Regularized Least Squares for the Building of a Sizing Criterion Based on Damage Mechanics

Orestis Friderikos, Emmanuel Baranger, Damien Guillon

▶ **To cite this version:**

Orestis Friderikos, Emmanuel Baranger, Damien Guillon. Regularized Least Squares for the Building of a Sizing Criterion Based on Damage Mechanics. Composite Structures, 2019, pp.111653. 10.1016/j.compstruct.2019.111653 . hal-02390625

HAL Id: hal-02390625
https://hal.science/hal-02390625

Submitted on 3 Dec 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.

Regularized Least Squares for the Building of a Sizing Criterion Based on Damage Mechanics

Orestis Friderikos[1*], Emmanuel Baranger[1], Damien Guillon[2]

1. LMT, ENS-Paris-Saclay, CNRS, Université Paris-Saclay, 61 avenue du Président Wilson, 94235 Cachan, France

2. CETIM, Technocampus Composites Z.I. du Chaffault, 44340, Bouguenais, France

Abstract

Utilization of the building-block approach for the design of composite parts emerged from the need to develop a true understanding of the complex composite structural response through experimentation and analysis. The present work aims to provide a numerical counterpart to the expensive experimental testing of advanced composite laminate structures. This detailed analysis can be fully integrated into the building-block approach to enhance its capability to examine the full-scale structural behavior and improve its reliability. To this end, a Virtual Testing Analysis (VTA) based on the structural decomposition into a number of complicated laminates is introduced for determining premature damage evolution. VTA augment FEM with a hierarchical procedure which goes down to the meso-scale where a detailed modeling of each component is examined using non-linear FEM. Design of Experiments is utilized for an optimal sampling of the design space related to the laminate subcomponents. Linear least squares regression using Tikhonov regularization (L_2-norm penalty function) and Truncated Singular Value Decomposition (TSVD) are used to combat overfitting and stabilize possible ill-conditioned matrix inverse problems. Furthermore, a Multi-Stage Least Squares methodology is implemented as an alternative approximation which utilizes a separation of the design variables into sets. This method yields optimal solutions in each subspace in a hope to obtain the final global optimal solution from the composition of locally optimal solutions. Perfor-

*Corresponding author.

Email address: friderikos@lmt.ens-cachan.fr (Orestis Friderikos[1])

mance and efficiency is evaluated both for training and validation data using error measures based on the residuals. Preliminary results showed acceptable accuracy for the linear least squares and Multi-Stage Least Squares methods. An important key aspect of VTA is that off-line computations for the advanced laminates can be stored and used for different full-scale structures as a simple failure criterion.

Keywords: Virtual Testing Analysis, Cohesive zone modeling, Surrogate analysis, Regularization, Multi-Stage Least Squares, Virtual charts

---

1. Introduction

Design and development of advanced composite materials is a challenging procedure which involves a wide range of mechanical testing and numerous standards. Providing robust and accurate tools for cost effective design of composite parts remains a challenging task. Within the engineering community, the structural substantiation process, which uses testing and analysis at increasingly complex levels, has become known as the 'building block approach' [1]. This approach for verifying a structural design for certification consisting of coupons, elements, structural details, subcomponents and full-scale prototype experimental testing.

Based on the building block approach, in this paper a design/analysis methodology is proposed which exploits high power computing along with computational structural mechanics to achieve certification-by-analysis (CBA) for composite structures (Fig. 1). The CBA process involves an accurate simulation of physical tests using a Virtual Testing Analysis (VTA). CBA provides assistance to planning and reduction of experimental testing at coupon and large component levels. It can also be used to perform robust design of structures by minimizing a design's sensitivity to certain types of failures such as delamination. VTA can be considered as a numerical 'twin' to the experimental testing of composites. Here, the choice is to work at the limit between the scale of the material coupon including simple details (holes, ply drops, etc) and the scale of a part.

Under this framework, a damage analysis to determine the early stages of degradation evolution under service loading is introduced. Physics based composite damage and failure mechanisms are used to interrogate the structural response. With the current approach, the evaluation of damage and failure criteria are implemented at the important structural components. These

laminate substructures (e.g., stiffeners, tapering of the thickness of a laminate by terminate, or dropping, internal or external plies, corners [2]) (see Fig. 1), henceforth called Representative Structural Component (RSC) problems are modeled using nonlinear FEM (Abaqus/Standard). Then, appropriate stiffness degradation functions with a predefined tolerance are reflected in the full-scale structural stiffness. The stiffness changes can be evaluated in various reference directions and can be proposed as a damage criterion. As a benchmark test case, simulation of interlaminar failures, namely delamination initiation/growth mechanisms of an internal ply drop laminate using the Cohesive Zone Modeling (CZM) formulation [3]-[4] in Abaqus is investigated. It has to emphasized that the method can be extended to other degradation mechanisms such as transverse cracking.

An efficient sampling of a high dimensional design space for each laminate substructure is an essential part of VTA. Design of Experiments (DOE) provides the basic theory on where these design points should be placed in the design space in order to maximize the information from a limited sample size. Classical designs include factorial/fractional factorial designs, central composite designs, Box-Behnken designs, Plackett-Burman designs, Koshal designs and D-optimal designs [5]. Here, design variables corresponding to the operating loading conditions, material orientation angles and geometric features are introduced. Several design methods are implemented, i.e., uniform sampling, Latin Hypercube Sampling (LHS) and sampling subjected to geometrical constraints. Hence, a large-scale system identification problem arise in the estimation of the polynomial regression approximations.

Surrogate-based analysis is an effective approach for the design of computationally expensive models such as those found in fluid and structural mechanics, among other disciplines. Different model estimation have been shown to be effective in the context of surrogate analysis, namely polynomial regression, Gaussian radial basis functions, Kriging, Support Vector Machines and Neural Networks [6]. In this work, the classical Tikhonov regularization and the Truncated Singular Value Decomposition (TSVD) are applied to Linear Least Squares (LLS) to balance the bias-variance trade-off and stabilize a possibly ill-conditioned matrix inversion problem. SVD and TSVD provide the basic tools of regularization techniques and their main properties are summarized.

Moreover, a Multi-Stage Least Squares (MSLS) approach is proposed as an alternative regression method. MSLS has been used for the prediction of fatigue life in composite materials using the SVD [7]. The central idea

is to separate the variables into sets and yield locally optimal solutions in each related subspace. The separation of variables reduce the dimensionality of the problem and in turn, the computational complexity. Essentially, MSLS defines LLS minimization problems in each subspace and flexibility is offered for using various polynomial order approximations. Numerical results demonstrate similar accuracy for the LLS and MSLS, however with a computational complexity reduction of more than two orders of magnitude for the latter method.

In this introduction, only general information is cited. A more detailed literature survey will be given in each section. The remainder of this paper is organized as follows. In Section 2 the local scale modeling approach for the determination of interlaminar composite failures is presented. Design of Experiments is discussed in Section 3. In Section 4, the LLS method is presented followed with a basic introduction to SVD. Section 5 introduces Tikhonov and TSVD regularization. The MSLS subspace approximation method is described in Section 6. Next, in Section 7, we discuss several important computational aspects of the methods, whereas in Section 8, the LLS and MSLS model validation is addressed. Finally, in Section 9 the main outcomes and some important perspectives are highlighted.

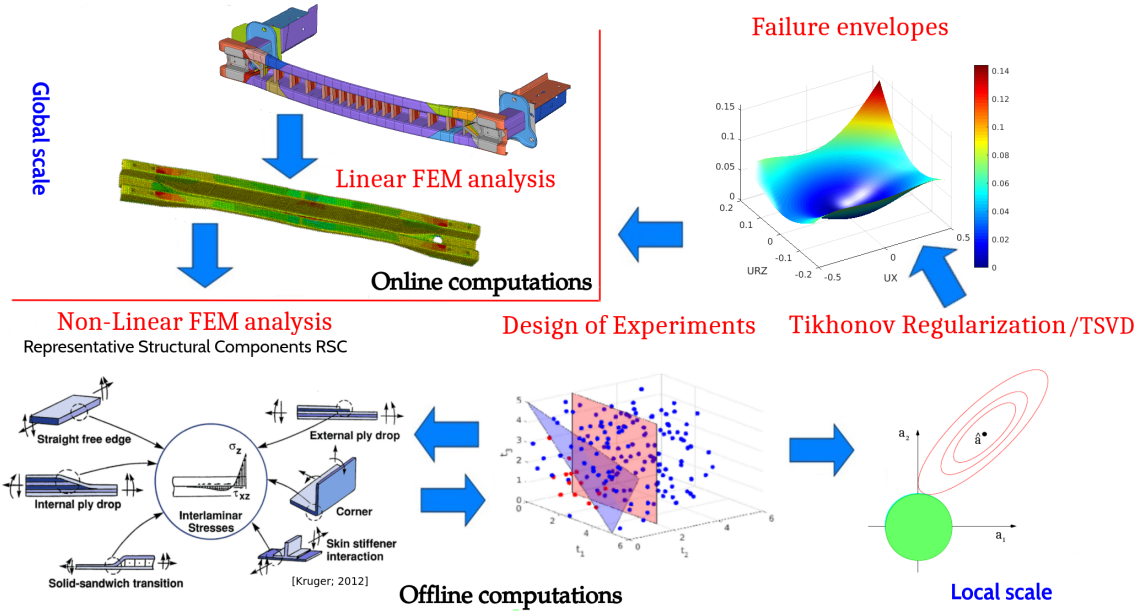

Figure 1: Virtual Testing Analysis flowchart.

2. Local scale analysis

Simulation of interlaminar composite failures, namely delamination initiation/growth mechanisms of an internal ply drop laminate using the cohesive zone modeling formulation is the benchmark model problem under consideration. Note, that no intralaminar failures are taken into account in the present work.

2.1. Simulation of the Debonding Mechanisms of Internal Ply Drop Laminates

Tapered laminated structures, which are formed by dropping off some of the plies at discrete positions over the laminate, have received attention because of their structural tailoring capabilities, damage tolerance, and their potential for creating significant weight savings in engineering applications. Unfortunately, the inherent weakness of this construction is the presence of material and geometric discontinuities at the ply drop region which induce premature interlaminar failure at interfaces between dropped and continuous plies and may cause a significant loss of structural integrity.

First illustration of a reference ply drop geometry is shown in Fig. 2 which concerns a 10-mm-long and 2-mm-wide specimen. The resin pocket is idealized to be right triangular shaped. This geometry represents the worst-case scenario, i.e., it is more prone to delamination than other triangular shapes (especially in the thin section) [8]. Hence, a minimization of the number of parameters needed to describe the ply-drop geometry is accomplished. Furthermore, the resin pocket is considered as a void to increase the stress intensity factors in the local models. This promotes thin section delamination, while the onset and growth of interlaminar cracks in the thick section are both unaffected by the material/geometrical properties of the resin pocket [9],[10].

Based on the cohesive zone modeling, a bi-linear triangular cohesive law for failure analysis is used for pure Mode I, Mode II or Mode III [2], [3], [4], [11], [12], [13], [14], [15]. A mixed-mode criterion accounting for the effect of the interaction of the traction components in the onset of delaminaton is used. Damage initiation is predicted using a quadratic failure criterion considering nominal stresses where compressive normal tractions do not affect delamination onset. In order to accurately account for the variation of fracture toughness as a function of mode ratio, a mixed-mode energy-based

damage evolution criterion proposed by Benzeggagh and Kenane (BK) [16] is used.

In all simulations a displacement-controlled procedure is applied since it is more stable than load control for nonlinear analysis. The model is subjected to prescribed displacements imposed at the free gauge length (thin ply drop section) of 2 mm length while the constraint gauge length (thick section) is 3 mm. The thick section is comprised of three composite 0° plies with uniform thickness of 1 mm. The drop off zone is corresponding to a tapered angle of 15°. Material properties of Celstran/Ticona/PA66 used in the model are as follows: $E_{11} = 38600$ MPa, $E_{22} = E_{33} = 8800$ MPa, $\nu_{12} = \nu_{13} = \nu_{23} = 0.3$, $G_{12} = G_{13} = G_{23} = 3500$ MPa. The elastic properties of the interface material are defined using uncoupled traction-separation behavior with interface stiffness of $K_{nn} = K_{ss} = K_{tt} = 1 \times 10^6$ N/mm^3. The relevant material data are as follows: $t_n^\circ = t_s^\circ = t_t^\circ = 40$ MPa represent the nominal cohesive strengths in normal and the two shear directions, respectively, $\boldsymbol{\mathcal{G}}_n^c = 2.2$ N/mm , $\boldsymbol{\mathcal{G}}_s^c = \boldsymbol{\mathcal{G}}_t^c = 3.0$ N/mm are the critical fracture energies along the normal and the first and the second shear directions, and $n = 1.57$ is the semi-empirical criterion exponent applied to delamination initiation and growth [16].

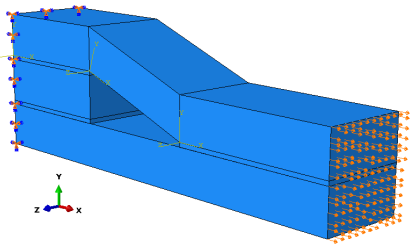

Figure 2: Geometry of the benchmark ply drop model.

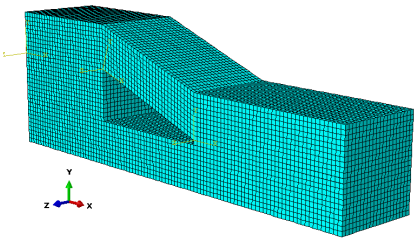

Figure 3: Mesh of the benchmark ply drop model using cohesive elements.

Fig. 4 shows a final damage state and the associated failure mechanisms of the benchmark ply drop model under static tensile loading using prescribed displacements of 0.25 mm in direction-x. The scalar damage variable of the cohesive elements (SDEG) presents the failure state of the cohesive zones (red color identifies completely damaged elements). Delamination starts from the root of the ply drop (thin section) followed by delamination propagation at the two cohesive zones at the thick ply drop section. Displacement discontinuity in the longitudinal direction-1 indicates Mode II fracture in both

cohesive zones (see Fig. 4.c) whereas displacement discontinuity in the transverse direction-2 resulting to Mode I fracture occurs only in the ply drop thin section cohesive zone.

