



**HAL**  
open science

## Descent methods for design optimization under uncertainty

Fabrice Poirion, Quentin Mercier

► **To cite this version:**

Fabrice Poirion, Quentin Mercier. Descent methods for design optimization under uncertainty. Aerospace Lab, 2018, 14, pp.1-11. 10.12762/2018.AL14-05 . hal-01938525

**HAL Id: hal-01938525**

**<https://hal.science/hal-01938525>**

Submitted on 28 Nov 2018

**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.

F. Poirion, Q. Mercier  
(ONERA)

E-mail: [fabrice.poirion@onera.fr](mailto:fabrice.poirion@onera.fr)

DOI: 10.12762/2018.AL14-05

## Descent Methods for Design Optimization under Uncertainty

This paper is about optimization under uncertainty, when the uncertain parameters are modeled through random variables. Contrary to traditional robust approaches, which deal with a deterministic problem through a worst-case scenario formulation, the stochastic algorithms presented introduce the distribution of the random variables modeling the uncertainty. For single-objective problems such methods are currently classical, based on the Robbins-Monro algorithm. When several objectives are involved, the optimization problem becomes much more difficult and the few available methods in the literature are based on a genetic approach coupled with Monte-Carlo approaches, which are numerically very expensive. We present a new algorithm for solving the expectation formulation of stochastic smooth or non-smooth multi-objective optimization problems. The proposed method is an extension of the classical stochastic gradient algorithm to multi-objective optimization, using the properties of a common descent vector. The mean square and the almost-certain convergence of the algorithm are proven. The algorithm efficiency is illustrated and assessed on an academic example.

### Introduction

Manufacturers are ever looking for designing products with better performance, and higher reliability at lower cost and risk. One way to address these antagonistic objectives is to use multi-objective optimization approaches. However, real-world problems are rarely described through a collection of fixed parameters and uncertainty has to be taken into account, whether it appears in the system description itself, or in the environment and operational conditions. Indeed, the system behavior can be very sensitive to modifications in some parameters [1, 2, 3]. This is why uncertainty has to be introduced in the design process from the start. Optimization under uncertainty has undergone important advances since the second half of the 20<sup>th</sup> century [4, 5] and various approaches have been proposed, including robust optimization, where only the bounds of the uncertain parameters are used, and stochastic optimization where uncertain parameters are modeled through random variables with a given distribution and where the probabilistic information is directly introduced into the numerical approaches. In that context, the uncertain multi-objective problems are written in terms of the expectation of each objective. Considering single objective stochastic optimization problems, a large variety of numerical approaches [6, 7] can be found in the literature, with the first results appearing in the late 50's [4, 8, 5]. With regard to aerospace applications, optimization problems under uncertainty are

either considered as robust optimization problems or as reliability ones. In both cases, the numerical procedures that are the most frequently used are purely deterministic ones: this is indeed the case for robust optimization, since it is written as a "worst case" deterministic optimization problem, but it is also true when reliability is addressed. In this last situation, the chance constraint is transformed into a deterministic constraint using FORM or SORM approximations [9]. In both cases classical deterministic algorithms, such as the SQP algorithm [10], are eventually used to numerically solve the optimization problem. There is another route: it uses the probabilistic distribution of the random variables modeling the uncertainty [11, 12, 13]. There are two main approaches: the stochastic gradient algorithm, based on stochastic approximations such as the Robbins Monro algorithm [14, 15, 16], which is a descent method, and a second one based on scenario approaches [17, 18], the latter being more frequently applied for chance-constrained problems.

After briefly presenting the now-classical stochastic gradient algorithm, we illustrate its potential for being used in structural optimization on a reliability optimization problem in aeroelasticity. We pursue this by presenting the problem of optimizing several objective functions when uncertainty, modeled through random variables,

is introduced in part of the objective function. After providing some necessary mathematical elements for comprehension of the method, we present a general algorithm based on the existence of a descent vector common to each objective, which can be used in a broad context: for regular or non-regular, convex or non-convex objectives, with or without constraints. An illustration on the optimal design of a sandwich will highlight the efficiency of the proposed approach compared to that of classical genetic algorithms.

## Single-Objective Stochastic Optimization

The following deterministic optimization problem

$$\text{Argmin}\{f(x) \mid g(x) \geq 0, x \in X\}; f, g: \mathbb{R}^n \rightarrow \mathbb{R}; X \subset \mathbb{R}^n \quad (1)$$

is a classical problem for any regular objective function  $f$  and constraint function  $g$ . However, when random parameters  $\xi(\omega) = (\xi_1(\omega), \dots, \xi_d(\omega)) \in \mathbb{R}^d$  defined on a probability space  $(\Omega, \mathcal{T}, \mathbb{P})$  are introduced into either one or both functions  $f$  and  $g$ , the meaning given to the random problem must be specified:

$$\text{Argmin}\{f(x, \xi(\omega)) \mid g(x, \xi(\omega)) \geq 0, x \in X\} \quad (2)$$

Depending on the nature of the practical applications considered, there are several approaches to deal with stochastic optimization problems. For instance, without being limited to these options, one can consider working with either:

- a mean value description:

$$\text{Argmin}\{\mathbb{E}[f(x, \xi(\omega))] \mid \mathbb{E}[g(x, \xi(\omega))] \geq 0\}$$

- a worst case scenario:

$$\text{Argmin}\{\mathbb{E}[f(x, \xi(\omega))] \mid g(x, \xi(\omega)) \geq 0, \forall \omega \in \Omega\}$$

- a robust context:

$$\text{Argmin}\{\mathbb{E}[f(x, \xi(\omega))] \mid g(x, \xi(\omega)) \geq 0, \forall \omega \in F \subset \Omega, \mathbb{P}(F) = p_0, p_0 \in [0, 1]\}$$

- or a chance-constraint formulation:

$$\text{Argmin}\{\mathbb{E}[f(x, \xi(\omega))] \mid \mathbb{P}[g(x, \xi(\omega)) \geq 0] \geq p_0\}$$

denoting the mathematical expectation as  $\mathbb{E}$ .

### The Stochastic Gradient Approach

Let  $(\Omega, \mathcal{A}, \mathbb{P})$  be an abstract probabilistic space, and  $\xi: \Omega \rightarrow \mathbb{R}^d$  a random vector. We denote as  $\mu$  the distribution of the random variable  $\xi$ , and as  $\mathcal{W}$  its image space  $\xi(\Omega)$ . Let  $\xi_1, \dots, \xi_k, \dots$  be independent copies of the random variable  $\xi$ , which will be used to generate independent random samples with the distribution  $\mu$ . We consider the case where the constraints and the optimization parameters are deterministic. In the stochastic optimization problem (the objective function is defined as the mathematical expectation of the random quantity  $f(x, \xi(\omega))$ ):

$$x^* = \text{Argmin}_{x \in X^{\text{ad}}} J(x); J(x) = \mathbb{E}[f(x, \xi(\omega))] \quad (3)$$

$X^{\text{ad}}$  denotes the admissible space. There exists a stochastic extension of the standard deterministic gradient method that is particularly suited to this problem: it does not necessitate the estimation of the expectation in relation (3) to be built at each optimization step.

The algorithm of the stochastic gradient method uses optimization iteration instead, in order to build an estimate of the gradient expectation:

- Choose  $x_0 \in X$  and  $\gamma_k > 0$  for  $k \in \mathbb{N}$ .
- Draw  $\xi_{n+1}$  under the law of  $\xi$  independently from  $\xi_k$  for  $k \leq n$ .
- Update

$$X_{n+1} = x_n - \gamma_n (f'_x(x_n, \xi_{n+1})) \quad (4)$$

- Project over the feasible space  $X^{\text{ad}}$

$$x_{n+1} = \Pi_{X^{\text{ad}}}(X_{n+1}) \quad (5)$$

$\Pi_{X^{\text{ad}}}$  defines the projection operator on the feasible space  $X^{\text{ad}}$ . The series  $(\gamma_n)$  must be divergent, and the series  $(\gamma_n^2)$  convergent. Typically,  $\gamma_n = a/(n^\alpha + b)$ ,  $\alpha \in ]0.5, 1]$ . Like its deterministic version, the sequence  $(x_n)$  converges to the solution  $x^*$  of the problem under the Robbins-Monroe assumptions applied to  $f'_x$  [14]. When the gradient is easily available the method is very efficient (see the discussion section and Figure 10). Some results on convergence speed and enhancements can be found in the book [19].

