html. Accessed : 2014-11-18.
- [3] An open source, platform-independent, flight dynamics & control software library in C++. <http://jsbsim.sourceforge.net/>.
- [4] Scimago journal & country rank. <https://www.scimagojr.com/>.
- [5] Web of knowledge. <https://apps.webofknowledge.com/>.
- [6] B. M. Adams, L. E. Bauman, W. J. Bohnhoff, K. R. Dalbey, J. P. Eddy, M. S. Ebeida, M. S. Eldred, P. D. Hough, K. T. Hu, J. D. Jakeman, L. P. Swiler, J. A. Stephens, D. M. Vigil, and T. M. Wildey. Dakota, a multi-level parallel object-oriented framework for design optimization, parameter estimation, uncertainty quantification, and sensitivity analysis : Version 6.1 users manual. Technical Report SAND2014-4633, Sandia National Laboratories, Albuquerque, NM, November 2014. Available online from <http://dakota.sandia.gov/documentation.html>.
- [7] Jeremy S Agte, Jaroslaw Sobieszczanski-Sobieski, and Robert R Sandusky. Supersonic business jet design through bi-level integrated system synthesis. Technical report, SAE Technical Paper, 1999.
- [8] John David Anderson. *Aircraft performance and design*, volume 1. WCB/McGraw-Hill Boston, MA, 1999.
- [9] Miguel F Anjos and Donald R Jones. Mopta 2008 benchmark, 2009.
- [10] C. Audet, S. Le Digabel, and C. Tribes. NOMAD user guide. Technical Report G-2009-37, Les cahiers du GERAD, 2009.

- [11] Charles Audet and Warren Hare. *Derivative-free and blackbox optimization*. Springer, 2017.
- [12] D. Bailly, N. Bartoli, C. Blondeau, P. Klotz, and J.-S. Schotté. Projet de recherche fédérateur DOOM : comparaison de modèles réduits sur bases de données fixes. Technical Report RT 3/12607 DTIM/DPRS, ONERA, Juin 2008.
- [13] Mathieu Balesdent, Jérôme Morio, and Julien Marzat. Kriging-based adaptive importance sampling algorithms for rare event estimation. *Structural Safety*, 44 :1–10, 2013.
- [14] N. Bartoli, V. Baudoui, and P. Klotz. Plate-forme pour la construction de modèles réduits. Technical Report RT 1/13619 DTIM, ONERA, Décembre 2008.
- [15] N. Bartoli and D. Bettebghor. Image : Surrogate models for stacking sequence optimization. Technical Report RT 2/14190 DMAE/DTIM, ONERA, Octobre 2011.
- [16] N. Bartoli, D. Bettebghor, F-H Leroy, and F-X Irisarri. Artemis : Construction de métamodèles pour des problèmes de flambement de panneaux raidis. Technical Report RT 2/16855 DCPS/DTIM/DMSC, ONERA, 2012.
- [17] N. Bartoli, D. Bettebghor, F. Luraghi, and M. Samuelides. On some applications of surrogate model-based optimization techniques to the design of composite panels for aeronautical applications. AFG’11, 15th Austrian-French-German conference on Optimization, Toulouse, September 19-23, 2011 2011.
- [18] N. Bartoli, G. Billet, M. Borrel, V. Chiaruttini, P. Delorme, X. Ferrières, F. Feyel, S. Laurens, B. Michielsen, P. Mazet, V. Mouysset, C. Peyret, F. Rogier, J. Ryan, and E. Savin. PRF MAHPSO. Technical Report ONERA RF 1/15040 DTIM ; DTIM T54550, ONERA, 2009.
- [19] N. Bartoli, T. Druot, T. Lefebvre, F. Gallard, A. Gazaix, D. Guénot, and C. Vanaret. Construction, analysis and management of the overall aircraft constraints. Technical Report LIV-S-015-M2.4.4-218-V0, IRT Saint-Exupéry, projet MDA-MDO, December 2016.
- [20] N. Bartoli, M. Samuelides, D. Bailly, and M. Marcelet. Revue de modèles réduits pour l’optimisation. Technical Report RT 2/11576 DPRS/DTIM, ONERA, Décembre 2006.
- [21] Nathalie Bartoli, Mohamed-Amine Bouhleb, Igor Kurek, Rémi Lafage, Thierry Lefebvre, Joseph Morlier, Rémy Priem, Vivien Stilz, and Rommel

- Regis. Improvement of efficient global optimization with application to aircraft wing design. In *17th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, page 4001, Washington D.C., USA, 2016.
- [22] Nathalie Bartoli, Thierry Lefebvre, Sylvain Dubreuil, Romain Olivanti, Nicolas Bons, Joaquim Martins, Mohamed-Amine Bouhlef, and Joseph Morlier. An adaptive optimization strategy based on mixture of experts for wing aerodynamic design optimization. In *18th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, page 4433, 2017.
- [23] Nathalie Bartoli, Thierry Lefebvre, Sylvain Dubreuil, Romain Olivanti, Rémy Priem, Nicolas Bons, Joaquim R.R.A. Martins, and Joseph Morlier. Adaptive modeling strategy for constrained global optimization with application to aerodynamic wing design. *Aerospace Science and Technology*, 90 :85–102, 2019.
- [24] Nathalie Bartoli, Thierry Lefebvre, Sylvain Dubreuil, Marco Panzeri, Roberto D’Ippolito, Kirill Anisomov, and Andrey Savelyev. Robust nacelle optimization design investigated in the AGILE european project. In *19th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, Atlanta, United States, 2018.
- [25] Nathalie Bartoli and Armand Oyzel. Amélioration des performances du modèle acoustique CARTONOISE de la chaîne informatique IESTA - phase 2 : Introduction de techniques RSM pour la génération d’empreintes acoustiques. Technical Report RTI 1/20838, ONERA, 2013.
- [26] Peter Schmollgruber Nathalie Bartoli and Yves Gourinat. Virtual flight testing in an aircraft sizing and optimization process. Dallas, June 2015. 15th AIAA Aviation Technology, Integration, and Operations Conference.
- [27] Pierre-Marie Basset, Arnault Tremolet, Nathalie Bartoli, and Thierry Lefebvre. Helicopter presizing by multidisciplinary-multi objective optimization. In *1st International Conference on Engineering and Applied Sciences Optimization*, 2014.
- [28] Gaspard Berthelin, Sylvain Dubreuil, Michel Salaün, Christian Gogu, and Nathalie Bartoli. Couplage optimisation et réduction de modèles par projection : Application à la conception avion avant-projet. Giens (Var), Mai 2019. 14e Colloque National en calcul des structures.
- [29] D. Bettebghor. Description des développements méthodologiques et informatiques de la bibliothèque de modèles réduits par mélanges d’experts image (improved metamodelling approximation through gaussian mixture of experts). Technical Report RT 1/20166 DTIM/DADS, ONERA, 2012.

