UNIVERSALITY FOR NON-LINEAR CONVEX VARIATIONAL PROBLEMS

PABLO M. BERNÁ[1,*], ANTONIO FALCÓ[2]

[1]Departamento de Métodos Cuantitativos, CUNEF Universidad, Madrid, 28040 Spain

[2]ESI International Chair@CEU-UCH, Universidad Cardenal Herrera-CEU, CEU Universities, San Bartolomé 55, 46115, Alfara del Patriarca, Valencia, Spain

Abstract. This article introduces an innovative mathematical framework designed to tackle non-linear convex variational problems in reflexive Banach spaces. Our approach employs a versatile technique that can handle a broad range of variational problems, including standard ones. To carry out the process effectively, we utilize specialized sets known as radial dictionaries, where these dictionaries encompass diverse data types, such as tensors in Tucker format with bounded rank and Neural Networks with fixed architecture and bounded parameters. The core of our method lies in employing a greedy algorithm through dictionary optimization defined by a multivalued map. Significantly, our analysis shows that the convergence rate achieved by our approach is comparable to the Method of Steepest Descend implemented in a reflexive Banach space, where the convergence rate follows the order of $O(m^{-1})$.

Keywords. Non-linear Convex Functional, Variational Problem, Greedy Algorithm, Radial Dictionary

1. Introduction

Related to the numerical methods of Partial Differential Equations (PDEs), two particular frameworks are attracted in recent years to the scientific computing community: the application of Tensor Numerical Methods [3, 4, 13, 15] and the Deep Neural Networks (DNNs) [22, 24, 25, 27, 28]. Concerning the use of these methods in engineering and industrial applications, among the family of tensor-based methods, the Proper Generalized Decomposition (PGD) has been used , among others, in surgery simulations [20], design optimization [2, 19], data-driven applications [16], elastodynamic [5], as well in structural damage identification [10].

Here, we propose a generalization of the use of the Proper Generealized Decomposition studied in [15]. Concretely, we would to propose a mathematical framework that includes, among others, the use of tensors formats and DNNs to solve the problem

$$\min_{x\in\mathbb{X}} \mathscr{E}(x), \tag{1.1}$$

where $\mathscr{E}:\mathbb{X}\longrightarrow\mathbb{R}$ is an elliptic and differentiable functional defined over a reflexive Banach space $\mathbb{X}$.

A model problem of the above framework is the following one (see for example [11] for more details): let Ω be a bounded domain in $\mathbb{R}^d$ $(d\geq 2)$ with Lipschitz continuous

*Corresponding author.

E-mail addresses: pablo.berna@cunef.edu (P. M. Berná).

Received -; Accepted -

boundary. For some fixed $p>2$, we let $H_0^{1,p}(\Omega)$, which is the closure of $C_c^\infty(\Omega)$ (functions in $C^\infty(\Omega)$ with compact support in Ω) with respect the norm

$$\|f\| = \left(\sum_{\ell=1}^{d} \|\partial_{x_\ell} f\|_{L^p(\Omega)}^p\right)^{1/p}.$$

We then introduce the functional $\mathscr{E}: H_0^{1,p}(\Omega) :\to \mathbb{R}$ defined by

$$\mathscr{E}(u) = \frac{1}{p}\|u\|^p - \langle \varphi, u\rangle,$$

with $\varphi \in H_0^{1,p}(\Omega)^*$. Its Fréchet differential is $\mathscr{E}'(u) = -\Delta_p - \varphi$ where

$$\Delta_p(u) = \sum_{\ell=1}^{d} \frac{\partial}{\partial x_\ell}\left(\left|\frac{\partial u}{\partial x_\ell}\right|^{p-2}\frac{\partial u}{\partial x_\ell}\right),$$

and Δ_p is known as the p-Laplacian.

In this paper, to obtain the solution u^* of (1.1), we will use the following "greedy algorithm":

(1) $u_0 = 0$;
(2) for $m\geq 1$, $u_m \in u_{m-1} + \nabla_{u_{m-1}}(\mathscr{E};\mathscr{D})$,

where

$$\nabla_u(\mathscr{E};\mathscr{D}) = \arg\min_{z\in u+\mathscr{D}} \mathscr{E}(z),$$

and $\mathscr{D}\subset \mathbb{X}$ is a universal dictionary (see Definition 3.4). The objectives that we have in the paper are the next ones.

(S1) We need to fix conditions on $\mathscr{E}$ in order to guarantee the existence of $u^*\in\mathbb{X}$ satisfying

$$\inf_{u\in\mathbb{X}} \mathscr{E}(u) = \min_{u\in\mathbb{X}} \mathscr{E}(u) = \mathscr{E}(u^*).$$

To this end, we choose the classical framework developed by Akilov and Kantorovich in [1, Chapter XV] (where, for some applied problems, they proved the convergence of the Method of Steepest Descend in a reflexive Banach space).

(S2) We introduce and characterize what is a radial (and universal) dictionary $\mathscr{D}$ giving some examples extracted from the literature.

(S3) Finally, under the above conditions, we prove that the sequence $\{u_m\}_{m\in\mathbb{N}}$ generated by the greedy algorithm over $\mathscr{D}$ "minimizes" $\mathscr{E}$ in the sense that

$$\lim_{m\to\infty} \mathscr{E}(u_m) = \mathscr{E}(u^*),$$

and $\lim_{m\to\infty} u_m = u^*$, and also it satisfies

$$\mathscr{E}(u_m) - \mathscr{E}(u^*) = O(m^{-1}),$$

which is the same rate of convergence obtained by Akilov and Kantorovich in [1] for the Method of Steepest Descend.

The organization of the paper is the following: in the next section we introduce some preliminary definitions and results used along the text related with (S1). In Section 3, related to the objective (S2), we introduce and characterize radial (universal) dictionaries where, moreover, we construct some examples by using well-known results from the literature. Section 4 is devoted to explain the greedy algorithm by (universal) dictionary optimization and to prove its convergence and that its rate of convergence is the same that the Method of Steepest Descend implemented in a reflexive Banach space, that ends (S3). Some conclusions and final remarks will be given in Section 5 and, finally, we prove some techincal results in the Appendix A.

2. Preliminary definitions and results

Let $\mathbb{X}$ be a reflexive Banach space endowed with a norm $\|\cdot\|$. We denote by $\mathbb{X}^*$ be the dual space of $\mathbb{X}$, endowed as usual with the dual norm $\|\cdot\|_*$, and $\langle\cdot,\cdot\rangle : \mathbb{X}^* \times \mathbb{X} \to \mathbb{R}$ is the duality pairing. Along this paper, given $A \subset \mathbb{X}$ and $r > 0$, we will denote by

$$B_{A,r} = \{x \in A : \|x\| \leq r\},$$

and the unit sphere in $\mathbb{X}$ as

$$S_{\mathbb{X}} = \{x \in \mathbb{X} : \|x\| = 1\},$$

which is closed and bounded in $\mathbb{X}$.

Here, we are going to attack the following optimization problem

$$\min_{x \in \mathbb{X}} \mathscr{E}(x), \tag{2.1}$$

where $\mathscr{E} : \mathbb{X} \to \mathbb{R}$ is a Fréchet differentiable functional satisfying the following properties:

(A) The Fréchet derivative $\mathscr{E}' : \mathbb{X} \longrightarrow \mathbb{X}^*$ is locally Lipschitz continuous and
(B) $\mathscr{E}$ is $\mathbb{X}$-elliptic, that is, there are two real numbers $\alpha > 0$ and $1 < s \leq 2$ such that for every $x, y \in \mathbb{X}$,

$$\langle \mathscr{E}'(x) - \mathscr{E}'(y), x - y \rangle \geq \alpha \|x - y\|^s.$$

Remark 2.1. Since any Lipschitz continuous function is uniformly continuous, (A) implies that the Fréchet derivative $\mathscr{E}'$ is also locally uniformly continuous.

Conditions (A) and (B) are used in [1, Chapter XV] to prove the convergence and also to compute the rate of convergence of the Method of Steepest Descend in reflexive Banach spaces. Moreover, in [1, Chapter XV §3] some applications to elliptic PDEs are also given.

Remark 2.2. Taking the model problem posed in the last section, following [8], assumption (A) is satisfied by the functional $\mathscr{E}$. Recall that if a functional $F : V \longrightarrow W$, where V and W are Banach spaces, is Fréchet differentiable at $v \in V$, then it is also locally Lipschitz continuous at $v \in V$. Since the map $G : H_0^{1,p}(\Omega) \subset L^p(\Omega) \longrightarrow \mathbb{R}$, given by $G(u) = \|u\|$, is $\mathscr{C}^2$ for $p \geq 2$ then $\mathscr{E}$ is also of class $\mathscr{C}^2$. Hence, $\mathscr{E}'$ is locally Lipschitz continuous in $H_0^{1,p}(\Omega)$ and assumption (B) holds.

Now, we give some results and definitions that will be useful along this paper.

Lemma 2.1. [8, Lemma 2.2] Under assumptions (A)-(B), we have

(a) For all $x,y \in \mathbb{X}$,
$$\mathscr{E}(x)-\mathscr{E}(y) \geq \langle \mathscr{E}'(y), x-y\rangle + \frac{\alpha}{s}\|x-y\|^s. \tag{2.2}$$
(b) $\mathscr{E}$ is strictly convex.
(c) $\mathscr{E}$ is bounded from below and coercive, i.e. $\lim_{\|x\|\to\infty}\mathscr{E}(x)=+\infty$.

Remark 2.3. From (2.2) we deduce
$$\mathscr{E}(x)-\mathscr{E}(y) \geq \langle \mathscr{E}'(y), x-y\rangle, \tag{2.3}$$
and hence
$$\mathscr{E}(y)-\mathscr{E}(x) \leq \langle \mathscr{E}'(y), y-x\rangle, \tag{2.4}$$
holds for all $x,y \in \mathbb{X}$.

The above lemma is the key of the following classical result.

