On the convergence of a Greedy Rank-One Update Algorithm for a class of linear systems

A. Ammar · F. Chinesta · A. Falcó

Abstract In this paper we study the convergence of the well-known Greedy Rank-One Update Algorithm. It is used to construct the rank-one series solution for full-rank linear systems. The existence of the rank one approximations is also not new, but surprisingly the focus there has been more on the applications side more that in the convergence analysis. Our main contribution is to prove the convergence of the algorithm and also we study the required rank one approximation in each step. We also give some numerical examples and describe its relationship with the Finite Element Method for High-Dimensional Partial Differential Equations based on the tensorial product of one-dimensional bases. We illustrate this situation taking as a model problem the multidimensional Poisson equation with homogeneous Dirichlet boundary condition.

A.Ammar
Laboratoire de Rhéologie, INPG, UJF, CNRS (UMR 5520)
1301 rue de la piscine
BP 53 Domaine universitaire
F-38041 Grenoble Cedex 9, France. E-mail: Amine.Ammar@ujf-grenoble.fr

F.Chinesta
Institut de Recherche en Génie Civil et Mécanique (GeM)
Ecole Centrale de Nantes
1 rue de la Noë
BP 92101 44 321 Nantes Cédex 3, France. E-mail: Francisco.Chinesta@ec-nantes.fr

A. Falcó
Departamento de Ciencias, Físicas, Matemáticas y de la Computación
Universidad CEU Cardenal Herrera
San Bartolomé 55
46115 Alfara del Patriarca (Valencia), Spain. E-mail: afalco@uch.ceu.es

1 Introduction

In [1] and [2], some of the authors of the present paper propose the use of a separated representation, which allows to define a tensor product approximation basis as well as to decouple the numerical integration of a high dimensional model in each dimension. The milestone of this methodology is the use of shape functions given by a tensorial based construction. This fact has advantages as the manipulation of only one dimensional polynomials and its derivatives, that provides a better computational performance and simplified implementation and use one-dimensional integration rules. Moreover, it makes possible the solution of models defined in spaces of more than hundred dimensions in some specific applications. This problem is closely related with the decomposition of a tensor as a sum of rank-one tensors, that it can be considered as a higher order extension of the matrix Singular Value Decomposition.

The purpose of this work is to formalize and analyze the above strategy in the framework of methods for solving linear systems by means tensor decompositions. As we will show the approximation given in [1] and [2], is closely related with the best low-rank approximation problem for high order tensors (see [4]). Unfortunately, in [4] it has been proved that tensors of order 3 or higher can fail to have best rank-r approximation for $r \geq 2$. Our strategy, with the perspective of [4] in mind, is to use the fact that tensors of order 3 or higher have best rank-1 approximation.

In this context, we propose the use of a Greedy Rank-One Update Algorithm to construct, for a full rank linear system, a rank-r approximate solution. This approach is based in the so-called by the signal processing community as the *Matching Pursuit Algorithm* of Mallat and Zhang [12], also known as *Projection Pursuit* by the statistics community (see Friedman and Stuezle [7] and Huber [8]) or as a *Pure Greedy Algorithm* (see the recent survey of Temlyakov [14]) in the approximation theory community. Our main contribution, stated in Theorem 1, is to prove the convergence of this Greedy rank-one update algorithm. and characterize the speed of convergence in terms of a sequence of angles. This strategy depends strongly on the computation of the best rank-1 approximation of the residual obtained at each step of the proposed algorithm. To solve this we will propose the use of a Block Coordinate Descend Method, because it has global convergence. In particular, we will show that for the class of invertible matrices, this problem collapses, for each selected direction, to an ordinary least-squares problem. We remark that this strategy appears to be identical to so-called Alternating Least Squares (ALS) method proposed in [3] for the class of separable matrices (see statement of Corollary 1 below).

In [9] an orthogonal Greedy tensor decomposition has been used in order to compute a rank-r approximation. However, as the author points out, the computational difficulty of this approach arises in enforcing the constrains of the rank-one approximation needed at each step of the proposed algorithm. Zhang and Golub [16], also explores various computational techniques when the tensor has a completely orthogonal decomposition, in which case the problem is more simpler. All these methods use an ALS Approach for computing the rank-one approach. On the other hand, the approach following in this paper differs from the [11] based in the associated Lagrange Equations to the minimum least-squares cost function.

This paper is organized as follows. In the next section we introduce the notation used in this paper and give our main result, the convergence of the Greedy Rank-One Update Algorithm for solving full rank linear systems. In Section 3 we study the rank-one approach, in particular we prove that a Block Cyclic Coordinate Descend strategy

implies an ALS Algorithm. Section 4.1 is dedicated to give some numerical examples of the above algorithms and describe its relationship with the Finite Element Method for High-Dimensional Partial Differential Equations based on the tensorial product of one-dimensional bases. We illustrate this situation taking as a model problem the multidimensional Poisson equation with homogeneous Dirichlet boundary condition. We conclude with some comments and remarks.

2 Definitions and Statement of Main Result

First at all we introduce some notation. We denote by $\mathbb{R}^{N \times M}$, the set of $N \times M$-matrices and by A^T the transpose of a given matrix A. As usual we use

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y} = \mathbf{y}^T \mathbf{x}$$

to denote the Euclidean inner product in $\mathbb{R}^N$, and its corresponding 2-norm, by $\|x\|_2 = \langle \mathbf{x}, \mathbf{x} \rangle^{1/2}$. Let I_N be the $N \times N$-identity matrix and when the dimension is clear from the context, we simply denote it by I. Given a sequence $\{\mathbf{u}_j\}_{j=0}^{\infty} \subset \mathbb{R}^N$, we say that a vector $\mathbf{u} \in \mathbb{R}^N$ can be written as

$$\mathbf{u} = \sum_{j=0}^{\infty} \mathbf{u}_j$$

if and only if

$$\lim_{n \to \infty} \sum_{j=0}^{n} \mathbf{u}_j = \mathbf{u}$$

in the $\|\cdot\|_2$-topology. Now, we recall the definition and some properties of the Kronecker product. The Kronecker product of $A \in \mathbb{R}^{N_1' \times N_1}$ and $B \in \mathbb{R}^{N_2' \times N_2}$, written $A \otimes B$, is the tensor algebraic operation defined as

$$A \otimes B = \begin{bmatrix} A_{1,1}B & A_{1,2}B & \cdots & A_{1,N_1'}B \\ A_{2,1}B & A_{2,2}B & \cdots & A_{2,N_1'}B \\ \vdots & \vdots & \ddots & \vdots \\ A_{N_1,1}B & A_{N_1,2}B & \cdots & A_{N_1,N_1'}B \end{bmatrix} \in \mathbb{R}^{N_1'N_2' \times N_1 N_2}.$$

Also, it can be defined by

$$(A \otimes B)_{(j_1-1)N_2'+j_2;(i_1-1)N_2+i_2} = A_{j_1;i_1}B_{j_2;i_2}. \tag{1}$$

Assume that $A_i \in \mathbb{R}^{N_i' \times N_i}$ for $1 \leq i \leq d$. Proceeding inductively we show that

$$(A_1 \otimes \cdots \otimes A_d)_{s,t} = (A_1)_{j_1,i_1} \cdots (A_d)_{j_d,i_d}$$

if and only if s and t satisfy

$$s = j_d + \sum_{l=1}^{d-1} \left[(j_l - 1) \prod_{l'=l+1}^{d} N_{l'}' \right] \text{ and } t = i_d + \sum_{l=1}^{d-1} \left[(i_l - 1) \prod_{l'=l+1}^{d} N_{l'} \right]. \tag{2}$$

Finally, we list some of the well-know properties of the Kronecker product (see for example [5] or [15]).

1. $A \otimes (B \otimes C) = (A \otimes B) \otimes C$.
2. $(A + B) \otimes (C + D) = (A \otimes C) + (B \otimes C) + (A \otimes D) + (B \otimes D)$.
3. $AB \otimes CD = (A \otimes C)(B \otimes D)$.
4. $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$.
5. $(A \otimes B)^T = A^T \otimes B^T$.
6. If A and B are banded, then $A \otimes B$ is banded.
7. If A and B are symmetric, then $A \otimes B$ is symmetric.
8. If A and B are definite positive, then $A \otimes B$ is definite positive.

The concept of separated representation was introduced by Beylkin and Mohlenkamp in [3] and it is related with the problem of constructing the approximate solutions of some classes of problems in high-dimensional spaces by means a separable function. In particular, for a given map

$$u : [0,1]^d \subset \mathbb{R}^d \longrightarrow \mathbb{R},$$

we say that it has a *separable representation* if

$$u(x_1, \ldots, x_d) = \sum_{j=1}^{\infty} u_1^{(j)}(x_1) \cdots u_d^{(j)}(x_d) \tag{3}$$

Now, consider a mesh of $[0,1]$ in the x_k-variable given by N_k-mesh points, $1 \leq k \leq d$, then we can write a discrete version of (3) by

$$u(x_{i_1}, \ldots, x_{i_d}) = \sum_{j=1}^{\infty} u_1^{(j)}(x_{i_1}) \cdots u_d^{(j)}(x_{i_d}), \tag{4}$$

where $1 \leq i_k \leq N_k$ for $1 \leq k \leq d$. Observe that for each $1 \leq k \leq d$, if $\mathbf{x}_k^j \in \mathbb{R}^{N_k}$ denotes the vector with components $u_k^{(j)}(x_{i_k})$ for $1 \leq i_k \leq N_k$, then (4) it is equivalent to

$$\mathbf{u} = \sum_{j=1}^{\infty} \mathbf{x}_1^j \otimes \cdots \otimes \mathbf{x}_d^j. \tag{5}$$

We point out that (5) is an useful expression to implemented numerical algorithms using the MATLAB and OCTAVE function `kron`.