## An Aeroelasticity Illustration

### Flutter Equation

We consider the classical context of aeroelasticity, where the aerodynamic forces are calculated using a linearized assumption together with a doublet-lattice method [20], where the airplane structure is described through a finite-element model, and when uncertainty is introduced in the mass and stiffness matrices through a vector valued random variable  $\xi$ . The finite-element discretization for the aeroelastic analysis can be formulated in the frequency domain as follows:

$$L(\xi)^T \left[ p^2 \Phi^T(\xi) M(\xi) \Phi(\xi) + \Phi^T(\xi) K(\xi) \Phi(\xi) + \frac{1}{2} \rho V^2 \Phi^T(\xi) A(p/V) \Phi(\xi) \right] R(\xi) = 0 \quad (6)$$

where  $M$  and  $K$  are the structural mass and stiffness matrices,  $\rho$  is the air density,  $V$  is the flow speed,  $A$  is the aerodynamic load matrix and  $\Phi$  is the modal basis of the structure  $(M, K)$ . Assuming the air-flow speed  $V$  to be constant, the solution  $p \in \mathbb{C}$  of the flutter equation depends on the aerodynamic parameter  $\rho$  and on the uncertain parameters  $\xi$ . The sign of the real part  $\Re(p)$  specifies the stability of the coupled system. We define the critical pressure  $q_c$  as the smallest pressure value  $q$  such that  $\Re(p(q)) = 0$ , if any. The critical pressure depends on the uncertain parameters  $\xi$  and, therefore, is itself a random variable. The vectors  $L$  and  $R$  are the associated pseudo left and right eigenvectors. The dimension of Problem (6) is equal to the number of eigenmodes retained for the aeroelastic analysis.

### Gradient Calculation

We shall address the problem of optimizing the mass distribution of a given number of concentrated masses  $m_i, i = 1, q$  of the finite element model, in order to maximize the critical pressure value. We shall denote by  $m$  the vector  $m = (m_1, \dots, m_q)$ . Therefore, we shall need to

evaluate the gradients  $\frac{\partial p}{\partial m_i}$  and  $\frac{\partial p}{\partial q}$ . The derivation of such quantities is classical [21, 22]; they are obtained by differentiating Equation (6):

$$\frac{\partial p}{\partial m_i} = - \frac{p^2 L^T \Phi^T \frac{\partial M}{\partial m_i} \Phi R}{L^T \Phi^T \left( 2pM + \frac{1}{2} q V^2 A'(p/V) \right) \Phi R'} \quad (7)$$

$$\frac{\partial p}{\partial q} = - \frac{\frac{1}{2} L^T \Phi^T V^2 A(p/V) \Phi R}{L^T \Phi^T \left( 2pM + \frac{1}{2} q V^2 A'(p/V) \right) \Phi R'} \quad (8)$$

where  $A'$  stands for the derivative of  $A$ . These two relations will allow the derivation of the critical pressure gradient expression. For each mass distribution  $m$ , the critical pressure is defined by  $\Re(p(m, q_c)) = 0$ . Using the implicit function theorem in the neighborhood of a point  $(m^0, q_c^0)$ , under the assumption that  $p$  is a regular function, there exists a function  $\phi$  such that  $\phi(m) = q_c$  (with  $q_c^0 = \phi(m^0)$ ). Moreover, in the neighborhood of  $(m^0, q_c^0)$ , we have, for each mass point  $m_i$ :

$$\frac{\partial \phi}{\partial m_i}(m^0) = \frac{\partial q_c}{\partial m_i}(m^0) = - \frac{\Re \left( \frac{\partial p}{\partial m_i}(m^0, q_c^0) \right)}{\Re \left( \frac{\partial p}{\partial q}(m^0, q_c^0) \right)} \quad (9)$$

The aerodynamic load matrix is modeled by a matrix-valued rational function using the "Minimum State" approach [23]. The analytical expression of the aerodynamic matrix gradients can then be readily derived, since their calculation involves rational function differentiation.

## Wing Model

The goal of this section is to show numerically the applicability of these two algorithms to an aeroelastic optimization problem. We shall consider a finite-element model of a simple wing and introduce uncertainty in several structural parameters. We consider then two different optimization problems: the first one involving a probabilistic objective function and deterministic constraints, which will be solved using the stochastic gradient algorithm, and the second one, which is a chance constraint optimization problem, and which will be solved using the stochastic Arrow-Hurwicz algorithm.

## Description of the Model

We consider a wing model that was defined as a wind-tunnel model similar to a heavy-carrier airplane wing, in order to evaluate and compare different CFD codes among various partners in the late '80s [24]: Aerospatiale, ONERA, DLR and MBB. The structural model is given in Figure 1. It is a stick model with concentrated masses. This model has the advantage of being numerically more tractable for testing stochastic algorithms. During the design and optimization stages, stick models are, in fact, used by manufacturers because they give a clear and synthetic overview of the structure properties. The stick model, including the super element modelling the mounting bracket, is defined with 93 beams and 97 concentrated masses  $m_i$ .

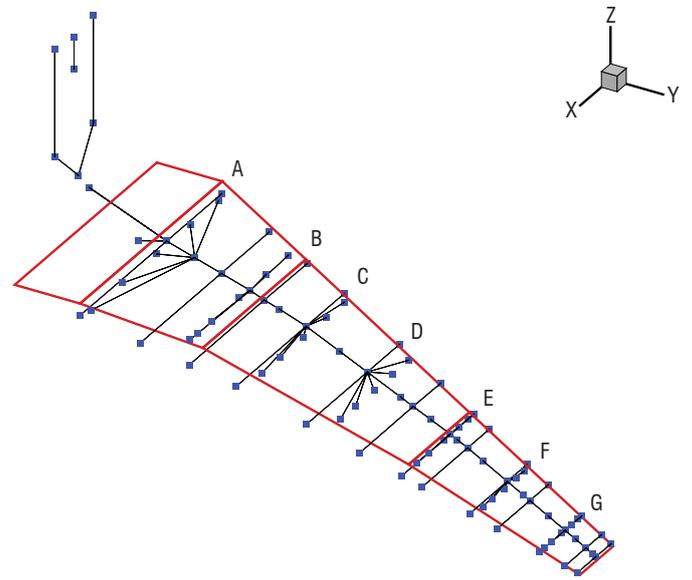


Figure 1 – AMP stick model

The root and tip chord lengths are, respectively, equal to 0.42 m and 0.10 m. The sweep angle is equal to 32 degrees, and the span length is equal to 1 m.

A flutter analysis is performed using a doublet-lattice method for computing the aerodynamic matrix. In Figure 2, the frequency and damping evolution of the first bending mode (23.4 Hz) and first torsion mode (31.85 Hz) with respect to the pressure are shown. The first torsion mode becomes unstable for  $q_c \approx 12 \times 10^4$ .

