Theoretical Computer Science 664 (2017) 12–28

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Sparse decomposition by iterating Lipschitzian-type mappings

Alessandro Adamo [a], Giuliano Grossi [b,*], Raffaella Lanzarotti [b], Jianyi Lin [b]

[a] Dipartimento di Matematica, Università degli Studi di Milano, Via Cesare Saldini 50, I-20133 Milano, Italy
[b] Dipartimento di Informatica, Università degli Studi di Milano, Via Comelico 39, I-20135 Milano, Italy

ARTICLE INFO

Article history:
Received 13 March 2015
Received in revised form 22 February 2016
Accepted 22 April 2016
Available online 26 April 2016

Keywords:
Underdetermined linear systems
Sparse recovery
Lipschitzian mappings
Smooth ℓ_0-norm
Smooth ℓ_1-norm
Fixed-point iteration scheme

ABSTRACT

This paper provides the analysis of a fast iterative method for finding sparse solutions to underdetermined linear systems. It is based on a fixed-point iteration scheme which combines nonconvex Lipschitzian-type mappings with canonical orthogonal projectors. The former are aimed at uniformly enhancing the sparsity level by shrinkage effects, the latter are used to project back onto the space of feasible solutions. The iterative process is driven by an increasing sequence of a scalar parameter that mainly contributes to approach the sparsest solutions. It is shown that the minima are locally asymptotically stable for a specific smooth ℓ_0-norm. Furthermore, it is shown that the points yielded by this iterative strategy are related to the optimal solutions measured in terms of a suitable smooth ℓ_1-norm. Numerical simulations on phase transition show that the performances of the proposed technique overcome those yielded by well known methods for sparse recovery.

1. Introduction

Finding sparse solutions, i.e. with a few non-zero entries, of underdetermined systems of linear equations is a topic extensively studied in signal processing [1–3]. Many problems across several disciplines can take advantages in finding sparse solutions as for instance data compression, reconstruction, separation, and transmission. In general, all these tasks amount to solving the problem

$$\Phi\alpha = s, \tag{1}$$

where $s = (s_1, \ldots, s_n)^T$ is a vector (for instance interpreted as discrete-time signal) and Φ is a $n \times m$ full-rank matrix with $n < m$. Depending on the various applicative contexts, $\Phi = [\phi_1, \ldots, \phi_m]$ is a collection of m basic waveforms or vectors in $\mathbb{R}^n$, usually called atoms, and the full-rank matrix Φ is called dictionary or frame. With the above assumptions, the dictionary will result in an overcomplete frame leading to infinite solutions of system (1) and $\alpha \in \mathbb{R}^m$ represents the coefficients of linear combination of the atoms in Φ.

Thus, the attitude of finding sparse decomposition of the signal s, that is solutions with very few non-zero terms, can be rephrased into the combinatorial optimization problem

$$\min_{\alpha \in \mathbb{R}^m} \| \alpha \|_0 \quad \text{subject to} \quad \Phi\alpha = s, \tag{P_0}$$

where $\| \alpha \|_0 = |\{j : \alpha_j \neq 0\}|$ denotes the ℓ_0-norm, or more properly said the ℓ_0-pseudonorm. This attitude is often referred to as ℓ_0-minimization.

* Corresponding author.
E-mail addresses: alessandro.adamo@unimi.it (A. Adamo), grossi@di.unimi.it (G. Grossi), lanzarotti@di.unimi.it (R. Lanzarotti), jianyi.lin@unimi.it (J. Lin).

http://dx.doi.org/10.1016/j.tcs.2016.04.025
0304-3975/

Problem (P_0) is combinatorial in nature and hence NP-hard [4]. Due to its intractability, many researchers have developed sparse decomposition solvers alternative to the exhaustive search. One of the most important classes is represented by the family of greedy iterative algorithms as, for instance, Matching Pursuit (MP) [5], Orthogonal Matching Pursuit (OMP) and Compressive Sampling Matching Pursuit (CoSaMP) [6]. MP is a fast method for approximating the problem (P_0) which optimizes one component of the sparse solution at a time. In the first step the closest atom to the input vector s is selected and its corresponding coefficient is determined, such that the representation error is minimum with only one non-zero element. At each subsequent step, a further coefficient corresponding to a new atom selected is optimized. The process goes on until the representation error is less than a threshold or the number of non-zero elements is more than a limit. To improve the accuracy of MP while maintaining its low complexity, some algorithms have been proposed. Among them, the most popular is the Orthogonal MP (OMP) [7] that at each step, after selecting the best atom, ignores the coefficients already computed and adopts the pseudo-inverse to determine the best representation (i.e., the least square error) of vector s on the selected atoms. CoSaMP is at heart a greedy pursuit method for reconstructing a signal embracing the compressed sensing theory[1]. Ultimately, it shares the OMP strategy, while offering rigorous bounds on computational cost and storage, and being extremely more efficient for practical problems.

A completely different (non-greedy) approach is the Smoothed-ℓ_0 algorithm (SL0) [8]. It tackles directly the problem (P_0) by minimizing the ℓ_0-norm in a coarse to fine approach. To overcome the discrete nature of the ℓ_0-norm, the method introduces a parametric family of smoothed functionals which converge to ℓ_0-norm in the limit of the controlling parameter.

Another category of approaches is represented by convex relaxations of (P_0), such as those relying to the ℓ_1-norm. This allows to recast (P_0) as a convex optimization problem solvable by linear programming (also referred to as Basis Pursuit – BP [2,3])

$$\min_{\alpha\in\mathbb{R}^m} \|\alpha\|_1 \quad \text{subject to} \quad \Phi\alpha = s, \tag{P_1}$$

where $\|\alpha\|_1 = \sum_i |\alpha_i|$. BP is based on the observation that for large systems of equations, the minimum ℓ_1-norm solution is also the minimum ℓ_0-norm solution under mild conditions [9]. By utilizing fast Linear Programming (LP) algorithms (specifically, interior-point LP solvers), large-scale problems, such as those with mixture of thousands of sources, become tractable. Since its introduction, this method gained much popularity demonstrating that ℓ_1-optimization still represents a preferable way to attack ℓ_0-minimization problems arising from signal and imaging processing. Another well-known method based on this minimization goal is the so called LASSO [10], which utilizes the ℓ_1-penalty to automatically select significant variables via continuous shrinkage, thus retaining the good features of both the best subset selection and the ridge regression. Other than the two mentioned methods, the literature is scattered of papers presenting various computational methods under this viewpoint [11–15].

Many works have addressed the existent relationship between ℓ_0 and ℓ_1-minimizations, as for instance in [3,16–19]. Among the numerous results, the most fruitful are those based on the so called Restrict Isometry Property (RIP [3,20]) and Null Space Property (NSP [21]) which have been extensively investigated in both noisy and noiseless data cases.

Finally, in the range of minimization norms, it is worth to mention also a special case of interest that comes from the well-known ℓ_p-norms [9]. Indeed, for $0 < p < 1$, they represent another way to drive the solutions to become sparse on the affine space despite the lack of convexity, thus solving the problem

$$\min_{\alpha\in\mathbb{R}^m} \|\alpha\|_p^p \quad \text{subject to} \quad \Phi\alpha = s. \tag{P_p}$$

Unfortunately, each choice of $0 < p < 1$ leads to a non-convex optimization problem. This raises some weakness due to the nonexistence of efficient algorithms to optimize it. Nevertheless, from the heuristic point of view, it represents a useful formulation for sparsity request. Analogously to ℓ_p-norms, sparsity can be promoted by other functionals of variable α. In particular, it would be worth to mention any function $J(\alpha)$ being component-wise symmetric, monotonically non-decreasing, and with monotonic non-increasing partial derivatives. Classic examples of this family are for instance $J(\alpha) = 1 - \exp(|\alpha|)$, $J(\alpha) = \tanh(|\alpha|)$ and $J(\alpha) = |\alpha|/(1 + |\alpha|)$, and their parametrized variants [9].

In this paper, we propose a theoretical analysis of LiMapS (which stands for Lipschitzian Mappings for Sparsity), which is an iterative method based on Lipschitzian mappings that on the one hand promotes sparsity and on the other hand restores feasibility of iterated solutions. Specifically, LiMapS adopts a nonlinear parametric family $\mathcal{F} = \{f_\lambda : \lambda > 0\}$ of shrinkage mappings acting on coefficients in order to drive the search toward highly sparse solutions. Moreover, LiMapS uses an orthogonal projector to map the obtained near-feasible point onto the affine space of solutions. The combination of these two mappings induces the iterative system to find the solution in the subspace spanned by as smaller as possible number of dictionary atoms.

The role of the parameter λ in the proposed family $\mathcal{F}$ of Lipschitzian-type mappings is crucial in determining how strong the shrinkage effect is within each step of the iterative process. On the one hand, for small parameter values the iterative mappings involved promote sparse solutions forcing the magnitude of all coefficients to become closer and closer to zero. On the other hand, for high parameter values, the chance of reducing the magnitude of each coefficient diminishes, fixing their values over the iterations.

[1] Compressed sensing theory asserts that one can recover certain signals and images from far fewer samples or measurements than traditional methods use, this relying on two principles: sparsity, which pertains to the signals of interest, and incoherence, which pertains to the sensing modality.

The main results of this work can be summarized as follows.

- We find that sparse solutions of underdetermined linear systems are well recovered by shrinkage functions having small Lipschitz constant L. For this reason we use functions f_λ having constant L independent from λ since it is scheduled to diverge (Section 2).
- We provide sufficient conditions for the sequence $\{\lambda_t\}$ of the control parameter λ in order to guarantee the iterative scheme convergence to admissible solutions (Section 2).
- We introduce a smooth version of ℓ_0-minimization problem, and we prove that the minima are locally asymptotically stable fixed points of the iterative scheme for a fixed large $\lambda > 0$ (Section 3).
- Finally, we prove that both smooth ℓ_0 and ℓ_1 minimization problems, as λ goes to infinite, "tend" to the exact versions, i.e. (P_0) and (P_1) problems. (Section 3).

To show the performances of the proposed method, we conducted many numerical experiments choosing signals and frames of various sizes $n \times m$, independently drawn from stochastic models frequently employed in this analysis, as for instance the Gaussian–Bernoulli model. These tests provide a useful comparison between LiMapS algorithm and classical well-known methods used in the sparse representation area. In fact, we empirically compute the phase transition threshold as relationship between the normalized sparsity and the problem indeterminacy n/m, thus distinguishing the regions with reconstruction-error probability approaching 1 from those in which such a probability vanishes.

The LiMapS algorithm was empirically proven effective in a preliminary work [22], and in recent works we showed that it is successful also when adopted in some applications such as image processing and information-theoretic tasks. Specifically, we applied the sparse representation technique to the face recognition problem with both noisy instances and poor galleries, that is with only few samples per subject [23], and we tackled the electrocardiogram (ECG) signal compression task using the natural basis extracted from the ECG signal, without any training or burdensome feature construction activity [24]. Good results in these applications are possible thank to the iterative nature of the method, that encompasses both low time complexity and significant performances (in terms of SNR). This behavior is achieved even in case of sparsity level near the rank of the coefficient matrix, a challenging condition that makes other methods perform poorly.

