mathematics

MDPI

Article

Towards a Vector Field Based Approach to the Proper Generalized Decomposition (PGD)

Antonio Falcó [1,*], Lucía Hilario [1], Nicolás Montés [1], Marta C. Mora [2] and Enrique Nadal [3]

[1] ESI International Chair@CEU-UCH, Departamento de Matemáticas, Física y Ciencias Tecnológicas, Universidad Cardenal Herrera-CEU, CEU Universities San Bartolomé 55, 46115 Alfara del Patriarca, Spain; luciah@uchceu.es (L.H.); nicolas.montes@uchceu.es (N.M.)

[2] Departamento de Ingeniería Mecánica y Construcción, Universitat Jaume I, Avd. Vicent Sos Baynat s/n, 12071 Castellón, Spain; mmora@uji.es

[3] Departamento de Ingeniería Mecánica y de Materiales, Universitat Politècnica de València Camino de Vera, s/n, 46022 Valencia, Spain; ennaso@upvnet.upv.es

* Correspondence: afalco@uchceu.es

Abstract: A novel algorithm called the Proper Generalized Decomposition (PGD) is widely used by the engineering community to compute the solution of high dimensional problems. However, it is well-known that the bottleneck of its practical implementation focuses on the computation of the so-called best rank-one approximation. Motivated by this fact, we are going to discuss some of the geometrical aspects of the best rank-one approximation procedure. More precisely, our main result is to construct explicitly a vector field over a low-dimensional vector space and to prove that we can identify its stationary points with the critical points of the best rank-one optimization problem. To obtain this result, we endow the set of tensors with fixed rank-one with an explicit geometric structure.

Keywords: proper generalised decomposition; alternating least squares; greedy rank one update algorithm; tensor numerical methods

MSC: 15A69; 15A23; 57R25; 65N30

Citation: Falcó, A.; Hilario, L.; Montés, N.; Marta, M.C.; Nadal, E. Towards a Vector Field Based Approach to the Proper Generalized Decomposition (PGD). *Mathematics* **2021**, *9*, 34. https://dx.doi.org/10.3390/math9010034

Received: 11 October 2020
Accepted: 21 December 2020
Published: 25 December 2020

1. Introduction

Many problems in science and engineering are hard to compute due their numerical complexity. Moreover, in the analysis of complex systems under a real time constraint, the evaluation of all possible scenarios appears as a necessity [1]. Despite the improvements in techniques used in high dimensional problems, some challenging questions remain unresolved due to the efficiency of our computers. However, a novel technique called Proper Generalized Decomposition (PGD) [2,3] has been developed to provide an answer of these difficult tasks. It was initially proposed to compute, in a separated representation framework, the variational solution of partial differential equations (PDE) defined over a tensor product space [4]. It is possible to distinguish two different benefits. The first one is the possibility of managing high dimensional problems, and the second is the possibility to include the model's parameters as extra-coordinates. This last fact gives a powerful strategy to deal with classical problems because the PGD framework facilitates an efficient design and a real-time decision-making [5,6].

This novel technique allows for computing the whole set of solutions of a parametrized problem. The strategy is to include in an equivalent non-parametrized problem all possible parameter values as extra-coordinates. The name of this particular PGD based approach is Progressive Variational Vademecum [1], and it can be implemented offline. As a consequence, the PGD based approach opens the possibility of solving problems in industry with a different strategy not envisioned until now.

The mathematical analysis of the PGD was given by Falcó and Nouy, Ref. [4] in a Hilbert space framework and in [7] for a more general setting.

The Greedy Rank One Algorithm (GROA) [8] used to solve high-dimensional linear systems (with a full rank matrix) is the procedure of choice in the engineering community to implement the PGD. It is an iterative method made up of two steps that we can cyclically repeat until convergence. The first one consists of computing the minimal residual of the linear system over the set of tensors with bounded rank-one. In the second step, we use this optimal rank-one solution to update the residual. In the following, we return to the first step. Ammar et al. [8] propose an Alternating Least Squares (ALS) Algorithm for the practical implementation of the first step of the GROA. In the aforementioned paper, the authors justify the choice of the ALS showing that its convergence to a critical point of the optimization problem (not necessarily an optimal one) is assured under very weak conditions. In addition, the convergence has been studied by El Hamidi, Osman and Jazar [9] in the framework of Sobolev tensor spaces.

In this work, we want to study the optimization problem of the first step the GROA. To this end, even though our problem is not convex, we will take into account the relationship between the underlying convex optimization problem and the behaviour of its associated gradient flow. It is well-known that the vector field constructed by using a convex functional (for example related with a convex minimization problem defined over a finite dimensional vector space), has a gradient flow that provides a dynamical system with a unique stationary point. Moreover, it can be shown that it is a sink and its stable manifold coincides with the whole domain of the convex functional. This fact motivates the classical paradigm about the convergence of gradient-based numerical optimization algorithms.

In this work, we want to adapt the above paradigm in the framework of the GROA [8]. The main goal is to find a vector field in a low-dimensional vector space related to the gradient flow of a convex functional defined over the set of tensors of fixed rank one. The idea is to use this vector field to characterize the behaviour of the solutions of the non-convex optimization problem associated with the first step of GROA. To achieve this, we will prove that the set of critical points of the optimization problem over the set of tensors with fixed rank-one can be identified with the set of stationary points of that vector field. In order to construct it, we will proceed as follows. First, we will endow the set tensors of fixed rank-one with an explicit structure of smooth manifold. Second, by the help of this geometric structure, we will explicitly construct a vector field over a low dimensional vector space related with the first step of GROA. Finally, we will show that the set of stationary points of this vector field coincides with the set of critical points of the optimization problem associated with the PGD algorithm. In consequence, this vector field allows us to explain the dynamical behaviour around each of its stationary points. Moreover, we can get explicit information, in a neighbourhood of each of these stationary points, about the structure of its stable and unstable manifolds. In our opinion, a more precise knowledge of these invariant sets can help us develop better and more efficient PGD approaches.

The paper is organised as follows. Section 2 provides some preliminary definitions and results used along this paper. Section 3 shows a geometric approach to the PGD. In Section 4, the characterization of the smooth manifold of the set of tensors of fixed rank-one is given. After that, Section 5 shows the first order optimality conditions for the PGD which is the main result of this paper. Finally, Section 6 provides some conclusions of the work.

2. Preliminary Definitions and Results

First of all, we introduce some notation used along this paper. 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

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

To conclude, we list some of the well-known properties of the Kronecker product (see, for example, [10] or [11]).

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 [12], and it is related to the problem of constructing the approximate solutions of some classes of problems in high-dimensional spaces by means of 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

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

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 (1) by

$$u(x_{i_1}, \ldots, x_{i_d}) = \sum_{j=1}^{\infty} u_1^{(j)}(x_{i_1}) \cdots u_d^{(j)}(x_{i_d}), \tag{2}$$

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 (2) is equivalent to

$$\mathbf{u} = \sum_{j=1}^{\infty} \mathbf{x}_1^j \otimes \cdots \otimes \mathbf{x}_d^j. \tag{3}$$

We point out that (3) is a useful expression to implement numerical algorithms using the MATLAB and OCTAVE function `kron`.

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

$$A\mathbf{u} = \mathbf{f}, \tag{4}$$

where A is a $(N_1 \cdots N_d) \times (N_1 \cdots N_d)$-dimensional invertible matrix, for some $N_1, \ldots, N_d \in \mathbb{N}$, that is, $A \in \mathrm{GL}(N_1 \cdots N_d)$. 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 show that

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

that is,

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

Thus, in a first approach to solve it, we would like 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

$$\mathrm{argmin}_{\mathrm{rank}_{\otimes}\mathbf{u} \leq n} \|\mathbf{f} - A\mathbf{u}\|_2, \tag{6}$$

by using the notation introduced in [13].