Theorem 2.1. [9, Theorem 7.4.4] Under assumptions (A)-(B), the problem (2.1) admits a unique solution $x \in \mathbb{X}$ which is equivalently characterized by
$$\langle \mathscr{E}'(x), y\rangle = 0 \quad \forall y \in \mathbb{X} \tag{2.5}$$

As we have explanied in the last section, the idea is to construct the solution of (2.1) by a "greedy algorithm" using a set $\mathscr{A} \subset \mathbb{X}$ with some conditions and the milestone of this procedure is the following optimization program
$$\inf_{v\in\mathscr{A}} \mathscr{E}(v). \tag{2.6}$$
We mention [12] to study some results about the existence of a minimizer of these type of problems and some classical definitions and results to justify the existence of solutions of (2.6) are the following ones.

Definition 2.1. We recall that a sequence $x_m \in \mathbb{X}$ is weakly convergent if $\lim_{m\to\infty}\langle \varphi, x_m\rangle$ exists for all $\varphi \in \mathbb{X}^*$. We say that $(x_m)_{m\in\mathbb{N}}$ converges weakly to $x \in \mathbb{X}$ if $\lim_{m\to\infty}\langle \varphi, x_m\rangle = \langle \varphi, x\rangle$ for all $\varphi \in \mathbb{X}^*$. In this case, we write $x_m \rightharpoonup x$.

Definition 2.2. A subset $\mathscr{A} \subset \mathbb{X}$ is called weakly closed if $x_m \in \mathscr{A}$ and $x_m \rightharpoonup x$ implies $x \in \mathscr{A}$.

Remark 2.4. Of course, the condition of weakly closed is stronger than the condition to be close, so if $\mathscr{A} \subset \mathbb{X}$ is weakly closed, then it is closed.

Now, we focus our attention in some known properties about the functional $\mathscr{E}$.

Definition 2.3. We say that a map $\mathscr{E}:\mathbb{X} \longrightarrow \mathbb{R}$ is weakly sequentially lower semi-continuous (respectively, weakly sequentially continuous) in $\mathscr{A} \subset \mathbb{X}$ if for all $x \in \mathscr{A}$ and for all $x_m \in \mathscr{A}$ such that $x_m \rightharpoonup v$, it holds $\mathscr{E}(x) \leq \liminf_{m\to\infty}\mathscr{E}(x_m)$ (respectively, $\mathscr{E}(v)=\lim_{m\to\infty}\mathscr{E}(x_m)$).

Proposition 2.1. [29, Proposition 41.8 (H1)] Let $\mathbb{X}$ be a reflexive Banach space and let $\mathscr{E}:\mathbb{X}\to\mathbb{R}$ be a functional. If $\mathscr{E}$ is a convex and lower semi-continuous functional, then $\mathscr{E}$ is weakly sequentially lower semi-continuous.

One interesting consequence of Lemma 2.1(b) is the following corollary.

Corollary 2.1. Let $\mathbb{X}$ be a reflexive Banach space and let $\mathscr{E}:\mathbb{X}\to\mathbb{R}$ be a functional satisfying (A) and (B). Then $\mathscr{E}$ is weakly sequentially lower semi-continuous and coercive.

The next theorem says us that in order to have a solution of (2.6), the set $\mathscr{A}\subset\mathbb{X}$ should be weakly closed.

Theorem 2.2. [15, Theorem 2] Let $\mathbb{X}$ be a reflexive Banach space, and $\mathscr{A}\subset\mathbb{X}$ a weakly closed set. If $\mathscr{E}:\mathscr{A}\to\mathbb{R}\cup\{\infty\}$ is weakly sequentially lower semi-continuous and coercive on $\mathscr{A}$, then problem (2.6) has a solution.

3. Radial dictionaries in reflexive Banach spaces

The aim of this section is to introduce the class of sets that we use in (2.6) to construct the solution of (2.1). We will call these sets radial (universal) dictionaries.

Recall that a dictionary $\mathscr{D}\subset\mathbb{X}$, where $\mathbb{X}$ is a particular Banach space, is usually defined as a countable family of unit vectors and it is complete when the closure of $\operatorname{span}\mathscr{D}$ is equal to $\mathbb{X}$. In some sense, a complete dictionary is a redundant basis of a vector space.

In this paper, we introduce a different definition, where essentially the countable family of unit vectors is substituted by a closed set of norm one elements. Our definition takes into account two properties. The first one is geometric: a radial dictionary should be a cone (it contains all the straight lines generated by its owns elements). The second property is topological: the whole cone should be generated by some non-empty closed and bounded set. As we will see below, any radial radial dictionary is a weakly closed cone, and as consequence, the aforementioned set can be identified with a cone generated by a non-empty closed set of norm one elements. Before to introduce the concept of radial dictionary we need the following definitions.

Definition 3.1. A set $C\subset\mathbb{X}$ is an extended cone if for all $x\in C$ the one dimensional subspace $\operatorname{span}\{x\}\subset C$, that is, $\lambda x\in C$ for all $\lambda\in\mathbb{R}$.

Definition 3.2. Given a non-empty set $A\subset\mathbb{X}$ we define the extended cone generated by A, denoted by $\langle A\rangle$, as

$$\langle A\rangle=\{\lambda a:\lambda\in\mathbb{R}\text{ and }a\in A\}.$$

Observe that for any non-empty set $A\subset\mathbb{X}$ it holds $\langle\langle A\rangle\rangle=\langle A\rangle$. Moreover, in order to construct the closed subspace containing the set A we can follow the next scheme given by the inclusions

$$A\subset\langle A\rangle\subset\operatorname{span}\langle A\rangle\subset\overline{\operatorname{span}\langle A\rangle}^{\|\cdot\|}.$$

In consequence, the cone generating by the set A is the first geometric object that appears towards the construction of the closed linear subspace $\overline{\operatorname{span}\langle A\rangle}^{\|\cdot\|}$.

Definition 3.3. A subset $\mathscr{D}$ is called a radial dictionary of $\mathbb{X}$ if it is a cone generated by some non-empty closed and bounded set in $\mathbb{X}$, that is, there exists a closed and bounded set $\mathscr{K}\subset\mathbb{X}$ such that $\mathscr{D}=\langle\mathscr{K}\rangle$.

Since in a Banach space any compact set is closed and bounded we have the following consequence.

Lemma 3.1. Let $\mathbb{X}$ be a reflexive Banach space and $\mathscr{K} \subset \mathbb{X}$ be a compact set. Then the set $\langle \mathscr{K} \rangle$ is a radial dictionary.

The next result characterizes a radial dictionary as a weakly closed cone.

Theorem 3.1. Let $\mathbb{X}$ be a reflexive Banach space and $\mathscr{D} \subset \mathbb{X}$. Then the following statements are equivalent.

(a) $\mathscr{D}$ is a radial dictionary of $\mathbb{X}$.
(b) $\mathscr{D}$ is a weakly closed cone in $\mathbb{X}$.

Proof. Assume $\mathscr{D} = \langle \mathscr{K} \rangle$ is a radial dictionary for some $\mathscr{K} \subset \mathbb{X}$ closed and bounded. We only need to prove to prove that $\mathscr{D}$ is weakly closed. To this end, take a sequence $(f_n)_{n\in\mathbb{N}} \subset \mathscr{D}$ be such that $f_n \rightharpoonup f$. Then $f_n = \lambda_n h_n$, for $\lambda_n \in \mathbb{R}$ and $h_n \in \mathscr{K}$, is bounded in $\mathbb{X}$ and hence $(\lambda_n)_{n\in\mathbb{N}}$ is bounded in $\mathbb{R}$ because $\mathscr{K}$ is. In consequence, the sequence $(\lambda_n, h_n)_{n\in\mathbb{N}} \subset \mathbb{R}\times\mathbb{X}$ is also bounded, and there exists a subsequence, also denoted by $(\lambda_n, h_n)_{n\in\mathbb{N}}$, such that $\lim_{n\to\infty}(\lambda_n, h_n) = (\lambda, h)$. Moreover, since $\mathbb{R}\times\mathscr{K}$ is closed in $\mathbb{R}\times\mathbb{X}$, then $(\lambda, h) \in \mathbb{R}\times\mathscr{K}$.

Now, the bilinear map $\Phi : \mathbb{R}\times\mathbb{X} \longrightarrow \mathbb{X}$ given by $\Phi(\lambda, x) = \lambda\, x$ is clearly continuous. Then

$$\lim_{n\to\infty} f_n = \lim_{n\to\infty} \Phi(\lambda_n, h_n) = \Phi(\lim_{n\to\infty}(\lambda_n, h_n)) = \Phi(\lambda, h) = \lambda\, h,$$

and hence $f = \lambda\, h$, because $f_n \rightharpoonup \lambda\, h$. This proves (b)

Assume that $\mathscr{D} \subset \mathbb{X}$ is a weakly closed cone and hence $\mathscr{D}$ is also closed. Consider the closed and bounded set $S_{\mathbb{X}} \subset \mathbb{X}$. Then $\mathscr{K} = \mathscr{D} \cap S_{\mathbb{X}}$ is closed and bounded in $\mathbb{X}$. Clearly, $\langle \mathscr{K} \rangle \subset \mathscr{D}$. To conclude, take $x \in \mathscr{D}$, if $x = 0$ then $x \in \langle \mathscr{K} \rangle$, otherwise $\mathrm{span}\{x\} \subset \mathscr{D}$, and then $x/\|x\| \in \mathscr{K}$. Hence $x \in \langle \mathscr{K} \rangle$ and $\langle \mathscr{K} \rangle = \mathscr{D}$ and the proof is done. □

From the proof of Theorem 3.1 we deduce that every radial dictionary of $\mathbb{X}$ can be written as $\mathscr{D} = \langle \mathscr{D} \cap S_{\mathbb{X}} \rangle$, where $\mathscr{D} \cap S_{\mathbb{X}}$ is a closed set of norm one elements.

Definition 3.4. A subset $\mathscr{D}$ is called universal dictionary of $\mathbb{X}$, if $\mathscr{D}$ is a radial dictionary of $\mathbb{X}$ with the following extra condition: the linear subspace span $\mathscr{D}$ is dense in $\mathbb{X}$.

