HAL
archives-ouvertes.fr

Iterative Cosparse Projection Algorithms for the Recovery of Cosparse Vectors

Raja Giryes, Sangnam Nam, Rémi Gribonval, Mike E. Davies

To cite this version:

Raja Giryes, Sangnam Nam, Rémi Gribonval, Mike E. Davies. Iterative Cosparse Projection Algorithms for the Recovery of Cosparse Vectors. The 19th European Signal Processing Conference (EUSIPCO-2011), 2011, Barcelona, Spain. 2011. <inria-00611592>

HAL Id: inria-00611592

https://hal.inria.fr/inria-00611592

Submitted on 26 Jul 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ITERATIVE COSPARSE PROJECTION ALGORITHMS FOR THE RECOVERY OF COSPARSE VECTORS

Raja Giryes[1], Sangnam Nam[2], Rémi Gribonval[2] and Mike E. Davies[3]

[1] Computer Science Department,
Technion – Israel Institute of Technology
Haifa 32000, Israel

email: raja@cs.technion.ac.il

[2] INRIA Rennes - Bretagne Atlantique
Campus de Beaulieu
F-35042 Rennes Cedex, France

email: firstname.lastname@inria.fr

[3] School of Engineering and Electronics
The University of Edinburgh
The King's Buildings, Mayfield Road,
Edinburgh EH9 3JL, UK
email: Mike.Davies@ed.ac.uk

ABSTRACT

Recently, a cosparse analysis model was introduced as an alternative to the standard sparse synthesis model. This model was shown to yield uniqueness guarantees in the context of linear inverse problems, and a new reconstruction algorithm was provided, showing improved performance compared to analysis ℓ_1 optimization. In this work we pursue the parallel between the two models and propose a new family of algorithms mimicking the family of Iterative Hard Thresholding algorithms, but for the cosparse analysis model. We provide performance guarantees for algorithms from this family under a Restricted Isometry Property adapted to the context of analysis models, and we demonstrate the performance of the algorithms on simulations.

1. INTRODUCTION

Many natural signals and images have been observed to be inherently low dimensional despite their possibly very high ambient signal dimension. It is by now well understood that this phenomenon lies at the heart of the success of numerous methods of signal and image processing.

The sparse synthesis data model, which has enjoyed much popularity in recent years, is the leading model associated to this observation. To show more concretely how useful such a model can be, the following generic linear inverse problem is considered in this paper : For some unknown signal $\mathbf{x}_0 \in \mathbb{R}^d$, an incomplete set of linear observations $\mathbf{y} \in \mathbb{R}^m$ is available via $\mathbf{y} = \mathbf{M}\mathbf{x}_0 + \mathbf{e}$, where $\mathbf{e} \in \mathbb{R}^m$ is an additive bounded noise that satisfies $\|\mathbf{e}\|_2^2 \leq \varepsilon^2$. The task is to recover or approximate $\mathbf{x}_0$. In the noiseless setting where $\mathbf{e} = 0$, this amounts to solving

$$\mathbf{y} = \mathbf{M}\mathbf{x}. \tag{1.1}$$

Of course, a simple fact in linear algebra tells us that (1.1) admits infinitely many solutions. Therefore, when all we have is the observation $\mathbf{y}$ and the measurement/observation matrix $\mathbf{M}$, we are in a hopeless situation to recover $\mathbf{x}_0$.

Sparse recovery guarantees. This is where the utility of what we may call in general the 'sparse signal models' comes into play. In the sparse synthesis model, the signal $\mathbf{x}_0$ is assumed to have a very sparse representation in a given fixed dictionary $\mathbf{D} \in \mathbb{R}^{d\times n}$. In other words, there exists $\mathbf{z}_0$ with few nonzero entries as counted by the "ℓ_0-norm" $\|\mathbf{z}_0\|_0$ such that

$$\mathbf{x}_0 = \mathbf{D}\mathbf{z}_0, \quad \text{and} \quad k := \|\mathbf{z}_0\|_0 \ll d. \tag{1.2}$$

