Iranian Journal of Operations Research
Vol. 10, No. 2, 2019, pp. 01-17

Forward and Backward Uncertainty Quantification in Optimization

B. Mohammadi[1]

This contribution gathers some of the ingredients presented during the Iranian Operational Research community gathering in Babolsar in 2019. It is a collection of several previous publications on how to set up an uncertainty quantification (UQ) cascade with ingredients of growing computational complexity for both forward and reverse uncertainty propagation.

***Keywords**: Uncertainty, Optimization, Complexity.*

Manuscript was received on 10/07/2019, revised on 14/10/2019 and accepted for publication on 02/12/2019.

1. Introduction

This contribution uses data analysis ingredients in a context of existing deterministic simulation platforms. It starts with a complexity-based splitting of the independent variables and the definition of a parametric optimization problem. Geometric characterization of global sensitivity spaces through their dimensions and relative positions through principal angles between vector spaces bring a first set of information on the impact of uncertainties of the functioning parameters on the optimal solution. Joining the multi-point descent direction and Probability Density Function (PDF) quantiles of the optimization parameters permits to define the notion of Directional Extreme Scenarios (DES) without sampling of large dimension design spaces. One goes beyond DES with Ensemble Kalman Filters (EnKF) after the multi-point optimization algorithm is cast into an ensemble simulation environment. This formulation accounts for the variability in large dimension. The UQ cascade continues with the joint application of the EnKF and DES leading to the concept of Ensemble Directional Extreme Scenarios (EDES) which provides a more exhaustive description of the possible extreme scenarios. The different ingredients developed for this cascade also permits to quantify the impact of state uncertainties on the design and provide confidence bounds for the optimal solution. This is typical of inverse designs where the target should be assumed uncertain. Our proposal uses the previous DES strategy applied this time to the target data. We use these scenarios to define a matrix having the structure of the covariance matrix of the optimization parameters. This construction can be compared to another one using available adjoint-based gradients of the functional. Eventually, we go beyond inverse design and apply the method to general optimization problems.

2. Motivation

We consider a generic situation where the simulation aims at predicting a given quantity of interest $j(\mathbf{x}, \alpha)$ and there are a few functioning or operating parameters α and several design parameters $\mathbf{x}$

[1] Univ. Montpellier, IMAG, Montpellier, France, Email: bijan.mohammadi@umontpellier.fr.

involved. The ranges of the functioning parameters define the global operating/functioning conditions of a given design. This splitting of the independent variables in two sets is important in the sequel.

We propose a cascade of ingredients to account for uncertainties avoiding any sampling of large dimensional spaces. A sampling will be only necessary for the functioning parameters **u** range leading to a multi-point optimization problem.

The literature on uncertainty quantification (UQ) is huge. In short, forward propagation aims at defining a probability density function for j knowing those of **x** and α [15, 18, 25]. This can be done, for instance, through Monte Carlo simulations or a separation between deterministic and stochastic features using Karhunen-Loeve theory (polynomial chaos theory belongs to this class) [12, 13, 16, 48, 50]. Examples of shape optimization with polynomial chaos and surrogate models during optimization are given in [6, 39].

Backward propagation aims at reducing the model's bias or calibrating model's parameters knowing the probability density function of j (or other constraints and observations) [5, 23, 45]. This can be seen as a minimization problem and Kalman filters [24] give, for instance, an elegant framework for this inversion assimilating the uncertainties on the observations.

Our aim is to propose a geometric framework to address the curse of dimensionality of existing approaches related to the explosion of their computational complexity due to the sampling necessary to access probabilistic information, even if this can be improved with intelligent sampling techniques [4, 43]. The different ingredients presented here can be applied with either high-fidelity or reduced order models, when available [38, 40, 41, 47]. Low-order models are often used instead of the full models to overcome the computational complexity of UQ.

After the splitting of the independent variables mentioned above, we define a multi-point formulation to account for the variability on α. This is feasible because the size of α is assumed to be small. We define a global sensitivity space using the sensitivities of j with respect to **x** for the multi-point problem. Once this space is built, we analyze its dimension. We previously showed how to perform this task and how to use this information for adaptive sampling [28, 36].

The next step is to analyze the impact of different modelling or discretizations on the results. Different models or solution procedures lead to different sensitivity spaces. Beyond their respective dimensions, principal angles between the respective sensitivity vector spaces permit to measure the deviation due to such changes. The dimensions of the spaces and the angles are interesting measures for both the epistemic and aleatory uncertainties. Indeed, suppose that, for a given model the dimensions of the sensitivity spaces remain unchanged when enriching the sampling of the functioning parameter range. This stability would be a first indication of a low level of sensitivity of the simulations with respect to this parameter. Once this is established, principal angles between subspaces permit to analyze both the impact of a given evolution of the modelling on the sensitivity spaces or an enrichment of our sampling. Eventually, constant dimension and low angles will clearly indicate a situation of low uncertainty.

These ingredients can be used in a context of multi-point robust analysis of a system to define worst-case scenarios for its functioning. To this end, we combine a multi-point sensitivity with the probabilistic features of the control parameters through their quantiles [22, 31] to define the concept of Directional Extreme Scenarios (DES) without a sampling of large dimension design spaces.

Ensemble Kalman filters (EnKF) [2, 8, 9, 10, 11, 24] permit to go beyond the directional uncertainty quantification concept when accounting for the uncertainties in large dimension. They also permit backward uncertainty propagation assimilating the uncertainty on the functional and constraints during the design. We cast our multi-point optimization problem into the ensemble formulation. A joint application of the EnKF and DES leads to the concept of Ensemble Directional Extreme Scenarios (EDES) to provide a more exhaustive description of possible extreme scenarios.

Despite these approaches avoiding the sampling of a large dimensional space, the computing cost remains high and the procedures turn to be difficult to simply explain in engineering environments. We propose a low-complexity approach for the inversion of uncertain data where the target state $\mathbf{u}^*$ used in an inverse problem is uncertain. In this situation, we consider functional of the form $j(\mathbf{x}, \alpha, \mathbf{u}^*) = \|\mathbf{u}(\mathbf{x}, \alpha) - \mathbf{u}^*(\alpha)\|$ to reduce the distance between a model state $\mathbf{u}(\mathbf{x}, \alpha)$ and observations.

Targeting uncertain data is a realistic situation as the acquired data are usually uncertain. It is therefore interesting to be able to quantify the impact of this uncertainty on the inversion results. An important information will be the sensitivity of the design to a given level of uncertainty on the data at some location. Indeed, if this sensitivity is low, this would be an indication that a more accurate acquisition is unnecessary.

Considering the target as uncertain is also interesting because we do not always have existence of a solution for an inversion problem as $\mathbf{u}^*$ is not necessarily solution of the state equation making an exact or deterministic inversion pointless. Also, the approach permits to go beyond inversions based on least squares minimization involving a mean state target.