- [30] D. Bettebghor and N. Bartoli. Application des mélanges d’experts à l’approximation des contraintes de flambage en calcul de structures. ONERA JSO Modèles de Substitution & Optimisation Multi-Disciplinaire, Janvier 2012.
- [31] D. Bettebghor, N. Bartoli, and M. Samuelides. Embed surrogate model for multi-level optimization for composite fuselage barrel. Technical Report RT 3/14190 DMAE/DTIM, ONERA, Novembre 2011.
- [32] D. Bettebghor, N. Bartoli, and M. Samuelides. Enhancing approximation of critical buckling factor for laminated composites. Braunschweig, Germany, 4th to 6th June 2012 2012. ODAS 2012, 12th Onera-DLR Aerospace Symposium.
- [33] D. Bettebghor, N. Bartoli, M. Samuelides, J. Morlier, and S. Grihon. Approche en paramètres de stratification pour l’optimisation biniveau de structures composites. Giens (var), 9-13 mai 2011 2011. CSMA, 10e Colloque national en calcul des structures.
- [34] Dimitri Bettebghor. *Bilevel optimization of large scale composite structures*. PhD thesis, ISAE, Toulouse, december 2011.
- [35] Dimitri Bettebghor and Nathalie Bartoli. Approximation of the critical buckling factor for composite panels. *Structural and Multidisciplinary Optimization*, 46(4) :561–584, 2012. article soumis en aout 2011 et modifié en janvier 2012.
- [36] Dimitri Bettebghor, Nathalie Bartoli, Stéphane Grihon, Joseph Morlier, and Manuel Samuelides. Surrogate modeling approximation using a mixture of experts based on em joint estimation. *Structural and Multidisciplinary Optimization*, 43(2) :243–259, 2011. 10.1007/s00158-010-0554-2.
- [37] Barron J Bichon, Michael S Eldred, Laura Painton Swiler, Sandaran Mahadevan, and John M McFarland. Efficient global reliability analysis for nonlinear implicit performance functions. *AIAA journal*, 46(10) :2459–2468, 2008.
- [38] G Blatman and B Sudret. Efficient computational of global sensitivity indices using sparse polynomial chaos expansions. *Reliability Engineering & System Safety*, 95 (11) :1216–1229, 2010.
- [39] Mohamed A Bouhlel and Joaquim RRA Martins. Gradient-enhanced kriging for high-dimensional problems. *Engineering with Computers*, pages 1–17, 2017.
- [40] Mohamed-Amine Bouhlel. *Optimisation auto-adaptative en environnement d’analyse multo-disciplinaire via les modèles de Krigeage combinés à la*

- méthode PLS*. PhD thesis, ISAE-SUPAERO, 2016. <https://hal.archives-ouvertes.fr/tel-01293319>.
- [41] Mohamed-Amine Bouhleb and Nathalie Bartoli. Avancement de la thèse au cours de la première année. Technical Report RA 1/21734 DCPS, ONERA, septembre 2013.
- [42] Mohamed-Amine Bouhleb and Nathalie Bartoli. Second year of the PhD thesis, progress report : SEGOKPLS, a new algorithm for derivative-free optimization of expensive black-box objective functions subject to inequality constraints. Technical Report RA 2/21734 DCPS, ONERA, september 2014.
- [43] Mohamed Amine Bouhleb, Nathalie Bartoli, Joseph Morlier, and Abdelkader Otsmane. Kriging combining partial least squares for optimization in high dimension. EngOpt 2014, 4th International Conference on Engineering Optimization, Lisbon, 8-11 September 2014.
- [44] Mohamed Amine Bouhleb, Nathalie Bartoli, Joseph Morlier, and Abdelkader Otsmane. Surrogate models for optimization in high dimension using a mixed kriging/pls method. PLS2014, 8th International Conference on Partial Least Squares and Related Method, Paris, 26-28 May 2014.
- [45] Mohamed Amine Bouhleb, Nathalie Bartoli, Joseph Morlier, and Abdelkader Otsmane. Surrogate models for optimization in high dimension using a mixed PLS/Kriging method. OPT-i 14, Greece, 4-6 June 2014.
- [46] Mohamed Amine Bouhleb, Nathalie Bartoli, Abdelkader Otsmane, and Joseph Morlier. An improved approach for estimating the hyperparameters of the kriging model for high-dimensional problems through the partial least squares method. *Mathematical Problems in Engineering*, 2016, 2016.
- [47] Mohamed Amine Bouhleb, Nathalie Bartoli, Abdelkader Otsmane, and Joseph Morlier. Improving kriging surrogates of high-dimensional design models by partial least squares dimension reduction. *Structural and Multidisciplinary Optimization*, 53(5) :935–952, 2016.
- [48] Mohamed Amine Bouhleb, Nathalie Bartoli, Rommel G. Regis, Abdelkader Otsmane, and Joseph Morlier. Efficient global optimization for high-dimensional constrained problems by using the kriging models combined with the partial least squares method. *Engineering Optimization*, 0(0) :1–16, 2018.
- [49] Mohamed Amine Bouhleb, John T Hwang, Nathalie Bartoli, Rémi Lafage, Joseph Morlier, and Joaquim RRA Martins. A python surrogate modeling framework with derivatives. *Advances in Engineering Software*, page 102662, 2019.