The Proper Generalised Decomposition (PGD in short) appears when we consider solving the linear Equation (4) as an optimization problem as follows. For each fixed $A \in \mathrm{GL}(N_1 N_2 \cdots N_d)$ and $\mathbf{f} \in \mathbb{R}^{N_1 N_2 \cdots N_d}$, we define a map

$$J_{A,\mathbf{f}} : \mathbb{R}^{N_1 \cdots N_d} \to \mathbb{R}, \quad J_{A,\mathbf{f}}(\mathbf{u}) = \frac{1}{2}\|\mathbf{f} - A\mathbf{u}\|_2^2,$$

hence

$$\mathrm{argmin}_{\mathbf{u} \in \mathbb{R}^{N_1 N_2 \cdots N_d}} J_{A,\mathbf{f}}(\mathbf{u}) = \{A^{-1}\mathbf{f}\}$$

holds. The goal is to use (5) to approximate the solution of (4). To this end, 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 [13], 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 [13], 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 [13], it follows that $\mathcal{S}_1$ is a closed set in any norm-topology. This fact implies (see Lemma 1 in [8]), that given $A \in \mathrm{GL}(N_1 N_2 \cdots N_d)$, then, for every $\mathbf{f} \in \mathbb{R}^{N_1 \cdots N_d}$, we have that the set

$$\mathfrak{C}(A, \mathbf{f}) := \mathrm{argmin}_{\mathbf{x} \in \mathcal{S}_1} J_{A,\mathbf{f}}(\mathbf{x}) \neq \emptyset. \tag{7}$$

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

$$\begin{aligned} \mathbf{r}_{n-1} &= \mathbf{f} - A\mathbf{u}_{n-1}, & (8)\\ \mathbf{u}_n &= \mathbf{u}_{n-1} + \mathbf{y}_n \text{ where } \mathbf{y}_n \in \mathfrak{C}(A, \mathbf{r}_{n-1}). & (9) \end{aligned}$$

Note that, given $A \in \mathrm{GL}(N_1 N_2 \cdots N_d)$ and $\mathbf{f} \in \mathbb{R}^{N_1 \cdots N_d}$, we can construct for each n, by using (8) and (9), 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 Algorithms (8) and (9) as

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

The following theorem (see Theorem 1 in [8]) 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 $A \in \mathrm{GL}(N_1 N_2 \cdots N_d)$. Then, by using the iterative scheme (8) and (9), we obtain that the sequence $\{\|\mathbf{r}_n\|_2\}_{n=0}^{\mathrm{rank}^G_{\otimes}\,(A^{-1}\mathbf{f})}$, is strictly decreasing and*

$$A^{-1}\mathbf{f} = \lim_{n \to \infty} \mathbf{u}_n = \sum_{j=0}^{\mathrm{rank}^G_{\otimes}\,(A^{-1}\mathbf{f})} \mathbf{y}_j. \tag{10}$$

Moreover, the rate of convergence is given by

$$\frac{\|\mathbf{r}_n\|_2}{\|\mathbf{r}_0\|_2} = \prod_{j=1}^{n} \sin\theta_j \tag{11}$$

for $1 \leq n \leq \mathrm{rank}^G_{\otimes}\,(A^{-1}\mathbf{f})$ where

$$\theta_j = \arccos\left(\frac{\langle \mathbf{r}_{j-1}, A\mathbf{y}_j \rangle}{\|\mathbf{r}_{j-1}\|_2 \|A\mathbf{y}_j\|_2} \right) \in (0, \pi/2) \tag{12}$$

for $1 \leq j \leq n$.

From (10), we obtain that, if $\mathrm{rank}^G_{\otimes}\,(A^{-1}\mathbf{f}) < \infty$, then $\|\mathbf{r}_n\|_2 = 0$ for all $n > \mathrm{rank}^G_{\otimes}\,(A^{-1}\mathbf{f})$. Thus, the above theorem allows for us to construct a procedure, which 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 (7) (see the step 5 in Algorithm 1) and that we introduce below.

Algorithm 1 Greedy Rank-One Update

1: **procedure** GROU($\mathbf{f}, A, \varepsilon, \texttt{tol}, \texttt{rank\_max}$)
2: $\quad \mathbf{r}_0 = \mathbf{f}$
3: $\quad \mathbf{u} = \mathbf{0}$
4: $\quad$ **for** $i = 0, 1, 2, \ldots, \texttt{rank\_max}$ **do**
5: $\quad\quad \mathbf{y} =$ **procedure** $(\min_{\mathbf{x} \in \mathcal{S}_1} J_{A,\mathbf{r}_i}(\mathbf{x}))$
6: $\quad\quad \mathbf{r}_{i+1} = \mathbf{r}_i - A\mathbf{y}$
7: $\quad\quad \mathbf{u} \leftarrow \mathbf{u} + \mathbf{y}$
8: $\quad\quad$ **if** $\|\mathbf{r}_{i+1}\|_2 < \varepsilon$ or $|\|\mathbf{r}_{i+1}\|_2 - \|\mathbf{r}_i\|_2| < \texttt{tol}$ **then goto** 13
9: $\quad\quad$ **end if**
10: $\quad$ **end for**
11: $\quad$ **return** $\mathbf{u}$ and $\|\mathbf{r}_{\texttt{rank\_max}}\|_2$.
12: $\quad$ **break**
13: $\quad$ **return** $\mathbf{u}$ and $\|\mathbf{r}_{i+1}\|_2$
14: **end procedure**

3. A Geometric Approach to the PGD

In this section, to study the procedure given in the line 5 of the Algorithm 1, we introduce a smooth manifold. To this end, introduce first the set of tensors of fixed rank-one in the tensor space $\mathbb{R}^{N_1 \cdots N_d} = \bigotimes_{j=1}^{d} \mathbb{R}^{N_j}$ defined as

$$\mathcal{M}_{N_1 \cdots N_d} = \left\{ \mathbf{u} \in \bigotimes_{j=1}^{d} \mathbb{R}^{N_j} : \mathbf{u} = \lambda \bigotimes_{j=1}^{d} \mathbf{u}_j, \lambda \in \mathbb{R}_*, \mathbf{u}_j \in \mathbb{R}^{N_j} \setminus \{\mathbf{0}\}, 1 \leq j \leq d \right\}$$

where $\mathbb{R}_* = \mathbb{R} \setminus \{0\}$. Observe that the set

$$\mathcal{S}_1 = \{\mathbf{0}\} \cup \mathcal{M}_{N_1 \cdots N_d}.$$

Then, our first result is the following theorem of the alternative.