Of course, every universal dictionary is indeed a radial dictionary, but the converse could be false (see Section 3.1 below). However, it is easy to see that given a radial dictionary $\mathscr{D} \subset \mathbb{X}$, taking into account the Banach space $\overline{\mathrm{span}\,\mathscr{D}}^{\|\cdot\|_{\mathbb{X}}}$, generates an universal dictionary as the following corollary shows.

Corollary 3.1. Let $(\mathbb{Y}, \|\cdot\|_{\mathbb{Y}})$ be a reflexive Banach space and $\mathscr{K} \subset \mathbb{Y}$ be a closed and bounded set. Then the set $\langle \mathscr{K} \rangle$ is an universal dictionary of $\mathbb{X} = \overline{\mathrm{span}\,\langle \mathscr{K} \rangle}^{\|\cdot\|_{\mathbb{Y}}} \subset \mathbb{Y}$.

Some examples of universal dictionaries are the following.

3.1. Basis based radial dictionaries. Let us consider $(\mathbb{Y}, \|\cdot\|)$ be a reflexive Banach space and $\mathscr{K} = \{x_1, \ldots, x_n\} \subset \mathbb{Y}$ be a set of n-linearly independent vectors in $\mathbb{Y}$. Clearly, it is a closed and bounded set. Thus

$$\langle \mathscr{K} \rangle = \{\lambda\, x_i : \lambda \in \mathbb{R} \text{ and } 1 \le i \le n\}$$

is an universal dictionary of $\mathbb{X}_n = \mathrm{span}\,\langle \mathscr{K} \rangle$.

Assume now that $(\mathbb{Y},\|\cdot\|)$ is a separable Hilbert space. Consider $\mathscr{B}=\{(g_i)_{i\in\mathbb{N}}:\|g_i\|=1\}$ be an orthogonal basis of $\mathbb{Y}$. Then $\mathscr{K}=\overline{\mathscr{B}}^{\|\cdot\|}$ is closed and bounded in $\mathbb{Y}$ and hence $\langle\mathscr{K}\rangle$ is an universal dictionary of $\mathbb{Y}$.

3.2. A natural radial dictionary for a non-linear convex problem. Let $\mathbb{X}$ be a reflexive Banach space. Assume that $\mathscr{E}:\mathbb{X}\longrightarrow\mathbb{R}$ satisfies (A)-(B) and $u^*\in\mathbb{X}$ solves (2.1). Consider the closed set

$$\Omega_0=\{u\in\mathbb{X}:\mathscr{E}(u^*)\leq\mathscr{E}(u)\leq\mathscr{E}(0)\}=\mathscr{E}^{-1}([\mathscr{E}(u^*),\mathscr{E}(0)]),$$

where we can use Proposition 4.1 to define the set above taking the 0. From Lemma 2.1(c), Ω_0 is also bounded. Then $\langle\Omega_0\rangle$ is an universal dictionary of $\mathbb{X}_{\mathscr{E}}=\overline{\mathsf{span}\,\langle\Omega_0\rangle}^{\|\cdot\|}$. Moreover,

$$\mathscr{E}(u^*)=\min_{x\in\mathbb{X}}\mathscr{E}(x)=\min_{x\in\mathbb{X}_{\mathscr{E}}}\mathscr{E}(x).$$

3.3. A Radial dictionary for an Approximation Laplacian Fractional Algorithm. In [6] the authors consider the weakly closed cone

$$\mathscr{D}=\{u\in L^2([0,1]):u(x)=\lambda\,x^{\beta}\text{ for }\lambda\in\mathbb{R}\text{ and }\beta\in[0,2]\}$$

in the Hilbert space $L^2[(0,1])$ to explain the convergence of a Approximation Laplacian Fractional Algorithm (ALFA). This algorithm was previously introduced in the framework of a fractional deconvolution model used in image restoration [7]. Observe that

$$\|u\|_{L^2}=\frac{|\lambda|}{\sqrt{1+2\beta}}\text{ holds for all }u=\lambda\,x^{\beta}\in\mathscr{D}.$$

Thus, $\mathscr{D}=\langle\mathscr{K}\rangle$, where

$$\mathscr{K}=\{u\in L^2([0,1]):u(x)=x^{\beta}\text{ for }\beta\in[0,2]\},$$

is a closed and bounded set in $L^2[0,1]$. We conclude that $\mathscr{D}$ is an universal dictionary for the Hilbert space $\overline{\mathsf{span}\,\mathscr{D}}^{\|\cdot\|_{L^2}}$.

3.4. Tensor based radial dictionaries. Let $D:=\{1,2,\ldots,d\}$ be a finite index set and $(V_\alpha,\|\cdot\|_\alpha)$ be a Banach space for each index $\alpha\in D$. we refer to Greub [17] for the definition of the algebraic tensor space ${}_a\bigotimes_{\alpha\in D}V_\alpha$ generated from vector spaces V_α $(\alpha\in D)$. As underlying field we choose $\mathbb{R}$, the suffix 'a' in ${}_a\bigotimes_{\alpha\in D}V_\alpha$ refers to the 'algebraic' nature of the tensor space. By definition, all elements of

$$\mathbf{V}_D:={}_a\bigotimes_{\alpha\in D}V_\alpha$$

are finite linear combinations of elementary tensors $\mathbf{v}=\bigotimes_{\alpha\in D}v_\alpha$ $(v_\alpha\in V_\alpha)$.

Next, we introduce the injective norm $\|\cdot\|_\vee$ for $\mathbf{v}\in\mathbf{V}={}_a\bigotimes_{\alpha\in D}V_\alpha$, by

$$\|\mathbf{v}\|_\vee:=\sup\left\{\frac{\left|\left(\varphi^{(1)}\otimes\varphi^{(2)}\otimes\ldots\otimes\varphi^{(d)}\right)(\mathbf{v})\right|}{\prod_{\alpha\in D}\|\varphi^{(\alpha)}\|_\alpha^*}:0\neq\varphi^{(\alpha)}\in V_\alpha^*,\,\alpha\in D\right\}. \tag{3.1}$$

Assume that $\|\cdot\|$ is another norm in $\mathbf{V}_D$ not weaker than the injective norm, that is, there exists a constant C be such that $\|\mathbf{v}\|_\vee\leq C\|\mathbf{v}\|$ holds for all $\mathbf{v}\in\mathbf{V}_D$.

Let us consider $\mathbb{Y} = \overline{\mathbf{V}_D}^{\|\cdot\|}$ a tensor Banach space given by the completion of $\mathbf{V}_D$ under $\|\cdot\|$. Given $\mathbf{r} = (r_\alpha)_{\alpha\in D} \in \mathbb{N}^{\#D}$ (where $r_\alpha \geq 1$ for all $\alpha \in D$), we introduce the set of tensors in Tucker format with bounded rank $\mathbf{r}$ as

$$\mathscr{M}_{\leq \mathbf{r}}(\mathbf{V}_D) = \left\{ \mathbf{v} \in \mathbf{V}_D : \begin{array}{l} \mathbf{v} \in {}_a\bigotimes_{\alpha\in D} U_\alpha \text{ where } U_\alpha \text{ is a subspace} \\ \text{in } V_\alpha \text{ with } \dim U_\alpha \leq r_\alpha \text{ for each } \alpha \in D \end{array} \right\}.$$

In [14] (Proposition 4.3) the following result has been proved.

Proposition 3.1. Let $(\overline{\mathbf{V}_D}^{\|\cdot\|}, \|\cdot\|)$ be a Banach tensor space with a norm not weaker than the injective norm. Then the set $\mathscr{M}_{\leq \mathbf{r}}(\mathbf{V}_D)$ is weakly closed.

The set $\mathscr{M}_{\leq \mathbf{r}}(\mathbf{V}_D)$ is clearly a cone and, by the above proposition, it is a weakly closed cone. Theorem 3.1 implies that the set $\mathscr{M}_{\leq \mathbf{r}}(\mathbf{V}_D)$ is a radial dictionary. The linear subspace $\operatorname{span} \mathscr{M}_{\leq \mathbf{r}}(\mathbf{V}_D)$ dense in $\mathbb{Y}$, because it contains the whole set of elementary tensors. Thus, $\mathscr{M}_{\leq \mathbf{r}}(\mathbf{V}_D)$ is an universal dictionary in $\mathbb{Y}$.

3.5. Neural networks based radial dictionaries. To introduce the radial dictionary composed of Neural Networks it necessary to distinguish between a neural network as a set of weights and the associated function implemented by the network, which we call its realization. To explain this distinction, let us fix numbers $L, N_0, N_1, \ldots, N_L \in \mathbb{N}$. We say that a family $\mathbf{\Phi} = \big((A_\ell, b_\ell)\big)_{\ell=1}^L$ of matrix-vector tuples of the form $A_\ell \in \mathbb{R}^{N_\ell \times N_{\ell-1}}$ and $b_\ell \in \mathbb{R}^{N_\ell}$ is a neural network.

We call $S := (N_0, N_1, \ldots, N_L)$ the architecture of $\mathbf{\Phi}$; furthermore $N(S) := \sum_{\ell=0}^L N_\ell$ is called the number of neurons of S and $L = L(S)$ is the number of layers of S and $d := N_0$ the input dimension of $\mathbf{\Phi}$. For a given architecture S, we denote by $\mathscr{N}\mathscr{N}(S)$ the set of neural networks with architecture S.

Defining the realization of such a network $\mathbf{\Phi} = \big((A_\ell, b_\ell)\big)_{\ell=1}^L$ requires two additional ingredients: a so-called activation function $\rho : \mathbb{R} \to \mathbb{R}$, and a domain of definition $\Omega \subset \mathbb{R}^{N_0}$. Given these, the realization of the network $\mathbf{\Phi} = \big((A_\ell, b_\ell)\big)_{\ell=1}^L$ is the function

$$\mathrm{R}_\rho^\Omega(\mathbf{\Phi}) : \Omega \to \mathbb{R}, \quad x \mapsto x_L,$$

where x_L results from the following scheme:

$$\begin{aligned} x_0 &:= x, \\ x_\ell &:= \rho(A_\ell x_{\ell-1} + b_\ell), \qquad \text{for } \ell = 1, \ldots, L-1, \\ x_L &:= A_L x_{L-1} + b_L, \end{aligned}$$

and where ρ acts component-wise; that is, $\rho(x_1, \ldots, x_d) := (\rho(x_1), \ldots, \rho(x_d))$.