The paper is organized as follows. In Section 1.1 some preliminary notions are given. In Section 2 we introduce the shrinkage mapping adopted in LiMapS, we show the convergence properties of the iterative system used for finding sparse solutions and we present the overall pseudo-code. In Section 3 we give some minimality conditions and analyze the stability for solutions of the smooth ℓ_0 and ℓ_1-minimization problems. In Section 4 we discuss computational issues of the LiMapS algorithm and we report the numerical results about the transition phase phenomenon that characterizes this kind of sparsity solvers. In Section 5 some conclusions are drawn and possible future works are sketched.

1.1. Notations and preliminary notions

In this section we recall some standard geometry notions and fix the notation. Let $A \in \mathbb{R}^{n\times m}$ be a full rank matrix; when $n \le m$ the (Moore–Penrose) pseudo-inverse of A is $A^\dagger = A^T(AA^T)^{-1}$, while if $n \ge m$ the pseudo-inverse of A is $A^\dagger = (A^TA)^{-1}A^T$. If A has not full rank, its pseudo-inverse is $A^\dagger = \lim_{\varepsilon\downarrow 0}(A^TA + \varepsilon I)^{-1}A^T = \lim_{\varepsilon\downarrow 0} A^T(AA^T + \varepsilon I)^{-1}$. Given a linear map represented by $A : x \mapsto Ax$, we denote the orthogonal projections onto $\ker A$ and $(\ker A)^\perp$ with $P = \mathrm{Proj}_{\ker A}$ and $Q = \mathrm{Proj}_{(\ker A)^\perp} = I - P$ respectively. The two projectors can be effectively computed as $P = I - A^\dagger A$ and $Q = A^\dagger A$. Let us recall the direct sum $\mathbb{R}^m = \ker A \oplus (\ker A)^\perp$ yielding $x = Px + Qx$ for any $x \in \mathbb{R}^m$, and $\mathrm{rank}(P) = \dim(\ker A) \le m - n$ and $\mathrm{rank}(Q) = \dim(\mathrm{im}A) \le n$.

Given a linear system $\Phi x = s$ the set of solutions x forms the affine set $\mathcal{A}_{\Phi,s} = \ker\Phi + \bar{x}$ with $\bar{x}$ being any particular solution of the system. When the system is solvable, the solution with minimum norm is $\nu = \mathrm{argmin}_{x\in\mathcal{A}_{\Phi,s}} \|x\| = \Phi^\dagger s \in (\ker\Phi)^\perp$.

The real vector space $\mathbb{R}^m$ endowed with the inner product $\langle x, y\rangle = \sum_1^m x_i y_i$ is a Hilbert space, and hence a Banach space with the induced norm $\|x\| = \sqrt{\langle x, x\rangle}$. This means that convergence in the space is guaranteed for any Cauchy sequence, i.e. any $\{x_i\} \subset \mathbb{R}^m$ s.t. $\forall\varepsilon > 0$ there exists $N = N(\varepsilon)$ satisfying $\|x_i - x_j\| < \varepsilon$ for all $i, j > N$. Consider the functional $\|x\|_p = (\sum_{i=1}^m |x_i|^p)^{1/p}$; it is known that this functional is a norm for $p \ge 1$, called ℓ_p-norm, and is a quasi-norm [25] for $0 < p < 1$. Moreover, the functional $\lim_{p\downarrow 0}\|x\|_p^p = \|x\|_0$ satisfies the axioms of the norm except the

$$\text{absolute homogeneity: } \forall\lambda \in \mathbb{R}, x \in \mathbb{R}^m \ \|\lambda x\| = |\lambda|\|x\|$$

and hence is called a pseudo-norm. This pseudo-norm, also improperly called the ℓ_0-norm, is the main measure of sparsity.

We denote the Hadamard (or element-wise) product between two vectors $x, y \in \mathbb{R}^m$ by $x \odot y = (x_1y_1, x_2y_2, \ldots, x_my_m)^T \in \mathbb{R}^m$. Such a product operator has precedence over matrix multiplication (in particular over the projections P, Q). Moreover, the absolute value of a vector is understood element-wise, i.e. $|x| = (|x_1|, |x_2|, \ldots, |x_m|)^T \in \mathbb{R}^m$ for any $x \in \mathbb{R}^m$.

Let $T : A \to A$ be a self-mapping on a set A; T can be considered a simple *discrete-time dynamical system* or *iterative system* [26]. Indeed, given an element $a_0 \in A$, the sequence $\{a_n\}$ generated from a_0 by

$$a_n = T(a_{n-1}) = \underbrace{T \circ T \circ \ldots \circ T}_{n \text{ times}}(a_0) \quad \text{for } n = 1, 2, \ldots$$

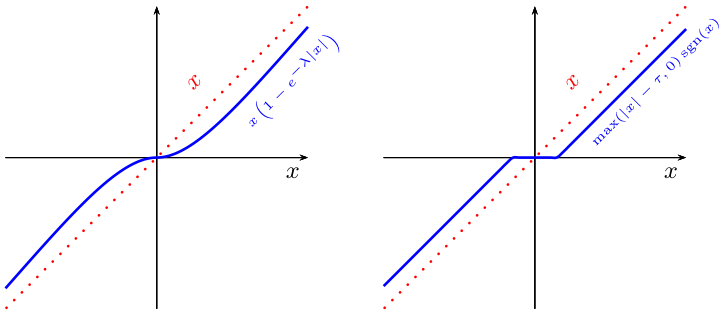

Fig. 1. The graphs of shrinking functions $f_\lambda(x)$ for $\lambda = 1$ and soft threshold function $s(x)$ for $\tau = 5$.

is the orbit of a_0 under T. Let $T : A \to A$ and $S : B \to B$ be two continuous iterative systems; a (topological) *conjugacy* is a homeomorphism $h : B \to A$ such that $T \circ h = h \circ S$. Equivalently, $T = h \circ S \circ h^{-1}$ and $S = h^{-1} \circ T \circ h$, which are hence said to be (topologically) conjugate. Conjugacy is useful for instance to study T by analyzing its conjugate S, that may be less complex. Moreover, an element $p \in A$ s.t. $T(p) = p$ is called a fixed point of T; the set of fixed points of T is denoted Fix $T \subseteq A$.

In context of sparsity, given a vector $\alpha \in \mathbb{R}^m$ we denote by $\mathrm{supp}(\alpha) = \{j : \alpha_j \neq 0\}$ the index set of non-zero entries in α, also called support of α and we say that α is k-sparse if $\|\alpha\|_0 = |\mathrm{supp}(\alpha)| \leq k$. Given a matrix A, the minimum number of linearly dependent columns in A is denoted $\mathrm{spark}(A) > \mathrm{rank}(A)$.

2. Sparsity promotion mappings

The aim of this section is to build a convergent iterative scheme for variable selection based on shrinkage mappings. This is done relying on a family of uniformly Lipschitzian nonlinear mappings $\mathcal{F} = \{f_\lambda \mid \lambda \in \mathbb{R}_+\}$, where the choice of the parameter λ is important in controlling the shrinkage effects, i.e. to drive the search towards the sparsest solutions. After a pure magnitude reduction, in order to restore the feasibility we apply the classical linear orthogonal projectors built on the Moore–Penrose pseudo-inverse. Notice that, in literature shrinkage functions are used to optimize the regularization term of the objective function [27,28]. Conversely, in our approach the shrinkage mapping aims at penalizing nonlinearly the coefficients close to zero.

2.1. Lipschitzian shrinkage mappings

The key feature of all shrinkage functions [29] is to attenuate coefficients adaptively with respect to scale and time, taking properly into account the general constraint strength of the signal. In the present work this task is committed to a function we first introduce in the univariate form and then extend to the multivariate case.

Let $f_\lambda : \mathbb{R} \to \mathbb{R}$ be a function depending on a real parameter $\lambda > 0$, defined as

$$f_\lambda(x) = x\,(1 - e^{-\lambda|x|}). \tag{2}$$

The function f_λ is odd and continuously differentiable in $\mathbb{R}$, while its derivative $f'_\lambda(x) = (\lambda|x| - 1)e^{-\lambda|x|} + 1$ is even and non-negative. Since it holds that $\sup_{x\in\mathbb{R}} |f'_\lambda(x)| = 1 + e^{-2}$, as a direct consequence of the intermediate value theorem from calculus $|f_\lambda(x) - f_\lambda(y)| \leq (1 + e^{-2})|x - y|$, for every $\lambda > 0$ and $x, y \in \mathbb{R}$. Thus, mapping (2) is uniformly Lipschitzian with respect to λ with Lipschitz constant $L = 1 + e^{-2}$. Moreover, in any compact subinterval of $(-1/\lambda, 1/\lambda)$, the mapping f_λ is a contraction with unique fixed point at the origin [30]. Naturally, it obeys to the shrinkage rule $|f_\lambda(x)| \leq |x|$ for all $x \in \mathbb{R}$ and it is clearly non-decreasing (see for instance its graph in Fig. 1). For sake of comparison, it is interesting also to show the well-known [31] soft threshold function $s(x) = \max(|x| - \tau, 0)\,\mathrm{sgn}(x)$ which arises frequently in sparse signal processing and compressed sensing. Conversely, the latter function is discontinuous and maps small values (at most τ in absolute value) to zero while it shrinks large values (greater than τ in absolute value) toward zero.

To understand the meaning of f_λ, it should be observed the role played by the mapping $f_\lambda(x)/x$: it behaves like a symmetric sigmoid function, where larger values of λ give sharper sigmoids, becoming in the limit a Heaviside step function. Roughly speaking, this asymptotic behavior has nonlinear shrinking effects for values near zero because the function $f_\lambda(x) \sim x^2$, while far from the origin it is asymptotic to the identity function, being $\lim_{\lambda\to+\infty} |f_\lambda(x)| = x$ for all $x \in \mathbb{R}$.

To deal with high dimensional data, we extend the mapping f_λ to many dimensions by applying it separately to all variables. Precisely, given $\lambda > 0$ we define the vector-valued mapping $F_\lambda : \mathbb{R}^m \to \mathbb{R}^m$ by

$$F_\lambda(x) = (f_\lambda(x_1), f_\lambda(x_2), \ldots, f_\lambda(x_m))^T = x \odot \left(\mathbf{1} - e^{-\lambda|x|}\right) \tag{3}$$

Analogously to the scalar case, the function $(f_\lambda(x_1)/x_1, \ldots, f_\lambda(x_m)/x_m)$ represents a origin symmetric sigmoid in m dimensions, where larger values of λ give sharper sigmoids. The mapping F_λ has the effect of shrinking toward $\underline{0}$, but posing in this way a feasibility issue for the shrinked point, i.e. generally $y = F_\lambda(x) \notin \mathcal{A}_{\Phi,s}$.

In order to address the feasibility issue, we use a suitable projection. Given a point $y = F_\lambda(x) \in \mathbb{R}^m$, we choose the unique point α in $\mathcal{A}_{\Phi,s}$ minimizing the distance from y:

$$\alpha = \mathrm{Proj}_{\mathcal{A}_{\Phi,s}}(y) = \underset{\alpha \in \mathcal{A}_{\Phi,s}}{\operatorname{argmin}} d(\alpha, y) = Py + \nu \in \mathcal{A}_{\Phi,s},$$

where $P = I - \Phi^\dagger\Phi$ is the usual orthogonal projector onto $\ker\Phi$, and $\nu = \Phi^\dagger s \in (\ker\Phi)^\perp$.

