Sparsity Recovery by Iterative Orthogonal Projections of Nonlinear Mappings

Alessandro Adamo
Dipartimento di Matematica
Università degli Studi di Milano
Via Saldini 50, I-20133 Milan, Italy
Email: alessandro.adamo@unimi.it

Giuliano Grossi
Dipartimento di Scienze dell'Informazione
Università degli Studi di Milano
Via Comelico 39, I-20135 Milan, Italy
Email: grossi@dsi.unimi.it

***Abstract*—This paper provides a new regularization method for sparse representation based on a fixed-point iteration schema which combines two Lipschitzian-type mappings, a nonlinear one aimed to uniformly enhance the sparseness level of a candidate solution and a linear one which projects back into the feasible space of solutions. It is shown that this strategy locally minimizes a problem whose objective function falls into the class of the ℓ^p-norm and represents an efficient approximation of the intractable problem focusing on the ℓ^0-norm. Numerical experiments on randomly generated signals using classical stochastic models show better performances of the proposed technique with respect to a wide collection of well known algorithms for sparse representation.**

I. Introduction

Finding sparse solutions of underdetermined systems of linear equations is a topic extensively studied in signal processing [1], [2]. Even though one is dealing with systems having infinitely many solutions, it is shown that sufficiently sparse solutions may be identified uniquely [3].

In this paper we propose a new algorithm to promote high level of sparsity within underdetermined system solutions in which Lipschitzian mappings and fixed-point iteration schemes [4] play a fundamental role. In particular, we introduce the following two kinds of Lipschitzian maps. An asymptotically nonexpansive [5] parametric family of nonlinear functions able to select near-feasible solutions having the sparsity property. A nonexpansive linear map consisting on orthogonal projection of near-feasible points along the space of the affine solutions of an inhomogeneous linear system. The main purpose of this interaction is that of reducing the gap existing between the two image spaces, being the first a subspace containing sparsest coefficients (in the sense of the ℓ^0-norm which measures the total number of nonzero elements) while the second is a set containing infinite solutions for the problem at hand.

By composing the two mappings we are faced with nonlinear problems in Banach space for which we show the existence of fixed-points. Moreover we prove they may be obtained as the limit of a fixed-point iteration scheme defined by repeated images under the mapping of an arbitrary starting point in the space.

A critical point in our method is to provide right values for the continuous parameter characterizing the nonlinear mapping on which depends the sparsity level gained at each iteration as well as at the end of the entire process of sparsification. We provide a scheduling policy of incremental values in order to have a good chance to converge to the sparsest solution.

To show the performances of the proposed method we executed a lot of numerical experiments choosing signals and frames of various size whose elements are all independently drawn using stochastic models frequently employed in these analysis as, for instance, the Gaussian-Bernoulli model. These tests have the aim of comparing our algorithm with a plethora of other algorithms representing the state-of-the-art of the strategies used in this field.

By expressing the algorithm performances in terms of signal-to-noise-ratio SNR and computation time, we found that the proposed strategy outperforms the other algorithms in almost all experiments conducted, carrying out the sparsest solutions with arbitrary precision error and achieving a comparable computation time with the fastest one.

II. Sparse Solutions to Undetermined Systems

Let $s = (s_1, \ldots, s_n)$ be a discrete-time signal of length n, i.e., a vector in $\mathbb{R}^n$, and $\Phi = [\phi_1, \ldots, \phi_m]$ be a collection of m basic waveforms or vectors in $\mathbb{R}^n$, usually called atoms of the dictionary Φ. Assuming $m > n$, the dictionary will result in an overcomplete frame leading to infinite solutions of the underdetermined linear system:

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

where $\alpha \in \mathbb{R}^m$ represents the coefficients of the atoms in Φ. This system has more unknowns than equations, and thus it has either no solutions, if s is not in the span of the columns of the matrix Φ, or infinitely many solutions. In order to avoid the anomaly of having no solutions, we shall hereafter assume that Φ is a full-rank matrix, implying that its columns span the entire space $\mathbb{R}^n$.