In more realistic cases $\mathbf{x}_0 \approx \mathbf{D}\mathbf{z}_0$. Having this knowledge we solve (1.1) using $\mathbf{x}_S = \mathbf{D}\mathbf{z}_S$ where

$$\mathbf{z}_S = \operatorname*{argmin}_{\mathbf{z}} \|\mathbf{z}\|_0 \quad \text{subject to} \quad \|\mathbf{y} - \mathbf{M}\mathbf{D}\mathbf{z}\|_2^2 \leq \varepsilon^2. \tag{1.3}$$

The upshot of the model (1.2) in the context of the inverse problem (1.1), in the case that $\mathbf{e} = 0$, is that we know exactly when we can guarantee the uniqueness of sparse solution to the resulting equation $\mathbf{y} = \mathbf{M}\mathbf{D}\mathbf{z}$. Precisely, it is shown [5, 9] that there exists k_0 such that if $\mathbf{y} = \mathbf{M}\mathbf{D}\mathbf{z}_0$ with $k = \|\mathbf{z}_0\|_0 < k_0$, then the linear system $\mathbf{y} = \mathbf{M}\mathbf{D}\mathbf{z}$ has a unique k-sparse solution, which is necessarily $\mathbf{z}_0$. The largest such constant k_0 is exactly known to be $\mathrm{spark}(\mathbf{M}\mathbf{D})/2$, where $\mathrm{spark}(\mathbf{M}\mathbf{D})$ is the minimum number of rows of $\mathbf{M}\mathbf{D}$ that are linearly dependent. Consequently, if $\mathbf{x}_0$ is sufficiently sparse—in the sense of (1.2)—then we can recover it by solving (1.3) with $\varepsilon = 0$.

Convex optimization algorithms. Unfortunately, solving (1.3) was shown to be NP hard. A convex relaxation using an ℓ_1-norm instead of an "ℓ_0-norm" can overcome this computational issue [5, 9]. For example, one can solve

$$\mathbf{z}_S = \operatorname*{argmin}_{\mathbf{z}} \|\mathbf{z}\|_1 \quad \text{subject to} \quad \|\mathbf{y} - \mathbf{M}\mathbf{D}\mathbf{z}\|_2^2 \leq \varepsilon^2. \tag{1.4}$$

The ℓ_1-minimization has gained much popularity both in theory and practice. Interestingly, a slightly different form of ℓ_1-minimization has been used in practice and studied [6]. In this 'altered form,' one solves

$$\mathbf{x}_A = \operatorname*{argmin}_{\mathbf{x}} \|\mathbf{\Omega}\mathbf{x}\|_1 \quad \text{subject to} \quad \|\mathbf{y} - \mathbf{M}\mathbf{x}\|_2^2 \leq \varepsilon^2, \tag{1.5}$$

where $\mathbf{\Omega} \in \mathbb{R}^{p\times d}$ is an *analysis operator*. Examples of such operators include finite difference operators, curvelet transforms, and undecimated wavelet transforms. Despite apparent similarities between (1.4) and (1.5), some fundamental differences have been demonstrated [6].

Cosparse signal model and recovery guarantees. Recently, it has been pointed out that a new signal model called *cosparse analysis model* [11] is more relevant to (1.5) than the sparse synthesis model. In particular, the uniqueness property in the context of linear inverse problem (1.1), in the noiseless case, has been obtained [11] for this model, exploiting general results on union of subspace models [10]. Therefore, one has a partial but theoretically firm foundation for using (1.5) for signal recovery.