These initial considerations suggest the definition of a new mapping by composing the shrinkage mapping F_λ and the feasibility mapping $y \mapsto \alpha$, i.e. we construct the self-mapping $G_\lambda : \mathcal{A}_{\Phi,s} \to \mathcal{A}_{\Phi,s}$ as

$$G_\lambda(x) = PF_\lambda(x) + \nu = x - Px \odot e^{-\lambda|x|}. \tag{4}$$

We can provide the Lipschitz constant also for mapping G_λ. Indeed, by recalling that P is non-expansive and each component of F_λ is f_λ (whose Lipschitz constant is $L = 1 + e^{-2}$), it is easy to see that for any $\alpha, \beta \in \mathcal{A}_{\Phi,s}$

$$\begin{aligned} \| G_\lambda(\alpha) - G_\lambda(\beta) \| &= \| P[F_\lambda(\alpha) - F_\lambda(\beta)] \| \\ &\le \| F_\lambda(\alpha) - F_\lambda(\beta) \| \\ &\le \left(1 + e^{-2}\right) \| \alpha - \beta \| \end{aligned} \tag{5}$$

and the constant is tight. Hence G_λ has also Lipschitz constant $L = 1 + e^{-2}$. Lipschitz continuity quantifies the boundedness of derivatives for a function by the value L. Note that L is independent of λ. This is a useful property since the distance between the images of G_λ does not increase as $\lambda \to \infty$, which is required in the following to prove the convergence of our proposed scheme.

Remark 1. Notice that by definition of G_λ,

$$\alpha \in \mathrm{Fix}\, G_\lambda \qquad \Longleftrightarrow \qquad P\alpha \odot e^{-\lambda|\alpha|} = \underline{0}.$$

2.2. An iterative system

Consider an increasing sequence of real numbers $\{\lambda_t\}_{t\ge0} \subseteq \mathbb{R}_+$ and the corresponding family of self-mappings $\{G_{\lambda_t}\}_{t\ge0}$ as defined by eq. (4). We construct the sequence of iterates $\{\alpha_t\}_{t\ge0}$ generated by the family $\{G_{\lambda_t}\}_{t\ge0}$ from $\alpha_0 \in \mathcal{A}_{\Phi,s}$ as

$$\alpha_t = G_{\lambda_{t-1}}(\alpha_{t-1}), \quad \text{for } t \ge 1. \tag{6}$$

Notice that $\alpha_t \in \mathcal{A}_{\Phi,s}$ for all $t \ge 0$. We now study the convergence of the sequence α_t. First, we provide a lemma expressing the general term α_t in inductive form.

Lemma 1. *Let $\{\alpha_t\}$ be the sequence generated by (6), then it holds:*

$$\alpha_t = \alpha_0 - P\left(\sum_{k=0}^{t-1} \alpha_k \odot e^{-\lambda_k|\alpha_k|}\right) \quad \textit{for } t \ge 1. \tag{7}$$

Proof. To prove the lemma we proceed by induction on t. The case $t = 1$ is trivial. For $t > 1$, by definition we have

$$\alpha_t = G_{\lambda_{t-1}}(\alpha_{t-1}) = \alpha_{t-1} - P\left(\alpha_{t-1} \odot e^{-\lambda_{t-1}|\alpha_{t-1}|}\right).$$

By induction hypothesis the equality (7) is true for $t-1$, implying that

$$\begin{aligned} \alpha_t &= \alpha_0 - P\left(\sum_{k=0}^{t-2} \alpha_k \odot e^{-\lambda_k|\alpha_k|}\right) - P\left(\alpha_{t-1} \odot e^{-\lambda_{t-1}|\alpha_{t-1}|}\right) = \\ &= \alpha_0 - P\left(\sum_{k=0}^{t-1} \alpha_k \odot e^{-\lambda_k|\alpha_k|}\right). \end{aligned}$$

Hence the equality (7) holds for all $t \ge 1$. □

The following result gives a sufficient condition on the sequence of parameters $\{\lambda_t\}$, for the sequence $\{\alpha_t\}$ to be convergent in $\mathcal{A}_{\Phi,s}$.

Proposition 1. *Let $\{\lambda_t\}$ be a real positive sequence. Then, if $\sum_{t=0}^{\infty} 1/\lambda_t < +\infty$, the sequence $\{\alpha_t\}$ generated by the recurrence (6) from any $\alpha_0 \in \mathcal{A}_{\Phi,s}$ converges in $\mathcal{A}_{\Phi,s}$.*

Proof. First of all observe that the scalar-valued mapping

$$x \mapsto |x| e^{-\lambda|x|} \leq \frac{1}{e\lambda}$$

is bounded, which implies

$$\left\| \alpha_k \odot e^{-\lambda_k|\alpha_k|} \right\| \leq \frac{\sqrt{m}}{e} \frac{1}{\lambda_k}. \tag{8}$$

Take two iterates $\alpha_t, \alpha_{t'}$, with $t < t'$; by Lemma 1 their usual distance can be written as

$$\|\alpha_t - \alpha_{t'}\| = \left\| P \sum_{k=t}^{t'-1} \alpha_k \odot e^{-\lambda_k|\alpha_k|} \right\|.$$

Since P is an orthogonal projection, $\|Pv\| \leq \|v\|$ for any $v \in \mathbb{R}^m$, and it holds

$$\begin{aligned} \|\alpha_t - \alpha_{t'}\| &\leq \left\| \sum_{k=t}^{t'-1} \alpha_k \odot e^{-\lambda_k|\alpha_k|} \right\| \\ &\leq \sum_{k=t}^{t'-1} \|\alpha_k \odot e^{-\lambda_k|\alpha_k|}\| \\ &\leq \frac{\sqrt{m}}{e} \sum_{k=t}^{t'-1} \frac{1}{\lambda_k}, \end{aligned}$$

using in the last two steps the triangle inequality and the bound (8). Since by hypothesis the series of terms $1/\lambda_k$ converges, its remainder can be made arbitrarily small, hence for any $\varepsilon > 0$ there always exists an index $N > 0$ such that, for all $t, t' > N$, it holds that

$$\sum_{k=t}^{t'} \frac{1}{\lambda_k} < \frac{e}{\sqrt{m}} \varepsilon.$$

Then $\|\alpha_t - \alpha_{t'}\| < \varepsilon$ for $t, t' > N$, that is $\{\alpha_t\}$ is a Cauchy sequence.

Since the finite-dimensional space $\mathbb{R}^m$ endowed with $\|\cdot\|$ is complete, the Cauchy sequence $\{\alpha_t\}$ converges to an $\alpha_\infty \in \mathbb{R}^m$. The affine set $\mathcal{A}_{\Phi,s}$ is closed in $\mathbb{R}^m$ and hence $\alpha_\infty = \lim \alpha_t$ is also in $\mathcal{A}_{\Phi,s}$. □

A direct consequence of these two results is the following:

Corollary 1. *If $\sum_{t=0}^{\infty} 1/\lambda_t < +\infty$ the sequence $\{\alpha_t\}$ generated by (6) satisfies*

$$P\alpha_t \odot e^{-\lambda_t|\alpha_t|} \longrightarrow \underline{0} \quad \text{as } t \to \infty$$

Proof. Since $\{\alpha_t\}$ converges, by using Lemma 1 there exists the limit

$$\lim_{t\to\infty} P\left(\sum_{k=0}^{t-1} \alpha_k \odot e^{-\lambda_k|\alpha_k|} \right) = \sum_{k=0}^{\infty} P\alpha_k \odot e^{-\lambda_k|\alpha_k|}$$

and hence the general term of the series goes to $\underline{0}$. □

Notice that by eq. (4) this corollary corresponds to the fact that the distance between a t-th term of the sequence $\{\alpha_t\}$ and its iterate through G_{λ_t} can be made arbitrary small, i.e. for some large integer $N = N(\varepsilon)$

$$\|\alpha_t - \alpha_{t+1}\| = \|\alpha_t - G_{\lambda_t}(\alpha_t)\| = \|P\alpha_t \odot e^{-\lambda_t|\alpha_t|}\| < \varepsilon \quad \text{for } t > N$$

which can be viewed as a relaxation of fixed point condition of Remark 1 for α_t.

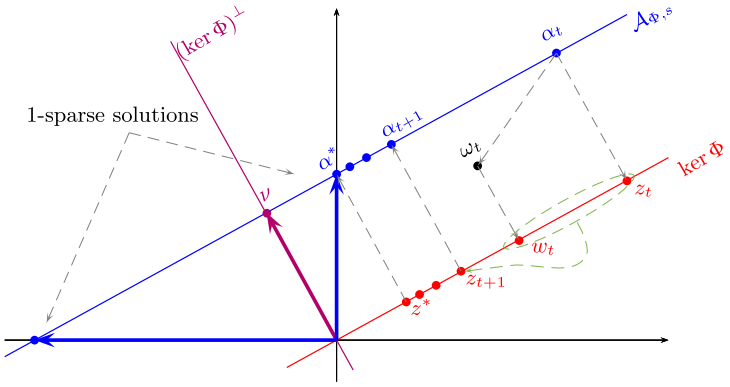

Fig. 2. Convergence process of LIMAPS algorithm.

2.3. *The* LIMAPS *algorithm*

The "mechanics" of the overall iterative process described in previous section may be summed up as follows. Picked an initial point $\alpha_0 \in \mathcal{A}_{\Phi,s} \subset \mathbb{R}^m$, the convergence process so established may be explained as interplay between two alternate actions: shrinkage (by eq. (3)) and projection (by P). Fig. 2 shows graphically the case $\mathcal{A}_{\Phi,s} \subset \mathbb{R}^2$. Fixed a sequence of positive reals $\{\lambda_t\}_{t\geq 0}$, given the point $\alpha_t = z_t + \nu$ with $\nu = \Phi^\dagger s$, a new point α_{t+1} is generated by eq. (6) using the parameter value λ_t. Actually, the projected new point z_{t+1} is computed as the difference between z_t and $w_t = P\omega_t \in \ker\Phi$, with $\omega_t = \alpha_t \odot e^{-\lambda_t|\alpha_t|} \in \mathbb{R}^m$, and then adding the vector $\nu \in (\ker\Phi)^\perp$, i.e., $\alpha_{t+1} = z_{t+1} + \nu$, where $z_{t+1} = z_t - w_t$. Due to shrinkage effects and successive re-projection, the point sequence tends to be closer and closer to the sparse solution $\alpha^* = z^* + \nu$.

Motivated by previous observations and based on the iterative system suggested in the above sections, we can design an algorithm, called LIMAPS, for approximately solving the problem (P_0). In many cases, when the linear system admits a sparse solution of problem (P_0), LIMAPS is able to recover it in very few iterations adopting as parameter scheduling the geometric sequence $\lambda_t = \gamma^t$, with a base γ slightly greater than 1. Still, on the empirical side it should be useful to start the process just from solution ν (marked in Fig. 2) which is in general far from being sparse without favoring any a priori choice of the support. The overall pseudo-code of LIMAPS is reported in Algorithm 1.

Algorithm 1 LIMAPS.