Theorem 2. *Let $A \in \mathrm{GL}(N_1 N_2 \cdots N_d)$ and $\mathbf{f} \in \mathbb{R}^{N_1 N_2 \cdots N_d}$. Either $\mathfrak{C}(A, \mathbf{f}) \subset \mathcal{M}_{N_1 \cdots N_d}$, or $\mathbf{0} \in \mathfrak{C}(A, \mathbf{f})$ but not both.*

Proof. Assume that $\mathbf{0} \in \mathfrak{C}(A, \mathbf{f})$ and that there exists $\mathbf{u} \in \mathfrak{C}(A, \mathbf{f}) \cap \mathcal{M}_{N_1 \cdots N_d}$. Since we can write

$$J_{A,\mathbf{f}}(\mathbf{u}) = \frac{1}{2}\|\mathbf{f}\|_2^2 - \mathbf{f}^T A\mathbf{u} + \frac{1}{2}\|A\mathbf{u}\|_2^2$$

and $J_{A,\mathbf{f}}(\mathbf{0}) = J_{A,\mathbf{f}}(\mathbf{u}) = \frac{1}{2}\|\mathbf{f}\|_2^2 \leq J_{A,\mathbf{f}}(\mathbf{u}')$ holds for all $\mathbf{u}' \in \mathcal{M}_{N_1 \cdots N_d}$, we have

$$\mathbf{f}^T A\mathbf{u} = \frac{1}{2}\|A\mathbf{u}\|_2^2 > 0.$$

Now, consider the map $f : \mathbb{R} \longrightarrow \mathbb{R}$ defined as

$$f(\lambda) := J_{A,\mathbf{f}}(\lambda\mathbf{u}) = \frac{1}{2}\|\mathbf{f}\|_2^2 - \lambda\, \mathbf{f}^T A\mathbf{u} + \frac{\lambda^2}{2}\|A\mathbf{u}\|_2^2 = \frac{1}{2}\|\mathbf{f}\|_2^2 - \mathbf{f}^T A\mathbf{u}\, \lambda + \mathbf{f}^T A\mathbf{u}\, \lambda^2.$$

Then, $f(0) = J_{A,\mathbf{f}}(\mathbf{0}) = J_{A,\mathbf{f}}(\mathbf{u}) = f(1) \leq f(\lambda)$ for all $\lambda \in \mathbb{R}$ holds. Observe that

$$f'(\lambda) = -\mathbf{f}^T A\mathbf{u} + 2\, \mathbf{f}^T A\mathbf{u}\, \lambda$$

and hence the map f has a global minimum for $\lambda = \frac{1}{2}$, a contradiction. $\square$

The main consequence of this result is the following. It says that the output of the procedure given in step 5 in Algorithm 1 always remains in the set $\mathcal{M}_{N_1\cdots N_d}$ before it gives us the final output.

Corollary 1. *Let $\mathbf{f} \in \mathbb{R}^{N_1N_2\cdots N_d}$ and $A \in \mathrm{GL}(N_1N_2\cdots N_d)$ such that $A^{-1}\mathbf{f} \notin \mathcal{S}_1$. Then, $\mathrm{rank}^G_{\otimes}(A^{-1}) = i-1$ for some $i > 1$ if and only if $\mathfrak{C}(A, \mathbf{r}_i) = \{\mathbf{0}\}$, and hence $\mathfrak{C}(A, \mathbf{r}_j) \subset \mathcal{M}_{N_1\cdots N_d}$ for all $j < i-1$.*

As a consequence of the above corollary, the situation of interest is when given $A \in \mathrm{GL}(N_1N_2\cdots N_d)$ and $\mathbf{f} \in \mathbb{R}^{N_1N_2\cdots N_d}$, we have that $\mathfrak{C}(A, \mathbf{f}) \subset \mathcal{M}_{N_1\cdots N_d}$. Thus, in order to study the vectors in $\mathfrak{C}(A, \mathbf{f})$, we need to characterize the structure of the critical points of the map $\mathbf{u} \mapsto \|\mathbf{f} - A\mathbf{u}\|$ restricted to the set $\mathcal{M}_{N_1\cdots N_d}$. To see this in the next section, we provide to $\mathcal{M}_{N_1\cdots N_d}$ of a structure of smooth manifold.

4. The Set of Tensors of Fixed Rank-One as a Smooth Manifold

Along this paper, we will consider a manifold as a pair $(\mathbb{M}, \mathcal{A})$, where $\mathbb{M}$ is a subset of some finite-dimensional vector space V and $\mathcal{A}$ is an atlas representing the local coordinate system of $\mathbb{M}$. We recall the definition of an atlas associated with a set $\mathbb{M}$.

Definition 1. *Let $\mathbb{M}$ be a set. An atlas of class C^p $(p \geq 0)$ or analytic on $\mathbb{M}$ is a family of charts with some indexing set A, namely $\{(U_\alpha, \varphi_\alpha) : \alpha \in A\}$, having the following properties (see [14]):*

AT1 *$\{U_\alpha\}_{\alpha\in A}$ is a covering of $\mathbb{M}$, that is, $U_\alpha \subset \mathbb{M}$ for all $\alpha \in A$ and $\cup_{\alpha\in A} U_\alpha = \mathbb{M}$.*
AT2 *For each $\alpha \in A$, $(U_\alpha, \varphi_\alpha)$ stands for a bijection $\varphi_\alpha : U_\alpha \to W_\alpha$ of U_α onto an open set W_α of a finite dimensional normed space $(X_\alpha, \|\cdot\|_\alpha)$, and for any α and β the set $\varphi_\alpha(U_\alpha \cap U_\beta)$ is open in X_α.*
AT3 *Finally, if we let $U_\alpha \cap U_\beta = U_{\alpha,\beta}$ and $\varphi_\alpha(U_{\alpha,\beta}) = U_{\alpha,\beta}$, the transition mapping $\varphi_\beta \circ \varphi_\alpha^{-1} : U_{\alpha,\beta} \to U_{\beta,\alpha}$ is a diffeomorphism of class $\mathcal{C}^p$ $(p \geq 0)$ or analytic.*

Observe that the condition of an open covering is not used, see [14]. Moreover, in AT2, we do not require that the normed spaces to be the same for all indices α, or even to be isomorphic. If X_α is linearly isomorphic to some finite dimensional normed space X for all α, we have the following definition.

Definition 2. *Let $\mathbb{M}$ be a set and X be a finite dimensional normed space. We say that $\mathbb{M}$ is a C^p (respectively, analytic) manifold modelled on X if there exists an atlas of class C^p (respectively, analytic) over $\mathbb{M}$ with X_α linearly isomorphic to X for all $\alpha \in A$.*

Since different atlases can give the same manifold, we say that two atlases are *compatible* if each chart of one atlas is compatible with the charts of the other atlas in the sense of AT3. One verifies that the relation of compatibility between atlases is an equivalence relation.

Definition 3. *An equivalence class of atlases of class C^p on $\mathbb{M}$, also denoted by $\mathcal{A}$, is said to define a structure of a C^p-manifold on $\mathbb{M}$, and hence we say that $(\mathbb{M}, \mathcal{A})$ is a finite dimensional manifold. In a similar way, if an equivalence class of atlases is given by analytic maps, then we say that $(\mathbb{M}, \mathcal{A})$ is an analytic finite dimensional manifold.*

For each $\mathbf{u} = \lambda \bigotimes_{j=1}^d \mathbf{u}_j \in \mathcal{M}_{N_1\cdots N_d}$, we construct a local chart as follows. Let be

$$\mathrm{span}\{\mathbf{u}_j\}^\perp = \{\mathbf{v}_j \in \mathbb{R}^{N_j} : \mathbf{v}_j^T \cdot \mathbf{u}_j = 0\},$$

the orthogonal complement of the linear space $\mathrm{span}\{\mathbf{u}_j\}$ for $1 \leq j \leq d$. Let us consider the set

$$\mathcal{U}_{\mathbf{u}} := \mathrm{span}\{\mathbf{u}_1\}^\perp \times \cdots \times \mathrm{span}\{\mathbf{u}_d\}^\perp \times \mathbb{R}_*,$$

which is an open and dense set of the finite-dimensional vector space

$$X_{\mathbf{u}} := \operatorname{span}\{\mathbf{u}_1\}^{\perp} \times \cdots \times \operatorname{span}\{\mathbf{u}_d\}^{\perp} \times \mathbb{R}.$$

Observe that the vector space $X_{\mathbf{u}}$ is linearly isomorphic to the vector space $\mathbb{R}^{N_1-1} \times \cdots \times \mathbb{R}^{N_d-1} \times \mathbb{R}$ for all $\mathbf{u} \in \mathcal{M}_{N_1 \cdots N_d}$.