Suppose that for given a linear Partial Differential Equation, and after a discretization by means Finite Elements, we need to solve the linear system

$$A\mathbf{u} = \mathbf{f}, \tag{6}$$

where A is a $(N_1 \cdots N_d) \times (N_1 \cdots N_d)$-dimensional invertible matrix, for some $N_1, \ldots, N_d \in \mathbb{N}$. Then from all said above a low rank approximation

$$A^{-1}\mathbf{f} \approx \mathbf{u}_n = \sum_{j=1}^{n} \mathbf{x}_1^j \otimes \cdots \otimes \mathbf{x}_d^j$$

with sufficient approximation exists, for some $n \geq 1$ and where $\mathbf{x}_i^j \in \mathbb{R}^{N_i}$ for $i = 1, 2, \ldots, d$ and $j = 1, 2, \ldots, n$. Moreover, we would to show that

$$\lim_{n \to \infty} \left\| A^{-1}\mathbf{f} - \mathbf{u}_n \right\|_2 = 0,$$

that is,

$$A^{-1}\mathbf{f} = \sum_{j=1}^{\infty} \mathbf{x}_1^j \otimes \cdots \otimes \mathbf{x}_d^j.$$

Thus, in a first approach to solve it, we would to determine vectors $\mathbf{x}_1^j, \ldots, \mathbf{x}_d^j$ for $j = 1, 2, \ldots, n$ that minimizes

$$\left\| \mathbf{f} - A \left(\sum_{j=1}^{n} \mathbf{x}_1^j \otimes \cdots \otimes \mathbf{x}_d^j \right) \right\|_2,$$

or, in short

$$\mathrm{argmin}_{\mathrm{rank}_\otimes \mathbf{u} \leq n} \|\mathbf{f} - A\mathbf{u}\|_2, \tag{7}$$

by using the notation introduced in [4]. Note that this problem is closely related with the multi-linear generalization of the best low rank approximation of high-order tensors (see also [10], [11] and references therein).

For each $n \in \mathbb{N}$, we define the set

$$\mathcal{S}_n = \{ \mathbf{x} \in \mathbb{R}^{N_1 \cdots N_d} : \mathrm{rank}_\otimes \mathbf{x} \leq n \},$$

introduced in [4], in the following way. Given $\mathbf{x} \in \mathbb{R}^{N_1 \cdots N_d}$ we say that $\mathbf{x} \in \mathcal{S}_1 = \mathcal{S}_1(N_1, N_2, \ldots, N_d)$ if $\mathbf{x} = \mathbf{x}_1 \otimes \mathbf{x}_2 \otimes \cdots \otimes \mathbf{x}_d$, where $\mathbf{x}_i \in \mathbb{R}^{N_i}$, for $i = 1, \ldots, d$. For $n \geq 2$ we define inductively $\mathcal{S}_n = \mathcal{S}_n(N_1, N_2, \ldots, N_d) = \mathcal{S}_{n-1} + \mathcal{S}_1$, that is,

$$\mathcal{S}_n = \left\{ \mathbf{x} : \mathbf{x} = \sum_{i=1}^{k} \mathbf{x}^{(i)}, \mathbf{x}^{(i)} \in \mathcal{S}_1 \text{ for } 1 \leq i \leq k \leq n \right\}.$$

Note that $\mathcal{S}_n \subset \mathcal{S}_{n+1}$ for all $n \geq 1$.

Unfortunately, from Proposition 4.1 (a) of [4], we have that the set $\mathcal{S}_n$ is not necessarily (or even usually) closed for each $n \geq 2$. However, from Proposition 4.2 of [4] it follows that $\mathcal{S}_1$ is a closed set in any norm-topology. This fact implies, as we will see below in Lemma 2, that given an invertible matrix $A \in \mathbb{R}^{N_1 N_2 \cdots N_d \times N_1 N_2 \cdots N_d}$, then for every $\mathbf{b} \in \mathbb{R}^{N_1 \cdots N_d}$ we have

$$\mathrm{argmin}_{\mathbf{x} \in \mathcal{S}_1} \|\mathbf{b} - A\mathbf{x}\|_2 \neq \emptyset. \tag{8}$$

This allow to consider the following iterative scheme. Let $\mathbf{u}_0 = \mathbf{y}_0 = \mathbf{0}$, and for each $n \geq 1$ take

$$\mathbf{r}_{n-1} = \mathbf{f} - A\mathbf{u}_{n-1}, \tag{9}$$

$$\mathbf{u}_n = \mathbf{u}_{n-1} + \mathbf{y}_n \text{ where } \mathbf{y}_n \in \mathrm{argmin}_{\mathbf{y} \in \mathcal{S}_1} \|\mathbf{r}_{n-1} - A\mathbf{y}\|_2. \tag{10}$$

Note that for each vector $\mathbf{f} \in \mathbb{R}^{N_1 \cdots N_d}$ and each invertible matrix $A \in \mathbb{R}^{N_1 N_2 \cdots N_d \times N_1 N_2 \cdots N_d}$, we can construct for each n, by using (9)-(10), a vector

$$\mathbf{u}_n = \sum_{j=1}^{n} \mathbf{y}_n \in \mathcal{S}_n \setminus \mathcal{S}_{n-1},$$

here we assume that $\mathbf{y}_j \neq \mathbf{0}$ for $1 \leq j \leq n$, that is, $\mathrm{rank}_{\otimes} \mathbf{u}_n = n$. Since $\mathbf{u}_n \approx A^{-1}\mathbf{f}$, we define the $\mathrm{rank}_{\otimes}$ for $A^{-1}\mathbf{f}$ obtained by the Greedy Rank-One Update Algorithm (9)-(10) as

$$\mathrm{rank}_{\otimes}^{G}\left(A^{-1}\mathbf{f}\right) = \begin{cases} \infty & \text{if } \{j \geq 1 : \mathbf{y}_j = \mathbf{0}\} = \emptyset, \\ \min\{j \geq 1 : \mathbf{y}_j = \mathbf{0}\} - 1 & \text{otherwise.} \end{cases}$$

The following theorem, which is the main result of this paper, gives the convergence of the Greedy Rank-One Update Approximation for solving linear systems with full rank matrix.

Theorem 1 *Let* $\mathbf{f} \in \mathbb{R}^{N_1 N_2 \cdots N_d}$ *and* $A \in \mathbb{R}^{N_1 N_2 \cdots N_d \times N_1 N_2 \cdots N_d}$, *be an invertible matrix. Then, by using the iterative scheme (9)-(10), we obtain that the sequence* $\{\|\mathbf{r}_n\|_2\}_{n=0}^{\mathrm{rank}_{\otimes}^{G}(A^{-1}\mathbf{f})}$, *is strictly decreasing and*

$$A^{-1}\mathbf{f} = \lim_{n\to\infty} \mathbf{u}_n = \sum_{j=0}^{\mathrm{rank}_{\otimes}^{G}(A^{-1}\mathbf{f})} \mathbf{y}_j. \tag{11}$$

Moreover, the rate of convergence is given by

$$\frac{\|\mathbf{r}_n\|_2}{\|\mathbf{r}_0\|_2} = \prod_{j=1}^{n} \sin\theta_j \tag{12}$$

for $1 \leq n \leq \mathrm{rank}_{\otimes}^{G}\left(A^{-1}\mathbf{f}\right)$ *where*

$$\theta_j = \arccos\left(\frac{\langle \mathbf{r}_{j-1}, A\mathbf{y}_j \rangle}{\|\mathbf{r}_{j-1}\|_2 \|A\mathbf{y}_j\|_2}\right) \in (0, \pi/2) \tag{13}$$

for $1 \leq j \leq n$.

We remark that Grasedyck [6] provides the existence and computation of a low Kronecker rank approximate for a linear system where the matrix A possesses a particular tensor structure and $\mathbf{f} \in \mathcal{S}_1$.

From (11) we obtain that if $\mathrm{rank}_{\otimes}^{G}\left(A^{-1}\mathbf{f}\right) < \infty$, then $\|\mathbf{r}_n\|_2 = 0$ for all $n > \mathrm{rank}_{\otimes}^{G}\left(A^{-1}\mathbf{f}\right)$. Thus, the above theorem allow to us to construct a procedure, that we give in the pseudo-code form in Algorithm 1, under the assumption that we have a numerical method in order to find a $\mathbf{y}$ solving (8) (see the step 5 in Algorithm 1) and that we introduce below.

In order to prove Theorem 1 we need the following two lemmas. From Proposition 4.2 of [4] it follows the next result.

Lemma 1 $\mathcal{S}_1$ *is a closed set in the* $\|\cdot\|_2$-*topology of* $\mathbb{R}^{N_1 \cdots N_d}$.

The next lemma give us the existence (but not unicity) of a minimizer for the map $\Phi(\mathbf{x}) = \|\mathbf{b} - A\mathbf{x}\|_2$ defined from $\mathcal{S}_1$ to $[0, \infty)$.

Lemma 2 *Assume that* A *invertible. Then there exists* $\mathbf{x}^* \in \mathcal{S}_1$ *such that*

$$\|\mathbf{b} - A\mathbf{x}^*\|_2 = \min_{\mathbf{x} \in \mathcal{S}_1} \|\mathbf{b} - A\mathbf{x}\|_2. \tag{14}$$

Algorithm 1 Greedy Rank-One Update

1: **procedure** $\mathrm{GROU}(\mathbf{f}, A, \varepsilon, \mathtt{tol}, \mathtt{rank\_max})$
2:   $\mathbf{r}_0 = \mathbf{f}$
3:   $\mathbf{u} = \mathbf{0}$
4:   **for** $i = 0, 1, 2, \ldots, \mathtt{rank\_max}$ **do**
5:    $\mathbf{y} = $ **procedure** $(\min_{\mathrm{rank}_\otimes \mathbf{x} \leq 1} \|\mathbf{r}_i - A\mathbf{y}\|_2^2)$
6:    $\mathbf{r}_{i+1} = \mathbf{r}_i - A\mathbf{y}$
7:    $\mathbf{u} \leftarrow \mathbf{u} + \mathbf{y}$
8:    **if** $\|\mathbf{r}_{i+1}\|_2 < \varepsilon$ or $|\|\mathbf{r}_{i+1}\|_2 - \|\mathbf{r}_i\|_2| < \mathtt{tol}$ **then goto** 13
9:    **end if**
10:   **end for**
11:   **return** $\mathbf{u}$ and $\|\mathbf{r}_{\mathtt{rank\_max}}\|_2$.
12:   **break**
13:   **return** $\mathbf{u}$ and $\|\mathbf{r}_{i+1}\|_2$
14: **end procedure**

Proof Let $\mathfrak{S} = \{\|\mathbf{b} - A\mathbf{x}\|_2 : \mathbf{x} \in \mathcal{S}_1\} \subset \mathbb{R}$. Since it is bounded below, there exists $\gamma = \inf \mathfrak{S} = \inf_{\mathbf{x} \in \mathcal{S}_1} \|\mathbf{b} - A\mathbf{x}\|_2 \geq 0$. Moreover, for each $\mathbf{y}_0 \in \mathbb{R}^{N_1 \cdots N_d}$ such that $\|\mathbf{y}_0\|_2 = \gamma$, then $\mathbf{x}_0 = A^{-1}(\mathbf{b} - \mathbf{y}_0)$ satisfy that

$$\|\mathbf{b} - A\mathbf{x}_0\|_2 = \gamma = \inf_{\mathbf{x} \in \mathcal{S}_1} \|\mathbf{b} - A\mathbf{x}\|_2.$$