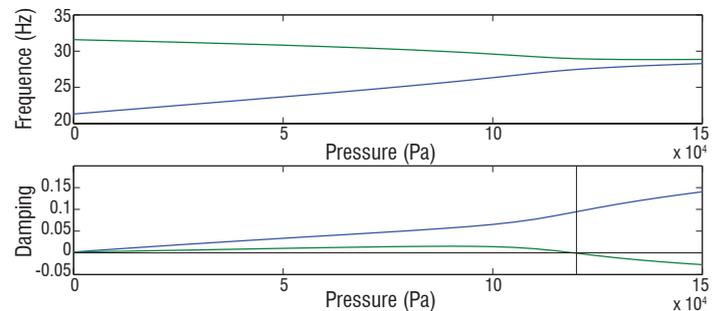


Figure 2 – Flutter diagram for the AMP wing

## Optimization of the Critical Pressure

The illustration goal is to test and assess the gradient-based method applied to our simple stick model. The optimization problem purpose is to modify the value of each of the 89 mass points lying on the wing, in order to increase the value of the critical pressure. The mass points defining the mounting bracket are not considered. Several constraints are introduced. The first one is to keep the global mass of the model constant. The other set of constraints is related to the range of variation of each mass point: their value must stay within a bounded interval in order to avoid physical aberration (negative or null mass). The optimization problem is then written as:

$$\text{Argmax} \left\{ \mathbb{E} [q_c(m, \xi(\omega))] \mid \sum_{j=1}^N m_j = c ; m_i \in [a_i, b_i], \forall i \right\} \quad (10)$$

The feasible space

$$X^{\text{ad}} = \left\{ m = (m_1, \dots, m_N) \in \mathbb{R}^N \mid \sum_{j=1}^N m_j = c; m_i \in [a_i, b_i], \forall i \right\}$$

is convex. The gradient-based algorithm iterates on the values of vector  $m$  and is written:

$$m^{n+1} = \Pi_{X^{\text{ad}}} \left( m^n + \gamma_n \frac{\partial q_c}{\partial m} (m^n, \xi_{n+1}) \right) \quad (11)$$

The divergent series  $(\gamma_n)$  is chosen as:  $\frac{K_1}{K_2 + n}$ , where  $K_1$  and  $K_2$  are

parameters that need to be tuned. Indeed, these two parameters have an impact on the convergence speed. For this particular application, only a couple of tests were necessary to obtain an acceptable convergence speed. More precisely, we have taken  $K_1 = 0.01$  and  $K_2 = 100$ . Under classical assumptions [19], the gradient-based algorithm converges to the best solution, either almost surely or as a mean square for the norm  $\mathbb{L}^2(\Omega)$ .

In the numerical experiment, seven regions of the stick model have been considered, for which a random stiffness coefficient  $\xi$  is introduced in order to model the stiffness uncertainty of each region. Those seven regions (indicated in Figure 1 by the letters A through G) contain the beams connecting the mass points lying on a same chord. The uncertainty is modeled as a uniform random variable over  $[0.75 \times \xi_{i,0}, 1.25 \times \xi_{i,0}]$ , where  $\xi_{i,0}$  are the stiffness nominal values for each region. Eighty nine grid mass points  $m_j$  are chosen as optimization parameters, and a maximum variation of 25% of the initial values:  $m_j \in [0.75 \times m_{j,0}, 1.25 \times m_{j,0}]$  is allowed.

Five hundred iterations of the stochastic gradient algorithm have been considered. In order to illustrate the quality of the optimization result, we have performed an uncertainty propagation study by drawing 1000 random stiffness realizations for the initial and final mass configurations and by constructing the critical pressure histogram.

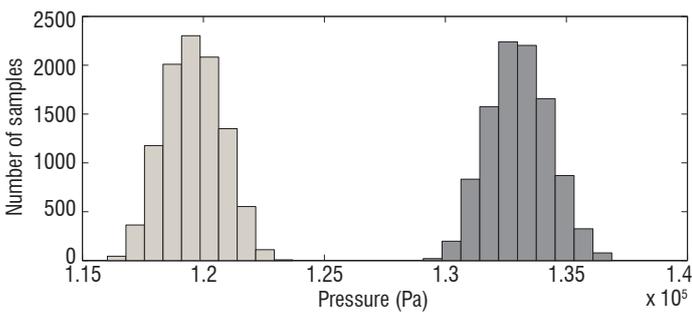


Figure 3 – Random critical pressure distribution before and after optimization

The critical pressure histogram corresponding to the initial values of the optimization parameters is represented on the left in Figure 3, and that corresponding to the final values is on the right. The gain obtained is clearly visible. This result shows that the critical pressure of the wing can be significantly increased by modifying the mass distribution, without modifying the total weight. Such a result can be interesting for updating the numerical model of a wing with uncertain parameters, in order to match an experimental critical speed value for a wind tunnel mockup.

The locus of the wing centers has hardly been modified by the optimization procedure.

## Multi-Criteria Stochastic Optimization

Let  $m$  functions  $f_i: \mathbb{R}^n \times \mathcal{W} \rightarrow \mathbb{R}$ ,  $i=1, \dots, m$  depending on uncertain parameters be modeled through a random vector  $W(\omega)$ . We consider the following stochastic optimization problem:

$$\min_{x \in \mathbb{R}^n} \left\{ \mathbb{E}[f_1(x, W(\omega))], \mathbb{E}[f_2(x, W(\omega))], \dots, \mathbb{E}[f_m(x, W(\omega))] \right\} \quad (12)$$

More precisely, we want to construct the associated Pareto set: multi-objective optimization is based on the notion of Pareto-optimal and weak Pareto-optimal solutions. Consider  $m$  convex functions  $f_i: \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $i=1, \dots, m$  and the unconstrained optimization problem

$$\min_{x \in \mathbb{R}^n} \{f_1(x), \dots, f_m(x)\} \quad (13)$$

A solution  $x^*$  of Problem (13) is Pareto-optimal if no point  $x$  such that  $f_i(x) \leq f_i(x^*) \forall i=1, \dots, m$  and  $f_j(x) < f_j(x^*)$  for an index  $j \in \{1, \dots, m\}$  exists. It is weakly Pareto-optimal if no point  $x$  such that  $f_i(x) < f_i(x^*) \forall i=1, \dots, m$  exists. A complete review of multi-objective optimization can be found in [25]. Before continuing with the algorithm description that will be used to solve the previous problem, we shall recall definitions of some notions appearing in the context of non-smooth analysis and multi-objective optimization. Throughout the paper, the standard inner product on  $\mathbb{R}^n$  will be used and denoted as  $\langle \cdot, \cdot \rangle$ , with the norm being denoted as  $\|\cdot\|$ .

### Some Definitions and Results in Convex Analysis

**Definition 1** – A function  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  is locally Lipschitz-continuous at point  $x$  if there exists scalars  $K > 0$  and  $\varepsilon > 0$  such that, for all  $y, z \in B(x, \varepsilon)$

$$|f(y) - f(z)| \leq K \|y - z\|$$

where  $B(x, \varepsilon)$  denotes the open ball of center  $x$  and radius  $\varepsilon$ .

**Definition 2** – A function  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  is convex if for all  $x, y \in \mathbb{R}^n$  and  $\lambda \in [0, 1]$  the following inequality holds:

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$$

**Definition 3** – The directional derivative at  $x$  along the direction  $v \in \mathbb{R}^n$  of a function  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  is defined by the limit:

$$f'(x; v) = \lim_{t \downarrow 0} \frac{f(x + tv) - f(x)}{t}$$

Any convex function  $f$  is continuous and differentiable almost everywhere. Moreover, there exists at each point  $x$  a lower affine function that is identical to  $f$  at  $x$ . This affine function defines the equation of a plane called a tangent plane. When the function  $f$  is differentiable at  $x$ , there is only one tangent plane characterized by the gradient  $\nabla f(x)$ . When  $f$  is non-differentiable at  $x$ , there exists an infinity of tangent planes that define the subdifferential.

**Definition 4** – The subdifferential of a function  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  at  $x$  is the set

$$\partial f(x) = \{s \in \mathbb{R}^n : f(y) \geq f(x) + \langle s, y - x \rangle \forall y \in \mathbb{R}^n\} \quad (14)$$

This set is non-empty, convex, closed and reduced to  $\nabla f(x)$  when  $f$  is differentiable. The following result allows the notion of subdifferential to be used for characterizing the optima of convex functions.

**Theorem 1 ([26])** – Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  a convex function. The following statements are equivalent

- $f$  is minimized at  $x^*$ :  $f(y) \geq f(x^*) \forall y \in \mathbb{R}^n$ ,
- $0 \in \partial f(x^*)$ ,
- $f'(x^*, d) \geq 0 \forall d \in \mathbb{R}^n$ .