- [50] Clément Chevalier. *Fast uncertainty reduction strategies relying on Gaussian process models*. Theses, Universität Bern, September 2013.
- [51] Ankit Chiplunkar. *Incorporating Prior Information from Engineering Design into Gaussian Process Regression : with applications to Aeronautical Engineering*. PhD thesis, Toulouse, ISAE, 2017.
- [52] Rudy Chocat, Loïc Brevault, Mathieu Balesdent, and Sébastien Defoort. Modified covariance matrix adaptation–evolution strategy algorithm for constrained optimization under uncertainty, application to rocket design. *International Journal for Simulation and Multidisciplinary Design Optimization*, 6 :A1, 2015.
- [53] Pier Davide Ciampa and Björn Nagel. Towards the 3rd generation mdo collaborative environment. *30th ICAS*, 2016.
- [54] J. Clément, M. Masmoudi, N. Bartoli, and J. Hermetz. Multi-disciplinary optimization methods studies. 2nd European Conference for Aero-Space Sciences EUCASS, Brussels, Belgium, July 1–6, 2007.
- [55] J. Clément, M. Masmoudi, J. Hermetz, N. Bartoli, and Y. Parte. DIVE : une méthode d’utilisation de méta-modèles dans un processus d’optimisation multi-disciplinaire. 9th Colloque national en calcul des structures, Giens, 25–29 mai 2009.
- [56] Joël Clément. *Optimisation multidisciplinaire : étude théorique et application à la conception des avions en phase d’avant projet*. PhD thesis, ISAE-Supaéro, juin 2009.
- [57] Joan Mas Colomer. Progress of the phd thesis during the 1st year : Aerostructural optimization of a fixed-wing drone. Technical Report RA 1/25487 DCPS, ONERA, October 2016.
- [58] Joan Mas Colomer. Progress of the phd thesis during the 2nd year : Aerostructural optimization of a fixed-wing drone. Technical Report RA 2/25487 DCPS, October 2017.
- [59] Joan Mas colomer. *Aeroelastic Similarity of a Flight Demonstrator via Multidisciplinary Optimization*. PhD thesis, ISAE-SUPAERO, December 2018.
- [60] Joan Mas Colomer, Nathalie Bartoli, Thierry Lefebvre, Sylvain Dubreuil, Peter Schmollgruber, and Joseph Morlier. Static and dynamic aeroelastic scaling of the crm wing via multidisciplinary optimization. In *WCSMO12 12th World Congress of Structural and Multidisciplinary Optimisation*, Braunschweig, Germany, 2017.

- [61] Joan Mas Colomer, Nathalie Bartoli, Thierry Lefebvre, and Joseph Morlier. Static aeroelastic scaling with non-similar flow through multidisciplinary optimization. Lisboa, Portugal, September 2018. Engopt : 6th International Conference on Engineering Optimization.
- [62] Simone Coniglio. *Optimisation topologique à formalisme Eulérien et Lagrangien appliquée à la conception d'un ensemble propulsif*. PhD thesis, Toulouse, ISAE, 2019.
- [63] Paul G Constantine. *Active subspaces : Emerging ideas for dimension reduction in parameter studies*. SIAM, 2015.
- [64] R I Cukier, J H Schaibly, and K E Shuler. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. iii. analysis of the approximations. *The Journal of Chemical Physics*, 63 :1140–1149, 1975.
- [65] Benoit Dabas, Nathalie Bartoli, Thierry Lefebvre, François Gallard, Anne Gazaix, Thierry Druot, and Damien Guénot. Error-based adaptive coupling process between multipoint high-fidelity aerodynamics and mission performance for shape optimization in the MDA-MDO project. In *20th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, 2019.
- [66] Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and TAMT Meyarivan. A fast and elitist multiobjective genetic algorithm : Nsga-ii. *Evolutionary Computation, IEEE Transactions on*, 6(2) :182–197, 2002.
- [67] Scott Delbecq. *Knowledge-Based Multidisciplinary Sizing and Optimization of Embedded Mechatronic Systems - Application to Aerospace Electro-Mechanical Actuation Systems*. PhD thesis, INSA Toulouse, 2018.
- [68] Jean-Antoine Désidéri. Multiple-gradient descent algorithm (mgda) for multiobjective optimization. *Comptes Rendus Mathématique*, 350(5-6) :313–318, 2012.
- [69] Y Diouane, S Gratton, and LN Vicente. Globally convergent evolution strategies for constrained optimization. *Computational Optimization and Applications*, pages 1–24, 2015.
- [70] Mark Drela. Xfoil : An analysis and design system for low reynolds number airfoils. In *Low Reynolds number aerodynamics*, pages 1–12. Springer, 1989.
- [71] S. Dubreuil, N. Bartoli, C. Gogu, and T. Lefebvre. Multidisciplinary optimization by surrogate models : handling epistemic uncertainties by polynomial chaos expansion. Crete Island, Greece, June 2016. ECCOMAS Congress, European Congress on Computational Methods in Applied Sciences and Engineering.

- [72] S Dubreuil, N Bartoli, C Gogu, and T Lefebvre. Propagation of modeling uncertainty by polynomial chaos expansion in multidisciplinary analysis. *Journal of Mechanical Design*, 138(11) :111411, 2016.
- [73] S. Dubreuil, N. Bartoli, C. Gogu, T. Lefebvre, and J. Mas Colomer. Extreme value oriented random field discretization based on an hybrid polynomial chaos expansion — kriging approach. *Computer Methods in Applied Mechanics and Engineering*, 332 :540 – 571, 2018.
- [74] S Dubreuil, M Berveillier, F Petitjean, and M Salaün. Construction of bootstrap confidence intervals on sensitivity indices computed by polynomial chaos expansion. *Reliability Engineering & System Safety*, 121 (Supplement C) :263–275, 2014.
- [75] Sylvain Dubreuil, Nathalie Bartoli, Thierry Lefebvre, and Christian Gogu. Towards an efficient global mdo based on surrogate models. In *1st European Workshop on MDO for Industrial Applications in Aeronautics – Challenges and Expectations*, Braunschweig, Germany, October 2017.
- [76] Sylvain Dubreuil, Nathalie Bartoli, Thierry Lefebvre, and Christian Gogu. Efficient global multidisciplinary optimization based on surrogate models. In *19th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, Atlanta, United States, 2018.
- [77] Sylvain Dubreuil, Nathalie Bartoli, Thierry Lefebvre, and Christian Gogu. Multidisciplinary optimization based on surrogate models. In *OLA 2018, Workshop on Optimization and Learning : Challenges and Applications*, Alicante, Spain, February 2018.
- [78] Sylvain Dubreuil, Christian Gogu, Nathalie Bartoli, and Thierry Lefebvre. An adaptive kriging based approach to robust optimization. Greece, 15-17 June 2017. UNCECOMP 2017, 2nd International Conference on Uncertainty Quantification in Computational Sciences and Engineering.
- [79] Sylvain Dubreuil, Christian Gogu, Nathalie Bartoli, and Thierry Lefebvre. Adaptive random field discretization for optimization under uncertainties. In *WCSMO12 12th World Congress of Structural and Multidisciplinary Optimisation*, Braunschweig, Germany, 2017.
- [80] Sylvain Dubreuil, Joan Mas-Colomer, Michel Salaün, Nathalie Bartoli, and Thierry Lefebvre. Évaluation de l’incertitude associée à l’interpolation de maillage dans un calcul couplé partitionné. Giens (Var), Mai 2017. 13e Colloque National en calcul des structures.