Input: Matrix P, initial solution ν, sequence $\{\lambda_t\}_{t\geq 0}$

1: $\alpha \leftarrow \nu$
2: $t \leftarrow 0$
3: **while** [loop condition] **do**
4: $\quad\lambda \leftarrow \lambda_t$ <*parameter update*>
5: $\quad\alpha \leftarrow \alpha - P\alpha \odot e^{-\lambda|\alpha|}$ <*iteration step*>
6: $\quad t \leftarrow t + 1$ <*step update*>
7: **end while**

Output: an sparse approximate solution $\alpha : \Phi\alpha \approx s$

Remark 2. It is interesting to briefly discuss the sensitivity of the scheme against noise. Suppose to take the model $\Phi\tilde{\alpha} = s + \eta$ where η is an additive noise in the measurement. In this case the iterative scheme would be initialized from $\tilde{\alpha}_0 = \tilde{\nu} = \Phi^\dagger s + \Phi^\dagger\eta$. The sequence $\tilde{\alpha}_t$ generated by scheme (6) from $\tilde{\nu}$ has a distance from noiseless sequence α_t

$$\|\alpha_t - \tilde{\alpha}_t\| \leq 2\frac{\sqrt{m}}{e}\sum_1^t \frac{1}{\lambda_k} + \|\Phi^\dagger\eta\|$$

Since Φ is a frame, it has suitable *frame bounds* $0 < a \leq b < \infty$:

$$a\|x\|^2 \leq \|\Phi x\|^2 \leq b\|x\|^2 \quad \text{for all } x \in \mathbb{R}^m$$

which are in fact the least and the greatest singular value: $a = \sigma_n(\Phi), b = \sigma_1(\Phi)$. The transpose pseudoinverse $(\Phi^\dagger)^T$ is the so-called canonical dual frame, which is still a frame for $\mathbb{R}^n$ with frame bounds $0 < \frac{1}{b} \leq \frac{1}{a} < \infty$ [32, Theor. 5.5]. Therefore

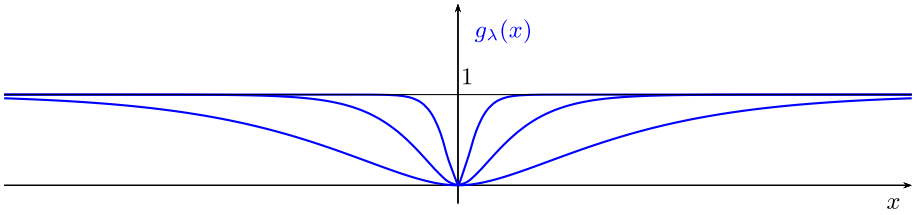

Fig. 3. Graph of parametric functions $g_\lambda(x)$, for $\lambda = 1, 4, 10$.

the sequence distortion introduced by an additive noise η in measurement is $\frac{1}{\sqrt{a}}\|\eta\|$ with a term K constant w.r.t. the noise magnitude:

$$\|\alpha_t - \tilde{\alpha}_t\| \leq \frac{1}{\sqrt{a}}\|\eta\| + K.$$

3. Sparsity minimization

In the context of sparsity promotion techniques, many functionals have been used to measure the sparsity level of a solution, such as those mentioned in the paper [33], where a large class of admissible sparsity measures is introduced and characterized. Following these ideas, in this section we study the sparsity minimization property of the limit points carried out by the iterative scheme (6), and measured by suitable smooth ℓ_0 and ℓ_1-norms.

3.1. Smooth ℓ_0-minimization

For sake of simplicity, in this section we use the notation $\alpha(t)$ to mean the t-th iterate point of an iterative system, $\alpha_k(t)$ to denote its k-th component, and α_k to denote the k-th component of a vector α.

Let us consider the family of functions

$$g_\lambda(x) = 1 - e^{-\lambda|x|}(1 + \lambda|x|)$$

with parameter $\lambda > 0$. It may be useful to look at its graphical representation (Fig. 3) which puts in evidence that $g_\lambda(x)$ is even, increasing for $x \geq 0$ and with image $[0, 1)$. Its behavior strongly depends on the parameter λ since $g_\lambda(x) \approx \lambda^2 x^2$ when $|x| \ll \frac{1}{\lambda}$, while left and right asymptotes are $y = 1$ approached from below. The limit function of the family is the pseudo-norm $\|\cdot\|_0$ in one dimension:

$$\lim_{\lambda\to\infty} g_\lambda(x) = \begin{cases} 0 & \text{if } x = 0 \\ 1 & \text{if } x \neq 0 \end{cases}$$

Moreover, the first two derivatives

$$g'_\lambda(x) = \lambda^2 x e^{-\lambda|x|} \quad \text{and} \quad g''_\lambda(x) = \lambda^2(1 - \lambda|x|)e^{-\lambda|x|} \tag{9}$$

are continuous functions, i.e. $g_\lambda \in C^2(\mathbb{R})$. In order to construct our sparsity functional, we need the triangle inequality for g_λ stated by the following:

Lemma 2 *(Triangle inequality). For every $x, y \in \mathbb{R}$ and $\lambda > 0$ it holds:*

$$g_\lambda(x + y) \leq g_\lambda(x) + g_\lambda(y).$$

Proof. Since $g_\lambda(x) = g_1(\lambda x)$, it is sufficient to study the case of $\lambda = 1$. Let us consider the function

$$\delta(a, b) = e^{-a-b}(1 + a + b) - e^{-a}(1 + a) - e^{-b}(1 + b) + 1$$

defined for $a, b \geq 0$. In $(0, 0)$ it attains its minimum, i.e. it holds: $0 = \delta(0, 0) \leq \delta(a, b)$, that can be rewritten as

$$1 - e^{-a-b}(1 + a + b) \leq 1 - e^{-a}(1 + a) + 1 - e^{-b}(1 + b). \tag{10}$$

This allows us to conclude that

$$\begin{aligned} g_1(x + y) &= 1 - e^{-|x+y|}(1 + |x + y|) \\ &\leq 1 - e^{-|x|-|y|}(1 + |x| + |y|) \\ &\leq 1 - e^{-|x|}(1 + |x|) + 1 - e^{-|y|}(1 + |y|) \qquad \text{(by eq. (10))} \\ &= g_1(x) + g_1(y). \quad \square \end{aligned}$$

Based on g_λ, let us now introduce the functional $J_\lambda : \mathbb{R}^m \to \mathbb{R}_+$ defined as

$$J_\lambda(\alpha) = \sum_{i=1}^{m} g_\lambda(\alpha_i).$$

This $C^2(\mathbb{R}^m)$ functional has "separable" variables and has the tendency as $\|\cdot\|_0$ to promote sparse points. Moreover, analogously to $\|\cdot\|_0$ it is a pseudo-norm as stated in the following:

Lemma 3. *For any $\lambda > 0$, J_λ is a pseudonorm.*

Proof. It is easy to see that $J_\lambda(\alpha) = 0$ if and only if $\alpha = 0$. For all $\alpha, \beta \in \mathbb{R}^m$, the triangle inequality is proved as follows:

$$\begin{aligned} J_\lambda(\alpha+\beta) &= \sum_{k=1}^{m} g_\lambda(\alpha_k+\beta_k) \\ &\le \sum_{k=1}^{m} [g_\lambda(\alpha_k) + g_\lambda(\beta_k)] \qquad \text{(by Lemma 2)} \\ &= J_\lambda(\alpha) + J_\lambda(\beta). \quad \square \end{aligned}$$

The limit functional of J_λ is the ℓ_0-pseudonorm, i.e.

$$\lim_{\lambda\to\infty} J_\lambda(\alpha) = \sum_{k=1}^{m} \lim_{\lambda\to\infty} g_\lambda(\alpha_k) = \|\alpha\|_0.$$

Because of its smoothness, we study J_λ (for large $\lambda > 0$) as a reasonable approximation of ℓ_0-pseudonorm.

3.1.1. Relative minima

On the basis of the stated properties for J_λ, it would seem natural to relax the original problem (P_0) into a problem of the form

$$\min_{\alpha\in\mathbb{R}^m} J_\lambda(\alpha) \quad \text{s.t. } \Phi\alpha = s \tag{P_{J0}}$$

with a large value of λ. We want now to study the minima of this problem. In particular we obtain a necessary condition for the minima in terms of the self-mapping G_λ used in the iterative scheme.

To this end, consider an orthonormal basis $\{a_1, \ldots, a_r\} \subset \mathbb{R}^m$ of the null-space $\ker\Phi$, where $a_k = (a_{k,1}, \ldots, a_{k,m})^T$, $1 \le k \le r = \dim\ker\Phi \le m - n$. Let $\nu = \Phi^\dagger s$; since $\mathcal{A}_{\Phi,s} = \ker\Phi + \nu$, each $\alpha \in \mathcal{A}_{\Phi,s}$ can be represented in unique manner as the linear combination of the basis $\{a_1, ..., a_r\}$ plus ν, i.e. there exists a bijection $\phi = (\phi_1, ..., \phi_m)^T : \mathbb{R}^r \to \mathcal{A}_{\Phi,s}$:

$$\begin{aligned} x &\overset{\phi}{\longmapsto} \alpha = \sum_{k=1}^{r} x_k a_k + \nu \\ \alpha &\overset{\phi^{-1}}{\longmapsto} x = (\langle\alpha, a_1\rangle, ..., \langle\alpha, a_r\rangle)^T. \end{aligned}$$

It is straight-forward to check that the mapping ϕ is a conjugacy with the $m \times r$ Jacobian matrix $\mathbf{D}\phi(x) = [\partial\phi_j(x)/\partial x_k]_{j,k} = [a_{k,j}]_{j,k}$. Using this mapping the problem (P_{J0}) can be reformulated as

$$\min_{\alpha\in\mathcal{A}_{\Phi,s}} J_\lambda(\alpha) = \min_{x\in\mathbb{R}^r} \Gamma_\lambda(x) := J_\lambda(\phi(x)). \tag{P_Γ}$$

Theorem 1. *Let α^* be a local minimum solution of problem (P_{J0}) with a fixed $\lambda > 0$, then*

$$P\alpha^* \odot e^{-\lambda|\alpha^*|} = \underline{0}.$$

It follows that the point $\alpha^* \in \mathcal{A}_{\Phi,s}$ is a fixed point of G_λ.

Proof. To prove the theorem it is thus sufficient to show that given any local minimum solution x of the problem (P_Γ), its corresponding $\alpha = \phi(x)$ satisfies $P\alpha \odot e^{-\lambda|\alpha|} = \underline{0}$. Indeed, if $x = \phi^{-1}(\alpha)$ is a minimum point of Γ_λ, then we have necessarily the first-order condition $\nabla\Gamma_\lambda(x) = \nabla J_\lambda \circ \mathbf{D}\phi(x) = \underline{0}$, i.e. for all $k = 1, ..., r$

$$0 = \frac{\partial}{\partial x_k} \sum_{j=1}^{m} g_\lambda(\alpha_j) = \sum_{j=1}^{m} g'_\lambda(\alpha_j) \frac{\partial\phi_j}{\partial x_k}(x) = \sum_{j=1}^{m} g'_\lambda(\alpha_j) a_{k,j} = \langle\nabla J_\lambda(\alpha), a_k\rangle$$

where $\nabla J_\lambda(\alpha) = (g'_\lambda(\alpha_1), \ldots, g'_\lambda(\alpha_m))^T$. This first-order condition corresponds to zero coordinates w.r.t. the basis $\{a_1, \ldots, a_r\}$ of $\ker \Phi$, i.e. null projection $P\nabla J_\lambda(\alpha) = \underline{0}$.