Before to give the next result, let us note that the set $\mathscr{N}\mathscr{N}(S)$ of all neural networks (that is, the network weights) with a fixed architecture forms a finite-dimensional vector space, which we equip with the norm

$$\|\mathbf{\Phi}\|_{\mathscr{N}\mathscr{N}(S)} := \|\mathbf{\Phi}\|_{\text{scaling}} + \max_{\ell=1,\ldots,L} \|b_\ell\|_{\max}$$

$$\text{for} \qquad \mathbf{\Phi} = \big((A_\ell, b_\ell)\big)_{\ell=1}^L \in \mathscr{N}\mathscr{N}(S),$$

where $\|\Phi\|_{\text{scaling}} := \max_{\ell=1,\dots,L} \|A_\ell\|_{\max}$. If the specific architecture of Φ does not matter, we simply write $\|\Phi\|_{\text{total}} := \|\Phi\|_{\mathcal{N}\mathcal{N}(S)}$. In addition, if ρ is continuous, we denote the realization map by

$$\mathrm{R}_\rho^\Omega : \mathcal{N}\mathcal{N}(S) \to C(\Omega;\mathbb{R}^{N_L}),\ \Phi \mapsto \mathrm{R}_\rho^\Omega(\Phi). \tag{3.2}$$

In [23] the authors proves the following result (Proposition 3.5).

Proposition 3.2. Let $S = (d, N_1, \dots, N_L)$ be a neural network architecture, let $\Omega \subset \mathbb{R}^d$ be compact, let furthermore $p \in (0,\infty)$, and let $\rho : \mathbb{R} \to \mathbb{R}$ be continuous. For $C > 0$, let

$$\Theta_C := \left\{\Phi \in \mathcal{N}\mathcal{N}(S) : \|\Phi\|_{\text{total}} \leq C\right\}.$$

Then the set $\mathrm{R}_\rho^\Omega(\Theta_C)$ is compact in $C(\Omega)$ as well as in $L_\mu^p(\Omega)$, for any finite Borel measure μ on Ω and any $p \in (0,\infty)$.

Under the assumptions of Proposition 3.2 let us consider the compact set $\mathrm{R}_\rho^\Omega(\Theta_C) \subset L^p(\mu)$ for $p \in (1,\infty)$. Then, from Lemma 3.1, the set $\langle \mathrm{R}_\rho^\Omega(\Theta_C)\rangle$ is a radial dictionary in $L^p(\mu)$ for $p \in (1,\infty)$. Recall that

$$\langle \mathrm{R}_\rho^\Omega(\Theta_C)\rangle = \{\lambda \mathrm{R}_\rho^\Omega(\Phi) : \lambda \in \mathbb{R} \text{ and } \Phi \in \Theta_C\}.$$

Since $\lambda \mathrm{R}_\rho^\Omega(\Phi)(x_0) = \lambda x_L = \lambda A_L x_{L-1} + \lambda b_L$, we obtain that $\lambda \mathrm{R}_\rho^\Omega(\Phi) = \mathrm{R}_\rho^\Omega(\Phi^*)$ with $\Phi^* = \left((A_\ell^*, b_\ell^*)\right)_{\ell=1}^L$ where $A_\ell^* = A_\ell$, $b_\ell^* = b_\ell$ for $1 \leq \ell \leq L-1$ and $A_L^* = \lambda A_L$, $b_L^* = \lambda b_L$. In consequence, the radial dictionary $\langle \mathrm{R}_\rho^\Omega(\Theta_C)\rangle \subset \mathrm{R}_\rho^\Omega(\mathcal{N}\mathcal{N}(S))$ is a cone included in the set of realizations of neural networks with architecture S. Thus, $\langle \mathrm{R}_\rho^\Omega(\Theta_C)\rangle$ is a universal dictionary of the reflexive Banach space $\overline{\operatorname{span}\langle \mathrm{R}_\rho^\Omega(\Theta_C)\rangle}^{\|\cdot\|_{L_\mu^p}} \subset L_\mu^p(\Omega)$ for $p \in (1,\infty)$.

3.6. $\mathscr{E}$-dictionary optimization over $\mathscr{D}$. Now, to complete (S2) we need to introduce the following definition.

Definition 3.5 (Dictionary optimization). Let $\mathbb{X}$ be a reflexive Banach space and $\mathscr{D}$ be radial dictionary of $\mathbb{X}$. Assume that $\mathscr{E} : \mathbb{X} \longrightarrow \mathbb{R}$ satisfies (A)-(B). A $\mathscr{E}$-dictionary optimization over $\mathscr{D}$ is a multivalued map defined as follows:

$$\nabla_\cdot(\mathscr{E};\mathscr{D}) : \mathbb{X} \rightrightarrows \mathscr{D}, \quad u \mapsto \nabla_u(\mathscr{E};\mathscr{D}) := \arg \min_{z \in u+\mathscr{D}} \mathscr{E}(z).$$

Remark 3.1. Observe that for radial dictionaries in reflexive Banach spaces,

$$\nabla_u(\mathscr{E};\mathscr{D}) := \arg\min_{v\in\mathscr{D}} \mathscr{E}(u+v).$$

The next result gives us some interesting properties of this multivalued map.

Theorem 3.2. Let $\mathbb{X}$ be a reflexive Banach space and $\mathscr{D}$ be radial dictionary of $\mathbb{X}$. Assume that $\mathscr{E} : \mathbb{X} \longrightarrow \mathbb{R}$ satisfies (A)-(B). Then, for all $u \in \mathbb{X}$, the following statements hold.

(a) The set $\nabla_u(\mathscr{E};\mathscr{D}) \neq \emptyset$ and it is weakly closed in $\mathbb{X}$.
(b) If $\mathscr{D}$ is an universal dictionary in $\mathbb{X}$ and $u^* \in \mathbb{X}$ satisfies $0 \in \nabla_{u^*}(\mathscr{E};\mathscr{D})$, then u^* solves (2.1).

Proof. (a) We claim that $u + \mathscr{D}$ is also weakly closed. To see this assume that $u + y_n \rightharpoonup y$ for some $\{y_n\}_{n\geq 1} \subset \mathscr{D}$, then $y_n \rightharpoonup y - u$ and since $\mathscr{D}$ is weakly closed, $y - u \in \mathscr{D}$. In consequence $y \in u + \mathscr{D}$ and the claim follows.

Next, we will show that $\nabla_u(\mathscr{E};\mathscr{D}) \neq \emptyset$ for all $u \in \mathbb{X}$. Applying the Fréchet differentiability of $\mathscr{E}$, we know that $\mathscr{E}$ is continuous. Since $\mathscr{E}$ is also convex, $\mathscr{E}$ is weakly sequentially lower semi-continuous by Proposition 2.1. Moreover, $\mathscr{E}$ is coercive on $\mathbb{X}$ by Lemma 2.1(c). By the claim $u+\mathscr{D}$ is a weakly closed subset in $\mathbb{X}$. Then, the existence of a minimizer follows from Theorem 2.2, that is, $\nabla_u(\mathscr{E};\mathscr{D}) \neq \emptyset$.

Now, we show that $\nabla_u(\mathscr{E};\mathscr{D})$ is weakly closed in $\mathbb{X}$. For that, consider the map

$$\mathscr{E}_u : \mathbb{X} \longrightarrow \mathbb{R}$$

given by $\mathscr{E}_u(x) = \mathscr{E}(x+u)$. Let $a = \min_{\xi \in \mathscr{D}} \mathscr{E}_u(\xi) \in \mathbb{R}$. Then, $\mathscr{E}_u^{-1}(\{a\})$ is a closed set in $\mathbb{X}$, thus $\nabla_u(\mathscr{D};\mathscr{E}) = \mathscr{E}_u^{-1}(\{a\}) \cap \mathscr{D}$ is also closed in $\mathbb{X}$. Consider a sequence $\{u_n\} \subset \nabla_u(\mathscr{D};\mathscr{E})$ such that $u_n \rightharpoonup u$. Then $\{u_n\}$ is a bounded sequence in $\mathbb{X}$ that is a reflexive Banach space. Hence, there exists a subsequence, also denoted by $\{u_n\}$, convergent to $z \in \nabla_u(\mathscr{D};\mathscr{E})$. By the unicity of limits, $z = u$. Thus, the set $\nabla_u(\mathscr{E};\mathscr{D})$ is weakly closed in $\mathbb{X}$. This ends the proof of (a).

(b) Finally, to prove the last part of the theorem, assume that $u^* \in \mathbb{X}$ satisfies $0 \in \nabla_{u^*}(\mathscr{E};\mathscr{D})$. Then for all $\gamma \in \mathbb{R}_+$ and $z \in \mathscr{D}$, it holds $\mathscr{E}(u^*+\gamma z) \geq \mathscr{E}(u^*)$ and therefore

$$\langle \mathscr{E}'(u^*), z \rangle = \lim_{\gamma \searrow 0} \frac{1}{\gamma}(\mathscr{E}(u^*+\gamma z) - \mathscr{E}(u^*)) \geq 0$$

holds for all $z \in \mathscr{D}$. Since $\mathscr{D}$ is a cone, we have $\langle \mathscr{E}'(u^*), z \rangle = 0$ for all $z \in \mathscr{D}$. Finally, by using the universality of $\mathscr{D}$, we obtain $\langle \mathscr{E}'(u^*), v \rangle = 0$, for all $v \in \mathbb{X}$, and (b) follows from Theorem 2.1. □

4. Greedy Sequences by dictionary optimization

Let $\mathbb{X}$ be a reflexive Banach space and $\mathscr{D}$ be a universal dictionary of $\mathbb{X}$. Assume that $\mathscr{E} : \mathbb{X} \longrightarrow \mathbb{R}$ satisfies (A)-(B) and $u^* \in \mathbb{X}$ solves (2.1). Given $u \in \mathbb{X}$, from the above theorem, if $0 \in \nabla_u(\mathscr{E};\mathscr{D})$ then $u = u^*$, roughly speaking $\mathscr{E}'(u^*) = 0$. Otherwise, $\mathscr{E}(v) < \mathscr{E}(u)$ holds for all $v \in u + \nabla_u(\mathscr{E};\mathscr{D})$. Intuitively, the set $\nabla_u(\mathscr{E};\mathscr{D})$ contains feasible optimal corrections to reduce the value of $\mathscr{E}(u+\cdot)$ over the dictionary $\mathscr{D}$. This discussion allows us to propose an iterative multivalued procedure that motivates the following definition.