- [81] B Echard, N Gayton, and M Lemaire. AK-MCS : an active learning reliability method combining kriging and monte carlo simulation. *Structural Safety*, 33(2) :145–154, 2011.
- [82] Bradley Efron and Robert Tibshirani. *An introduction to the Bootstrap*. Chapman & Hall, 1993.
- [83] Michael A Epton and Alfred E Magnus. PANAIR, a computer program for predicting subsonic and supersonic potential flows about arbitrary configurations using a high order panel method, volume i, theoretical manual, version 3.0. *Theoretical Manual, Version 3.0. NASA-CR-3251, Revision 1.*, 1, 1992.
- [84] M. Giselle Fernández-Godino. *Quantifying Particle Departure from Axisymmetry in Multiphase Cylindrical Detonation*. PhD thesis, University of Florida, 2018.
- [85] M Giselle Fernández-Godino, Sylvain Dubreuil, Nathalie Bartoli, Christian Gogu, S Balachandar, and Raphael T Haftka. Linear regression based multi-fidelity surrogate for disturbance amplification in multi-phase explosion. *Structural and Multidisciplinary Optimization, Gainesville, FL*, 2019.
- [86] M Giselle Fernández-Godino, Raphael T Haftka, Sivaramakrishnan Balachandar, Christian Gogu, Nathalie Bartoli, and Sylvain Dubreuil. Noise filtering and uncertainty quantification in surrogate based optimization. In *2018 AIAA Non-Deterministic Approaches Conference*, page 2176, Kissimmee, United States, 2018.
- [87] X. Ferrieres, J.B. Laurent, V. Mouysset, T. Volpert, N. Bartoli, F. Renac, C. Marmignon, J. Ryan, S. Feld-Payet, and V. Chiaruttini. Rapport de synthèse ARF DMORPH concernant les travaux de la première année. Technical Report ONERA RTS 1/18669 DEMR/DSNA/DMSM/DTIM, ONERA, 2012.
- [88] Alexander Forrester, Andras Sobester, and Andy Keane. Multi-fidelity optimization via surrogate modelling. 463 :3251–3269, 12 2007.
- [89] Alexander Forrester, Andras Sobester, and Andy Keane. *Engineering design via surrogate modelling : a practical guide*. John Wiley & Sons, 2008.
- [90] Peter I Frazier. A tutorial on bayesian optimization. *arXiv preprint arXiv :1807.02811*, 2018.
- [91] Anne Gazaix, François Gallard, Vincent Ambert, Damien Guénot, Maxime Hamadi, Patrick Sarouille, Vincent Gachelin, Justin Plakoo, Nicolas Desfachelles, Nathalie Bartoli, Thierry Lefebvre, and Stéphane Grihon. MDA-MDO project : Advanced MDO processes, achievements and results. In

- 20th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, 2019.
- [92] Anne Gazaix, François Gallard, Vincent Gachelin, Thierry Druot, Stéphane Grihon, Vincent Ambert, Damien Guénot, Rémi Lafage, Charlie Vanaret, Benoit Pauwels, Nathalie Bartoli, Thierry Lefebvre, Patrick Sarouille, Nicolas Desfachelles, Joel Brézillon, Maxime Hamadi, and Selime Gurol. Towards the industrialization of new mdo methodologies and tools for aircraft design. In *18th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, page 3149, 2017.
- [93] Marie Gibert. Towards an optimization for an aircraft wing design. Technical report, ISAE-SUPAERO/ONERA, February 2017.
- [94] Justin Gray, Kenneth T. Moore, Tristan A. Hearn, and Bret A. Naylor. Standard platform for benchmarking multidisciplinary design analysis and optimization architectures. *AIAA Journal*, 51(10) :2380–2394, Oct 2013.
- [95] Justin S. Gray, John T. Hwang, Joaquim R. R. A. Martins, Kenneth T. Moore, and Bret A. Naylor. OpenMDAO : An Open-Source Framework for Multidisciplinary Design, Analysis, and Optimization. *Structural and Multidisciplinary Optimization*, 2019. In Press.
- [96] Marina Gruet. *Intelligence Artificielle et prévision de l'impact de l'activité solaire sur l'environnement magnétique terrestre*. PhD thesis, ISAE-SUPAERO, Avril 2018.
- [97] Marina Gruet, Nathalie Bartoli, Sandrine Rochel, Rémi Benacquista, and Guy Rolland. Prediction of the geomagnetic index am based on the development and performance comparisons of static and dynamic neural networks. *submitted to Space Weather Space Climate*, 2018. 1st revision juillet 2018, 2nd revision dec 2018.
- [98] Marina Gruet, Nathalie Bartoli, Sandrine Rochel, Aurelie Marchaudon, Pierre-Louis Blelly, and Guy Rolland. Neural networks and am predictions, a preliminary study. In *13th European Space Weather Week*, Oostende, Belgium, 2016.
- [99] Marina Gruet, Sandrine Rochel, Nathalie Bartoli, Rémi Benacquista, Angélica Sicard, Guy Rolland, and Thomas Pelligrini. Prediction of the magnetic index am based on development and the performance comparisons of performances of static and dynamic neural networks. In *Space Weather, a multidisciplinary approach*, Leiden, The Netherlands, 2017.
- [100] Marina Gruet, Sandrine Rochel, Nathalie Bartoli, Angélica Sicard, Thomas Pelligrini, and Guy Rolland. Prediction of the magnetic index am based