Finally, the uncertainty in measurements is also an interesting way to account for the presence of variability in the state. More generally, as the model and numerical procedures are by nature imperfect and partial, we can consider this uncertainty as a representation or estimation of the imperfections. These imperfections are even more present in inverse problems where one cannot afford the same level of resolution thus for a single simulation. We therefore need to be able to quantify the impact of these weaknesses on the design. The approach presented here is therefore also useful to account for epistemic uncertainties related to possible model or solution procedure deficiency.

Concerning the computational cost of these analyses, one can say that, when using the same calculation ingredients as in a high-fidelity simulation (i.e., without calling for low-order models or cheaper discretizations), the best calculation complexity for a simulation under uncertainty is when its cost is comparable to the deterministic situation. This is clearly unreachable, except if all the extra effort can be achieved in a fully parallel manner and parallel to the initial deterministic calculation so that the time to solution to remain unchanged when accounting for the presence of uncertainties. This is the case with the Monte Carlo approaches. But these are quite expensive and do not take advantage of available simulation environments. In particular, when an adjoint-based optimization environment exists. Our proposal consists of upgrading existing platforms without abandoning what has been built for the deterministic situations and with keeping the time to solution unchanged in the presence of uncertainties with two sources of parallelism coming from the multi-point formulation to account for the uncertainties on the functioning parameters and from the EnKF formulation for those on the optimization variables and observation data.

3. Parametric Optimization

We are interested in a class of optimization problems where the cost function involves a functioning parameter α not considered as a design parameter:

$$\min_{\mathbf{x}\in\mathbf{O}_{ad}} j(\mathbf{x},\alpha),\ \ \alpha\in\mathbf{I}\subset\mathbb{R}^n,\qquad \mathbf{O}_{ad}\subset\mathbb{R}^N, \tag{1}$$

where $\mathbf{x}$ is the design vector belonging to $\mathbf{O}_{ad}$, the optimization admissible domain. Usually, the functioning parameters (or operating conditions) α are just a few. On the other hand, the size N of $\mathbf{x}$ is usually large. Together, $\mathbf{x}$ and α fully describe our system and we have $n << N$. This splitting between functioning parameters (or operating conditions) and design variables is central to our discussion.

In [28, 29] we showed how to use multi-point optimization to address such optimization problem. The aim is to remove, during optimization, the dependency in α. This is done minimizing a functional $J(\mathbf{x})$ encapsulating this dependency expressed through $\mathbf{A}=\{j(\mathbf{x},\alpha_k),\alpha_k\in\mathbf{I}_M\}$ over $\mathbf{I}_M$ a given sampling of $\mathbf{I}$:

$$J=\mathbf{J}(\mathbf{A}),\qquad \text{such that } \mathbf{G}(\mathbf{A})\le 0. \tag{2}$$

Several choices are possible for $\mathbf{J}$ and $\mathbf{G}$ to address the issue of robust design. For instance, following Taguchi's definition, one can look for minimal-variance design or only a given level for the variance. Indeed, a classical approach to extend the single point design and improve off-design points is to control μ, mean performance, and σ, variance of the functional [44] as in First-Order Second Moment (FOSM) methods [27]. One can also look for information about the tails of the distributions which can be linked to the variance in the Gaussian framework and we use this relationship in quantile-based extreme scenarios.

Often it might be interesting to go beyond the first two moments and in particular consider the first four moments of j during the design. Going beyond the first two moments is important when the PDF of j deviates from a pure Gaussian distribution. Indeed, even with interval-based (with uniform PDF) or Gaussian entries there is no reason the PDF of the solution of a simulation to remain uniform or Gaussian.

The third and fourth moments are the skewness, γ, and the kurtosis, κ. One can consider that a robust design should favor symmetry in the distribution which means a lower absolute value of skewness. For instance, in a Gaussian distribution we have $\gamma=0$. Also, in a normal distribution the mean and median coincide and if a PDF is not too far from a normal distribution, the median will be near $\mu-\gamma\sigma/6$. Therefore, if $|\gamma|\to 0$ then the PDF tends toward a normal distribution. This provides an inequality constraint on $|\gamma|$ as γ can be positive or negative. For a unimodal PDF a reduction of the skewness comes when the mean and the mode of the distribution converge towards each other at a given standard deviation.

Concerning the fourth moment, a robust design should favor higher density near the mean which means a higher kurtosis, but this is more subtle. Indeed, in spite of the fact that a higher kurtosis means a higher concentration of the probability mass around the mean, it could also imply thicker tails in the PDF. This means that more of the variance is the result of infrequent extreme deviations. We need therefore to define what we mean by a more robust design: acceptance of frequent modest

deviations or acceptance of infrequent extreme ones. If operational security is a major concern, then the latter should be definitely avoided. Hence, a reasonable requirement would be to have a design reducing the initial kurtosis value, $\kappa \leq \kappa_0$, together with a constraint on the variance σ.

4. Gradients, Sensitivity Spaces and Admissible Search Directions

Monte Carlo simulations permit to recover the moments with an error decreasing as $\sigma M^{-\frac{1}{2}}$, with M being the number of functional evaluations and the rate being independent of n. But, for small n, classical numerical integration outperforms Monte Carlo simulation in terms of complexity based on the number of functional evaluations to attain at a given accuracy these moments. As we are concerned with small values of n (typically, $n = 2$ or 3 in our applications), this latter is therefore preferred. Anyway, both Monte Carlo trials and numerical integration lead to the introduction of weighted sums over an M-point sampling $\mathbf{I}_M$ of I as estimators of the previous moments.

The linearity in the sums permits to access the gradient of the moments with respect to the control parameters $\mathbf{x}$ from the gradient of the functional at the sampling point α_k. These are four vectors in $S_M = Span\{\nabla_{\mathbf{x}} j(\mathbf{x}, \alpha_k),\ \alpha_k \in I_M\} \subset \mathbb{R}^N$. In applications of interest, N is large. However, we showed that often $\dim(S_M) << N$ [28, 29, 30]. This analysis also permits to posteriori give confidence bounds on the choice of the sampling size M which should be clearly larger than $\dim(S_M)$.

Let us denote by $C_{i=1,2,3}$ the three constraints on the second, third and fourth moments and let us consider the subspace $s_M = Span\{\nabla_{\mathbf{x}} C_{i=1,2,3}\} \subset \mathbb{R}^3 \subset \mathbb{R}^N$. Obviously $p = dim(s_M) \leq 3$. Let us denote by $\{\mathbf{q}_{i=1,\dots,p}\}$, an orthonormal basis for s_M obtained, for instance, orthonormalizing the three gradient vectors by the Gram-Schmidt procedure. The gradients G of the constraints can therefore be expressed as a linear combination of the q_i: $G = \left(\nabla_{\mathbf{x}} C_{i=1,2,3}\right) = P^{-1}\left(\mathbf{q}_{i=1,\dots,p}\right)$ with P being the matrix expressing the coordinates of $\mathbf{q}$ in G.