Since γ is the infimum of the set $\mathfrak{S}$, for each $n > 0$ there exists $\mathbf{x}^n \in \mathcal{S}_1$ such that

$$\gamma + \frac{1}{n} \geq \|\mathbf{b} - A\mathbf{x}^n\|_2 \geq \gamma. \tag{15}$$

Now, let $\mathbf{y}^n = \mathbf{b} - A\mathbf{x}^n$, then for all $n > 0$,

$$\mathbf{y}^n \in \{\mathbf{y} : (\gamma + 1) \geq \|\mathbf{y}\|_2 \geq \gamma\},$$

which is a compact set in the $\|\cdot\|_2$-topology. Thus, there exists a subsequence $\{\mathbf{y}^{n_k}\}$ such that $\mathbf{y}^{n_k} \to \mathbf{y}^*$ as $n_k \to \infty$ in the $\|\cdot\|_2$-topology. In consequence,

$$\mathbf{x}^{n_k} = A^{-1}(\mathbf{b} - \mathbf{y}^{n_k}) \in \mathcal{S}_1 \longrightarrow \mathbf{x}^* = A^{-1}(\mathbf{b} - \mathbf{y}^*) \in \mathcal{S}_1.$$

as $n_k \to \infty$, because $\mathcal{S}_1$ is a closed set the $\|\cdot\|_2$-topology. Finally, substituting in (15) n by n_k and taking limits as $n_k \to \infty$, we obtain that $\|\mathbf{b} - A\mathbf{x}^*\|_2 = \gamma$ and this ends the proof of lemma. $\square$

Proof of Theorem 1

It is clear that if $\mathbf{f} = \mathbf{0}$ the theorem follows by consider $\mathbf{u}_0 = \mathbf{y}_0 = \mathbf{0}$. Thus, from now one we assume that $\mathbf{f} \neq \mathbf{0}$.

  Recall that $\mathbf{u}_0 = \mathbf{y}_0 = \mathbf{0}$, and

$$\mathbf{r}_n = \mathbf{f} - A\mathbf{u}_n = \mathbf{r}_{n-1} - A\mathbf{y}_n,$$

for all $n \geq 0$. By using Lemma 2, the residual at the n-th step $\mathbf{r}_n$ satisfy

$$\|\mathbf{r}_n\|_2 = \|\mathbf{r}_{n-1} - A\mathbf{y}_n\|_2 = \min_{\mathbf{y} \in \mathcal{S}_1} \|\mathbf{r}_{n-1} - A\mathbf{y}\|_2. \tag{16}$$

Since, $\mathbf{0} \in \mathcal{S}_1$ it follows that

$$\|\mathbf{r}_n\|_2 \leq \|\mathbf{r}_{n-1}\|_2 \tag{17}$$

for all $n \geq 1$. From (17) we have that either there exists $k \in \mathbb{Z}^+$ such that

$$\|\mathbf{r}_{k+1}\|_2 = \|\mathbf{r}_k\|_2 < \|\mathbf{r}_j\|_2 \text{ for } j < k, \tag{18}$$

that is, $k = \operatorname{rank}_{\otimes}^G (A^{-1}\mathbf{f}) < \infty$, or

$$\|\mathbf{r}_{n+1}\|_2 < \|\mathbf{r}_n\|_2, \tag{19}$$

for all $n \geq 0$. The later condition implies that $\operatorname{rank}_{\otimes}^G (A^{-1}\mathbf{f}) = \infty$. In particular we obtain that the sequence $\{\mathbf{r}_n\}_{n=0}^{\operatorname{rank}_{\otimes}^G (A^{-1}\mathbf{f})}$ is strictly decreasing.

Assume that $\operatorname{rank}_{\otimes}^G (A^{-1}\mathbf{f}) = k < \infty$. Since if $\mathbf{x}_i \in \mathbb{R}^{N_i}$ for $i = 1, 2, \ldots, d$, then $\lambda \mathbf{x}_1 \otimes \mathbf{x}_2 \otimes_k \cdots \otimes \mathbf{x}_d \in \mathcal{S}_1$ for all $\lambda \in \mathbb{R}$. Consequently, we have

$$\|\mathbf{r}_k - A(\lambda \mathbf{x}_1 \otimes \mathbf{x}_2 \otimes_k \cdots \otimes \mathbf{x}_d)\|_2 \geq \|\mathbf{r}_k\|_2 \tag{20}$$

for every $\lambda \in \mathbb{R}$. This implies that $\mathbf{r}_k$ and $A(\mathbf{x}_1 \otimes \mathbf{x}_2 \otimes_k \cdots \otimes \mathbf{x}_d)$ are orthogonal. Since the vectors $\mathbf{x}_i \in \mathbb{R}^{N_i}$ for $i = 1, 2, \ldots, d$, are arbitrary and A has full rank, this means that $\mathbf{r}_k = \mathbf{0}$ and in consequence $\mathbf{u}_k = A^{-1}\mathbf{f}$ and $\mathbf{r}_j = \mathbf{0}$ for all $j \geq k$.

Let be
$$\mathcal{D} = \{\mathbf{w} : \mathbf{w} = A\mathbf{y}, \|\mathbf{w}\|_2 = 1 \text{ and } \mathbf{y} \in \mathcal{S}_1\},$$

then
$$\min_{\mathbf{y} \in \mathcal{S}_1} \|\mathbf{b} - A\mathbf{x}\|_2 = \min_{\mathbf{w} \in \mathcal{D}, \lambda \in \mathbb{R}} \|\mathbf{b} - \lambda \mathbf{w}\|_2$$

Note that
$$\|\mathbf{b} - \lambda \mathbf{w}\|_2^2 = \|\mathbf{b}\|_2^2 - 2\lambda \langle \mathbf{b}, \mathbf{w} \rangle + \lambda^2, \tag{21}$$

then the minimum for $\|\mathbf{b} - \lambda \mathbf{w}\|_2^2$ when $\mathbf{w} \in \mathcal{D}$ is obtained for $\lambda = \langle \mathbf{b}, \mathbf{w} \rangle > 0$ and
$$\|\mathbf{b} - \lambda \mathbf{w}\|_2^2 = \|\mathbf{b}\|_2^2 - |\langle \mathbf{b}, \mathbf{w} \rangle|^2.$$

Thus,
$$\min_{\mathbf{y} \in \mathcal{S}_1} \|\mathbf{b} - A\mathbf{y}\|_2^2 = \min_{\lambda \in \mathbb{R}, \, \mathbf{w} \in \mathcal{D}} \|\mathbf{b} - \lambda \mathbf{w}\|_2^2 \tag{22}$$

$$= \min_{\mathbf{w} \in \mathcal{D}} \|\mathbf{b} - \langle \mathbf{b}, \mathbf{w} \rangle \mathbf{w}\|_2^2 \tag{23}$$

$$= \|\mathbf{b}\|_2^2 - \max_{\mathbf{w} \in \mathcal{D}} |\langle \mathbf{b}, \mathbf{w} \rangle|^2. \tag{24}$$

By using (16) and (22)-(24) we have that for each $1 \leq n \leq \operatorname{rank}_{\otimes}^G (A^{-1}\mathbf{f})$ there exists $\mathbf{w}_n \in \mathcal{D}$, such that
$$A\mathbf{y}_n = \langle \mathbf{r}_{n-1}, \mathbf{w}_n \rangle \mathbf{w}_n. \tag{25}$$

Thus,
$$\|A\mathbf{y}_n\|_2 = \langle \mathbf{r}_{n-1}, \mathbf{w}_n \rangle > 0,$$

and
$$\langle \mathbf{r}_{n-1}, \mathbf{w}_n \rangle \geq |\langle \mathbf{r}_{n-1}, \mathbf{w} \rangle|, \tag{26}$$

for all $\mathbf{w} \in \mathcal{D}$.

From (22)-(24) and (25) we obtain
$$\|\mathbf{r}_n\|_2^2 = \|\mathbf{r}_{n-1} - A\mathbf{y}_n\|_2^2 \tag{27}$$

$$= \|\mathbf{r}_{n-1}\|_2^2 - |\langle \mathbf{r}_{n-1}, \mathbf{w}_n \rangle|^2 \tag{28}$$

$$= \|\mathbf{r}_{n-1}\|_2^2 \left(1 - \rho_n^2\right), \tag{29}$$

for all $1 \leq n \leq \operatorname{rank}_{\otimes}^{G}(A^{-1}\mathbf{f})$ and where

$$\rho_n = \frac{\langle \mathbf{r}_{n-1}, \mathbf{w}_n \rangle}{\|\mathbf{r}_{n-1}\|_2} = \frac{\langle \mathbf{r}_{n-1}, A\mathbf{y}_n \rangle}{\|\mathbf{r}_{n-1}\|_2 \|A\mathbf{y}_n\|_2} = \cos\theta_n \in (0,1).$$

Proceeding inductively from (27)-(29) follows (12).

From (28) we obtain that

$$\|\mathbf{r}_n\|_2^2 = \|\mathbf{f}\|_2^2 - \sum_{j=1}^{n} |\langle \mathbf{r}_{j-1}, \mathbf{w}_j \rangle|^2 \tag{30}$$

for all $1 \leq n \leq \operatorname{rank}_{\otimes}^{G}(A^{-1}\mathbf{f})$. Since the sequence $\{\|\mathbf{r}_n\|_2\}_{n=0}^{\infty}$ is decreasing and bounded below, there exists

$$\lim_{n\to\infty} \|\mathbf{r}_n\|_2^2 = R^2 = \|\mathbf{f}\|_2^2 - \lim_{n\to\infty} \sum_{j=1}^{n} |\langle \mathbf{r}_{j-1}, \mathbf{w}_j \rangle|^2.$$

In consequence,

$$\sum_{j=1}^{\infty} |\langle \mathbf{r}_{j-1}, \mathbf{w}_j \rangle|^2 = \sum_{j=1}^{\operatorname{rank}_{\otimes}^{G}(A^{-1}\mathbf{f})} |\langle \mathbf{r}_{j-1}, \mathbf{w}_j \rangle|^2$$

is a convergent series. Thus $\lim_{n\to\infty} |\langle \mathbf{r}_{n-1}, \mathbf{w}_n \rangle|^2 = 0$, and from (26) we have that

$$\lim_{n\to\infty} |\langle \mathbf{r}_{n-1}, \mathbf{w} \rangle|^2 = 0 \tag{31}$$

for all $\mathbf{w} \in \mathcal{D}$. Recall that if $\operatorname{rank}_{\otimes}^{G}(A^{-1}\mathbf{f}) < \infty$ then $\mathbf{r}_j = \mathbf{0}$ for all $j \geq \operatorname{rank}_{\otimes}^{G}(A^{-1}\mathbf{f})$. Thus, to end the proof of theorem we need to show that if $\operatorname{rank}_{\otimes}^{G}(A^{-1}\mathbf{f}) = \infty$, then the sequence $\{\mathbf{r}_n\}_{n=0}^{\infty}$ is convergent in $\mathbb{R}^{N_1 N_2 \cdots N_d}$ with the $\|\cdot\|_2$-topology. Because if it is true, let $\mathbf{g} = \lim_{n\to\infty} \mathbf{r}_{n-1}$. From (31), we have $\langle \mathbf{g}, \mathbf{w} \rangle = 0$ for all $\mathbf{w} \in \mathcal{D}$. Since $\mathcal{S}_1$ contains a basis of $\mathbb{R}^{N_1 N_2 \cdots N_d}$ and A has full rank, then we obtain $\mathbf{g} = \mathbf{0}$. Finally, $\lim_{n\to\infty} \mathbf{r}_{n-1} = \mathbf{0}$ and the theorem follows.