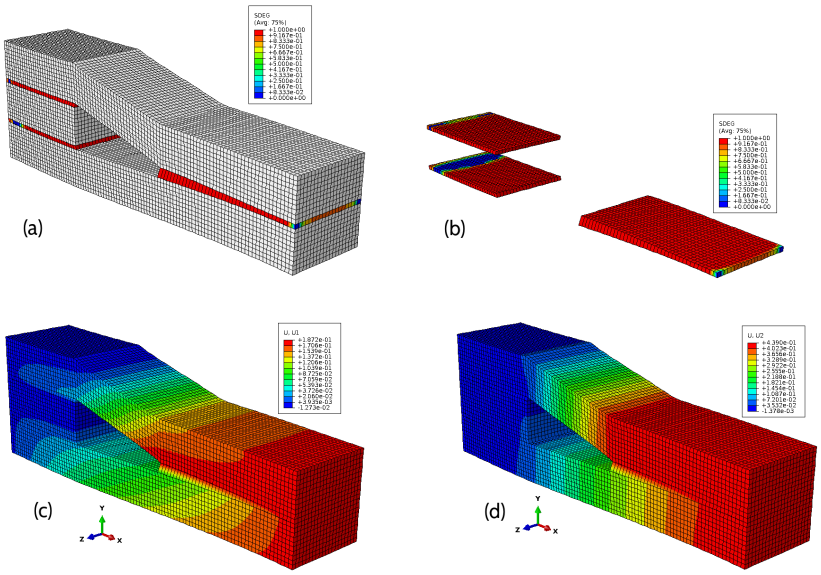

Figure 4: (a) Damage evolution of the benchmark ply drop model under tensile loading using the scalar damage variable (SDEG) of the cohesive elements (red color identifies completely damaged elements); (b) Detailed profile of the damaged cohesive surfaces; (c) U1 displacement field indicating discontinuity in longitudinal direction-1 (analogous to Mode II fracture); (d) U2 displacement field indicating discontinuity in the transverse direction-2 analogous to Mode I fracture only in the ply drop thin section cohesive zone.

In order to reduce the computational cost, an equivalent 2D ply drop model is implemented using Surface-Based Cohesive Behavior (SBCB) [17]. SBCB provides a simplified way to model cohesive connections with negligibly small interface thickness using the traction-separation constitutive law. Using this approach, cohesive behavior is defined as a surface interaction property and a pure master-slave formulation is enforced for surfaces with cohesive behavior. The 2D model has 2300 bilinear plane stress quadrilateral elements of type CPS4 with a total number of 2563 nodes (Fig. 5 and Fig. 6). The size of elements in the direction of crack propagation is 0.1 mm. It has to be mentioned that in Abaqus/Standard it is not possible to take into account the different fiber orientations of the plies for the 2D case. Thus, a transformation of the material axes is needed to determine the

full anisotropic tensor. Nevertheless, some three-dimensional effects, such as anticlastic (saddle-shaped) curvature or elastic coupling are omitted. And besides, not all loading conditions can be applied. To this end, special caution must be taken using a 2D model in a more general case. However, withing the prescribed loading range, numerical investigations revealed not significant discrepancies between the 3D and 2D models. Hence, the 2D model offers the desirable trade-off between complexity and accuracy given that the computational cost is reduced by a factor in the range of 5 up to 7 times at the latter case.

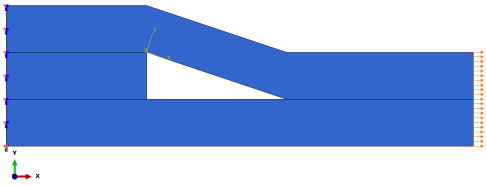

Figure 5: Geometry of the benchmark 2D equivalent ply drop model using SBCB.

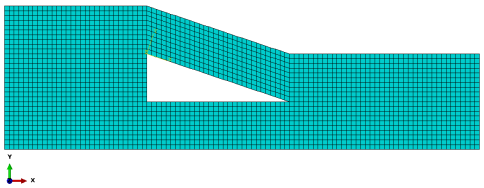

Figure 6: Mesh of the benchmark 2D equivalent ply drop model using SBCB.

2.2. Linear Perturbation Analysis (LPA) - Directional Stiffness Loss criterion

For building a robust sizing criterion the structure's integrity due to damage accumulation related to delamination of the cohesive zone layers is investigated. A precise characterization of the structural response requires the evaluation of the stiffness loss in various reference directions even it is subjected to an arbitrary loading state. To this end, a natural choice is to use a Linear Perturbation Analysis (LPA) as a measure of the structure's integrity. LPA is employed using a general analysis step to provide the linear response of the system. It is applied from time to time during the nonlinear analysis by including the linear perturbation steps between the general response steps. The linear perturbation response has no effect on the general analysis since the perturbation load is taken arbitrarily to be a very small number. LPA steps are performed in directions 1, 2 and rotation in direction 3 (see Fig 7) using prescribed displacements/rotations resulting to the evaluation of the Directional Stiffness Loss (DSL) functions DSL_a, $a = \{\mathrm{UX}, \mathrm{UY}, \mathrm{URZ}\}$ defined as:

$$\mathrm{DSL}_a = \frac{\mathrm{K}_0^a - \mathrm{K}^a}{\mathrm{K}_0^a}, \qquad \mathrm{K}^a = \frac{R^a}{\tilde{p}^a} \tag{1}$$

where K_0^a and K^a represent the initial and current structural stiffness (damaged) at state n, respectively, R^a is the reaction force or moment and $\tilde{p}^a$ is the perturbation load (displacements or rotation). The stiffness K^a in LPA is associated with the evaluation of the secant modulus. Using definitions (Eq.1) for the DSL criterion, a percentage of stiffness loss (for example 10%) can be defined as a design limit. Note also that implementation of LPA increases the computational time due to multiple loading-unloading steps as occurs in the experimental case.

2.3. Parametric modeling of Internal Ply Drop Laminates

A sketch of the parametric internal ply drop model used in FEM simulations is presented in Fig. 7. The geometric design variables t_1, t_2 and t_3 are associated to the thickness of the ply drop plies. At the examined case, several geometrical constraints are enforced according to the final net shape design. These constraints concern not only the thickness of the individual plies of the laminate but also the overall structural thickness. In particular, the geometric variables are subjected to the following design constraints: $t_1+t_2+t_3 \leq 5$, $0.2 \leq t_1 \leq 4.6$, $0.2 \leq t_2 \leq 2.0$ and $0.2 \leq t_3 \leq 4.6$ (dimensions in mm). The drop-off zone length l_2 is varied in each ply drop configuration corresponding to a constant tapered angle $\gamma = 15^\circ$. The length of the cohesive zones are constant in all configurations with a length of $l_1 = 3$ mm and $l_3 = 5$ mm for the thick and the thin section of the ply drop, respectively. The design variables $\theta_1, \theta_2, \theta_3$ correspond to fiber orientations angles of the three plies from the bottom to the upper ply, respectively. The following design limits have been considered: $\theta_1 \in [0, \pi/2]$, $\theta_2 \in [0, \pi/2]$, $\theta_3 \in [0, \pi/2]$. Finally, prescribed displacements UX, UY and rotation URZ are imposed to the thin section of the ply drop laminate as shown in Fig 7. The following design limits have been considered: UX $\in [-0.5, 0.5]$ mm, UY $\in [-0.5, 0.5]$ mm, URZ $\in [0, \pi/18]$.

Note also that in the general case of multilayered internal ply drops, a high dimensional design space may be required, and moreover SBCB has to be defined in several ply drop layers. Therefore, the computational cost may be prohibitive for this type of problems. To circumvent this issue, a convenient choice is to use a representation of the k-composite plies into an equivalent single orthotropic ply which results to the model shown in Fig 7.

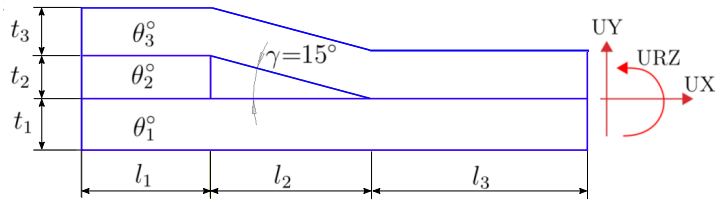

Figure 7: Sketch of the internal ply drop laminate model used for the parametric FEM simulations.

3. Design of Experiments

The choice of sampling points is a challenging and difficult problem in general. DOE is the theory for optimal sampling of the design variable space. It is well known that the sample set is of paramount importance to the success of the approximation function modeling task [18]. By efficiently selecting the sample set, a model can be more accurate with less information. Intuitively, the data points must be spread over the design space $\mathbb{R}^d$ in such a way as to convey a maximum amount of information. This is a non-trivial task, since little or nothing is known about the model structure a priori. In this work, classical DOE methodologies and little effort heuristic strategies adapted to the specific features and constraints of the input variables. However, since the sampling data sets are obtained separately in their related subspaces $\mathbb{R}^n (n < d)$, their cross product cannot guarantee an efficient coverage of the global design space. In this way, certain regions might be under-sampled while others might be over-sampled since the distribution of the training points is highly inhomogeneous. Therefore, it is questionable if subspace sampling leads close to an optimal space filling of the global design space $\mathbb{R}^d$. It has to be stressed that although the adopted methodology is simple, it may lead to poor performance on the approximation accuracy. However answering this issue is far beyond the scope of this work.

Latin hypercube sampling is a statistical method for generating a near-random sample of parameter values from a multidimensional distribution. LHS is an efficient technique especially if the number of design variables is greater than $n > 5$. Numerous modifications have been proposed for LHS to minimize correlation error and to ensure the unidimensional uniformity [19]-[21]. The LHS design is used for generating randomly distributed points in $\mathbb{R}^3$ associated to the fiber orientation angles $\theta_1, \theta_2, \theta_3$ within each ply. Fig. 8 shows the scatter plot defined with $d = 3$ dimensions and $N_\Theta = 40$ training

points. In the same figure, the sampling points for the validation analysis are also plotted.

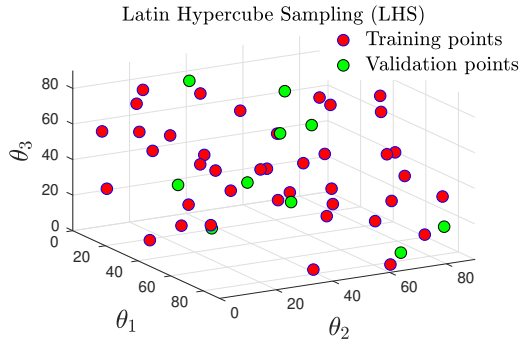

Figure 8: Training sampling points (in red) of an orthogonal array based on LHS ($N_\Theta = 40$) and validation points (in green). The design variables corresponds to fiber orientations angles $\theta_1, \theta_2, \theta_3$ of the transverse isotropic RSC plies.

For the loading variables, a detailed information of the damage state at maximum loading is of significant importance. Here the strategy is to sample at the boundary of the design space since all intermediate states are determined due to proportional loading in FEM computations. In a trivial case, one may consider sampling points uniformly distributed on the surface of a $\mathbb{R}^d$ sphere. Nevertheless, this choice could quickly become too computationally expensive due to the explosion of samples needed to cover the surface of a sphere in $\mathbb{R}^d$. In the investigated design, $N_L = 62$ training sampling points are equi-distributed on the surface of sphere in $\mathbb{R}^3$ with polar angle $\theta = \pi/6$ and azimuth angle $\phi = \pi/6$ using a parameterization $x = r\sin(\theta)\cos(\phi)$, $y = r\sin(\theta)\sin(\phi)$ and $z = r\cos(\theta)$. Moreover, for obtain the chosen number of points for validation analysis, the random number generator of Matlab is used.

Finally, sampling the design space $\mathbb{R}^3$ for the ply drop off laminate thickness t_1, t_2, t_3 cannot be addressed with common DOE methodologies due to geometrical constraints. Application of the LHS in a constraint space requires complex algorithms which cannot in general guarantee an optimal sampling, especially in high dimensional spaces [22]. Here, a heuristic non-optimal but simple procedure is employed: first, LHS is utilized considering the minimum/maximum values of the design variables and at a next step, the points which are not fulfill the geometrical constraints are rejected from the design space (Fig.9). Additionally, rejected points lying within a specified thresh-

old adjacent to the constrained space boundaries are projected back on the boundaries. $N_G = 20$ training sampling points have been used corresponding to the ply thickness of the RSC part. Note that the same procedure have been used to sample the validation points. Again, since the number of points may be exceeding the chosen number of points for validation analysis, the random number generator of Matlab is used.

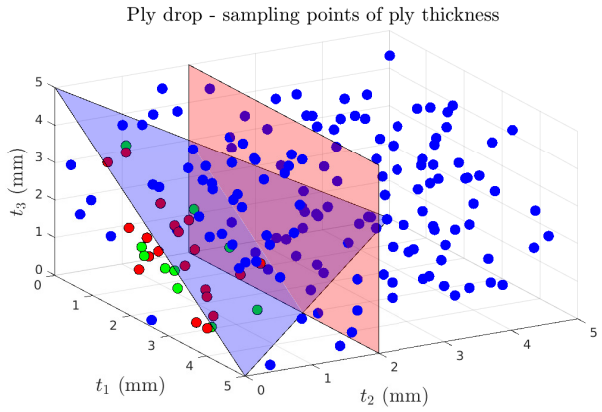

Figure 9: Sampling points of ply drop off laminate thickness subjected to geometrical constraints. After LHS using min/max values, a rejection of points which are not fulfill the constraints is performed. Blue color indicates initial LHS points, red color training points and green color validation (test) points.

4. Surrogate Modeling

4.1. Approximation models

Despite the steady and continuing growth of computational power, the complexity of high nonlinear FEM seems to keep pace with computing advances. These computationally expensive models are often used for optimization, design space approximation and sensitivity analysis, making concept exploration time consuming [23]. Surrogate models can be used instead of the actual FEM analysis to approximate the multivariate input/output behavior of complex systems, based on a limited set of simulations.

One approximation method is the polynomial regression, also known as Response Surface Models [5], using linear or high-order functions, developed by performing least squares curve fit to a set of training data. The data consists of one or more responses related to one or more independent variables.

