A least-squares method for sparse low rank approximation of multivariate functions*

M. Chevreuil† R. Lebrun‡ A. Nouy†§ P. Rai†

Abstract

In this paper, we propose a low-rank approximation method based on discrete least-squares for the approximation of a multivariate function from random, noisy-free observations. Sparsity inducing regularization techniques are used within classical algorithms for low-rank approximation in order to exploit the possible sparsity of low-rank approximations. Sparse low-rank approximations are constructed with a robust updated greedy algorithm which includes an optimal selection of regularization parameters and approximation ranks using cross validation techniques. Numerical examples demonstrate the capability of approximating functions of many variables even when very few function evaluations are available, thus proving the interest of the proposed algorithm for the propagation of uncertainties through complex computational models.

1 Introduction

Uncertainty quantification has emerged as a crucial field of investigation for various branches of science and engineering. Over the last decade, considerable efforts have been made in the development of new methodologies based on a functional point of view in probability, where random outputs of simulation codes are approximated with suitable functional expansions. Typically, when considering a function $u(\xi)$ of input random parameters $\xi = (\xi_1 \dots \xi_d)$, an approximation is searched under the form $u(\xi) \approx \sum_{i=1}^{P} u_i \phi_i(\xi)$ where the $\phi_i(\xi)$ constitute a suitable basis of multiparametric functions (e.g. polynomial chaos basis).

Several methods have been proposed for the evaluation of functional expansions (see [19, 22, 17]). Non intrusive discrete projection methods allow the estimation of expansion coefficients by using evaluations of the numerical model at certain sample points, thus allowing the simple use of existing deterministic simulation codes. However the dimension P of classical approximation spaces has an exponential (or factorial) increase with dimension d and hence the computational cost becomes prohibitively high as one needs to evaluate the model for a large number of samples of the order of P. The objective is to construct an approximation of the high dimensional function u, given the fact that we have only limited information on it. We are particularly interested in the case where the dimension d is large but the "effective dimensionality" of the function is fairly small.

In order to handle high-dimensional models, we here propose a low rank tensor approximation method using a discrete least-squares approach, which exploits the tensor structure of the stochastic function spaces and the possible sparsity of low rank approximations. The underlying assumption is that the model output functional can be well approximated using sparse low-rank tensor approximations.

*This work was supported by EADS Innovation Works and the French National Research Agency (Grant TYCHE ANR-2010-BLAN-0904).

†Ecole Centrale Nantes, Université de Nantes, GeM UMR CNRS 6183, Nantes, France

‡Department of Applied Mathematics and Modeling, EADS Innovation Works, Suresnes, France

§Corresponding author (anthony.nouy@ec-nantes.fr).

Low rank approximation methods have recently been applied to many areas of scientific computing for approximating elements in high dimensional tensor spaces [15, 13, 12, 10], with several applications in uncertainty propagation [21, 6, 23, 14, 20]. In the context of uncertainty quantification, for problems involving very high stochastic dimension d, instead of evaluating the coefficients of an expansion in a given approximation basis (e.g. polynomial chaos), function u is approximated in suitable low-rank tensor subsets of the form

$$\mathcal{M} = \left\{ v = F_{\mathcal{M}}(\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(d)}); \mathbf{p}^{(k)} \in \mathbb{R}^{n_k} \right\}$$

where $F_{\mathcal{M}}$ is a multilinear map which constitutes a parametrization of the subset $\mathcal{M}$ and the $\mathbf{p}^{(k)}$ are the parameters. Low rank tensor subsets have nice approximation properties in the sense that they are able to approximate with a good precision a large class of functions that can be encountered in practical applications. The dimension of the parametrization $\sum_{k=1}^{d} n_k$ typically grows linearly with the dimension d, thus making possible approximation in high dimension. Here, we rely on least-squares methods in order to construct approximations in these tensor subsets, using only sample evaluations of the function u. Methods based on least-squares have already been proposed in [2, 24, 8] for the construction of tensor approximations of multivariate functionals. Here, we propose an alternative construction of tensor approximations using greedy algorithms and least-squares methods with sparse regularization allowing the construction of sparse low rank approximations with only a few function evaluations.

The proposed method consists in approximating the model with a m-term representation $u_m(\xi) = \sum_{i=1}^{m} \alpha_i w_i(\xi)$ where the α_i are real coefficients and where the w_i are successively selected in a sparse low-rank (typically rank-one) tensor subset, ideally

$$\mathcal{M}^{\boldsymbol{m}\text{-sparse}} = \left\{ v = F_{\mathcal{M}}(\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(d)}); \mathbf{p}^{(k)} \in \mathbb{R}^{n_k}, \|\mathbf{p}^{(k)}\|_0 \leq m_k \right\}$$

where $\|\cdot\|_0$ is the zero "norm" counting the number of non zero coefficients. Although $\mathcal{M}^{\boldsymbol{m}\text{-sparse}}$ may have a very small effective dimension $\sum_{k=1}^{d} m_k \ll \sum_{k=1}^{d} n_k$, the structure of this set makes optimization in $\mathcal{M}^{\boldsymbol{m}\text{-sparse}}$ a combinatorial problem. Therefore, we replace the ideal sparse tensor subset by

$$\mathcal{M}^{\boldsymbol{\gamma}} = \left\{ v = F_{\mathcal{M}}(\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(d)}); \mathbf{p}^{(k)} \in \mathbb{R}^{n_k}, \|\mathbf{p}^{(k)}\|_1 \leq \gamma_k \right\}$$

by introducing a convex regularization of the constraints using ℓ_1-norm. In practice, optimal approximations in subset $\mathcal{M}^{\boldsymbol{\gamma}}$ are computed using an alternating minimization algorithm that exploits the specific low dimensional parametrization of the subset $\mathcal{M}^{\boldsymbol{\gamma}}$ and that involves the solution of successive least-squares problems with sparse ℓ_1-regularization. Cross validation techniques are introduced in order to select optimal regularization parameters $\boldsymbol{\gamma}$. The progressive (greedy) construction of the m-term representation has the advantages of being adaptive and also of reducing the dimension of successive least-squares problems, thus improving the robustness of the least-squares method when only few samples are available. As a result, the proposed technique allows to approximate the response of models with a large number of random inputs even with a limited number of model evaluations. A sparse regularization technique is also used in order to retain only the most significant functions w_i, which results in an improvement of robustness of the greedy construction when only a limited number of samples are available. In this paper, we restrict the presentation to the case where successive corrections are computed in the set $\mathcal{M}$ of rank-one tensors. The extension of the methodology to other low-rank tensor subsets is straightforward.

The outline of the paper is as follows. In section 2, we introduce some basic concepts about functional approaches in uncertainty propagation. We also detail methods based on least-squares for the computation of approximate functional expansions. In section 3, we introduce the proposed sparse low rank approximation method based on regularized least-squares. Finally the ability of the proposed method to detect and exploit low rank and sparsity of functions is illustrated on numerical applications in section 4.

2 Functional representation and least-squares methods

2.1 Stochastic function spaces and their tensor structure

We here introduce the definitions of stochastic functions spaces and approximation spaces. We consider a set ξ of d random variables and we denote by $(\Xi, \mathcal{B}, P_\xi)$ the associated probability space, where $\Xi \subset \mathbb{R}^d$ and where P_ξ is the probability law of ξ. We suppose that ξ can be split into r mutually independent sets of random variables $\{\xi_k\}_{k=1}^r$, i.e. $\xi = \{\xi_1, \ldots, \xi_r\}$, where ξ_k takes values in $\Xi_k \subset \mathbb{R}^{d_k}$. We have $d = \sum_{k=1}^r d_k$. We denote by $(\Xi_k, \mathcal{B}_k, P_{\xi_k})$ the probability space associated with ξ_k, with P_{ξ_k} the probability law of ξ_k. Therefore, the probability space $(\Xi, \mathcal{B}, P_\xi)$ associated with $\xi = (\xi_1, \ldots, \xi_r)$ has a product structure with $\Xi = \times_{k=1}^r \Xi_k$ and $P_\xi = \otimes_{k=1}^r P_{\xi_k}$.

We denote by $L^2_{P_\xi}(\Xi)$ the Hilbert space of second order random variables defined on $(\Xi, \mathcal{B}, P_\xi)$, defined by

$$L^2_{P_\xi}(\Xi) = \left\{ u : y \in \Xi \mapsto u(y) \in \mathbb{R}; \int_\Xi u(y)^2 P_\xi(dy) < \infty \right\},$$

which is a tensor Hilbert space with the following tensor structure:

$$L^2_{P_\xi}(\Xi) = L^2_{P_{\xi_1}}(\Xi_1) \otimes \ldots \otimes L^2_{P_{\xi_r}}(\Xi_r).$$