Thus, to conclude the proof we only need to show that the residuals sequence $\{\mathbf{r}_n\}_{n=0}^{\infty}$ is a Cauchy sequence. To prove it the following three lemmas will be useful.

Lemma 3 *For each $m, n \geq 1$,*

$$|\langle \mathbf{r}_{n-1}, A\mathbf{y}_m \rangle| \leq \langle \mathbf{r}_{n-1}, \mathbf{w}_n \rangle \langle \mathbf{r}_{m-1}, \mathbf{w}_m \rangle$$

Proof Since

$$|\langle \mathbf{r}_{n-1}, A\mathbf{y}_m \rangle| = |\langle \mathbf{r}_{n-1}, \langle \mathbf{r}_{m-1}\mathbf{w}_m \rangle \mathbf{w}_m \rangle| = |\langle \mathbf{r}_{n-1}, \mathbf{w}_m \rangle| \langle \mathbf{r}_{m-1}\mathbf{w}_m \rangle \leq \langle \mathbf{r}_{n-1}, \mathbf{w}_n \rangle \langle \mathbf{r}_{m-1}\mathbf{w}_m \rangle,$$

the lemma follows. $\square$

Lemma 4 *For every $\varepsilon > 0$ and $N \in \mathbb{N}$, there exists $\tau \geq N$ such that*

$$\langle \mathbf{r}_{\tau-1}, \mathbf{w}_\tau \rangle \sum_{k=1}^{\tau} \langle \mathbf{r}_{k-1}, \mathbf{w}_k \rangle \leq \varepsilon. \tag{32}$$

Proof Since $\sum_{j=1}^{\infty}\langle\mathbf{r}_{k-1},\mathbf{w}_k\rangle^2 < \infty$, for a given $\varepsilon > 0$ and $N \in \mathbb{N}$ we choose $n \geq N$ such that $\sum_{j=n+1}^{\infty}\langle\mathbf{r}_{k-1},\mathbf{w}_k\rangle^2 \leq \varepsilon/2$. Since $\lim_{k\to\infty}\langle\mathbf{r}_{k-1},\mathbf{w}_k\rangle = 0$, we can construct a map $\tau : \mathbb{N} \longrightarrow \mathbb{N}$, defined inductively by $\tau(1) = 1$ and

$$\tau(k+1) = \min_{m \geq \tau(k)}\left\{\langle\mathbf{r}_{m-1},\mathbf{w}_m\rangle \leq \langle\mathbf{r}_{\tau(k)-1},\mathbf{w}_{\tau(k)}\rangle\right\}$$

for all $k \geq 1$, such that τ is strictly increasing and $\lim_{k\to\infty}\tau(k) = \infty$. Observe that for all $k \geq 1$ and j satisfying that $\tau(k) < j < \tau(k+1)$ it follows

$$\langle\mathbf{r}_{\tau(k+1)-1},\mathbf{w}_{\tau(k+1)}\rangle \leq \langle\mathbf{r}_{\tau(k)-1},\mathbf{w}_{\tau(k)}\rangle \leq \langle\mathbf{r}_{j-1},\mathbf{w}_j\rangle.$$

Thus

$$\langle\mathbf{r}_{\tau(k+1)-1},\mathbf{w}_{\tau(k+1)}\rangle \leq \langle\mathbf{r}_{j-1},\mathbf{w}_j\rangle,$$

for all $1 \leq j < \tau(k+1)$. Now, we can choose $\tau = \tau(k+1) > n$ large enough, satisfying that

$$\langle\mathbf{r}_{\tau-1},\mathbf{w}_\tau\rangle\sum_{k=1}^{n}\langle\mathbf{r}_{k-1},\mathbf{w}_k\rangle \leq \varepsilon/2,$$

because $\lim_{k\to\infty}\langle\mathbf{r}_{\tau(k)-1},\mathbf{w}_{\tau(k)}\rangle = 0$. Then

$$\langle\mathbf{r}_{\tau-1},\mathbf{w}_\tau\rangle\sum_{k=1}^{\tau}\langle\mathbf{r}_{k-1},\mathbf{w}_k\rangle = \langle\mathbf{r}_{\tau-1},\mathbf{w}_\tau\rangle\sum_{k=1}^{n}\langle\mathbf{r}_{k-1},\mathbf{w}_k\rangle + \langle\mathbf{r}_{\tau-1},\mathbf{w}_\tau\rangle\sum_{k=n+1}^{\tau}\langle\mathbf{r}_{k-1},\mathbf{w}_k\rangle$$

$$\leq \varepsilon/2 + \sum_{k=n+1}^{\tau}\langle\mathbf{r}_{\tau-1},\mathbf{w}_\tau\rangle\langle\mathbf{r}_{k-1},\mathbf{w}_k\rangle$$

$$\leq \varepsilon/2 + \sum_{k=n+1}^{\tau}\langle\mathbf{r}_{k-1},\mathbf{w}_k\rangle^2$$

$$\leq \varepsilon/2 + \sum_{k=n+1}^{\infty}\langle\mathbf{r}_{k-1},\mathbf{w}_k\rangle^2 \leq \varepsilon.$$

This proves the lemma. $\square$

Lemma 5 *For each $M > N > 0$ it follows that*

$$\|\mathbf{r}_{N-1} - \mathbf{r}_{M-1}\|_2^2 \leq \|\mathbf{r}_{N-1}\|_2^2 - \|\mathbf{r}_{M-1}\|_2^2 + \langle\mathbf{r}_{M-1},\mathbf{w}_M\rangle\sum_{k=1}^{M}\langle\mathbf{r}_{k-1},\mathbf{w}_k\rangle.$$

Proof Since $\mathbf{r}_{N-1} = \mathbf{r}_{M-1} + \sum_{k=N}^{M-1}A\mathbf{y}_k$ we have

$$\|\mathbf{r}_{N-1} - \mathbf{r}_{M-1}\|_2^2 = \|\mathbf{r}_{N-1}\|_2^2 + \|\mathbf{r}_{M-1}\|_2^2 - 2\left\langle\mathbf{r}_{M-1} + \sum_{k=N}^{M-1}A\mathbf{y}_k, \mathbf{r}_{M-1}\right\rangle$$

$$= \|\mathbf{r}_{N-1}\|_2^2 - \|\mathbf{r}_{M-1}\|_2^2 - 2\sum_{k=N}^{M-1}\langle A\mathbf{y}_k,\mathbf{r}_{M-1}\rangle$$

$$\leq \|\mathbf{r}_{N-1}\|_2^2 - \|\mathbf{r}_{M-1}\|_2^2 + 2\sum_{k=N}^{M-1}\langle\mathbf{r}_{M-1},\mathbf{w}_M\rangle\langle\mathbf{r}_{k-1},\mathbf{w}_k\rangle \text{ by Lemma 3,}$$

$$\leq \|\mathbf{r}_{N-1}\|_2^2 - \|\mathbf{r}_{M-1}\|_2^2 + 2\sum_{k=1}^{M}\langle\mathbf{r}_{M-1},\mathbf{w}_M\rangle\langle\mathbf{r}_{k-1},\mathbf{w}_k\rangle \text{ by adding positive terms.}$$

This ends the proof of the lemma. □

Since $\lim_{n \to \infty} \|\mathbf{r}_n\|_2^2 = R^2$, and it is a decreasing sequence, for a given $\varepsilon > 0$ there exists $k_\varepsilon > 0$ such that

$$R^2 \le \|\mathbf{r}_{m-1}\|_2^2 \le R^2 + \varepsilon^2/8$$

for all $m \ge k_\varepsilon$. Assume that $m > k_\varepsilon$. From Lemma 4, for each $m + p$ there exists $\tau > m + p$ such that

$$\langle \mathbf{r}_{\tau-1}, \mathbf{w}_\tau \rangle \sum_{k=1}^{\tau} \langle \mathbf{r}_{k-1}, \mathbf{w}_k \rangle \le \varepsilon^2/8.$$

Now, we would to estimate

$$\|\mathbf{r}_{m-1} - \mathbf{r}_{m+p-1}\|_2 \le \|\mathbf{r}_{m-1} - \mathbf{r}_{\tau-1}\|_2 + \|\mathbf{r}_{\tau-1} - \mathbf{r}_{m+p-1}\|_2.$$

By using Lemma 5 with $M = \tau$ and $N = m$ and $m + p$, respectively, we obtain that

$$\|\mathbf{r}_{m-1} - \mathbf{r}_{\tau-1}\|_2^2 \le R^2 + \varepsilon^2/8 - R^2 + \varepsilon^2/8 = \varepsilon^2/4,$$

and

$$\|\mathbf{r}_{m+p-1} - \mathbf{r}_{\tau-1}\|_2^2 \le R^2 + \varepsilon^2/8 - R^2 + \varepsilon^2/8 = \varepsilon^2/4,$$

respectively. In consequence $\{\mathbf{r}_n\}_{n=0}^{\infty}$ is a Cauchy sequence and it converges to $\mathbf{0}$. □

3 A Block Coordinated Descent Approach for the Rank-One Minimization Problem

In this section we study the Rank-One minimization problem

$$\min_{\mathbf{x} \in \mathcal{S}_1} \|\mathbf{b} - A\mathbf{x}\|_2, \tag{33}$$

that we can write as the following unconstrained optimization problem:

$$\min_{(\mathbf{x}_1, \cdots, \mathbf{x}_d) \in \mathbb{R}^{N_1 + \cdots + N_d}} \|\mathbf{b} - A\left(\mathbf{x}_1 \otimes \cdots \otimes \mathbf{x}_d\right)\|_2. \tag{34}$$

A popular method for minimizing a real-valued continuously differentiable function Φ of $N_1 + \cdots + N_d$ real variables, subject to bound constrains, is the (block) coordinate descend method. In this method, the coordinates are partitioned into N_k blocks and, at each iteration, $\mathbf{b}$ is minimized with respect to one of the coordinate blocks while the others are held fixed (see Algorithm 2). These cyclic methods have the advantage of not requiring any information about the gradient to determine the descent directions. However, their convergence properties are poorer than steepest descend methods. Moreover, its are attractive because of their easy implementation in some particular cases as we will see below.