4.2. Linear Least Squares (LLS) estimates

The linear model is the main technique in regression problems and the primary tool for least squares fitting. LLS minimize a sum of squared errors, or equivalently the sample average of squared errors. That is a natural choice when one is interested in finding the regression function which minimizes the corresponding expected squared error. From the observed response values $y_1, y_2, ..., y_m$ and features x_{ij}, $(i = 1, ..., m,\ j = 1, ..., n)$, LLS is obtain the parameter values $\alpha_1, ..., \alpha_n$ that minimize

$$\underset{\alpha_1,...,\alpha_n \in \mathbb{R}}{\arg\min}\ S(\alpha_1, ..., \alpha_n) = \sum_{i=1}^{m} \left(y_i - \sum_{j=1}^{n} x_{ij}\alpha_j \right)^2 = \|\mathbf{Y} - \mathbf{X}\boldsymbol{\alpha}\|_{L_2}^2 = \|\hat{\boldsymbol{r}}\|_{L_2}^2 \quad (2)$$

where $\mathbf{X}$ is the design matrix, x_{ij} are the regressors and $\hat{\boldsymbol{r}}$ is the m vector of residuals. The minimizing values are denoted with hat symbols: $\hat{\alpha}_1, ..., \hat{\alpha}_n$. To minimize S, its partial derivative with respect to each α_j is set to 0

$$\frac{\partial S}{\partial \alpha_j} = -2 \sum_{i=1}^{m} \left(y_i - \sum_{j=1}^{n} x_{ij}\hat{\alpha}_j \right) x_{ij} = 0, \qquad j = 1, .., n \quad (3)$$

The n equations in Eq. (3) are also known as the normal equations. Let us also write $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_n)^T$ and let $\hat{\boldsymbol{\alpha}}$ denote the minimizer of the sum of squares. Eq. (2) can be written $S(\boldsymbol{\alpha}) = \hat{\boldsymbol{r}}^T\hat{\boldsymbol{r}} = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\alpha}})^T(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\alpha}})$ and the normal equations (Eq. (3)) become

$$\hat{\boldsymbol{\alpha}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \quad (4)$$

Note that the last relation is valid only when $\mathbf{X}^T\mathbf{X}$ is invertible. The estimated coefficient $\hat{\boldsymbol{\alpha}}$ is a fixed linear combination of $\mathbf{Y}$, meaning that we get it by multiplying $\mathbf{Y}$ by the matrix $\mathbf{X}^+ = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ where the superscript $'+'$ stands for the Moore-Penrose inverse. It has to be noted that the LLS solution obtained directly from the normal equations is rather susceptible to round off error and, more importantly, to the singularity of these equations. Normal equations is the fastest method but at the same time numerically unstable.

Another issue of LLS is associated with its stability and accuracy. This problem is related to the conditions under which Eq. (2) is stable and accurate for a given sequence of linear spaces (V_n) with $dim(V_n) = P \leq N$, where P is the binomial coefficient (number of regression coefficients) and N is

the number of random sampling points [24]. It is well known that such approximations can be inaccurate when P is too close to N, even when the samples are noiseless. Regularization by taking P substantially smaller than N may therefore be needed even in a noise-free context. Hence, the concern is to use a value of P which provides the needed amount of regularization in order to ensure that the LLS method is stable and accurate.

4.3. Accuracy of prediction of Linear Least Squares

The fraction explained by the least squares is denoted by the quantity R^2 known as the coefficient of determination

$$R^2 = 1 - \frac{\|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\alpha}}\|_{L_2}^2}{\|\mathbf{Y} - \bar{\mathbf{Y}}\|_{L_2}^2} \tag{5}$$

where $\bar{y} = \frac{1}{m}\sum_{i=1}^{m} y_i$ is the mean of the observed data. Its non-negative square root R is called the coefficient of multiple correlation. It is one measure of how well the response $\mathbf{Y}$ correlates with the n predictors in $\mathbf{X}$ taken collectively. Among the several error measures that can be evaluated based on the residuals, the Root Mean Square value (RMS) is defined as

$$\text{RMS} = \sqrt{\frac{\|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\alpha}}\|_{L_2}^2}{m}} \tag{6}$$

where m is the number of sampling points. RMS provide a measure of the overall accuracy but can be biased as the residuals are not relatively measured.

4.4. Singular Value Decomposition (SVD) and Matrix Rank Reduction

The SVD can quantify the sensitivity of a linear system to numerical error or obtain a matrix inverse. Additionally, it provides solutions to least-squares problems and handles situations when matrices are either singular or numerically very close to singular.

Theorem 3.1. Let a real matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$. Then there exist orthogonal matrices $\mathbf{U} \in \mathbb{R}^{m\times m}$ and $\mathbf{V} \in \mathbb{R}^{n\times n}$ such that

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T \tag{7}$$

where

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{S} & 0 \\ 0 & 0 \end{bmatrix}$$

$$\mathbf{U} = [\mathbf{u}_1\ \mathbf{u}_2\ ...\ \mathbf{u}_r] \qquad \mathbf{V} = [\mathbf{v}_1\ \mathbf{v}_2\ ...\ \mathbf{v}_r] \tag{8}$$

and $\mathbf{S} = \text{diag}(\sigma_1, \sigma_2, ..., \sigma_r)$ with $\sigma_1 \geq ... \geq \sigma_r > 0$ is the diagonal matrix formed by the singular numbers. Vectors $\mathbf{v}_i$ are called right singular vectors of $\mathbf{A}$, vectors $\mathbf{u}_i$ are called left singular vectors of $\mathbf{A}$ while σ_i are called singular values of $\mathbf{A}$. One of the key applications of the SVD is the construction of low-rank approximations to a matrix. Eq. (7) represents a decomposition into rank-one matrices with the property that the k partial sum captures as much of the 'energy' of $\mathbf{A}$ as possible defined by either the L^2-norm or the Frobenius norm.

Theorem 3.2. For any k with $0 \leq k \leq r$, define

$$\mathbf{A}_k = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^T \tag{9}$$

Then

$$\|\mathbf{A} - \mathbf{A}_k\|_2 = \inf_{\substack{\mathbf{B} \in \mathbb{R}^{m \times n} \\ rank(\mathbf{B}) \leq k}} \|\mathbf{A} - \mathbf{B}\|_2 = \sigma_{k+1} \tag{10}$$

Additionally, $\mathbf{A}_k$ provides a unique solution to the problem in the Frobenius norm:

$$\|\mathbf{A} - \mathbf{A}_k\|_F = \inf_{\substack{\mathbf{B} \in \mathbb{R}^{m \times n} \\ rank(\mathbf{B}) \leq k}} \|\mathbf{A} - \mathbf{B}\|_F = \sqrt{\sigma_{k+1}^2 + ... + \sigma_r^2} \tag{11}$$

Thus, $\mathbf{A}_k$ constructed from the k largest singular values of $\mathbf{A}$ is the optimal rank-k approximation to $\mathbf{A}$ with respect to both $\|\cdot\|_F$ and $\|\cdot\|_2$. More generally, one can also show that $\|\mathbf{A}\|_2 = \sigma_1$ and that $\|\mathbf{A}\|_F^2 = \sum_{i=1}^{r} \sigma_i^2$. It should be emphasized that SVD is expensive to compute, but is numerically stable and can handle rank deficiency. Hence, SVD provides solution to least squares in cases of numerical instability. More details on the mathematics and computation of SVD can be found in [25]-[26].

5. Regularization methods/Filtering

Linear models often incorporate an increasing number of regression coefficients due to a loose knowledge for the model structure. As the basis set of regressors increases in size, the complexity of the surrogate model increases in a way to overfit the training data. This results with models that have a low bias due to the increased degrees of freedom in the regression model, but at the same time the variance may grow with no control. This means that the predicted values can be very sensitive with small changes in the training data set and the accuracy of the model on out of training data might be too small [27]. For deterministic computer simulations the primary interest is minimizing the bias error because the numerical noise is small. Regularization algorithms are used to balance the bias-variance trade-off and preventing overfitting in data-driven models. This is of fundamental importance across modeling techniques in order to produce models that will generalize to unseen data [28], [29]. Moreover, regularization can also stabilize a possibly ill-conditioned matrix inversion problem.

5.1. Truncated Singular Value Decomposition (TSVD)

A well-known method for dealing with ill-conditioned matrices in problems defined in (Eq.2) is the TSVD [30], [31]. The basic idea of TSVD is to define a new well-posed problem related to the ill-posed problem, with a solution which is less sensitive to perturbations than the coefficient vector $\boldsymbol{\alpha}$ defined in Eq. (2). This can be done by replacing the smaller singular values by zero [32],[33]. The use of the TSVD has certain similarities with the use of regularization in standard form, and it is generally known that the two methods often produce very similar results [34]. The SVD solution to Eq. (2) can be defined as:

$$\underset{\alpha_1,\ldots,\alpha_n\in\mathbb{R}}{\arg\min}\ S(\alpha_1,\ldots,\alpha_n) = \|\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T\boldsymbol{\alpha}-\mathbf{Y}\|_{L_2}^2 = \|\mathbf{U}^T(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T\boldsymbol{\alpha}-\mathbf{Y})\|_{L_2}^2 =$$

$$\|\boldsymbol{\Sigma}\mathbf{V}^T\boldsymbol{\alpha}-\mathbf{U}^T\mathbf{Y}\|_{L_2}^2 = \sum_{i=1}^{n}(\sigma_i\mathbf{v}_i^T\boldsymbol{\alpha}-\mathbf{u}_i^T\mathbf{Y})^2 + \sum_{i=n+1}^{m}(\mathbf{u}_i^T\mathbf{Y})^2 \qquad (12)$$

The optimal solution is given using the spectral presentation as

$$\boldsymbol{\alpha} = \sum_{i=1}^{n}\frac{\mathbf{u}_i^T\mathbf{Y}}{\sigma_i}\mathbf{v}_i, \qquad i=1,\ldots,n \qquad (13)$$

or in matrix form

$$\boldsymbol{\alpha} = \mathbf{X}^{+}\mathbf{Y} = \mathbf{V}\boldsymbol{\Sigma}^{+}\mathbf{U}^{T}\mathbf{Y}, \qquad \boldsymbol{\Sigma}^{+} = \mathrm{diag}(\sigma_1^{-1}, \ldots, \sigma_n^{-1}, 0, \ldots, 0) \tag{14}$$

Therefore

$$\underset{\alpha_1,\ldots,\alpha_n \in \mathbb{R}}{\arg\min} \|\mathbf{X}\boldsymbol{\alpha} - \mathbf{Y}\|_{L_2}^2 = \sum_{i=n+1}^{m} (\mathbf{u}_i^T \mathbf{Y})^2 \tag{15}$$

The TSVD solution is given by truncating the last $n-k$ singular values ($k < n$) in Eq.13. The underlying problem of finding the effective rank of a system arises frequently, as discussed in [35].

5.2. Tikhonov regularization

A classical regularization in standard form is the method of Philips [36] and Tikhonov [37] for solving the minimization problem

$$\boldsymbol{\alpha}_\lambda = \underset{\alpha_1,\ldots,\alpha_n \in \mathbb{R}}{\arg\min} \left\{ \|\mathbf{Y} - \mathbf{X}\boldsymbol{\alpha}\|_{L_2}^2 + \lambda^2 \|\mathbf{L}(\boldsymbol{\alpha} - \boldsymbol{\alpha_0})\|_{L_2}^2 \right\} \tag{16}$$

where λ is a regularization parameter that must be chosen, $\mathbf{L}$ is a matrix that in a standard form is $\mathbf{L} = \mathbf{I}$, and vector $\boldsymbol{\alpha_0}$ is an a priori estimate of $\boldsymbol{\alpha}$ which is set to zero if no information is available. It can be shown that $\boldsymbol{\alpha}_\lambda$ is the least squares solution to the problem

$$\boldsymbol{\alpha}_\lambda = \underset{\alpha_1,\ldots,\alpha_n \in \mathbb{R}}{\arg\min} \left\| \begin{bmatrix} \mathbf{X} \\ \lambda\mathbf{I} \end{bmatrix} \boldsymbol{\alpha} - \begin{bmatrix} \mathbf{Y} \\ \lambda\mathbf{L}\boldsymbol{\alpha_0} \end{bmatrix} \right\|_{L_2}^2 \tag{17}$$

The graphical tool for displaying the norm of a regularized solution versus the residual norm is the L-curve which gives insight into the regularization properties of the underlying method in a log-log scale is shown in Fig.10. It also aids to choose an appropriate regularization parameter λ for the given training data.

By using the SVD to solve the least squares formulation defined in Eq. (17) it can be shown that the Tikhonov solution is

$$\boldsymbol{\alpha}_\lambda = \sum_{i=1}^{n} f_i \frac{\mathbf{u}_i^T \mathbf{Y}}{\sigma_i} \mathbf{v}_i \tag{18}$$

where the Tikhonov filter factors $f_1, \ldots, f_n$ have the form

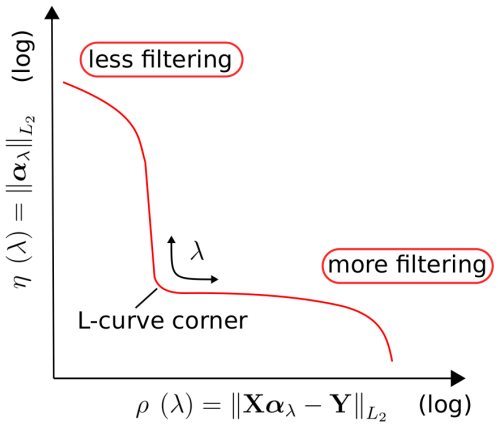

Figure 10: The generic form of the L-curve.

$$f_i = \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2} \approx \begin{cases} 1, & \sigma_i \gg \lambda \\ \sigma_i^2/\lambda^2, & \sigma_i \ll \lambda \end{cases} \tag{19}$$

corresponding to a smooth filter that dampens out the small singular values, while leaving almost unaffected the SVD components associated to large singular values. For the TSVD, the filter factor has the form

$$f_i = \begin{cases} 1, & \sigma_i \geq \sigma_k \\ 0, & \sigma_i < \sigma_k \end{cases} \tag{20}$$

which is similar to a sharp filter that simply cuts off the last $n - k$ components. When k is chosen such that $\sigma_k = \lambda$, the sharp filter of the TSVD can be seen as an approximation to the smooth filter of the regularization. The residual vector corresponding to $\boldsymbol{\alpha_\lambda}$ is given in terms of the SVD by

$$\mathbf{r} = \mathbf{X}\boldsymbol{\alpha}_\lambda - \mathbf{Y} = \sum_{i=1}^{n}(1 - f_i)\mathbf{u}_i^T\mathbf{Y}\mathbf{u} + \mathbf{r}_\perp \tag{21}$$

where $\mathbf{r}_\perp = \mathbf{Y} - \sum_{i=1}^{n} \mathbf{u}_i^T\mathbf{Y}\mathbf{u}$ is least squares residual, i.e., the component of $\mathbf{Y}$ that lies outside the range of the column space of $\mathbf{X}$. From Eq. (18) and Eq. (21) the solution and residual norms in terms of SVD can be written as

$$\eta^2(\lambda) = \|\boldsymbol{\alpha}_\lambda\|_{L_2}^2 = \sum_{i=1}^{n}\left(f_i\frac{\mathbf{u}_i^T\mathbf{Y}}{\sigma_i}\right)^2 \tag{22}$$

$$\rho^2(\lambda) = \|\mathbf{X}\boldsymbol{\alpha}_\lambda - \mathbf{Y}\|_{L_2}^2 = \sum_{i=1}^{n} \left((1 - f_i)\mathbf{u}_i^T\mathbf{Y}\right)^2 + \mathbf{r}_\perp^2 \tag{23}$$

These two norms are the basis of the analysis using the L-curve. For example, Fig.12 shows the Tikhonov regularization for the problem addressed in this work, in which the appearance of L-curve is distinct. One can observe that if too much regularization is imposed on the solution (large values of the parameter λ) then it will not fit the data $\mathbf{Y}$ properly and the residual $\rho(\lambda)$ will be large. This trend corresponds to the right flat part of the L-curve. On the other hand, if too little regularization is imposed (small values of the parameter λ) then the fit is good but the solution is dominated by the superfluous regressors (observe the left vertical part of the L-curve).