When the function is no longer convex, but is locally Lipschitz-continuous, the directional derivative defined in Definition 3 does not necessarily exist and a generalized directional derivative must be considered. Moreover, the notion of subdifferential has to be replaced by the notion of Clarke subdifferential [27]. The Clarke subdifferential at point  $x$  is the set containing all of the convex combinations of limits of gradients at points located in the neighborhood of  $x$ :

$$\partial f(x) = \text{conv} \left\{ \lim_{i \rightarrow \infty} \nabla f(x_i); x_i \rightarrow x \text{ and } \nabla f(x_i) \text{ exists} \right\} \quad (15)$$

In order to define the Clarke subdifferential more formally, we give the definition of a generalized directional derivative in a first step:

**Definition 5** – Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be a locally Lipschitz-continuous function. The generalized directional derivative of  $f$  at  $x$  in the direction  $v \in \mathbb{R}^n$  is defined by:

$$f^\circ(x; v) = \limsup_{y \rightarrow x, t \downarrow 0} \frac{f(y + tv) - f(y)}{t}$$

**Definition 6** – Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be a locally Lipschitz-continuous function. The Clarke subdifferential of  $f$  at  $x$  is the set  $\partial f(x)$  of vectors defined by:

$$\partial f(x) = \left\{ s \in \mathbb{R}^n : f^\circ(x; v) \geq s^T v \forall v \in \mathbb{R}^n \right\} \quad (16)$$

Theorem 1 cannot be generalized to arbitrary non-convex functions: a locally Lipschitz continuous function  $f$  has a local minimum at  $x^*$  if  $0 \in \partial f(x^*)$ , but it is not a sufficient condition. There exist, however, classes of functions for which the result still holds, it is the case, for instance, of  $f^\circ$ -pseudoconvex functions, which are defined by:

**Definition 7** – A locally Lipschitz-continuous function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is  $f^\circ$ -pseudoconvex if

$$\forall x, y \in \mathbb{R}^n, f(y) < f(x) \Rightarrow f^\circ(x; y - x) < 0 \quad (17)$$

### Common Descent Direction

The algorithm presented in the next section is based on the existence and construction of a descent direction. We first recall its definition.

**Definition 8** – A vector  $d$  is called a descent direction if  $\exists t_0 > 0$ , such that  $f(x + td) < f(x)$  for all  $t \in [0, t_0]$ .

For smooth functions it is well known that the opposite direction of the gradient is a descent vector. In the non-smooth convex or non-convex context, not all elements of the subdifferential are a descent vector.

There are several techniques to construct such a descent vector: proximal bundle methods [28, 29, 30], quasisecant methods [31], or gradient sampling methods [32, 33]. Considering now  $m$  functions  $f_1, \dots, f_m$ , we show that there exists a vector  $d$  that is a descent direction for each function. Its construction is based on properties of the following convex set  $C$ :

**Lemma 1 ([36])** – Let  $C$  be the convex hull of either

- the gradients  $\nabla f_i(x)$  of the objective functions when they are differentiable,
- or the union of the subdifferentials  $\partial f_i(x)$ ,  $i = 1, \dots, m$  when they are non-differentiable but convex, or
- the union of the Clarke subdifferentials  $\partial f_i(x)$ ,  $i = 1, \dots, m$  if they are non-convex.

Then, there exists a unique vector  $p^* = \text{Argmin}_{p \in C} \|p\|$  such that

$$\forall p \in C : p^T p^* \geq p^{*T} p^* = \|p^*\|^2$$

The existence of the common direction  $d$  and its construction is given by the following theorem:

**Theorem 2 ([36])** – Let  $C$  be the convex set defined in Lemma 1 and let  $p^*$  be its minimum norm element. Then, either we have

- $p^* = 0$  and the point  $x$  is Pareto-stationary or
- $p^* \neq 0$  and the vector  $-p^*$  is a common descent direction for every objective function.

We now have sufficient elements to present the SMGDA (Stochastic Multi-Gradient Descent Algorithm) algorithm.

### The SMGDA Algorithm

As written problem (12) is a deterministic problem, but the objective function expectations are seldom known. A classical approach, the sample average approximation (SAA) method, is to replace each expectancy by an estimator built using independent samples  $w_k$  of the random variable  $W$ , [34, 35]. The algorithm that we propose does not need the objective function expectancy to be calculated, and is based only on the construction of a common descent vector. Let  $\omega$  be given in  $\Omega$ , and consider the deterministic multi-objective optimization problem:

$$\min_{x \in \mathbb{R}^n} \{f_1(x, W(\omega)), f_2(x, W(\omega)), \dots, f_m(x, W(\omega))\} \quad (18)$$

Pursuant to Theorem 2 there exists a descent vector common to each objective function  $f_k(x, W(\omega))$ ,  $k = 1, \dots, m$  at point  $x$ .

The common descent vector depends on  $x$  and  $\omega$ , and therefore will be considered as a random vector denoted by  $d(\omega)$  defined on the probability space  $(\Omega, \mathcal{A}, \mathbb{P})$ .

### The Algorithm

We now list the successive steps of the algorithm that we propose.

1. Choose an initial point  $x_0$  in the design space, a number  $N$  of iterations and a  $\sigma$ -sequence  $t_k : \sum t_k = \infty ; \sum t_k^2 < \infty$ ,
2. At each step  $k$ , draw a sample  $w_k$  of the random variable  $W_k(\omega)$ ,
3. Construct the common descent vector  $d(w_k)$  using Theorem 2 and the gradient sampling approximation method,
4. Update the current point :  $x_k = x_{k-1} + t_k d(w_k)$ .

The last step of the algorithm defines a sequence of random variables on the probability space  $(\Omega, \mathcal{A}, \mathbb{P})$  through the relation

$$X_k(\omega) = X_{k-1}(\omega) - t_k d(X_{k-1}(\omega), W_k(\omega)) \quad (19)$$

Initializing the algorithm with different points in the admissible space, for instance using a random or quasi-random distribution, allows different points located on the Pareto front to be constructed. This procedure is entirely parallelizable.

**Theorem 3 ([36])** – Under a set of assumptions,

1. The sequence of random variables  $X_k(\omega)$  defined by Relation (19) converges in a mean square towards a point  $X^*$  of the Pareto set:

$$\lim_{k \rightarrow +\infty} \mathbb{E} \left[ \|X_k(\omega) - X^*\|^2 \right] = 0$$

2. The sequence converges almost surely towards  $X^*$ .

$$\mathbb{P} \left( \left\{ \omega \in \Omega, \lim_{k \rightarrow \infty} X_k(\omega) = X^* \right\} \right) = 1$$

### Illustration: Optimal Designs of a Sandwich Plate with Uncertainties

We consider a sandwich panel whose constitutive materials are given but their mechanical properties are uncertain: solid foams present random, disordered micro-structure, while a honeycomb core may present uncertain geometrical characteristics, which may result in a

distinct scatter and unpredictability of the macroscopic material properties. These uncertainties will be introduced into the optimization problem by means of random variables.

More precisely, in this application we consider a three-layer non-symmetric sandwich panel with aluminum skins and a regular hexagonal honeycomb core. The mechanical properties of the plate are described by the Young modulus  $E_{(i)}$ , the elastic resistance  $\sigma_{(i)}$  and the mass density  $\rho_{(i)}$  of the upper and bottom skin and of the core constitutive material. We introduce the honeycomb wall thickness/length ratio  $R = t/\ell$ .

The relations yielding the honeycomb core material properties from its geometrical description and from its constitutive material properties are given in [37] and are recalled in Table 1.