The main result of this section is the following.

Algorithm 2 A Block Coordinated Descent Algorithm

1: Given $\Phi : \mathbb{R}^{N_1 + \cdots + N_d} \to \mathbb{R}$
2: Initialize $\mathbf{x}_i^0 \in \mathbb{R}^{N_i}$ for $i = 1, 2 \ldots, d$.
3: **for** $n = 1, 2, \ldots$ **do**
4:     **for** $k = 1, 2, \ldots, d$ **do**
5:         $\mathbf{x}_k^n \in \arg\min_{\mathbf{x}_k \in \mathbb{R}^{N_k}} \Phi(\mathbf{x}_1^n, \ldots, \mathbf{x}_{k-1}^n, \mathbf{x}_k, \mathbf{x}_{k+1}^{n-1}, \ldots, \mathbf{x}_d^{n-1})$
6:     **end for**
7: **end for**

Theorem 2 *Let* $\mathbf{b} \in \mathbb{R}^{N_1 N_2 \cdots N_d}$ *and* $A \in \mathbb{R}^{N_1 N_2 \cdots N_d \times N_1 N_2 \cdots N_d}$, *be an invertible matrix. Assume that for each* $k \in \{1, 2, \ldots, d\}$ *the* $(N_1 \cdots N_d) \times N_k$-*matrix*

$$Z_k = A \left(\mathbf{x}_1 \otimes \cdots \otimes \mathbf{x}_{k-1} \otimes I_{N_k} \otimes \mathbf{x}_{k+1} \cdots \otimes \mathbf{x}_d \right)$$

has linearly independent columns for every $(\mathbf{x}_1, \ldots, \mathbf{x}_d) \in \mathbb{R}^{N_1 + \cdots + N_d}$ *satisfying*

$$\|\mathbf{b} - A (\mathbf{x}_1 \otimes \cdots \otimes \mathbf{x}_d)\|_2 \leq \left\| \mathbf{b} - A \left(\mathbf{x}_1^0 \otimes \cdots \otimes \mathbf{x}_d^0 \right) \right\|_2. \tag{35}$$

Then every accumulation point $(\mathbf{x}_1^*, \ldots, \mathbf{x}_d^*)$ *of the sequence*

$$\left\{ \left(\mathbf{x}_1^n, \ldots, \mathbf{x}_d^n \right) \right\}_{n=0}^\infty,$$

generated by Algorithm 2 using the map

$$\Phi (\mathbf{x}_1, \ldots, \mathbf{x}_d) = \|\mathbf{b} - A(\mathbf{x}_1 \otimes \cdots \otimes \mathbf{x}_d)\|_2,$$

satisfies the equation $\nabla \Phi (\mathbf{x}_1^*, \ldots, \mathbf{x}_d^*) = \mathbf{0}$. *Moreover, assume that* $\mathbf{x}_1, \ldots, \mathbf{x}_{k-1}, \mathbf{x}_{k+1}, \ldots, \mathbf{x}_d$, *are fixed for some* $k \in \{1, 2, \ldots, d\}$, *then*

$$\mathbf{x}_k = (Z_k^T Z_k)^{-1} Z_k^T \mathbf{b},$$

is the global minimum of the directional minimization problem

$$\min_{\mathbf{x} \in \mathbb{R}^{N_k}} \|\mathbf{b} - A (\mathbf{x}_1 \otimes \cdots \otimes \mathbf{x}_{k-1} \otimes \mathbf{x} \otimes \mathbf{x}_{k+1} \otimes \cdots \otimes \mathbf{x}_d)\|_2. \tag{36}$$

Proof To prove the theorem we will use the following lemma, that it can be proved by using the same argument as the proof of Theorem 5.32 in [13].

Lemma 6 *Assume that the function* $\Phi : \mathbb{R}^{N_1 + \cdots + N_d} \to \mathbb{R}$ *is continuously differentiable and that the set*

$$X_1 = \left\{ (\mathbf{x}_1, \ldots, \mathbf{x}_d) \in \mathbb{R}^{N_1 + \cdots + N_d} : \Phi(\mathbf{x}_1, \ldots, \mathbf{x}_d) \leq \Phi(\mathbf{x}_1^0, \ldots, \mathbf{x}_d^0) \right\}$$

is bounded. Moreover, assume that for every $(\mathbf{x}_1, \ldots, \mathbf{x}_d) \in X_1$ *the directional minimization problem*

$$\min_{\mathbf{x}_k \in \mathbb{R}^{N_k}} \Phi(\mathbf{x}_1, \ldots, \mathbf{x}_{k-1}, \mathbf{x}_k, \mathbf{x}_{k+1}, \ldots, \mathbf{x}_d)$$

has a unique solution for each $k = 1, 2, \ldots, d$. *Then every accumulation point* $(\mathbf{x}_1^*, \ldots, \mathbf{x}_d^*)$ *of the sequence*

$$\left\{ \left(\mathbf{x}_1^n, \ldots, \mathbf{x}_d^n \right) \right\}_{n=0}^\infty,$$

generated by Algorithm 2 satisfies the equation $\nabla \Phi(\mathbf{x}_1^*, \ldots, \mathbf{x}_d^*) = \mathbf{0}$.

Since

$$\mathbf{x}_1 \otimes \cdots \otimes \mathbf{x}_{k-1} \otimes \mathbf{x} \otimes \mathbf{x}_{k+1} \cdots \otimes \mathbf{x}_d$$
$$= \left(\mathbf{x}_1 \otimes \cdots \otimes \mathbf{x}_{k-1} \otimes I_k \otimes \mathbf{x}_{k+1} \cdots \otimes \mathbf{x}_d\right) \left(1 \otimes \cdots \otimes 1 \otimes \mathbf{x} \otimes 1 \cdots \otimes 1\right)$$
$$= \left(\mathbf{x}_1 \otimes \cdots \otimes \mathbf{x}_{k-1} \otimes I_k \otimes \mathbf{x}_{k+1} \cdots \otimes \mathbf{x}_d\right) \mathbf{x},$$

we can write

$$\|\mathbf{b} - A\left(\mathbf{x}_1 \otimes \cdots \otimes \mathbf{x}_d\right)\|_2 = \|\mathbf{b} - Z_k \mathbf{x}_k\|_2 . \tag{37}$$

Then, for a fixed $\mathbf{x}_1, \ldots, \mathbf{x}_{k-1}, \mathbf{x}_{k+1}, \ldots, \mathbf{x}_d$ the directional minimization problem (36) is equivalent to the standard Least Squares problem

$$\min_{\mathbf{x} \in \mathbb{R}^{N_k}} \|\mathbf{b} - Z_k \mathbf{x}\|_2 . \tag{38}$$

Since if Z_k has linearly independent columns then (38) has a unique solution

$$\mathbf{x}_k = (Z_k^T Z_k)^{-1} Z_k^T \mathbf{b},$$

from Lemma 6, the theorem follows. □

Note that given a point $(\mathbf{x}_1, \ldots, \mathbf{x}_d)$, descend with respect to the coordinate $\mathbf{x}_k$ means, from Theorem 2, that we need to solve the standard least squares problem (38). In particular we minimize Φ cyclically with respect to the coordinate variables Thus, Theorem 2 allow to us to solve the rank-one minimization problem (33) by means the Alternate Least Squares (ALS) Algorithm 3. We point out that for high-dimensional problems the numerical implementation of solving the equation

$$Z_k^T Z_k \mathbf{x} = Z_k^T \mathbf{x}, \tag{39}$$

can be a hardly task. However, if the matrix A can be represented also in separated representation form, then as the following corollary shows (39) can be implemented in a more easy way by using the properties of the Kronecker product.

Algorithm 3 ALS Algorithm

1: **procedure** ALS($\mathbf{b}, A, \texttt{iter\_max}, \texttt{tol}$)
2:     Initialize $\mathbf{x}_i^0$ for $i = 1, 2 \ldots, d$.
3:     $\texttt{iter} = 1$
4:     **while** $\texttt{iter} < \texttt{iter\_max}$ **do**
5:         $\hat{\mathbf{x}}_k \leftarrow \mathbf{x}_k^0, k = 1, \ldots, d$
6:         **for** $k = 1, 2, \ldots, d$ **do**
7:             $Z = A(\mathbf{x}_1^0 \otimes \cdots \otimes \mathbf{x}_{k-1}^0 \otimes I_{N_k} \otimes \hat{\mathbf{x}}_{k+1} \otimes \cdots \otimes \hat{\mathbf{x}}_d)$
8:             $\mathbf{x}_k^0 = (Z^T Z)^{-1} Z^T \mathbf{b}$
9:         **end for**
10:       **if** $\prod_{k=1}^d \|\mathbf{x}_k^0 - \hat{\mathbf{x}}_k\|_2 < \texttt{tol}$ **then goto** 14
11:       **end if**
12:       $\texttt{iter} = \texttt{iter} + 1$
13:     **end while**
14:     **return** $\mathbf{x}^0 = (\mathbf{x}_1^0, \ldots, \mathbf{x}_d^0)$
15: **end procedure**

Corollary 1 *Assume that*

$$A = \sum_{j=1}^{r_A} A_1^j \otimes \cdots \otimes A_d^j,$$

where $A_i^j \in \mathbb{R}^{N_i \times N_i}$ for $i = 1, 2, \ldots, d$ and $j = 1, 2, \ldots, r_A$. Let $k \in \{1, 2, \ldots, d\}$ and assume that for $\mathbf{x}_1, \ldots, \mathbf{x}_{k-1}, \mathbf{x}_{k+1}, \ldots, \mathbf{x}_d$ fixed, the $(N_1 \cdots N_d) \times N_k$-matrix

$$Z_k = \sum_{j=1}^{r_A} A_1^{(j)} \mathbf{x}_1 \otimes \cdots \otimes A_{k-1}^{(j)} \mathbf{x}_{k-1} \otimes A_k^{(j)} \otimes A_{k+1}^{(j)} \mathbf{x}_{k+1} \cdots \otimes A_d^{(j)} \mathbf{x}_d,$$

has linearly independent columns. Then

$$\mathbf{x}_k^* = (Z_k^T Z_k)^{-1} Z_k^T \mathbf{b}. \tag{40}$$

is the global minimum of of the directional minimization problem (36).