The general goal would be to find a highly sparse decomposition of the signal s, that is, one with very few nonzero terms in α. This can be rephrased into the following NP-hard combinatorial optimization problem:

$$\min_{\alpha} \| \alpha \|_0 \quad \text{subject to} \quad \Phi\alpha = s, \tag{P0}$$

where $\| \alpha \|_0 = |\{k : \alpha_k \neq 0\}|$ denotes the ℓ^0-norm. The set $\mathcal{S}(\alpha) = \{k : \alpha_k \neq 0\}$ is also called support of α. Moreover, a

978-1-4673-0753-6/11/$26.00 ©2011 IEEE

signal s is called k-sparse if it admits a representation which combines at most k atoms of the dictionary Φ.

Owing to the intractability of problem (P0) [6], many researchers have developed alternatives to the exhaustive search which consider in many cases a relaxation to the ℓ^1-norm[1] allowing to recast (P0) as convex optimization solvable by linear programming [3].

A different approach has been developed in [7]. The main idea is to approximate the ℓ^0-norm by a smoothing function and then following a graduated non convexity [8] approach to avoid to get trapped into local minima. At the end of each step along the gradient ascent of a suitable objective function, the approximate solution found is projected back into the feasible region.

The approach followed in this paper represents an alternative minimizer to that proposed in [7]. It is based on a suitable sparseness measure which represents a sort of measure falling in between the ℓ^0 and ℓ^1 norms [9].

III. Sparsity Promotion Mappings

The aim of this section is to show a fundamental asymptotic fixed-point property for a uniformly Lipschitzian mapping based on a family of nonlinear maps $\mathcal{F} = \{f_\lambda \mid \lambda \in \mathbb{R}^+\}$, where the choice of λ is important in controlling the sparsity effects. A nonexpansive orthogonal operator built on the Moore-Penrose pseudo-inverse is then applied in order to restore the solution feasibility.

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, continuous and differentiable in $\mathbb{R}$ with positive and even derivative $f'_\lambda(x) = (\lambda|x| - 1)e^{-\lambda|x|} + 1$. Since we have that $\sup_{x\in\mathbb{R}} |f'_\lambda(x)| = 1 + e^{-2}$, as a direct consequence of the intermediate value theorem from calculus it holds that $|f_\lambda(x) - f_\lambda(y)| \leq (1+e^{-2})|x-y|$, for each $\lambda > 0$ and $x, y \in \mathbb{R}$. Thus, mapping (2) is uniformly Lipschitzian with respect to λ with Lipschitz constant $1 + e^{-2}$. Moreover, given that $|f'_\lambda(x)| < 1$ on the interval $(-1/\lambda, 1/\lambda)$, the mapping (2) is contractive within that interval with fixed-point at the origin [4]. For a sketch of the function (2) and its first derivative see Fig. 1.

In order to understand the meaning of f_λ, it should be observed the role played by the mapping $x \mapsto f_\lambda(x)/x$, which behaves like a symmetric sigmoid function, where larger values of λ give sharper sigmoids, in the limit becoming a Heaviside step function. Roughly speaking, this asymptotic behavior has shrinking effects for values near zero, where the function f_λ becomes a contraction (see derivative in Fig. 1 which is less then 1 for $x < 1/\lambda$) and behaves as a second-order infinitesimal operator, while far from the origin it is asymptotic to the identity function, being $\lim_{\lambda\to+\infty} |f_\lambda(x)| = x$ for all $x \in \mathbb{R}$.

[1]The ℓ^1-norm is defined as $\|\alpha\|_1 = \sum_i |\alpha_i|$.

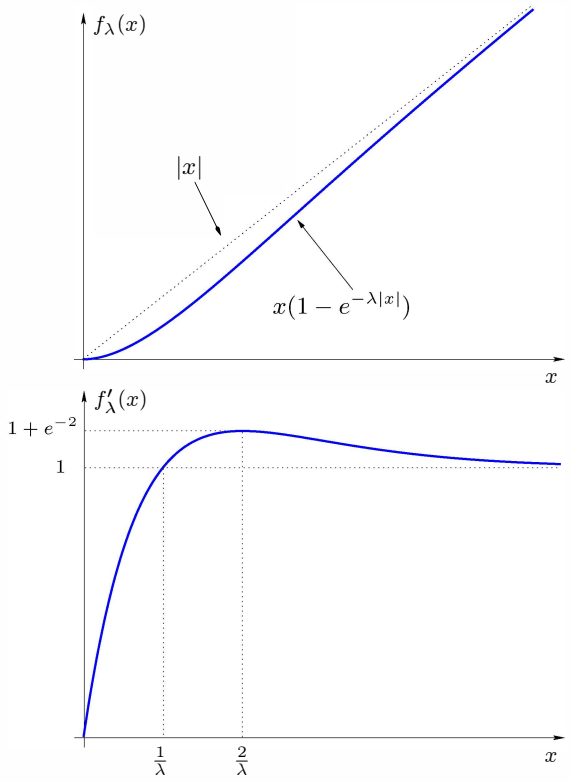

Fig. 1. The graph of function (2) and its first derivative.