Definition 4.1 (Greedy Sequence by Dictionary Optimization). We say that a greedy sequence of u^* by a $\mathscr{E}$-dictionary optimization over $\mathscr{D}$ is any sequence $\{u_m\}_{m\geq 1} \subset \mathbb{X}$ constructed by the following greedy algorithm:

(1) $u_0 = 0$;
(2) for $m \geq 1$, $u_m \in u_{m-1} + \nabla_{u_{m-1}}(\mathscr{E};\mathscr{D})$.

Observe that in a practical implementation, $u_m = u_{m-1} + \widetilde{\varepsilon}_m$ for some $\widetilde{\varepsilon}_m \in \nabla_{u_{m-1}}(\mathscr{E};\mathscr{D})$, where

$$d_m := \mathscr{E}(u_{m-1}) - \mathscr{E}(u_{m-1} + \widetilde{\varepsilon}_m) \geq 0 \tag{4.1}$$

is an optimal constant value over $\mathscr{D}$. The quantity $\widetilde{\varepsilon}_m = u_m - u_{m-1} \in \nabla_{u_{m-1}}(\mathscr{E};\mathscr{D})$ can be see as a conditional residual that belongs to $\mathscr{D}$, and depends on u_{m-1}. By definition, the first residual $\widetilde{\varepsilon}_1 = u_1 \in \nabla_0(\mathscr{E};\mathscr{D})$ is a minimum of $\mathscr{E}$ over $\mathscr{D}$, where in absence of any a priori information we may assume that $u_0 = 0$. Moreover, for all $m \geq 1$, we have $u_m = u_{m-1} + \widetilde{\varepsilon}_m = \sum_{\ell=1}^{m} \widetilde{\varepsilon}_\ell$ where $\widetilde{\varepsilon}_\ell \in \nabla_{u_{\ell-1}}(\mathscr{E};\mathscr{D})$ for $\ell \geq 1$ (see Figure 1). In consequence, the sequence $\{u_m\}_{m\geq 1}$ can be interpreted as an algebraic stack that progressively stores conditional residues.

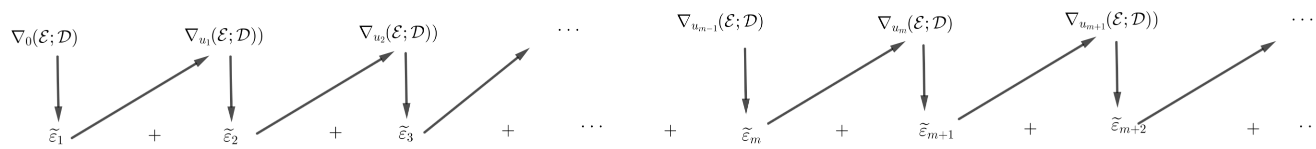

Figure 1. A scheme of the construction of a greedy sequence of u^* by a $\mathscr{E}$-dictionary optimization over $\mathscr{D}$.

A greedy sequence by a dictionary optimization can be see as an extension of the Method of Steepest Descend (see [1, Chapter XV]) because both methods share assumptions (A)-(B) to assure its convergence. To compare, recall that Steepest Descend is performed by means a sequence $\{u_m\}_{m\geq 1}$ constructed as follows:

(1) $u_0 = 0$, and for $m \geq 1$,
(2) given u_{m-1} compute:
  (2.1) $w_m \in \arg\min_{z\in S_{\mathbb{X}}} \langle \mathscr{E}'(u_{m-1}), z\rangle$;
  (2.2) $\lambda_m \in \arg\min_{\lambda\in\mathbb{R}} \mathscr{E}(u_{m-1} + \lambda\, w_m)$;
  put $u_m = u_{m-1} + \lambda_m w_m$.

Step (2.1) appears as the key point of this procedure. It implies $\langle \mathscr{E}(u_{m-1}), w_m\rangle = -\|\mathscr{E}'(u_{m-1})\|_*$. Moreover, a bound of $O(m^{-1})$ for the rate of convergence of the sequence $\{\mathscr{E}(u_m) - \mathscr{E}(u^*)\}_{m\geq 1}$, where u^* solves (2.1), has been obtained (see Theorem 3 p. 466 in [1]).

We are going to prove two results:

- The first one is to show that each greedy sequence $\{u_m\}_{m\geq 1}$ of u^* by a $\mathscr{E}$-dictionary optimization over $\mathscr{D}$ converges in $\mathbb{X}$ to u^*, where u^* is the solution of (2.1) as is stated in the next result.

Theorem 4.1. Let $\mathbb{X}$ be a reflexive Banach space and $\mathscr{D}$ be a universal dictionary of $\mathbb{X}$. Assume that $\mathscr{E}:\mathbb{X}\longrightarrow\mathbb{R}$ satisfies (A)-(B) and $u^*\in\mathbb{X}$ solves (2.1). Then for each greedy sequence $\{u_m\}_{m\geq 1}$ of u^* by a $\mathscr{E}$-dictionary optimization over $\mathscr{D}$ converges in $\mathbb{X}$ to u^*, that is, $\lim_{m\to\infty}\|u^* - u_m\| = 0$.

- The second one shows that a greedy sequence by dictionary optimization has the same rate of convergence as the well-known Method of the Steepest Descend implemented in a reflexive Banach space.

Theorem 4.2. Let $\mathbb{X}$ be a reflexive Banach space and $\mathscr{D}$ be a universal dictionary of $\mathbb{X}$. Assume that $\mathscr{E}:\mathbb{X}\longrightarrow\mathbb{R}$ satisfies (A)-(B) and $u^*\in\mathbb{X}$ solves (2.1). Then each greedy sequence $\{u_m\}_{m\geq 1}$ of u^* by a $\mathscr{E}$-dictionary optimization over $\mathscr{D}$ satisfies

$$\mathscr{E}(u_m) - \mathscr{E}(u^*) = O(m^{-1}).$$

To prove Theorem 4.1, the next proposition is needed. It gives us some useful properties of the sequences $\{u_m\}_{m\in\mathbb{N}}$, $\{\mathscr{E}(u_m)\}_{m\geq 1}$ and $\{\mathscr{E}'(u_m)\}_{m\geq 1}$.

Proposition 4.1. Let $\mathbb{X}$ be a reflexive Banach space and $\mathscr{D}$ be a universal dictionary of $\mathbb{X}$. Assume that $\mathscr{E}:\mathbb{X}\longrightarrow\mathbb{R}$ satisfies (A)-(B) and $u^*\in\mathbb{X}$ solves (2.1). Then for each greedy sequence $\{u_m\}_{m\geq 1}$ of u^* by a $\mathscr{E}$-dictionary optimization over $\mathscr{D}$ the following statements hold.

(a) The sequence $\{\mathscr{E}(u_m)\}_{m\geq 1}$, is non increasing and bounded below, that is,

$$\mathscr{E}(u^*) \leq \mathscr{E}(u_m) \leq \mathscr{E}(u_{m-1}),$$

holds for all $m \geq 1$.

(b) If $\lim_{m\to\infty} \mathscr{E}(u_m) = \mathscr{E}(u^*)$, then $\lim_{m\to\infty} \|u^* - u_m\| = 0$.

(c) If for some $m \geq 1$ it holds $u_m = u_{m-1}$ then $u_{m-1} = u^*$.

(d) The sequence $\{\mathscr{E}'(u_m)\}_{m\in\mathbb{N}}$ satisfies $\lim_{m\to\infty}\langle \mathscr{E}'(u_m), z\rangle = 0$, for all z in $\mathbb{X}$.

(e) There exists a subsequence $\{u_{m_k}\}_{k\in\mathbb{N}}$ such that $\lim_{k\to\infty}\langle \mathscr{E}'(u_{m_k}), u_{m_k}\rangle \to 0$.

Proof. See Appendix A. □

Now, we give the proof of the theorem.

Proof of Theorem 4.1. From Proposition 4.1(a), $\{\mathscr{E}(u_m)\}$ is a non increasing sequence. If there exists m such that $\mathscr{E}(u_m) = \mathscr{E}(u_{m-1})$. From Proposition 4.1(c), we have $u_m = u^*$, which ends the proof. Let us now suppose that $\mathscr{E}(u_m) < \mathscr{E}(u_{m-1})$ for all m. Now, $\{\mathscr{E}(u_m)\}_{m\in\mathbb{N}}$ is a strictly decreasing sequence of real numbers which is bounded below by $\mathscr{E}(u^*)$. Then, there exists

$$\mathscr{E}^* = \lim_{m\to\infty} \mathscr{E}(u_m) \geq \mathscr{E}(u^*).$$

If $\mathscr{E}^* = \mathscr{E}(u^*)$, Proposition 4.1(b) allows to conclude that $\{u_m\}$ strongly converges to u^*. Therefore, it remains to prove that $\mathscr{E}^* = \mathscr{E}(u^*)$. As a consequence of (2.4) we have

$$\mathscr{E}(u_m) - \mathscr{E}(u^*) \leq \langle \mathscr{E}'(u_m), u_m - u^*\rangle = \langle \mathscr{E}'(u_m), u_m\rangle - \langle \mathscr{E}'(u_m), u^*\rangle \tag{4.2}$$

By Proposition 4.1(e), we have that there exists a subsequence $\{u_{m_k}\}_{k\in\mathbb{N}}$ such that

$$\lim_{k\to\infty} \langle \mathscr{E}'(u_{m_k}), u_{m_k}\rangle = 0$$

and, from Proposition 4.1(d), also $\lim_{k\to\infty}\langle \mathscr{E}'(u_{m_k}), u^*\rangle = 0$. Therefore, putting $\{u_{m_k}\}_{k\in\mathbb{N}}$ in (4.2) and taking limits as $k \to \infty$, we obtain

$$\mathscr{E}^* - \mathscr{E}(u^*) = \lim_{k\to\infty} \mathscr{E}(u_{m_k}) - \mathscr{E}(u^*) \leq 0$$

Since we already had $\mathscr{E}^* \geq \mathscr{E}(u^*)$, this yields $\mathscr{E}^* = \mathscr{E}(u^*)$, which ends the proof. □

To prove now Theorem 4.2, we need the following three results. The first two lemmas are due to Akilov and Kantorovich [1].