This corollary implies that we can solve the minimization problem

$$\min_{(\mathbf{x}_1, \ldots, \mathbf{x}_d)} \left\| \mathbf{b} - \sum_{i=1}^{r_A} A_1^i \mathbf{x}_1 \otimes \cdots \otimes A_d^i \mathbf{x}_d \right\|_2, \tag{41}$$

by means the ALS Algorithm 4.

Algorithm 4 ALS with A is separable form

1: **procedure** $\mathrm{ALS}(\mathbf{b}, \sum_{i=1}^{r_A} A_1^i \otimes \cdots \otimes A_d^i, \texttt{iter\_max}, \texttt{tol})$
2:     Initialize $\mathbf{x}_i^0$ for $i = 1, 2 \ldots, d$.
3:     $\texttt{iter} = 1$
4:     **while** $\texttt{iter} < \texttt{iter\_max}$ **do**
5:         $\hat{\mathbf{x}}_k \leftarrow \mathbf{x}_k^0, k = 1, \ldots, d$
6:         **for** $k = 1, 2, \ldots, d$ **do**
7:             $Z = \sum_{j=1}^{r_A} A_1^{(j)} \mathbf{x}_1^0 \otimes \cdots \otimes A_{k-1}^{(j)} \mathbf{x}_{k-1}^0 \otimes A_k^{(j)} \otimes A_{k+1}^{(j)} \hat{\mathbf{x}}_{k+1} \cdots \otimes A_d^{(j)} \hat{\mathbf{x}}_d$
8:             $\mathbf{x}_k^0 = (Z^T Z)^{-1} Z^T \mathbf{b}$
9:         **end for**
10:         **if** $\prod_{k=1}^d \|\mathbf{x}_k^0 - \hat{\mathbf{x}}_k\|_2 < \texttt{tol}$ **then goto** 14
11:         **end if**
12:         $\texttt{iter} = \texttt{iter} + 1$
13:     **end while**
14:     **return** $\mathbf{x}^0 = (\mathbf{x}_1^0, \ldots, \mathbf{x}_d^0)$
15: **end procedure**

---

4 Numerical examples

In order to illustrate the Greedy Rank-One Update Algorithm 1 by using the Alternate Least Squares Algorithms 3 and 4 we give the following examples. First, we compute for a given vector $\mathbf{f} \in \mathbb{R}^{14^3}$ its approximation by a vector $\sum_{j=1}^n \mathbf{f}_j^1 \otimes \mathbf{f}_j^2 \otimes \mathbf{f}_j^3$. To this end we consider

$$A = I \otimes I \otimes I$$

where I is the identity matrix of size 14×14. We take as a choice of $\mathbf{f}$ a random vector. Then for the parameter values `iter_max` $= 10$, `rank_max` $= 1000$ `tol` $= 2.22e - 16$, and $\varepsilon = 1.0e - 08$, we compute the sequences $\{\theta_j\}_{j=1}^n$, given by (13), and $\{\|\mathbf{r}^j\|_2\}_{j=0}^n$ As can be seen in Figure 1, the first one is an oscillating sequence around approximately 1.39, and, as Theorem 1 shows, the norm residuals is strictly decreasing. The algorithm stopped at $n = 1000$ after 8.9 seconds with a relative error equal to $2.86238020895006e - 08$.
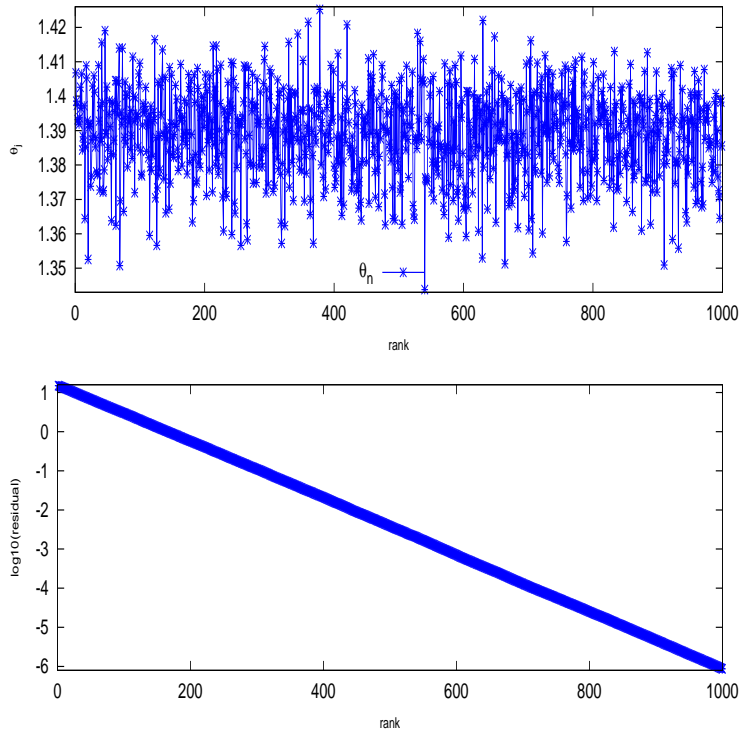
Fig. 1 The sequences $\{\theta_j\}_{j=1}^{1000}$, and $\{\log_{10} \|\mathbf{r}^j\|_2\}_{j=0}^{1000}$ when A is the identity matrix.

Now, in our second example we would to study the relationship between the sequence of angles $\{\theta_j\}_{j=1}^n \subset (0, \pi/2)$, given in the statement of Theorem1, and the sequence of the norm of the residuals $\{\|\mathbf{r}_j\|_2\}_{j=1}^n$. Here n is the $\text{rank}_{\otimes}^G \widehat{\mathbf{u}}$ where $\widehat{\mathbf{u}}$ is a numerical approximation of $A^{-1}\mathbf{f}$. Motivated by the above example we consider the mean angle, denoted by

$$\bar{\theta}_n = \frac{1}{n} \sum_{j=1}^n \theta_j,$$

of a finite sequence $\{\theta_j\}_{j=1}^n$. Now, we perform the following numerical experiment. We fix the values $N_1 = 3$, $N_2 = 4$, `tol` $= 2.22e - 16$, $\varepsilon = 1.0e - 08$, `iter_max` $= 10$ and `rank_max` $= 1000$. We repeat the following procedure 500-times. First, we generated randomly a matrix $A \in \mathbb{R}^{N_1 N_2 \times N_1 N_2}$, and a solution $A^{-1}\mathbf{f}$, in order to construct the right-side hand $\mathbf{f}$ of the linear system. Finally, we run Algorithm 1 to obtain an approximate solution $\widehat{\mathbf{u}}$ with $\text{rank}_{\otimes}^G \widehat{\mathbf{u}}$ and a mean angle $\overline{\theta}_{\text{rank}_{\otimes}^G \widehat{\mathbf{u}}}$.

In Figure 2 we can see a sample obtained in this experiment, for this particular case the approximate solution $\widehat{\mathbf{u}}$ had $\text{rank}_{\otimes}^G \widehat{\mathbf{u}} = 62$, with a relative error equal to $9.47228661927500e - 08$. Moreover, the the matrix A was nearly singular with a condition number equal to 488. Note, that the angle sequence, as we can see in Figure 2, oscillates around $\overline{\theta}_{\text{rank}_{\otimes}^G \widehat{\mathbf{u}}} = 0.949051119522545$.

In Figure 3 we plot the mean angle as a function of the $\text{rank}_{\otimes}^G$ of the approximate solution. From this picture we can deduce that some weakly linear dependence exists.
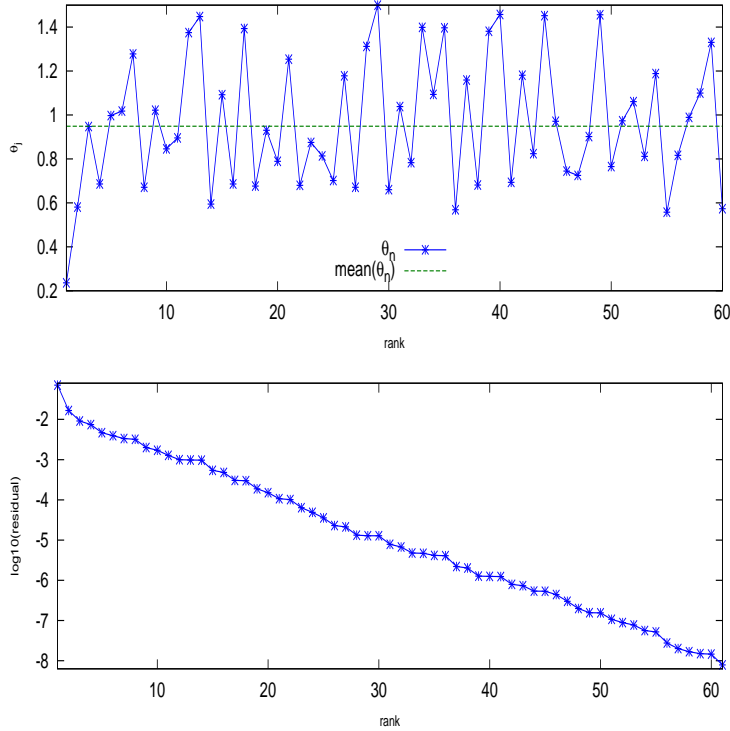

Fig. 2 The sequences $\{\theta_j\}_{j=1}^{62}$, and $\{\log_{10} \|\mathbf{r}^j\|_2\}_{j=0}^{62}$ when we take as a choice of A a 20×20 random matrix.
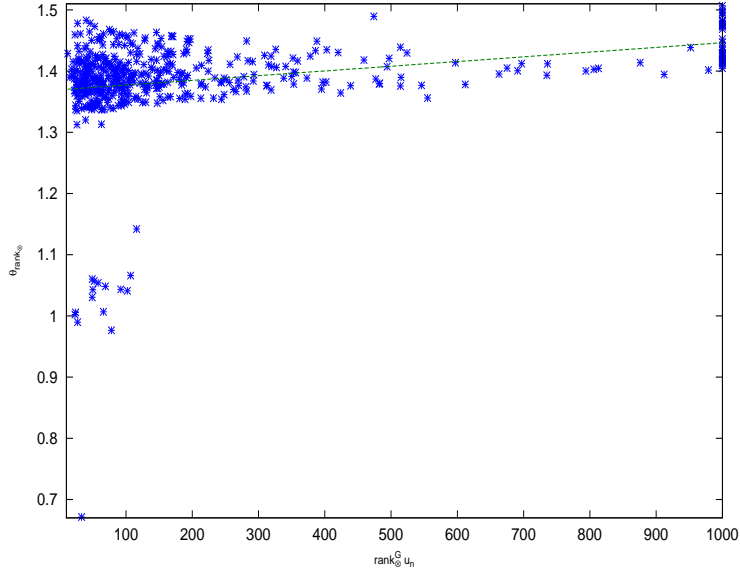
Fig. 3 The mean angle $\overline{\theta}_{\text{rank}_{\otimes}^G \, \hat{\mathbf{u}}}$ as a function of $\text{rank}_{\otimes}^G \, \hat{\mathbf{u}}$. The discontinuous line is the least squares straight line $y = (7.69351915752358e - 05)x + 1.36946939179450$.