To deal with high dimensional data, we extend mapping (2) to many dimensions via the elementwise Hadamard product of vectors, obtaining the one-parameter family of nonlinear functions $\mathcal{F} = \{f_\lambda : \mathbb{R}^m \to \mathbb{R}^m \mid \lambda \in \mathbb{R}^+\}$ where each component of $\mathcal{F}$ extends (2) as follows

$$f_\lambda(x) = x \odot (1 - e^{-\lambda|x|}), \tag{3}$$

where $\odot$ denotes the Hadamard product.

As previously noted, from the geometric point of view the vector $[f_\lambda(\alpha_1)/\alpha_1, \ldots, f_\lambda(\alpha_m)/\alpha_m]$ represents a symmetric sigmoid function in m dimensions, where larger values of λ give sharper sigmoids, in the limit becoming a Heaviside step function.

Let us now consider an orthogonal projector aimed to map every point falling in the range of (3) into the nearest point in the affine space $\mathcal{A}_{\Phi,s} = \{\alpha \in \mathbb{R}^m : \Phi\alpha = s\}$ (supposed not empty) given by equation (1).

Denote by $\Phi^\dagger = (\Phi^T\Phi)^{-1}\Phi^T$ the Moore-Penrose pseudo-inverse of a full-rank matrix Φ and by $\nu = \Phi^\dagger s$ the closed-form least-squares solution, i.e., those heaving minimum ℓ^2-norm. Given an approximate solution α of system (P0), in order to minimize the euclidean norm of the residual vector $s - \Phi\alpha$, or equivalently, the distance between α and the set of feasible solutions $\mathcal{A}_{\Phi,s}$, we use orthogonal projections as given by the following mapping:

$$\alpha \mapsto \alpha + \Phi^\dagger(s - \Phi\alpha) = P\alpha + \nu, \tag{4}$$

where $P = I - \Phi^\dagger\Phi$ is the orthogonal projector onto the kernel of Φ.

Putting all together, a new map $T_\lambda : \mathbb{R}^m \to \mathcal{A}_{\Phi,s}$, obtained by combining nonlinear mappings falling in the family $\mathcal{F}$ and the orthogonal projector (4), is given by

$$T_\lambda(\alpha) = P f_\lambda(\alpha) + \nu. \tag{5}$$

Also for mapping (5) we can provide an estimate of the Lipschitzianity by recalling that the operator P is orthogonal (and then firmly nonexpansive), which allows to derive:

$$\begin{aligned} \| T_\lambda(\alpha) - T_\lambda(\beta) \| &= \| P f_\lambda(\alpha) - P f_\lambda(\beta) \| \\ &\leq \| f_\lambda(\alpha) - f_\lambda(\beta) \| \\ &\leq \sqrt{m}\left(1 + e^{-2}\right) \| \alpha - \beta \| . \end{aligned} \tag{6}$$

Therefore the Lipschitz constant $\sqrt{m}\left(1 + e^{-2}\right)$ depends sublinearly on the greatest dimension of the spaces involved in (P0).

Here, we show that a fixed-point iterative schema involving mappings like (5), where the nonlinear function belongs to $\mathcal{F}$, becomes an effective procedure to find the sparsest solution of the linear system (1), provided that a suitable sequence of values would be supplied to the parameter λ.

Thus, fixed a positive sequence $\{\lambda_n\}_{n\geq 0}$, the iterates are inductively defined as

$$\left\{ \begin{array}{rcl} \alpha_0 & = & \alpha \in \mathbb{R}^m \\ \alpha_{n+1} & = & T_{\lambda_n}(\alpha_n) \end{array} \right. . \tag{7}$$

In order to study the convergence of the sequence $\{\alpha_n\}_{n\geq 0}$, we first provide an inductive form of the general term α_n.