6. Multi-Stage Linear Least Squares and Subspace Approximation

Here the basic concepts of the Multi-Stage Least Squares (MSLS) algorithm are presented. The central idea is to separate the design variables into sets and yield locally optimal solutions in each subspace using polynomial regression. Then, the method seeks to obtain a globally optimal solution using the composition of the local stage solutions. An important aspect of the method is its flexibility of using different polynomial order approximations in each subspace. Particularly, the method begins by defining a LLS problem in the first subspace and then uses the optimal regression coefficients as an input to solve a LLS problem in the second subspace. The same procedure is followed sequentially for all subspaces and finally a global minimum is found. Besides, due to subspace variable separation, regularization is more efficiently tuned in the local spaces as the number of superfluous regressors decrease. Note also that during MSLS, an additional smoothing of the initial local space approximation functions is performed through the sequential subspace LLS minimization problems. As discussed in the previous section, the stability and accuracy of MSLS related to the number of regression coefficients becomes also important due to the separation of the design variables. The formulation of MSLS for a 3-stage problem is described in the following.

6.1. Selection of Subsets of Regression Variables

Let us assume the vectors $\mathbf{x}^{\mathbf{L}}, \mathbf{x}^{\mathbf{\Theta}}, \mathbf{x}^{\mathbf{G}}$ corresponding to different sets of variables as follows

$$\mathbf{x} = \{\underbrace{x_1\ x_2\ \cdots\ x_p}_{\mathbf{x^L}}\ \underbrace{x_{p+1}\ x_{p+2}\cdots x_{p+q}}_{\mathbf{x^\Theta}}\ \underbrace{x_{p+q+1}\ x_{p+q+1}\cdots x_{p+q+r}}_{\mathbf{x^P}}\}^T$$

$$\mathbf{x} = \{\mathbf{x^L} \quad \mathbf{x^\Theta} \quad \mathbf{x^P}\}^T \tag{24}$$

where $\mathbf{x^L} \in \mathbb{R}^p$ is associated to the loading variables, $\mathbf{x^\Theta} \in \mathbb{R}^q$ is associated to the material orientation variables, $\mathbf{x^G} \in \mathbb{R}^r$ is associated to the geometric variables and $p + q + r = n$ in a way that the following LLS problems can be defined in different subspaces.

6.2. First-Stage Least Squares: Loading variables

In the subspace spanned by the variables $\mathbf{x^L}$ one needs to solve $(N_\Theta \times N_G)$ minimization problems for every sampling set $\{\mathbf{x^{\Theta}}_i, \mathbf{x^{G}}_j\}$ in the design space under consideration

$$\underset{\alpha_1,\ldots,\alpha_{P_1}\in\mathbb{R}}{\arg\min}\ S_1(\alpha_1, ..., \alpha_{P_1}) = \|\mathbf{Y} - \mathbf{L}(\mathbf{x^L})\boldsymbol{\alpha}(\mathbf{x^{\Theta}}_i, \mathbf{x^{G}}_j)\|^2_{L_2} \tag{25}$$

and estimate the regression coefficient vectors $\hat{\boldsymbol{\alpha}}(\mathbf{x^{\Theta}}_i, \mathbf{x^{G}}_j)$, $i = 1, 2, \ldots, N_\Theta$, $j = 1, 2, \ldots, N_G$. The design matrix $\mathbf{L}(\mathbf{x^L}) \in \mathbb{R}^{n_s \times P_1}$ can be presented as

$$\mathbf{L}(\mathbf{x^L}) = \begin{bmatrix} (x_1^{(1)})^d & (x_2^{(1)})^d & \cdots & (x_p^{(1)})^d & \cdots & (x_p^{(1)})^1 & 1 \\ (x_1^{(2)})^d & (x_2^{(2)})^d & \cdots & (x_p^{(2)})^d & \cdots & (x_p^{(2)})^1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ (x_1^{(n_s)})^d & (x_2^{(n_s)})^d & \cdots & (x_p^{(n_s)})^d & \cdots & (x_p^{(n_s)})^1 & 1 \end{bmatrix} \tag{26}$$

where d is the polynomial regression order, p is the number of variables in the first subspace, and n_s the number of sampling points. Notice that for each of these problems, a different matrix $\mathbf{L}(\mathbf{x^L})$ is calculated according to a polynomial regression of order d, and the distinction between them is taken into account by the regression coefficient vector $\hat{\boldsymbol{\alpha}}(\mathbf{x^{\Theta}}_i, \mathbf{x^{G}}_j)$.

6.3. Second-Stage Least Squares: Laminate angles

In the subspace spanned by the variables $\mathbf{x^\Theta}$ one needs to solve $(P_1 \times N_G)$ minimization problems for every sampling set $\{\mathbf{x^{G}}_j\}$ in the design space under consideration

$$\underset{\beta_1,...,\beta_{P_2}\in\mathbb{R}}{\arg\min}\ S_2(\beta_1,...,\beta_{P_2}) = \|\hat{\boldsymbol{\alpha}}_{n_1}^{\boldsymbol{\Theta},\mathbf{G}_j} - \boldsymbol{\Theta}(\mathbf{x}^{\boldsymbol{\Theta}})\boldsymbol{\beta}_{n1}(\mathbf{x}^{\mathbf{G}_j})\|_{L_2}^2 \tag{27}$$

with $\hat{\boldsymbol{\alpha}}_{n_1}^{\boldsymbol{\Theta},\mathbf{G}_j} = \{\hat{\alpha}_{n_1}^{\boldsymbol{\Theta}_1,\mathbf{G}_j} \quad \hat{\alpha}_{n_1}^{\boldsymbol{\Theta}_2,\mathbf{G}_j} \quad \cdots \quad \hat{\alpha}_{n_1}^{\boldsymbol{\Theta}_k,\mathbf{G}_j}\}^T$, $n_1 = 1, 2, \ldots, P_1$, $j = 1, 2, \ldots, N_G$, $\boldsymbol{\Theta}(\mathbf{x}^{\boldsymbol{\Theta}}) \in \mathbb{R}^{N_\Theta \times P_2}$ and estimate the regression coefficient vectors $\hat{\boldsymbol{\beta}}_{n_1}(\mathbf{x}^{\mathbf{G}_j})$.

6.4. Third-Stage Least Squares: Ply Thickness

In the subspace spanned by the variables $\mathbf{x}^{\mathbf{G}}$ one needs to solve $(P_1 \times P_2)$ minimization problems in the form

$$\underset{\gamma_1,...,\gamma_{P_3}\in\mathbb{R}}{\arg\min}\ S_3(\gamma_1,...,\gamma_{P_3}) = \|\hat{\boldsymbol{\beta}}_{n_1,n_2}^{\mathbf{G}} - \mathbf{G}(\mathbf{x}^{\mathbf{G}})\boldsymbol{\gamma}_{n_1,n_2}\|_{L_2}^2 \tag{28}$$

with $\hat{\boldsymbol{\beta}}_{n_1,n_2}^{\mathbf{G}} = \{\hat{\beta}_{n_1,n_2}^{\mathbf{G}_1} \quad \hat{\beta}_{n_1,n_2}^{\mathbf{G}_2} \quad \cdots \quad \hat{\beta}_{n_1,n_2}^{\mathbf{G}_m}\}^T$, $n_1 = 1, 2, \ldots, P_1$, $n_2 = 1, 2, \ldots, P_2$, $\mathbf{G}(\mathbf{x}^{\mathbf{G}}) \in \mathbb{R}^{N_G \times P_3}$ and finally estimate the regression coefficient vectors $\hat{\boldsymbol{\gamma}}_{n1,n2}$.

6.5. Reconstruction of the analytical function

The construction of the analytical function after the MSLS approximations is obtained using the composition of local stage solutions. This can be written as

$$\hat{y}(\mathbf{x}^{\mathbf{L}}, \mathbf{x}^{\boldsymbol{\Theta}}, \mathbf{x}^{\mathbf{G}}) = \sum_i^{P_1} \hat{\alpha}_i(\mathbf{x}^{\boldsymbol{\Theta}}, \mathbf{x}^{\mathbf{G}}) f_i(\mathbf{x}^{\mathbf{L}}) = \sum_i^{P_1} \left[\sum_j^{P_2} \hat{\beta}_j^i(\mathbf{x}^{\mathbf{G}}) h_j(\mathbf{x}^{\boldsymbol{\Theta}}) \right] f_i(\mathbf{x}^{\mathbf{L}}) =$$

$$\sum_i^{P_1} \left[\sum_j^{P_2} \left[\sum_k^{P_3} \hat{\gamma}_k^{i,j} q_k(\mathbf{x}^{\mathbf{G}}) \right] h_j(\mathbf{x}^{\boldsymbol{\Theta}}) \right] f_i(\mathbf{x}^{\mathbf{L}}) \tag{29}$$

where $\hat{\beta}_j^i$: j regression coefficient estimation of $\hat{\alpha}_i$ and $\hat{\gamma}_k^{i,j}$ is the k regression coefficient estimation of $\hat{\beta}_j^i$, $i = 1, ..., P_1$, $j = 1, ..., P_2$, $k = 1, ..., P_3$.

7. Implementation and computational results

In this section a wide set of numerical investigations is performed to shed light on the approximation properties of the LLS and MSLS using the training points. In order to gain further insight into the nonlinear structure of the 9D manifold, the $\mathbb{R}^3$ space is considered where the estimated DSL functions are plotted using two loading variables of the set $\mathbf{x^L}$ whereas the other variable is set to a constant value. The graphs are obtained for several training data sets for the variables $\mathbf{\Theta}_i(\mathbf{x^\Theta}); \mathbf{x^\Theta} = \{x_4, x_5, x_6\}$ and $\mathbf{G}_j(\mathbf{x^G}); \mathbf{x^G} = \{x_7, x_8, x_9\}$, corresponding to the material ply orientation angles and ply laminate thickness, respectively.

7.1. Linear Least Squares

When solving a linear regression problem, it is very unusual to know the model structure a priori. This means that one seldom knows the order of the model and the number of regression coefficients which are of importance. In this work, a polynomial regression of several variables is presented in the form

$$y(x_1, x_2, \ldots, x_n) = \sum_{k_1,k_2,\ldots,k_n=0}^{d} \alpha_{k_1,k_2,\ldots,k_n} x_1^{k1} x_2^{k2} \cdots x_n^{kn} \tag{30}$$

with $0 \leq \sum_{i=1}^{n} k_i \leq d$, where d is the polynomial order and n denotes the number of variables. The LLS problem is solved with an increasing complexity for several values of d. Thus, for example, the LLS approximation using a quadratic ($d = 2, n = 9$) and a fourth-order equation ($d = 4, n = 9$) requires the estimation of $\binom{n+d}{n}$=55 and 715 regression coefficients, respectively.

Since the DSL functions must be zero at zero values of the loading variables $\mathbf{x^L}$, a constraint LLS (see ref. [32]) should be implemented. To avoid this expensive computational cost, a selection of regressors is proposed. Using the variable separation defined in Eq. (24), between all polynomial terms determined by the binomial coefficient P, only the associated regressor coefficients $\alpha_{k_1,k_2,\cdots,k_n}$ subjected to the following constraint are chosen

$$P = \left\{ \binom{n+d}{n} \quad \middle| \quad \alpha_{k_1,k_2,\ldots,k_n} (\mathbf{x^L})^s (\mathbf{x^\Theta})^l (\mathbf{x^P})^m \quad \text{s.t.} \quad s{+}l{+}m \leq d, \quad s \geq 1 \right\} \tag{31}$$

In addition, the constant term is omitted to prevent a possible shifting of the estimated function. This selection finally leads to 27 and 505 regression coefficients for a quadratic and a fourth-order polynomial, respectively.

First, the regularization of the response function $\mathrm{DSL}_{\mathrm{UR3}}$ (LPA in UR3 direction) on the training data using a fourth-order polynomial regression is presented. The eigenvalue spectrum is shown in a semilog scale in Fig. 11. The condition number of the matrix is about $5.81e+05$. Throughout this section the notation (η, ρ) is used for the chosen measures of the size of the residual vector and the size of the solution vector, respectively (see Eq. (22) and Eq. (23)). The L-curve associated with Tikhonov regularization is introduced by plotting the norm of the solution η vs. the norm of the residual ρ in a log-log scale (Fig. 12). It consists of a number of discrete points corresponding to different values of the regularization parameter λ in a range of $[0.01 - 2000]$ at which η and ρ are evaluated. One can observe that the L-shaped appearance is very distinct. However, the rounded corner on the L-curve shows the need for a rigorous definition of the 'corner' of the L-curve. In order to locate the L-curve 'corner', the Regularization Toolbox in Matlab is used [38],[39],[40] and the optimal value is evaluated at $\lambda = 1.037$ corresponding to norms $\eta = 8.8$ and $\rho = 30.68$. This value of the regularization parameter indeed produces a regularized solution with optimal error. Additionally, the TSVD regularization L-curve is illustrated in Fig. 13. Norms η and ρ are evaluated for each rank-k approximation of the matrix $\mathbf{X}$. Using the adaptive pruning algorithm [38], the L-curve 'corner' is evaluated at the singular value $\sigma_k = 320$ corresponding to norms $\eta = 6.796$ and $\rho = 33.33$. One can observe that Tikhonov regularization and TSVD provide very similar results, as discussed in [34].