Extending the scalar case (9), it is easy to calculate the gradient $\nabla J_\lambda(\alpha) = \lambda^2 \alpha \odot e^{-\lambda|\alpha|}$, and hence the necessary condition of the local minimum solution turns out to be

$$P\nabla J_\lambda(\alpha) = \underline{0} \Leftrightarrow P\alpha \odot e^{-\lambda|\alpha|} = \underline{0}$$

since $\lambda > 0$. Moreover, by Remark 1 the last equality is equivalent to $\alpha \in \operatorname{Fix} G_\lambda$. □

This theorem gives necessary conditions, saying that the minimizers are among the fixed points of the scheme with any fixed λ. The converse condition and the relationship to Proposition 1 for time-varying λ are not trivial, as we can notice in the following remarks.

Remark 3. Let $\alpha \in \mathcal{A}_{\Phi,s}$ be any point satisfying the necessary condition $P\alpha \odot e^{-\lambda|\alpha|} = \underline{0}$ of previous theorem and let ϕ be the conjugacy introduced above. In fact, the point α can be either minimum, maximum or saddle point, since $J_\lambda \in C^2(\mathbb{R}^m)$. A sufficient condition for $\alpha = \phi(x)$ to be a local minimum is that the $r \times r$ Hessian matrix $\nabla^2\Gamma_\lambda(x) = (\mathbf{D}\phi(x))^T \nabla^2 J_\lambda(\phi(x)) \mathbf{D}\phi(x) + \sum_{k=1}^r \nabla^2\phi_k(x) \frac{\partial J_\lambda}{\partial \alpha_k}(\phi(x))$ be positive definite. By noticing that $\nabla^2\phi_k$, $k = 1, \ldots, r$, are zero matrices, with a direct calculation we have the general entry of $\nabla^2\Gamma_\lambda(x)$:

$$[\nabla^2\Gamma_\lambda(x)]_{h,k} = \frac{\partial^2}{\partial x_h \partial x_k} \sum_{i=1}^m g_\lambda(\alpha_i) = \sum_{i=1}^m g''_\lambda(\alpha_i) a_{h,i} a_{k,i} = \tag{11}$$

$$= \sum_{i=1}^m \lambda^2 (1 - \lambda|\alpha_i|) e^{-\lambda|\alpha_i|} a_{h,i} a_{k,i}$$

which forms a symmetric matrix for every x. Indeed, showing the positive definiteness of Hessian matrix (11) seems not to be straight-forward.

Remark 4. Proposition 1 provides a sufficient condition for time-varying λ. Moreover, for large λ, its corollary gives a relaxation of the fixed point condition $P\alpha^* \odot e^{-\lambda|\alpha^*|} = \underline{0}$ stated in Theorem 1, i.e. $\|P\alpha_t \odot e^{-\lambda|\alpha_t|}\| < \varepsilon$. Despite this relaxation gets tighter and tighter, it does not guarantee closeness between α_t and a fixed point α^*. However, Proposition 1 does not exclude the convergence for fixed λ, and in fact we empirically observed that the iterative system with fixed λ always converges to a limit point.

Since local minima α^* of the optimization problem (P_{J0}) are fixed points of G_λ, we are interested in studying the behavior of the iterative system G_λ in a neighborhood of α^*. This is the aim of the stability analysis in the next section.

3.1.2. Fixed point stability of minima

In this section we study Lyapunov stability [26] of relative minimum α^* of Theorem 1 which is a fixed point of the iterative system

$$\alpha(t+1) = G_\lambda(\alpha(t)) \tag{12}$$

for a fixed $\lambda > 0$. To this end, it is convenient to introduce the conjugate map $L_\lambda : \mathbb{R}^r \to \mathbb{R}^r$ such that $L_\lambda = \phi^{-1} \circ G_\lambda \circ \phi$, where the conjugacy ϕ is the same as defined in previous section.

Lemma 4. *Let $\nabla^2\Gamma_\lambda$ be the Hessian matrix of mapping Γ_λ and $\alpha^* = \phi(x^*)$ a relative minimum of optimization problem (P_{J0}). If every eigenvalue σ_i, $i = 1, \ldots, r$, of $I - \frac{1}{\lambda^2}\nabla^2\Gamma_\lambda(x^*)$ satisfies $|\sigma_i| < 1$, then α^* is a locally asymptotically stable fixed point of the system G_λ.*

Proof. Instead of (12) we can equivalently study the conjugate iterative system $x(t+1) = L_\lambda(x(t))$. In terms of components $k = 1, \ldots, r$:

$$x_k(t+1) = x_k(t) - \frac{1}{\lambda^2} \sum_{j=1}^m g'_\lambda(\alpha_j) a_{k,j}$$

with $\alpha_j = \phi_j(x) = \sum_{i=1}^r x_i a_{i,j} + \nu_j$.

Let α^* be a relative minimum point and $x^* = \phi^{-1}(\alpha^*)$ be its conjugate point. Then by Theor. 1 the point α^* (resp. x^*) is a fixed point of G_λ (resp. L_λ). The classic way to study the stability of L_λ in the neighborhood of a fixed point x^* is by linearization [26]. By using the substitution $x(t) = x^* + \Delta(t)$ and approximating L_λ with the 1st-order Taylor expansion at x^*, we obtain

$$\Delta_k(t+1) = \Delta_k(t) - \sum_{h=1}^{r} \left[\frac{1}{\lambda^2} \frac{\partial}{\partial x_h} \sum_{j=1}^{m} g'_\lambda(\alpha_j^*) a_{k,j} \right]_{\alpha=\alpha^*} \Delta_h(t). \tag{13}$$

Since

$$\frac{\partial}{\partial x_h} \sum_{j=1}^{m} g'_\lambda(\alpha_j^*) a_{k,j} = \sum_{j=1}^{m} g''_\lambda(\alpha_j^*) a_{k,j} a_{h,j} = (\nabla^2 \Gamma_\lambda(x^*))_{k,h}$$

the iterative system (13) turns out to be

$$\Delta_k(t+1) = \Delta_k(t) - \frac{1}{\lambda_2} \sum_{h=1}^{r} (\nabla^2 \Gamma_\lambda(x^*))_{k,h} \Delta_h(t)$$

or in vector form

$$\Delta(t+1) = \left(I - \frac{1}{\lambda^2} \nabla^2 \Gamma_\lambda(x^*) \right) \Delta(t). \tag{14}$$

From the theory of linear systems it is well-known that if the matrix $I - \frac{1}{\lambda^2}\nabla^2\Gamma_\lambda(x^*)$ has all eigenvalues < 1 in absolute value, then the fixed point x^* is asymptotically stable for the iterative system L_λ in a neighborhood of x^*. In such a case, since G_λ and L_λ are conjugate, the fixed point $\alpha^* = \phi(x^*)$ is also locally asymptotically stable for G_λ. □

Since the eigenvalues σ of the state-transition matrix $I - \frac{1}{\lambda^2}\nabla^2\Gamma_\lambda(x^*)$ are related to the eigenvalues η of $\nabla^2\Gamma_\lambda(x^*)$ by $\sigma = 1 - \frac{\eta}{\lambda^2}$, then we can study the eigenvalues of $\nabla^2\Gamma_\lambda(x^*)$.

Lemma 5. *For every $\alpha = \phi(x) \in \mathcal{A}_{\Phi,s}$, the maximum eigenvalue of $\frac{1}{\lambda^2}\nabla^2\Gamma_\lambda(x)$ is $\eta_{max} \leq 1$.*

Proof. Recall from eq. (11) that $\frac{1}{\lambda^2}(\nabla^2\Gamma_\lambda(x))_{k,h} = \sum_{j=1}^{m} \gamma_j a_{k,j} a_{h,j}$, with $\gamma_j := \frac{1}{\lambda^2} g''_\lambda(\alpha_j)$. With a simple study of the function g''_λ in eq. (9), it is easy to see that $\gamma_j \leq 1$. It is known that $\eta_{\max}$ is the maximum value of Rayleigh quotient:

$$\eta_{\max} = \max_{v \in \mathbb{R}^r : ||v||=1} v^T \frac{1}{\lambda^2} \nabla^2 \Gamma_\lambda(x) v.$$

It results that

$$\begin{aligned} v^T \frac{1}{\lambda^2} \nabla^2 \Gamma_\lambda(x) v &= \sum_{k,h} v_k \left(\sum_{j=1}^{m} \gamma_j a_{k,j} a_{h,j} \right) v_h \\ &= \sum_j \gamma_j \left(\sum_i v_i a_{i,j} \right)^2 \\ &\leq \sum_{j=1}^{m} \left(\sum_{i=1}^{r} v_i a_{i,j} \right)^2. \end{aligned}$$

Let us complete the basis of $\ker \Phi$ to form the orthonormal basis $\{a_1, \ldots, a_r, a_{r+1}, \ldots, a_m\}$ of $\mathbb{R}^m$; the matrix $A := (a_{i,j}) = [a_1 | \cdots | a_m]^T$ is orthogonal and then

$$\begin{aligned} \sum_{j=1}^{m} \left(\sum_{i=1}^{r} v_i a_{i,j} \right)^2 &= \|(v_1, \ldots, v_r, 0, \ldots, 0) A\|^2 \\ &= \|(v_1, \ldots, v_r, 0, \ldots, 0)\|^2 && \text{(by orthogonality of } A\text{)} \\ &= \|v\|^2 = 1 && \text{(by constraint in Rayleigh quotient)} \end{aligned}$$

We conclude that

$$\eta_{\max} = \max_{\|v\|=1} v^T \frac{1}{\lambda^2} \nabla^2 \Gamma_\lambda(x) v \leq 1 \quad \square$$

We are now able to state the main result of this section.

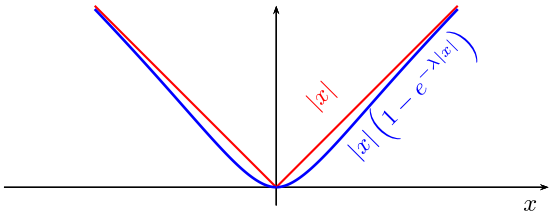

Fig. 4. Graphs of a component of both J_λ-norm (15) and ℓ_1-norm respectively.

Theorem 2. *Let $\alpha^* = \phi(x^*)$ be a relative minimum of (P_{J0}), with positive definite Hessian matrix $\nabla^2\Gamma_\lambda(x^*)$, then α^* is an asymptotically stable fixed point of the iterative system G_λ.*

Proof. Each eigenvalue of $I - \frac{1}{\lambda^2}\nabla^2\Gamma_\lambda(x^*)$ is $\sigma = 1 - \eta$, where η is an eigenvalue of $\frac{1}{\lambda^2}\nabla^2\Gamma_\lambda(x^*)$. By Lemma 5, we know that $\eta \leq 1$; furthermore, since $\nabla^2\Gamma_\lambda(x^*)$ is positive definite, it also holds that $\eta > 0$. Therefore σ is

$$0 \leq \sigma = 1 - \eta < 1.$$

Since each eigenvalue σ of $I - \frac{1}{\lambda^2}H(\alpha^*)$ satisfies $|\sigma| < 1$, by Lemma 4 we can conclude that α^* is an asymptotically stable fixed point of G_λ. □

3.2. Smooth ℓ_1-minimization

Among convex functions a particular interest is deserved to ℓ_1-norm, that became very popular since earlier works [2,3], both focusing on the problem (P_1).