Let $\{\alpha_n\}$ be the sequence in $\mathcal{A}_{\Phi,s}$ generated by the operator $T_\lambda(\cdot)$ in (7). Then it can be verify that

$$\begin{aligned} \alpha_1 &= P\alpha - P\left[\alpha \odot e^{-\lambda_\bullet |\alpha|}\right] + \nu \\ \alpha_2 &= P\alpha - P\left[\alpha \odot e^{-\lambda_\bullet |\alpha|} + \alpha_1 \odot e^{-\lambda_1 |\alpha_1|}\right] + \nu \\ &\vdots \\ \alpha_n &= P\alpha - P\left[\sum_{k=0}^{n-1} \alpha_k \odot e^{-\lambda_k |\alpha_k|}\right] + \nu \, . \end{aligned} \tag{8}$$

Thus, it is easy to see that the sequence converges when

$$P\alpha + \nu = (I - \Phi^\dagger \Phi)\alpha + \Phi^\dagger s = \alpha,$$

or equivalently, when

$$P\left[\sum_{k=0}^{\infty} \alpha_k \odot e^{-\lambda_k |\alpha_k|}\right] = 0 \tag{9}$$

that is, if and only if P maps the point $\sum_{k=0}^{\infty} \alpha_k \odot e^{-\lambda_k |\alpha_k|}$ onto the null vector.

Proving that $\{\alpha_n\}_{n\geq 0}$ converges in norm, implies to show that the series $\sum_{k=0}^{n-1} \alpha_k \odot e^{-\lambda_k |\alpha_k|}$ appearing in (8) converges in norm when $n \to +\infty$. But the proof can be easily derived by recalling that

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

which implies

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

Hence, by assuring a bound on the sum of reciprocals of $\{\lambda_n\}$, i.e., when $\sum_{n=0}^{\infty} 1/\lambda_n < +\infty$, it must be stated that

$$\| \alpha_n \| \to \| \alpha \| . \tag{10}$$

IV. The algorithm LiMapS

Stating the role of the parameter λ in the family of Lipschitzian-type mappings $\mathcal{F}$, we call it *sparsity ratio* because it determines how strong the overall increment of the sparsity level should be within each step of the iterative process. In fact, when applied iteratively, for small λ this kind of mappings should promote sparsity by forcing the magnitude of all components α_i to become more and more close to zero (recall that the map is contractive within $(-1/\lambda, 1/\lambda)$). On the other hand, for high values of λ, the chance to reduce the magnitudes of the α_i diminishes, fixing its value over the time. Hence, for gaining sparsity, the scheduling of sparsity ratio λ should start from small values and then increase according to the iteration step n.

This behavior is exhibited by the algorithm LiMapS (which stands for Lipschitzian Mappings for Sparsity), whose pseudo-code is sketched in Algorithm 1.

Algorithm 1 LiMapS

Require: - a dictionary $\Phi \in \mathbb{R}^{n\times m}$
- its pseudo-inverse $\Phi^\dagger$
- a signal $s \in \mathbb{R}^n$
- a sequence $\{\lambda_t\}_{t\geq 0}$

1: $t \leftarrow 0$
2: $\alpha \leftarrow \nu$
3: **while** [cond] **do**
4: $\quad \lambda \leftarrow \lambda_t$ <*sparsity ratio update*>
5: $\quad \beta \leftarrow f_\lambda(\alpha)$ <*increase sparsity*>
6: $\quad \alpha \leftarrow \beta - \Phi^\dagger(\Phi\beta - s)$ <*orthogonal projection*>
7: $\quad t \leftarrow t + 1$ <*step update*>
8: **end while**

Ensure: a fixed-point $\alpha = P\alpha + \nu$

Remark 1 As said informally above, its ability to find desired solutions is given by wise choices which will be adopted for the sequence $\{\lambda_t\}_{t\geq 0}$, together with choosing a good dictionary. Among many candidates respecting the constraints imposed by (10), one of the most promising sequences, at least on empirical grounds, is the geometric progression whose t-th term has the form

$$\lambda_t = \gamma \lambda_{t-1} = \theta \gamma^t \qquad \text{for } t \geq 1,$$

where $\lambda_0 = \theta$ and γ are positive and fixed constants.