- on development and comparisons of performances of static and dynamic neural networks. In *IAGA (International Association of Geomagnetism and Aeronomy)*, Cape Town, South Africa, 2017.
- [101] Marina A Gruet, M Chandorkar, Angélica Sicard, and Enrico Camporeale. Multiple-hour-ahead forecast of the dst index using a combination of long short-term memory neural network and gaussian process. *Space Weather*, 16(11) :1882–1896, 2018.
- [102] Raphael T. Haftka, Diane Villanueva, and Anirban Chaudhuri. Parallel surrogate-assisted global optimization with expensive functions – a survey. *Structural and Multidisciplinary Optimization*, pages 1–11, 2016.
- [103] Nikolaus Hansen, Andreas Ostermeier, and Andreas Gawelczyk. On the adaptation of arbitrary normal mutation distributions in evolution strategies : The generating set adaptation. In *ICGA*, pages 57–64. Citeseer, 1995.
- [104] Trevor Hastie, Robert Tibshirani, Jerome Friedman, and James Franklin. The elements of statistical learning : data mining, inference and prediction. *The Mathematical Intelligencer*, 27(2) :83–85, 2005.
- [105] Ali Hebbal, Loïc Brevault, Mathieu Balesdent, El-Ghazali Talbi, and Nouredine Melab. Multi-objective optimization using deep gaussian processes : Application to aerospace vehicle design. In *AIAA Scitech 2019 Forum*, page 1973, 2019.
- [106] J. T. Hwang. *A modular approach to large-scale design optimization of aerospace systems*. PhD thesis, University of Michigan, 2015.
- [107] John Hwang, Satadru Roy, Jason Kao, Joaquim Martins, and William A Crossley. Simultaneous aircraft allocation and mission optimization using a modular adjoint approach. In *56th AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, page 0900, 2015.
- [108] John T Hwang and Joaquim Martins. Parallel allocation-mission optimization of a 128-route network. In *16th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, page 2321, 2015.
- [109] John T Hwang and Joaquim RRA Martins. A fast-prediction surrogate model for large datasets. *Aerospace Science and Technology*, 75 :74–87, 2018.
- [110] John P Jasa, John T Hwang, and Joaquim RRA Martins. Open-source coupled aerostructural optimization using python. *Structural and Multidisciplinary Optimization*, pages 1–13, 2018.
- [111] Ruichen Jin, Wei Chen, and Agus Sudjianto. An efficient algorithm for constructing optimal design of computer experiments. *Journal of Statistical Planning and Inference*, 134(1) :268–287, 2005.

- [112] Donald R Jones, Matthias Schonlau, and William J Welch. Efficient global optimization of expensive black-box functions. *Journal of Global optimization*, 13(4) :455–492, 1998.
- [113] DR Jones. Large-scale multi-disciplinary mass optimization in the auto industry. In *MOPTA 2008 Conference (20 August 2008)*, 2008.
- [114] Michael I Jordan and Robert A Jacobs. Hierarchical mixtures of experts and the em algorithm. *Neural computation*, 6(2) :181–214, 1994.
- [115] K. Karhunen. *Über lineare Methoden in der Wahrscheinlichkeitsrechnung*. Annales Academiae scientiarum Fennicae : Mathematica - Physica. Universitat Helsinki, 1947.
- [116] Graeme Kennedy, Joaquim Martins, and Jorn Hansen. Aerostructural optimization of aircraft structures using asymmetric subspace optimization. In *12th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, page 5847, 2008.
- [117] Marc C Kennedy and Anthony O’Hagan. Predicting the output from a complex computer code when fast approximations are available. *Biometrika*, 87(1) :1–13, 2000.
- [118] Patricia Klotz, Nathalie Bartoli, Laurence Cornez, Renaud Lecourt, and Nicolas Savary. Database enrichment for multi-objective optimization. AAAF, 42ème Colloque d’Aérodynamique Appliquée : Couplages et optimisation multidisciplinaires, Nice, March 19-21, 2007.
- [119] Patricia Klotz, Nathalie Bartoli, Laurence Cornez, Manuel Samuelides, and Philippe Villedieu. Database enrichment for optimization with artificial neural networks. Second International Workshop on Surrogate Modelling and Space Mapping for Engineering Optimization (SMSMEO-06), Technical University of Denmark, November 9-11, 2006.
- [120] Patricia Klotz, Nathalie Bartoli, and Renaud Lecourt. Lpp optimization using response surface method. 2nd Colloque INCA, 2008.
- [121] Katerina Konakli and Bruno Sudret. Polynomial meta-models with canonical low-rank approximations : Numerical insights and comparison to sparse polynomial chaos expansions. *Journal of Computational Physics*, 321 :1144 – 1169, 2016.
- [122] Katerina Konakli and Bruno Sudret. Reliability analysis of high-dimensional models using low-rank tensor approximations. *Probabilistic Engineering Mechanics*, 46 :18 – 36, 2016.