$R$	=	$t/\ell$
$\rho_R$	=	$\frac{3}{2 \cos(\tilde{\theta})(1 + \sin(\tilde{\theta}))} \rho_c R$
$E_R$	=	$\rho_{RE_c} / \rho_c$
$\sigma_R$	=	$5.6(R)^{\frac{5}{3}} \sigma_c$

Table 1: Core material property function of  $R$

Two objectives are considered in the design process:

- Minimization of the mass per unit of surface

$$M = \rho_u t_u + \rho_c t_c + \rho_b t_b \quad (20)$$

- Maximization of the critical force leading to a failure mode when two modes are introduced: Mode  $F_{c,1}$  leading to the core indentation and Mode  $F_{c,2}$  leading to the lower-skin plastic stretching. We shall then consider the failure mode  $F_c$ , which appears first:

$$F_c = \min_{i=(1,2)} \{F_{c,i}\} \quad (21)$$

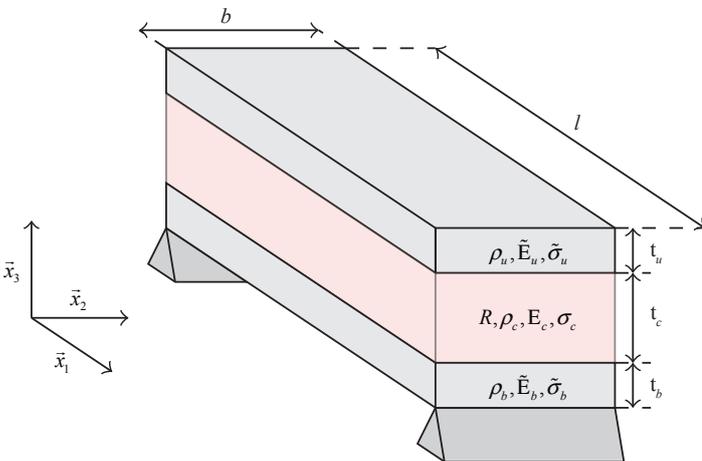
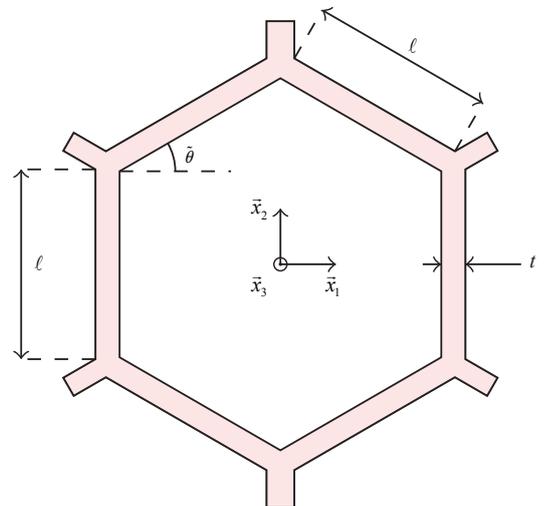


Figure 4 – Three-layer sandwich material beam



$$F_{c,1} = 2bt_u\sqrt{\sigma_c\tilde{\sigma}_u} + ab\sigma_c; F_{c,2} = \frac{4bt_b(t_c + (t_u + t_b)/2)}{4}\tilde{\sigma}_b \quad (22)$$

This last objective function is not differentiable, due to the presence of the minimum function.

Four design parameters are considered: the three thickness parameters of the sandwich plate ( $t_u, t_b, t_c$ ) and the honeycomb wall thickness/length ratio ( $R = \frac{t}{\ell}$ ).

We shall denote by  $x = (t_u, t_b, t_c, R)$  the vector containing the four design parameters. Two types of constraints are introduced into the problem, the first are bounding constraints on each design variable, which are handled using a projection method on the convex feasible set  $C$  defined by these constraints:

$$t_u, t_b \in [0.03, 14] \text{ cm}; t_c \in [0.05, 32] \text{ cm}; R \in [0.01, 0.2]$$

The second type is an inequality constraint for the total thickness  $e$  of the sandwich material:

$$e = t_u + t_b + t_c \leq 0.25$$

In the following numerical application, this last constraint is handled by introducing the exact penalty term [26]

$$g(x) = r \times \min\{0, 1 - e/.25\} \quad (23)$$

into each objective function, where  $r$  is a penalty term. Classically, an increasing sequence  $r = p^q, q = 1, 2, \dots$  is used, with  $q$  being the smallest integer for which the constraint is satisfied. For this application we have chosen  $r = 10^q$ .

We now introduce uncertainty in some parameters of the sandwich material. More precisely, 20% uncertainty is considered for the upper and bottom skin elastic resistance value:

$$(\tilde{\sigma}_u, \tilde{\sigma}_b) = \bar{\sigma}_{Alu} \times ((1, 1) + U_1) \quad (24)$$

where  $U_1$  is a uniform random variable on  $[-.2, .2] \times [-.2, .2]$  and where  $\bar{\sigma}_{Alu} = 350 \text{ MPa}$  is the nominal value. A second uncertain parameter is introduced: the value of the honeycomb angle  $\tilde{\theta} = \bar{\theta}(1 + U_2)$ , where  $U_2$  is a uniform random variable on  $[-.2, .2]$  and where  $\bar{\theta} = \pi/6$ . We shall denote by  $\xi = [U_1, U_2]$  the vector containing the various random variables introduced in the problem, which are assumed to be independent. In order to take into account these uncertainties in the design process, the following stochastic multi-objective problem is considered:

$$\min_{x \in C} \{\mathbb{E}[M(x)], \mathbb{E}[-F_c(x)]\} \text{ subject to } e(x) \leq 0.25 \quad (25)$$

This problem is rewritten using the exact penalty formulation:

$$\min_{x \in C} \{\mathbb{E}[M(x)] - rg(x), \mathbb{E}[-F_c(x)] - rg(x)\} \quad (26)$$

In order to compare the efficiency of the method assessed to the classical genetic algorithm NSGA-II, the expectancies appearing in Problem 12 are estimated, to be used in NSGA-II, through a sample-average method:

$$\mathbb{E}[f(x, \xi)] \approx \frac{1}{N} \sum_{i=1}^N f(w, \xi_i) \quad (27)$$

where  $\xi_i$  are independent samples of the random variable  $\xi$ . The number  $N$  of samples plays a crucial role in the efficiency of the algorithm: an excessively small value will give a wide confidence interval and a poor estimate of the objective function, while an excessively high value will dramatically increase the computational cost (see Figure 5).

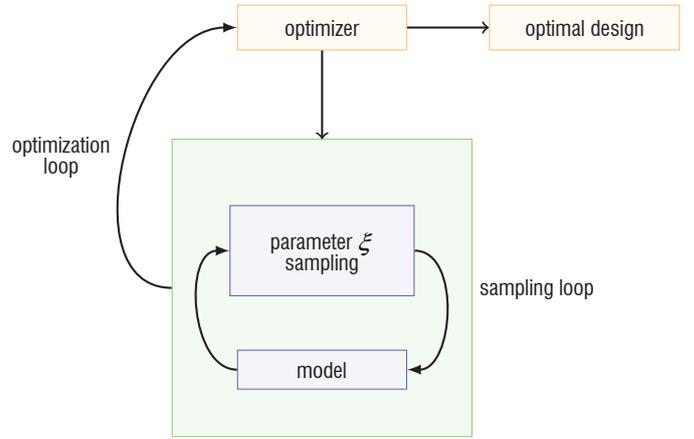


Figure 5 – Stochastic genetic optimization framework

In order to compare the two algorithms, we have chosen to compare results obtained for the same number of function calls  $f(w, \xi_i)$ . In the case of SMGDA, this number includes the number of starting points and the number of iterations per initial point. In the case of NSGA-II it includes the size of the initial population, the number  $N$  used for estimating the objective functions, and the number of generations.