Remark 2 The stop condition of the **while** loop (line 3: of Algorithm 1) may capture different events leading to a

correct termination of the iterative system to a solution having minimum error and, hopefully low sparsity. Even if to satisfy the ideal condition (9) may require infinite steps, in realistic computations, when the algorithm reaches values near machine precision we found solutions with very small error. Possible choices may include to bound the difference between two successive iterates, that is, until $\| \alpha_n - \alpha_{n-1} \| \geq \varepsilon$, or the discrepancy between the value $\left\| P \left[\sum_{k=0}^{\infty} \alpha_k \odot e^{-\lambda_k |\alpha_k|} \right] \right\|$ and zero.

V. Sparsity Minimization

The main goal of every sparsity recovery method is to find sub-optima to the hard problem (P0) in a fast way. In this work we face this problem in two stages:

1) relaxing the combinatorial objective function $\| \cdot \|_0$ to a continuous surrogate function characterized by a gradual approximation of the ℓ^0-norm leading to an asymptotic convergence with the latter during the process minimization;
2) optimizing such a relaxed objective function by invoking the nonlinear operator (5) which gives rise to a fixed-point algorithm for finding accurate solutions under plausible hypothesis of sparseness.

To make the discussion more formal, we follow the spirit of paper [9] in which a large class of admissible sparseness measures is introduced and characterized. This class include ℓ^p-norms ($0 \leq p \leq 1$) as a special case.

A meaningful functional directly linked to the sparsity promotion mapping (3) and approximatively measuring its support is the functional $g_\lambda : \mathbb{R}^m \to \mathbb{R}$, we call "$g_\lambda$-norm", given by

$$g_\lambda(|\alpha|) = \sum_{i=1}^{m} \left(1 - e^{-\lambda |\alpha_i|} \right) = m - \sum_{i=1}^{m} e^{-\lambda |\alpha_i|}, \qquad (11)$$

where the sharpness parameter λ is assumed to be positive. Its importance derives from the fact that g-norm behaves like the ℓ^p-norm for $p = 1/\lambda < 1$ and becomes $\| \alpha \|_0$ as λ tends to $+\infty$, as stated in the following properties:

$$g_\lambda(|\alpha|) \approx \| \alpha \|_p^p \quad \text{(with } p = 1/\lambda < 1\text{)}$$
$$\lim_{\lambda \to +\infty} g_\lambda(|\alpha|) = \| \alpha \|_0 \quad \text{(asymptotically).}$$

As far as the second stage is concerned, it would seem very natural to attempt to solve a new regularization problem of the form:

$$\min_{\alpha} \|\alpha\|_g \quad \text{subject to} \quad \Phi\alpha = s \qquad \text{(P1)}$$

by applying the mapping (5) and by letting λ tend to infinity. This approach leads to a family of fixed-point algorithms that starting by an arbitrary point in $\mathbb{R}^m$ converges to a solution of the problem (P1) by locally minimizing the g-norm.

In order to explain such a behaviour that appears to be characteristic of the iterative schema given in the previous section, due to space limit we do not report a formal rigorous analysis concerning the pseudo-norm minimization, but we highlight empirically the trend of the g-norm against various more or less sparse instances. In Fig. 2 we report the plottings of g-norm for random instances of size $n = 200$ and randomly generated Gaussian dictionary with $m = 800$ atoms, limiting to 2000 the number of iterations of the fixed-point schema. Such a metric provides a clear dichotomic behavior of the algorithm: when it converges to the optimum value k of a given k-sparse signal the g-norm values tends to k (e.g., see the plottings for $k \in \{10, 20, \ldots, 70\}$), otherwise they tend to diverge, as in cases of plottings for $k \in \{80, 90, 100\}$. It can be noted that this dichotomy also allows to detect, after a small iteration number, whether the process converges or not to a good solution.