- [123] Rémi Lafage and Sébastien Defoort. Whatsopt : a web application for multidisciplinary design analysis and optimization. In *20th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, Dallas, United States, 2019.
- [124] Rémi Lam, Douglas L Allaire, and Karen E Willcox. Multifidelity optimization using statistical surrogate modeling for non-hierarchical information sources. In *56th AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, page 0143, 2015.
- [125] Loic Le Gratiet and Josselin Garnier. Recursive co-kriging model for design of computer experiments with multiple levels of fidelity. *International Journal for Uncertainty Quantification*, pages 365–386, 2014.
- [126] T. Lefebvre, N. Bartoli, R. Lafage, and P. D. Ciampa. AGILE DC-1 MDO process using an efficient global optimization approach. In *6th EASN International Conference on Innovation in European Aeronautics Research*, Porto, Portugal, October 2016.
- [127] Thierry Lefebvre, Nathalie Bartoli, Sylvain Dubreuil, Marco Panzeri, Riccardo Lombardi, Roberto D’Ippolito, Pierluigi Della Vecchia, Fabrizio Nicolosi, and Pier Davide Ciampa. Methodological enhancements in mdo process investigated in the agile european project. In *18th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, page 4140, 2017.
- [128] Thierry Lefebvre, Nathalie Bartoli, Sylvain Dubreuil, Marco Panzeri, Riccardo Lombardi, Wim Lammen, Bert de Wit, Mengmeng Zhang, Imco van Gent, and Pier Davide Ciampa. Overview of mdo enhancement in agile project : a clustered and surrogate-based MDA use case. In *6th CEAS Aerospace Europe Conference*, paper no.956, Bucharest, Romania, October 2017.
- [129] Rhea P. Liem, Charles A. Mader, and Joaquim R. R. A. Martins. Surrogate models and mixtures of experts in aerodynamic performance prediction for mission analysis. *Aerospace Science and Technology*, 43 :126–151, 2015.
- [130] Rhea P. Liem and Joaquim R. R. A. Martins. Surrogate models and mixtures of experts in aerodynamic performance prediction for mission analysis. In *15th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, number AIAA-2014-2301., Atlanta, GA, June 2014 2014.
- [131] Haitao Liu, Yew Soon Ong, and Jianfei Cai. A survey of adaptive sampling for global metamodeling in support of simulation-based complex engineering design. *Structural and Multidisciplinary Optimization*, 57(1) :393–416, 2018.
- [132] Jason L Loeppky, Jerome Sacks, and William J Welch. Choosing the sample size of a computer experiment : A practical guide. *Technometrics*, 2012.

- [133] Paul Lévy. *Processus stochastiques et mouvement brownien / suivi d'une note de M. Loève "Fonctions aléatoires du second ordre"*. Gauthier-Villars, Paris, 1948.
- [134] Zhoujie Lyu and Joaquim RRA Martins. Aerodynamic design optimization studies of a blended-wing-body aircraft. *Journal of Aircraft*, 51(5) :1604–1617, 2014.
- [135] Charles A Mader and Joaquim RRA Martins. Stability-constrained aerodynamic shape optimization of flying wings. *Journal of Aircraft*, 50(5) :1431–1449, 2013.
- [136] Chu V Mai, Minas D Spiridonakos, Eleni N Chatzi, and Bruno Sudret. Surrogate modeling for stochastic dynamical systems by combining nonlinear autoregressive with exogenous input models and polynomial chaos expansions. *International Journal for Uncertainty Quantification*, 6(4), 2016.
- [137] Joaquim RRA Martins, Peter Sturdza, and Juan J Alonso. The complex-step derivative approximation. *ACM Transactions on Mathematical Software (TOMS)*, 29(3) :245–262, 2003.
- [138] JRRA Martins and Andrew B Lambe. Multidisciplinary design optimization : A survey of architectures. *AIAA Journal*, 59(9) :2049–2075, 2013.
- [139] Joan Mas Colomer, Nathalie Bartoli, Thierry Lefebvre, Sylvain Dubreuil, Joaquim Martins, Emmanuel Bénard, and Joseph Morlier. Similarity maximization of a scaled aeroelastic flight demonstrator via multidisciplinary optimization. In *58th AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, page 0573, Grapewine, United States, 2017.
- [140] Joan Mas Colomer, Nathalie Bartoli, Thierry Lefebvre, Joaquim Martins, and Joseph Morlier. Optimization approach for aeroelastic scaling of wings under non-similar flow. *submitted to Journal of Aircraft*, 2018. submitted in Sept 2018, rejected the 19/11/2018 with some comments.
- [141] Caleb W McCormick. *The nastran user's manual (level 15)*. 1972.
- [142] M D McKay. Nonparametric variance-based methods of assessing uncertainty importance. *Reliability Engineering & System Safety*, 57(7) :267–279, 1997.
- [143] Michael D McKay, Richard J Beckman, and William J Conover. Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2) :239–245, 1979.
- [144] M. Meheut, O. Atinault, P. Bardoux, A. Dumont, C. Francois, C. Blondeau I. Salah El Din, M. Marcelet, H. Rabia, P. Schmollgruber, T. Lefebvre, and

- N. Bartoli. Necst-2(new environmental compliant solutions for transportation - etape 2), rapport d'avancement des travaux onera à fin mai 2011 (tranche conditionnelle 1). Technical Report RA 3/16645 DAAP, ONERA - AIRBUS, Juin 2011.
- [145] M. Meheut, O. Atinault, P. Bardoux, A. Dumont, C. Francois, C. Blondeau I. Salah El Din, P. Schmollgruber, T. Lefebvre, and N. Bartoli. Necst-2(new environmental compliant solutions for transportation - etape 2), rapport d'avancement des travaux onera à fin décembre 2011 (tranche conditionnelle 1). Technical Report RA 6/16645 DAAP, ONERA - AIRBUS, Février 2012.
- [146] Mostafa Meliani, Nathalie Bartoli, Thierry Lefebvre, Mohamed-Amine Bouhleb, Joaquim R.R.A. Martins, and Joseph Morlier. Multi-fidelity efficient global optimization : Methodology and application to airfoil shape design. In *20th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, Dallas, United States, 2019.
- [147] Morgane Menz, Christian Gogu, Sylvain Dubreuil, Nathalie Bartoli, and Jérôme Morio. Vers des analyses de fiabilité à fidélité adaptative par l'utilisation de modèles en base réduite. In *10e Journées de Fiabilité des Matériaux et des Structures JFMS*, Bordeaux, France, Mars 2018.
- [148] M.Gruet, N.Bartoli, S.Rochel, T. Pellegrini, A. Sicard, R. Benacquista, and G.Rolland. From artificial intelligence to space weather. In *Séminaire à l'Institut royal d'Aéronomie Spatiale de Belgique (IASB)*, Bruxelles, Belgium, September 2017.
- [149] J. Močkus. *On Bayesian Methods for Seeking the Extremum*, pages 400–404. Springer Berlin Heidelberg, Berlin, Heidelberg, 1975.
- [150] Hossein Mohammadi, Rodolphe Le Riche, and Eric Touboul. Making ego and cma-es complementary for global optimization. In *Learning and Intelligent Optimization*, pages 287–292. Springer, 2015.
- [151] Menz Morgane, Christian Gogu, Nathalie Dubreuil, Sylvain Bartoli, and Jérôme Morio. Adaptive coupling of reduced basis modeling and kriging based active learning methods for reliability analyses. *submitted to Reliability Engineering and System Safety*, 2019. submitted 31/01/2019.
- [152] Max D. Morris and Toby J. Mitchell. Exploratory designs for computational experiments. *Journal of Statistical Planning and Inference*, 43(3) :381 – 402, 1995.