In the numerical illustration, the SMGDA algorithm is initiated from 50 starting points in  $\mathbb{R}^4$  and around 250 iterations were necessary to reach convergence. The same population number (50) is used for NSGA-II. The  $\sigma$ -sequence  $t_k = .03/(3 \times k + 10)$  is used in this illustration. Figures 6 and 7 illustrate the Pareto sets obtained by the two algorithms. The constraint is represented in the design space by a plane. Both methods give solutions that comply with the constraints. Variable  $R$  is represented in the figure by a variation of color according to the color scale in Figure 6. For a low number of function calls, NSGA-II coupled with a Monte Carlo estimator gives a less good result than SMGDA. It needs about one hundred times more calls to the objective functions to reach an identical Pareto set.

In order to evaluate the effect of uncertainty on the objective functions considered at optimal design points  $x^*$  located on the Pareto front, we have estimated the distribution of the random vector  $(M(x^*, \xi(\omega)), -F_c(x^*, \xi(\omega)))$  for two points  $x_1^*$  and  $x_2^*$ , by generating  $10^5$  samples of the random vector  $W$ . The corresponding probability distributions are drawn in Figure 8 and Figure 9, where the position of the chosen point  $x^*$  is indicated on the inner figure. The blue and red dots on the graph denote the mode of failure obtained for some of the samples used to estimate the distribution. A first result that can be drawn is that the distribution obtained is not a classical one, but there is no reason to obtain a classical distribution. The second observation is that the effect of the uncertainties is more important for the critical force objective than for the mass objective. Such a result could be valuable during the design stage of the material, knowing the high sensitivity of the critical force optimal value to uncertain parameters.

