



HAL
open science

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

Mathilde Chevreuril, Régis Lebrun, Anthony Nouy, Prashant Rai

► **To cite this version:**

Mathilde Chevreuril, Régis Lebrun, Anthony Nouy, Prashant Rai. A least-squares method for sparse low rank approximation of multivariate functions. *SIAM/ASA Journal on Uncertainty Quantification*, 2015, 3 (1), pp.897-921. 10.1137/13091899X . hal-00861913

HAL Id: hal-00861913

<https://hal.science/hal-00861913v1>

Submitted on 6 Jun 2024

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

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



Distributed under a Creative Commons Attribution 4.0 International License

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

M. Chevreuril[†] 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 [22, 26, 20]). 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 Airbus Group 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, Airbus Group, 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 [18, 15, 14, 12, 16], with several applications in uncertainty propagation [25, 8, 27, 17, 23]. 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

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

where $F_{\mathcal{M}}$ is a multilinear map which constitutes a parametrization of the subset \mathcal{M} and $\mathbf{p}^{(k)}$ 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, 28, 10] for the construction of tensor approximations of multivariate functions. Here, we propose an alternative construction that incorporates sparsity-inducing 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

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

where $\|\cdot\|_0$ is the “ ℓ_0 -norm” counting the number of non zero coefficients. Although $\mathcal{M}^{\text{m-sparse}}$ 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 $\mathcal{M}^{\text{m-sparse}}$ a combinatorial problem. Therefore, we replace the ideal sparse tensor subset by

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

where we introduce a convex regularization of the constraints using ℓ_1 -norm. In practice, optimal approximations in subset \mathcal{M}^{γ} are computed using an alternating minimization algorithm that exploits the specific low dimensional parametrization of the subset \mathcal{M}^{γ} 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 γ . 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. The results of this paper highlight the interest of exploiting both low-rank and sparse structures of functions for a better use of the available information on the function. In this paper, we restrict the presentation to the case where successive corrections are computed in the set \mathcal{M} of rank-one tensors. It is well known in practice that greedy rank-one constructions yield suboptimal low-rank canonical decompositions. 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. $\xi = \{\xi_1, \dots, \xi_r\}$, 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 , where P_{ξ_k} is the probability law of ξ_k . Therefore, the probability space $(\Xi, \mathcal{B}, P_\xi)$ associated with $\xi = (\xi_1, \dots, \xi_r)$ 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 $L_{P_\xi}^2(\Xi)$ the Hilbert space of second order random variables defined on $(\Xi, \mathcal{B}, P_\xi)$, defined by

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

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

$$L_{P_\xi}^2(\Xi) = L_{P_{\xi_1}}^2(\Xi_1) \otimes \dots \otimes L_{P_{\xi_r}}^2(\Xi_r).$$

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

$$\mathcal{S}_{n_k}^k = \left\{ v^{(k)}(y_k) = \sum_{j=1}^{n_k} v_j^k \phi_j^{(k)}(y_k); v_j^k \in \mathbb{R} \right\} = \left\{ v^{(k)}(y_k) = \phi^{(k)}(y_k)^T \mathbf{v}^{(k)}; \mathbf{v}^{(k)} \in \mathbb{R}^{n_k} \right\},$$

where $\mathbf{v}^{(k)}$ denotes the vector of coefficients of $v^{(k)}$ and where $\phi^{(k)} = (\phi_1^{(k)}, \dots, \phi_{n_k}^{(k)})^T$ denotes the vector of basis functions. An approximation space $\mathcal{S}_{\mathbf{n}} \subset L_{P_\xi}^2(\Xi)$ is then obtained by tensorization of approximation spaces $\mathcal{S}_{n_k}^k$:

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

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

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

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

Here, we suppose that the approximation space $\mathcal{S}_{\mathbf{n}}$ 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 $\mathcal{S}_{\mathbf{n}}$ for high dimensional applications and using only limited information on the functions. Note that in the case $r = d$, approximation space $\mathcal{S}_{\mathbf{n}}$ 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 $v \in \mathcal{S}_{\mathbf{n}}$ for high dimensional applications. In order to reduce the dimensionality, a classical strategy consists in introducing approximation subspaces $\mathcal{S}_{\mathbf{n},p} \subset \mathcal{S}_{\mathbf{n}}$ that are constructed using suitable tensorization rules: $\mathcal{S}_{\mathbf{n},p} = \{v = \sum_{i \in I_{\mathbf{n},p}} v_i \phi_i; v_i \in \mathbb{R}\}$, where $I_{\mathbf{n},p} \subset I_{\mathbf{n}}$ is an index set which can be chosen a priori. For $r = d$, a typical construction consists in taking for $\mathcal{S}_{n_k}^k$ the space of degree p polynomials $\mathbb{P}_p(\Xi_k)$, with $\phi_j^{(k)}$ the orthogonal polynomial

of degree $j - 1$, and for $I_{n,p} = \{i \in I_n; \sum_{k=1}^d (i_k - 1) \leq p\}$. Thus, $\mathcal{S}_{n,p}$ appears to be the so called polynomial chaos composed of multidimensional polynomials with total degree less than p [13, 30]. 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 $u : \Xi \rightarrow \mathbb{R}$. 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 $\mathcal{S}_P = \text{span}\{\phi_i\}_{i=1}^P$ 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:

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

Note that $\|\cdot\|_Q$ only defines a semi-norm on $L_{P,\xi}^2(\Xi)$ 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:

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

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

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

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

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

with $\mathbf{z} = (u(y^1), \dots, u(y^Q))^T \in \mathbb{R}^Q$ 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 $L : \mathbb{R}^P \rightarrow \mathbb{R}$ such that $\mathcal{L}(\sum_i v_i \phi_i) = L(\mathbf{v})$, and a function $J^\lambda : \mathbb{R}^P \rightarrow \mathbb{R}$ 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:

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

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 few sample evaluations. In general, a successful reconstruction of sparse solution vector depends on sufficient sparsity of the coefficient vector and additional properties (incoherence) depending on the samples and of the chosen basis (see [4, 7] or [9] in the context of uncertainty quantification). This strategy has been found to be effective for non-adapted sparse approximation of the solution of some PDEs [3, 9].