We now introduce approximation spaces $\mathcal{S}^k_{n_k} \subset L^2_{P_{\xi_k}}(\Xi_k)$ with orthonormal basis $\{\phi^{(k)}_j\}_{j=1}^{n_k}$, such that

$$\mathcal{S}^k_{n_k} = \left\{ v^{(k)}(y_k) = \sum_{j=1}^{n_k} v^k_j \phi^{(k)}_j(y_k); v^k_j \in \mathbb{R} \right\} = \left\{ v^{(k)}(y_k) = \boldsymbol{\phi}^{(k)}(y_k)^T \mathbf{v}^{(k)}; \mathbf{v}^{(k)} \in \mathbb{R}^{n_k} \right\},$$

where $\mathbf{v}^{(k)}$ denotes the vector of coefficients of $v^{(k)}$ and where $\boldsymbol{\phi}^{(k)} = (\phi^{(k)}_1, \ldots, \phi^{(k)}_{n_k})^T$ denotes the vector of basis functions. An approximation space $\mathcal{S}_{\boldsymbol{n}} \subset L^2_{P_\xi}(\Xi)$ is then obtained by tensorization of approximation spaces $\mathcal{S}^k_{n_k}$:

$$\mathcal{S}_{\boldsymbol{n}} = \mathcal{S}^1_{n_1} \otimes \ldots \otimes \mathcal{S}^r_{n_r} = \left\{ v = \sum_{i \in I_{\boldsymbol{n}}} v_i \phi_i \ ; \ v_i \in \mathbb{R} \right\},$$

where $I_{\boldsymbol{n}} = \times_{k=1}^r \{1 \ldots n_k\}$ and $\phi_i(y) = (\phi^{(1)}_{i_1} \otimes \ldots \otimes \phi^{(r)}_{i_r})(y_1, \ldots, y_r) = \phi^{(1)}_{i_1}(y_1) \ldots \phi^{(r)}_{i_r}(y_r)$. An element $v = \sum_i v_i \phi_i \in \mathcal{S}_{\boldsymbol{n}}$ can be identified with the algebraic tensor $\mathbf{v} \in \mathbb{R}^{n_1} \otimes \ldots \otimes \mathbb{R}^{n_r}$ such that $(\mathbf{v})_i = v_i$. Denoting $\boldsymbol{\phi}(y) = \boldsymbol{\phi}^{(1)}(y_1) \otimes \ldots \otimes \boldsymbol{\phi}^{(r)}(y_r) \in \mathbb{R}^{n_1} \otimes \ldots \otimes \mathbb{R}^{n_r}$, we have the identification $\mathcal{S}_{\boldsymbol{n}} \simeq \mathbb{R}^{n_1} \otimes \ldots \otimes \mathbb{R}^{n_r}$ with

$$\mathcal{S}_{\boldsymbol{n}} = \{ v(y) = \langle \boldsymbol{\phi}(y), \mathbf{v} \rangle; \mathbf{v} \in \mathbb{R}^{n_1} \otimes \ldots \otimes \mathbb{R}^{n_r} \},$$

where $\langle \cdot, \cdot \rangle$ denotes the canonical inner product in $\mathbb{R}^{n_1} \otimes \ldots \otimes \mathbb{R}^{n_r}$.

Here, we suppose that the approximation space $\mathcal{S}_{\boldsymbol{n}}$ is given and sufficiently rich to allow accurate representations of a large class of functions (e.g. by choosing polynomial spaces with high degree, wavelets with high resolution...). Then, the aim is to provide a method for the approximation of functions in $\mathcal{S}_{\boldsymbol{n}}$ for high dimensional applications and using only limited information on the functions. Note that in the case $r = d$, approximation space $\mathcal{S}_{\boldsymbol{n}}$ has a dimension $\prod_{k=1}^d n_k$ which grows exponentially with the dimension d, thus making impossible the numerical representation and computation of an element $v \in \mathcal{S}_{\boldsymbol{n}}$ for high dimensional applications. In order to reduce the dimensionality, a classical strategy consists in introducing approximation subspaces $\mathcal{S}_{\boldsymbol{n},p} \subset \mathcal{S}_{\boldsymbol{n}}$ that are constructed using suitable tensorization rules: $\mathcal{S}_{\boldsymbol{n},p} = \{ v = \sum_{i \in I_{\boldsymbol{n},p}} v_i \phi_i \ ; \ v_i \in \mathbb{R} \}$, where $I_{\boldsymbol{n},p} \subset I_{\boldsymbol{n}}$ is an index set which can be chosen a priori. For $r = d$, a typical construction consists in taking for $\mathcal{S}^k_{n_k}$ the space of degree p polynomials $\mathbb{P}_p(\Xi_k)$, with $\phi^{(k)}_j$ the orthogonal polynomial of degree $j-1$, and for $I_{\boldsymbol{n},p} = \{ i \in I_{\boldsymbol{n}}; \sum_{k=1}^d (i_k - 1) \leq p \}$. Thus, $\mathcal{S}_{\boldsymbol{n},p}$ appears to be the so called

polynomial chaos composed of multidimensional polynomials with total degree less than p [11, 26]. Other tensorization strategies have been proposed, based on a priori knowledge on the solution or based on adaptive strategies. In the present work, we are interested in alternative methods that try to approximate high dimensional functions using low-rank approximations. They will be introduced in section 3.

2.2 Least-squares methods

We here consider the case of a real-valued model output $u : \Xi \to \mathbb{R}$. We denote by $\{y^q\}_{q=1}^Q \subset \Xi$ a set of Q samples of ξ, and by $\{u(y^q)\}_{q=1}^Q \subset \mathbb{R}$ the corresponding function evaluations. We suppose that an approximation space $\mathcal{S}_P = span\{\phi_i\}_{i=1}^P$ is given. Classical least-squares method for the construction of an approximation $u_P \in \mathcal{S}_P$ then consists in solving the following problem:

$$\|u - u_P\|_Q^2 = \min_{v \in \mathcal{S}_P} \|u - v\|_Q^2 \quad \text{with} \quad \|u\|_Q^2 = \frac{1}{Q} \sum_{q=1}^{Q} u(y^q)^2. \tag{1}$$

Note that $\|\cdot\|_Q$ only defines a semi-norm on $L^2_{P_\xi}(\Xi)$ but it may define a norm on the finite dimensional subspace $\mathcal{S}_P$ if we have a sufficient number Q of model evaluations. A necessary condition is $Q \geq P$. However, this condition may be unreachable in practice for high dimensional stochastic problems and usual a priori (non adapted) construction of approximation spaces $\mathcal{S}_P$. Moreover, classical least-squares method may yield bad results because of ill-conditioning (solution very sensitive to samples). A way to circumvent these issues is to introduce a regularized least-squares functional:

$$\mathcal{J}^\lambda(v) = \|u - v\|_Q^2 + \lambda \mathcal{L}(v), \tag{2}$$

where $\mathcal{L}$ is a regularization functional and where λ refers to some regularization parameter. The regularized least-squares problem then consists in solving

$$\mathcal{J}^\lambda(u_P^\lambda) = \min_{v \in \mathcal{S}_P} \mathcal{J}^\lambda(v). \tag{3}$$

Denoting by $\mathbf{v} = (v_1, \ldots, v_P)^T \in \mathbb{R}^P$ the coefficients of an element $v = \sum_{i=1}^P v_i \phi_i \in \mathcal{S}_P$, we can write

$$\|u - v\|_Q^2 = \|\mathbf{z} - \mathbf{\Phi}\mathbf{v}\|_2^2, \tag{4}$$

with $\mathbf{z} = (u(y^1), \ldots, u(y^Q))^T \in \mathbb{R}^Q$ the vector of random evaluations of $u(\xi)$ and $\mathbf{\Phi} \in \mathbb{R}^{Q \times P}$ the matrix with components $(\mathbf{\Phi})_{q,i} = \phi_i(y^q)$. We can then introduce a function $L : \mathbb{R}^P \to \mathbb{R}$ such that $\mathcal{L}(\sum_i v_i \phi_i) = L(\mathbf{v})$, and a function $J^\lambda : \mathbb{R}^P \to \mathbb{R}$ such that $\mathcal{J}^\lambda(\sum_i v_i \phi_i) = J^\lambda(\mathbf{v}) = \|\mathbf{z} - \mathbf{\Phi}\mathbf{v}\|_2^2 + \lambda L(\mathbf{v})$. An algebraic version of least-squares problem (3) can then be written as follows:

$$\min_{\mathbf{v} \in \mathbb{R}^P} \|\mathbf{z} - \mathbf{\Phi}\mathbf{v}\|_2^2 + \lambda L(\mathbf{v}). \tag{5}$$

Regularization introduces additional information such as smoothness, sparsity, etc. Under some assumptions on the regularization functional $\mathcal{L}$, problem (3) may have a unique solution. However, the choice of regularization strongly influences the quality of the obtained approximation. Another significant component of solving (5) is the choice of regularization parameter λ. In this paper, we use cross validation for the selection of an optimal value of λ.

2.3 Sparse regularization

Over the last decade, sparse approximation methods have been extensively studied in different scientific disciplines. A sparse function is one that can be represented using few non zero terms when

expanded on a suitable basis. In the context of uncertainty quantification, if a stochastic function is known to be sparse on a particular function basis, e.g. polynomial chaos (or tensor basis), sparse regularization methods can be used for quasi optimal recovery with only a few sample evaluations. In general, a successful reconstruction of sparse solution vector depends on *sufficient* sparsity of the coefficient vector and on additional technical properties (e.g. incoherence). This strategy has been found to be effective for non-adapted sparse approximation of the solution of PDEs [3, 7].

More precisely, an approximation $\sum_{i=1}^{P} u_i \phi_i(\xi)$ of a function $u(\xi)$ is considered as sparse on a particular basis $\{\phi_i(\xi)\}_{i=1}^{P}$ if it admits a good approximation with only a few non zero coefficients. Under certain conditions, a sparse approximation can be computed accurately using only $Q \ll P$ random samples of $u(\xi)$ via sparse regularization.