With equality constraints, a descent direction d can be obtained writing the first order optimality condition stating that d needs to be orthogonal to s_M. Hence, using the local orthonormal basis $\left\{\mathbf{q}_{i=1,\dots,p}\right\}$, we consider d, given by

$$d = \nabla_{\mathbf{x}}\mu - \sum_{i=1}^{p} \langle \nabla_{\mathbf{x}}\mu, \mathbf{q}_i \rangle \mathbf{q}_i \,. \tag{3}$$

Denoting by Π the matrix of the projection operator $\langle \nabla_{\mathbf{x}}\mu, q \rangle$, we have

$$d = \nabla_{\mathbf{x}}\mu - (\Pi PG)^t\ PG = \nabla_{\mathbf{x}}\mu - (G^t P^t\, \Pi\, P)^t\ G = \nabla_{\mathbf{x}}\mu + \Lambda^t G,$$

with $\Lambda^t = (\lambda_1, \lambda_2, \lambda_3) \in \mathbb{R}^3$. We have $d \to 0$ with the optimization iterations converging. Stationarity in d therefore realizes the first order optimality conditions for the Lagrangian $L = J + \Lambda^t C$.

With inequality constraints, the solution of our minimization problem needs to verify the first order KKT conditions [37]. But, the optimality conditions for the Lagrangian will involve only positive Lagrange multipliers: $\Lambda \in \mathbb{R}^3_+$ and $\nabla_{\mathbf{x}} L = \nabla_{\mathbf{x}} J + \Lambda^t \nabla_{\mathbf{x}} C = 0$ with the complementarity condition $\Lambda^t C = 0$, meaning that $\lambda_i = 0$ if $C_i \leq 0$ and $\lambda_i > 0$ if $C_i = 0$ (i.e., C_i is an active constraint). To

define d we follow what would be placed for the equality constraints, only considering active constraint gradients in the definition of s_M which is not a subspace but a convex cone:

$$s_M = \{\mathbf{x} | \, \mathbf{x} = \sum_{i=1}^{3} \beta_i \, \nabla_{\mathbf{x}} C_i, \qquad \beta_i > 0 \mid C_i\} \subset \mathbb{R}^3 \subset \mathbb{R}^N. \tag{4}$$

At the solution, $\nabla_{\mathbf{x}} J$ is orthogonal to this cone. Before working on the cone, let us start defining a local orthonormal basis $\{\tilde{\mathbf{q}}_{i=1,\dots,p}\}$ for $\tilde{s}_M$ from (4) but with $\beta_i \in \mathbb{R}$. This is therefore a subspace and the basis can be defined, as before with $p = dim(s_M)$. Now, we have $q_i = \pm \tilde{q}_i$ with the sign chosen such that $< \mathbf{q}_{i=1,\dots,p}, \nabla_{\mathbf{x}} C_j > \geq 0$, if $C_j = 0$, for $j = 1, \dots, 3$ (i.e., pointing inside the cone).

Here, $\{q_{i=1,\dots,p}\}$ are therefore the generators of the cone s_M deduced from a basis of $\tilde{s}_M$. If the generators cannot be defined, then the problem is found to have no solution, as at least two of the constraints are incompatible with the gradients being parallel and pointing in opposite directions. These generators permit to define the admissible search direction d from (3) but taking into account that we only remove the non admissible contribution:

$$d = \nabla_{\mathbf{x}} \mu - \sum_{i=1}^{p} \chi_i \, \langle \mathbf{q}_i, \nabla_{\mathbf{x}} \mu \rangle \mathbf{q}_i, \tag{5}$$

with $\chi_i = 0$ if $\langle \mathbf{q}_i, \nabla_{\mathbf{x}} \mu \rangle \geq 0$ and $\chi_i = 1$ if $\langle \mathbf{q}_i, \nabla_{\mathbf{x}} \mu \rangle < 0$.

5. A Multi-point Descent Algorithm

Our aim is to use existing platforms. Hence, to compute the ingredients above we use an available single-point optimization environment which can easily be modified for parallel multi-point calculations. This platform involves a direct simulation chain linking the parameters $(\mathbf{x}, \alpha)$ to the state $\mathbf{u}$, solution of a state equation $F\left(\mathbf{u}\left(q(\mathbf{x}, \alpha)\right)\right) = 0$, and its adjoint $\mathbf{v}$ and to a functional j:

- Give $\mathbf{x}_0$, $0 < \rho$, $\mathbf{I}_M$, p_{max}.
- Optimization iterations: For $p = 1, \dots, p_{max}$ do
  - 1-M parallel state equation solutions $F\left(\mathbf{u}\left(q(\mathbf{x}_p), \alpha_k\right)\right) = 0$, $\alpha_k \in \mathbf{I}_M$,
  - 2-M parallel evaluations of $j(\mathbf{x}_p, \alpha_k)$, $\alpha_k \in \mathbf{I}_M$,
  - 3-M parallel solutions of the adjoint state v equation:
    $$v^t F_{\mathbf{u}}\left(\mathbf{u}\left(q(\mathbf{x}_p), \alpha_k\right)\right) = j_u^t, \qquad \alpha_k \in \mathbf{I}_M,$$
  - 4-M parallel evaluations of $\nabla_{\mathbf{x}} j(\mathbf{x}_p, \alpha_k) = j_{\mathbf{x}} + (\mathbf{v}^t F_{\mathbf{x}})^t$, $\alpha_k \in \mathbf{I}_M$,
  - 5-define d the descent direction using (5),
  - 6-minimization using d: (e.g., $\mathbf{x_{p+1}} = \mathbf{x}_p - \rho d$),
  - Stop if a given stopping criterion is achieved.

In multi-criteria problems, steps 2, 3 and 4 include the treatment of more than one functional inducing a different definition of the descent direction d to account for other constraints (mainly physical this time) than the moment-based ones mentioned above.

Despite the natural presence of parallelism due to the M independent evaluations of the state, functional and its gradient, computational complexity remains an issue. We have shown previously how to reduce this effort optimizing the sampling size [28] together with the use of incomplete sensitivity concept in the evaluation of the gradients which permits to avoid the solution of the M adjoint equations [35]. This is particularly suitable when using black-box state equation solvers not providing the adjoint of the state variables.

Such minimization problems have produced new interests to descent methods and this not only because of their lower computational complexity, as compared to gradient free methods [1, 19, 37]. Indeed, beyond minimization, we saw that gradients are useful to see what should actually be the search space in a context of robust multi-point design [29, 30]. Hence, beyond individual gradient accuracy (i.e. at each of the sampling point), what is important in multi-point problems is the global search space defined by the ensemble of the gradient vectors. This means that one might tolerate higher error levels in each of the gradient defined at the different sampling point than for a single-point optimization situation as what is important is for the global search space to remain nearly unchanged. An interesting mathematical concept which permits to measure the deviation between two subspaces is the principal angles between subspaces.

6. Angles Between Subspaces

We use the mathematical concept of principal angles between subspaces in the Euclidean space (here $\mathbb{R}^N$), initially introduced by Jordan [21]. If the maximum principle angle between two subspaces is small, then the two are nearly linearly dependent. Geometrically, this is the angle between two hyperplanes embedded in a higher dimensional space.