Now, we introduce the set

$$\mathfrak{U}(\mathbf{u}) := \left\{ \mathbf{u}' \in \mathcal{M}_{N_1 \cdots N_d} : \mathbf{u}' = \beta \bigotimes_{j=1}^{d} (\mathbf{u}_j + \mathbf{w}_j), \begin{array}{l} \mathbf{w}_j \in \operatorname{span}\{\mathbf{u}_j\}^{\perp}, 1 \leq j \leq d \\ \beta \in \mathbb{R}_* \end{array} \right\}$$

in $\mathcal{M}_{N_1 \cdots N_d}$ for which we can construct a natural bijection:

$$\varphi_{\mathbf{u}} : \mathfrak{U}(\mathbf{u}) \longrightarrow \mathcal{U}_{\mathbf{u}}, \quad \mathbf{u}' \mapsto (\mathbf{w}_1, \ldots, \mathbf{w}_d, \beta) \Leftrightarrow \mathbf{u}' = \beta \bigotimes_{j=1}^{d} (\mathbf{u}_j + \mathbf{w}_j).$$

Then, we can state the following result.

Theorem 3. *The set $\mathcal{A}_{N_1 \cdots N_d} = \{(\mathfrak{U}(\mathbf{u}), \varphi_{\mathbf{u}}) : \mathbf{u} \in \mathcal{M}_{N_1 \cdots N_d}\}$ is an atlas for $\mathcal{M}_{N_1 \cdots N_d}$ and hence $(\mathcal{M}_{N_1 \cdots N_d}, \mathcal{A}_{N_1 \cdots N_d})$ is a $\mathcal{C}^{\infty}$-manifold modelled on*

$$\mathbb{R}^{N_1-1} \times \cdots \times \mathbb{R}^{N_d-1} \times \mathbb{R} \cong \mathbb{R}^{\log_2(2^{N_1 \cdots N_d}) - d + 1}.$$

Proof. Clearly, AT1 holds. To prove AT2 and AT3, let us consider $\mathbf{u}, \mathbf{u}' \in \mathcal{M}_{N_1 \cdots N_d}$ be such that $\mathfrak{U}(\mathbf{u}) \cap \mathfrak{U}(\mathbf{u}') \neq \emptyset$. Without loss of generality we may assume that $\mathbf{u} = \lambda \bigotimes_{j=1}^{d} \mathbf{u}_i$ and $\mathbf{u}' = \lambda' \bigotimes_{j=1}^{d} \mathbf{u}'_i$, where $\|\mathbf{u}_i\|_{i,2} = \|\mathbf{u}'_i\|_{i,2} = 1$ for $1 \leq i \leq d$. Then, for each $\mathbf{z} \in \mathfrak{U}(\mathbf{u}) \cap \mathfrak{U}(\mathbf{u}')$, there exists a unique $(\mathbf{w}_1, \ldots, \mathbf{w}_d, \beta) \in \mathcal{U}_{\mathbf{u}}$ and a unique $(\mathbf{w}'_1, \ldots, \mathbf{w}'_d, \beta') \in \mathcal{U}_{\mathbf{u}'}$ such that

$$\mathbf{z} = \beta \bigotimes_{j=1}^{d} (\mathbf{u}_j + \mathbf{w}_j) = \beta' \bigotimes_{j=1}^{d} (\mathbf{u}'_j + \mathbf{w}'_j).$$

Since

$$\operatorname{span}\{\mathbf{u}_j + \mathbf{w}_j\} = \operatorname{span}\{\mathbf{u}'_j + \mathbf{w}'_j\}$$

holds, there exists a unique $\lambda_j \in \mathbb{R}_*$ such that

$$\mathbf{u}_j + \mathbf{w}_j = \lambda_j (\mathbf{u}'_j + \mathbf{w}'_j) \text{ for } 1 \leq j \leq d. \tag{13}$$

Thus, multiplying (13) on the left side by $\mathbf{u}_j^T$ and, by using that $\mathbf{u}_j^T \mathbf{u}_j = 1$ and $\mathbf{u}_j^T \mathbf{w}_j = 0$, we obtain

$$\lambda_j = \frac{1}{\mathbf{u}_j^T \mathbf{u}'_j + \mathbf{u}_j^T \mathbf{w}'_j} \text{ for } 1 \leq j \leq d.$$

Hence,

$$\mathbf{w}_j = \frac{\mathbf{u}'_j + \mathbf{w}'_j}{\mathbf{u}_j^T \mathbf{u}'_j + \mathbf{u}_j^T \mathbf{w}'_j} - \mathbf{u}_j$$

defines a $\mathcal{C}^{\infty}$-function from the open set $\mathcal{V}_j = \{\mathbf{w}_j \in \operatorname{span}\{\mathbf{u}'_j\}^{\perp} : \mathbf{u}_j^T \mathbf{u}'_j + \mathbf{u}_j^T \mathbf{w}'_j \neq 0\}$ to $\operatorname{span}\{\mathbf{u}_j\}^{\perp}$ for each $1 \leq j \leq d$. Moreover,

$$1 + \|\mathbf{w}_j\|_{j,2}^2 = \left\| \frac{\mathbf{u}'_j + \mathbf{w}'_j}{\mathbf{u}_j^T \mathbf{u}'_j + \mathbf{u}_j^T \mathbf{w}'_j} \right\|_{j,2}^2$$

holds for $1 \leq j \leq d$. Observe that $\mathbf{z}$ can be written as

$$\mathbf{z} = \beta \bigotimes_{j=1}^{d} (\mathbf{u}_j + \mathbf{w}_j) = \left(\beta \prod_{j=1}^{d} \sqrt{1 + \|\mathbf{w}_j\|_{j,2}^2} \right) \bigotimes_{j=1}^{d} \mathbf{z}_j,$$

where

$$\mathbf{z}_j := \frac{1}{\sqrt{1 + \|\mathbf{w}_j\|_{j,2}^2}} \mathbf{u}_j + \frac{1}{\sqrt{1 + \|\mathbf{w}_j\|_{j,2}^2}} \mathbf{w}_j$$

has norm one for $1 \leq j \leq d$. In addition,

$$\mathbf{z} = \left(\beta' \prod_{j=1}^{d} \sqrt{1 + \|\mathbf{w}'_j\|_{j,2}^2} \right) \bigotimes_{j=1}^{d} \mathbf{z}'_j,$$

where

$$\mathbf{z}'_j := \frac{1}{\sqrt{1 + \|\mathbf{w}'_j\|_{j,2}^2}} \mathbf{u}'_j + \frac{1}{\sqrt{1 + \|\mathbf{w}'_j\|_{j,2}^2}} \mathbf{w}'_j$$

has norm one for $1 \leq j \leq d$. Thus,

$$\beta = \beta' \frac{\prod_{j=1}^{d} \sqrt{1 + \|\mathbf{w}'_j\|_{j,2}^2}}{\prod_{j=1}^{d} \sqrt{1 + \|\mathbf{w}_j\|_{j,2}^2}} = \beta' \frac{\prod_{j=1}^{d} \sqrt{1 + \|\mathbf{w}'_j\|_{j,2}^2}}{\prod_{j=1}^{d} \left\| \frac{\mathbf{u}'_j + \mathbf{w}'_j}{\mathbf{u}_j^T \mathbf{u}'_j + \mathbf{u}_j^T \mathbf{w}'_j} \right\|_{j,2}}$$

clearly defines a $\mathcal{C}^\infty$-function from the open set $\mathcal{V}_1 \times \cdots \times \mathcal{V}_d \times \mathbb{R}_* \subset \mathcal{U}_{\mathbf{u}'}$ to $\mathcal{U}_{\mathbf{u}}$. Finally, we conclude that