- [153] M Moustapha and B Sudret. Surrogate-assisted reliability-based design optimization : a survey and a new general framework. *arXiv preprint arXiv :1901.03311*, 2019.
- [154] Jorge Nocedal and Stephen Wright. *Numerical optimization*. Springer Science & Business Media, 2006.
- [155] R. Olivanti, N. Bartoli, T. Lefebvre, and S. Dubreuil. Development of SEGO-MOE and application to wing aerodynamic optimization. Technical Report RA 5/22737, ONERA/DTIS, November 2017.
- [156] Abdelkader Otsmane, Mohamed Amine Bouhleh, Nathalie Bartoli, and Joseph Morlier. Procédé de conception de pièces mécaniques, notamment d’aubes de turbomachines, 2017. Brevet FR.15.57301 - cas 2625.
- [157] Jeong-Soo Park and Jangsun Baek. Efficient computation of maximum likelihood estimators in a spatial linear model with power exponential covariogram. *Computers & Geosciences*, 27(1) :1–7, 2001.
- [158] JM Parr, AJ Keane, Alexander IJ Forrester, and CME Holden. Infill sampling criteria for surrogate-based optimization with constraint handling. *Engineering Optimization*, 44(10) :1147–1166, 2012.
- [159] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn : Machine learning in Python. *Journal of Machine Learning Research*, 12 :2825–2830, 2011.
- [160] Benjamin Peherstorfer, Karen Willcox, and Max Gunzburger. Survey of multifidelity methods in uncertainty propagation, inference, and optimization. *SIAM Review*, 60(3) :550–591, 2018.
- [161] Julien Pelamatti, Loïc Brevault, Mathieu Balesdent, El-Ghazali Talbi, and Yannick Guerin. Efficient global optimization of constrained mixed variable problems. *Journal of Global Optimization*, 73(3) :583–613, 2019.
- [162] Ruben E Perez, Hugh H T. Liu, and Kamran Behdinan. Multidisciplinary optimization framework for control-configuration integration in aircraft conceptual design. *Journal of Aircraft*, 43(6) :1937–1948, 2006.
- [163] Victor Picheny. A stepwise uncertainty reduction approach to constrained global optimization. In *AISTATS*, pages 787–795, 2014.
- [164] Victor Picheny. Multiobjective optimization using gaussian process emulators via stepwise uncertainty reduction. *Statistics and Computing*, 25(6) :1265–1280, 2015.

- [165] Victor Picheny, David Ginsbourger, Olivier Roustant, Raphael T. Haftka, and Nam-Ho Kim. Adaptive Designs of Experiments for Accurate Approximation of a Target Region. *Journal of Mechanical Design*, 132(7) :071008, jul 2010.
- [166] Victor Picheny, Robert B Gramacy, Stefan Wild, and Sebastien Le Digabel. Bayesian optimization under mixed constraints with a slack-variable augmented lagrangian. In *Advances in Neural Information Processing Systems*, pages 1435–1443, 2016.
- [167] Matthias Poloczek, Jialei Wang, and Peter Frazier. Multi-information source optimization. In *Advances in Neural Information Processing Systems*, pages 4291–4301, 2017.
- [168] Michael JD Powell. A direct search optimization method that models the objective and constraint functions by linear interpolation. In *Advances in optimization and numerical analysis*, pages 51–67. Springer, 1994.
- [169] R. Priem, N. Bartoli, T. Lefebvre, and J. Morlier. Surrogate modeling using mixture of experts - adaptive optimization process : Description of the MOE and SEGOMOE python toolboxes. Technical Report RA 3/22737, ONERA/DCPS, February 2016.
- [170] Rémy Priem, Nathalie Bartoli, and Youssef Diouane. Use of upper trust bounds in constrained bayesian optimization infill criteria. In *20th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, Dallas, United States, 2019.
- [171] Rémy Priem, Nathalie Bartoli, Youssef Diouane, and Sylvain Dubreuil. An adaptive feasibility approach for constrained bayesian optimization with application in aircraft design. Lisboa, Portugal, September 2018. Engopt : 6th International Conference on Engineering Optimization.
- [172] Rommel G Regis. Constrained optimization by radial basis function interpolation for high-dimensional expensive black-box problems with infeasible initial points. *Engineering Optimization*, 46(2) :218–243, 2014.
- [173] Andrea Da Ronch, Marco Panzeri, Noesis Solutions N V, M Anas Abd Bari, Roberto Ippolito, Matteo Franciolini, Ingegneria Industriale, and Università Politecnica. Adaptive design of experiments for efficient and accurate estimation of aerodynamic loads. 4(December 2016) :558–569, 2017.
- [174] Satadru Roy, Kenneth Moore, John T Hwang, Justin S Gray, William A Crossley, and Joaquim Martins. A mixed integer efficient global optimization algorithm for the simultaneous aircraft allocation-mission-design pro-