4.1 A model problem: The Poisson equation in $(0, 1)^d$

In this section we present the following example. Recall that the Poisson problem reads

$$\begin{cases} -\Delta u = f \text{ in } \Omega \subset \mathbb{R}^d \\ \quad u|_{\partial\Omega} = 0. \end{cases} \tag{42}$$

where $f = f(x_1, x_2, \ldots, x_d)$ is a given function and $\Delta = \sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2}$ is the Laplace operator. In order to find its variational formulation, we recall the following Green formula for the Laplacian:

$$-\int_{\Omega} \Delta u \, v \, d\mathbf{x} = \int_{\Omega} \nabla u \cdot \nabla v \, d\mathbf{x} - \int_{\partial\Omega} \frac{\partial u}{\partial n} \, v \, d\gamma. \tag{43}$$

Assume that $\Omega = (0, 1)^d$. and for $d = 1, 2, \ldots$ let be the bilinear form

$$a_d(u; v) = \int_{\Omega} \nabla u \cdot \nabla v \, d\mathbf{x}.$$

Take $d = 3$ in (42) and then we can easily deduce that u satisfies the following problem: Find $u \in H_0^1(\Omega)$ such that

$$a_3(u; v) = \int_{\Omega} f \, v d\mathbf{x} \text{ for all } v \in H_0^1(\Omega). \tag{44}$$

The Galerkin approximation to (44) reads:

$$\text{find } u_h \in V_h : a_3(u_h; v_h) = \int_\Omega f\, v_h d\mathbf{x} \ \forall v_h \in V_h. \tag{45}$$

Assume $V_h = P \otimes P \otimes P$ where $P = \text{span}\{\omega^1, \ldots, \omega^N\}$ and $\omega^1, \ldots, \omega^N$ in $H_0^1(0,1)$ are the following N−linearly independent maps. First, we partitioned the interval $[0,1]$ into N-parts

$$0 = \widehat{x}_1 < \widehat{x}_2 < \ldots < \widehat{x}_{N+1} = 1.$$

Denote $h_i = \widehat{x}_{i+1} - \widehat{x}_i$ and $h = \max_{1 \leq i \leq N} h_i$. For $i = 1, 2, \ldots, N-1$, let

$$\omega^{(i)}(x) = \begin{cases} (x - \widehat{x}_i)/h_i & \widehat{x}_i \leq x \leq \widehat{x}_{i+1} \\ (\widehat{x}_{i+2} - x)/h_{i+1} & \widehat{x}_{i+1} \leq x \leq \widehat{x}_{i+2}, \\ 0 & \text{otherwise} \end{cases}$$

These functions are continuous and piecewise linear. It is easy to see that they are linearly independent. The first order weak derivatives of the basis functions are piecewise constant. Indeed for $i = 1, 2, \ldots, N-1$

$$\frac{d}{dx}\,\omega^{(i)}(x) = \begin{cases} 1/h_i & \widehat{x}_i \leq x \leq \widehat{x}_{i+1} \\ -1/h_{i+1} & \widehat{x}_{i+1} \leq x \leq \widehat{x}_{i+2}, \\ 0 & \text{otherwise} \end{cases}$$

We assume that we have an uniform partition, that is, $h_i = h$ for $i = 1, 2, \ldots, N$. Then the following formulas are useful

$$\int_0^1 \frac{d}{dx}\,\omega^{(i)}(x)\frac{d}{dx}\,\omega^{(i-1)}(x)dx = -\frac{1}{h} \text{ for } 2 \leq i \leq N-1, \tag{46}$$

$$\int_0^1 \left(\frac{d}{dx}\,\omega^{(i)}(x)\right)^2 dx = \frac{2}{h}, \text{ for } 1 \leq i \leq N-1, \tag{47}$$

$$\int_0^1 \omega^{(i)}(x)\,\omega^{(i-1)}(x)dx = \frac{h}{6} \text{ for } 2 \leq i \leq N-1, \tag{48}$$

$$\int_0^1 \left(\omega^{(i)}(x)\right)^2 dx = \frac{2h}{3}, \text{ for } 1 \leq i \leq N-1, \tag{49}$$

Now, let the stiffness multilinear matrix $\mathbb{A}_3$ defined by

$$(\mathbb{A}_3)_{j_1,j_2,j_3;i_1,i_2,i_3} = a_3(\omega^{(i_1)} \otimes \omega^{(i_2)} \otimes \omega^{(i_3)}; \omega^{(j_1)} \otimes \omega^{(j_2)} \otimes \omega^{(j_3)})$$

$$= \int_\Omega \left(\frac{\partial \omega^{(i_1)}}{\partial x_1}\frac{\partial \omega^{(j_1)}}{\partial x_1}\omega^{(i_2)}\omega^{(j_2)}\omega^{(i_3)}\omega^{(j_3)}\right.$$

$$+\frac{\partial \omega^{(i_2)}}{\partial x_2}\frac{\partial \omega^{(j_2)}}{\partial x_2}\omega^{(i_1)}\omega^{(j_1)}\omega^{(i_3)}\omega^{(j_3)}$$

$$\left.\frac{\partial \omega^{(i_3)}}{\partial x_2}\frac{\partial \omega^{(j_3)}}{\partial x_2}\omega^{(i_1)}\omega^{(j_1)}\omega^{(i_2)}\omega^{(j_2)}\right) d\mathbf{x}$$

$$= A_{j_1;i_1}B_{j_2;i_2}B_{j_3;i_3} + B_{j_1;i_1}A_{j_2;i_2}B_{j_1;i_1} + B_{j_1;i_1}B_{j_2;i_2}A_{j_3;i_3},$$

where

$$A_{j,i} = a_1(\omega^{(i)}; \omega^{(j)}) \text{ and } B_{j;i} = \int_0^1 \omega^{(i)}(x)\omega^{(j)}(x)dx.$$

Then, by using (46)-(49), we have that

$$A = \begin{bmatrix} \frac{2}{h} & -\frac{1}{h} & & & \\ -\frac{1}{h} & \frac{2}{h} & -\frac{1}{h} & & \\ \ddots & \ddots & \ddots & & \\ & & -\frac{1}{h} & \frac{2}{h} & -\frac{1}{h} \\ & & & -\frac{1}{h} & \frac{2}{h} \end{bmatrix} \text{ and } B = \begin{bmatrix} \frac{2h}{3} & \frac{h}{6} & & & \\ \frac{h}{6} & \frac{2h}{3} & \frac{h}{6} & & \\ \ddots & \ddots & \ddots & & \\ & & \frac{h}{6} & \frac{2h}{3} & \frac{h}{6} \\ & & & \frac{h}{6} & \frac{2h}{3} \end{bmatrix}.$$

Finally, by using (2), we obtain that

$$(\mathbb{A}_3)_{j_1,j_2,j_3;i_1,i_2,i_3} = (A \otimes B \otimes B + B \otimes A \otimes B + B \otimes B \otimes A)_{s,t}.$$

On the other hand we have that

$$\int_\Omega f\,(\omega^{(j_1)} \otimes \omega^{(j_2)} \otimes \omega^{(j_3)})d\mathbf{x} = \mathcal{F}_{j_1,j_2,j_3}.$$

In order to solve (45) we consider a map in V_h given by

$$u_h = \sum_{i_1,i_2,i_3} \mathcal{U}_{i_1,i_2,i_3}\, \omega^{(i_1)} \otimes \omega^{(i_2)} \otimes \omega^{(i_3)},$$

where $\mathcal{U}_{i_1,i_2,i_3} \approx u(\widehat{x}_{i_1+1}, \widehat{x}_{i_2+1}, \widehat{x}_{i_3+1})$. Thus,

$$\mathcal{U}_{k,i_2,i_3} = \mathcal{U}_{i_1,k,i_3} = \mathcal{U}_{i_1,i_2,k} = 0$$

for $k \in \{0, N\}$. Substitute u_h in (45) with $v_h = \omega^{(j_1)} \otimes \omega^{(j_2)} \otimes \omega^{(j_3)}$. Then we obtain that

$$\sum_{i_1,i_2,i_3=1}^{N-1} (\mathbb{A}_3)_{j_1,j_2,j_3;i_1,i_2.i_3}\mathcal{U}_{i_1,i_2,i_3} = \mathcal{F}_{j_1,j_2,j_3} \tag{50}$$

must be hold for all $j_1, j_2, j_3 \in \{1, \ldots, N-1\}$. This implies that (45) is equivalent to solve the following linear system

$$(A \otimes B \otimes B + B \otimes A \otimes B + B \otimes B \otimes A)\mathbf{u} = \mathbf{f}, \tag{51}$$

here we consider that a general multi-index tensor can be represented by a standard vector by using that

$$\mathcal{V}_{j_1,j_2,\ldots,j_d} = \mathbf{v}_s$$

if and only if

$$s = j_d + \sum_{l=1}^{d-1}\left[(j_l - 1)\prod_{l'=l+1}^{d} N'_{l'}\right].$$

For a general $d \geq 2$ it can be shown that in order to solve numerically (42) we need to solve the following linear system:

$$\left(\sum_{j=1}^{d} A_1^{(j)} \otimes \cdots \otimes A_d^{(j)}\right)\mathbf{u} = \mathbf{f}. \tag{52}$$

where

$$A_k^{(j)} = \begin{cases} A & \text{if} \quad k = j, \\ B & \text{if} \quad k \neq j. \end{cases}$$

Finally, it is not difficult to show that A and B are symmetric and definite positive matrices. Then, from the properties 7-8 listed at the end of Section 1, $\sum_{j=1}^{d} A_1^{(j)} \otimes \cdots \otimes A_d^{(j)}$ is a symmetric and definite positive matrix. In consequence, it is invertible.