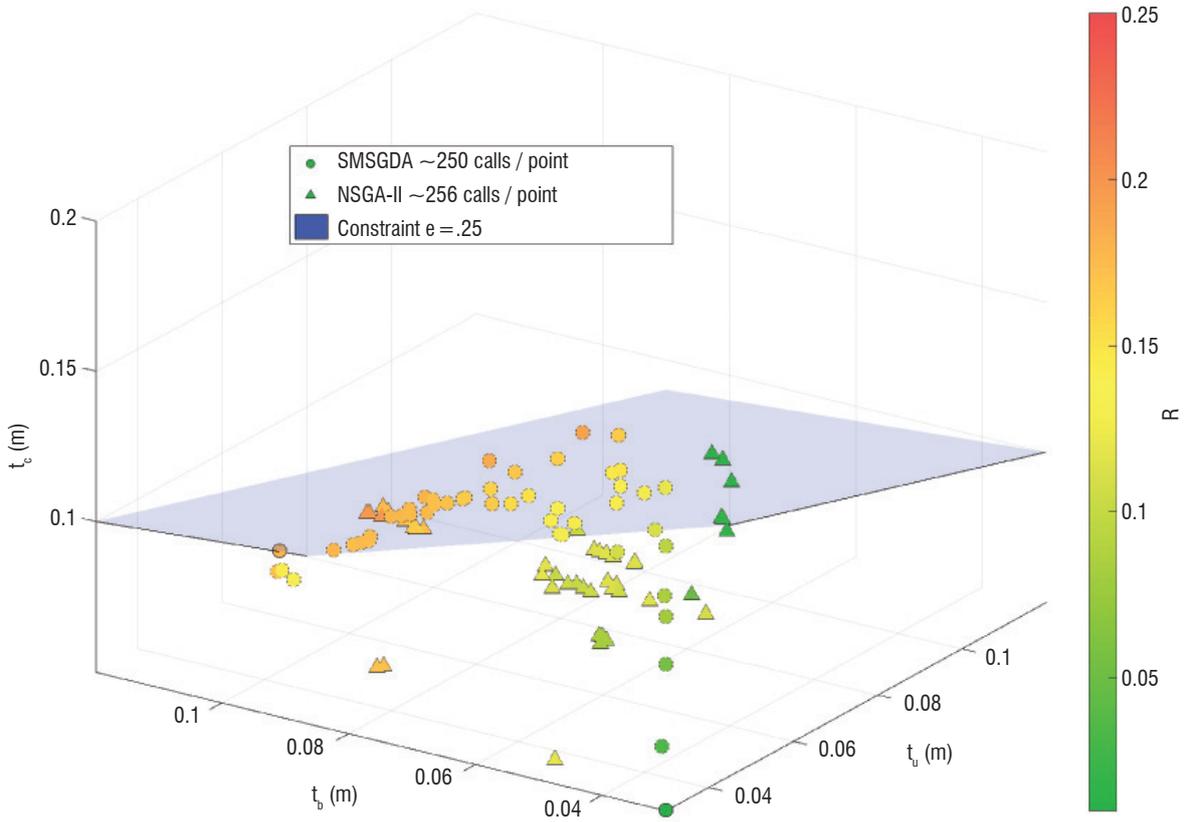


Figure 6 – Pareto set in the design space

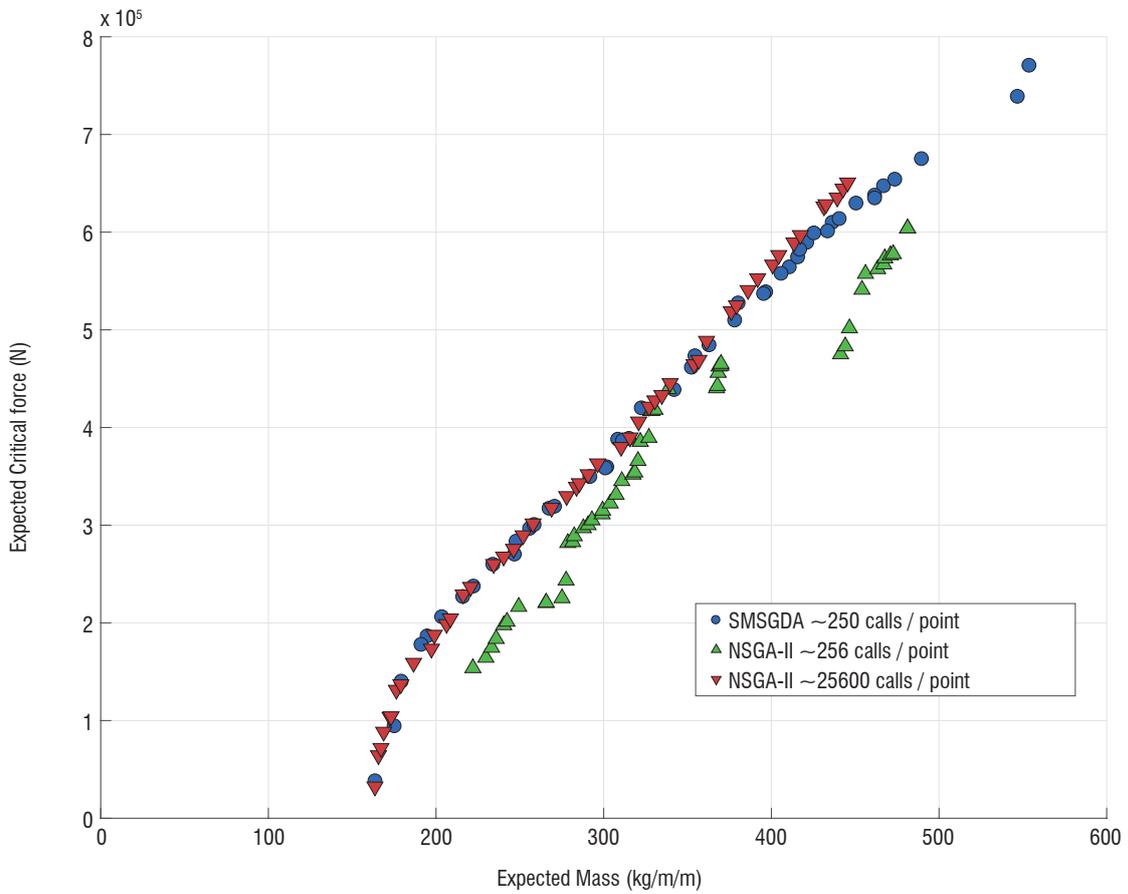


Figure 7 – Pareto set in the objective space

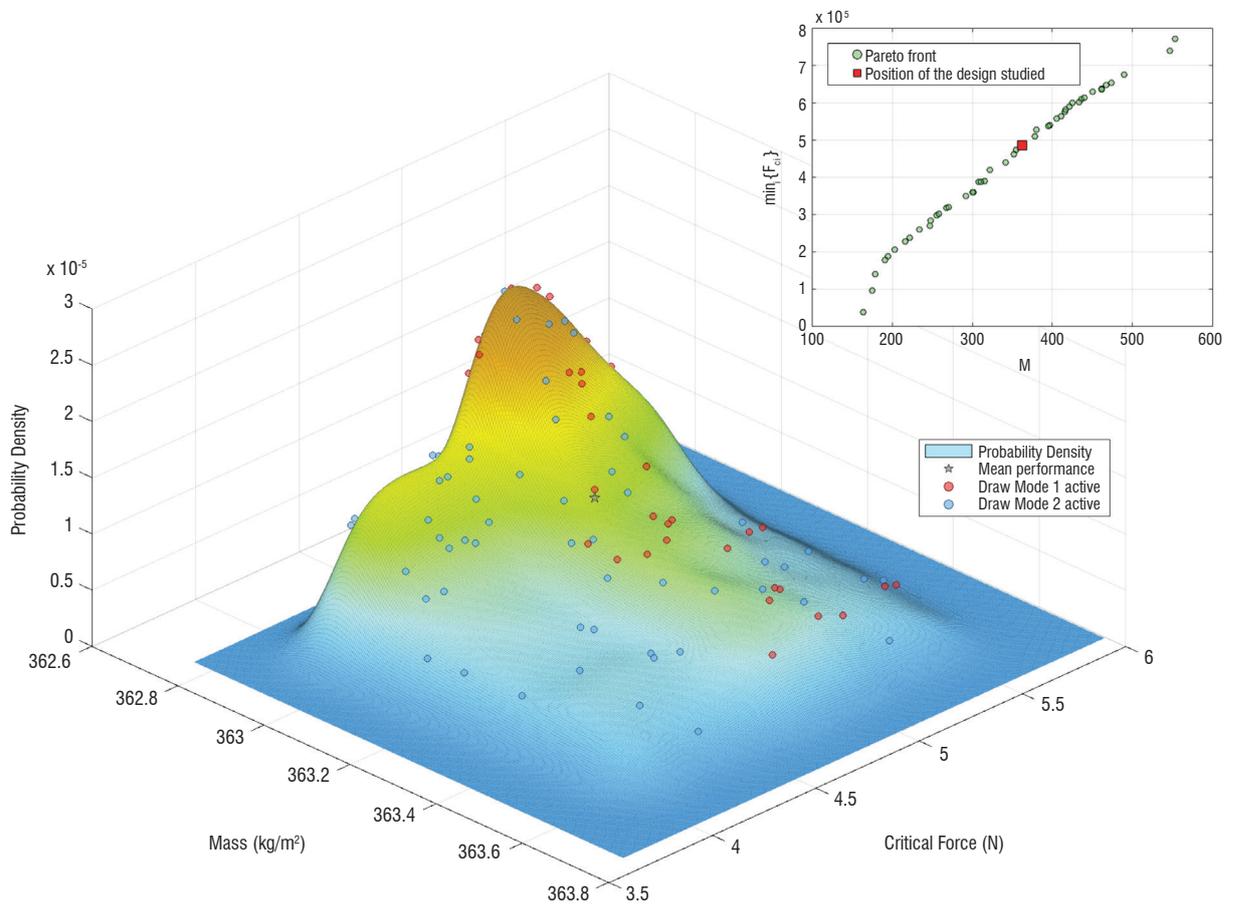


Figure 8 – Probability density of  $M(x_1^*, W(\omega)), -F_c(x_1^*, W(\omega))$

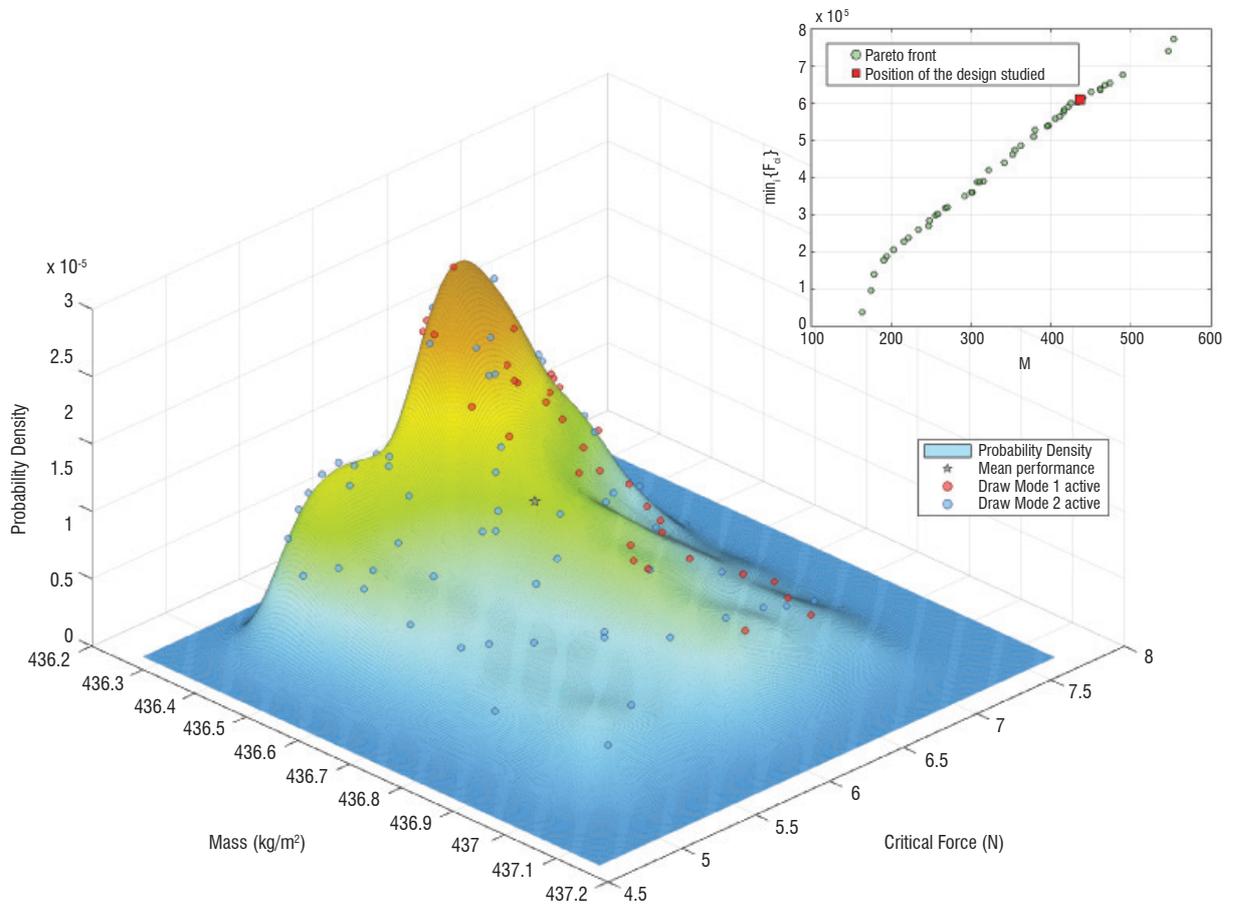


Figure 9 – Probability density of  $M(x_2^*, W(\omega)), -F_c(x_2^*, W(\omega))$

## Conclusion

Although the stochastic gradient algorithm is now a classical approach to deal with uncertain single-objective design optimization problems, it is much more difficult to deal with multiple uncertain objectives. The most classical approaches are based on the use of genetic algorithms, such as NSGA coupled with a scenario method, to construct estimates of the objective expectations, but their usefulness

is limited by the numerical cost induced by the estimator loop. Conversely, the SMGDA algorithm does not rely on the expectation estimation and converges relatively rapidly toward the Pareto boundary. It can, moreover, be entirely parallelizable. An illustration on the design optimization of a sandwich plate has shown its potential usefulness for engineering problems ■