Given the random samples $\mathbf{z} \in \mathbb{R}^Q$ of the function $u(\xi)$, a best m-sparse (or m-term) approximation of u can be ideally obtained by solving the constrained optimization problem

$$\min_{\mathbf{v}\in\mathbb{R}^P} \|\mathbf{z} - \mathbf{\Phi}\mathbf{v}\|_2^2 \quad \text{subject to} \quad \|\mathbf{v}\|_0 \leq m, \tag{6}$$

where $\|\mathbf{v}\|_0 = \#\{i \in \{1,\dots,P\} \ : \ v_i \neq 0\}$ is the so called ℓ_0-"norm" of $\mathbf{v}$ which gives the number of non zero components of $\mathbf{v}$. Problem (6) is a combinatorial optimization problem which is NP hard to solve. Under certain assumptions, problem (6) can be reasonably well approximated by the following constrained optimization problem which introduces a convex relaxation of the ℓ_0-"norm":

$$\min_{\mathbf{v}\in\mathbb{R}^P} \|\mathbf{z} - \mathbf{\Phi}\mathbf{v}\|_2^2 \quad \text{subject to} \quad \|\mathbf{v}\|_1 \leq \delta, \tag{7}$$

where $\|\mathbf{v}\|_1 = \sum_{i=1}^{P} |v_i|$ is the ℓ_1-norm of $\mathbf{v}$. Since the ℓ_2 and ℓ_1-norms are convex, we can equivalently consider the following convex optimization problem, known as Lasso [25] or basis pursuit [5]:

$$\min_{\mathbf{v}\in\mathbb{R}^P} \|\mathbf{z} - \mathbf{\Phi}\mathbf{v}\|_2^2 + \lambda\|\mathbf{v}\|_1, \tag{8}$$

where $\lambda > 0$ corresponds to a Lagrange multiplier whose value is related to δ. Problem (8) appears as a regularized least-squares problem. The ℓ_1-norm is a sparsity inducing regularization function in the sense that the solution $\mathbf{v}$ of (8) may contain components which are exactly zero. Several optimization algorithms have been proposed for solving (8) (see [1]). In this paper, we use the Lasso modified least angle regression algorithm (see LARS presented in [9]) that provides a set of N_r solutions, namely the regularization path, with increasing ℓ_1-norm. Let $\mathbf{v}^j$, with $j = 1,\dots,N_r$, denote this set of solutions, $A_j \subset \{1,\dots,P\}$ be the index set corresponding to non zero coefficients of $\mathbf{v}^j$, $\mathbf{v}^j_{A_j} \in \mathbb{R}^{\#A_j}$ the vector of the coefficients A_j of $\mathbf{v}^j$, and $\mathbf{\Phi}_{A_j} \in \mathbb{R}^{Q\times\#A_j}$ the submatrix of $\mathbf{\Phi}$ obtained by extracting the columns of $\mathbf{\Phi}$ corresponding to indices A_j. The optimal solution $\mathbf{v}$ is then selected using the fast leave-one-out cross validation error estimate [4] which relies on the use of the Sherman-Morrison-Woodbury formula (see [3] for its implementation within Lasso modified LARS algorithm). Algorithm 1 briefly outlines the cross validation procedure for the selection of the optimal solution. In this work, we have used Lasso modified LARS implementation of SPAMS software [18] for ℓ_1-regularization.

Algorithm 1 Algorithm to determine optimal LARS solution using leave-one-out cross validation.

Input: sample vector $\mathbf{z} \in \mathbb{R}^Q$ and matrix $\mathbf{\Phi} \in \mathbb{R}^{Q \times P}$
Output: vector of coefficients $\mathbf{v} \in \mathbb{R}^P$

1: Run the Lasso modified LARS procedure to obtain N_r solutions $\mathbf{v}^1, \dots, \mathbf{v}^{N_r}$ of the regularization path, with corresponding sets of non zeros coefficients $A_1, \dots, A_{N_r}$.
2: **for** $j = 1, \dots, N_r$ **do**
3: Recompute the non zero coefficients $\mathbf{v}^j_{A_j}$ of $\mathbf{v}^j$ using ordinary least-squares: $\mathbf{v}^j_{A_j} = \arg\min_{\mathbf{v} \in \mathbb{R}^{\#A_j}} \|\mathbf{z} - \mathbf{\Phi}_{A_j}\mathbf{v}\|_2^2$
4: Compute $h_q = (\mathbf{\Phi}_{A_j}(\mathbf{\Phi}^T_{A_j}\mathbf{\Phi}_{A_j})^{-1}\mathbf{\Phi}^T_{A_j})_{qq}$ using Sherman-Morrison-Woodbury formula.
5: Compute relative leave-one-out error $\epsilon_j = \frac{1}{Q}\sum_{q=1}^Q \left(\frac{(\mathbf{z})_q - (\mathbf{\Phi}_{A_j}\mathbf{v}^j_{A_j})_q}{(1-h_q)\hat{\sigma}(\mathbf{z})}\right)^2$, where $\hat{\sigma}(\mathbf{z})$ is the empirical standard deviation of $\mathbf{z}$.
6: **end for**
7: Return the optimal solution $\mathbf{v}$ such that $\mathbf{v}_{A_{j^*}} = \mathbf{v}^{j^*}_{A_{j^*}}$ with $j^* = \arg\min_j \epsilon_j$.

3 Sparse low-rank tensor approximations based on least-squares method

The aim is to find a sparse low rank approximation of a function $u(\xi)$ in the finite dimensional tensor space $\mathcal{S}_{\boldsymbol{n}} = \mathcal{S}^1_{n_1} \otimes \ldots \otimes \mathcal{S}^r_{n_r}$. The proposed method relies on a greedy algorithm where successive corrections of the approximation are computed in small low-rank tensor subsets using least-squares with sparsity inducing regularization. Here, we restrict the presentation to the case where successive corrections are computed in the elementary set of rank-one tensors $\mathcal{R}_1$, thus resulting in the construction of a low-rank canonical approximation of the solution. However, the methodology could be naturally extended in order to construct sparse representations in other tensor formats.

3.1 Sparse canonical tensor subsets

Let $\mathcal{R}_1$ denote the set of (elementary) rank-one tensors in $\mathcal{S}_{\boldsymbol{n}} = \mathcal{S}^1_{n_1} \otimes \ldots \otimes \mathcal{S}^r_{n_r}$, defined by

$$\mathcal{R}_1 = \left\{ w(y) = \left(\otimes_{k=1}^r w^{(k)}\right)(y) = \prod_{k=1}^r w^{(k)}(y_k) \ ; \ w^{(k)} \in \mathcal{S}^k_{n_k} \right\},$$

or equivalently by

$$\mathcal{R}_1 = \left\{ w(y) = \langle \boldsymbol{\phi}(y), \mathbf{w}^{(1)} \otimes \ldots \otimes \mathbf{w}^{(r)} \rangle ; \mathbf{w}^{(k)} \in \mathbb{R}^{n_k} \right\},$$

where $\boldsymbol{\phi}(y) = \boldsymbol{\phi}^{(1)}(y_1) \otimes \ldots \otimes \boldsymbol{\phi}^{(r)}(y_r)$, with $\boldsymbol{\phi}^{(k)} = (\phi^{(k)}_1, \dots, \phi^{(k)}_{n_k})^T$ the vector of basis functions of $\mathcal{S}^k_{n_k}$, and where $\mathbf{w}^{(k)} = (w^k_1, \dots, w^k_{n_k})^T$ is the set of coefficients of $w^{(k)}$ in the basis of $\mathcal{S}^k_{n_k}$, that means $w^{(k)}(y_k) = \sum_{i=1}^{n_k} w^k_i \phi^{(k)}_i(y_k)$.

Approximation in $\mathcal{R}_1$ using classical least-squares methods possibly enables to recover a good approximation of the solution using a reduced number of samples. However, it may not be sufficient in the case where the approximation spaces $\mathcal{S}^k_{n_k}$ have high dimensions n_k, thus resulting in a manifold of rank-one elements $\mathcal{R}_1$ with high dimension $\sum_{k=1}^r n_k$. This difficulty may be circumvented by introducing approximations in a $\boldsymbol{m}$-sparse rank-one subset defined as

$$\mathcal{R}_1^{\boldsymbol{m}\text{-sparse}} = \left\{ w(y) = \langle \boldsymbol{\phi}(y), \mathbf{w}^{(1)} \otimes \ldots \otimes \mathbf{w}^{(r)} \rangle ; \mathbf{w}^{(k)} \in \mathbb{R}^{n_k}, \|\mathbf{w}^{(k)}\|_0 \leq m_k \right\}$$

with effective dimension $\sum_{k=1}^r m_k \ll \sum_{k=1}^r n_k$. As mentioned in section 2.3, performing least-squares approximation in this set may not be computationally tractable. We thus introduce a

convex relaxation of the ℓ_0-"norm" to define the subset $\mathcal{R}_1^\gamma$ of $\mathcal{R}_1$ defined as

$$\mathcal{R}_1^\gamma = \left\{ w(y) = \langle \boldsymbol{\phi}(y), \mathbf{w}^{(1)} \otimes \ldots \otimes \mathbf{w}^{(r)} \rangle ; \mathbf{w}^{(k)} \in \mathbb{R}^{n_k}, \|\mathbf{w}^{(k)}\|_1 \leq \gamma_k \right\},$$

where the set of parameters $(\mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(r)})$ is now searched in a convex subset of $\mathbb{R}^{n_1} \times \ldots \times \mathbb{R}^{n_r}$. Finally, we introduce the set of canonical rank-m tensors $\mathcal{R}_m = \{v = \sum_{i=1}^m w_i \, ; w_i \in \mathcal{R}_1\}$ and the corresponding subset

$$\mathcal{R}_m^{\gamma^1,\ldots,\gamma^m} = \left\{ v = \sum_{i=1}^m w_i \, ; w_i \in \mathcal{R}_1^{\gamma^i} \right\}.$$

In the following, we propose algorithms for the construction of approximations in tensor subsets $\mathcal{R}_1^\gamma$ and $\mathcal{R}_m^{\gamma^1,\ldots,\gamma^m}$. These subsets are low-dimensional subsets of the approximation space $\mathcal{S}_{\boldsymbol{n}}$ but are not linear spaces nor convex sets, thus making more difficult the analysis and practical resolution of optimization problems in these sets. In practice, we rely on heuristic alternating minimization algorithms presented in the following sections.

Remark 1 *Other tensor subsets have been introduced that have better approximation properties, such as Tucker tensor sets or Hierarchical Tucker tensor sets (see [13]). These tensor formats are not considered here.*

3.2 Construction of sparse rank-one tensor approximation

The subset $\mathcal{R}_1^\gamma$ can be parametrized with the set of parameters $(\mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(r)}) \in \mathbb{R}^{n_1} \times \ldots \times \mathbb{R}^{n_r}$ such that $\|\mathbf{w}^{(k)}\|_1 \leq \gamma_k$ $(k = 1, \ldots, r)$, this set of parameters corresponding to an element $w = \otimes_{k=1}^r w^{(k)}$ where $\mathbf{w}^{(k)}$ denotes the vector of coefficients of an element $w^{(k)}$. With appropriate choice of bases, a sparse rank-one function w could be well approximated using vectors $\mathbf{w}^{(k)}$ with only a few non zero coefficients.