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

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

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”:

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

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 [29] or basis pursuit [6]:

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

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 [11]) 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 , $\mathbf{v}_{A_j}^j \in \mathbb{R}^{\#A_j}$ the vector of the coefficients A_j of \mathbf{v}^j , and $\Phi_{A_j} \in \mathbb{R}^{Q \times \#A_j}$ the submatrix of Φ obtained by extracting the columns of Φ corresponding to indices A_j . The optimal solution \mathbf{v} is then selected using the fast leave-one-out cross validation error estimate [5] 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 [21] 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 $\Phi \in \mathbb{R}^{Q \times P}$

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 $\mathbf{v}_{A_j}^j$ of \mathbf{v}^j using ordinary least-squares:

$$\mathbf{v}_{A_j}^j = \arg \min_{\mathbf{v} \in \mathbb{R}^{\#A_j}} \|\mathbf{z} - \Phi_{A_j} \mathbf{v}\|_2^2$$
 - 4: Compute $h_q = (\Phi_{A_j} (\Phi_{A_j}^T \Phi_{A_j})^{-1} \Phi_{A_j}^T)_{qq}$.
 - 5: Compute relative leave-one-out error $\epsilon_j = \frac{1}{Q} \sum_{q=1}^Q \left(\frac{(\mathbf{z})_q - (\Phi_{A_j} \mathbf{v}_{A_j}^j)_q}{(1-h_q)\hat{\sigma}(\mathbf{z})} \right)^2$, 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 $\mathbf{v}_{A_{j^*}} = \mathbf{v}_{A_{j^*}}^{j^*}$ 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 $\mathcal{S}_n = \mathcal{S}_{n_1}^1 \otimes \dots \otimes \mathcal{S}_{n_r}^r$. 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 $\mathcal{S}_n = \mathcal{S}_{n_1}^1 \otimes \dots \otimes \mathcal{S}_{n_r}^r$, defined by

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

or equivalently by

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

where $\phi(y) = \phi^{(1)}(y_1) \otimes \dots \otimes \phi^{(r)}(y_r)$, with $\phi^{(k)} = (\phi_1^{(k)}, \dots, \phi_{n_k}^{(k)})^T$ the vector of basis functions of $\mathcal{S}_{n_k}^k$, and where $\mathbf{w}^{(k)} = (w_1^k, \dots, w_{n_k}^k)^T$ is the set of coefficients of $w^{(k)}$ in the basis of $\mathcal{S}_{n_k}^k$, that means $w^{(k)}(y_k) = \sum_{i=1}^{n_k} w_i^k \phi_i^{(k)}(y_k)$.

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 $\mathcal{S}_{n_k}^k$ 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 \mathbf{m} -sparse rank-one subset defined as

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

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^γ of \mathcal{R}_1 defined as

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

where the set of parameters $(\mathbf{w}^{(1)}, \dots, \mathbf{w}^{(r)})$ is now searched in a convex subset of $\mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_r}$. 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

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

In the following, we propose algorithms for the construction of approximations in tensor subsets \mathcal{R}_1^γ and $\mathcal{R}_m^{\gamma^1, \dots, \gamma^m}$. These subsets are low-dimensional subsets of the approximation space \mathcal{S}_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 *For the construction of a rank- m approximation, we will introduce a progressive construction based on successive rank-one corrections. A direct approximation in \mathcal{R}_m would also be possible, with a straightforward extension of the algorithm presented in Section 3.2. Best approximation problem in \mathcal{R}_m for $m \geq 2$ is an ill-posed problem since \mathcal{R}_m is not closed (see e.g. [15] lemma 9.11 pg 255). However, using sparsity-inducing regularization (and also other types of regularizations) makes the best approximation problem well-posed. Indeed, it can be proven that the subset $\mathcal{R}_m^{\gamma^1, \dots, \gamma^m}$ of canonical tensors with bounded factors is a closed subset.*

Remark 2 *Other tensor subsets have been introduced that have better approximation properties, such as Tucker tensor sets or Hierarchical Tucker tensor sets (see [15] for a comprehensive review). These tensor formats are not considered here.*

3.2 Construction of sparse rank-one tensor approximation

The subset \mathcal{R}_1^γ can be parametrized with the set of parameters $(\mathbf{w}^{(1)}, \dots, \mathbf{w}^{(r)}) \in \mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_r}$ such that $\|\mathbf{w}^{(k)}\|_1 \leq \gamma_k$ ($k = 1, \dots, r$), 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

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

Problem (9) can be equivalently written

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

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 $\Phi^{(j)} \in \mathbb{R}^{Q \times n_j}$ the matrix whose components are $(\Phi^{(j)})_{qi} = \phi_i^j(y_j^q) \prod_{k \neq j} w^{(k)}(y_k^q)$, the minimization problem on $w^{(j)}$ can be written

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

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 $\mathbf{z} = (v(y^1), \dots, v(y^Q))^T \in \mathbb{R}^Q$.

Output: rank-one approximation $w(y) = \langle \phi(y), \mathbf{w}^{(1)}, \dots, \mathbf{w}^{(r)} \rangle$.

- 1: Initialize the vectors $\{\mathbf{w}^{(k)}\}_{k=1}^r$ and set $l = 0$.
 - 2: $l \leftarrow l + 1$.
 - 3: **for** $j = 1, \dots, r$ **do**
 - 4: Evaluate matrix $\Phi^{(j)}$.
 - 5: Solve problem (11) using Algorithm 1 for input $\mathbf{z} \in \mathbb{R}^Q$ and $\Phi^{(j)} \in \mathbb{R}^{Q \times n_j}$ to obtain $\mathbf{w}^{(j)}$.
 - 6: **end for**
 - 7: Compute $\hat{\mathbf{z}} = (w(y^1), \dots, w(y^Q))^T$.
 - 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 $\mathbf{w}^{(1)}, \dots, \mathbf{w}^{(r)}$.
-

Remark 3 (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 Squares (OLS), by replacing step 5 by the solution of*

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

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

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

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 $w_m = \otimes_{k=1}^r w_m^{(k)}$.

Remark 4 *This construction yields a suboptimal rank- M approximation but it has several advantages: successive minimization problems in \mathcal{R}_1 are well-posed (without any regularization), it requires the learning of a small number of parameters at each iteration. However, let us recall that a direct approximation in sparse canonical tensor format would also be possible (see remark 1).*

We start by setting $u_0 = 0$. Then, knowing an approximation u_{m-1} of u , we proceed as follows.

3.3.1 Sparse rank-1 correction step

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

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

which can be reformulated as

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

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 $\{\lambda_k\}_{k=1}^r$ are selected with the fast leave-one-out cross-validation.

3.3.2 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:

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

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 5 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 $\mathcal{W}_m^k = \text{span}\{w_i^{(k)}\}_{i=1}^m \subset \mathcal{S}_{n_k}^k$ (with dimension at most m), and look for an approximation of the form $u_m = \sum_{i_1=1}^m \dots \sum_{i_r=1}^m \alpha_{i_1 \dots i_r} \otimes_{k=1}^r w_{i_k}^{(k)} \in \otimes_{k=1}^r \mathcal{W}_m^k$ (namely Tucker tensor format). The update problem then consists in solving

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

where $\|\alpha\|_1 = \sum_{i_1, \dots, i_r} |\alpha_{i_1 \dots i_r}|$. 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 $\mathbf{z} = (u(y^1), \dots, u(y^Q))^T \in \mathbb{R}^Q$ 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 $\alpha = (\alpha_1, \dots, \alpha_m) \in \mathbb{R}^m$.