Lemma 4.1 ([1, Lemma 2 p. 464]). Let $\mathbb{X}$ be a reflexive Banach space and assume that $\mathscr{E} : \mathbb{X} \longrightarrow \mathbb{R}$ satisfies (A). Assume that $\mathscr{E}'$ satisfies the Lipschitz condition in $B_{\mathbb{X},r+s}$ for some $r, s > 0$ with a Lipschitz constant equal to $\mathscr{L}$. If $z \in B_{\mathbb{X},s}$ then it holds

$$\mathscr{E}(x+z) \leq \mathscr{E}(z) + \langle \mathscr{E}'(x), z\rangle + \frac{\mathscr{L}}{2}\|z\|^2.$$

Lemma 4.2 ([1, Lemma 4 p. 467]). Let $\{\lambda_m\}_{m\in\mathbb{N}}$ be a sequence of strictly positive real numbers such that, for some $\mu > 0$ we have $\lambda_m - \lambda_{m+1} \geq \mu\lambda_m^2$ for all $m \geq 1$. Then $\lambda_m = O(m^{-1})$.

The next proposition provides an inequality that will be useful to prove Theorem 4.2.

Proposition 4.2. Let $\mathbb{X}$ be a reflexive Banach space. Assume that $\mathscr{E}:\mathbb{X}\longrightarrow\mathbb{R}$ satisfies (A)-(B) and $u^*\in\mathbb{X}$ solves (2.1). Let $\Omega_0=\mathscr{E}^{-1}([\mathscr{E}(u^*),\mathscr{E}(0)])$, then there exists $c=c(\Omega_0)>0$ such that

$$\mathscr{E}(u)-\mathscr{E}(u^*)\le c\|\mathscr{E}'(u)\|_*$$

holds for all $u\in\Omega_0$.

Proof. See Appendix A. $\square$

From the above result and Proposition 4.1(a) we obtain the following.

Corollary 4.1. Let $\mathbb{X}$ be a reflexive Banach space and $\mathscr{D}$ be a universal dictionary of $\mathbb{X}$. Assume that $\mathscr{E}:\mathbb{X}\longrightarrow\mathbb{R}$ satisfies (A)-(B) and $u^*\in\mathbb{X}$ solves (2.1). Let $\Omega_0=\mathscr{E}^{-1}([\mathscr{E}(u^*),\mathscr{E}(0)])$, then there exists $c=c(\Omega_0)>0$ such that for each greedy sequence $\{u_m\}_{m\ge1}$ of u^* by a dictionary optimization over $\mathscr{D}$,

$$\mathscr{E}(u_m)-\mathscr{E}(u^*)\le c\|\mathscr{E}'(u_m)\|_*$$

holds for $m\ge1$.

Now, we have all ingredients to prove Theorem 4.2

Proof of Theorem 4.2. Put $\lambda_m=\mathscr{E}(u_m)-\mathscr{E}(u^*)$ for $m\ge1$. We claim that the sequence $(\lambda_m)_{m\ge1}$ satisfies the hypothesis of Lemma 4.2. Assume that $u_{m+1}=u_m+\lambda_{m+1}w_{m+1}$ where $\|w_{m+1}\|=1$ and $\lambda_{m+1}\in\mathbb{R}$.

Recall that Ω_0 is closed and bounded (see Section 3.2). Then the set

$$\Omega_0-\Omega_0=\{z\in X: z=z_1-z_2 \text{ where } z_i\in\Omega_0 \text{ for } i=1,2\}$$

is also bounded. Fix $s>0$ be such that $\Omega_0-\Omega_0\subset B_{\mathbb{X},s}$. Take $a=r+s$ for some $r>0$. Let $\mathscr{L}$ be the Lipschitz continuity constant of $\mathscr{E}'$ on the bounded set $B_{\mathbb{X},a}$. Then, for $m\ge1$,

$$\|\mathscr{E}'(u_m)\|_*=\|\mathscr{E}'(u_m)-\mathscr{E}'(u^*)\|\le\mathscr{L}\|u_m-u^*\|\le\mathscr{L}s,$$

holds.

Now, from Lemma 4.1,

$$\begin{aligned}\mathscr{E}(u_{m+1}) &\le \mathscr{E}(u_m+\mu w_{m+1})\\ &\le \mathscr{E}(u_m)+\mu\langle\mathscr{E}'(u_m),w_{m+1}\rangle+\tfrac{\mathscr{L}}{2}\mu^2\\ &\le \mathscr{E}(u_m)+\mu\|\mathscr{E}'(u_m)\|_*+\tfrac{\mathscr{L}}{2}\mu^2\end{aligned}$$

holds for $\|\mu w_{m+1}\|=|\mu|\le s$, in particular $|\mu_{m+1}|=\|u_{m+1}-u_m\|\le s$. The minimum for

$$\gamma(\mu)=\mathscr{E}(u_m)+\mu\|\mathscr{E}'(u_m)\|_*+\frac{\mathscr{L}}{2}\mu^2$$

is obtained at $\widetilde{\mu}=-\frac{1}{\mathscr{L}}\|\mathscr{E}'(u_m)\|_*$, where $|\widetilde{\mu}|\le s$, and $\gamma(\widetilde{\mu})=\mathscr{E}(u_m)-\frac{1}{2\mathscr{L}}\|\mathscr{E}'(u_m)\|_*^2$. Hence,

$$\mathscr{E}(u_{m+1})\le\mathscr{E}\left(u_m-\frac{1}{\mathscr{L}}\|\mathscr{E}'(u_m)\|_*w_{m+1}\right)\le\mathscr{E}(u_m)-\frac{1}{2\mathscr{L}}\|\mathscr{E}'(u_m)\|_*^2.$$

Thus,

$$\lambda_m-\lambda_{m+1}=\mathscr{E}(u_m)-\mathscr{E}(u_{m+1})\ge\frac{1}{2\mathscr{L}}\|\mathscr{E}'(u_m)\|_*^2.$$

On the other hand, by Corollary 4.1, we have for some $c>0$

$$\lambda_m=\mathscr{E}(u_m)-\mathscr{E}(u^*)\le c\|\mathscr{E}'(u_m)\|_*,$$

for $m \geq 1$. We conclude that

$$\lambda_m - \lambda_{m+1} \geq \frac{1}{2c^2\mathscr{L}}\lambda_m^2.$$

Then the theorem follows from Lemma 4.2 taking $\mu = \frac{1}{2c^2\mathscr{L}}$. □

5. Conclusions and final remarks

In this paper we propose a general mathematical framework to solve non-linear convex problems by a greedy sequence over a dictionary optimization. It appears as a supervised learning approach to the classical Method of Steepest Descend in a reflexive Banach space. Moreover, we introduce the radial dictionaries as a class of sets generated by closed and bounded sets. This class includes, among others, tensors in Tucker format with bounded rank and Neural Networks with fixed architecture and bounded parameters. We also point out that in our framework the convergence of the Proper Generalized Decomposition (PGD) in a reflexive Banach space, that has been proved in [15], is fulfilled. The novelty is that we are able to provide a convergence rate of $O(m^{-1})$ for the PGD.

Appendix A. Proof of Proposition 4.1 and Proposition 4.2

To prove Proposition 4.1 we need the following technical lemma.

Lemma A.1. Assume that $\mathbb{X}$ is a reflexive Banach space and take $\mathscr{D}$ a universal dictionary of $\mathbb{X}$. Assume that $\mathscr{E} : \mathbb{X} \longrightarrow \mathbb{R}$ satisfies (A)-(B) and $u^* \in \mathbb{X}$ solves (2.1). Then for each greedy sequence $\{u_m\}_{m\geq 1}$ of u^* by a $\mathscr{E}$-dictionary optimization over $\mathscr{D}$ the following statements hold.

(a) The sequence $\{\mathscr{E}'(u_m)\}_{m\geq 1}$, satisfies $\langle \mathscr{E}'(u_m), u_m - u_{m-1}\rangle = 0$ for all $m \geq 1$.

(b) It holds $\sum_{m=1}^{\infty} \|u_m - u_{m-1}\|^s < \infty$, for some $s > 1$, and thus,

$$\lim_{m\to\infty} \|u_m - u_{m-1}\| = 0. \tag{A.1}$$

(c) There exists $C > 0$ such that for $m \geq 1$,

$$|\langle \mathscr{E}'(u_{m-1}), z\rangle| \leqslant C\|u_m - u_{m-1}\|^2\|z\|,$$

holds for all $z \in \mathscr{D}$.

Proof. (a) Take $u_m - u_{m-1} = \widetilde{\varepsilon}_m \in \nabla_{u_{m-1}}(\mathscr{E};\mathscr{D})$ for some $m \geq 1$. Put $\widetilde{\varepsilon}_m = \lambda_m w_m$, with $\lambda_m \in \mathbb{R}^+$ and $\|w_m\| = 1$. Recall that $\widetilde{\varepsilon}_m \in \arg\min_{z\in\mathscr{D}} \mathscr{E}(u_{m-1} + z)$. Since $\mathscr{D}$ is a cone, we obtain

$$\mathscr{E}(u_{m-1} + \lambda_m w_m) \leq \mathscr{E}(u_{m-1} + \lambda w_m)$$

for all $\lambda \in \mathbb{R}$. Taking $\lambda = \lambda_m \pm \gamma$, with $\gamma \in \mathbb{R}^+$, we obtain for all cases

$$0 \leq \frac{1}{\gamma}\left(\mathscr{E}(u_{m-1} + \lambda_m w_m \pm \gamma w_m) - \mathscr{E}(u_{m-1} + \lambda_m w_m)\right).$$

Taking the limit $\gamma \searrow 0$, we obtain $0 \leq \pm\langle \mathscr{E}'(u_{m-1} + \lambda_m w_m), w_m\rangle$ and therefore

$$\langle \mathscr{E}'(u_{m-1} + \lambda_m w_m), w_m\rangle = 0,$$

which ends the proof.