We compute a rank-one approximation $w = \otimes_{k=1}^r w^{(k)} \in \mathcal{R}_1^\gamma$ of v by solving the least-squares problem

$$\min_{w \in \mathcal{R}_1^\gamma} \|v - w\|_Q^2 = \min_{\substack{\mathbf{w}^{(1)} \in \mathbb{R}^{n_1}, \ldots, \mathbf{w}^{(r)} \in \mathbb{R}^{n_r} \\ \|\mathbf{w}^{(1)}\|_1 \leq \gamma_1, \ldots, \|\mathbf{w}^{(r)}\|_1 \leq \gamma_r}} \|v - \langle \boldsymbol{\phi}, \mathbf{w}^{(1)} \otimes \ldots \otimes \mathbf{w}^{(r)} \rangle\|_Q^2. \tag{9}$$

Problem (9) can be equivalently written

$$\min_{\mathbf{w}^{(1)} \in \mathbb{R}^{n_1}, \ldots, \mathbf{w}^{(r)} \in \mathbb{R}^{n_r}} \|v - \langle \boldsymbol{\phi}, \mathbf{w}^{(1)} \otimes \ldots \otimes \mathbf{w}^{(r)} \rangle\|_Q^2 + \sum_{k=1}^r \lambda_k \|\mathbf{w}^{(k)}\|_1, \tag{10}$$

where the values of the regularization parameters $\lambda_k > 0$ (interpreted as Lagrange multipliers) are related to γ_k. In practice, minimization problem (10) is solved using an alternating minimization algorithm which consists in successively minimizing over $\mathbf{w}^{(j)}$ for fixed values of $\{\mathbf{w}^{(k)}\}_{k \neq j}$. Denoting by $\mathbf{z} \in \mathbb{R}^Q$ the vector of samples of function $v(\xi)$ and by $\boldsymbol{\Phi}^{(j)} \in \mathbb{R}^{Q \times n_j}$ the matrix whose components are $(\boldsymbol{\Phi}^{(j)})_{qi} = \phi_i^j(y_j^q) \prod_{k \neq j} w^{(k)}(y_k^q)$, the minimization problem on $w^{(j)}$ can be written

$$\min_{\mathbf{w}^{(j)} \in \mathbb{R}^{n_j}} \|\mathbf{z} - \boldsymbol{\Phi}^{(j)} \mathbf{w}^{(j)}\|_2^2 + \lambda_j \|\mathbf{w}^{(j)}\|_1, \tag{11}$$

which has a classical form of a least-squares problem with a sparsity inducing ℓ_1-regularization. Problem (11) is solved using the Lasso modified LARS algorithm where the optimal solution is selected using the leave-one-out cross validation procedure presented in Algorithm 1. Algorithm 2 outlines the construction of a sparse rank one approximation.

Algorithm 2 Algorithm to compute sparse rank one approximation of a function v.

Input: vector of evaluations $\mathbf{z} = (v(y^1), \ldots, v(y^Q))^T \in \mathbb{R}^Q$.
Output: rank-one approximation $w(y) = \langle \boldsymbol{\phi}(y), \mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(r)} \rangle$.
1: Initialize the vectors $\{\mathbf{w}^{(k)}\}_{k=1}^r$ and set $l = 0$.
2: $l \leftarrow l + 1$.
3: **for** $j = 1, \ldots, r$ **do**
4: Evaluate matrix $\boldsymbol{\Phi}^{(j)}$.
5: Solve problem (11) using Algorithm 1 for input $\mathbf{z} \in \mathbb{R}^Q$ and $\boldsymbol{\Phi}^{(j)} \in \mathbb{R}^{Q \times n_j}$ to obtain $\mathbf{w}^{(j)}$.
6: **end for**
7: Compute $\hat{\mathbf{z}} = (w(y^1), \ldots, w(y^Q))^T$.
8: **if** $\|\mathbf{z} - \hat{\mathbf{z}}\|_2 > \epsilon$ and $l \leq l_{max}$ **then**
9: Go to Step 2.
10: **end if**
11: Return the solution parameters $\mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(r)}$.

Remark 2 (Other types of regularization) *Different rank-one approximations can be defined by changing the type of regularization and constructed by replacing step 5 of Algorithm 2. First, one can consider a simple least-squares without regularization, namely Ordinary Least Square (OLS), by replacing step 5 by the solution of*

$$\min_{\mathbf{w}^{(j)} \in \mathbb{R}^{n_j}} \|\mathbf{z} - \boldsymbol{\Phi}^{(j)} \mathbf{w}^{(j)}\|_2^2. \tag{12}$$

Also, one can consider a regularization using ℓ_2-norm (ridge regression) by replacing step 5 by the solution of

$$\min_{\mathbf{w}^{(j)} \in \mathbb{R}^{n_j}} \|\mathbf{z} - \boldsymbol{\Phi}^{(j)} \mathbf{w}^{(j)}\|_2^2 + \lambda_j \|\mathbf{w}^{(j)}\|_2^2 \tag{13}$$

with a selection of optimal parameter λ_j using standard cross-validation (typically k-fold cross-validation). The approximations obtained with these different variants will be compared in the numerical examples of section 4.

3.3 Updated greedy construction of sparse rank-M approximation

We now wish to construct a sparse rank-M approximation $u_M \in \mathcal{R}_M$ of u of the form $u_M = \sum_{m=1}^M \alpha_m w_m$ by successive computations of sparse rank-one approximations $w_m = \otimes_{k=1}^r w_m^{(k)}$. We start by setting $u_0 = 0$. Then, knowing an approximation u_{m-1} of u, we proceed as follows.

Sparse rank-1 correction step. We first compute a correction $w_m \in \mathcal{R}_1$ of u_{m-1} by solving

$$\min_{w \in \mathcal{R}_1^\gamma} \|u - u_{m-1} - w\|_Q^2, \tag{14}$$

which can be reformulated as

$$\min_{w \in \mathcal{R}_1} \|u - u_{m-1} - \langle \boldsymbol{\phi}, \mathbf{w}^{(1)} \otimes \ldots \otimes \mathbf{w}^{(r)} \rangle\|_Q^2 + \sum_{k=1}^r \lambda_k \|\mathbf{w}^{(k)}\|_1. \tag{15}$$

Problem (15) is solved using an alternating minimization algorithm, which consists in successive minimization problems of the form (11) where $\mathbf{z} \in \mathbb{R}^Q$ is the vector of samples of the residual $(u - u_{m-1})(\xi)$. Optimal parameters $\{\lambda_k\}_{k=1}^r$ is selected with cross-validation.

Updating step. Once a rank-one correction w_m has been computed, it is normalized and the approximation $u_m = \sum_{i=1}^m \alpha_i w_i$ is computed by solving a regularized least-squares problem:

$$\min_{\boldsymbol{\alpha}=(\alpha_1,\ldots,\alpha_m)\in\mathbb{R}^m} \|u - \sum_{i=1}^m \alpha_i w_i\|_Q^2 + \lambda'\|\boldsymbol{\alpha}\|_1. \tag{16}$$

This updating step allows a selection of significant terms in the canonical decomposition, that means when some α_i are found to be negligible, it yields an approximation $u_m = \sum_{i=1}^m \alpha_i w_i$ with a lower effective rank representation. The selection of parameter λ' is also done with a cross-validation technique.

Remark 3 *Note that an improved updating strategy could be introduced as follows. At step m, denoting by $w_i = \otimes_{k=1}^r w_i^{(k)}$, $1 \leq i \leq m$, the computed corrections, we can define approximation spaces $\mathcal{W}_m^k = span\{w_i^{(k)}\}_{i=1}^m \subset \mathcal{S}_{n_k}^k$ (with dimension at most m), and look for an approximation of the form $u_m = \sum_{i_1=1}^m \cdots \sum_{i_r=1}^m \alpha_{i_1\ldots i_r} \otimes_{k=1}^r w_{i_k}^{(k)} \in \otimes_{k=1}^r \mathcal{W}_m^k$ (namely Tucker tensor format). The update problem then consists in solving*

$$\min_{\boldsymbol{\alpha}=(\alpha_{i_1\ldots i_r})\in\mathbb{R}^{m\times\ldots\times m}} \|u - \sum_{i_1,\ldots,i_r} \alpha_{i_1\ldots i_r} \otimes_{k=1}^r w_{i_k}^{(k)}\|_Q^2 + \lambda'\|\boldsymbol{\alpha}\|_1, \tag{17}$$

where $\|\boldsymbol{\alpha}\|_1 = \sum_{i_1,\ldots,i_r} |\alpha_{i_1\ldots i_r}|$. This updating strategy can yield significant improvements of convergence. However, it is clearly unpractical for high dimension r since the dimension m^r of the representation grows exponentially with r. For high dimension, other types of representations should be introduced, such as hierarchical tensor representations.

Algorithm 3 details the updated greedy construction of sparse low rank approximations.

Algorithm 3 Updated greedy algorithm for sparse low rank approximation of a function u.

Input: vector of evaluations $\mathbf{z} = (u(y^1), \ldots, u(y^Q))^T \in \mathbb{R}^Q$ and maximal rank M.
Output: Sequence of approximations $u_m = \sum_{i=1}^m \alpha_i w_i$, where $w_i \in \mathcal{R}_1$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m) \in \mathbb{R}^m$.