- 1: Set $u_0 = 0$
 - 2: **for** $m = 1, \dots, M$ **do**
 - 3: Evaluate the vector $\mathbf{z}_{m-1} = (u_{m-1}(y^1), \dots, u_{m-1}(y^Q))^T \in \mathbb{R}^Q$
 - 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 $\mathbf{W} \in \mathbb{R}^{Q \times m}$ with components $(\mathbf{W})_{qi} = w_i(y^q)$.
 - 6: Compute $\alpha \in \mathbb{R}^m$ with Algorithm 1 for input vector $\mathbf{z} \in \mathbb{R}^Q$ and matrix $\mathbf{W} \in \mathbb{R}^{Q \times m}$.
 - 7: **end for**
-

3.3.3 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 $S = \{1, \dots, Q\}$ 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 $\{\varepsilon_1^{S_i}, \dots, \varepsilon_M^{S_i}\}$ from the validation set of evaluations \mathbf{z}_{V_i} .
- For $m = 1, \dots, M$, compute the k -fold cross validation error $\bar{\varepsilon}_m = \frac{1}{k} \sum_{i=1}^k \varepsilon_m^{S_i}$.
- Select optimal rank $m_{op} = \operatorname{argmin}_{1 \leq m \leq M} \bar{\varepsilon}_m$.
- Run Algorithm 3 with the whole data set \mathbf{z} for computing $u_{m_{op}}$.

Remark 6 (Non greedy approximation) *A non greedy variant of Algorithm 3 for computing a rank- m approximation would consist in performing approximation directly in \mathcal{R}_m (as proposed in [10] without sparsity-inducing regularization). This strategy usually yields better approximations for the same rank m . However, the set \mathcal{R}_m for $m \geq 2$ being not closed (see [15] lemma 9.11 pg 255), minimization in \mathcal{R}_m is an ill-posed problem which usually requires regularization. The proposed construction based on successive rank-one corrections yields a suboptimal low-rank decomposition but it has several advantages: successive minimization problems in \mathcal{R}_1 are well-posed, it requires the learning of a small number of parameters at each iteration.*

4 Application examples

In this section, we validate the proposed algorithm on several benchmark problems. The purpose of the first example on Friedman function in section 4.1 is to highlight the benefit singly of the greedy low rank approximation by giving some hints on the number of samples needed for a stable approximation. The three following examples then exhibit the robustness of the l_1 -regularization within the low rank approximation:

- by correctly detecting sparsity when appropriate approximation space is introduced as in the checker board function case presented in section 4.2,
- or just by looking for the simplest representation with respect to the number of samples as in the examples of sections 4.3 and 4.4.

In all the examples, for the purpose of estimating the approximation errors, we introduce the relative error $\varepsilon(u_m, u)$ between the function u and a rank- m approximation u_m , estimated with Monte Carlo integration with $Q' = 1000$ samples:

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

Let us also define the sparsity ratio ϱ_1 of a rank-1 approximation $w = \otimes_{k=1}^d w^{(k)}$ as:

$$\varrho_1 = \frac{\sum_{k=1}^d \beta(w^{(k)})}{\sum_{k=1}^d n_k}$$

where $\beta(w^{(k)}) = \|\mathbf{w}^{(k)}\|_0$ gives the number of non zero coefficients in the vector of coefficients $\mathbf{w}^{(k)}$. In short, ϱ_1 is the ratio of total number of non zero parameters to the total number of

parameters in rank one tensor representation. We also define the total sparsity ratio ϱ_m and the partial sparsity ratio $\varrho_m^{(k)}$ in the dimension k of a rank- m approximation $u_m = \sum_{i=1}^m \otimes_{k=1}^d w_i^{(k)}$ as:

$$\varrho_m = \frac{\sum_{i=1}^m \sum_{k=1}^d \beta(w_i^{(k)})}{m \times \sum_{k=1}^d n_k} \quad \text{and} \quad \varrho_m^{(k)} = \frac{\sum_{i=1}^m \beta(w_i^{(k)})}{m \times n_k}.$$

4.1 Analytical model: Friedman function

Let us consider a simple benchmark problem namely the Friedman function of dimension $d = 5$ also considered in [2]:

$$u(\xi) = 10 \sin(\pi \xi_1 \xi_2) + 20(\xi_3 - 0.5)^2 + 10\xi_4 + 5\xi_5,$$

where $\xi_i, i = 1, \dots, 5$, are uniform random variables over $[0,1]$. In this section, we consider low rank tensor subsets without any sparsity constraint, that means we do not perform ℓ_1 regularization in the alternating minimization algorithm. The aim is to estimate numerically the number of function evaluations necessary in order to obtain a stable approximation in low rank tensor format.

The analysis on the number of function evaluations is here based upon results of [24] where it is proven that, for a stable approximation of a monovariate function with optimal convergence rate using ordinary least squares on polynomial spaces, the number of random sample evaluations required scales quadratically with the dimension of the polynomial space, i.e. $Q \sim (p+1)^2$ where p is the maximal polynomial degree. This result is supported with numerical tests on monovariate functions and on multivariate functions that show that indeed choosing Q scaling quadratically with the dimension of the polynomial space P is robust while Q scaling linearly with P may lead to an ill conditioned problem and hence an unstable approximation, depending on P and the dimension d .

The following numerical tests aim at bringing out a similar type of rule for choosing the number of samples Q for a low rank approximation of a multivariate function constructed in a greedy fashion according to Algorithm 3, but with ordinary least squares in step 5 of Algorithm 2, and given an isotropic tensor product polynomial approximation space with maximum degree p in all dimensions. We first consider a rank one approximation of the function in $\mathbb{P}_p \otimes \dots \otimes \mathbb{P}_p$ where \mathbb{P}_p denotes the polynomial approximation space of maximal degree p in each dimension from 1 to d . Given the features above and considering the algorithm for the construction of the rank one element of order d , we consider the following rule:

$$Q = cd(p+1)^\alpha$$

where c is a positive constant and $\alpha = 1$ (linear rule) or 2 (quadratic rule). In the following analyses of the current section, we plot the mean $\varepsilon(u_m, u)$ over 51 sample set repetitions in order to eliminate any dependence on the sample set of a given size. In figure 1, we compare the error of rank one approximation $\varepsilon(u_1, u)$ with respect to the Legendre polynomial degree p using both linear rule (left) and quadratic rule (right) for different values of c (ranging from 1 to 20 in the linear rule and 0.5 to 3 in quadratic rule). As could have been expected, we find that the linear rule yields a deterioration for small values of c whereas the quadratic rule gives a stable approximation with polynomial degree.