$$\varphi_{\mathbf{u}'}(\mathfrak{U}(\mathbf{u}) \cap \mathfrak{U}(\mathbf{u}')) = \mathcal{V}_1 \times \cdots \times \mathcal{V}_d \times \mathbb{R}_* \subset \mathcal{U}_{\mathbf{u}'},$$

the map

$$\varphi_{\mathbf{u}} \circ \varphi_{\mathbf{u}'}^{-1} : \varphi_{\mathbf{u}'}(\mathfrak{U}(\mathbf{u}) \cap \mathfrak{U}(\mathbf{u}')) \longrightarrow \varphi_{\mathbf{u}}(\mathfrak{U}(\mathbf{u}) \cap \mathfrak{U}(\mathbf{u}'))$$

is given by

$$(\mathbf{w}'_1, \ldots, \mathbf{w}'_d, \beta') \mapsto \left(\frac{\mathbf{u}'_1 + \mathbf{w}'_1}{\mathbf{u}_1^T \mathbf{u}'_1 + \mathbf{u}_1^T \mathbf{w}'_1} - \mathbf{u}_1, \ldots, \frac{\mathbf{u}'_d + \mathbf{w}'_d}{\mathbf{u}_d^T \mathbf{u}'_d + \mathbf{u}_d^T \mathbf{w}'_d} - \mathbf{u}_d, \beta' \frac{\prod_{j=1}^{d} \sqrt{1 + \|\mathbf{w}'_j\|_{j,2}^2}}{\prod_{j=1}^{d} \left\| \frac{\mathbf{u}'_j + \mathbf{w}'_j}{\mathbf{u}_j^T \mathbf{u}'_j + \mathbf{u}_j^T \mathbf{w}'_j} \right\|_{j,2}}, \right)$$

and it is $\mathcal{C}^\infty$. This follows AT2, AT3 and concludes the proof of the theorem. □

The construction of $\mathcal{M}_{N_1 \cdots N_d}$ as an algebraic variety is well-known (see, for example, [15]). More recently, in [16], a structure of smooth manifold is given, in the framework of Banach spaces, for the set of tensors of fixed rank-one. Following [16], it can be shown that the manifold $\mathcal{M}_{N_1 \cdots N_d}$ is also a principal bundle as follows. Consider the Grassmann manifold of one-dimensional, subspaces of $\mathbb{R}^{N_j}$, denoted by $\mathbb{G}_1(\mathbb{R}^{N_j})$, for $1 \leq j \leq d$ and define the surjective map

$$\pi : \mathcal{M}_{N_1 \cdots N_d} \longrightarrow \mathbb{G}_1(\mathbb{R}^{N_1}) \times \cdots \times \mathbb{G}_1(\mathbb{R}^{N_d}), \quad \mathbf{u} = \lambda \bigotimes_{j=1}^{d} \mathbf{u}_j \mapsto (\operatorname{span}\{\mathbf{u}_j\})_{j=1}^{d}$$

Then, for each $\mathbf{u} = \lambda \bigotimes_{j=1}^{d} \mathbf{u}_j \in \mathcal{M}_{N_1 \cdots N_d}$, it holds that

$$\pi^{-1}(\pi(\mathbf{u})) = \bigotimes_{j=1}^{d} \operatorname{span}\{\mathbf{u}_j\} \setminus \{\mathbf{0}\} = \operatorname{span}\{\mathbf{u}\} \setminus \{\mathbf{0}\} \cong \mathbb{R}_*.$$

Consequently, $\mathcal{M}_{N_1\cdots N_d}$ is also a principal bundle with base space $\mathbb{G}_1(\mathbb{R}^{N_1}) \times \cdots \times \mathbb{G}_1(\mathbb{R}^{N_d})$ and fibre $\mathbb{R}_*$. It allows for decomposing the tangent space at $\mathbf{u} \in \mathcal{M}_{N_1\cdots N_d}$, denoted $\mathrm{T}_{\mathbf{u}}\mathcal{M}_{N_1\cdots N_d}$ into the vertical and horizontal spaces:

$$\mathrm{T}_{\mathbf{u}}\mathcal{M}_{N_1\cdots N_d} = X_{\mathbf{u}} = \mathrm{H}_{\mathbf{u}}\mathcal{M}_{N_1\cdots N_d} + \mathrm{V}_{\mathbf{u}}\mathcal{M}_{N_1\cdots N_d},$$

where

$$\mathrm{H}_{\mathbf{u}}\mathcal{M}_{N_1\cdots N_d}. := \mathrm{span}\{\mathbf{u}_1\}^{\perp} \times \cdots \times \mathrm{span}\{\mathbf{u}_d\}^{\perp}$$

and $\mathrm{V}_{\mathbf{u}}\mathcal{M}_{N_1\cdots N_d} := \mathbb{R}$.

5. On the First Order Optimality Conditions for the PGD

The goal of this section is given $A \in \mathrm{GL}(N_1 N_2 \cdots N_d)$ and $\mathbf{f} \in \mathbb{R}^{N_1 N_2 \cdots N_d}$ characterize the points in the manifold $\mathcal{M}_{N_1\cdots N_d}$ satisfying the first order optimality conditions of the problem

$$\min_{\mathbf{z}\in\mathcal{M}_{N_1\cdots N_d}} J_{A,\mathbf{f}}(\mathbf{z}). \tag{14}$$

Recall that the map $J_{A,\mathbf{f}}$ is defined in the whole ambient space $\mathbb{R}^{N_1} \otimes \cdots \otimes \mathbb{R}^{N_d}$. We will denote its derivative at $\mathbf{u} \in \mathbb{R}^{N_1} \otimes \cdots \otimes \mathbb{R}^{N_d}$ by $J'_{A,\mathbf{f}}(\mathbf{u}) = (\mathbf{f} - A\mathbf{u})^T$, which is a bounded linear map from $\mathbb{R}^{N_1} \otimes \cdots \otimes \mathbb{R}^{N_d}$ to $\mathbb{R}$. From Theorem 3, we known that $\mathcal{M}_{N_1\cdots N_d}$ is a $\mathcal{C}^\infty$-manifold and hence it allows us to write the constrained map $J_{A,\mathbf{f}}|_{\mathcal{M}_{N_1\cdots N_d}}$ as follows. Since

$$J_{A,\mathbf{f}} : \mathbb{R}^{N_1} \otimes \cdots \otimes \mathbb{R}^{N_d} \longrightarrow \mathbb{R},$$

and $\mathcal{M}_{N_1\cdots N_d} \subset \mathbb{R}^{N_1} \otimes \cdots \otimes \mathbb{R}^{N_d}$, we can take into account the standard inclusion map

$$i : \mathcal{M}_{N_1\cdots N_d} \longrightarrow \mathbb{R}^{N_1} \otimes \cdots \otimes \mathbb{R}^{N_d}, \quad \mathbf{z} \mapsto \mathbf{z}$$

in order to write $J_{A,\mathbf{f}}|_{\mathcal{M}_{N_1\cdots N_d}} = (J_{A,\mathbf{f}} \circ i)$.

Definition 4. *We say that $\mathbf{u} \in \mathcal{M}_{N_1\cdots N_d}$ is a critical point for $J_{A,\mathbf{f}}$ in $\mathcal{M}_{N_1\cdots N_d}$ if*

$$T_{\mathbf{u}} J_{A,\mathbf{f}}(\mathbf{v}) := \left[D(J_{A,\mathbf{f}} \circ \varphi_{\mathbf{u}}^{-1})(\varphi_{\mathbf{u}}(\mathbf{u}))\right](\mathbf{v}) = 0$$

holds for all $\mathbf{v} \in \mathrm{T}_{\mathbf{u}}\mathcal{M}_{N_1\cdots N_d}$. Clearly, if $\mathbf{u}$ is an extremal point for $J_{A,\mathbf{f}}$ in $\mathcal{M}_{N_1\cdots N_d}$, then it is also a critical point for $J_{A,\mathbf{f}}$ in $\mathcal{M}_{N_1\cdots N_d}$.