1: Set $u_0 = 0$
2: **for** $m = 1, \ldots, M$ **do**
3: Evaluate the vector $\mathbf{z}_{m-1} = (u_{m-1}(y^1), \ldots, u_{m-1}(y^Q))^T \in \mathbb{R}^Q$
4: Compute a sparse rank-one approximation $w_m = \otimes_{k=1}^r w_m^{(k)}$ using Algorithm 2 for input vector of evaluations $\mathbf{z} - \mathbf{z}_{m-1}$.
5: Evaluate matrix $\boldsymbol{W} \in \mathbb{R}^{Q\times m}$ with components $(\boldsymbol{W})_{qi} = w_i(y^q)$.
6: Compute $\boldsymbol{\alpha} \in \mathbb{R}^m$ with Algorithm 1 for input vector $\mathbf{z} \in \mathbb{R}^Q$ and matrix $\boldsymbol{W} \in \mathbb{R}^{Q\times m}$.
7: **end for**

Rank Selection. Let us note that Algorithm 3 not only generates an approximation of the function but a sequence of approximations $\{u_m\}_{m=1}^M$ with increasing canonical ranks. To select the best rank, we use a k-fold cross validation method. The overall procedure is as follows:

- Split sample set $S = \{1, \ldots, Q\}$ into k disjoint subsamples $\{V_i\}_{i=1}^k$, $V_i \subset S$, of approximately the same size, and let $S_i = S \setminus V_i$.
- For each subsample, run Algorithm 3 from the reduced vector of evaluations $\mathbf{z}_{S_i}$ (test set) to obtain the sequence of model approximations $\{u_m^{S_i}\}_{1\leq m\leq M}$. Compute the corresponding mean squared errors $\{\varepsilon_1^{S_i}, \ldots, \varepsilon_M^{S_i}\}$ from the validation set of evaluations $\mathbf{z}_{V_i}$.
- For $m = 1, \ldots, M$, compute the k-fold cross validation error $\bar{\varepsilon}_m = \frac{1}{k}\sum_{i=1}^k \varepsilon_m^{S_i}$.
- Select optimal rank $m_{op} = \text{argmin}_{1\leq m\leq M} \bar{\varepsilon}_m$.
- Run Algorithm 3 with the whole data set $\mathbf{z}$ for computing $u_{m_{op}}$.

4 Application examples

4.1 Analytical model: Checker-board function

Function and approximation spaces. We first test Algorithm 3 on the so-called checker-board function $u(\xi_1, \xi_2)$ illustrated in figure 1. Random variables ξ_1 and ξ_2 are independent and uniformly distributed on $[0, 1]$. This function has a rank 2 and the functions of the rank-2 decomposition are plotted in figure 2.

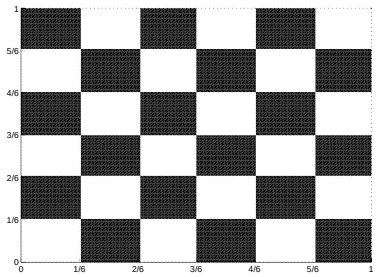
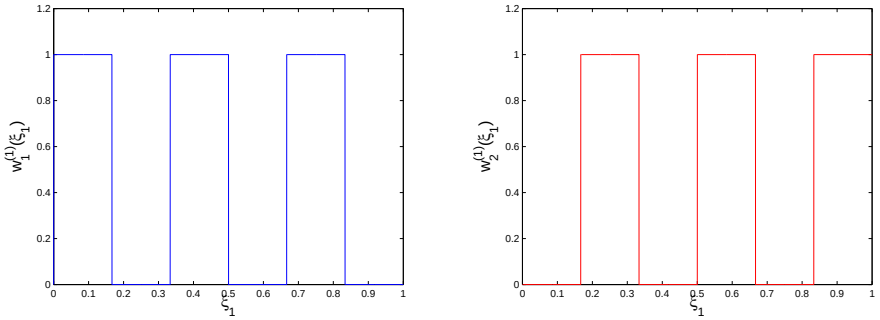


Figure 1: Checker-board function $u(\xi_1, \xi_2) = \sum_{i=1}^{2} w_i^{(1)}(\xi_1) w_i^{(2)}(\xi_2)$ (See Fig. 2).



(a) $w_1^{(1)}(\xi_1)$ (b) $w_2^{(1)}(\xi_1)$

(c) $w_1^{(2)}(\xi_2)$ (d) $w_2^{(2)}(\xi_2)$

Figure 2: Functions of the decomposition of the checker-board function $u(\xi_1, \xi_2) = \sum_{i=1}^{2} w_i^{(1)}(\xi_1) w_i^{(2)}(\xi_2)$ (rank-2).

For approximation spaces $\mathcal{S}_{n_k}^k$, $k \in \{1, 2\}$, we introduce piecewise polynomials of degree p defined on a uniform partition of Ξ_k composed by s intervals, corresponding to $n_k = s(p + 1)$. We denote by $\mathcal{S}_{n_k}^k = \mathbb{P}_{p,s}$ the corresponding space (for ex. $\mathbb{P}_{2,3}$ denotes piecewise polynomials of

degree 2 defined on the partition $\{(0, \frac{1}{3}), (\frac{1}{3}, \frac{2}{3}), (\frac{2}{3}, 1)\}$). We use an orthonormal basis composed of functions whose supports is one element of the partition and whose restriction on this support is a rescaled Legendre polynomial.

Note that when using a partition into $s = 6n$ intervals, $n \in \mathbb{N}$, then the checker-board function can be exactly represented, that means $u \in \mathbb{P}_{p,6n} \otimes \mathbb{P}_{p,6n}$ for all p and n. Also, the solution admits a sparse representation in $\mathbb{P}_{p,6n} \otimes \mathbb{P}_{p,6n}$ since an exact representation is obtained by only using piecewise constant basis functions ($u \in \mathbb{P}_{0,6n} \otimes \mathbb{P}_{0,6n}$). The effective dimensionality of the ckecker-board function is $2 \times 2 \times 6 = 24$, which corresponds to the number of coefficients required for storing the rank-2 representation of the function. We expect our algorithm to detect the low-rank of the function and also to detect its sparsity.

Results. Algorithm 3 allows the construction of a sequence of sparse rank-m approximations u_m in $\mathcal{S}^1_{n_1} \otimes \mathcal{S}^2_{n_2}$. We estimate optimal rank-m_{op} using 3-fold cross validation (See *rank selection* in Section 3.3).

In order to illustrate the accuracy of approximations in sparse low rank tensor subsets, we compare the performance of ℓ_1-regularization within the alternating minimization algorithm (step 4 of Algorithm 3) with no regularization (OLS) and the ℓ_2-regularization (see Remark 2 for the description of these alternatives). We introduce the relative error $\epsilon(u_m, u)$ which is estimated with Monte Carlo integration with $Q' = 1000$ samples:

$$\varepsilon(u_m, u) = \frac{\|u_m - u\|_{Q'}}{\|u\|_{Q'}}. \tag{18}$$

Table 1 shows the error $\varepsilon(u_{m_{op}}, u)$ obtained for the selected optimal rank m_{op}, without and with updating step 6 of Algorithm 3, and for the different types of regularization during the correction step 4 of Algorithm 3. The results are presented for a sample size $Q = 200$ and for different function spaces $\mathcal{S}^1_{n_1} = \mathcal{S}^2_{n_2} = \mathbb{P}_{p,s}$. P denotes the dimension of the space $\mathcal{S}^1_{n_1} \otimes \mathcal{S}^2_{n_2}$. We observe that, for $\mathbb{P}_{2,3}$, the solution is not sparse on the corresponding basis and ℓ_1-regularization does not provide a better solution than ℓ_2-regularization since the approximation space is not adapted. However, when we choose function spaces that are sufficiently rich for the solution to be sparse, we see that ℓ_1-regularization within the alternating minimization algorithm outperforms other types of regularization and yields low rank approximations of the function almost at the machine precision. In this example, ℓ_1-regularization allows recovering the exact rank-2 approximation of the function.

Table 1: Relative error $\varepsilon(u_{m_{op}}, u)$ and optimal rank m_{op} estimation of Checker-board function with various regularizations for $Q = 200$ samples. P is the dimension of the approximation space. ('-' indicates that none of the rank-one elements were selected during the update step).

	Ordinary Least Square				ℓ_2				ℓ_1			
	No update		Update		No update		Update		No update		Update	
Approximation space	Error	m_{op}	Error	m_{op}	Error	m_{op}	Error	m_{op}	Error	m_{op}	Error	m_{op}
$\mathcal{R}_m(\mathbb{P}_{2,3} \otimes \mathbb{P}_{2,3}), P = 9^2$	0.527	2	0.527	2	0.508	2	0.508	2	0.507	2	0.507	2
$\mathcal{R}_m(\mathbb{P}_{2,6} \otimes \mathbb{P}_{2,6}), P = 18^2$	0.664	2	0.664	2	0.061	8	0.061	8	$1.96\,10^{-12}$	4	$2.41\,10^{-13}$	2
$\mathcal{R}_m(\mathbb{P}_{2,12} \otimes \mathbb{P}_{2,12}), P = 36^2$	20.92	1	-	-	0.568	10	0.566	4	$1.93\,10^{-12}$	2	$1.50\,10^{-12}$	3
$\mathcal{R}_m(\mathbb{P}_{5,6} \otimes \mathbb{P}_{5,6}), P = 36^2$	31.27	1	-	-	0.624	10	0.623	3	$1.22\,10^{-12}$	2	$7.93\,10^{-13}$	2
$\mathcal{R}_m(\mathbb{P}_{10,6} \otimes \mathbb{P}_{10,6}), P = 66^2$	9648.8	1	-	-	0.855	10	0.855	10	$1.21\,10^{-12}$	2	$7.88\,10^{-13}$	2

From this first analytical example, several conclusions can be drawn:

- ℓ_1-regularization in alternating least square algorithm is able to detect sparsity and hence gives very accurate approximations using few samples as compared to OLS and ℓ_2-regularizations.
- Updating step selects the most pertinent rank-one elements and gives an approximation of the function with a lower effective rank.