For higher rank approximations, the total number of samples needed will have a dependence on the approximation rank m . Thus we modify sample size estimates and consider the rule

$$Q = cdm(p+1)^\alpha$$

with $\alpha = 1$ (linear rule) or 2 (quadratic rule). In figure 2, we plot approximation error $\varepsilon(u_m, u)$ using linear rule (left) and quadratic rule (right) for $m = 2, 3, 4$ and different values of c . We find again that quadratic rule gives a stable approximation for $c \geq 1$.

In order to analyze the accuracy of the rank- m approximation with respect to m , in figure 3 we plot the error $\varepsilon(u_m, u)$ with respect to the polynomial degree p for different values of m using

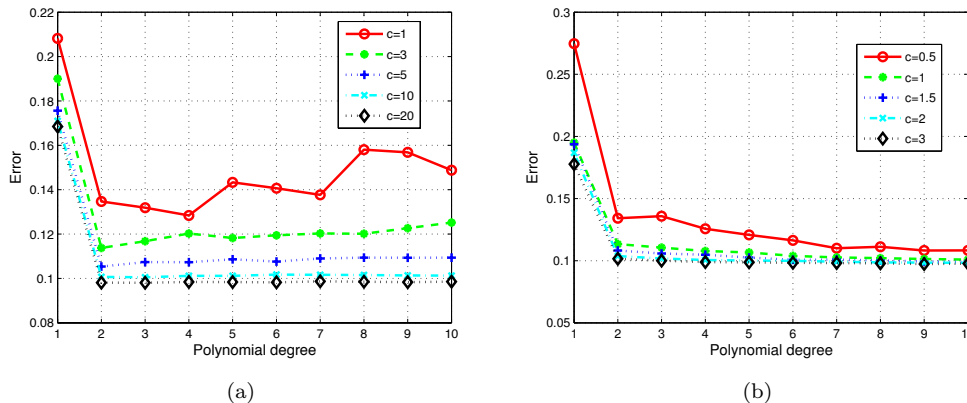


Figure 1: Evolution of rank one approximation error $\varepsilon(u_1, u)$ with respect to polynomial degree p with (a) $Q = cd(p+1)$ and (b) $Q = cd(p+1)^2$ samples with several values of c .

$Q = dm(p+1)^2$, that is $c = 1$. As the number of samples Q increases with rank m using this rule, more information on the function is given enabling for higher rank approximations to better represent the possible local features of the solution. We thus find that the approximation is more accurate as the rank increases.

From this example, we can draw the following conclusions:

- a heuristic rule to determine the number of samples needed in order to have a stable low rank approximation grows only linearly with dimension d and rank m and is given by $Q = dm(p+1)^2$,
- better solutions are obtained with high rank approximations, provided that enough model evaluations are available.

Quite often in practice, we do not have enough model evaluations and hence we may not be able to achieve good approximations with limited sample size. This is particularly true for certain classes of non smooth functions. One possible solution is a good choice of bases that are sufficiently rich (such as piecewise polynomials or wavelets) and that can capture simultaneously both global and local features of the model function. However, the sample size may not be sufficient enough to obtain good approximations with ordinary least squares in progressive rank one corrections due to large number of basis functions. We illustrate in section 4.2 and 4.3 that, in such cases, performing approximation in *sparse* low rank tensor subsets (i.e. using ℓ_1 regularization in alternating minimization algorithm) allows more accurate approximation of the model function. In addition, we illustrate in section 4.4 that approximation in sparse low rank tensor subsets leads to a relatively stable approximation with limited number of samples even for high degree polynomial spaces.

4.2 Analytical model: Checker-board function

4.2.1 Function and approximation spaces

We now test Algorithm 3 on the so-called checker-board function $u(\xi_1, \xi_2)$ of dimension $d = 2$ illustrated in figure 4. The purpose of this test is to illustrate that, given appropriate bases, in this case piecewise polynomials, Algorithm 3 allows the detection of sparsity and hence construction of a sequence of optimal sparse rank- m approximations with few samples.

Random variables ξ_1 and ξ_2 are independent and uniformly distributed on $[0, 1]$. The checker-