Next, the LLS approximation is assessed. To this effect, several combinations between the sampling point sets of the design variables in $\mathbb{R}^9$ are used. First, the dependency of the ply drop geometry for a given configuration of material orientation angles is investigated. Fig. 14 and Fig. 15 show a surface plot of the LLS approximation using a 4-order polynomial regression for the $\mathrm{DSL}_{\mathrm{U1}}$ function (LPA in U1 direction) corresponding to the material orientation sampling point $\mathbf{\Theta}_1(\theta_1 = 40.48°, \theta_2 = 53.26°, \theta_3 = 25.70°)$ for two different geometric sampling points $\mathbf{G_3}(t_1 = 0.85, t_2 = 1.30, t_3 = 1.39)$ and $\mathbf{G_6}(t_1 = 2.52, t_2 = 0.84, t_3 = 0.53)$, respectively. Note that these points are coming from arbitrary sampling of the design space and do not correspond to real manufacturing values. Plots of $\mathrm{DSL}_{\mathrm{U1}}$ are given as a function of UX and URZ at constant UY=0.

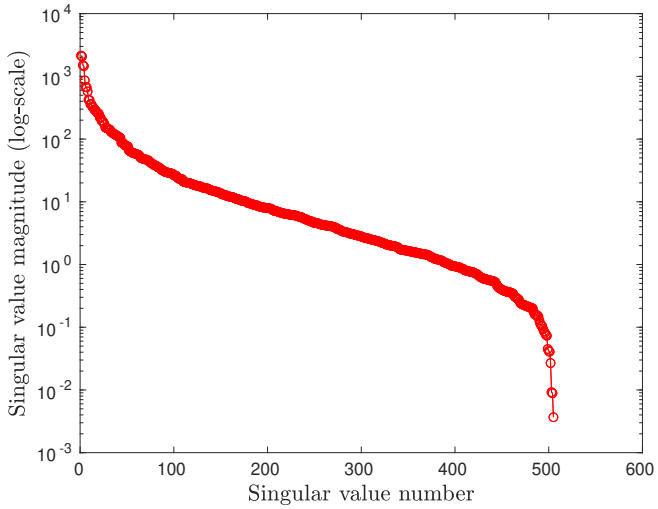

Figure 11: Singular values magnitude.

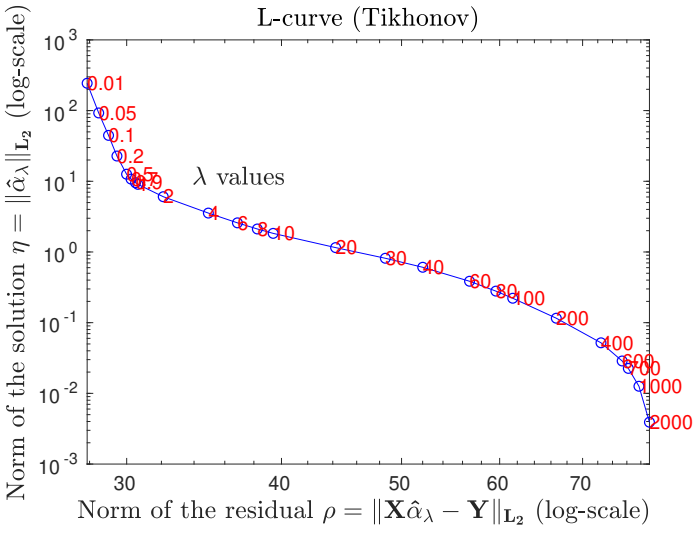

Figure 12: The Tikhonov L-curve for the model problem.

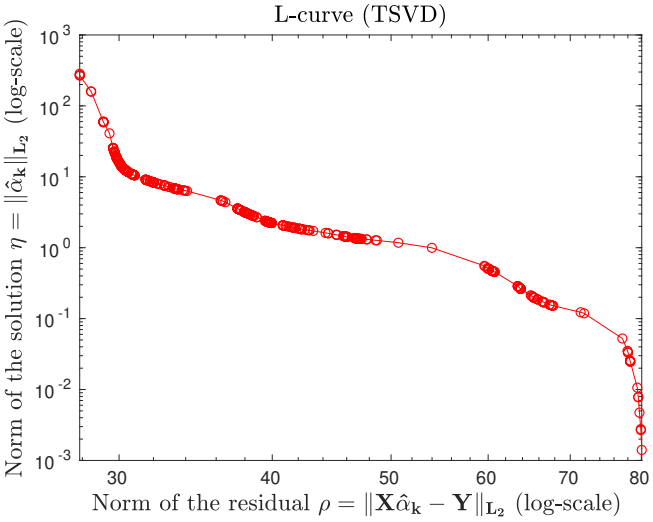

Figure 13: The TSDV L-curve for the model problem.

It is also interesting to visualize the dependency of the ply drop material orientations for a specific geometry. To this end, Fig. 16 presents selected contour plots of DSL_{U1} for the sampling points $\mathbf{\Theta_1}(\theta_1 = 40.48°, \theta_2 = 53.26°, \theta_3 = 25.70°)$, $\mathbf{\Theta_{17}}(\theta_1 = 7.22°, \theta_2 = 6.18°, \theta_3 = 57.37°)$, $\mathbf{\Theta_{32}}(\theta_1 = 59.03°, \theta_2 = 78.47°, \theta_3 = 22.6°)$, $\mathbf{\Theta_{40}}(\theta_1 = 8.46°, \theta_2 = 40.13°, \theta_3 = 70.94°)$, considering the geometric sampling point $\mathbf{G_1}(t_1 = 1.61, t_2 = 1.17, t_3 = 1.19)$. A variation of DSL_{U1} due to different material orientation angles can be clearly observed. Again, plots of DSL_{U1} are given as a function of UX and URZ at constant UY=0.

The quality of polynomial regression with increasing complexity is evaluated by comparing the coefficient of determination R^2 and RMS error. As an example, the DSL_{U1} function is used and no regularization method is applied. To illustrate the approximation accuracy, Fig. 17 and Fig. 18 show the coefficient of determination R^2 and the RMS value of the second and fourth-order polynomial regression, for each configuration $\mathbf{\Theta}_i = \mathbf{\Theta}_i(\theta_1, \theta_2, \theta_3)$, $i = 1, 2, ..., N_\Theta$, where N_Θ is the number of different composite layups, considering the sampling point $\mathbf{G_1}(t_1 = 1.61, t_2 = 1.17, t_3 = 1.19)$. Note that R^2 will in general increase with an increase in the number of regression coefficients. The mean values of R^2 are 0.654 and 0.918 for the second and fourth-order model, respectively. The mean values of RMS are 0.012 and 0.006 for the second and fourth-order model, respectively. As evidenced by

the high R^2 values close to unity and the small RMS values, the 4-order model appears to capture a large portion of the observed variance. As a general conclusion, even the 4-order polynomial approximation is not capable to accurately capture the high non-linearity of the data set. Higher order polynomial approximations have to be used to increase the prediction accuracy, however, an exhaustive investigation of this aspect is out of the scope of the present paper.

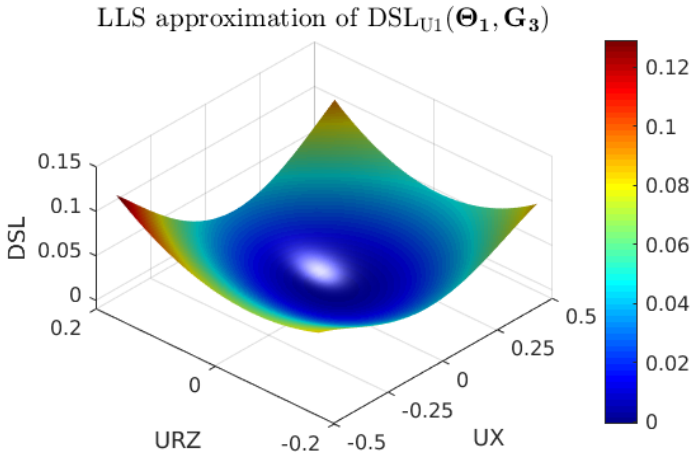

Figure 14: LLS approximation using a 4-order polynomial regression; $\mathrm{DSL}_{\mathrm{U1}}$ $(\mathbf{\Theta_1}, \mathbf{G_3}; \mathbf{\Theta}_1(\theta_1 = 40.48°, \theta_2 = 53.26°, \theta_3 = 25.70°), \mathbf{G_3}(t_1 = 0.85, t_2 = 1.30, t_3 = 1.39))$.

7.2. Multi-Stage Least Squares

Here the results of applying the MSLS method are presented. For the First-Stage approximation, let us consider the subspace spanned by the loading variables $\mathbf{x^L}$. In this space, one needs to solve $(N_\Theta = 40 \times N_G = 20)$ minimization problems for every sampling set $\{\mathbf{x}^{\mathbf{\Theta}_i}, \mathbf{x}^{\mathbf{G}_j}\}$ (eq. (25)). Using a 4-order polynomial equation $(d = 4, n = 3)$, results to the estimation of $P_1 = \binom{n+d}{n} = 35$ regression coefficients for the vector $\hat{\boldsymbol{\alpha}}(\mathbf{x}^{\mathbf{\Theta}_i}, \mathbf{x}^{\mathbf{G}_j})$. Note that the constant term is dropped to prevent curve shifting at zero loading.

In the following, numerical results are given for the response function $\mathrm{DSL}_{\mathrm{UR3}}$ (LPA in UR3 direction). The eigenvalue spectrum of $\mathbf{L}(\mathbf{x^L})$ corresponding to selected sets of variables $\{\mathbf{x}^{\mathbf{\Theta}_i}, \mathbf{x}^{\mathbf{G}_j}\}$ $(i = 1,\ j = 3, 10, 14, 19)$ is shown in a semilog scale in Fig. 19. The first regularization method is TSVD where one simply truncates the summation in Eq. (18) at an upper

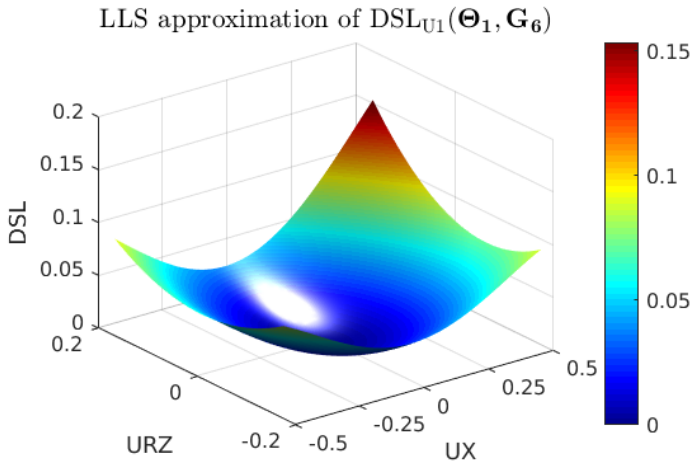

Figure 15: LLS approximation using a 4-order polynomial regression; DSL$_{U1}$ $(\mathbf{\Theta_1}, \mathbf{G_6}; \mathbf{\Theta}_1(\theta_1 = 40.48°, \theta_2 = 53.26°, \theta_3 = 25.70°), \mathbf{G_6}(t_1 = 2.52, t_2 = 0.84, t_3 = 0.53))$.

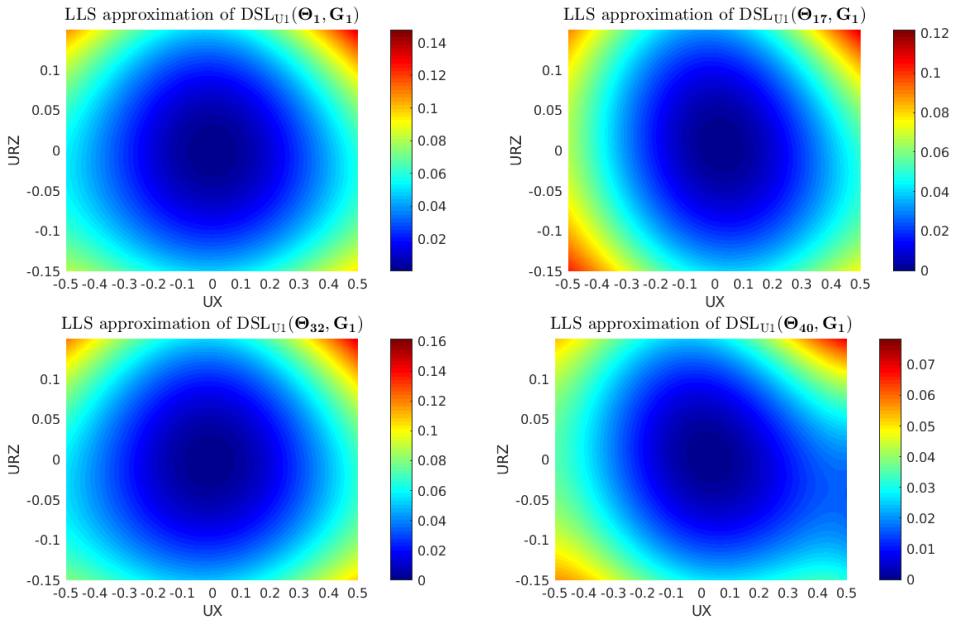

Figure 16: Contour plots of DSL$_{U1}$ for the sampling points $\mathbf{\Theta_1}, \mathbf{\Theta_{17}}, \mathbf{\Theta_{32}}, \mathbf{\Theta_{40}}$, for the geometrical sampling point $\mathbf{G_1}(t_1 = 1.61, t_2 = 1.17, t_3 = 1.19)$.

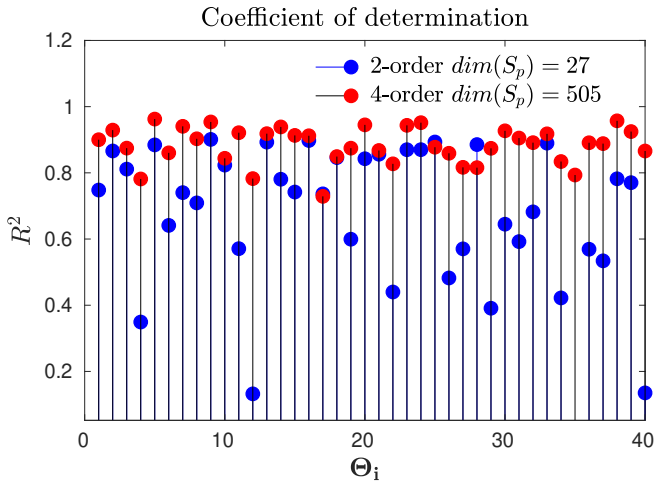

Figure 17: Coefficient of determination R^2 of the DSL_{U1} corresponding to a quadratic and a 4-order polynomial regression for each configuration $\mathbf{\Theta}_i = \mathbf{\Theta}_i(\theta_1, \theta_2, \theta_3)$, $i = 1, 2, ...k$, where k is the number of different composite layups considering the sampling point $\mathbf{G_1}(t_1 = 1.61, t_2 = 1.17, t_3 = 1.19)$. No regularization is applied.