4.2 Analytical model: Rastrigin function

For certain classes of non smooth functions, classical polynomial chaos bases may not be a good choice since they have global support and hence cannot capture local features accurately. One possible remedy is to use wavelet bases that allow the simultaneous description of global and local features [16]. In order to demonstrate the use of our algorithm with polynomial wavelet bases, we study the 2-dimensional Rastrigin function given by

$$f(\xi) = 20 + \sum_{i=1}^{2} (\xi_i^2 - 10\cos(2\pi\xi_i))$$

where ξ_1, ξ_2 are independent random variables uniformly distributed in [-4,4]. The 2-D Rastrigin function is shown in figure 3.

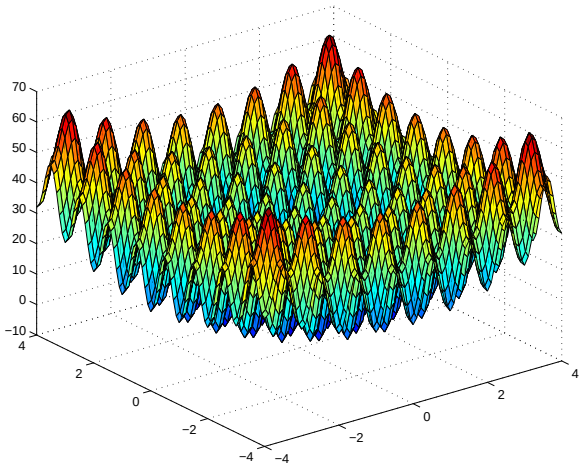


Figure 3: 2-D Rastrigin function.

We consider two types of approximation spaces $\mathcal{S}_{n_k}^k$, $k \in \{1, 2\}$:

- spaces of polynomials of degree 7, using Legendre polynomial chaos basis, denoted $\mathbb{P}_7$,
- spaces of polynomial wavelets with degree 4 and resolution level 3, denoted $\mathbb{W}_{4,3}$.

We compute a sequence of sparse rank-m approximations u_m in $\mathcal{S}_{n_1}^1 \otimes \mathcal{S}_{n_2}^2$ using Algorithm 3. We compare the approximations obtained with the different types of regularization during the correction step 4 of Algorithm 3: ℓ_1-regularization, ℓ_2-regularization and no regularization (OLS). For each case, an optimal rank approximation $u_{m_{op}}$ is selected using 3-fold cross validation (see the rank selection strategy in section 3.3). Figure 4 shows the convergence of this optimal approximation with respect to the sample size Q for the two different approximation spaces and different regularizations within the alternated minimization algorithm of the correction step. We find that the solution obtained with classical polynomial basis functions is inaccurate and does not improve with increase in sample size. Thus, polynomial basis functions are not a good choice to obtain a reasonably accurate estimate. On the other hand, when we use wavelet approximation bases, the approximation error reduces progressively with increase in sample size. Also, ℓ_1-regularization is more accurate when compared to OLS or ℓ_2-regularization, particularly for few function evaluations. We can thus conclude that a good choice of basis functions is important in order to fully realize the potential of ℓ_1-regularization in the tensor approximation algorithm.

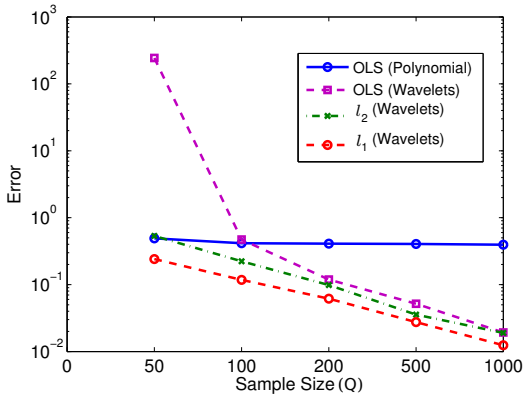


Figure 4: Evolution of error $\varepsilon(u_{m_{op}}, u)$ with respect to the number of samples Q. Approximations obtained with Algorithm 3 for different regularizations in alternating minimization algorithm (ℓ_1-regularization, ℓ_2-regularization, or no regularization (OLS)) with optimal rank selection, and for the two different approximation spaces $\mathbb{P}_7 \otimes \mathbb{P}_7$ ($P = 64$) and $\mathbb{W}_{4,3} \otimes \mathbb{W}_{4,3}$ ($P = 1600$).

Figure 5 shows the convergence of the approximation obtained with Algorithm 3 using different sample sizes. We find that as the sample size increases, we get better approximations with increasing rank. We also observe the superiority of ℓ_1-regularization compared to the ℓ_2-regularization, with a significant gain in terms of accuracy, especially when a small number of function evaluations are available.

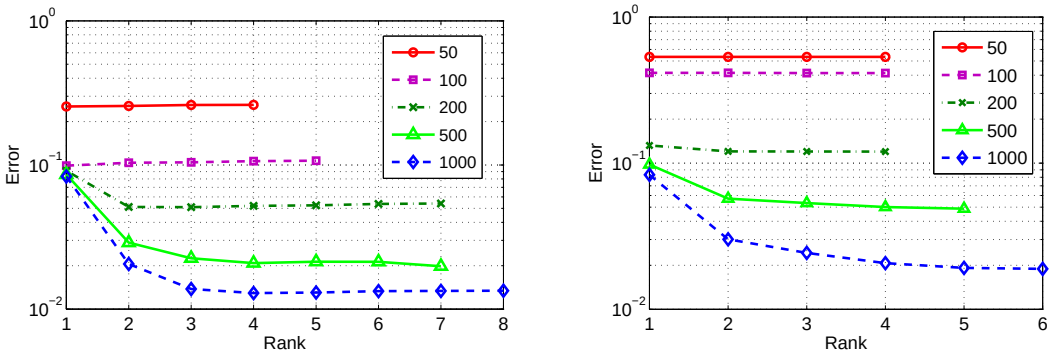


Figure 5: Evolution of error $\varepsilon(u_{m_{op}}, u)$ with respect to rank m for different sample sizes. Wavelet approximation space $\mathbb{W}_{4,3} \otimes \mathbb{W}_{4,3}$ ($P = 1600$). Approximations obtained with Algorithm 3 with ℓ_1-regularization in update step and (a) ℓ_1-regularization or (b) ℓ_2-regularization in correction step.

We finally analyze the robustness of Algorithm 3 (with ℓ_1-regularization within the alternated minimization algorithm) with respect to the sample sets. We use wavelet bases. An optimal rank approximation $u_{m_{op}}$ is selected using 3-fold cross validation (see the rank selection strategy in section 3.3). We compare this algorithm with a direct sparse least-squares approximation in the tensorized polynomial wavelet space (no low-rank approximation), using ℓ_1-regularization (use of Algorithm 1). Figure 6 shows the evolution of the relative error with respect to the sample size Q for these two strategies. The vertical lines represent the scattering of the error when different sample sets are used. We observe a smaller variance of the obtained approximations when exploiting low-rank representations. This can be explained by the lower dimensionality

of the representation, which is better estimated with a few number of samples. On this simple example, we see the interest of using low-rank representations when only a small number of samples is available. The interest of using low-rank representations should also become clear when dealing with higher dimensional problems.

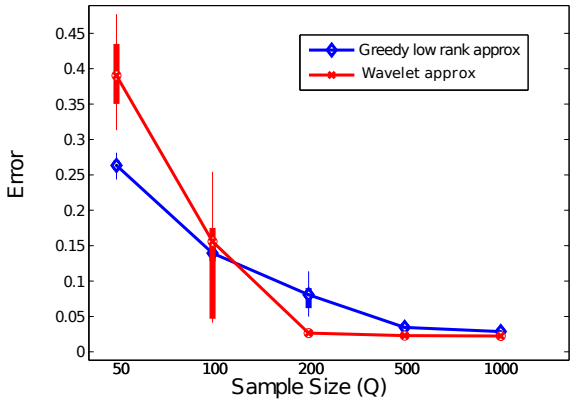


Figure 6: Evolution of error $\varepsilon(u_{m_{op}}, u)$ with respect to sample size Q. (Red line) approximation obtained with direct least-squares approximation with ℓ_1-regularization on the full polynomial wavelet approximation space $\mathbb{W}_{4,3} \otimes \mathbb{W}_{4,3}$, (blue line) approximation obtained with Algorithm 3 (with ℓ_1-regularization) and with optimal rank selection.

4.3 Diffusion equation with multiple inclusions

We consider a stationary diffusion problem defined on a two dimensional domain $\Omega = (0,1) \times (0,1)$ (see figure 7):

$$-\nabla \cdot (\kappa \nabla u) = I_D(x) \quad \text{on} \quad \Omega, \quad u = 0 \quad \text{on} \quad \partial\Omega,$$

where $D = (0.4, 0.6) \times (0.4, 0.6) \subset \Omega$ is a square domain and I_D is the indicator function of D. The diffusion coefficient is defined by

$$\kappa = \begin{cases} \xi_k & \text{on } C_k,\ 1 \leq k \leq 8 \\ 1 & \text{on } \Omega \backslash (\cup_{k=1}^{8} C_k) \end{cases},$$

where the C_k are circular domains and where the $\xi_k \sim U(0.9, 1.1)$ are independent uniform random variables. We define the quantity of interest

$$I(u)(\xi) \ = \ \int_D u(x, \xi) dx.$$

4.3.1 Influence of tensor format

We exploit two different tensor structures for the construction of low-rank representations (by regrouping some variables). First, we consider the finest tensor structure $\mathcal{S}^1 \otimes \ldots \otimes \mathcal{S}^8$. Secondly, we consider the tensor structure $\mathcal{S}^{(1,2,3,4)} \otimes \mathcal{S}^{(5,6,7,8)}$ where $\mathcal{S}^{(1,2,3,4)}$ and $\mathcal{S}^{(5,6,7,8)}$ are spaces of functions of 4 variables. We introduce polynomial approximation spaces of total degree $p = 2$, 3 or 5. In the first case, a rank-one function is the product of eight one-dimensional polynomials with degree p. In the second case, a rank-one function is the product of two 4-dimensional polynomials with total degree p.