In this spirit, within this section we study a minimization problem based on a functional $J_\lambda : \mathbb{R}^m \to \mathbb{R}_+$, that we call "$J_\lambda$-norm", strictly connected to our shrinkage function (3), given by

$$J_\lambda(\alpha) = \|F_\lambda(\alpha)\|_1 = \sum_{i=1}^{m} |\alpha_i| \left(1 - e^{-\lambda|\alpha_i|}\right), \tag{15}$$

where the sharpness parameter λ is assumed to be positive. It can be checked that $J_\lambda \in C^2(\mathbb{R}^m)$. Its importance comes from the fact that J_λ-norm converges to the ℓ_1-norm as λ approaches $+\infty$, i.e. $\lim_{\lambda\to+\infty} J_\lambda(\alpha) = \|\alpha\|_1$ for any α (see related graphs in Fig. 4). Thanks to this convergence we study by means of Lagrangian multipliers the minimization problem for this J_λ-norm, i.e.

$$\min_{\alpha\in\mathbb{R}^m} J_\lambda(\alpha) \quad \text{subject to } \alpha \in \mathcal{A}_{\Phi,s} \qquad (P_{J1})$$
$$T_\lambda(\alpha) := P\alpha \odot e^{-\lambda|\alpha|} = \underline{0}$$

that can be accomplished by seeking necessary and sufficient condition on the minimizers of its Lagrangian $\mathcal{L}_\lambda : \mathbb{R}^{3m} \to \mathbb{R}$. Notice that the second constraint is the fixed point condition $\alpha \in \operatorname{Fix} G_\lambda$ as in Remark 1, while first constraint can be turned into $Q\alpha - \nu = \underline{0}$, where $Q = I - P = \Phi^\dagger\Phi$ and $\nu = \Phi^\dagger s$. Hence, the Lagrangian takes the form:

$$\mathcal{L}(\alpha, \theta, \rho) = J_\lambda(\alpha) + \theta^T T_\lambda(\alpha) + \rho^T(Q\alpha - \nu). \tag{16}$$

Let α^* be a local optimum solution of (P_{J1}), and suppose that the Jacobian matrix of equality constraints has full rank. This guarantees the first order necessary conditions, that is there exist some multipliers $\theta, \rho \in \mathbb{R}^m$ such that $\nabla_\theta\mathcal{L}_\lambda(\alpha^*, \theta, \rho) = \underline{0}$, $\nabla_\rho\mathcal{L}_\lambda(\alpha^*, \theta, \rho) = \underline{0}$ and for $\alpha = \alpha^*$

$$\begin{aligned}\nabla_\alpha\mathcal{L}_\lambda(\alpha, \theta, \rho) &= \nabla J_\lambda(\alpha) + \mathbf{D}T_\lambda(\alpha)^T\theta + Q\rho \\ &= \mu_{\lambda,\alpha} + D_{\lambda,\alpha}P\theta + Q\rho = \underline{0}\end{aligned} \tag{17}$$

where $\mu_{\lambda,\alpha} := \nabla J_\lambda(\alpha)$ and $D_{\lambda,\alpha} := \mathbf{D}T_\lambda(\alpha)^T$ is the transposed Jacobian matrix of T_λ at α. Note that both $\mu_{\lambda,\alpha} \in \mathbb{R}^m$ and $D_{\lambda,\alpha} \in \mathbb{R}^{m\times m}$ have separable variables and moreover $D_{\lambda,\alpha}$ is diagonal. Their entries are

$$[\mu_{\lambda,\alpha}]_i = \operatorname{sgn}(\alpha_i)(1 + (\lambda|\alpha_i| - 1)e^{-\lambda|\alpha_i|}) \quad \text{and} \quad [D_{\lambda,\alpha}]_{ii} = (1 - \lambda|\alpha_i|)e^{-\lambda|\alpha_i|},$$

and therefore $\mathcal{L} \in C^1$. Let us denote $\mu_{\lambda,\alpha^*} = \mu_{\lambda,\alpha}|_{\alpha=\alpha^*}$ and $D_{\lambda,\alpha^*} = D_{\lambda,\alpha}|_{\alpha=\alpha^*}$. If $\alpha_i^* \neq \pm 1/\lambda$ for all $i = 1, \ldots, m$, the matrix D_{λ,α^*} has full rank m. Sylvester's Rank Inequality implies that $\operatorname{rank}(D_{\lambda,\alpha^*}P) \geq \operatorname{rank}(D_{\lambda,\alpha^*}) + \operatorname{rank}(P) - m = m + (m-n) - m = m - n$. The first order condition (17) evaluated at the local optimum $\alpha = \alpha^*$ consists in the underdetermined linear system of m equations in $2m$ unknowns (θ, ρ)

$$D_{\lambda,\alpha^*} P\theta + Q\rho = -\mu_{\lambda,\alpha^*}. \tag{18}$$

The coefficients matrix of this linear system is $[D_{\lambda,\alpha^*} P \mid Q]$, which is the Jacobian matrix of the constraints. Since $\operatorname{rank}(Q) = n$, we expect that in non-degenerate case this Jacobian matrix has full rank $(m-n)+n = m$, which hence is a reasonable assumption. In this case the system (18) admits ∞^m many solutions (θ^*, ρ^*).

Remark 5. We can in fact explicitly express the regularity condition: let α^* be a local optimum solution of the problem (P_{J1}), then

$$\alpha^* \text{ is regular point} \Leftrightarrow [D_{\lambda,\alpha^*} P \mid Q] \text{ has full rank.}$$

In such a case $\alpha_i^* \neq \pm 1/\lambda$, for all $i = 1, \ldots, m$.

Sufficient conditions for an extremal point can be studied by a close glance at the Hessian matrix $\nabla^2 \mathcal{L}_\lambda(\alpha, \theta, \rho)$ evaluated at the stationary point $(\alpha^*, \theta^*, \rho^*)$.

Lemma 6. *Let α^* be a regular stationary point of the Lagrangian $\mathcal{L}$ associated to problem (P_{J1}), with the respective multipliers (θ^*, ρ^*), and denote $S := \operatorname{supp}(\alpha^*)$. Then the Hessian matrix of $\mathcal{L}$ w.r.t. variables α at $(\alpha^*, \theta^*, \rho^*)$ is diagonal and has the diagonal asymptotic expressions:*

$$[\nabla^2_{\alpha\alpha} \mathcal{L}_\lambda(\alpha^*, \theta^*, \rho^*)]_{ii} \sim \begin{cases} 2\lambda, & \text{if } i \notin S \\ \lambda(1+\gamma_i^*), & \text{if } i \in S, \end{cases}$$

for $\lambda \to \infty$.

Proof. It is straight-forward to see that the Hessian matrix takes the form

$$\nabla^2_{\alpha\alpha} \mathcal{L}_\lambda(\alpha^*, \theta^*, \rho^*) = M_{\lambda,\alpha^*\alpha^*} + \operatorname{diag}(D_{\lambda,\alpha^*\alpha^*} P\theta^*),$$

where $\operatorname{diag}(v)$ denotes the diagonal matrix having vector v as diagonal, $M_{\lambda,\alpha^*\alpha^*}$ and $D_{\lambda,\alpha^*\alpha^*}$ are diagonal matrices with entries

$$[M_{\lambda,\alpha^*\alpha^*}]_{ii} = \lambda(2 - \lambda|\alpha_i^*|)e^{-\lambda|\alpha_i^*|} \quad \text{and} \quad [D_{\lambda,\alpha^*\alpha^*}]_{ii} = \lambda(\lambda\alpha_i^* - 2\operatorname{sgn}(\alpha_i^*))e^{-\lambda|\alpha_i^*|}$$

and therefore $\nabla^2_{\alpha\alpha} \mathcal{L}_\lambda(\alpha^*, \theta^*, \rho^*)$ is a diagonal matrix.

Since α^* is regular, $\alpha_i^* \neq \pm 1/\lambda$ by Remark 5 and hence D_{λ,α^*} is non-singular; therefore we can rewrite eq. (18) for $(P\theta^*)$ and plug it into the expression of the Hessian matrix $\nabla^2_{\alpha\alpha} \mathcal{L}_\lambda(\alpha^*, \theta^*, \rho^*)$. The diagonal element is

$$\begin{aligned} [\nabla^2_{\alpha\alpha} \mathcal{L}_\lambda(\alpha^*, \theta^*, \rho^*)]_{ii} &= [M_{\lambda,\alpha^*\alpha^*} - \operatorname{diag}(D_{\lambda,\alpha^*\alpha^*} D^{-1}_{\lambda,\alpha^*} (\mu_{\lambda,\alpha^*} + Q\rho^*))]_{ii} \\ &= \lambda(2 - \lambda|\alpha_i^*|)e^{-\lambda|\alpha_i^*|} + \\ &\quad + \lambda(\lambda\alpha_i^* - 2\operatorname{sgn}(\alpha_i^*)) \frac{\operatorname{sgn}(\alpha_i^*)(1 + (\lambda|\alpha_i^*| - 1)e^{-\lambda|\alpha_i^*|} + \gamma_i^*)}{\lambda|\alpha_i^*| - 1}, \end{aligned}$$

where γ_i^* is the i-th entry of vector $\gamma^* := Q\rho^*$. Now, it is easy to check the asymptotic expressions

$$[\nabla^2_{\alpha\alpha} \mathcal{L}_\lambda(\alpha^*, \theta^*, \rho^*)]_{ii} \sim \begin{cases} 2\lambda, & \text{if } \alpha_i^* = 0 \\ \lambda(1+\gamma_i^*), & \text{if } \alpha_i^* \neq 0, \end{cases} \tag{19}$$

as $\lambda \to \infty$. □

We hence can easily derive sufficient conditions for positive definiteness of the Hessian matrix $\nabla^2_{\alpha\alpha} \mathcal{L}_\lambda(\alpha^*, \theta^*, \rho^*)$ that we summarize in the following

Proposition 2. *Let α^*, θ^*, ρ^* and S be defined as in Lemma 6. If $\gamma_i^* = [Q\rho^*]_i > -1$ for $i \in S$, for every sufficiently large $\lambda > 0$ the matrix $\nabla^2_{\alpha\alpha} \mathcal{L}_\lambda(\alpha^*, \theta^*, \rho^*)$ is positive definite.*

Note that in general $\operatorname{spark}(Q) = \operatorname{spark}(\Phi^\dagger \Phi) > k$ is a sufficient condition for having an $x \in \mathbb{R}^m$ s.t. $[Qx]_i > -1, i \in \{i_1, i_2, \ldots, i_k\}$. We can now state the sufficient conditions for minimal point of problem (P_{J1}).