- blem. In *58th AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, page 1305, 2017.
- [175] A Saltelli, K Chan, and E M Scott. *Sensitivity analysis*. John Wiley & Sons, 1 edition, 2000.
- [176] M. Samuelides, D. Bettebghor, S. Grihon, N. Bartoli, J. Morlier, and A. Merval. Using mixture of experts surrogate models and EM learning for multi-level structural optimization. Palais des Congrès, Paris, May 16-21 2010. IV European Conference on Computational Mechanics.
- [177] Michael Sasena, Panos Papalambros, and Pierre Goovaerts. Global optimization of problems with disconnected feasible regions via surrogate modeling. In *9th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, Atlanta, GA, Paper No. AIAA-2002-5573*, 2002.
- [178] M.J. Sasena. *Flexibility and efficiency enhancements for constrained global design optimization with Kriging approximations*. PhD thesis, University of Michigan, 2002.
- [179] P Schmollgruber et al. Conceptual design of the aircraft and aircraft systems using the gabriel concept. *Deliverable D3*, 6, 2013.
- [180] Peter Schmollgruber. *Enhancement of the conceptual aircraft design process through certification constraints management and full mission simulations*. PhD thesis, ISAE-SUPAERO, December 2018.
- [181] Peter Schmollgruber, Nathalie Bartoli, Judicael Bedouet, Emmanuel Benard, and Yves Gourinat. Improvement of the aircraft design process for air traffic management evaluations. In *2018 AIAA Aerospace Sciences Meeting*, page 0283, Kissimmee, United States, 2018.
- [182] Peter Schmollgruber, Nathalie Bartoli, Judicaël Bedouet, Sebastien Defoort, Yves Gourinat, Emmanuel Benard, Rémi Lafage, and Alessandro Sgueglia. Use of a certification constraints module for aircraft design activities. In *17th AIAA Aviation Technology, Integration, and Operations Conference*, page 3762, 2017.
- [183] Peter Schmollgruber, Nathalie Bartoli, and Yves Gourinat. Impact of certification constraints in a multi-disciplinary design process. Workshop EASN on Aircraft Design (Aachen), October 2014.
- [184] Peter Schmollgruber, Judicaël Bedouet, Nathalie Bartoli, and Yves Gourinat. Development of a certification module tailored to aircraft multi disciplinary optimization. Delft, September 2015. CEAS Air & Space.
- [185] Alessandro Sgueglia, Sylvain Dubreuil, and Nathalie Bartoli. Technologies sensitivity analysis of an hybrid aircraft. In L. Brevault, M. Balesdent, and

- J. Morio, editors, *Aerospace System Analysis and Optimization in Uncertainty*, chapter 10. Springer, Optimization and Its Applications, New York, 2019. In Press.
- [186] Alessandro Sgueglia, Peter Schmollgruber, Nathalie Bartoli, Olivier Atinault, Emmanuel Benard, and Joseph Morlier. Exploration and sizing of a large passenger aircraft with distributed ducted electric fans. In *2018 AIAA Aerospace Sciences Meeting*, page 1745, Kissimmee, United States, 2018.
- [187] Alessandro Sgueglia, Peter Schmollgruber, Nathalie Bartoli, Emmanuel Benard, Joseph Morlier, John Jasa, Joaquim R.R.A. Martins, John T. Hwang, and Justin S. Gray. Development of a multidisciplinary design optimization framework with gradient calculation applied to hybrid aircraft. *submitted to Journal of Aircraft*, 2019. submitted in 6 march 2019.
- [188] Alessandro Sgueglia, Peter Schmollgruber, Emmanuel Benard, Nathalie Bartoli, , and Joseph Morlier. Preliminary sizing of a medium range blended wing-body using a multidisciplinary design analysis approach. In *8th EASN-CEAS International Workshop on Manufacturing for Growth & Innovation*, volume 233. MATEC Web Conf., 2018.
- [189] Jaroslaw Sobieszczanski-Sobieski and Srinivas Kodiyalam. Bliss/s : a new method for two-level structural optimization. *Structural and Multidisciplinary Optimization*, 21(1) :1–13, 2001.
- [190] I M Sobol. Sensitivity estimated for non linear mathematical models. *Math. Mod. Comp. Exp.*, 1(4) :407–414, 1993.
- [191] V. Stilz, N. Bartoli, T. Lefebvre, and J. Morlier. Optimization process with SEGOMOE : application to different test cases for aircraft design. Technical Report RA 4/22737, ONERA/DCPS - ISAE, August 2016.
- [192] B Sudret. Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering & System Safety*, 93 (7) :964–979, 2008.
- [193] David JJ Toal, NW Bressloff, AJ Keane, and CME Holden. The development of a hybridized particle swarm for kriging hyperparameter tuning. *Engineering optimization*, 43(6) :675–699, 2011.
- [194] University of Michigan, ONERA, ISAE-SUPAERO, and NASA. Surrogate model toolbox. <https://github.com/SMTorg/SMT/>.
- [195] J. Vankan, R. Maas, B. de Wit, S. Hernandez, D. Bettebghor, N. Bartoli, and S. Grihon M. Samuelides. Multi-level optimisation framework for composite fuselage barrel design. Technical Report MAAXIMUS-WP2-4-NLR-DEL-D2.4.2-MLOforCompFusDes, MAAXIMUS, mars 2011.

- [196] John C Vassberg, Mark A DeHaan, S Melissa Rivers, and Richard A Wahls. Development of a common research model for applied cfd validation studies. *AIAA paper*, 6919 :2008, 2008.
- [197] Rémi Vauclin. Multi-fidelity surrogate models for structural optimization. Technical report, ISAE-SUPAERO/ONERA, July 2014.
- [198] Felipe A. C. Viana, Raphael T. Haftka, and Layne T. Watson. Efficient global optimization algorithm assisted by multiple surrogate techniques. *Journal of Global Optimization*, 56(2) :669–689, 2013.
- [199] Felipe AC Viana, Raphael T Haftka, and Layne T Watson. Why not run the efficient global optimization algorithm with multiple surrogates. In *Proceedings of the 51th AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics, and materials conference, AIAA, Orlando, FL, USA. AIAA-2010-3090*, 2010.
- [200] Ziyu Wang, Frank Hutter, Masrour Zoghi, David Matheson, and Nando de Freitas. Bayesian optimization in a billion dimensions via random embeddings. *Journal of Artificial Intelligence Research*, 55 :361–387, 2016.
- [201] Graham Warwick. Zunum picks safran turbine for hybrid-electric airliner : Safran ardiden 32 turboshaft selected as range extender ; rockwell turbo commander acquired as flying testbed. *Aviation Week & Space Technology*, 2018.
- [202] Alan G Watson and Randal J Barnes. Infill sampling criteria to locate extremes. *Mathematical Geology*, 27(5) :589–608, 1995.
- [203] Garrett Waycaster. *Stratégies d’optimisation par modélisation explicite des différents acteurs influençant le processus de conception*. PhD thesis, Université de Toulouse, Université Toulouse III-Paul Sabatier, 2015.
- [204] Yiming Zhang, Nam H Kim, Chanyoung Park, and Raphael T Haftka. Multifidelity surrogate based on single linear regression. *AIAA Journal*, pages 1–9, 2018.