Observe that we can write

$$J_{A,\mathbf{f}} \circ \varphi_{\mathbf{u}}^{-1} = J_{A,\mathbf{f}} \circ (i \circ \varphi_{\mathbf{u}}^{-1}),$$

where on the left side of the equality, we consider $J_{A,\mathbf{f}}$ over $\mathcal{M}_{N_1\cdots N_d}$, whereas, in the right one, $J_{A,\mathbf{f}}$ is considered defined over the whole space. Thus, by using the chain rule, we have

$$\begin{aligned} D(J_{A,\mathbf{f}} \circ \varphi_{\mathbf{u}}^{-1})(\varphi_{\mathbf{u}}(\mathbf{u})) &= D(J_{A,\mathbf{f}} \circ (i \circ \varphi_{\mathbf{u}}^{-1}))(\varphi_{\mathbf{u}}(\mathbf{u})) \\ &= J'_{A,\mathbf{f}}(\mathbf{u}) \circ D(i \circ \varphi_{\mathbf{u}}^{-1})(\varphi_{\mathbf{u}}(\mathbf{u})), \end{aligned}$$

that is,

$$T_{\mathbf{u}} J_{A,\mathbf{f}} = J'_{A,\mathbf{f}}(\mathbf{u}) \circ T_{\mathbf{u}} i = (\mathbf{f} - A\mathbf{u})^T \, T_{\mathbf{u}} i. \tag{15}$$

In order to compute (15), we consider first the standard inclusion map i that in local coordinates in a neighbourhood of $\mathbf{u} = \lambda \bigotimes_{j=1}^{d} \mathbf{u}_j \in \mathcal{M}_{N_1 \cdots N_d}$ looks like

$$(i \circ \varphi_{\mathbf{u}}^{-1})(\mathbf{w}_1, \ldots, \mathbf{w}_d, \beta) = \beta \bigotimes_{j=1}^{w} (\mathbf{u}_j + \mathbf{w}_j).$$

Here, $(i \circ \varphi_{\mathbf{u}}^{-1}) : \mathcal{U}_{\mathbf{u}} \longrightarrow \mathbb{R}^{N_1} \otimes \cdots \otimes \mathbb{R}^{N_d}$. Hence, its derivative as a morphism (a map between manifolds)

$$\mathrm{T}_{\mathbf{u}} i = D(i \circ \varphi_{\mathbf{u}}^{-1})(\mathbf{0}, \ldots, \mathbf{0}, \lambda) : \mathrm{T}_{\mathbf{u}} \mathcal{M}_{N_1 \cdots N_d} \longrightarrow \mathbb{R}^{N_1} \otimes \cdots \otimes \mathbb{R}^{N_d},$$

is given by

$$\mathrm{T}_{\mathbf{u}} i(\beta, \mathbf{w}_1, \ldots, \mathbf{w}_d) = \sum_{j=1}^{d} \mathbf{w}_j \otimes \mathbf{u}_{[j]} + \frac{\beta}{\lambda} \mathbf{u},$$

where

$$\mathbf{w}_j \otimes \mathbf{u}_{[j]} := \lambda\, \mathbf{u}_1 \otimes \cdots \otimes \mathbf{u}_{j-1} \otimes \mathbf{w}_j \otimes \mathbf{u}_{j+1} \otimes \cdots \otimes \mathbf{u}_d.$$

for $1 \leq j \leq d$. From (15), we have that $\mathbf{u}^* = \lambda^* \bigotimes_{j=1}^{d} \mathbf{u}_i^*$ is a critical point for $J_{A,\mathbf{f}}$ in $\mathcal{M}_{N_1 \cdots N_d}$ if and only if

$$(\mathbf{f} - A\mathbf{u}^*)^T\, T_{\mathbf{u}^*} i(\mathbf{w}_1, \ldots, \mathbf{w}_d, \beta) = \mathbf{0}, \text{ holds for all } (\mathbf{w}_1, \ldots, \mathbf{w}_d, \beta) \in \mathrm{T}_{\mathbf{u}^*} \mathcal{M}_{N_1 \cdots N_d}, \quad (16)$$

that is, it is equivalent to state that

$$(\mathbf{f} - A\,\mathbf{u}^*)^T \left(\sum_{j=1}^{d} \mathbf{w}_j \otimes \mathbf{u}_{[j]}^* + \frac{\beta}{\lambda} \mathbf{u}^* \right) = \mathbf{0}, \text{ holds for all } (\mathbf{w}_1, \ldots, \mathbf{w}_d, \beta) \in \mathrm{T}_{\mathbf{u}^*} \mathcal{M}_{N_1 \cdots N_d}. \quad (17)$$

Now, we will prove the following result that characterizes the set of critical points for $J_{A,\mathbf{f}}$ in $\mathcal{M}_{N_1 \cdots N_d}$ as the stationary points of a vector field in $\mathbb{R}^{N_1} \times \cdots \times \mathbb{R}^{N_d}$, and it is the main result of this paper.

Theorem 4. *Given $A \in \mathrm{GL}(N_1 N_2 \cdots N_d)$ and $\mathbf{f} \in \mathbb{R}^{N_1 N_2 \cdots N_d}$. Then, $\mathbf{u}^* = \bigotimes_{j=1}^{d} \mathbf{u}_i^* \in \mathcal{M}_{N_1 \cdots N_d}$ is a critical point for $J_{A,\mathbf{f}}$ in $\mathcal{M}_{N_1 \cdots N_d}$ if and only if $\mathbf{X}_{A,\mathbf{f}}(\mathbf{u}_1^*, \ldots, \mathbf{u}_d^*) = \mathbf{0}$, where $\mathbf{X}_{A,\mathbf{f}}$ is a vector field in $\mathcal{M}_{N_1 \cdots N_d}$ given by*

$$\mathbf{X}_{A,\mathbf{f}}(\mathbf{u}_1, \ldots, \mathbf{u}_d) = (\mathbf{y}_1, \ldots, \mathbf{y}_d).$$

Here,

$$\mathbf{y}_j := (\mathbf{u}_1 \otimes \cdots \otimes \mathbf{u}_{j-1} \otimes id_{\mathbb{R}^{N_j}} \otimes \mathbf{u}_{j+1} \otimes \cdots \otimes \mathbf{u}_d)^T (\mathbf{f} - A\,(\mathbf{u}_1 \otimes \cdots \otimes \mathbf{u}_d))$$

for $1 \leq j \leq d$.