Theorem 3. *Let α^*, θ^*, ρ^* and S be defined as in Lemma 6. If $\gamma_i^* = [Q\rho^*]_i > -1$ for $i \in S$, then α^* is a strict local minimum point of problem (P_{J1}) for sufficiently large $\lambda > 0$.*

Proof. It can be checked that the Hessian matrix of the Lagrangian function takes the block form

$$\nabla^2 \mathcal{L}_\lambda(\alpha^*, \theta^*, \rho^*) = \begin{pmatrix} \nabla^2_{\alpha\alpha} \mathcal{L}_\lambda(\alpha^*, \theta^*, \rho^*) & P D_{\lambda,\alpha^*} & Q \\ D_{\lambda,\alpha^*} P & 0 & 0 \\ Q & 0 & 0 \end{pmatrix}$$

Let us study the asymptotic behavior of the Rayleigh quotient $z^T \nabla^2 \mathcal{L}_\lambda(\alpha^*, \theta^*, \rho^*) z$ with $\|z\| = 1$:

$$z^T \nabla^2 \mathcal{L}_\lambda(\alpha^*, \theta^*, \rho^*) z = u^T \nabla^2_{\alpha\alpha} \mathcal{L}_\lambda(\alpha^*, \theta^*, \rho^*) u + 2u^T P D_{\lambda,\alpha^*} v + 2u^T Q w \tag{20}$$

where $(u, v, w) = z$. Since $[D_{\lambda,\alpha^*}]_{ii}$ goes to either 0 or 1 as $\lambda \to \infty$, the second and third terms are bounded. Since λ is sufficiently large, by Proposition 2 the first term is positive definite, and in particular is asymptotic to

$$\lambda \sum_{i \in S} (1 + \gamma_i^*) u_i^2 + 2\lambda \sum_{i \notin S} u_i^2 \geq \lambda \left(\varepsilon \sum_{i \in S} u_i^2 + 2 \sum_{i \notin S} u_i^2 \right)$$

as $\lambda \to \infty$, where $\varepsilon := \min_{i \in S} \gamma_i^* + 1$. Hence eq. (20) is null iff $u = \underline{0}$, and in such a case the Rayleigh quotient is also null. Otherwise, since the two remaining terms are bounded, the Rayleigh quotient is asymptotically proportional to λ.

Therefore, since the quadratic form is null if and only if the increasing direction from α^* is $u = \underline{0}$, the point α^* is local minimum satisfying the constraints, i.e. a local solution of (P_{J1}) for sufficiently large $\lambda > 0$. □

Given previous results, we can draw some conclusions about the minima so characterized.

- The J_λ-norm in (P_{J1}) is a non-convex functional but it uniformly converges to the convex ℓ_1-norm of (P_1), as $\lambda \to \infty$. Indeed, we have

$$\lim_{\lambda \to \infty} \left(\sup_\alpha |J_\lambda(\alpha) - \|\alpha\|_1| \right) = \lim_{\lambda \to \infty} \left(\frac{m}{e\lambda} \right) = 0.$$

  Moreover, if $\alpha_t \to \alpha_\infty \in \mathcal{A}_{\Phi,s}$ and $\lambda_t \to \infty$ as $t \to \infty$, then

$$\lim_{t \to \infty} J_{\lambda_t}(\alpha_t) = \lim_{t \to \infty} \|F_{\lambda_t}(\alpha_t)\|_1 = \|\alpha_\infty\|_1.$$

- The operator T_λ in (P_{J1}) also converges uniformly to the zero operator:

$$\lim_{\lambda \to \infty} \left(\sup_\alpha \|P\alpha \odot e^{-\lambda|\alpha|} - \underline{0}\| \right) = \lim_{\lambda \to \infty} \left(\frac{\sqrt{m}}{e\lambda} \right) = 0$$

  Hence, the second constraint of problem (P_{J1}) in a certain sense vanishes as $\lambda \to \infty$.

Summarizing, these two convergence properties of the objective function and the constraint suggest some evidence that problem (P_{J1}) is a reasonable relaxation of problem (P_1).

4. Performances of LiMapS

In this section we present the main aspects about the computational complexity of the LiMapS algorithm and some numerical simulations which compare this method with other sparse recovery methods, highlighting the phenomenon of phase transition.

4.1. Computational complexity

As far as the computational complexity is concerned, we report here a brief analysis related to the number of arithmetic operations needed to carry out the approximate solution, once a stop criterion is fixed. In particular, if we use as usual measure of convergence for a sequence $\{\alpha_t\}_{t \geq 0}$ the distance between two consecutive iterates, i.e. $d(t) = \|\alpha_{t+1} - \alpha_t\|$, it is easy to exhibit an upper bound to this discrepancy at each step t. In fact, by eqns. (4) and (6) we have

$$d(t) = \|\alpha_{t+1} - \alpha_t\| = \|P\alpha_t \odot e^{-\lambda_t |\alpha_t|}\| \leq \|\alpha_t \odot e^{-\lambda_t |\alpha_t|}\| \leq \frac{\sqrt{m}}{e\lambda_t}$$

where last inequality is given by eq. (8).

We study the case of parameter scheduling $\lambda_t = \gamma^t$, with fixed $\gamma > 1$. In order to guarantee $d(t) \leq \epsilon$, for an arbitrary error $\varepsilon > 0$, it is sufficient to iterate $t \geq (\frac{1}{2} \log m + \log \varepsilon^{-1} + \log e^{-1}) / \log \gamma$ steps of the while loop in the Algorithm 1. Assuming floating-point arithmetic, numerical exponentiation $\exp(\cdot)$ through the classical AGM technique [34, sec. 4.8] with guaranteed B-bit precision requires $O(\log B)$ time. Therefore, $O(m^2 + m \log B)$ is the time for one iteration of line 5, that dominates algorithm's computational time. Hence, the overall time complexity of Algorithm 1 is $O((\log m + \log \varepsilon^{-1})(m^2 + m \log B))$. In

case of floating-point arithmetic with fixed $B = 64$ bits, it reduces to $O(m^2(\log m + \log \varepsilon^{-1}))$. The space complexity of the algorithm is simply dominated by that needed for storing the given projector matrix P, values with precisions ε or δ, and hence is asymptotically the same as time complexity.

To gain a little speed-up, since $P = I - \Phi^\dagger \Phi$, we could pre-compute once the matrix $\Phi^\dagger \in \mathbb{R}^{m \times n}$ using a thin SVD (up to $\text{rank}(\Phi) = n$) with a cost of $O(mn^2)$ time. Indeed, line 5 can be replaced with

$$5': \quad \alpha \leftarrow \alpha - \alpha \odot e^{-\lambda|\alpha|} + \Phi^\dagger(\Phi\alpha \odot e^{-\lambda|\alpha|})$$

where the vector $\alpha \odot e^{-\lambda|\alpha|}$ can be computed once each iteration. In line 5′ the most expensive operations are two matrix–vector multiplications requiring $O(nm)$ time. Therefore, forming the full matrix P is unnecessary and, assuming pre-computation of $\Phi^\dagger$, the overall time complexity of the algorithm reduces to $O((\log m + \log \varepsilon^{-1})(nm + m \log B))$, or $O(nm(\log m + \log \varepsilon^{-1}))$ if $B = 64$ bits are fixed for the arithmetic.

4.2. Phase transition in numerical simulations

In order to quantitatively evaluate LiMapS, and to compare it with well-known sparse recovery methods in literature, we adopt the experimental analysis proposed in [35]. Specifically, Donoho and Tunner stated that, assuming the solution to (P_0) is k-sparse, and the dimensions (k, n, m) of the linear problem are large, many sparse recovery algorithms exhibit the phenomenon of phase transition.

According to this analysis, using randomly generated instances of the matrix Φ and k-sparse vector α^*, we build instances (Φ, s) of (P_0) such that $\Phi\alpha^* = s$. We experimentally show that both LiMapS algorithm and the other methods considered exhibit a phase transition by measuring the Signal-to-Noise-Ratio between α^* and the recoverd solution α, i.e., $\text{SNR} = 20 \log_{10} \|\alpha\| / \|\alpha - \alpha^*\|$. In particular, the elements of atoms collected in matrix Φ are i.i.d. random variables drawn from standard Gaussian distribution, while sparse coefficients α^* are randomly generated by the so-called Bernoulli-Gaussian model. Let $\omega = (\omega_1, \ldots, \omega_m)$ be a vector of i.i.d. standard Gaussian variables and $\theta = (\theta_1, \ldots, \theta_m)$ be a vector of i.i.d. Bernoulli variables with parameter $0 < \rho \leq 0.5$. The Bernoulli-Gaussian vector $\alpha^* = (\alpha_1^*, \ldots, \alpha_m^*)$ is then given by $\alpha_i^* = \theta_i \cdot \omega_i$, for all $i = 1, \ldots, m$. Regarding the instance size, we fix $n = 100$, and we let the sparsity level k and the number of unknowns m range in the intervals $[1, 50]$ and $[101, 1000]$ respectively. The SNR is achieved by averaging on 100 randomly generated trials for every $\delta = \frac{n}{m}$ and $\rho = \frac{k}{n}$, that are the normalized measure of problem indeterminacy and the normalized measure of the sparsity, respectively.

In Fig. 5 we report the 3D phase transitions of LiMapS and other sparsity solvers. Specifically, we refer to both ℓ_0-norm methods such as OMP [7], CoSaMP [6], and SL0 [8], and to ℓ_1-norm methods LASSO [10] and BP [2,3]. The figure clearly shows the existence of a sharp phase transition that partitions the phase space into a *unrecoverable* region, with vanishing exact-recovery probability, from an *recoverable* region in which the probability to recover the sparse vector α^* will eventually approach to one. This former case corresponds to high normalized sparsity and high problem indeterminacy. Qualitatively, it is evident that LiMapS algorithm reaches the best results, having the largest area of high recoverability. A quantitative evaluation is given by the volume V under the surface by summing up the SNRs of each method in correspondence of the discrete mesh in the δ–ρ plane. These measures, normalized to that of LiMapS, are reported in Fig. 5, next to the method name.

Our simulations were performed on AMD Athlon II X4 630 64-bit 2.8 GHz processor with 4 GB of memory, using publicly available MATLAB implementation of the algorithms[2]. The algorithm LiMapS is available online at the URL http://phuselab.di.unimi.it/limaps.html.

5. Conclusions

This article provides the theoretical justification of a recently proposed sparsity recovery method, namely LiMapS [22], that attempts to find the sparse decomposition of a given signal iteratively by applying two suitable Lipschitzian maps aiming at achieving sparsity and feasibility. In particular, it is shown that LiMapS works well adopting shrinkage functions having small Lipschitz constant not bound to the sequence of the control parameter λ driving the search. Furthermore, it is proved that, for fixed λ, minima of smooth versions of ℓ_0 and ℓ_1-minimization problems are among the fixed points of the derived iterative scheme.

From the empirical viewpoint, we showed by numerical simulations on synthetic data the effectiveness of the proposed method with respect to other well-known ones in literature in terms of phase transition.

Further theoretical studies should be addressed to apply classical tools for analyzing the performance of recovery algorithms, such as the restricted isometry properties (RIP), the null space properties (NSP) or the mutual coherence that measures the maximum correlation between any two distinct columns of the dictionary. Future works should also be devoted to the definition of general conditions to impose on the scheduling of the fundamental parameter λ which drives the search toward optimal solutions. Moreover in this regard, since we established necessary conditions for the minimal solutions of problem (P_{J0}), one of the goals in future works consists in extending these intermediate results, i.e. finding sufficient conditions for the minimality of fixed points yielded by the scheme.