(b) By the condition (B) of $\mathscr{E}$, we have

$$\mathscr{E}(u_{m-1}) - \mathscr{E}(u_m) \geq \langle -\mathscr{E}'(u_m), u_m - u_{m-1}\rangle + \frac{\alpha}{s}\|u_m - u_{m-1}\|^s$$

for some $s > 1$ and $\alpha > 0$. Using (a), we obtain

$$\mathscr{E}(u_{m-1}) - \mathscr{E}(u_m) \geq \frac{\alpha}{s}\|u_m - u_{m-1}\|^s$$

Now, summing on m, and using $\lim_{m\to\infty}\mathscr{E}(u_m) = L < \infty$ (by (a)), we obtain

$$\frac{\alpha}{s}\sum_{m=1}^{\infty}\|u_m - u_{m-1}\|^s \leq \sum_{m=1}^{\infty}(\mathscr{E}(u_{m-1}) - \mathscr{E}(u_m)) = \mathscr{E}(0) - L < +\infty.$$

which implies $\lim_{m\to\infty}\|u_m - u_{m-1}\|^s = 0$. The continuity of the map $x \mapsto x^{1/s}$ at $x = 0$ proves (A.1).

(c) By using that $\{\mathscr{E}(u_m)\}_{m\in\mathbb{N}}$ is bounded sequence and the functional $\mathscr{E}$ is coercive we obtain that $\{u_m\}_{m\geq 1}$ is bounded. Then there exists $r > 0$ be such that $\|u_m\| \leq r$ holds for all $m \geq 1$. From (b), $\|u_m - u_{m-1}\| \to 0$ as $m \to \infty$, hence $\{u_m - u_{m-1}\}_{m\geq 1}$ is also a bounded sequence and hence there exists $s > 0$ be such that $\|u_m - u_{m-1}\| \leq s$ holds for all $m \geq 1$. Take $a = r + s > 0$ and let $\mathscr{L}$ be the Lipschitz continuity constant of $\mathscr{E}'$ on the bounded set $B_{\mathbb{X},a}$. Then

$$\begin{aligned} -\langle \mathscr{E}'(u_{m-1}), z\rangle &= \langle \mathscr{E}'(u_{m-1}+z) - \mathscr{E}'(u_{m-1}), z\rangle - \langle \mathscr{E}'(u_{m-1}+z), z\rangle \\ &\leq \|\mathscr{E}'(u_{m-1}+z) - \mathscr{E}'(u_{m-1})\|_*\|z\| - \langle \mathscr{E}'(u_{m-1}+z), z\rangle \\ &\leq \mathscr{L}\|z\|^2 - \langle \mathscr{E}'(u_{m-1}+z), z\rangle \end{aligned}$$

for all $z \in B_{\mathscr{D},s} \subset B_{\mathbb{X},a}$. By (2.3) and since $\mathscr{E}(u_m) \leq \mathscr{E}(u_{m-1}+z)$ holds for all $z \in \mathscr{D}$, we have

$$\langle \mathscr{E}'(u_{m-1}+z), (u_m - u_{m-1}) - z\rangle \leq \mathscr{E}(u_m) - \mathscr{E}(u_{m-1}+z) \leq 0.$$

Therefore, for all $z \in B_{\mathscr{D},s}$, we have

$$\begin{aligned} -\langle \mathscr{E}'(u_{m-1}), z\rangle &\leq \mathscr{L}\|z\|^2 - \langle \mathscr{E}'(u_{m-1}+z), u_m - u_{m-1}\rangle \\ &\leq \mathscr{L}\|z\|^2 - \langle \mathscr{E}'(u_{m-1}+z) - \mathscr{E}'(u_{m-1}), u_m - u_{m-1}\rangle \quad \text{(by (a))} \\ &\leq \mathscr{L}\|z\|^2 + \mathscr{L}\|z - (u_m - u_{m-1})\|\|u_m - u_{m-1}\| \quad \text{(by (b))} \\ &\leq \mathscr{L}\left(\|z\|^2 + \|z\|\|u_m - u_{m-1}\| + \|u_m - u_{m-1}\|^2\right) \end{aligned}$$

Let $z = w\|u_m - u_{m-1}\| \in B_{\mathscr{D},s}$, with $\|w\| = 1$. Then

$$|\langle \mathscr{E}'(u_{m-1}), w\rangle| \leq 3\mathscr{L}\|u_m - u_{m-1}\|^2 \quad \text{for all } w \in \mathscr{D} \cap S_{\mathbb{X}}.$$

Taking $w = z/\|z\|$, with $z \in \mathscr{D}$, and $C = 3\mathscr{L} > 0$ we obtain

$$|\langle \mathscr{E}'(u_{m-1}), z\rangle| \leq C\|u_m - u_{m-1}\|^2\|z\| \quad \text{for all } z \in \mathscr{D}.$$

This concludes the proof of (c) and the lemma. □

Proof of Proposition 4.1. (a) Since $u_m = u_{m-1} + \nabla_{u_{m-1}}(\mathscr{E};\mathscr{D})$ and $0 \in \mathscr{D}$, we have

$$\mathscr{E}(u_m) = \min_{\xi \in \mathscr{D}} \mathscr{E}(u_{m-1}+\xi) \le \mathscr{E}(u_{m-1}).$$

Then, $\{\mathscr{E}(u_m)\}_{m\ge 1}$ is non increasing.

(b) By using the ellipticity property (2.2) of $\mathscr{E}$, we have

$$\mathscr{E}(u_m) - \mathscr{E}(u^*) \ge \langle \mathscr{E}'(u^*), u_m - u^*\rangle + \frac{\alpha}{s}\|u^* - u_m\|^s = \frac{\alpha}{s}\|u^* - u_m\|^s.$$

Therefore,

$$\frac{\alpha}{s}\|u^* - u_m\|^s \le \mathscr{E}(u_m) - \mathscr{E}(u^*) \underset{m\to\infty}{\longrightarrow} 0,$$

which ends the proof.

(c) Assume $u_m = u_{m-1}$ for some $m \ge 1$. Then $u_m - u_{m-1} = 0 \in \nabla_{u_{m-1}}(\mathscr{E};\mathscr{D})$. Hence, from Theorem 3.2, we conclude that u_{m-1} solves (2.1).

(d) Thanks to the reflexivity of the space $\mathbb{X}$, we can identify $\mathbb{X}^{**}$ with $\mathbb{X}$ and the duality pairing $\langle \cdot,\cdot\rangle_{\mathbb{X}^{**},\mathbb{X}^*}$ with $\langle \cdot,\cdot\rangle_{\mathbb{X}^*,\mathbb{X}}$ (recall that weak and weak-$*$ topologies coincide on $\mathbb{X}^*$). The sequence $\{u_m\}_{m\in\mathbb{N}}$ being bounded, and since by condition (A) $\mathscr{E}'$ is locally Lipschitz continuous, we have that there exists a constant $C > 0$ such that

$$\|\mathscr{E}'(u_m)\| = \|\mathscr{E}'(u_m) - \mathscr{E}'(u)\| \le C\|u - u_m\|.$$

This inequality proves that $\{\mathscr{E}'(u_m)\}_{m\in\mathbb{N}} \subset \mathbb{X}^*$ is a bounded sequence. Since $\mathbb{X}^*$ is also reflexive, from any subsequence of $\{\mathscr{E}'(u_m)\}_{m\in\mathbb{N}}$, we can extract a further subsequence $\{\mathscr{E}'(u_{m_k})\}_{k\in\mathbb{N}}$ that weakly-$*$ converges to an element $\varphi \in \mathbb{X}^*$. By using Lemma A.1(c), we have for all $z \in \mathscr{D}$,

$$|\langle \mathscr{E}'(u_{m_k}), z\rangle| \le C\|u_{m_{k+1}} - u_{m_k}\|^2\|z\|.$$

Taking the limit with $k\to\infty$, and using Lemma A.1(b), then $\varphi \in \mathbb{X}^*$ satisfies $\langle \varphi, z\rangle = 0$ for all $z \in \mathscr{D}$. Since $\mathscr{D}$ is a universal dictionary, we conclude that $\langle \varphi, z\rangle = 0$ holds for all $z \in \mathbb{X}$. Since from any subsequence of the initial sequence $\{\mathscr{E}'(u_m)\}_{m\in\mathbb{N}}$ we can extract a further subsequence that weakly-$*$ converges to the same limit 0, then the whole sequence converges to 0.

(e) Let $s^* > 1$ be such that $1/s^* + 1/s = 1$. From Lemma A.1(b), we have

$$\sum_{m=1}^{\infty} \|u_m - u_{m-1}\|^s < \infty.$$

Since $\lim_{m\to\infty}\|u_m - u_{m-1}\|^s = 0$ we can extract a subsequence $\{\|u_{m_k} - u_{m_{k-1}}\|^s\}_{k\ge 1}$ monotonically decreasing and such that

$$\sum_{k=1}^{\infty} \|u_{m_k} - u_{m_{k-1}}\|^s < \infty.$$

From Theorem 1 p.16 in [18], $\lim_{k\to\infty} m_k\|u_{m_k} - u_{m_{k-1}}\|^s \to 0$. For $1 < s \le 2$, we have $2 \le s^*$. Since $\lim_{k\to\infty}\|u_{m_k} - u_{m_{k-1}}\| = 0$, then

$$\|u_{m_k} - u_{m_{k-1}}\|^{2s^*} \le \|u_{m_k} - u_{m_{k-1}}\|^{2s}$$

for k large enough, and therefore $\lim_{k\to\infty} m_k\|u_{m_k} - u_{m_{k-1}}\|^{2s^*} = 0$.