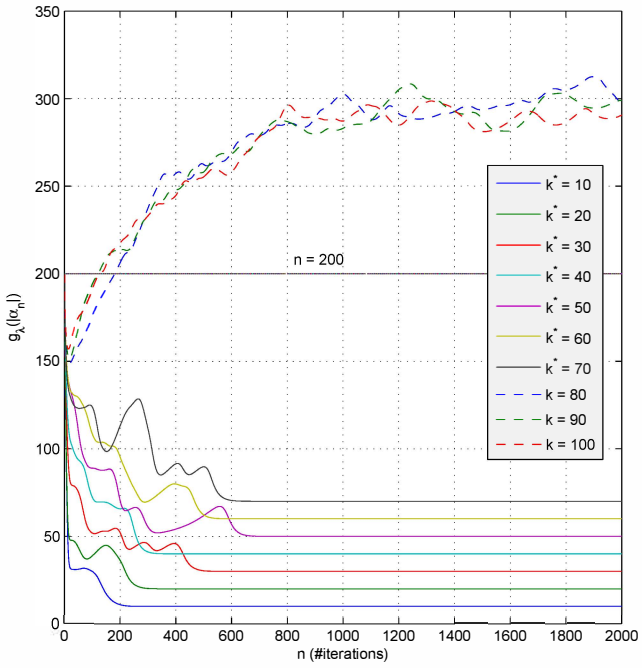

Fig. 2. Plotting of the g_λ-norm functional versus iterations of the **while** loop of Algorithm 1.

VI. Numerical Results

To show the effectiveness of our algorithm we directly compared it with some algorithms for sparsity recovery well-known in literature, as Matching Pursuit (MP) [10], Orthogonal Matching Pursuit (OMP) [11], Stagewise Orthogonal Matching Pursuit (StOMP) [12], LASSO [13], LARS [13], Smoothed L0 (SL0) [7] and Improved SL0 (ISL0[2]) [14].

In all tests, the frames Φ and the optimum coefficients α^* are randomly generated using the noiseless Gaussian-Bernoulli

[2]In order to make the algorithm behavior more stable, in our ISL0 implementation we used the explicit pseudo inverse calculation instead of the conjugate gradient method, so penalizing its time performances in case of big size instances.

stochastic model, i.e., for all $i, j \in [1, \ldots, m]$:

$$\Phi_{ij} \sim \mathcal{N}(0, n^{-1}) \quad \text{and} \quad \alpha_i^* \sim x_i \cdot \mathcal{N}(0, \sigma),$$

where $x_i \sim \text{Bern}(p)$. In this way each coefficient α_i^* has probability p to be active and probability $1-p$ to be inactive. When the coefficient α_i^* is active, its value is randomly drawn with a Gaussian distribution having zero mean and standard deviation σ. Conversely, if the coefficient is not active the value is set to zero. As far as the parameters are concerned, we fix $\lambda_0 = 10^{-3}$ and $\gamma = 1.01$ because they have given good results in all considered instances, coming out essentially independent from the size $n \times m$ of the frames and size m of the coefficient vectors.

We evaluate the performances of the algorithms measuring relative error and computation time:

1) as errors we consider the Signal-to-Noise-Ratio (SNR) and the Sum of Squares Error (SSE) of an approximate solution found α with respect the optimum α^*, defined as usual:

$$\text{SNR} = 20 \log_{10} \frac{\|\alpha\|}{\|\alpha - \alpha^*\|}, \quad \text{SSE} = \|s - \Phi\hat{\alpha}\|^2;$$

2) as computation time we take the CPU time spent in the execution of the algorithm cores, without including the computation of instances generation or the pseudo-inverse matrix of the dictionary in our and SL0 algorithms.

The simulations were performed on AMD Athlon II X4 630 Processor 64 bit, 2.8 GHz processor with 4 GB of memory, using MATLAB with SparseLab (http://sparselab.stanford.edu) and Toolbox Sparse Optimization (http://www.ceremade.dauphine.fr/~peyre/matlab/) for algorithms implementation. The algorithm LIMAPS is available online at the URL http://dalab.dsi.unimi.it/limaps.

Among the many experiments done, in Figure 3, Figure 4 and Figure 5 we report the average SNR, times and the relative number (in %) of correctly recovered atoms values respectively, obtained from executions on instances of $n = 200$ equations and m ranging from 300 to 1400 variables, moving the percentage of sparsity k from 10% to 50% over n. For each n, m and k 100 instances of dictionary and coefficients were randomly generated.

As can be noted, our algorithm outperforms all the others with regard to the reconstruction quality, reaching arbitrary precision and keeping a CPU execution time comparable with the others. The most interesting results are obtained with the sparsity levels between the 30% and the 50% over n, where our algorithm keeps a good accuracy in terms of SNR.