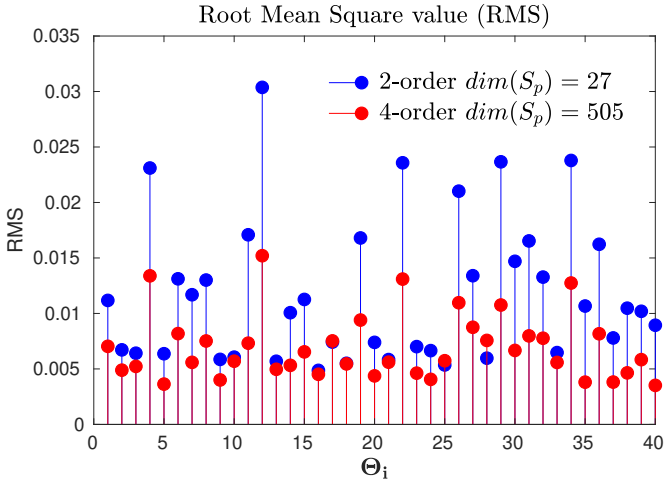

Figure 18: RMS value of the DSL_{U1} corresponding to a quadratic and a 4-order polynomial regression for each configuration $\mathbf{\Theta}_i = \mathbf{\Theta}_i(\theta_1, \theta_2, \theta_3)$, $i = 1, 2, ...k$, where k is the number of different composite layups considering the sampling point $\mathbf{G_1}(t_1 = 1.61, t_2 = 1.17, t_3 = 1.19)$. No regularization is applied.

limit $k < n$ using the filter defined in Eq. (20), before the small singular values start to dominate. The TSVD regularization L-curve in Fig. 20 shows a parametric plot of (η, ρ), where norms η and ρ measure the size of the regularized solution and the corresponding residual.

Additionally, the L-curves associated with Tikhonov regularization for the same selected sets of variables are shown in Fig. 21. For locating the L-curve corner the two-stage approach algorithm is applied using the Regularization Toolbox in Matlab [38]. First, a list of corner candidates is found by successively pruning the points on the discrete L-curve, and then the 'best' corner is found from this list.

At the Second-Stage approximation the subspace spanned by the variables $\mathbf{x}^{\mathbf{\Theta}}$ is considered. In this space, one needs to solve $(P_1 = 34 \times N_G = 20)$ minimization problems for every sampling set $\{\mathbf{x}^{\mathbf{G}_j}\}$ (Eq. (27)). Using as response the estimated regression coefficient vectors $\hat{\boldsymbol{\alpha}}_{n1}^{\mathbf{\Theta},\mathbf{G}_j}$ (First-Stage), the dependence on variables $\mathbf{x}^{\mathbf{\Theta}}$ is taken into account by the estimation of $P_2 = \binom{n+d}{n}$ regression coefficients for the vector $\hat{\boldsymbol{\beta}}_{n1}(\mathbf{x}^{\mathbf{G}_j})$. In order to ensure stability and accuracy a cubic $(d = 3, n = 3)$ polynomial regression is used resulting to the estimation of $P_2 = 20$ coefficients. It is worthy to mention that using a 4-order polynomial regression at this stage did not improve the results and the MSLS failed to compute a reasonable stable solution. This can be attributed to the case of a linear space (V_n) with $dim(V_n) = P_2 = 35 \approx N_\Theta$, thus violating the stability condition $P_2 \ll N_\Theta$. The eigenvalue spectrum of matrix $\mathbf{\Theta}(\mathbf{x}^{\mathbf{\Theta}})$ is shown in a semilog scale in Fig. 22. As an example, the TSVD regularization L-curve for the regression coefficient vector $\hat{\boldsymbol{\alpha}}_{32}^{\mathbf{\Theta},\mathbf{G}_1}$ is shown in Fig. 23. This is a typical behavior of an ill-determined numerical rank where there is obviously no intuitive way of choosing a suitable regularization parameter. In this case, the Matlab Regularization toolbox determines the leftmost point of the curve, i.e., the point with the highest norm η and the lowest norm ρ.

Finally, at the Third-Stage approximation, only the subspace spanned by the geometric variables $\mathbf{x}^{\mathbf{G}}$ is considered. In this space, one needs to solve $(P_1 = 34) \times (P_2 = 20) = 680$ minimization problems (see Eq. (28)). Using as a response the estimated regression coefficient vectors $\hat{\boldsymbol{\beta}}_{n1,n2}^{\mathbf{G}} = \{\hat{\beta}_{n1,n2}^{\mathbf{G}_1} \quad \hat{\beta}_{n1,n2}^{\mathbf{G}_2} \quad \cdots \quad \hat{\beta}_{n1,n2}^{\mathbf{G}_m}\}^T$, $n_1 = 1, 2, \ldots, P_1$, $n_2 = 1, 2, \ldots, P_2$ (Second-Stage), the dependence on variables $\mathbf{x}^{\mathbf{G}}$ is taken into account by a quadratic $(d = 2, n = 3)$ polynomial equation resulting to the estimation of $P_3 = \binom{n+d}{n} =$ 10 coefficients for the vector $\hat{\boldsymbol{\gamma}}_{n1,n2}$. The eigenvalue spectrum of matrix

$\mathbf{G}(\mathbf{x}^{\mathbf{G}})$ is shown in a semilog scale in Fig. 24. The TSVD regularization L-curve for a selected regression coefficient vector $\hat{\boldsymbol{\gamma}}_{n_1,n_2}$ is shown in Fig. 25.

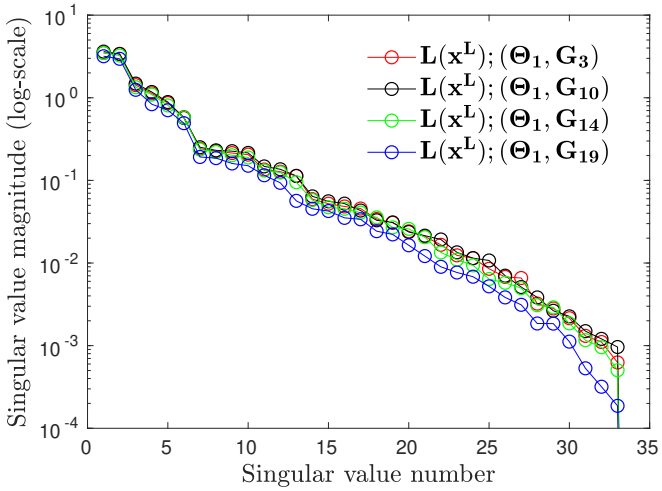

Figure 19: The eigenvalue spectrum of $\mathbf{L}(\mathbf{x}^{\mathbf{L}})$ corresponding to selected sets of variables $\{\mathbf{x}^{\boldsymbol{\Theta}_i},\mathbf{x}^{\mathbf{G}_j}\}$ $(i=1,\ j=3,10,14,19)$; First-Stage approximation.

The coefficient of determination R^2 of the $\mathrm{DSL}_{\mathrm{U1}}$ and $\mathrm{DSL}_{\mathrm{U2}}$ for each configuration $\boldsymbol{\Theta}_i=\boldsymbol{\Theta}_i(\theta_1,\theta_2,\theta_3)$, $i=1,2,...,N_\Theta$, where $N_\Theta=40$ is the number of different composite layups considering the geometric sampling point $\mathbf{G_1}(t_1=1.61,t_2=1.17,t_3=1.19)$ is presented in Fig. 26 and Fig. 27, respectively. High values of R^2 (close to 1.0) indicating an accurate approximation. As expected, one can observe a general decrease of R^2 due to the multi-stage approximation of the analytical function. This indicates an indirect smoothing of the local approximation functions due to multi-stage subspace evaluation. However, there are also a number of exceptions where the coefficient of determination R^2 has been increased (see Fig. 27).

7.3. Computational Complexity of LLS and MSLS

To facilitate a comparison of the computational complexity of LLS and MSLS methods, a preliminary analysis is made considering the LLS solution using the normal equations. In this example, regularization is not taken into account and the 9 design variables are expanded at different polynomial order approximations d. For the LLS, Eq. (32) provides the necessary arithmetic operations (Appendix A). In MSLS, since the design variables are separated

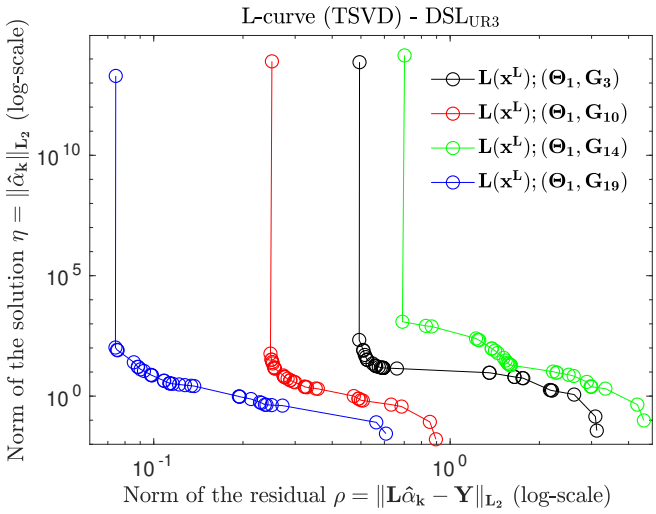

Figure 20: The TSVD regularization L-curve corresponding to selected sets of variables $\{\mathbf{x}^{\boldsymbol{\Theta}_i}, \mathbf{x}^{\mathbf{G}_j}\}$ $(i = 1,\ j = 3, 10, 14, 19)$; First-Stage approximation.

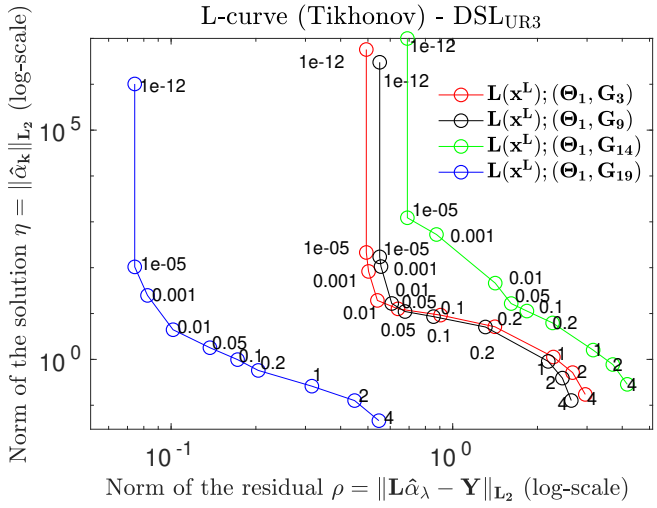

Figure 21: The L-curves associated with Tikhonov regularization corresponding to selected sets of variables $\{\mathbf{x}^{\boldsymbol{\Theta}_i}, \mathbf{x}^{\mathbf{G}_j}\}$ $(i = 1,\ j = 3, 10, 14, 19)$; First-Stage approximation.

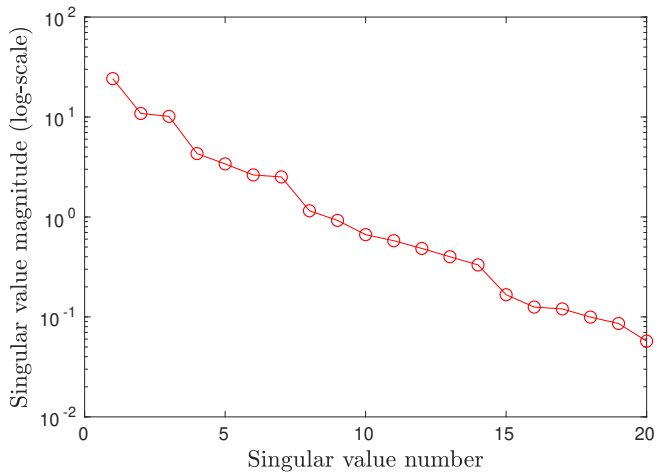

Figure 22: The eigenvalue spectrum of matrix $\boldsymbol{\Theta}(\mathbf{x}^{\boldsymbol{\Theta}})$; Second-Stage approximation.

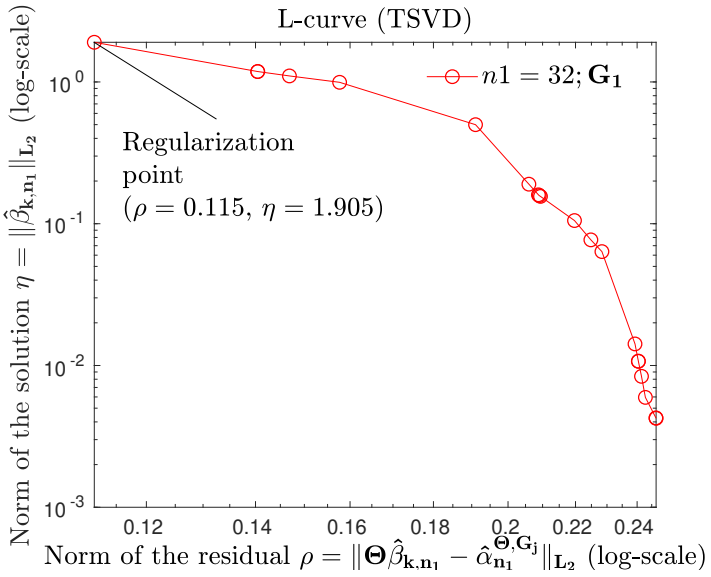

Figure 23: The TSVD regularization L-curve considering eq. (27) corresponding to regression coefficient vector $\hat{\boldsymbol{\alpha}}_{n_1}^{\boldsymbol{\Theta},\mathbf{G}_j}$ ($n_1 = 32$, $j = 1$); Second-Stage approximation.

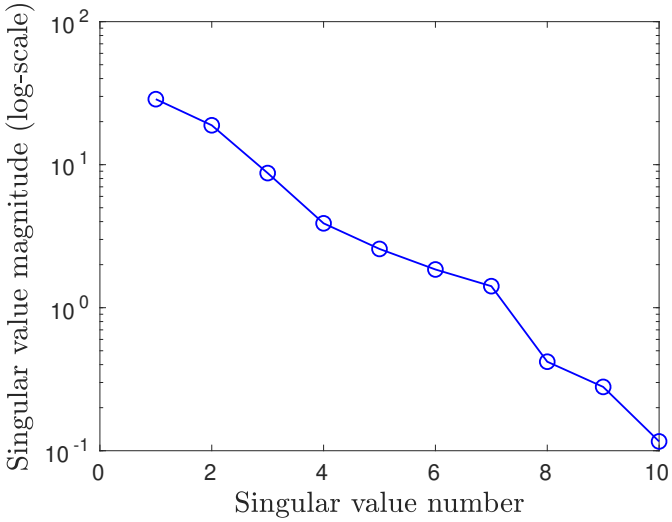

Figure 24: The eigenvalue spectrum of $\mathbf{G}(\mathbf{x}^{\mathbf{G}})$; Third-Stage approximation.