Let us briefly recall the concept of principal angles and how to practically compute them [14, 20]. For simplicity, suppose A and B are two subspaces of dimension k of $\mathbb{R}^N$, $N \geq 2k$, although this is not a prerequisite to define the principal angles. The k principal angles $\{\theta_i, i = 1, \dots, k\}$ are recursively defined as

$$\cos(\theta_i) = \frac{\langle a_i, b_i \rangle}{\|a_i\| \|b_i\|} = \max \left\{ \frac{\langle a, b \rangle}{\|a\| \|b\|} : \; a \perp a_m, b \perp b_m : m = 1, \dots, i-1 \right\},$$

where $a_j \in A$ and $b_j \in B$.

The principal angles θ_i are between 0 and $\pi/2$. This is an important point and will be used later to take advantage of the positivity of the cosine of the angles. The procedure finds unit vectors $a_1 \in A$ and $b_1 \in B$ minimizing the angle θ_1 between them. It then takes the orthogonal complement of a_1 in A and b_1 in B and iterates. This procedure is not useful, in practice, as computationally inadequate. We would like to be able to find the angles θ_i from the inner products $\langle a_i, b_j \rangle$ of the elements of two bases of A and B [42]. This would be interesting in our multi-point optimization context where we can exhibit an orthonormal basis of the global search space for the multi-point optimization problem using the Gram-Schmidt orthonormalization.

Now, let $\{a_i, i = 1, \dots, k\}$ and $\{b_i, i = 1, \dots, k\}$ be two arbitrary orthonormal bases for A and B. Orthonormal bases are easy to obtain through the Gram-Schmidt orthonormalization procedure. Consider M being the matrix of the projection operator Pr_A of B onto A, defined by

$$Pr_A(b_i) = \sum_{j=1}^{k} \langle b_i, a_j \rangle a_j \,, \qquad M = \left(\langle b_i, a_j \rangle\right)_{i,j}.$$

The principal angles can be linked to this operator [42] through

$$M = G\Sigma H^t,$$

where G and H are orthogonal matrices and $\Sigma = diag(\cos(\theta_i))$.

With G and H being orthogonal matrices, this is a Singular Vector Decomposition (SVD) of M. G and H are unknown at this point. But we will show that we do not need them to get the θ_i. Otherwise, the approach will again be computationally useless.

We recall that the columns of G are the left-singular vectors of M and eigenvectors of MM^t and the columns of H are the right-singular vectors of M and eigenvectors of M^tM. And most importantly, the $\cos^2(\theta_i)$ are the eigenvalues of $Pr_A^t Pr_A$, written in matrix form as: $M^tM = (G\Sigma H^t)^t(G\Sigma H^t) = H\Sigma^2 H^t$ with $\Sigma^2 = diag(\cos^2(\theta_i))$.

Therefore, to find the principal angles between subspaces A and B, knowing an orthonormal basis in each subspace, one should calculate M and find the eigenvalues of M^tM and take the square root of them. This last operation is valid as the angles are between 0 and $\pi/2$, and the cosine is therefore always positive.

We presented the approach for subspaces of the same dimension k, but it is not necessary for the two subspaces to be of the same size in order to find the angles between them. We need $N \geq 2k$ to be able to exhibit two orthogonal subspaces. If $N < 2k$, then some principal angles necessarily vanish, and for $N = k$, they all vanish. This procedure is still valid if the subspaces have different dimensions. The projection operator can be defined as well as its transpose; and the eigenvalues of M^tM are real as this is a symmetric square matrix.

In our optimization applications, we always proceed first with a reduction in size of the search space using a sampling reduction size algorithm [28]. This makes the calculation of the whole set of eigenvalues feasible in terms of computing complexity. However, if this is out of reach, one can evaluate the bounds on the angles to see the global pertinence of our reduced order models and gradient approximations. This can be done without an exact calculation of all the eigenvalues. It is sufficient to use the Gershgorin circle theorem to find the bounds, as every eigenvalue of M^tM lies within at least one of the Gershgorin discs $D((M^tM)_{ii}, R_i)$ centred on $(M^tM)_{ii}$ and with radius $R_i = \sum_{j \neq i} |(M^tM)_{ij}|$. And because M^tM is symmetric, the eigenvalues being real, we only consider the intersection of the discs with the horizontal axis. Alternatively, the largest and smallest principal angles can be found using the iterative power and the inverse power methods applied to M^tM.

Principal angles between multi-point search spaces are interesting to measure the pertinence of sensitivity definitions based on different models or numerics. Indeed, the design will be unaffected by a reduction in the model's complexity if the search subspaces, generated by the gradients at the sampling points of the functioning parameter interval and their approximations, remain the same. This is therefore an original quantification tool for epistemic uncertainties.

7. Inversion for Uncertain Data

Let us expand the class of problems introduced in Section 3 to the following

$$\min_{\mathbf{x}\in\mathbf{O}_{ad}} j(\mathbf{x},\alpha,\mathbf{u}^*), \qquad \mathbf{u}^*\in\mathbb{R}^p, \qquad \alpha\in\mathbf{I}\subset\mathbb{R}^n, \qquad \mathbf{O}_{ad}\subset\mathbb{R}^N, \tag{6}$$

where $\mathbf{u}^*$ represents either measurements or state estimations which is a vector of random variables. We are interested in functionals j of the following form

$$j(\mathbf{x},\alpha,\mathbf{u}^*) = \tilde{j}(\mathbf{x},\alpha) + \frac{1}{2}\|\Pi\mathbf{u}(\mathbf{x},\alpha) - u^*(\alpha)\|^2. \tag{7}$$

The first term is what has been discussed up to now. Operator $\Pi\colon \mathbb{R}^N \to \mathbb{R}^p$ (typically, a linear interpolation operator) makes the state available at data locations. Inverse problems are in this class [17, 45]. This formulation also permits to see the state as uncertain as a whole with Π the identity operator. One can also introduce zoning techniques to discriminate through the level of confidence one might have on the state evaluation following the variability one observes in practice.

To summarize, we assume the components of $\mathbf{u}^*$ to be independent, uncertain and given by their Gaussian PDF, for instance, $\mathcal{N}(\mu_i,\sigma_i^2)$, $i=1,\dots,p$ with mean μ_i and variance of σ_i^2. $Cov_{\mathbf{u}^*}$ is therefore a diagonal matrix.

The simplest way to measure the effect of these uncertainties on the inversion result is again to proceed with Monte Carlo simulations. This implies a sampling of the variation domain of the data consistent with their PDF. This means we proceed with M independent inversions for M data sets defined by independent choices compatible with the PDF of $\mathbf{u}^*$ given by

$$\mathcal{N}(\mu_i,\sigma_i^2) \to (\mathbf{u}_i^*)^m, \qquad i=1,\dots,p, \qquad m=1,\dots,M.$$

These independent inversions will produce M optimal control parameters $\mathbf{x}_{opt}^m$, $m=1,\dots,M$, from which statistical moments can be defined (typically, the mean and variance) with again a rate of convergence in $M^{-1/2}$ being independent of p. Such generation of scenarios is already very demanding when involving only a direct simulation chain. In our problem, each of the scenarios involves an inversion, each requiring several solutions of the direct and adjoint problems. This complexity makes the approach be clearly out of reach, even if the calculations were independent and could be carried out in parallel.