A second kind of experiment was aimed to study the algorithm behavior when the sparsity level k is low (e.g., 50% over n), that is when algorithms find more difficulties in general to converge toward the sparsest solution. To this end, we have generated random instances of dimensions $n = 400$ and $m = 800$ with a sparsity level $k = 200$, doing also for this case 1000 trials. The results are outlined in Table I, listed by error averages $\mu_{\text{SSE}}, \mu_{\text{SNR}}$ and time averages μ_{time} together

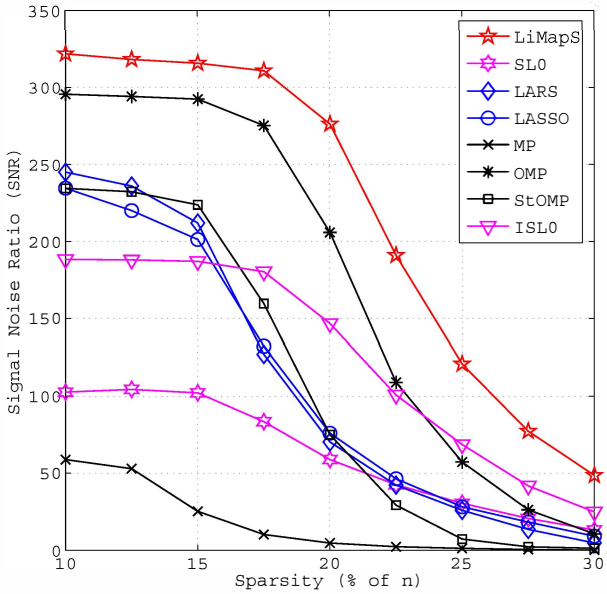

Fig. 3. Averages SNR of the algorithms vs. sparsity, expressed in percentage of the number of equations n.

TABLE I
AVERAGES AND STANDARD DEVIATIONS OF THE RESULTS OBTAINED BY THE ALGORITHMS FROM 1000 TRIALS WITH INSTANCES OF DIMENSIONS $m = 800$, $n = 400$ AND $k = 200$.

	μ_{SSE}	σ_{SSE}	μ_{SNR}	σ_{SNR}	μ_{time}	σ_{time}
LIMAPS	**1.5e-24**	1.3e-24	**249.8**	114.9	0.79	0.23
SL0	4.8e-24	7.6e-25	24.7	38.3	**0.15**	0.01
ISL0	4.3e-16	3.7e-15	82.7	89.5	9.10	12.70
LASSO	1.3e+02	1.1e+03	8.3	2.0	1.79	0.20
LARS	2.3e-10	2.3e-09	6.3	2.2	0.79	0.07
MP	2.4e+04	4.3e+03	1.9	0.7	0.18	0.01
OMP	2.4e+00	2.8e-01	1.6	5.5	11.4	0.79
StOMP	3.8e+05	1.4e+05	2.4	0.7	0.02	0.01

with their relative standard deviations σ. Again LIMAPS gives the best results in terms of SNR and of SSE with lower standard deviations while the times remain comparable with other algorithms. Finally, it must be noted that the SSE of solutions found by LIMAPS vanishes at each iteration of while cycle (statement 1: in Algorithm 1) since they are remapped every time onto the feasible space of (P0).

VII. CONCLUSIONS

We developed a new heuristic to solve efficiently the sparsity coding of signals described by underdetermined linear systems. It consists on alternating two lipschitzian maps: one promotes the sparsity of each near-feasible solution (or point) falling outside the affine space associated with the linear

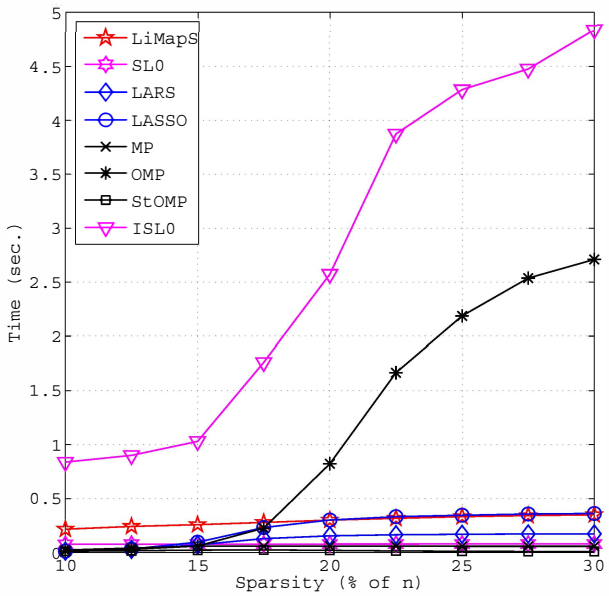

Fig. 4. Averages computational times of the algorithms vs. sparsity, expressed in percentage of the number of equations n.