Since $u_{m_k} = \sum_{\ell=1}^{k} u_{m_\ell} - u_{m_{\ell-1}}$, we have

$$\begin{aligned}|\langle \mathscr{E}'(u_{m_k}), u_{m_k}\rangle| &\leq \sum_{\ell=1}^{k} |\langle \mathscr{E}'(u_{m_k}), u_{m_\ell} - u_{m_{\ell-1}}\rangle| \\ &\leq C \sum_{\ell=1}^{k} \|u_{m_{k+1}} - u_{m_k}\|^2 \|u_{m_\ell} - u_{m_{\ell-1}}\| \quad \text{(By Lemma A.1(c))}.\end{aligned}$$

By Holder's inequality, we have

$$\begin{aligned}|\langle \mathscr{E}'(u_{m_k}), u_{m_k}\rangle| &\leq C \left(m_k \|u_{m_{k+1}} - u_{m_k}\|^{2s^*} \right)^{1/s^*} \left(\sum_{\ell=1}^{k} \|u_{m_\ell} - u_{m_{\ell-1}}\|^s \right)^{1/s} \\ &\leq C \left(m_{k+1} \|u_{m_{k+1}} - u_{m_k}\|^{2s^*} \right)^{1/s^*} \left(\sum_{\ell=1}^{k} \|u_{m_\ell} - u_{m_{\ell-1}}\|^s \right)^{1/s}.\end{aligned}$$

By taking limits as $k \to \infty$ in the above inequality we obtain the desired result. This ends the proof of proposition. □

Proof of Proposition 4.2. Recall that Ω_0 is closed and bounded (see Section 3.2). Then the set

$$\Omega_0 - \Omega_0 = \{z \in \mathbb{X} : z = z_1 - z_2 \text{ where } z_i \in \Omega_0 \text{ for } i = 1, 2\}$$

is also bounded. Fix $c > 0$ be such that $\Omega_0 - \Omega_0 \subset B_{c,\mathbb{X}}$. Take $u \in \Omega_0$ and, by (2.3),

$$\mathscr{E}(u+z) - \mathscr{E}(u) \geq \langle \mathscr{E}'(u), z\rangle$$

holds for all $z \in \mathbb{X}$. Hence,

$$\min_{z \in B_{\mathbb{X},c}} \mathscr{E}(u+z) - \mathscr{E}(u) \geq \min_{z \in B_{\mathbb{X},c}} \langle \mathscr{E}'(u), z\rangle. \tag{A.2}$$

As $u \in \Omega_0$, we have $u - \Omega_0 \subset \Omega_0 - \Omega_0 \subset B_{\mathbb{X},c}$, so that $\Omega_0 \subset u + B_{\mathbb{X},c}$. Therefore,

$$\mathscr{E}(u^*) = \min_{z \in X} \mathscr{E}(z) \leq \min_{z \in B_{\mathbb{X},c}} \mathscr{E}(u+z) \leq \min_{z \in \Omega_0} \mathscr{E}(z) = \mathscr{E}(u^*).$$

Thus, from (A.2),

$$\mathscr{E}(u^*) - \mathscr{E}(u) \geq \min_{z \in B_{\mathbb{X},c}} \langle \mathscr{E}'(u), z\rangle$$

holds. Now, by using

$$\min_{z \in B_{\mathbb{X},c}} \langle \mathscr{E}'(u), z\rangle = c \min_{z \in B_{\mathbb{X},1}} \langle \mathscr{E}'(u), z\rangle = -c \max_{z \in B_{\mathbb{X},1}} \langle \mathscr{E}'(u), z\rangle = -c\|\mathscr{E}'(u)\|_*,$$

the proposition follows. □

Acknowledgments
Antonio Falcó would thank to Marie Billaud-Friess, Régis Lebrun, Bertrand Michel, Anthony Nouy and Philipp Petersen their helpful discussions along the days in the Research School on High-dimensional Approximation and Deep Learning held in Nantes from May 16 to 20, 2022.
Funding : P. M. Berná was partially supported by the Grant PID2022-142202NB-I00 (Agencia Estatal de Investigación, Spain). A. Falcó was partially supported by the grant

number INDI21/15 from Universidad CEU Cardenal Herrera and by ESI International Chair@ CEU-UCH.

References

[1] G. P. Akilov and L. V. Kantorovich, Functional analysis, 2d ed, Pergamon Press, 1982.
[2] Ammar A., Huerta A., Chinesta F., Cueto E., and Leygue A., Parametric solutions involving geometry: A step towards efficient shape optimization, Computer Methods in Applied Mechanics and Engineering 268 (2014), 178–193, DOI 10.1016/j.cma.2013.09.003.
[3] M. Bachmayr, R. Schneider, and A. Uschmajew, Tensor Networks and Hierarchical Tensors for the Solution of High-Dimensional Partial Differential Equations, Found. Comput. Math. 16 (2016), 1423–1472, DOI 10.1007/s10208-016-9317-9.
[4] M. Bachmayr and V. Kazeev, Stability of Low-Rank Tensor Representations and Structured Multilevel Preconditioning for Elliptic PDEs, Found. Comput. Math. 20 (2020), 1175–1236, DOI 10.1007/s10208-020-09446-z.
[5] Boucinha L., Gravouil A., and Ammar A., Space–time proper generalized decompositions for the resolution of transient elastodynamic models, Computer Methods in Applied Mechanics and Engineering 255 (2013), 67-88, DOI 10.1016/j.cma.2012.11.003.
[6] V. Candela, A. Falcó, and P. D. Romero, A general framework for a class of non-linear approximations with applications to image restoration, Journal of Computational and Applied Mathematics 330 (2018), 982-994.
[7] V. Candela and P. D. Romero, Blind Deconvolution Models Regularized by Fractional Powers of the Laplacian, J. Math. Imaging. Vis. 32 (2008), 181–191.
[8] C. Canuto and K. Urban, Adaptive optimization of convex functionals in Banach spaces, SIAM J. Numer. Anal. 42 (2005), 2043-2075.
[9] P. G. Ciarlet, Introduction to Numerical Linear Algebra and Optimization, Cambridge Texts in applied Mathematics, Cambridge University Press, 1989.
[10] Coelho Lima I., Robens-Radermacher A., Titscher T., Kadoke D., Koutsourelakis P-S., and Unger J. F., Bayesian inference for random field parameters with a goal-oriented quality control of the PGD forward model's accuracy, Computational Mechanics, posted on 2022, DOI 10.1007/s00466-022-02214-6.
[11] G. Dinca, P. Jebelean, and J. Mawhin, Variational and topological methods for Dirichlet problems with p-Laplacian, Portugaliae Mathematica. Nova Série 58 (2001), 339-378.
[12] I. Ekeland and R. Teman, Convex Analysis and Variational Problems, Classics in Applied Mathematics. SIAM, 1999.
[13] A. Falcó, W. Hackbusch, and A. Nouy, On the Dirac–Frenkel Variational Principle on Tensor Banach Spaces, Found. Comput. Math. 19 (2019), 159–204, DOI 10.1007/s10208-018-9381-4.
[14] A. Falcó and W. Hackbusch, On Minimal Subspaces in Tensor Representations, Found. Comp. Math. 12 (2012), no. 6, 765–803, DOI 10.1016/j.jfa.2013.02.003. MR3029146
[15] A. Falcó and A. Nouy, Proper generalized decomposition for nonlinear convex problems in tensor Banach spaces, Number. Math. 121 (2012), 503–530, DOI 10.1016/j.jfa.2013.02.003. MR3029146
[16] González D., Masson F., Poulhaon F., Leygue A., Cueto E., and Chinesta F., Proper Generalized Decomposition based dynamic data driven inverse identification, Mathematics and Computers in Simulation 82 (2012), no. 9, 1677–1695, DOI 10.1016/j.matcom.2012.04.001.
[17] W. H. Greub, Linear Algebra, 4th ed., Graduate Text in Mathematics, Springer-Verlag, 1981.
[18] K. Knop, Infinite Sequences and Series, Dover Publications, Inc., 1956.
[19] Leygue A. and Verron E., A first step towards the use of proper general decomposition method for structural optimization, Arch. Computat. Methods. Eng. 17 (2010), 465–472, DOI 10.1007/ s11831-010-9052-3.
[20] Niroomandi S., González D., Alfaro I., Bordeu F., Leygue A., Cueto E., and Chinesta F., Real-time simulation of biological soft tissues: A PGD approach, Int. J. Numer. Meth. Biomed. Engng. 29 (2013), 586–600, DOI 10.1002/cnm.2544.

[21] M. Mohri, A. Rostamizadeh, and A. Talwalkar, Foundations of Machine Learning, Second Edition, Adaptive Computation and Machine Learning, The MIT Press, 2018.

[22] C. Marcati, J.A.A. Opschoor, P.C. Petersen, and Ch. Schwab, Exponential ReLU Neural Network Approximation Rates for Point and Edge Singularities, Found. Comput. Math., posted on 2022, DOI 10.1007/s10208-022-09565-9.

[23] P. Petersen, M. Raslan, and F. Voigtlaender, Topological Properties of the Set of Functions Generated by Neural Networks of Fixed Size, Found. Comput. Math. 21 (2021), 375–444.

[24] M. Raissi and G.E. Karniadakis, Hidden physics models: machine learning of nonlinear partial differential equations, J. Comput. Phys. 357 (2018), 125–141, DOI 10.1016/j.jcp.2017.11.039.

[25] M. Raissi, P. Perdikaris, and G.E. Karniadakis, Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, J. Comput. Phys. 378 (2019), 686–707, DOI 10.1016/j.jcp.2018.10.045.

[26] A. Sawicki and K. Karnas, Universality of Single-Qudit Gates, Ann. Henri Poincaré 18 (2017), 3515–3552, DOI 0.1007/s00023-017-0604-z.

[27] H. Sheng and C. Yang, PA penalty-free neural network method for solving a class of second-order boundary-value problems on complex geometries, J. Comput. Phys. 428 (2021), 110085, DOI 10.1016/j.jcp.2020.110085.

[28] Weinan E. and Bing Y., The Deep Ritz Method: A Deep Learning-Based Numerical Algorithm for Solving Variational Problems, Communications in Mathematics and Statistics 6 (2018), no. 1, 1-12, DOI 10.1007/s40304-018-0127-z.

[29] E. Zeidler, Nonlinear Functional Analysis and its Applications III. Variational Methods and Optimization., CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, vol. 26, Springer-Verlag, 1985.