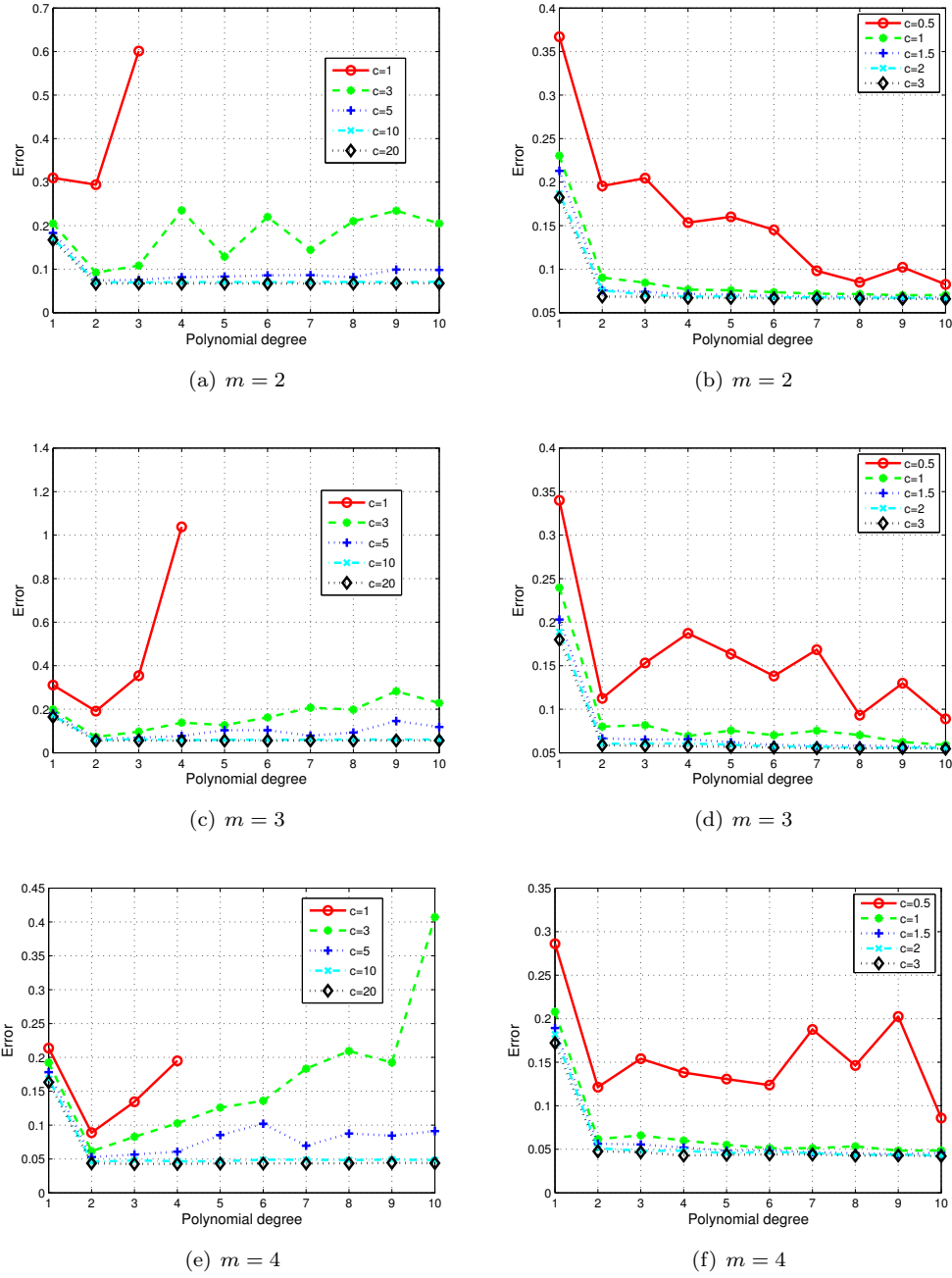


Figure 2: Evolution of approximation error $\varepsilon(u_m, u)$ with respect to polynomial degree p with $Q = cd(p+1)$ (left column) and $Q = cd(p+1)^2$ (right column) with several values of c and for $m = 2, 3, 4$.

board function is a rank-2 function

$$u(\xi_1, \xi_2) = \sum_{i=1}^2 w_i^{(1)}(\xi_1) w_i^{(2)}(\xi_2)$$

with $w_1^{(1)}(\xi_1) = c(\xi_1)$, $w_1^{(2)}(\xi_2) = 1 - c(\xi_2)$, $w_2^{(1)}(\xi_1) = 1 - c(\xi_1)$ and $w_2^{(2)}(\xi_2) = c(\xi_2)$ where $c(\xi_k)$

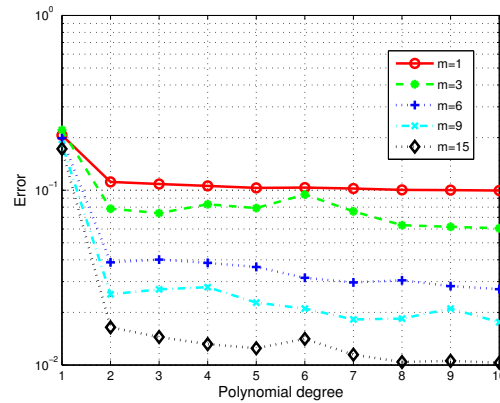


Figure 3: Evolution of error $\varepsilon(u_m, u)$ with respect to polynomial degree p for different values of m using sample size given by quadratic rule $Q = dm(p + 1)^2$.

is the crenel function defined by:

$$c(\xi_k) = \begin{cases} 1 & \text{on } [0, \frac{1}{6}[+ 2n\frac{1}{6}, n = 0, 1, 2 \\ 0 & \text{on } [\frac{1}{6}, \frac{2}{6}[+ 2n\frac{1}{6}, n = 0, 1, 2 \end{cases} .$$

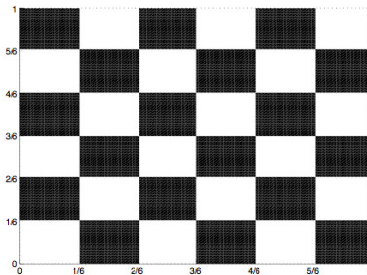


Figure 4: Checker-board function.

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 are one element of the partition and whose restrictions on these supports are rescaled Legendre polynomials.

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 checker-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.

4.2.2 Results

Algorithm 3 allows the construction of a sequence of sparse rank- m approximations u_m in $\mathcal{S}_{n_1}^1 \otimes \mathcal{S}_{n_2}^2$. We estimate optimal rank- m_{op} using 3-fold cross validation (see Section 3.3.3).

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 3 for the description of these alternatives). 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 $\mathcal{S}_{n_1}^1 = \mathcal{S}_{n_2}^2 = \mathbb{P}_{p,s}$. P denotes the dimension of the space $\mathcal{S}_{n_1}^1 \otimes \mathcal{S}_{n_2}^2$. 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. This is because ℓ_1 -regularization is able to select non zero coefficients corresponding to appropriate basis functions of the piecewise polynomial approximation space. For instance, when $\mathbb{P}_{5,6}$ is used as the approximation space, only 3 (out of 36) non zero coefficients corresponding to piecewise constant bases are selected by ℓ_1 regularization along each dimension in each rank one element (that is the sparsity ratio is $\varrho_1 \approx 0.08$ for each rank-one element), thus yielding an almost exact recovery. We also find that 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).

Approximation space	Ordinary Least Square				ℓ_2				ℓ_1			
	No update		Update		No update		Update		No update		Update	
	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	$1.96 \cdot 10^{-12}$	4	$2.41 \cdot 10^{-13}$	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	$1.93 \cdot 10^{-12}$	2	$1.1 \cdot 10^{-12}$	2
$\mathcal{R}_m(\mathbb{P}_{5,6} \otimes \mathbb{P}_{5,6}), P = 36^2$	31.27	1	-	-	0.624	10	0.623	3	$1.22 \cdot 10^{-12}$	2	$7.93 \cdot 10^{-13}$	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	$1.21 \cdot 10^{-12}$	2	$7.88 \cdot 10^{-13}$	2

From this analytical example, several conclusions can be drawn:

- ℓ_1 -regularization in alternating least squares 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.

4.3 Analytical model: Rastrigin function

For certain classes of non smooth functions, wavelet bases form an appropriate choice as they allow the simultaneous description of global and local features [19]. In this example, we demonstrate the use of our algorithm with polynomial wavelet bases by studying 2-dimensional Rastrigin function. The function is given by

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

where ξ_1, ξ_2 are independent random variables uniformly distributed in $[-4,4]$.

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 and an optimal rank approximation $u_{m_{op}}$ is selected using 3-fold cross validation (see the rank selection strategy in section 3.3.3). Figure 5(a) shows the convergence of this optimal approximation with respect to the sample size Q for the two different approximation spaces. 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. Figure 5(b) shows the convergence of the optimal wavelet approximation with respect to the sample size Q for different regularizations within the alternated minimization algorithm of the correction step. The ℓ_1 regularization is more accurate when compared to both OLS and ℓ_2 regularization, particularly for few model evaluations. We can thus conclude that a good choice of basis functions is important in order to fully realize the potential of sparse ℓ_1 regularization in the tensor approximation algorithm.