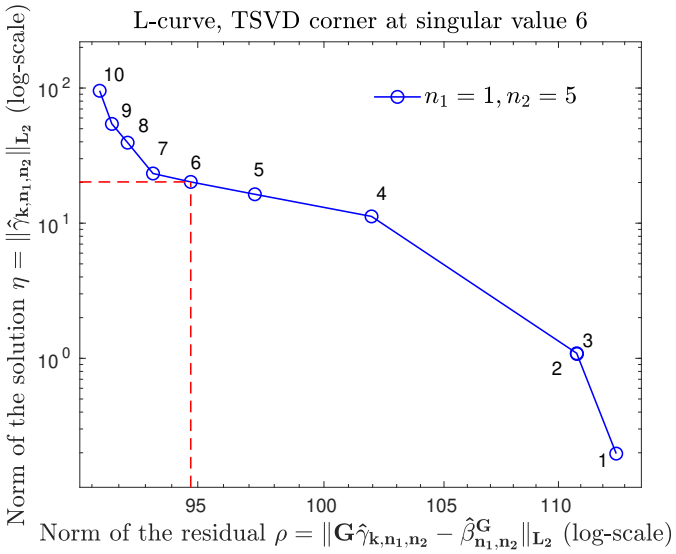

Figure 25: The TSVD regularization L-curve considering eq. (28) corresponding to regression coefficient vector $\hat{\boldsymbol{\beta}}^{\mathbf{G}}_{n1,n2}$ ($n_1 = 1$, $n_2 = 5$); Third-Stage approximation.

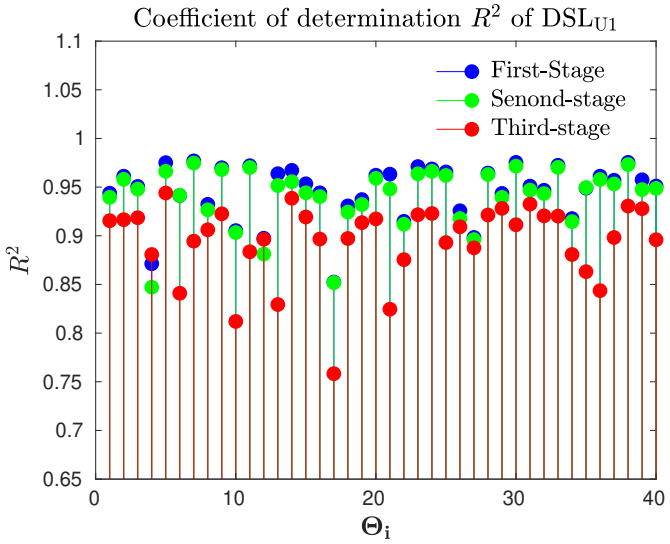

Figure 26: Coefficient of determination R^2 of the $\mathrm{DSL_{U1}}$ for each configuration $\mathbf{\Theta}_i = \mathbf{\Theta}_i(\theta_1, \theta_2, \theta_3)$, $i = 1, 2, ...k$, where k is the number of different composite layups corresponding to sampling point $\mathbf{G_1}(t_1 = 1.61, t_2 = 1.17, t_3 = 1.19)$.

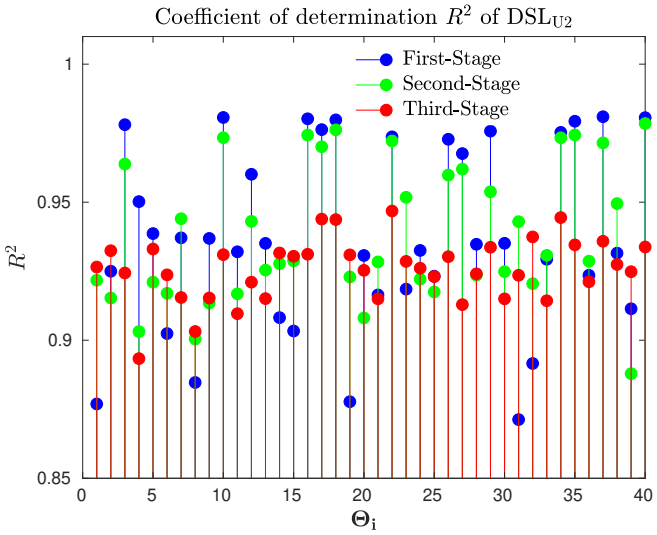

Figure 27: Coefficient of determination R^2 of the $\mathrm{DSL_{U2}}$ for each configuration $\mathbf{\Theta}_i = \mathbf{\Theta}_i(\theta_1, \theta_2, \theta_3)$, $i = 1, 2, ...k$, where k is the number of different composite layups corresponding to sampling point $\mathbf{G_1}(t_1 = 1.61, t_2 = 1.17, t_3 = 1.19)$.

in three subspaces ($n_1 = n_2 = n_3 = 3$), the necessary arithmetic operations can be estimated from Eq. (33)-(35) (Appendix A). It is worth to underline that in MSLS the computational complexity is reduced over two orders of magnitude in comparison to LLS for the same polynomial order as shown in a log-log scale in Fig. 28.

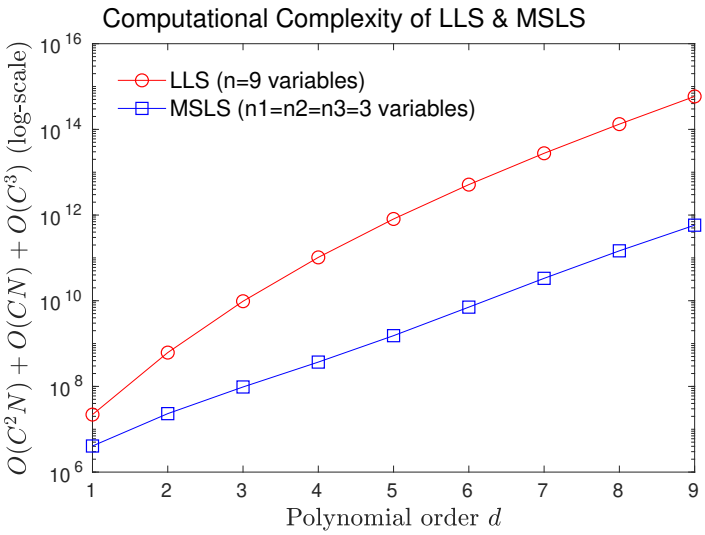

Figure 28: Computational complexity of the LLS and MSLS methods.

8. Model Validation of LLS and MSLS

8.1. Error Analysis of LLS and MSLS Models

This section has the purpose of establishing the predictive capabilities of the LLS and MSLS models away from the available training data. It is essential that the validation point set is spread over the design space to provide a reliable picture of the accuracy and not in close proximity to the training points, since that can lead to an over-optimistic evaluation of the model [41]. In the present study, 10 validation points are randomly chosen inside the range of the design space, which avoids extrapolations. The procedures described in Section 3 have been used for sampling the validation points (see Fig.8 and Fig.9). Error is defined as the difference between the actual response from FEM analysis and the predicted values from the LLS or MSLS models. The mean RMS error of the validation points is summarized in Table 1 and Table 2 for the DSL functions.

The LLS surrogate models for the DSL_{UR3} damage function are shown against the FEM simulation for the validation point VP5 in Fig. 29. The results here are very interesting. Damage evolution curves show the different behavior on the approximation using TSVD and Tikhonov regularization in comparison with a non-regularized solution. As a general trend, we observe numerical oscillations on the results when all energy modes are present in the model. These oscillations can be normally attributed to overfitting which leads to poor performance on the validation data set. It is evident that small singular values introduce 'noise' so that in some validation points the approximation deviates from the reference solution. On the other hand, TSVD and Tikhonov regularization based on the optimal L-curve corner shows an improvement on the approximation accuracy. Nevertheless, regularization didn't provide acceptable results for other validation points, as Fig.30 shows. The oscillatory solutions do not appear in these cases and the non-regularized results are closer to the high fidelity FEM solutions. Notice also that a particular choice of the optimal L-curve corner deserves special attention as may work well on some cases and not on others.

Using appropriate regularization in MSLS is not a straightforward task. To demonstrate that, the following regularization cases are investigated: a) Tikhonov regularization in all stages, and b) Tikhonov regularization at the First and Second-Stage and TSVD regularization at the Third-Stage by dropping a number of the last (smaller) singular values in all minimization problems. In a similar manner, a graphical comparison between the reference FEM solution and the DSL_{UR3} damage function for the same validation point (VP5) presented before is illustrated in Fig. 31. Results are also compared with the non regularized solution. It can been seen that the latter regularization case works well and provides reliable solutions in most of the validation points. Again there are exceptions for some validation points in which regularization failed to provide accurate results (Fig. 32). It needs to emphasized that there are many other possible combinations for regularization at the MSLS approach, however this is an interesting subject for future experimentation.

Table 1: Linear Least Squares

Damage function	DSL_{U1}	DSL_{U2}	DSL_{UR3}
RMS	0.0109	0.0582	0.0282

Table 2: Multi Stage Least Squares

Damage function	DSL_{U1}	DSL_{U2}	DSL_{UR3}
RMS	0.0066	0.0415	0.0255

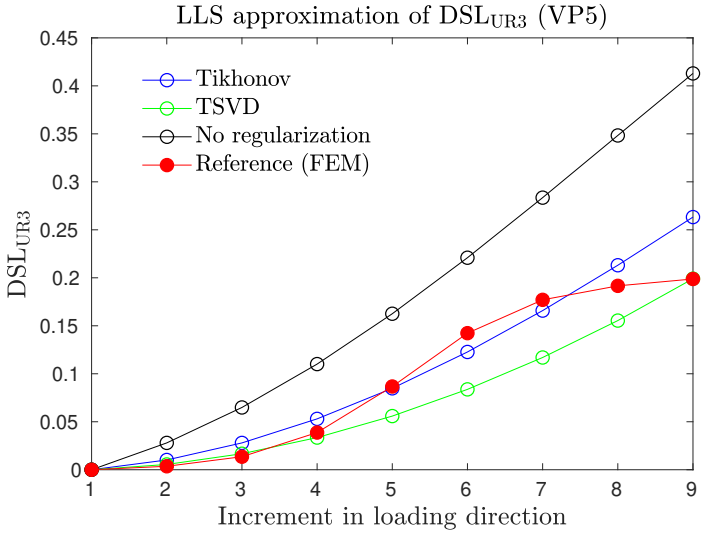

Figure 29: A graphical comparison between the reference (FEM) solution and the LLS approximated function DSL_{UR3} for validation point 5 using TSVD and Tikhonov regularization.

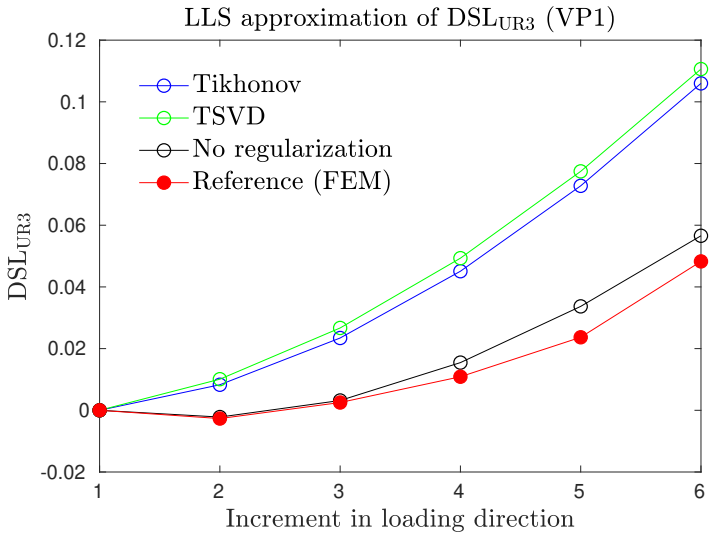

Figure 30: A graphical comparison between the reference (FEM) solution and the LLS approximated function DSL$_{UR3}$ for validation point 1 using TSVD and Tikhonov regularization.

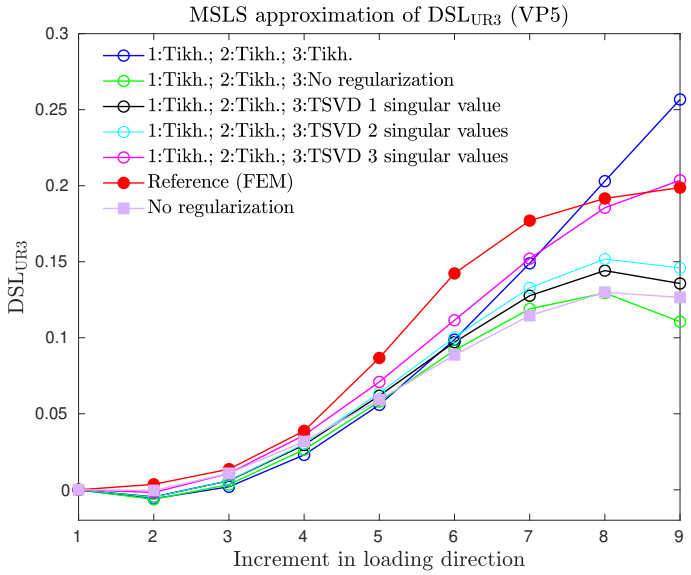

Figure 31: A graphical comparison between the reference (FEM) solution and the MSLS approximated function DSL$_{UR3}$ for validation point 5 using different regularization methods.

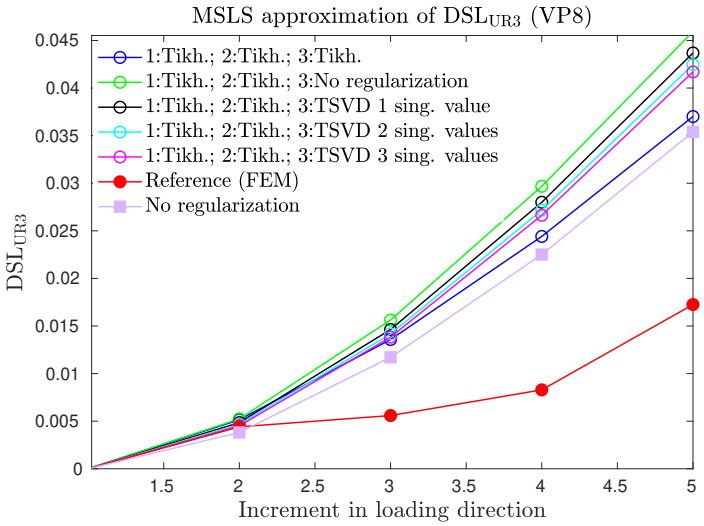

Figure 32: A graphical comparison between the reference (FEM) solution and the MSLS approximated function DSL$_{UR3}$ for validation point 8 using different regularization methods.