7.1. Low-complexity uncertainty evaluation

In the sequel, we discuss two low-complexity constructions of $Cov_{\mathbf{x}}$, the covariance matrix of the control parameters from $Cov_{\mathbf{u}^*}$, the covariance matrix of the data. We want these constructions to have a cost comparable to a deterministic inversion and, again, we want to avoid any sampling of a large dimension space.

8. α-Quantile

Consider a random variable v with its PDF known (either analytic or tabulated). The tail of the PDF can be characterized defining, for a given probability level ($0 < a < 1$), the following threshold value:

$$\mathrm{VaR}_a = \inf\{l \in \mathbb{R} : P(v > l) \leq 1 - a\}.$$

Different a-quantiles are available. One very well known is the Value at Risk (VaR) which has been widely used in financial engineering as a measure of risk of loss on a given asset [22]. We do not need the time dependency issue here, this is however interesting as it permits to account for possible improvement of measurement accuracy as discussed in [31].

8.1. Bounding the uncertainty domain

We would like to use the concept of a-quantile (we call it in the sequel VaR) to define a closed domain of variation for the uncertain data [31]. Given a threshold $0 \leq a < 1$, a data $\mathbf{u}_i^*, i = 1, \ldots, p$, belongs to the interval $[\mu_i + \mathrm{VaR}_a^-, \mu_i + \mathrm{VaR}_a^+]$, $\mathrm{VaR}_a^- \leq 0 \leq \mathrm{VaR}_a^+$ with probability a. For Gaussian probability density functions, we have $\mathrm{VaR}_a^- = -\mathrm{VaR}_a^+$ and the values at risk are explicitly known to be

$$\mathrm{VaR}_{0.99}\big(N(0,1)\big) = 2.33 \text{ and } \mathrm{VaR}_{0.95}\big(N(0,1)\big) = 1.65,$$

and $\mathrm{VaR}_a\big(N(0,\sigma^2)\big) = \sigma^2 \mathrm{VaR}_a\big(N(0,1)\big)$. We have therefore, with probability a, an uncertainty domain for the data, given by

$$B_a(\mu) = \prod_{i=1}^{p} \left[\mu_i - 1.65\sigma_i^2, \mu_i + 1.65\sigma_i^2\right] \subset \mathbb{R}^p.$$

This is a large domain and we do not want to proceed with any sampling.

8.2. Directional Extreme Scenarios (DES)

However, using the sensitivity of the functional with respect to the data, we can identify two directional extreme sets of data corresponding to the intersection of $B_a(\mu)$ and $d = \mu + t\partial j/\partial u^*$, $t \in \mathbb{R}$. Let us call these two data sets $(\mathbf{u}^*)^{\pm}$, defined as

$$(\mathbf{u}^*)^{\pm} = \mu \pm 1.65\sigma_i^2 \left(\frac{\partial j/\partial u^*}{\|\partial j/\partial u^*\|}\right)_i. \tag{8}$$

To measure of the impact of this variability on the result of the inversion, we proceed with two minimizations with the target data given by $(\mathbf{u}^*)^{\pm}$ starting from $\mathbf{x}^* = \mathbf{x}_{opt}(\mathbf{u}^* = \mu)$. Let us call $(\mathbf{x}^*)^{\pm}$ the results of these inversions.

We assume monotonic behaviour of the outcome of the inversion with respect to the data. That is

$$\|\mathbf{x}^*(\mu) - \mathbf{x}^*(\nu)\| \nearrow \quad if \quad \|\mu - \nu\| \nearrow. \tag{9}$$

This assumption is reasonable and means that larger deviations in data sets bring larger variations in the outcome of the optimization. This also suggests that the maximum deviation for the results of the inversion due to the uncertainty of the data can be estimated by $X^{\pm} = (\mathbf{x}^*)^+ - (\mathbf{x}^*)^-$. Hence, we introduce a first approximation to the covariance matrix $Cov_{\mathbf{x}^{\pm}}$ [49] for $\mathbf{x}$ as follows

$$Cov_{\mathbf{x}^{\pm}} = \mathbb{E}\left((X^{\pm})(X^{\pm})^t\right) - \mathbb{E}(X^{\pm})\mathbb{E}(X^{\pm})^t \sim (X^{\pm})(X^{\pm})^t - \left(\overline{X^{\pm}}\right)\left(\overline{X^{\pm}}\right)^t, \tag{10}$$

with $\left(\overline{X^{\pm}}\right) = ((\mathbf{x}^*)^+ + (\mathbf{x}^*)^- - 2\mathbf{x}^*)/2$.

The monotonicity hypothesis can be checked a posteriori, at least partially, measuring the distance between $(\mathbf{x}^*)^{\pm}$ and $\mathbf{x}^* \pm \rho\nabla_{\mathbf{x}} j(\mathbf{x}^*, \mathbf{u}^*), \rho > 0$. This expression permits to identify two bounds $\rho^{\pm}$ and two intervals $[0, \rho^{\pm}]$ on which the monotonicity is verified. Larger values of parameters $\rho^{\pm}$ a posteriori enforce the hypothesis.

If one looks at optimization from the view point of controllability for dynamical systems [41,35], quantiles can be introduced in optimization algorithms [31]. The notion of over-solving appears then naturally where it becomes useless to solve accurately near an optimum when the variations in control parameters between two iterations of the optimizer fall below the uncertainties defined through a local uncertainty ball: all the points inside this ball being indeed equivalent in term of the confidence one can have on their performance.

We have presented the concept of Directional Extreme Scenarios for various applications in [29, 30, 34]. Directional Extreme Scenarios can be defined for $\mathbf{x}$ as well, considering the components of the design vector as random variables. It is indeed interesting to account for uncertainties in large dimensional spaces. We have also extended the DES considering ensemble-based simulations after casting the multi-point optimization algorithm into the Ensemble Kalman Filters (EnKF) formalism (see [32] for the details). The joint application of EnKF and DES leads to the concept of Ensemble Directional Extreme Scenarios (EDES) which provides more exhaustive possible extreme scenarios knowing the Probability Density Function of our optimization parameters.

9. From the Adjoint to the Covariance Matrix of the Optimization Parameters

Another construction of $Cov_{\mathbf{x}}$ takes advantage of our adjoint calculation leading to $\nabla_{\mathbf{x}} j$, the gradient of the functional with respect to the optimization parameters [33].