[2] SparseLab from Stanford University at http://web.stanford.edu/group/sparselab, SL0 from http://ee.sharif.edu/~SLzero, and CoSaMP from http://mathworks.com/matlabcentral.

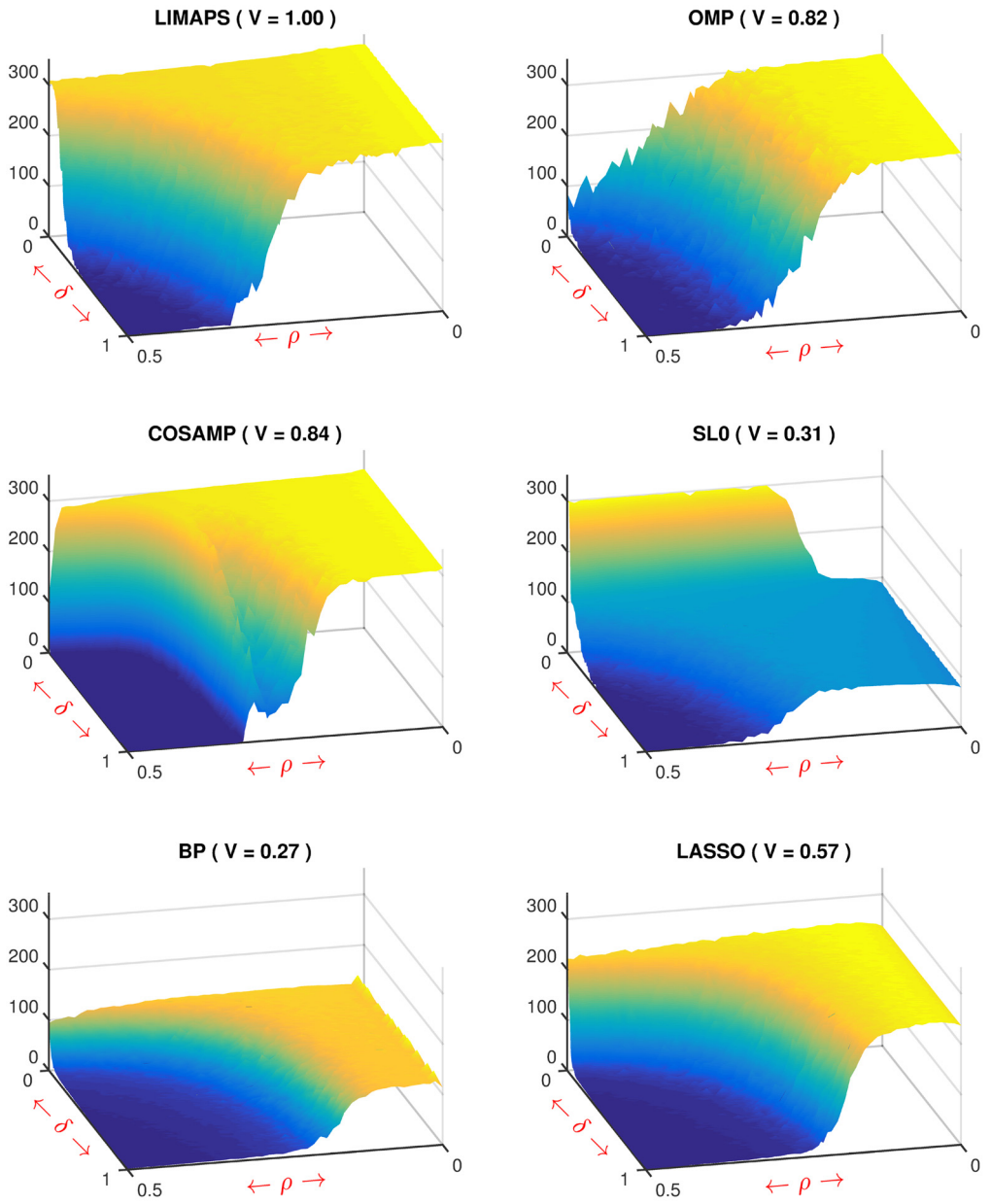

Fig. 5. SNR of phase transitions of both ℓ_0-minimizers (first two rows) and ℓ_1-minimizers (third row) methods. The domain is defined by $(\delta, \rho) \in [0, 1] \times [0, 0.5]$. Next to the method name, V represents the volume under the surface normalized to that of LiMapS.

Two key open issues remain to study. On the one hand, proving the convergence in case of fixed λ, which is a very difficult task but would be of great help in finding relationships between LiMapSand the minimizers of standard problems relying on ℓ_0 and ℓ_1 norms. On the other hand, a step ahead would be given by a clear characterization of the local minima of both smooth ℓ_0 and ℓ_1 minimization problems here defined, and by specifying the way they relate to the exact versions, namely for (P_0) and (P_1) problems. In particular, for theoretically estimating the distance between minimizers of the smooth and exact problems, it might be useful to restrict the analysis to k-sparse points. In general, difficulties are also due to the strong non-linearity in the mapping used here for shrinkage effect.

Acknowledgements

We are grateful to Alberto Bertoni for the useful suggestions and sharp intuitions he gave us during this study. We also thank the anonymous referees for helping us to improve the manuscript.

References

[1] D.L. Donoho, For most large underdetermined systems of linear equations the minimal ℓ^1-norm solution is also the sparsest solution, Comm. Pure Appl. Math. 59 (2004) 797–829.
[2] S.S. Chen, D.L. Donoho, Michael A. Saunders, Atomic decomposition by basis pursuit, SIAM J. Sci. Comput. 20 (1) (1998) 33–61.
[3] E. Candès, T. Tao, Decoding by linear programming, IEEE Trans. Inform. Theory 52 (2006) 5406–5425.
[4] B.K. Natarajan, Sparse approximate solutions to linear systems, SIAM J. Comput. 24 (1995) 227–234.
[5] S. Mallat, Z. Zhang, Matching pursuit with time-frequency dictionaries, IEEE Trans. Signal Process. 41 (1993) 3397–3415.
[6] D. Needell, J.A. Tropp, Cosamp: iterative signal recovery from incomplete and inaccurate samples, Appl. Comput. Harmon. Anal. 26 (3) (2009) 301–321.
[7] J. Tropp, A. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, IEEE Trans. Inform. Theory 53 (12) (2007) 4655–4666.
[8] H. Mohimani, M. Babaie-Zadeh, C. Jutten, A fast approach for overcomplete sparse decomposition based on smoothed ℓ^0 norm, IEEE Trans. Signal Process. 57 (1) (2009) 289–301.
[9] M. Elad, Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing, Springer, 2010.
[10] B.E. Trevor, T. Hastie, I. Johnstone, R. Tibshirani, Least angle regression, Ann. Stat. 32 (2002) 407–499.
[11] E.T. Hale, W. Yin, Y. Zhang, Fixed-point continuation for l1-minimization: methodology and convergence, SIAM J. Optim. 19 (3) (2008) 1107–1130.
[12] W. Yin, S. Osher, D. Goldfarb, J. Darbon, Bregman iterative algorithms for l1-minimization with applications to compressed sensing, SIAM J. Imaging Sci. 1 (1) (2008) 143–168.
[13] A.M. Bruckstein, D.L. Donoho, M. Elad, From sparse solutions of systems of equations to sparse modeling of signals and images, SIAM Rev. 51 (1) (2009) 34–81.
[14] J. Tropp, S. Wright, Computational methods for sparse solution of linear inverse problems, Proc. IEEE 98 (6) (2010) 948–958.
[15] Y. Zhao, D. Li, Computational methods for sparse solution of linear inverse problems, SIAM J. Optim. 22 (3) (2012) 1065–1088.
[16] D. Donoho, M. Elad, Optimality sparse representation in general (non-orthogonal) dictionaries via l1 minimization, Proc. Natl. Acad. Sci. USA 100 (2003) 2197–2202.
[17] M. Elad, A. Bruckstein, A generalized uncertainty principle and sparse representation in pairs of bases, IEEE Trans. Inform. Theory 48 (2002) 2558–2567.
[18] E. Candès, J. Romberg, T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, IEEE Trans. Inform. Theory 52 (2006) 489–509.
[19] E. Candès, J. Romberg, T. Tao, Stable signal recovery from incomplete and inaccurate measurements, Comm. Pure Appl. Math. 59 (8) (2006) 1207–1223.
[20] E. Candès, Compressive sampling, in: Proc. Int. Congr. of Mathematicians, vol. III, Madrid, Spain, 2006, pp. 1433–1452.
[21] A. Cohen, W. Dahmen, R. Devore, Compressed sensing and best k-term approximation, J. Amer. Math. Soc. 22 (2009) 211–231.
[22] A. Adamo, G. Grossi, A fixed-point iterative schema for error minimization in k-sparse decomposition, in: 2011 IEEE Int. Symp. on Signal Processing and Information Technology, ISSPIT, 2011, pp. 167–172.
[23] A. Adamo, G. Grossi, R. Lanzarotti, Local features and sparse representation for face recognition with partial occlusions, in: 2013 IEEE Int. Conf. on Image Processing, ICIP, 2013, pp. 3008–3012.
[24] A. Adamo, G. Grossi, R. Lanzarotti, J. Lin, ECG compression retaining the best natural basis k-coefficients via sparse decomposition, Biomed. Signal Process. Control 15 (2015) 11–17.
[25] Y. Benyamini, J. Lindenstrauss, Geometric Nonlinear Functional Analysis, vol. 1, Amer. Math. Soc. Colloq. Publ., vol. 48, American Mathematical Soc., 1998.
[26] S. Wiggins, Introduction to Applied Nonlinear Dynamical Systems and Chaos, 2nd edition, Springer, 2003.
[27] P.-Y. Chen, I.W. Selesnick, Translation-invariant shrinkage/thresholding of group sparse signals, Signal Process. 94 (2014) 476–489.
[28] R. Chartrand, Shrinkage mappings and their induced penalty functions, in: IEEE Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP 2014, Florence, Italy, May 4–9, 2014, IEEE, 2014, pp. 1026–1029.
[29] D.L. Donoho, I.M. Johnstone, Ideal spatial adaptation by wavelet shrinkage, Biometrika 81 (3) (1994) 425–455.
[30] S. Banach, Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales, Fund. Math. 3 (1) (1922) 133–181.
[31] D.L. Donoho, De-noising by soft-thresholding, IEEE Trans. Inform. Theory 41 (3) (1995) 613–627.
[32] S. Mallat, A Wavelet Tour of Signal Processing: The Sparse Way, Academic Press, 2008.
[33] R. Gribonval, M. Nielsen, Highly sparse representations from dictionaries are unique and independent of the sparseness measure, Appl. Comput. Harmon. Anal. 22 (3) (2007) 335–355.
[34] R. Brent, P. Zimmermann, Modern Computer Arithmetic, Cambridge University Press, New York, NY, USA, 2010.
[35] D.L. Donoho, J. Tunner, Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing, Philos. Trans. A Math. Phys. Eng. Sci. 367 (1906) (2009) 4273–4293.