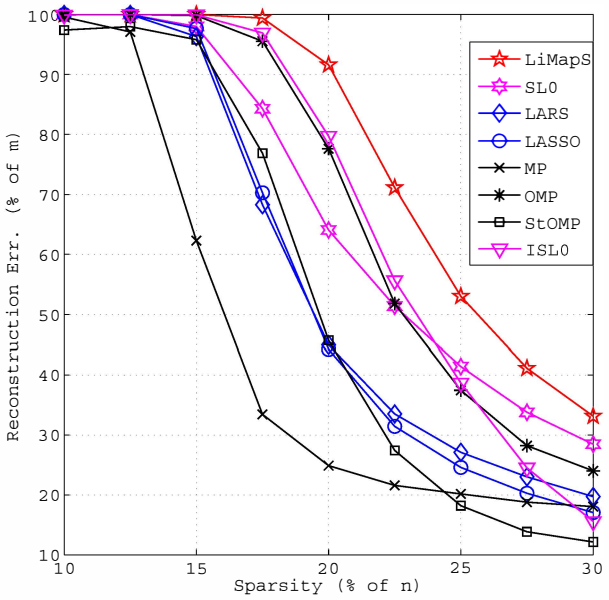

Fig. 5. Relative number (in %) of correctly recovered atoms not equal to zero. An atom is considered correctly reconstructed if the deviation from the true value of the estimated value is less than 5%.

transformation and the other remaps such a point in the nearest solution of the feasible space.

The so derived heuristic consists on a fixed-point iteration scheme which converges to a good solution coinciding, in many cases, with the sparsest (or optimum) solution admitted. The trajectory followed by the system evolution is leaded by a unique parameter which assumes growing values given by empirical analysis. To make the sparsification request effective, we introduce a new metric into the search space with two aims: to measure the goodness of the sparsity level reached and to act as objective function directly implied in the optimization process. Such a critical aspect of the framework proposed deserves a deeper analysis in order to make the search strategy more efficient, reducing the number of iterations of the main cycle and avoiding to get stuck into local minima as well.

With the experimental results we highlight the high solution quality and a good average time complexity in practice, comparable with the fastest well-known reconstruction algorithms; in particular, such technique is promising because it exhibits very good performances (high SNR) also in case of very low sparsity (near $n/2$), values for which many others fail.

REFERENCES

[1] D. L. Donoho, For most large underdetermined systems of linear equations the minimal l1-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 Journal on Scientific Computing 20 (1998) 33–61.

[3] E. Candès, T. Tao, Decoding by linear programming, IEEE Trans. Inform. Theory 52 (2006) 5406–5425.

[4] S. Banach, Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales, Fundamenta Mathematicae III (1922) 133–181.

[5] K. Goebel, W. A. Kirk, A fixed point theorem for asymptotically nonexpansive mappings, Proc. Amer. Math. Soc. 35 (1972) 171–174.

[6] B. K. Natarajan, Sparse approximate solutions to linear systems, SIAM J. Comput. 24 (1995) 227–234.

[7] H. Firouzi, M. Farivar, M. Babaie-Zadeh, C. Jutten, Approximate sparse decomposition based on smoothed l0-norm, CoRR abs/0811.2868.

[8] A. Blake, A. Zisserman, Visual Reconstruction, MIT Press, 1987.

[9] R. Gribonval, M. Nielsen, Highly sparse representations from dictionaries are unique and independent of the sparseness measure, Applied and Computational Harmonic Analysis 22 (3) (2007) 335–355.

[10] S. Mallat, Z. Zhang, Matching pursuit with time-frequency dictionaries, IEEE Transactions on Signal Processing 41 (1993) 3397–3415.

[11] J. A. Tropp, Anna, C. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, IEEE Trans. Inform. Theory 53 (2007) 4655–4666.

[12] D. L. Donoho, Y. Tsaig, I. Drori, J. luc Starck, Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit, Tech. rep., Standford University (2006).

[13] B. E. Trevor, T. Hastie, L. Johnstone, R. Tibshirani, Least angle regression, Annals of Statistics 32 (2002) 407–499.

[14] M. Hyder, K. Mahata, An improved smoothed ℓ^0 approximation algorithm for sparse representation, IEEE Transactions on Signal Processing 58 (4) (2010) 2194–2205.