Let us recall the adjoint formulation for a generic state equation $F\left(\mathbf{u}(\mathbf{x}, \alpha)\right) = 0$. The gradient of j with respect to $\mathbf{x}$ is given by

$$\nabla_{\mathbf{x}} j = \frac{\partial j}{\partial \mathbf{x}} + \left(\left(\frac{\partial j}{\partial \mathbf{u}}\right)^t \frac{\partial \mathbf{u}}{\partial \mathbf{x}}\right)^t = \frac{\partial j}{\partial \mathbf{x}} + \left(\left(\frac{\partial j}{\partial \mathbf{u}}\right)^t \left(\frac{\partial F}{\partial \mathbf{u}}\right)^{-1} \frac{\partial F}{\partial \mathbf{x}}\right)^t = \frac{\partial j}{\partial \mathbf{x}} + \left(\mathbf{v}^t \frac{\partial F}{\partial \mathbf{x}}\right)^t,$$

where we have introduced the adjoint variable $\mathbf{v}$ to be the solution of

$$\mathbf{v}^t \frac{\partial F}{\partial u} = \left(\frac{\partial j}{\partial u}\right)^t, \tag{11}$$

and used in the algorithm given in Section 5. When the governing equations are self adjoint (i.e., $\frac{\partial F}{\partial \mathbf{u}} = \left(\frac{\partial F}{\partial \mathbf{u}}\right)^t$), one can use the corresponding solver with $\frac{\partial j}{\partial \mathbf{u}}$ as the right-hand side and simply solve:

$$\frac{\partial F}{\partial \mathbf{u}} v = \frac{\partial j}{\partial \mathbf{u}},$$

Also, if F is linear, then $\frac{\partial F}{\partial \mathbf{u}}$ is a constant operator independent of $\mathbf{u}$. The interest of the adjoint formulation is that the cost of getting $\nabla_{\mathbf{x}} j$ becomes independent of the size of $\mathbf{x}$. But, the problem with the adjoint approach is that, except for the two situations we mentioned (linear or self adjoint state equations), it needs the development (and maintenance) of a new code. This is why we use automatic differentiation when possible.

In multi-criteria problems the functional j is minimized under equality or inequality constraints $c_{i=1,\dots,q}$, and we need to solve an adjoint problem for the functional and each of the active constraints (needed to express the first order KKT conditions). This can be seen as a block diagonal matrix inversion with all blocks being similar and the right-hand side given by $\left(\partial_{\mathbf{u}} j, \partial_{\mathbf{u}} c_1, \dots, \partial_{\mathbf{u}} c_q\right)^t$ with q active constraints. Automatic differentiation in reverse mode with multiple right-hand side capacity can be used to address this problem. Otherwise, deflation techniques for linear systems with multiple right-hand sides can be applied [26, 47] taking advantage of the fact that the blocks being the same the Krylov decomposition needs to be conducted only once.

The functional j involves the least square deviation at data location between model and data, and $\partial_{\mathbf{u}} j$ in the right-hand side of (11) can be obtained by

$$\begin{aligned} j(\mathbf{x}, \mathbf{u}^*) &= \tilde{j} + \frac{1}{2}\|\Pi \mathbf{u} - \mathbf{u}^*\|^2 = \tilde{j} + \frac{1}{2}\langle \Pi\, \mathbf{u} - \mathbf{u}^*, \Pi\, \mathbf{u} - \mathbf{u}^* \rangle \\ &= \tilde{j} + \frac{1}{2}\langle \Pi^{\mathrm{t}} \Pi\, \mathbf{u}, \mathbf{u} \rangle - \langle \Pi^{\mathrm{t}} \mathbf{u}^*, \mathbf{u} \rangle + \frac{1}{2}\langle \mathbf{u}^*, \mathbf{u}^* \rangle, \end{aligned}$$

Where we have $\partial_{\mathbf{u}} j = \partial_{\mathbf{u}} \tilde{j} + \Pi^t \Pi u - \Pi^t \mathbf{u}^*$. On the other hand, the sensitivity of j with respect to the data $\partial_{\mathbf{u}^*} j$ needed in (8) is given by $\partial j / \partial \mathbf{u}^* = -(\Pi \mathbf{u} - \mathbf{u}^*)$.

With $\nabla_{\mathbf{x}} j$ at hand, let us establish another expression for the covariance matrix of $\mathbf{x}$ considered as a vector of zero-mean random variables. Denote, for simplicity, by $\mathbf{u}$ the model solution (zero-mean valued: $\mathbf{u} \leftarrow \mathbf{u} - \mu$) at the data location and suppose it is linked to the parameters through a linear model: $\boldsymbol{u} = L\mathbf{x}$. The covariance matrix for $\mathbf{u}$ is therefore

$$Cov_{\mathbf{u}} = \mathbb{E}(\mathbf{u}\mathbf{u}^t) = \mathbb{E}(L\ \mathbf{x}\mathbf{x}^t\ L^t) = L\ \mathbb{E}(\mathbf{x}\mathbf{x}^t)\ L^t = L\ Cov_{\mathbf{x}}\ L^t.$$

If the dependency of $\mathbf{u}$ with respect to the parameter $\mathbf{x}$ is nonlinear, then the analysis still holds for the linearized model. Introducing $\mathcal{J} = \partial \mathbf{u} / \partial \mathbf{x}$, we have

$$Cov_{\mathbf{u}} = \mathcal{J}\ Cov_{\mathrm{x}}\ \mathcal{J}^t.$$

To get $Cov_{\mathbf{x}}$, we need therefore to invert this expression and because the amount of data can be large and probably impossible to exactly fit, we proceed with a least squares formulation looking for $Cov_{\mathbf{x}}$ minimizing

$$\frac{1}{2}\langle \mathcal{J}\ Cov_{\mathrm{x}}\ \mathcal{J}^t, \mathcal{J}\ Cov_{\mathrm{x}}\ \mathcal{J}^t\rangle - \langle Cov_{\mathrm{u}}, \mathcal{J}\ Cov_{\mathrm{x}}\ \mathcal{J}^t\rangle.$$

The first order optimality conditions with respect to Cov_{x} gives

$$\mathcal{J}^t\mathcal{J}\ Cov_{\mathrm{x}}\ \mathcal{J}^t\mathcal{J} - \mathcal{J}^t\ Cov_{\mathrm{u}}\ \mathcal{J} = 0,$$

which leads to

$$Cov_{\mathrm{x}} = (\mathcal{J}^t\mathcal{J})^{-1}\ \mathcal{J}^t\ Cov_{\mathrm{u}}\ \mathcal{J}\ (\mathcal{J}^t\mathcal{J})^{-1},$$

and eventually, to

$$Cov_{\mathrm{x}} = \mathcal{J}^{-1}\ Cov_{\mathrm{u}}\ \mathcal{J}^{-t} = (\mathcal{J}^t\ Cov_{\mathrm{u}}\ \mathcal{J})^{-1}. \tag{12}$$

To get Cov_{x} and knowing Cov_{u}, it is therefore sufficient to evaluate $\mathcal{J} = \partial\mathbf{u}/\partial\mathbf{x}$. The second expression in (12) is interesting as it involves the inversion of a square matrix and gives a least squares sense to the inversion of rectangular matrices. Also, if the optimization is successful and the model u and the data u^* close, then we can assume that the data are independently collected and use the covariance matrix of the observation instead of $Cov_{\mathbf{u}}$, that is

$$Cov_{\mathrm{u}} \sim Cov_{\mathrm{u}^*},$$

which is then diagonal and its inversion is straightforward.