Some numerical examples

Example 1 Firstly, we consider the following problem in 3D: Solve for

$$(x_1, x_2, x_3) \in \Omega = (0,1)^3 :$$

$$-\Delta u = (2\pi)^2 \cdot 3 \cdot \sin(2\pi x_1 - \pi) \sin(2\pi x_2 - \pi) \sin(2\pi x_3 - \pi), \qquad (53)$$

$$u|_{\partial \Omega} = 0, \qquad (54)$$

which has as closed form solution

$$u(x_1, x_2, x_3) = \sin(2\pi x_1 - \pi) \sin(2\pi x_2 - \pi) \sin(2\pi x_3 - \pi).$$

We used the separable representation algorithm given in Section 2 with parameter values `iter_max` $= 5$, `rank_max` $= 1000$ and $\varepsilon = 0.001$. The algorithm give us an approximated solution $\mathbf{u}_1 \in \mathcal{S}_1$. In Figure 4 we represent the relative error of the solution computed using the separable representation algorithm, using logarithmic scale, as a function of the number of nodes used in the discretization of the Poisson equation. All the computations were performed using the GNU software OCTAVE in a AMD 64 Athlon K8 with 2Gib of RAM.

In Figure 5 we represent the CPU time, in logarithmic scale, used in solving the linear system (52) against the separable representation algorithm. In both cases all the linear systems involved were solved using the standard linear system solver (`A\b`) of OCTAVE.

Example 2 Finally we are addressing some highly multidimensional models. To this end we solve numerically (42) for $(x_1, \ldots, x_d) \in \Omega = (0,\pi)^d$ where

$$f = \sum_{k=1}^{d} -(1+k)\sin^{(-1+k)}(x_k)\left(-k\cos^2(x_k) + \sin^2(x_k)\right) \prod_{k'=1, k'\neq k}^{d} \sin^{(1+k')}(x_{k'}),$$

which has as closed form solution

$$u(x_1, \ldots, x_d) = \prod_{k=1}^{d} \sin^{(k+1)}(x_k).$$

Here we consider the true solution $\mathbf{u}$ given by $\mathcal{U}_{i_1, \ldots, i_d} = u(\widehat{x}_{i_1+1}, \ldots, \widehat{x}_{i_d+1})$. For $d = 10$ we use the parameter values `iter_max` $= 2$, `rank_max` $= 10$ and $\varepsilon = 0.001$. In a similar way as above the algorithm give us an approximated solution $\widehat{\mathbf{u}} \in \mathcal{S}_1$. In Figure 6 we represent the absolute error $\|\widehat{\mathbf{u}} - \mathbf{u}\|_2$ as a function of $h = \pi/N$ for $N = 5, 10, 20, \ldots, 160$ in $\log_{10}$-scale. By using similar parameters values the problem has been solved for $d = 100$ in about 20 minutes.
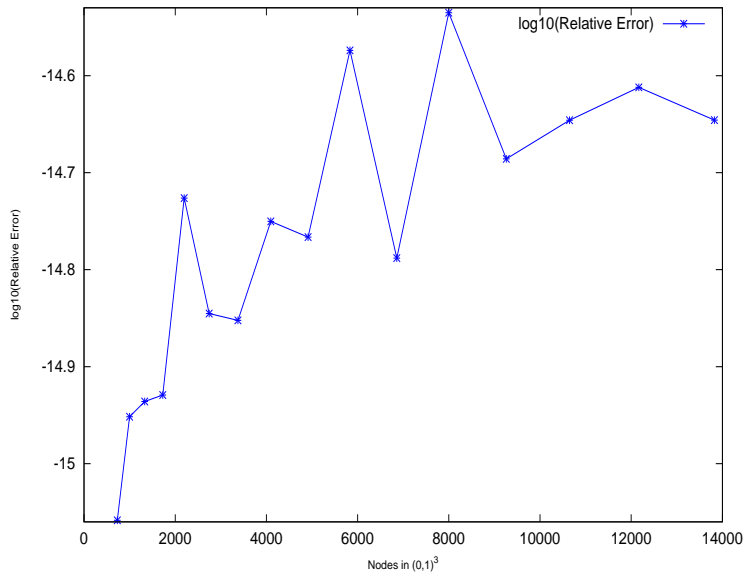
Fig. 4 The relative error $\|\mathbf{u}_1 - A^{-1}\mathbf{f}\|_2 / \|A^{-1}\mathbf{f}\|_2$ in logarithmic scale.

5 Concluding Remarks

In this paper we prove the convergence of the Greedy Rank-One Update Algorithm for solving linear systems with a full rank matrix. Moreover, we study the Rank-One Minimization Problem and show that a Block Cyclic Coordinate Descend strategy implies an Alternating Least Squares Algorithm. As we can show the method runs under very weak conditions, recall that we only use the assumption that the linear system has a an invertible matrix. However, its efficiency depends strongly on the matrix form (symmetric, tridiagonal, full, sparse, ...).

References

1. A. Ammar. B. Mokdad, F. Chinesta and R. Keunings. *A new family of solvers for some classes of multidimensional partial differential equations encountered in Kinetic Theory modelling Complex Fluids*, Journal of Non–Newtonian Fluid Mechanics, **139**, (2006), pp. 153-176.
2. A. Ammar. B. Mokdad, F. Chinesta and R. Keunings. *A new family of solvers for some classes of multidimensional partial differential equations encountered in Kinetic Theory modelling Complex Fluids. Part II: Transient Simulations using Space-Time Separated Representations* Journal of Non-Newtonian Fluid Mechanics, **144**, (2007), pp. 98-121
3. G. Beylkin and M. J. Mohlenkamp. *Algorithms for Numerical Analysis in High Dimensions.* SIAM J. Sci. Comput. Vol 26, No. 6 (2005). pp. 2133–2159.
4. V. de Silva and L.-H. Lim. *Tensor Rank and Ill-posedness of the best low-rank approximation problem*, SIAM Journal of Matrix Analysis & Appl. Vol. 30, no 3 (2008), pp. 1084–1127.
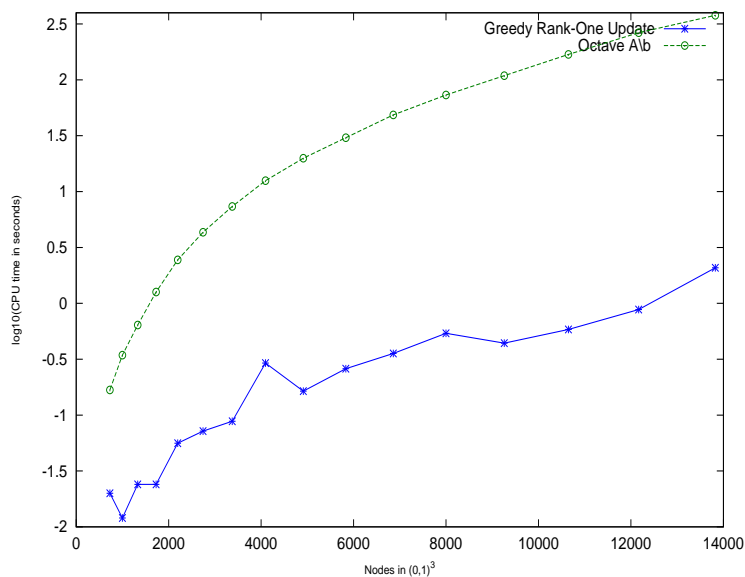
Fig. 5 The CPU time, in seconds, used in solving the linear system as a function of the number of nodes employed in the discretization of the Poisson Equation.

5. A. Graham. *Kronecker Products and Matrix Calculus with Applications*. John Wiley, 1981.
6. L. Grasdyck. *Existence and computation of low Kronecker-rank approximations for large linear systems of tensor product structure*, Computing 72 (2004), pp. 247-265
7. J. H. Friedman and W. Stuezle, *Projection pursuit regression*. J. Amer. Statist. Asoc., vol 76 (1981), pp. 817–823
8. P. J. Huber, *Projection pursuit*, Ann. Statist., vol 13, no. 2 (1985), pp. 435–475.
9. T. Kolda, *Orthogonal Tensor Decompositions*. SIAM J. Matrix Anal. Appl. Vol 23, No. 1 (2001), pp. 243–255.
10. L. de Lathauwer, B. de Moor and J. Vandewalle. *A Multilinear Singular Value Decomposition*. SIAM J. Matrix Anal. Appl. Vol 21, No. 4 (2000), pp. 1253–1278.
11. L. de Lathauwer, B. de Moor and J. Vandewalle. *On the best rank-1 and rank-$(R_1, \ldots, R_N)$ approximations of High Order Tensors* , SIAM J. Matrix Anal. Appl. Vol 21, No. 4 (2000), pp. 1324–1342.
12. S. Mallat and Z. Zhang, *Matching Pursuit with time-frequency dictionaries*. IEEE Trans. Signal Processing, **41**, Dec. 1993, pp. 3397–3415.
13. A. Ruszczyński. *Nonlinear Optimization*. Princeton University Pres, 2006.
14. V. N Temlyakov, *Greedy Approximation*. Acta Numerica (2008), pp. 235–409.
15. C. F. Van Loan. *The ubiquitous Kronecker product*, J. Comput. Appl. Math. **123** (2000), pp. 85-100.
16. T. Zhang and G. Golub, *Rank-one approximation to high order tensors*, SIAM J. Matrix Anal. Appl. Vol 21, No. 4 (2000), pp. 1253–1278.
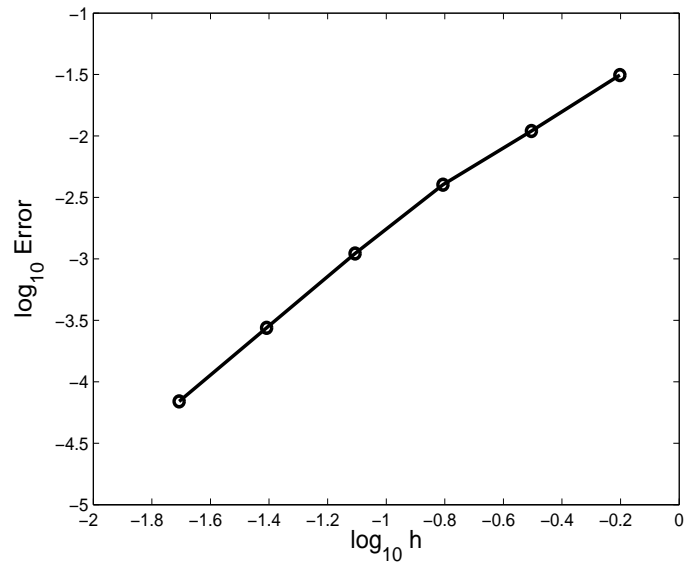
Fig. 6 The absolute error $\|\widehat{\mathbf{u}} - \mathbf{u}\|_2$ as a function of $h = \pi/N$ for $N = 5, 10, 20, \ldots, 160$ in $\log_{10}$-scale.