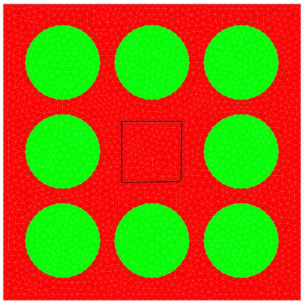


Figure 7: Diffusion problem with multiple inclusions.

As in previous examples, we compare approximations obtained with different types of regularization in the correction step of Algorithm 3. In order to better understand the performance of algorithms using different low-rank tensor structures, we determine the relative error for both rank-one approximation (table 2) and optimal rank approximation (table 3).

Table 2: Relative error $\varepsilon(u_1, u) \times 10^5$ for rank one approximations for different regularizations. $\mathbb{P}_p^{(i)}$ corresponds to a polynomial space with total degree p in i dimensions. Results are reported for $p = 2$, 3 and 5, for different sample sizes Q and for different regularizations in the greedy correction step: ordinary least square (OLS), ℓ_2-regularization and ℓ_1-regularization.

Approximation	Q=50			Q=100			Q=1000		
	OLS	ℓ_2	ℓ_1	OLS	ℓ_2	ℓ_1	OLS	ℓ_2	ℓ_1
$\mathcal{R}_1(\mathbb{P}_2^{(1)} \otimes \ldots \otimes \mathbb{P}_2^{(1)})$	2.78	2.77	2.78	2.68	2.66	2.66	2.42	2.38	2.38
$\mathcal{R}_1(\mathbb{P}_2^{(4)} \otimes \mathbb{P}_2^{(4)})$	1.90	1.84	1.72	1.19	1.22	1.26	1.03	1.04	1.01
$\mathcal{R}_1(\mathbb{P}_3^{(1)} \otimes \ldots \otimes \mathbb{P}_3^{(1)})$	2.94	2.91	2.80	2.96	2.89	2.75	2.42	2.42	2.40
$\mathcal{R}_1(\mathbb{P}_3^{(4)} \otimes \mathbb{P}_3^{(4)})$	1421	1421	18.42	16.95	15.60	1.30	1.03	1.11	0.97
$\mathcal{R}_1(\mathbb{P}_5^{(1)} \otimes \ldots \otimes \mathbb{P}_5^{(1)})$	12.20	12.45	4.45	3.16	3.17	2.97	2.42	2.42	2.40
$\mathcal{R}_1(\mathbb{P}_5^{(4)} \otimes \mathbb{P}_5^{(4)})$	-	-	-	-	-	24.5	1.15	2.14	1.36

We can derive several conclusions from table 2. We find that, for few samples, the minimum error is obtained for approximation spaces with smallest degree. Also, in these spaces, the solution obtained with variables regrouping gives a better approximation. As we may have expected, the error reduces with increase in sample size for both cases. However, we find that, with variables regrouping, OLS and ℓ_2-regularizations are unstable for approximation with high degree polynomials. Indeed for few samples and with variables regrouping, more coefficients are required to be evaluated (as compared to low rank-approximation in the finest tensor space). ℓ_1-regularization performs better by retaining few significant polynomial basis functions. This observation is particularly significant since in actual practice, we do not know in advance the degree of polynomials that would give an optimal approximation and high dimensional approximation spaces may be used if no adaptive strategy is considered. When we have a sufficient number of samples with respect to the number of basis functions in each group, variables regrouping gives better approximation than a complete separation (finest tensor structure) whereas complete separation is the most relevant when having only few samples (see $Q = 50$ and approximation in $\mathcal{R}_m(\mathbb{P}_5^{(1)} \otimes \ldots \otimes \mathbb{P}_5^{(1)})$). This observation features the existence of an optimal variables regrouping (or equivalently optimal variable separation) with respect to the number of samples and to the approximation space. A challenge would be to design automatic procedures for variables regrouping, avoiding algorithms

Table 3: Relative error $\varepsilon(u_{m_{op}}, u) \times 10^5$ for optimal rank-m_{op} approximation (with update and optimal rank selection using 3-fold cross validation) of multi-inclusion problem with various regularizations.

Approximation	Q=50						Q=100						Q=1000					
	OLS		ℓ_2		ℓ_1		OLS		ℓ_2		ℓ_1		OLS		ℓ_2		ℓ_1	
	ϵ	m_{op}	ϵ	m_{op}	ϵ	m_{op}	ϵ	m_{op}	ϵ	m_{op}	ϵ	m_{op}	ϵ	m_{op}	ϵ	m_{op}	ϵ	m_{op}
$\mathcal{R}_m(\mathbb{P}_2^{(1)} \otimes \ldots \otimes \mathbb{P}_2^{(1)})$	2.78	1	2.68	8	2.78	1	2.68	1	2.36	10	2.66	1	0.55	20	0.53	20	1.98	4
$\mathcal{R}_m(\mathbb{P}_2^{(4)} \otimes \mathbb{P}_2^{(4)})$	1.90	1	1.83	2	1.72	1	0.84	2	0.91	2	0.88	3	0.42	15	0.42	15	0.41	2
$\mathcal{R}_m(\mathbb{P}_3^{(1)} \otimes \ldots \otimes \mathbb{P}_3^{(1)})$	2.94	1	2.85	6	2.79	2	2.96	1	2.81	10	2.67	2	0.61	18	0.66	20	1.93	3
$\mathcal{R}_m(\mathbb{P}_3^{(4)} \otimes \mathbb{P}_3^{(4)})$	1230	1	1250	1	18.3	2	16.95	1	15.54	3	1.05	2	0.26	7	0.25	9	0.22	4
$\mathcal{R}_m(\mathbb{P}_5^{(1)} \otimes \ldots \otimes \mathbb{P}_5^{(1)})$	12.208	1	12.40	3	4.42	2	3.16	1	3.11	9	2.97	1	1.99	4	1.34	20	2.40	1
$\mathcal{R}_m(\mathbb{P}_5^{(4)} \otimes \mathbb{P}_5^{(4)})$	-	-	-	-	-	-	-	-	-	-	24.5	1	0.71	7	0.73	4	0.41	4

with combinatorial complexity.

Table 3 shows the effect of greedy construction for both tensor structures. With few exceptions, we find that ℓ_1-regularization is the best.

4.3.2 Strategy for the approximation of vector valued functions

We wish to obtain a reduced model for the random field $u(x,\xi) \in \mathcal{V}_N \otimes L^2_{P_{\xi_1}}(\Xi_1) \otimes \ldots \otimes L^2_{P_{\xi_d}}(\Xi_d)$, where $\mathcal{V}_N$ is a N-dimensional finite element approximation space used for the discretization of the partial differential equation. A straightforward application of the previous methodology would require to evaluate u at certain number of realizations of the input random vector ξ and to perform least-squares approximation for each node of the finite element mesh. However, this may not be feasible for large number of degrees of freedom N. Hence we wish to obtain a reasonably accurate low rank representation of the random field $u(x,\xi)$. For that purpose, we apply an empirical Karhunen-Loeve decomposition.

Let $\mathbf{u}(\xi^q) \in \mathbb{R}^N$, $q \in \{1 \ldots Q\}$, represent the finite element solutions (nodal values) associated with the Q realizations of the input random vector ξ. We denote by $\mathbf{u}_0 = \frac{1}{Q}\sum_{q=1}^{Q} \mathbf{u}(\xi^q)$ the empirical mean, and by $\widetilde{\mathbf{u}}(\xi^q) = \mathbf{u}(\xi^q) - \mathbf{u}_0$. We gather the centered realizations in a matrix $\widetilde{\boldsymbol{U}} = (\widetilde{\mathbf{u}}(\xi^1) \ldots \widetilde{\mathbf{u}}(\xi^Q))$. We compute the Singular Value Decomposition (SVD) $\widetilde{\boldsymbol{U}} = \sum_{i=1}^{N} \sigma_i \mathbf{v}_i \otimes \mathbf{z}_i$, where the $\mathbf{v}_i \in \mathbb{R}^N$ and the $\mathbf{z}_i \in \mathbb{R}^Q$ are respectively the left and right singular vectors of $\widetilde{\boldsymbol{U}}$, and the σ_i are the corresponding singular values ordered with decreasing values. Then, this decomposition is truncated by retaining only the $N^\star$ dominant singular values: $\widetilde{\boldsymbol{U}} \approx \sum_{i=1}^{N^\star} \sigma_i \mathbf{v}_i \otimes \mathbf{z}_i$. It corresponds to the following approximation of the finite element solution (truncated empirical Karhunen-Loeve decomposition):

$$\mathbf{u}(\xi) \approx \mathbf{u}_0 + \sum_{i=1}^{N^\star} \sigma_i \mathbf{v}_i z_i(\xi) \tag{19}$$

where the $z_i(\xi)$ are random variables whose evaluations at samples $\{\xi^q\}_{q=1}^{Q}$ are gathered in the vectors $\mathbf{z}_i$.

This procedure reduces the effective dimensionality of the solution field, which is now a function of only $N^\star$ random variables. Our least-squares-based tensor approximation method can be applied to each random variable z_i separately.

We perform the above procedure for multi-inclusion problem and compare the solution obtained for sample sets of three different sizes ($Q = 100, 500, 1500$). Figure 8 shows the relative error v/s the rank of the SVD decompostion of $\widetilde{\boldsymbol{U}}$ (with respect to Frobenius norm). As can be seen, the

error drops significantly after 8 modes. Thus, we choose $N^\star = 8$ and we apply the least-squares-based tensor approximation method for the random variables $\{z_i(\xi)\}_{i=1}^8$, using Algorithm 3 for each input vector $\mathbf{z}_i$ independently. We consider the finest tensor structure $\mathcal{S}^1 \otimes \ldots \otimes \mathcal{S}^8$ with polynomial spaces of total degree $p = 5$.