The question is, therefore, how to efficiently evaluate $\mathcal{J} = \partial\mathbf{u}/\partial\mathbf{x}$. The model at data locations $\Pi\mathbf{u}$ is obtained applying, for instance, a linear interpolation operator Π to the model solution $\mathbf{u}$ on the mesh. Therefore, we have

$$\mathcal{J} = \Pi\frac{\partial\mathbf{u}}{\partial\mathbf{x}}.$$

Now, recall that $\nabla_{\mathbf{x}}j$ is available and has been computed with an adjoint approach. We now use it to access $\partial\mathbf{u}/\partial\mathbf{x}$ without extra calculation as follows

$$\nabla_{\mathbf{x}}j = \frac{\partial j}{\partial\mathbf{x}} + \left(\left(\frac{\partial j}{\partial\mathbf{u}}\right)^t\ \frac{\partial\mathbf{u}}{\partial\mathbf{x}}\right)^t = \frac{\partial j}{\partial\mathbf{x}} + \left(\left(\frac{\partial j}{\partial\mathbf{u}}\right)^t\ \Pi^{-1}\mathcal{J}\right)^t,$$

where the first terms in the right-hand side is zero if there is no direct dependency on $\mathbf{x}$ in j, and it is non-zero, for instance, if a Tykhonov regularization term is introduced in the functional [46]. This leads to

$$\left(\frac{\partial j}{\partial\mathbf{u}}\right)^t\ \Pi^{-1}\mathcal{J} = \left(\nabla_x j - \frac{\partial j}{\partial\mathbf{x}}\right)^t,$$

and eventually to,

$$\mathcal{J} = \Pi\left(\frac{\partial j}{\partial \mathbf{u}}\right)^{-t}\left(\nabla_x j - \frac{\partial j}{\partial \mathbf{x}}\right)^{t}. \tag{13}$$

where the components of $(\partial j/\partial \mathbf{u})^{-t}$, which is a line vector, are given by the inverse of those of $(\partial j/\partial \mathbf{u})$ and scaled by the inverse of its size in order to have $(\partial j/\partial \mathbf{u})^{-t}.(\partial j/\partial \mathbf{u}) = 1$. Alternatively, to avoid numerical difficulties with small components of $(\partial j/\partial \mathbf{u})$, (13) can again be seen in a least squares sense with the inverse of a normal matrix as follows:

$$\mathcal{J} = \Pi\left(\left(\frac{\partial j}{\partial \mathbf{u}}\right)\left(\frac{\partial j}{\partial \mathbf{u}}\right)^{t}\right)^{-1}\frac{\partial j}{\partial \mathbf{u}}\left(\nabla_x j - \frac{\partial j}{\partial \mathbf{x}}\right)^{t} \tag{14}$$

This expression involves the inverse of the information matrix $\left(\left((\partial j/\partial \mathbf{u})\right)\left((\partial j/\partial \mathbf{u})\right)^{t}\right)$. One should be aware that the condition number of this matrix can be very high. We do not discuss this issue here but typically the Bunch and Kaufman [3] algorithm should be used in order to account for this possible deficiency. In particular, if rank deficiency is detected, then the Moore-Penrose inverse should be used based on the eigenvalue decomposition of the information matrix [7].

Under the hypothesis of the validity of the physical model, this analysis gives indications on the level of backward sensitivity of the optimization parameters with respect to the model solution at data locations which is also the sensitivity with respect to the deviation between the model and data at the data locations (as the data are independent of the optimization parameters):

$$\frac{\partial \mathbf{u}}{\partial \mathbf{x}} = \frac{\partial(\mathbf{u} - \mathbf{u}^*)}{\partial \mathbf{x}}.$$

These ingredients have been used in a generic way in several applications [28, 29, 30, 33] in the presence of various sources of variability.

10.Concluding Remarks

In order to be easily integrated in engineering environments to quantify our confidence in optimal solutions without intensive sampling of large dimensional parameter spaces, a cascade of geometric uncertainty quantification concepts has been presented. The cascade is based on the application of data analysis concepts together with existing deterministic simulation platforms.

The analysis starts with the geometric characterization of global sensitivity spaces through their dimensions and relative positions by the principal angles between global search subspaces. Then, joining a multi-point descent direction and extreme values information from the probability density functions of design variables the concept of Directional Extreme Scenarios (DES) has been introduced.

The construction goes beyond DES with Ensemble Kalman Filters (EnKF) after the multi-point optimization algorithm is cast into an ensemble simulation environment. This permits to account for

the variability on the functioning parameters through the multi-point formulation and for the variability on the optimization parameters and observation data through the ensemble Kalman filter formulation.

The joint application of the EnKF and DES leads to the concept of Ensemble Directional Extreme Scenarios (EDES) which provides exhaustive possible extreme scenarios knowing the PDF of the optimization parameters, and this being achieved without a sampling of the admissible space.

The UQ cascade ends with low-complexity solutions for reverse propagation of aleatory uncertain target data in inverse design with two approximations of the covariance matrix of the optimization parameters. These provide uncertainty quantification analysis for the inversion solution with confidence margins in the design parameters in very large design spaces. The constructions also permit to account for epistemic uncertainties considering a model or a solution procedure as always being imperfect. Hence, seeing the associated error as uncertainty, these reverse propagation constructions provide a quantification of the impact of these weaknesses on the design.

Acknowledgement

The author would like to thank Prof. N. Mahdavi-Amiri and the Iranian Operational Research community for the invitation to the annual Iranian Operational Research conference.

References

[1] Conn, A. R., Scheinberg, K., and Vicente, L. N. (2009), Introduction to derivative-free optimization, Society for Industrial and Applied Mathematics.
[2] Anderson, B. and Moore, J. (1979), Optimal filtering. Prentice-Hall, Englewood Cliffs, New Jersey.
[3] Bunch, J. R. and Kaufman, L. (1997), Some stable methods for calculating inertia and solving symmetric linear systems, *Mathematics of Computation*, 31(137), 163-179.
[4] Bungartz, H. J. and Griebel, M. (2004), Sparse grids, *Acta numerica*, 13(1), 147-269.
[5] Casella, G. and Berger, R.L. (2002), Statistical inference (Vol. 2, pp. 337-472), Pacific Grove, CA: Duxbury.
[6] Cinnella, P., and Hercus, S. J. (2010), Robust optimization of dense gas flows under uncertain operating conditions, *Computers and Fluids*, 39(10), 1893-1908.
[7] Courrieu, P. (2008), Fast computation of Moore-Penrose inverse matrices, *CoRR*, abs/0804.4809, 2008.
[8] Evensen, G. (1997), Advanced data assimilation for strongly nonlinear dynamics, *Monthly weather review*, 125(6), 1342-1354.
[9] Evensen, G. (2002), Sequential data assimilation for nonlinear dynamics: the ensemble Kalman filter, In *Ocean Forecasting* (pp. 97-116). Springer, Berlin, Heidelberg.
[10] Gelb, A. (1970), Stochastic Processes and Filtering Theory, Academic Press, NY.
[11] Gelb, A. (Ed.), (1974), *Applied optimal estimation*. MIT press.
[12] Ghanem, R.G. and Doostan, A. (2006), On the construction and analysis of stochastic models: characterization and propagation of the errors associated with limited data, *Journal of Computational Physics*, 217(1), 63-81.