Proof. Since $\mathbf{u}^* = \bigotimes_{j=1}^{d} \mathbf{u}_i^* \in \mathcal{M}_{N_1 \cdots N_d}$ is a critical point for $J_{A,\mathbf{f}}$ in $\mathcal{M}_{N_1 \cdots N_d}$ if and only if (17) holds for $(\mathbf{w}_1, \ldots, \mathbf{w}_d, \beta) \in \mathrm{T}_{\mathbf{u}^*} \mathcal{M}_{N_1 \cdots N_d}$. Then, (17) is equivalent to prove that

$$(\mathbf{f} - A\,\mathbf{u}^*)^T \mathbf{u}^* = \mathbf{0}, \quad (18)$$

and

$$(\mathbf{f} - A\,\mathbf{u}^*)^T (\mathbf{w}_j \otimes \mathbf{u}_{[j]}^*) = \mathbf{0} \text{ holds for all } \mathbf{w}_j \in \operatorname{span}\{\mathbf{u}_j\}^{\perp} \text{ and } 1 \leq j \leq d. \quad (19)$$

Since $\mathbf{w}_j \otimes \mathbf{u}^*_{[j]} = \mathbf{u}^*$ when $\mathbf{w}_j = \mathbf{u}^*_j$ for $1 \leq j \leq d$, we conclude that $\mathbf{u}^* = \bigotimes_{j=1}^d \mathbf{u}^*_i \in \mathcal{M}_{N_1 \cdots N_d}$ is a critical point for $J_{A,\mathbf{f}}$ in $\mathcal{M}_{N_1 \cdots N_d}$ if and only if

$$(\mathbf{f} - A\,\mathbf{u}^*)^T(\mathbf{w}_j \otimes \mathbf{u}^*_{[j]}) = \mathbf{0} \text{ holds for all } \mathbf{w}_j \in \mathbb{R}^{N_j} \text{ and } 1 \leq j \leq d. \tag{20}$$

Writing the equality in (20) as

$$(\mathbf{f} - A\,\mathbf{u}^*_1 \otimes \cdots \otimes \mathbf{u}^*_d)^T(\mathbf{u}^*_1 \otimes \cdots \otimes \mathbf{u}^*_{j-1} \otimes id_{\mathbb{R}^{N_j}} \otimes \mathbf{u}^*_{j+1} \otimes \cdots \otimes \mathbf{u}^*_d)\mathbf{w}_j = \mathbf{0},$$

we conclude that it is equivalent to state that

$$(\mathbf{f} - A\,\mathbf{u}^*_1 \otimes \cdots \otimes \mathbf{u}^*_d)^T(\mathbf{u}^*_1 \otimes \cdots \otimes \mathbf{u}^*_{j-1} \otimes id_{\mathbb{R}^{N_j}} \otimes \mathbf{u}^*_{j+1} \otimes \cdots \otimes \mathbf{u}^*_d) = \mathbf{0}^T$$

for $1 \leq j \leq d$. This concludes the proof of theorem □

Remark 1. *From (17), we can deduce that $\mathbf{u}^* = \bigotimes_{j=1}^d \mathbf{u}^*_i$ is a critical point for $J_{A,\mathbf{f}}$ in $\mathcal{M}_{N_1 \cdots N_d}$ if and only if*

$$(\mathbf{f} - A\,\mathbf{u}^*)^T\left(\sum_{j=1}^d \mathbf{w}_j \otimes \mathbf{u}^*_{[j]}\right) = \mathbf{0} \textit{ holds for all } (\mathbf{w}_1, \ldots, \mathbf{w}_d) \in \mathbb{R}^{N_1} \times \cdots \times \mathbb{R}^{N_d}. \tag{21}$$

Statement (21) was first introduced in [2,3] as a step to enrich the approximation basis in the PGD algorithm. From the proof of Theorem 4, we have that (21) is also equivalent to (18) and (19).

There are several strategies in the literature that can be used to solve $\mathbf{X}_{A,\mathbf{f}}(\mathbf{u}) = \mathbf{0}$. The first one, closely related to the one used in [2,3], is to find a fixed point of the map

$$F_{A,\mathbf{f}}(\mathbf{u}_1, \ldots, \mathbf{u}_d) := \mathbf{X}_{A,\mathbf{f}}(\mathbf{u}_1, \ldots, \mathbf{u}_d) - (\mathbf{u}_1, \ldots, \mathbf{u}_d)$$

One of the most popular numerical strategies to compute an approximated value of $\mathbf{u}^*$ such that $\mathbf{X}_{A,\mathbf{f}}(\mathbf{u}^*) = \mathbf{0}$ is founded under the following argument. Since the $\mathbf{y}_1$ is equal to

$$\begin{aligned}(id_{\mathbb{R}^{N_1}} \otimes \mathbf{u}_2 \otimes \cdots \otimes \mathbf{u}_d)^T(\mathbf{f} - A\,\mathbf{u}_1 \otimes \cdots \otimes \mathbf{u}_d) &= \\ (id_{\mathbb{R}^{N_1}} \otimes \mathbf{u}_2 \otimes \cdots \otimes \mathbf{u}_d)^T\mathbf{f} - (id_{\mathbb{R}^{N_1}} \otimes \mathbf{u}_2 \otimes \cdots \otimes \mathbf{u}_d)^T A\,(\mathbf{u}_1 \otimes \cdots \otimes \mathbf{u}_d) &= \\ (id_{\mathbb{R}^{N_1}} \otimes \mathbf{u}_2 \otimes \cdots \otimes \mathbf{u}_d)^T\mathbf{f} - (id_{\mathbb{R}^{N_1}} \otimes \mathbf{u}_2 \otimes \cdots \otimes \mathbf{u}_d)^T A\,(id_{\mathbb{R}^{N_1}} \otimes \mathbf{u}_2 \cdots \otimes \mathbf{u}_d)\mathbf{u}_1,&\end{aligned}$$

we can choose randomly some vectors $\mathbf{u}_j \in \mathbb{R}^{N_j}$ for $2 \leq j \leq d$ and then try to solve the linear system

$$(id_{\mathbb{R}^{N_1}} \otimes \mathbf{u}_2 \otimes \cdots \otimes \mathbf{u}_d)^T A\,(id_{\mathbb{R}^{N_1}} \otimes \mathbf{u}_2 \cdots \otimes \mathbf{u}_d)\mathbf{u}_1 = (id_{\mathbb{R}^{N_1}} \otimes \mathbf{u}_2 \otimes \cdots \otimes \mathbf{u}_d)^T\mathbf{f}.$$

By using Least Squares, we can compute a $\mathbf{u}'_1$, and next we can proceed in a similar way iteratively with each of the other components $\mathbf{y}_2, \ldots, \mathbf{y}_d$ as follows. After the step $i \geq 1$, we know that $\mathbf{u}'_1, \mathbf{u}'_2, \ldots, \mathbf{u}'_i$. Then, by choosing randomly $\mathbf{u}_{i+1}, \ldots \mathbf{u}_d$, we can solve, by using Least Squares, the linear system

$$\begin{aligned}(\mathbf{u}'_1 \otimes \cdots \otimes \mathbf{u}'_i \otimes id_{\mathbb{R}^{N_i}} \otimes \mathbf{u}_{i+1} \otimes \cdots \otimes \mathbf{u}_d)^T A\,(id_{\mathbb{R}^{N_1}} \otimes \mathbf{u}_2 \cdots \otimes \mathbf{u}_d)\mathbf{u}_i \\ = (\mathbf{u}'_1 \otimes \cdots \otimes \mathbf{u}'_i \otimes id_{\mathbb{R}^{N_i}} \otimes \mathbf{u}_{i+1} \otimes \cdots \otimes \mathbf{u}_d)^T\mathbf{f},\end{aligned}$$

to compute $\mathbf{u}'_i$. We can proceed cyclically until $\|\mathbf{f} - A(\mathbf{u}'_1 \otimes \cdots \otimes \mathbf{u}'_d)\| < \varepsilon$ holds. This strategy is known by the name of the Alternating Least Squares (ALS).