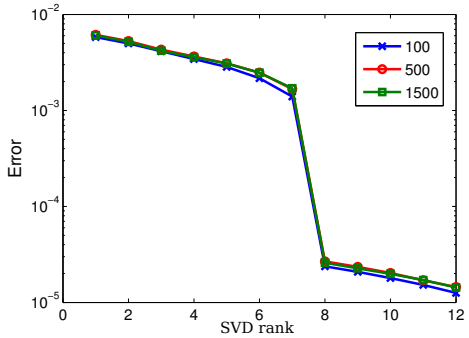


Figure 8: Error of the truncated SVD with respect to the rank for different sample size.

Figure 9 shows the cross validation error of the low rank approximations of $\{z_i(\xi)\}_{i=1}^8$ with different sample sizes. As can be expected, the cross validation error reduces with increase in sample size and we can select the best low-rank approximation (with optimal rank selection) for each mode to construct an approximation of $\mathbf{u}(\xi)$ under the form (19).

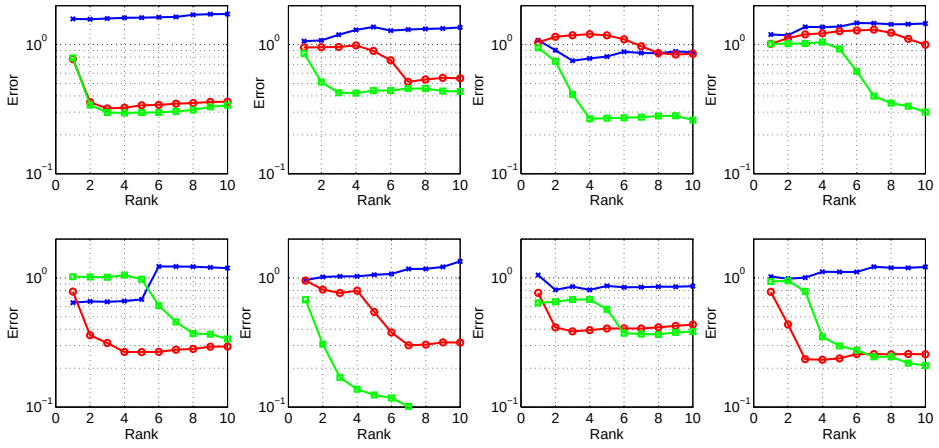


Figure 9: Evolution of the 3-fold cross validation error of the low rank approximation of the first 8 random variables $\{z_i(\xi)\}_{i=1}^8$ with the rank (left to right) for sample size 100(blue), 500(red) and 1500(green).

Let $\bar{I}$ denote the approximation of quantity of interest I (mean of the solution field in a square domain defined in section 4.3.1) obtained by the post treatment of the low rank approximation of $\mathbf{u}(\xi)$. Figure 10 shows the error $\varepsilon(\bar{I}, I)$ on the quantity of interest obtained with the reduced model for different sample sizes. The same figure also plots the empirical mean of the quantity of interest. We observe that, for few samples (i.e. 100), the mean value estimated from reduced model is not good since the corresponding approximation of the stochastic modes are inaccurate. However, as we increase the sample size, this strategy gives better approximation of stochastic modes and hence more accurate estimation of the quantity of interest.

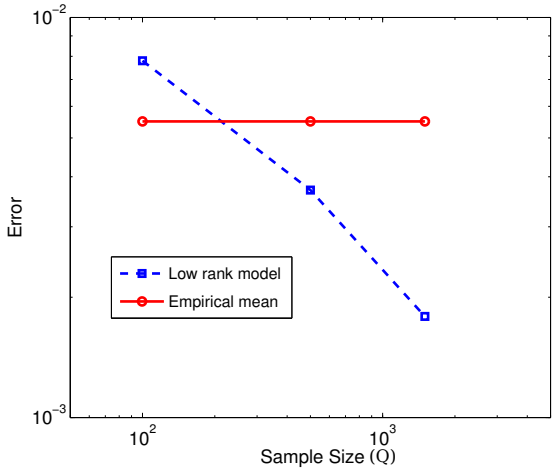


Figure 10: Evolution of the error $\varepsilon(\bar{I}, I)$ on the quantity of interest with the sample size Q and the empirical mean of I.

5 Conclusion

A non-intrusive least-squares-based sparse low-rank tensor approximation method has been proposed for propagation of uncertainty in high dimensional stochastic models. Greedy algorithms for low-rank tensor approximation have been combined with sparse least-squares approximation methods in order to obtain a robust construction of sparse low-rank tensor approximations in high dimensional approximation spaces. The ability of the proposed method to detect and exploit low-rank and sparsity was illustrated on two analytical models and on a stochastic partial differential equation (diffusion problem). The results have illustrated the significant differences between approximations obtained using different tensor structures. It reveals the need to devise algorithms that are able to automatically detect optimal low-rank tensor structures and exploit at best the few samples available in practical applications in uncertainty propagation. Also, the design of sampling strategies that are adapted to the construction of low-rank approximations could further improve the performances of these techniques.

References

[1] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, *Optimization with sparsity-inducing penalties*, Foundations and Trends in Machine Learning, 4 (2012), pp. 1–106.

[2] G. Beylkin, B. Garcke, and M.J. Mohlenkamp, *Multivariate regression and machine learning with sums of separable functions*, Journal of Computational Physics, 230 (2011), pp. 2345–2367.

[3] G. Blatman and B. Sudret, *Adaptive sparse polynomial chaos expansion based least angle regression*, Journal of Computational Physics, 230 (2011), pp. 2345–2367.

[4] G.C. Cawley and N.L.C. Talbot, *Fast exact leave-one-out cross-validation of sparse least-squares support vector machines*, Neural Networks, 17 (2004), pp. 1467–1475.

[5] S. S. Chen, D. L. Donoho, and M. A. Saunders, *Atomic decomposition by basis pursuit*, SIAM Journal on Scientific Computing, 20 (1999), pp. 33–61.

[6] A. Doostan and G. Iaccarino, *A least-squares approximation of partial differential equations with high-dimensional random inputs*, Journal of Computational Physics, 228 (2009), pp. 4332–4345.

[7] A. Doostan and H. Owhadi, *A non-adapted sparse approximation of pdes with stochastic inputs*, Journal of Computational Physics, 230 (2011), pp. 3015–3034.

[8] A. Doostan, A. Validi, and G. Iaccarino, *Non-intrusive low-rank separated approximation of high-dimensional stochastic models*, http://arxiv.org/abs/1210.1532v1, (2012).

[9] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, *Least angle regression*, The Annals of Statistics, 32 (2004), pp. 407–499.

[10] A. Falcó and A. Nouy, *Proper generalized decomposition for nonlinear convex problems in tensor banach spaces*, Numerische Mathematik, 121 (2012), pp. 503–530.

[11] R. Ghanem and P. Spanos, *Stochastic finite elements: a spectral approach*, Springer, Berlin, 1991.

[12] L. Grasedyck, D. Kressner, and C. Tobler, *A literature survey of low-rank tensor approximation techniques*, arXiv:1302.7121, (2013).

[13] W. Hackbusch, *Tensor Spaces and Numerical Tensor Calculus*, vol. 42 of Series in Computational Mathematics, Springer, 2012.

[14] B.N. Khoromskij and C. Schwab, *Tensor-structured galerkin approximation of parametric and stochastic elliptic pdes*, Tech. Report Research Report No. 2010-04, ETH, 2010.

[15] T. G. Kolda and B. W. Bader, *Tensor decompositions and applications*, SIAM Review, 51 (2009), pp. 455–500.

[16] O.P. Le Maître, O. M. Knio, H. N. Najm, and R. G. Ghanem, *Uncertainty propagation using Wiener-Haar expansions*, Journal of Computational Physics, 197 (2004), pp. 28–57.

[17] O. P. Le Maître and O. M. Knio, *Spectral Methods for Uncertainty Quantification With Applications to Computational Fluid Dynamics*, Scientific Computation, 2010.

[18] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, *Online learning for matrix factorization and sparse coding*, Journal of Machine Learning Research, 11 (2010).

[19] H. G. Matthies, *Stochastic finite elements: Computational approaches to stochastic partial differential equations*, Zamm-Zeitschrift Fur Angewandte Mathematik Und Mechanik, 88 (2008), pp. 849–873.

[20] H. G. Matthies and E Zander, *Solving stochastic systems with low-rank tensor compression*, Linear Algebra and its Applications, 436 (2012), pp. 3819–3838.

[21] A. Nouy, *A generalized spectral decomposition technique to solve a class of linear stochastic partial differential equations*, Computer Methods in Applied Mechanics and Engineering, 196 (2007), pp. 4521–4537.

[22] ——, *Recent developments in spectral stochastic methods for the numerical solution of stochastic partial differential equations*, Archives of Computational Methods in Engineering, 16 (2009), pp. 251–285.

[23] ——, *Proper generalized decompositions and separated representations for the numerical solution of high dimensional stochastic problems*, Archives of Computational Methods in Engineering, 17 (2010), pp. 403–434.

[24] P. Rai, M. Chevreuil, A. Nouy, and J. Sen Gupta, *A regression based non-intrusive method using separated representation for uncertainty quantification*, in ASME 2012, 11th Biennial Conference on Engineering Systems Design and Analysis (ESDA 2012), Nantes, France, July 2-4 2012.

[25] R. Tibshirani, *Regression shrinkage and selection via the lasso*, Journal of the Royal Statistical Society Series B, 58 (1996), pp. 267–288.

[26] D. Xiu and G. E. Karniadakis, *The Wiener-Askey polynomial chaos for stochastic differential equations*, SIAM J. Sci. Comput., 24 (2002), pp. 619–644.