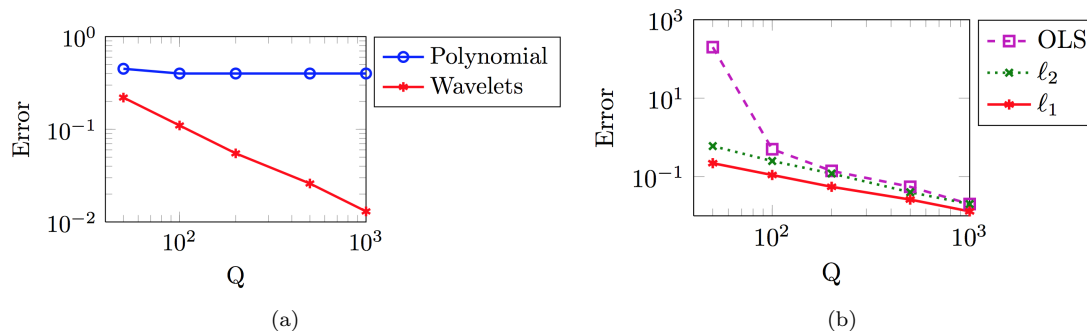


Figure 5: Evolution of error $\varepsilon(u_{m_{op}}, u)$ with respect to the number of samples Q . Approximations obtained with Algorithm 3 with optimal rank selection: (a) for the two different approximation spaces $\mathbb{P}_7 \otimes \mathbb{P}_7$ ($P = 64$) and $\mathbb{W}_{4,3} \otimes \mathbb{W}_{4,3}$ ($P = 1600$) and (b) for different types of regularizations with approximation space $\mathbb{W}_{4,3} \otimes \mathbb{W}_{4,3}$.

Figure 6 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. Conversely, if only very few samples are available, then a very low rank approximation, even rank one, is able to capture the global features. The proposed method provides the simplest representation of the function with respect to the available information.

We finally analyze the robustness of Algorithm 3 with respect to the sample sets. We use wavelet bases. An optimal rank approximation $u_{m_{op}}$ is selected using 3-fold cross validation as described in section 3.3.3. 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 7 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 greedy constructions of sparse low-rank representations when only a small number of samples is available, indeed the problem is reduced to one where elements of subsets of small dimension are to be learnt. The interest of using low-rank representations should also become clear when dealing with higher dimensional problems.

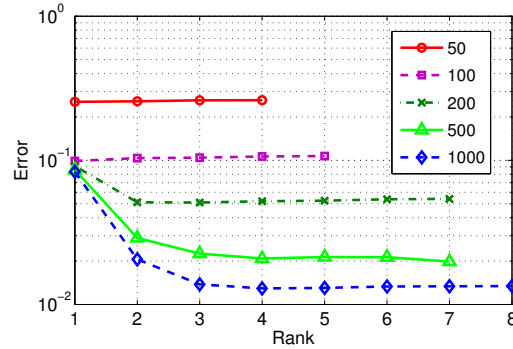


Figure 6: Evolution of error $\varepsilon(u_{m_{op}}, u)$ of approximations obtained using Algorithm 3 with respect to rank m for different sample sizes on wavelet approximation space $\mathbb{W}_{4,3} \otimes \mathbb{W}_{4,3}$ ($P = 1600$)

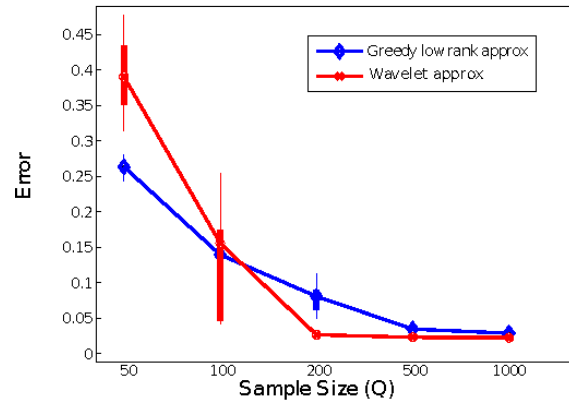


Figure 7: Evolution of error $\varepsilon(u_{m_{op}}, u)$ with respect to sample size Q . (Red line) approximation obtained with direct least-squares approximation with ℓ_1 -regularization on the full polynomial wavelet approximation space $\mathbb{W}_{4,3} \otimes \mathbb{W}_{4,3}$, (blue line) approximation obtained with Algorithm 3 (with ℓ_1 -regularization) and with optimal rank selection.

4.4 A model problem in structural vibration analysis

4.4.1 Model problem

We consider a forced vibration problem of a slightly damped random linear elastic structure. The structure composed of two plates is clamped on part Γ_1 of the boundary and submitted to a harmonic load on part Γ_2 of the boundary as represented in figure 8(a). A finite element approximation is introduced at the spatial level using a mesh composed of 1778 DKT plate elements (see figure 8(b)) and leading to a discrete deterministic model with $N = 5556$ degrees of freedom.

The resulting discrete problem at frequency ω writes

$$(-\omega^2 \mathbf{M} + i\omega \mathbf{C} + \mathbf{K})\mathbf{u} = \mathbf{f},$$

where $\mathbf{u} \in \mathbb{C}^N$ is the vector of coefficients of the approximation of the displacement field and \mathbf{M} , $\mathbf{K} = E\mathbf{K}$ and $\mathbf{C} = i\omega\eta E\mathbf{K}$ are the mass, stiffness and damping matrices respectively. The non-dimensional analysis considers a unitary mass density and a circular frequency $\omega = 0.67 \text{ rad.s}^{-1}$.

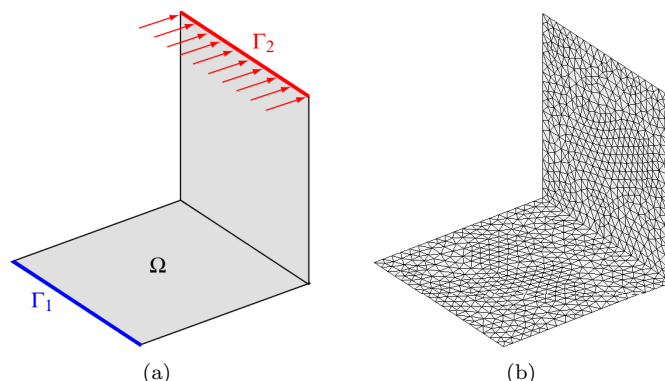


Figure 8: Elastic plate structure under harmonic bending load. Geometry and boundary conditions (a) and finite element mesh (b).