[13] Ghanem, R.G., and Spanos, P.D. (1991), Stochastic finite element method: Response statistics, In *Stochastic finite elements: a spectral approach* (pp. 101-119), Springer, New York, NY.
[14] Gluck, H., and Warner, F.W. (1983), Great circle fibrations of the three-sphere, *Duke Mathematical Journal*, 50(1), 107-132.
[15] Hoel, P.G. (1971), Introduction to Mathematical Statistics, John Wiley, London.
[16] Iaccarino, G. (2008), Quantification of Uncertainty in Flow Simulations Using Probabilistic Methods, VKI Lecture Serie. *Mechanical Engineering, Institute for Computational Mathematical-Engineering, Stanford University*.
[17] Chery, J., Mohammadi, B., Peyret, M. and Joulain, C. (2011), Plate rigidity inversion in southern California using interseismic GPS velocity field, *Geophysical Journal International*, 187(2), 783-796.
[18] Onez, R.O., Spooner, J.T., Maggiore, M. and Passino, K.M. (2002), Stable Adaptive Control and Estimation for Nonlinear Systems: Neural and Fuzzy Approximator Techniques. John, Wiley, New York.
[19] Jahn, J. (2004), Vector Optimization: Theory, Applications and Extensions. Springer, Berlin.
[20] Jiang, S. (1996), Angles between euclidean subspaces, *Geometricae Dedicata*, 36(2), 113-121.
[21] Jordan, C. (1875), Essay on geometry in n dimensions, *Bull. Soc. Math. France*, 3, 103-174, 1875.
[22] Jorion, Ph. (2006), Value at Risk: The New Benchmark for Managing Financial Risk, McGraw-Hill, New York.
[23] Ghil, M., Ide, K., Courtier, P. and Lorenc, A. (1997), Unified notation for data assimilation: operational, sequential and variational, *Journal of the Meteorological Society of Japan*, 75(1B), 181-189.
[24] Kalman, R.E. (1960), A new approach to linear filtering and prediction problems. *Transactions of the ASME - Journal of Basic Engineering*, 82, 35-45.
[25] Lindman, H.R. (1974), Analysis of Variance in Complex Experimental Designs. Freeman, New York.
[26] Dywayne, A., Nicely, A., Abdou, M., Morgan, R.B. and Wilcox, W. (2010), Deflated and restarted symmetric lanczos methods for eigenvalues and linear equations with multiple right-hand sides, *SIAM Journal on Scientific Computing*, 32(1), 129-149.
[27] Melchers, R.E. (1999), Structural Reliability Analysis and Prediction. John Wiley and Sons, Chichester.
[28] Mohammadi, B. (2013), Reduced sampling and incomplete sensitivity for low- complexity robust parametric optimization, *International Journal for Numerical Methods in Fluids*, 73(4), 307-322.
[29] Mohammadi, B. (2014), Principal angles between subspaces and reduced order modelling accuracy in optimization, *Structural and Multidisciplinary Optimization*, 50(2), 237-252.
[30] Mohammadi, B. (2014), Uncertainty quantification by geometric characterization of sensitivity spaces, *Computer Methods in Applied Mechanics and Engineering*, 280, 197-221.
[31] Mohammadi, B. (2014), Value at risk for confidence level quantifications in robust engineering optimization, *Optimal Control: Applications and Methods*, 35(2), 179-190.
[32] Mohammadi, B. (2015), Ensemble Kalman filters and geometric characterization of sensitivity spaces for uncertainty quantification in optimization, *Computer Methods in Applied Mechanics and Engineering*, 290, 228-249.
[33] Mohammadi, B. (2016), Backward uncertainty propagation in shape optimization, *International Journal for Numerical Methods in Fluids*, 80(5), 285-305.

[34] Mohammadi, B. and Bouchette, F. (2014), Extreme scenarios for the evolution of a soft bed interacting with a fluid using the Value at Risk of the bed characteristics, *Computers and Fluids*, 89, 78-87.
[35] Mohammadi, B. and Pironneau, O. (2010), Applied shape optimization for fluids, Oxford university press.
[36] Gallard, F., Mohammadi, B., Montagnac, M. and Meaux, M. (2015), An adaptive multipoint formulation for robust parametric optimization. *Journal of Optimization Theory and Applications*, 167(2), 693-715.
[37] Nocedal, J. and Wright, S. (2006), Numerical Optimization, Springer, NY.
[38] Obinata, G. and Anderson, B. (2000), Model Reduction for Control System Design, Springer Verlag, Berlin.
[39] Congedo, P. M., Corre, C., and Martinez, J.M. (2011), Shape optimization of an airfoil in a BZT flow with multiple-source uncertainties, *Computer Methods in Applied Mechanics and Engineering*, 200(1-4), 216-232.
[40] Qu, Z.Q. (2004), Model Order Reduction Techniques with Applications in Finite Element Analysis: With Applications in Finite Element Analysis, Springer Science and Business Media.
[41] Mohammadi, B., and Redont, P. (2009), Improving the identification of general Pareto fronts by global optimization, *Comptes Rendus Mathematique*, 347(5-6), 327-331.
[42] Schilders, W.H., Van der Vorst, H.A. and Rommes, J. (2008), Model order reduction: theory, research aspects and applications (Vol. 13), Berlin, Springer.
[43] Shonkwiler, C. (2009), Poincare duality angles for Riemannian manifolds with boundary, PhD thesis, Univ. Pennsylvania.
[44] Smolyak, S.A. (1963), Quadrature and interpolation formulas for tensor products of certain classes of functions, In *Doklady Akademii Nauk* (Vol. 148, No. 5, pp. 1042-1045), Russian Academy of Sciences.
[45] Tang, Z., and Périaux, J. (2012), Uncertainty based robust optimization method for drag minimization problems in aerodynamics. *Computer methods in applied mechanics and engineering*, 217, 12-24.
[46] Tarantola, A. (2005), Inverse problem theory and methods for model parameter estimation, Society for Industrial and Applied Mathematics.
[47] Curioni, A., Kalantzis, V., Bekas, C. and Gallopoulos, E. (2014), Accelerating data uncertainty quantification by solving linear systems with multiple right-hand sides. *Numerical Algorithms*, 62(2), 637-653.
[48] Veroy, K., and Patera, A.T. (2005), Certified real- time solution of the parametrized steady incompressible Navier–Stokes equations: rigorous reduced- basis a posteriori error bounds, *International Journal for Numerical Methods in Fluids*, 47(8- 9), 773-788.
[49] Wan, X., and Karniadakis, G.E. (2006), Multi-element generalized polynomial chaos for arbitrary probability measures. *SIAM Journal on Scientific Computing*, 28(3), 901-928.
[50] Wasserman, L. (2004), All of Statistics: A Concise Course in Statistical Inference. ISBN 0-387-40272-1, Springer.
[51] Xiu, D. (2010), Numerical methods for stochastic computations: a spectral method approach, Princeton university press.