From (17), we have that $\mathbf{u}^*$ is a critical point of $J_{A,\mathbf{f}}$ in $\mathcal{M}_{N_1 \cdots N_d}$ if and only if

$$(\mathbf{f} - A\,\mathbf{u}^*) \perp \mathrm{T}_{\mathbf{u}^*} i(\mathrm{T}_{\mathbf{u}^*}\mathcal{M}_{N_1 \cdots N_d}) \tag{22}$$

holds, that is, the residual at $\mathbf{u}^*$ is orthogonal to linear subspace $\mathrm{T}_{\mathbf{u}^*} i(\mathrm{T}_{\mathbf{u}^*}\mathcal{M}_{N_1\cdots N_d})$. Observe that $\mathbf{w}_j \otimes \mathbf{u}_{[j]} \in V_j(\mathbf{u})$, where

$$V_j(\mathbf{u}) := \mathrm{span}\{\mathbf{u}_1\} \otimes \cdots \otimes \mathrm{span}\{\mathbf{u}_{j-1}\} \otimes \mathrm{span}\{\mathbf{u}_j\}^{\perp} \otimes \mathrm{span}\{\mathbf{u}_{j+1}\} \otimes \cdots \otimes \mathrm{span}\{\mathbf{u}_d\}$$

is an $N_j - 1$-dimensional subspace of $\mathbb{R}^{N_1} \otimes \cdots \otimes \mathbb{R}^{N_d}$ for $1 \leq j \leq d$. Then, the next result explicitly describes the linear subspace $\mathrm{T}_{\mathbf{u}^*} i(\mathrm{T}_{\mathbf{u}^*}\mathcal{M}_{N_1\cdots N_d})$.

Proposition 1. *The linear map $T_{\mathbf{u}} i : \mathrm{T}_{\mathbf{u}}\mathcal{M}_{N_1\cdots N_d} \longrightarrow \mathbb{R}^{N_1} \otimes \cdots \otimes \mathbb{R}^{N_d}$ is injective. Moreover,*

$$\mathrm{T}_{\mathbf{u}} i(\mathrm{T}_{\mathbf{u}}\mathcal{M}_{N_1\cdots N_d}) = \bigoplus_{j=1}^{d} V_j(\mathbf{u}) \oplus \mathrm{span}\{\mathbf{u}\} \cong \mathrm{T}_{\mathbf{u}}\mathcal{M}_{N_1\cdots N_d}.$$

Proof. Assume that

$$\mathrm{T}_{\mathbf{u}} i(\beta, \mathbf{w}_1, \ldots, \mathbf{w}_d) = \sum_{j=1}^{d} \mathbf{w}_j \otimes \mathbf{u}_{[j]} + \frac{\beta}{\lambda}\mathbf{u} = \mathbf{0}.$$

Since $\mathbf{u} \perp (\mathbf{w}_j \otimes \mathbf{u}_{[j]})$ for $1 \leq j \leq d$ and $(\mathbf{w}_i \otimes \mathbf{u}_{[j]}) \perp (\mathbf{w}_j \otimes \mathbf{u}_{[j]})$ for all $i \neq j$, then

$$\beta = 0 \text{ and } \mathbf{w}_i \otimes \mathbf{u}_{[j]} = \mathbf{0} \text{ for } 1 \leq j \leq d.$$

Since $\mathbf{u}_{[j]} \neq \mathbf{0}$, then $\mathbf{w}_j = \mathbf{0}$ for $1 \leq j \leq d$. This follows the first statement. To prove the second one, we remark that the inclusion

$$\mathrm{T}_{\mathbf{u}} i(\mathrm{T}_{\mathbf{u}}\mathcal{M}_{N_1\cdots N_d}) \subset \bigoplus_{j=1}^{d} V_j(\mathbf{u}) \oplus \mathrm{span}\{\mathbf{u}\}$$

between both subspaces is trivial. Clearly,

$$\dim \mathrm{T}_{\mathbf{u}} i(\mathrm{T}_{\mathbf{u}}\mathcal{M}_{N_1\cdots N_d}) = \dim \left(\bigoplus_{j=1}^{d} V_j(\mathbf{u}) \oplus \mathrm{span}\{\mathbf{u}\} \right)$$

holds and then the second statement is also proved. □

6. Conclusions

In this paper, a geometric upgrade of the PGD algorithm is proposed. To this end, we endow a set of tensors of fixed rank-one with a smooth manifold structure. This construction provides the solutions of a non-convex optimization problem by using a set of stationary points of a vector field. This new perspective allows for the characterization of the behaviour of the solutions of the Greedy Rank One Algorithm which could not be achieved with the ALS strategy used in [8].

Author Contributions: Investigation, writing and revision: A.F., L.H., N.M., M.C.M. and E.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the GVA/2019/124 grant from Generalitat Valenciana and by the RTI2018-093521-B-C32 grant from the Ministerio de Ciencia, Innovación y Universidades.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Falcó, A.; Montés, N.; Chinesta, F.; Hilario, L.; Mora, M.C. On the existence of a Progressive Variational Vademecum based on the Proper Generalized Decomposition for a Class of Elliptic Parametrized Problems. *J. Comput. Appl. Math.* **2018**, *330*, 1093–1107. [CrossRef]
2. Ammar, A.; Mokdad, B.; Chinesta, F.; Keunings, R. A new family of solvers for some classes of multidimensional partial differential equations encountered in Kinetic Theory modelling Complex Fluids. *J. Non-Newton. Fluid Mech.* **2006**, **139**, 153–176. [CrossRef]
3. Ammar, A.; Mokdad, B.; Chinesta, F.; Keunings, R. 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. *J. Non–Newton. Fluid Mech.* **2007**, **144**, 98–121. [CrossRef]
4. Falcó, A.; Nouy, A. A Proper Generalized Decomposition for the solution of elliptic problems in abstract form by using a functional Eckart- Young approach. *J. Math. Anal. Appl.* **2011**, *376*, 469–480. [CrossRef]
5. Chinesta, F.; Leygue, A.; Bordeu, F.; Aguado, J.V.; Cueto, E.; González, D.; Alfaro, I.; Ammar, A.; Huerta, A. PGD-Based computational vademecum for efficient Design, Optimization and Control. *Arch. Comput. Methods Eng.* **2013**, *20*, 31–49. [CrossRef]
6. Chinesta, F.; Ladevéze, P. *Separated Representations and PGD-Based Model Reduction*; Springer International Centre for Mechanical Sciences: Cham, Switzerland, 2014.
7. Falcó, A.; Nouy, A. Proper Generalized Decomposition for Nonlinear Convex Problems in Tensor Banach Spaces. *Numer. Math.* **2012**, *121*, 503–530. [CrossRef]
8. Ammar, A.; Chinesta, F.; Falcó, A. On the convergence of a greedy rank-one update algorithm for a class of linear systems. *Arch. Comput. Methods Eng.* **2010**, *17*, 473–486. [CrossRef]
9. El Hamidi, A.; Osman, H.; Jazar, M. On the convergence of alternating minimization methods in variational PGD. *Comput. Optim. Appl.* **2017**, *68*, 455–472. [CrossRef]
10. Graham, A. *Kronecker Products and Matrix Calculus with Applications*; John Wiley: Hoboken, NJ, USA, 1981.
11. Van Loan, C.F. The ubiquitous Kronecker product, *J.Comput. Appl. Math.* **2000**, *123*, 85–100. [CrossRef]
12. Beylkin, G.; Mohlenkamp, M.J. Algorithms for Numerical Analysis in High Dimensions. *SIAM J. Sci. Comput.* **2005**, *26*, 2133–2159. [CrossRef]
13. de Silva, V.; Lim, L.-H. Tensor Rank and Ill-posedness of the best low-rank approximation problem. *SIAM J. Matrix Anal. Appl.* **2008**, *30*, 1084–1127. [CrossRef]
14. Lang, S. *Differential and Riemannian Manifolds*; Graduate Texts in Mathematics; Springer: Berlin/Heidelberg, Germany, 1995; Volume 160.
15. Landsberg, J.M. Tensor: Geometry and Applications. Graduate Studies in Mathematics; American Mathematical Society: Providence, Rhode Island, 2012; Volume 128.
16. Falcó, A.; Hackbusch, W.; Nouy, A. On the Dirac-Frenkel Variational Principal on Tensor Banach Spaces. *Found. Comput. Math.* **2019**, *19*, 159–204. [CrossRef]