The Young modulus E and the damping parameter η are defined by

$$E = \begin{cases} 0.975 + 0.025\xi_1 & \text{on horizontal plate,} \\ 0.975 + 0.025\xi_2 & \text{on vertical plate,} \end{cases}$$

$$\eta = \begin{cases} 0.0075 + 0.0025\xi_3 & \text{on horizontal plate,} \\ 0.0075 + 0.0025\xi_4 & \text{on vertical plate,} \end{cases}$$

where the $\xi_k \sim U(-1, 1)$, $k = 1, \dots, 4$, are independent uniform random variables (here $d = 4$). We define the quantity of interest

$$I(u)(\xi) = \log \|u_c\|,$$

where u_c is the displacement of the top right node of the two plate structure.

4.4.2 Impact of regularization and stochastic polynomial degree

In this example, we illustrate that the approximation in sparse low rank tensor subsets is robust when increasing the degree of underlying polynomial approximation spaces with a fixed number of samples Q .

We use Legendre polynomial basis functions with degree $p = 1$ to 20 and denote by \mathbb{P}_p the corresponding space of polynomials of maximal degree p in each dimension. A rank- m approximation is searched in the isotropic tensor space $\mathbb{P}_p \otimes \dots \otimes \mathbb{P}_p$.

Figure 9(left column) shows the error $\varepsilon(I_m, I)$ as a function of the polynomial degree p for different ranks m for three different sizes of the sample set, $Q = 80, 200$ and 500. The low rank approximation I_m is computed with and without sparsity constraint, *i.e.* we compare OLS (dashed lines) and ℓ_1 regularization (solid lines) in correction step 4 of Algorithm 3. Figure 10 summarizes the error $\varepsilon(I_m, I)$ for different sizes of sample sets for the rank-10 approximation when using ℓ_1 -regularization (solid lines) and for the rank- m approximation giving the best approximation when using OLS (dashed lines). On the one hand, we find that OLS yields a deterioration with higher polynomial order. This is consistent with the conclusions in section 4.1 and the quadratic rule according to which convergence is observed for $Q \geq dm(p+1)^2$ and a deterioration is expected otherwise. On the other hand, we see that ℓ_1 -regularization gives a more stable approximation with increasing polynomial order and also gives a more accurate best approximation than the best approximation obtained with OLS. This can be attributed to the selection of pertinent basis functions obtained by imposing sparsity constraint. Indeed, we clearly see in figure 9(right column) that the sparsity ratio ρ_5 for sparse low rank approximation (solid black line) decreases with increasing polynomial degree. Along with the total sparsity ratio, the partial sparsity ratios $\rho_5^{(k)}$

in each dimension $k = 1$ to 4 are plotted in figure 9(right column). We see that ℓ_1 -regularization exploits sparsity especially in dimensions 3 and 4 corresponding to the damping coefficients, indeed the quantity of interest has smooth dependence on variables ξ_3 and ξ_4 whereas it has a high non linear behavior with respect to ξ_1 and ξ_2 . Figure 11 shows the reference quantity of interest and the rank-3 approximation I_3 obtained using ℓ_1 -regularization and polynomial degree 5 constructed from $Q = 200$ samples.

This illustration also points out that a small number of model evaluations, for instance $Q = 80$, does not enable to capture correctly local features of the function and a low rank approximation ($m = 3$) is selected as the best approximation regarding the available information. As the number of samples increases, higher rank approximation are selected that capture the local features of the function more accurately.

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 when having only very few information on the function. The ability of the proposed method to detect and exploit low-rank and sparsity was illustrated on three analytical models and on a partial differential equation with random coefficients. In order to exploit at best the few samples available in practical applications in uncertainty propagation, algorithms that are able to automatically detect low-rank structures of a function (e.g. by finding an optimal tree in hierarchical tensor representations) should be developed. 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] E.J. CANDÈS, J. ROMBERG, AND T. TAO, *Near optimal signal recovery from random projections: Universal encoding strategies?*, IEEE Transactions on information theory, 52(12) (2006), pp. 5406–5425.
- [5] 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.
- [6] 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.
- [7] D.L. DONOHO, *Compressed Sensing*, IEEE Transactions on information theory, 52(4) (2006), pp. 1289–1306.
- [8] 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.