## References

- [1] C. PAPADIMITRIOU, L. S. KATAFYGIOTIS, S.-K. AU - *Effects of Structural Uncertainties on tmd Design : A Reliability-Based Approach*. Journal of Structural Control, 4:65-88, 1997.
- [2] H. G. MATTHIES, C. E. BRENNER, C. G. BUCHER, C. G. SOARES - *Uncertainties in Probabilistic Numerical Analysis of Structures and Solids-Stochastic Finite Elements*. Structural safety, 3:283-336, 1997.
- [3] R. ARNAUD, F. POIRION - *Optimization of an Uncertain Aeroelastic System Using Stochastic Gradient Approaches*. Journal of Aircraft, doi: arc.aiaa.org/doi/abs/10.2514/1.C032142., 2014.
- [4] B. G. DANTZIG - *Linear Programming Under Uncertainty*. Management Science, 1:197-206, 1955.
- [5] R. E. BELLMAN, L. A. ZADEH - *Decision-Making in a Fuzzy Environment*. Management Science, 17:141-161, 1970.
- [6] N. V. SAHINIDIS - *Optimization Under Uncertainty: State-of-the-Art and Opportunities*. Computers & Chemical Engineering, 28(6-7):971-983, 2004.
- [7] R. ROY, S. HINDUJA, R. TETI - *Recent Advances in Engineering Design Optimisation: Challenges and Future Trends*. Manufacturing Technology, 57:697-715, 2008.
- [8] R. E. BELLMAN - *Dynamic Programming*. Princeton University Press, 1957.
- [9] A. D. KIUREGHIAN - *First and Second Order Reliability Methods*. Chapter 14, CRC Press, 2005.
- [10] K. SCHITTKOWSKI - *A FORTRAN Subroutine for Solving Constrained Nonlinear Programming Problems*. Ann Operations Research, 5:485-500, 1985.
- [11] B. AROUNA - *Robbins-Monro Algorithms and Variance Reduction in Finance*. The Journal of Computational Finance, 7(2), Winter 2003/2004., 7(2), Winter 2003/2004.
- [12] P. GLASSERMAN - *Monte Carlo Methods in Financial Engineering*. Stochastic Modelling and Applied Probability, volume 53 of Applications of Mathematics (New York), Springer-Verlag, New York, 2004, 2004.
- [13] J. LELONG - *Asymptotic Properties of Stochastic Algorithms and Pricing of Parisian Options*. PhD thesis, Ecole Nationale des Ponts et Chaussées, 2007.
- [14] H. ROBBINS, S. MONRO - *A Stochastic Approximation Method*. Ann. Math. Statistics, 22:400-407, 1951.
- [15] Y. ERMOLIEV - *Stochastic Quasigradient Methods and their Application to Systems Optimization*. Stochastics, 9:1-36, 1983.
- [16] Y. ERMOLIEV, R. WETS - *Numerical Techniques for Stochastic Optimization*. Springer Verlag, 1988.
- [17] J. DUPACOV - *Stochastic Programming: Approximation via Scenarios*. Proceedings of 3<sup>rd</sup> Caribbean Conference on Approximation and Optimization, Puebla, 1995.
- [18] A. SHAPIRO - *Handbooks in OR & MS*. Vol. 10, chapter 6, Monte Carlo Sampling Methods, pp. 353-425. Elsevier Science B.V., 2003.
- [19] M. DUFLO - *Random Iterative Models*. Springer-Verlag, 1997.
- [20] E. ALBANO, W. RODDEN - *A Doublet Lattice Method for Calculating Lift Distributions on Surfaces in Subsonic Flows*. AIAA Journal, 7(2):279-285, Feb 1969. Errata vol 7, no. 11, Nov. 1969.
- [21] R. L. FOX, M. P. KAPPOR - *Rates of Change of Eigenvalues and Eigenvectors*. AIAA Journal, 6(12):2427-2429, 1968.
- [22] C. S. RUDISILL, K. BHATIA - *Optimization of Complex Structures to Safety Flutter Requirements*. AIAA Journal, 9(8):1487-1491, 1971.
- [23] S. H. TIFFANY, M. KARPEL - *Aeroservoelastic Modeling and Applications Using Minimum-State Approximations of the Unsteady Aerodynamics*. Technical Report TM-101574, NASA, 1989.
- [24] H. ZINGEL - *Measurement of Steady and Unsteady Airloads on a Stiffness Scaled Model of Modern Transport Aircraft Wing*. Proc. Int. Forum on Aeroelasticity Structural Dynamics, volume DGLR 91-06, pp. 120-131, 1991.
- [25] K. MIETTINEN - *Nonlinear Multiobjective Optimization, Volume 12 of International Series in Operations Research & Management Science*. Springer US, 1998.
- [26] A. BAGIROV, N. KARMITSA, M. M. MKEL - *Introduction to Nonsmooth Optimization: Theory, Practice and Software*. Springer, 2014.
- [27] F. H. CLARKE - *Optimization and Nonsmooth Analysis*. Wiley, 1983.
- [28] K. C. KIWIEL - *Methods of Descent for Nondifferentiable Optimization*. Number 1133 in Lecture Notes in Mathematics. Berlin edition, 1985.
- [29] O. WILPPU, N. KARMITSA, M. M. MKEL - *New Multiple Subgradient Descent Bundle Method for Nonsmooth Multiobjective Optimization*. Technical Report, Turku Centre for Computer Science, 2014.

- [30] M. M. MÄKELÄ, N. KARMITSA, O. WILPPU - *Mathematical Modeling and Optimization of Complex Structures*. Chapter Proximal Bundle Method for Nonsmooth and Nonconvex Multiobjective Optimization, pp. 191-204. Springer, 2016.
- [31] A. BAGIROV, L. JIN, N. KARMITSA, A. AL NUIMAT, N. SULTANOVA - *Subgradient Method for Nonconvex Nonsmooth Optimization*. Journal of Optimization Theory and Applications, 157:416-435, 2013.
- [32] J. V. BURKE, A. S. LEWIS, M. L. OVERTON - *Approximating Subdifferentials by Random Sampling of Gradients*. Mathematics of Operation Research, 27:567-584, 2002.
- [33] J. V. BURKE, A. S. LEWIS, M. L. OVERTON - *A Robust Gradient Sampling Algorithm for Nonsmooth, Nonconvex Optimization*. SIAM J. Optim., 15:751-779, 2005.
- [34] H. BONNEL, J. COLLONGE - *Stochastic Optimization Over a Pareto Set Associated with a Stochastic Multi-Objective Optimization Problem*. J. Optim. Theory Appl., 162:405-427, 2014.
- [35] J. FLIEGE, H. XU - *Stochastic Multiobjective Optimization: Sample Average Approximation and Applications*. Journal of Optimization Theory and Applications, 151:135-162, 2011.
- [36] F. POIRION, Q. MERCIER, J. A. DÉSIDÉRI - *Descent Algorithm for Nonsmooth Stochastic Multiobjective Optimization*. Computational Optimization and Applications, pp. 10.1007/s10589-017-9921-x, 2017.
- [37] L. J. GIBSON, M. F. ASHBY - *Cellular Solids. Structure and Properties*. Cambridge University Press, 1997.
- [38] J. A. DÉSIDÉRI - *Multiple-Gradient Descent Algorithm (MGDA) for Multiobjective Optimization*. CRAS Paris, Ser. I, 350:313-318, 2012.

## AUTHORS

---



**Fabrice Poirion** has a PhD in Mathematics from the Pierre et Marie Curie University (Paris 6), and obtained the *Habilitation à Diriger des Recherches* (HDR) at the same University in 1999. He currently works in stochastic structural dynamics, and aeroelasticity, and manages an internal project on probability and statistics, at ONERA.



**Quentin Mercier** is a PhD student at ONERA and an ENS Cachan graduate.