9. Synopsis and Perspectives

The aim of the present work is to increase the knowledge of damage evolution in advanced composite laminates so that balanced design decisions can be made. This methodology can be integrated into the building block approach to improve product performance. A multiscale VTA has been introduced based on the decomposition of a full-scale structure into a number of subcomponents. This approach aims to augment FEM analysis with a hierarchical modeling which goes down to the mesoscale where damage and delamination have their origins.

At the current stage, only delamination failures are taken into account at the local nonlinear computations, however, other interlaminar mechanisms described at the meso-scale can be incorporated into the model as well. By integrating a Linear Perturbation Analysis during the course of non-linear FEM computations, stiffness degradation functions associated to various reference directions can be evaluated for the corresponding laminate substructures.

Optimal space sampling of a high dimensional design space is an essential part of VTA. So far, several DOE methods have been used in this study regarding the different subsets of the design variables. It is evident that randomization can be helpful for making designs more robust. Hence, a po-

tential perspective could be to utilize alternative strategies for obtaining the training data, i.e., use several sets of points between the different subspaces of the design variables ($\{\mathbf{x^L}\}_i$, $\{\mathbf{x^\Theta}\}_j$, $\{\mathbf{x^G}\}_k$, $(i \neq j \neq k)$). Besides, using advanced space filling experimental designs for the constraint space spanned by the geometric variables $\mathbf{x^G}$, much more accuracy can be obtained.

Linear Least Squares and Multi-Stage Least Squares methods have been used for generating analytical damage functions. Regularization methods using the L_2-norm that have naturally associated L-curves defined in terms of norms for the particular methods are utilized. Numerical investigations demonstrated the computational efficiency and accuracy of the VTA. Validation points addressed the approximation error versus the different regularization methods. Satisfactory results have been achieved but there is room for further work especially in area of DOE as well as in implementing high-order polynomial approximations to increase accuracy. Although there is no baseline model for comparison, it can be concluded that both methods performed well and provide quite similar results for the majority of the training points. However, exceptions exist for validation points where regularization fails to provide accurate results. It has to be remarked that the similarity of different model predictions (LLS and MSLS) serves as an alternative way for model validation.

An interesting area for future work would be to solve the LLS problem considering different norms for regularization, as for example, using the L_1 norm (lasso), combinations of L_1 and L_2 norms (elastic net regression) or using the L_0 norm (best subset selection). Especially, the MSLS method could reduce the computational complexity associated with the combinatorial problem of the L_0 norm. Furthermore, an alternative way to obtain an analytical solution for the damage functions is to use Artificial Neural Networks (ANN) which become relatively competitive to conventional regression models.

The proposed method is expected to pave the way for a new computational strategy based on a concept of 'damage databases' (at a premature state) associated with complex laminate composite structures. Indeed, assuming that the design space has been efficiently sampled and relevant failure criteria are precomputed using the training points, when a request is formulated for a new full scale structure, the developed predictive model is enabling the computation of failure criteria associated with the desired substructural components in near real-time.

10. Acknowledgments

The authors would like to acknowledge CETIM for its financial support in the framework of the common lab COMPINNOV.

11. Data availability

The raw/processed data required to reproduce these findings cannot be shared at this time due to technical or time limitations.

12. Appendix A

12.1. Computational Complexity of Least Squares

A Least Squares regression with N training examples and P features, using the normal equations takes:

- $\tilde{\mathcal{O}}(NP^2)$ time to multiply $\mathbf{X}^T$ by $\mathbf{X}$
- $\tilde{\mathcal{O}}(NP)$ time to multiply $\mathbf{X}^T$ by $\mathbf{Y}$
- $\tilde{\mathcal{O}}(P^3)$ time to compute the LU (or Cholesky) factorization of $\mathbf{X}^T\mathbf{X}$ and use that to compute the product $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$

Therefore, the total time complexity is

$$\tilde{\mathcal{O}}(NP^2) + \tilde{\mathcal{O}}(NP) + \tilde{\mathcal{O}}(P^3) \tag{32}$$

Asymptotically, $\tilde{\mathcal{O}}(NP^2)$ dominates $\tilde{\mathcal{O}}(NP)$ so the $\tilde{\mathcal{O}}(NP)$ term can be neglected. Since the normal equations are used one can assume that $N \gg P$ otherwise the matrix $\mathbf{X}^T\mathbf{X}$ would be singular (and hence non-invertible), which means that $\tilde{\mathcal{O}}(NP^2)$ asymptotically dominates $\tilde{\mathcal{O}}(P^3)$. Therefore the total time complexity is $\tilde{\mathcal{O}}(NP^2)$.

12.2. Computational complexity of the MSLS

In this section the time complexity of MSLS is addressed. The total time complexity for the LLS using the normal equations is $\tilde{\mathcal{O}}(C^2N) + \tilde{\mathcal{O}}(CN) + \tilde{\mathcal{O}}(C^3)$ (Eq. (32)). Thus, for the MSLS, the time complexity for the different approximation stages can be defined as follows:

First-Stage Least Squares. In the subspace spanned by the variables $\mathbf{x}^{\mathbf{L}}$ one needs to solve $(N_\Theta \times N_G)$ minimization problems for every sampling set $\{\mathbf{x}^{\mathbf{\Theta}_i}, \mathbf{x}^{\mathbf{G}_j}\}$ in the design space. Assume a polynomial order d_1 and n_1 variables, the number of regressors is given by the binomial coefficient $P_1 = \binom{n_1+d_1}{n_1}$. The arithmetic operations can be determined as

$$N_\Theta \cdot N_G \cdot \tilde{\mathcal{O}}([N(\mathbf{x}^{\mathbf{L}})P_1^2 + N(\mathbf{x}^{\mathbf{L}})P_1 + P_1^3]) \tag{33}$$

Second-Stage Least Squares. In the subspace spanned by the variables $\mathbf{x}^{\mathbf{\Theta}}$ one needs to solve $(P_1 \times N_G)$ minimization problems for every sampling set $\{\mathbf{x}^{\mathbf{G}_j}\}$ in the design space. Assume a polynomial order d_2 and n_2 variables for the Second-Stage approximation, the number of combinations between the regressors is given by the binomial coefficient $P_2 = \binom{n_2+d_2}{n_2}$. The arithmetic operations can be determined as

$$P_1 \cdot N_G \cdot \tilde{\mathcal{O}}([N(\mathbf{x}^{\mathbf{\Theta}})P_2^2 + N(\mathbf{x}^{\mathbf{\Theta}})P_2 + P_2^3]) \tag{34}$$

Third-Stage Least Squares. In the subspace spanned by the variables $\mathbf{x}^{\mathbf{G}}$ one needs to solve $(P_1 \times P_2)$ minimization problems. Again assume a polynomial order d_3 and n_3 variables for the Third-Stage approximation, the number of combinations between the regressors is given by the binomial coefficient $P_3 = \binom{n_3+d_3}{n_3}$. The arithmetic operations can be determined as

$$P_1 \cdot P_2 \cdot \tilde{\mathcal{O}}([N(\mathbf{x}^{\mathbf{G}})P_3^2 + N(\mathbf{x}^{\mathbf{G}})P_3 + P_3^3]) \tag{35}$$

References

[1] ASM Handbook, Volume 21 Composites, ISBN: 978-0-87170-703-1, ASM International, 2001.

[2] Krueger R. Development of benchmark examples for quasi-static delamination propagation and fatigue growth predictions. SIMULIA Community Conference, Providence, RI, May 14-17, 2012.

[3] Camanho PP, Davila CG, Ambur DR. Numerical Simulation of Delamination Growth in Composite Materials. NASATP-2001-211041, 2001.

[4] Camanho PP, Davila CG. Mixed-Mode Decohesion Finite Elements for the Simulation of Delamination in Composite Materials. NASA/TM-2002-211737, 2002.

[5] Myers RH, Montgomery DC, Andersson-Cook CM. Response surface methodology: Process and product optimization using designed experiments (Third ed.), Hoboken, New Jersey, USA: Wiley, 2008.

[6] Sacks J, Welch WJ, Mitchell TJ, Wynn HP. Design and Analysis of Computer experiments. Statistical Science, 4, 409-435, 1989.

[7] Babaei H, Mostofi TM. Modeling and prediction of fatigue life in composite materials by using the singular value decomposition method. Journal of Materials: Design and Applications, 0(0)1-9, 2016.

[8] Allegri G, Wisnom MR, Hallett SR. A simplified approach to the damage tolerance design of asymmetric tapered laminates. Part II: methodology validation. Composites Part A, 41(10), 1395-1402, 2010.

[9] Cui W, Wisnom MR, Jones MI. An experimental and analytical study of delamination of unidirectional specimens with cut central plies, Journal of Reinforced Plastics and Composites. 13, 722-739, 1994.

[10] Petrossian Z, Wisnom MR. Parametric study of delamination in composites with discontinuous plies using an analytical solution based on fracture mechanics. Composites Part A, 29A, 403-414, 1998.

[11] Camanho PP, Davila CG, De Moura MF. Numerical Simulation of Mixed-mode Progressive Delamination in Composite Materials. Journal of Composite Materials, 37(16), 2003.

[12] Turon A, Camanho PP, Costa J, Renart J. Accurate simulation of delamination growth under mixed-mode loading using cohesive elements: Definition of interlaminar strengths and elastic stiffness. Composite Structures, 92, 1857-1864, 2010.

[13] Xie D, Garg M, Huang D, Abdi F. Cohesive Zone Model for Surface Cracks using Finite Element Analysis. 49th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference, 7-10 April, Schaumburg, IL, 2008.

[14] Krueger R, Cvitkovich MK, O'Brien TK, Minguet PJ. Testing and Analysis of Composite Skin/Stringer Debonding under Multi-Axial Loading. Journal of Composite Materials, 34(15), 1263-1300, 2000.

[15] Reeder J, Kyongchan S, Chunchu PB, Ambur DR. Postbuckling and Growth of Delaminations in Composite Plates Subjected to Axial Compression. 43rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, Denver, Colorado, 1746, 10, 2002.

[16] Benzeggagh ML, Kenane M. Measurement of mixed-mode delamination fracture toughness of unidirectional glass/epoxy composites with mixed-mode bending apparatus. Composites Science and Technology, 56:439-449, 1996.

[17] ABAQUS 6.14 User's Manual, ABAQUS, Dassault Systèmes Simulia Corp., Providence, RI, USA.

[18] Koehler JR, Owen AB. Computer Experiments. Handbook of Statistics (Ghosh, S. and Rao, C. R., eds.), Elsevier Science, New York, 261-308, 1996.

[19] Owen AB. Orthogonal arrays for computer experiments, integration and visualization. Statistica Sinica, 2, 439-452, 1992.

[20] McKay MD, Beckman RJ, Conover WJ. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics, 21(2), 1979.

[21] Tang B, Orthogonal array-based Latin hypercubes. Journal of the American Statistical Association, 88(424), 1993.

[22] Stinstra E, Stehouwer P, Hertog D, Vestjens A. Constrained Maximin Designs for Computer Experiments. Technometrics, 45, 340-346, 2003.

[23] Simpson TW. Comparison of Response Surface and Kriging Models in the Multidisciplinary Design of an Aerospike Nozzle. Institute for Computer Applications in Science and Engineering, NASA Langley Research Center Hampton, VA. ICASE Report N.98-16, 1998.

[24] Cohen A, Davenport MA, Leviatan D. On the stability and accuracy of least squares approximations. Foundations of Computational Mathematics, 13(5):819-834, 2013.

[25] Strang G. Introduction to Linear Algebra. 3rd ed., Wellesley-Cambridge, ISBN 0-9614088-5-5, 1998.

[26] Golub GH, Reinsch C. Singular value decomposition and least squares solutions. Numerische Mathematik, 14, 403-420, 1970.

[27] Friedman J, Hastie T, Tibshirani R. The Elements of Statistical Learning, (10th ed.), New York, NY, USA: Springer-Verlag, 2013.

[28] Furnival GM, Wilson RW. Regressions by leaps and bounds. Technometrics, 6, 499-511, 1974.

[29] Wilson ZT, Sahinidis NV. The ALAMO approach to machine learning. Computers and Chemical Engineering, 106, 785–795, 2017.

[30] Hanson RJ. A numerical method for solving Fredholm integral equations of the first kind using singular values. SIAM Journal on Numerical Analysis. 8, 616-622, 1971.

[31] Varah JM. On the numerical solution of ill-conditioned linear systems with applications to ill-posed problems. SIAM Journal on Numerical Analysis, 10, 257-267, 1973.

[32] Golub G. Numerical Methods for Solving Linear Least Squares Problems. Numerische Mathematik, 7, 206-216, 1965.

[33] Bjork A. Iterative refinement of linear least squares solutions. II, BIT, 8, 8-30, 1968.

[34] Varah JM. A practical examination of some numerical methods for linear discrete ill-posed problems. SIAM Review, 21, 100-111, 1979.

[35] Peters G, Wilkinson JH. The least squares problem and pseudoinverses. Comput. J., 13, 309-316, 1970.

[36] Phillips DL. A technique for the numerical solution of certain integral equations of the first kind. J. ACM, 9, 84-97, 1962.

[37] Tikhonov AN. Solution of incorrectly formulated problems and the regularization method. DokL Akad. Nauk. SSSR 151, 501 504 = Soviet Math. Dokl. 4, 1035-1038, 1963.

[38] Regularization Tools: A Matlab Package for Analysis and Solution of Discrete Ill-Posed Problems. Numerical Algorithms, 6, 1-35, 1994.

[39] Hansen PC, Jensen TK, Rodriguez G. An adaptive pruning algorithm for the discrete L-curve criterion. Journal of Computational and Applied Mathematics, 198, 483-492, 2007.

[40] Hansen PC, O'Leary DP. The use of the L-curve in the regularization of discrete ill-posed problems. SIAM Journal on Scientific Computing, 14, 1487-1503, 1993.

[41] Iooss B, Boussouf L, Feuillard V, Marrel A. Numerical studies of the metamodel fitting and validation processes. International Journal on Advances in Systems and Measurements, 3(1 & 2), 11-21, 2010.