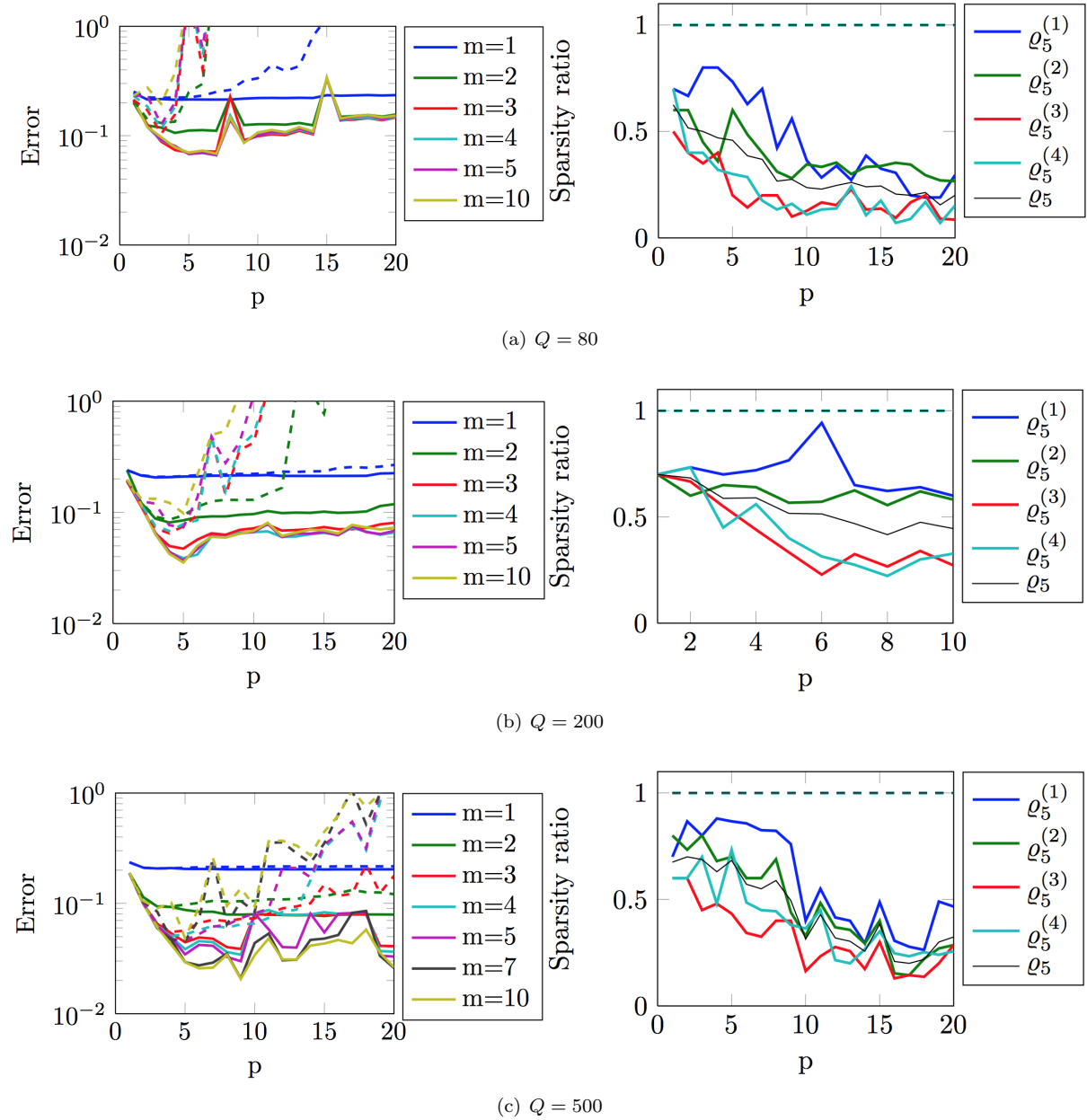


Figure 9: Left column: evolution of approximation error $\varepsilon(I_m, I)$ with respect to polynomial degree p with a fixed number of samples $Q = 80, Q = 200$ and $Q = 500$. Polynomial approximations obtained using ℓ_1 -regularization (solid lines) and using OLS (dashed lines). Right column: total and partial sparsity ratios with respect to polynomial degree for $m = 5$.

- [9] A. DOOSTAN AND H. OWHADI, *A non-adapted sparse approximation of pdes with stochastic inputs*, Journal of Computational Physics, 230 (2011), pp. 3015–3034.
- [10] 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).
- [11] B. EFRON, T. HASTIE, I. JOHNSTONE, AND R. TIBSHIRANI, *Least angle regression*, The Annals of Statistics, 32 (2004), pp. 407–499.

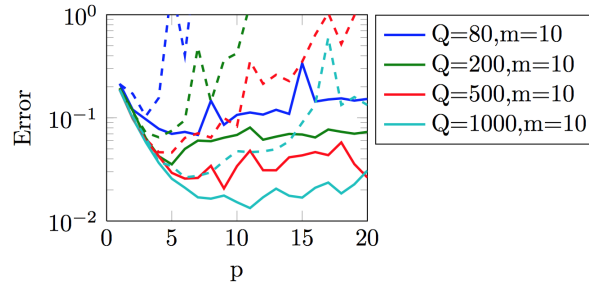


Figure 10: Evolution of approximation error $\varepsilon(I_m, I)$ with respect to polynomial degree p for different sizes of sample sets (random sampling). Polynomial approximations obtained using ℓ_1 -regularization (solid lines) and using OLS (dashed lines).

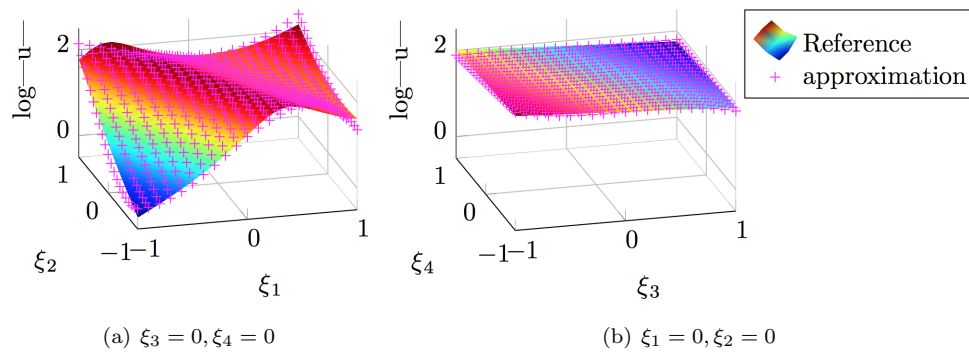


Figure 11: Response surface : (surface) reference and (dotted) rank-4 approximation obtained with polynomial basis $p = 4$, $Q = 200$ using l_1 regularization.

- [12] A. FALCÓ AND A. NOUY, *Proper generalized decomposition for nonlinear convex problems in tensor banach spaces*, Numerische Mathematik, 121 (2012), pp. 503–530.
- [13] R. GHANEM AND P. SPANOS, *Stochastic finite elements: a spectral approach*, Springer, Berlin, 1991.
- [14] L. GRASEDYCK, D. KRESSNER, AND C. TOBLER, *A literature survey of low-rank tensor approximation techniques*, arXiv:1302.7121, (2013).
- [15] W. HACKBUSCH, *Tensor Spaces and Numerical Tensor Calculus*, vol. 42 of Series in Computational Mathematics, Springer, 2012.
- [16] B.N. KHOROMSKIJ, *Tensor structured numerical methods in scientific computing: Survey on recent advances*, Chemometrics and Intelligent laboratory Systems, 110 (2012), pp. 1–19.
- [17] 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.
- [18] T. G. KOLDA AND B. W. BADER, *Tensor decompositions and applications*, SIAM Review, 51 (2009), pp. 455–500.
- [19] 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.
- [20] O. P. LE MAÎTRE AND O. M. KNIO, *Spectral Methods for Uncertainty Quantification With Applications to Computational Fluid Dynamics*, Scientific Computation, Springer, 2010.

- [21] J. MAIRAL, F. BACH, J. PONCE, AND G. SAPIRO, *Online learning for matrix factorization and sparse coding*, Journal of Machine Learning Research, 11 (2010).
- [22] 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.
- [23] H. G. MATTHIES AND E ZANDER, *Solving stochastic systems with low-rank tensor compression*, Linear Algebra and its Applications, 436 (2012), pp. 3819–3838.
- [24] G. MIGLIORATI, F. NOBILE, E. VON SCHWERIN, AND R. TEMPONE, *Analysis of the discrete L_2 projection on polynomial spaces with random evaluations*, Tech. Report 46, MATHICSE, 2011.
- [25] 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.
- [26] ———, *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.
- [27] ———, *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.
- [28] 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.
- [29] R. TIBSHIRANI, *Regression shrinkage and selection via the lasso*, Journal of the Royal Statistical Society Series B, 58 (1996), pp. 267–288.
- [30] D. XIU AND G. E. KARNIADAKIS, *The Wiener-Askey polynomial chaos for stochastic differential equations*, SIAM J. Sci. Comput., 24 (2002), pp. 619–644.