In the analysis model we aim at solving the following minimization problem:

$$\mathbf{x}_A = \operatorname*{argmin}_{\mathbf{x}} \|\mathbf{\Omega}\mathbf{x}\|_0 \quad \text{subject to} \quad \|\mathbf{y} - \mathbf{M}\mathbf{x}\|_2^2 \leq \varepsilon^2. \tag{1.6}$$

This work was supported in part by the EU FP7-ICT-FET-Open program, project SMALL, grant number: 225913

The model and the uniqueness result can be summarized as follows: For a fixed analysis operator $\mathbf{\Omega} \in \mathbb{R}^{p \times d}$, a signal $\mathbf{x}_0 \in \mathbb{R}^d$ is said to satisfy the cosparse analysis model if

$$\ell := p - \|\mathbf{\Omega}\mathbf{x}_0\|_0 \text{ is large.} \tag{1.7}$$

The quantity ℓ is the *cosparsity* of $\mathbf{x}_0$ and $\mathbf{x}_0$ is said to be ℓ-cosparse, or simply cosparse. As the definition of cosparsity ℓ suggests, the emphasis of the cosparse analysis model is on 'many zeros' of the analysis representation $\mathbf{\Omega}\mathbf{x}_0$. This contrasts to the emphasis on 'few non-zeros' of a synthesis representation $\mathbf{z}_0$ in the synthesis model (1.2). The uniqueness result for the analysis model reads as expected: There exists ℓ_0 such that if $\mathbf{y} = \mathbf{M}\mathbf{x}_0$ for ℓ-cosparse $\mathbf{x}_0$ with $\ell \geq \ell_0$, then the problem (1.6) when $\varepsilon = 0$ has a unique ℓ-cosparse solution. We refer the readers to [11] for more discussion.

New algorithms for cosparse recovery. The two ℓ_1-minimization principles (1.4) and (1.5) can now be seen as methods to solve or approximate the solutions of (1.3) and (1.6) for the recovery of sparse and cosparse signals, respectively. Having this parallelism, we ask immediately: what about a plethora of algorithms that also aims to solve (1.3)? Are there analogous counterparts for the analysis model? For some algorithms, the answer is definitely yes. Indeed, a greedy algorithm called Greedy Analysis Pursuit (GAP) has been developed [11]. It somehow mimics Orthonormal Matching Pursuit [12], and its effectiveness was demonstrated for the cosparse signal recovery problem.

Contributions. Another avenue exists for an analysis algorithm in the direction of the iterative hard thresholding algorithms [2, 7, 4]. The contributions in this paper are:

- Algorithms in the cosparse analysis framework are defined in Section 2 in the spirit of Iterative Hard Thresholding (IHT, [2]) and Hard Thresholding Pursuit (HTP, [7]). Note that the main novelty is in the definition of an appropriate *cosparse projection* replacing the *hard thresholding* step used in the synthesis framework.
- A success guarantee based on an RIP-like property for the analysis model is provided in Section 3. As far as we know, no uniform guarantees exists for a practical algorithm in the analysis model. This guarantee, unlike the one in [3] gives a bound for all cosparse signals under the appropriate conditions.
- Empirical performance is demonstrated in Section 4 in the context of the cosparse signal recovery problem.

2. ALGORITHMS DESCRIPTION

It is quickly observed that the solution $\mathbf{z}_0$ of the ideal problem (1.3) for the case $\varepsilon = 0$ satisfies: a) $\|\mathbf{z}_0\|_0 = k$ (assuming prior knowledge of the optimum sparsity k); and b) $\|\mathbf{y} - \mathbf{M}\mathbf{D}\mathbf{z}_0\|_2^2 = 0$. This observation somehow motivates the basic ideas of IHT and HTP as follows: We look for a coefficient vector $\mathbf{z}_0$ that minimizes an objective function $f(\mathbf{z}) := \|\mathbf{y} - \mathbf{M}\mathbf{D}\mathbf{z}\|_2^2$ under the constraint $\|\mathbf{z}_0\|_0 = k$. The main difficulty here is that the constraint is not only non-convex but also non-smooth.

2.1 Quick review of IHT and HTP.

Fortunately, there are two very simple tasks intimately related to the considered problem; the first one is when we have an estimate, say $\hat{\mathbf{z}}_1$, of $\mathbf{z}_0$, it is easy to find a better estimate in terms of reducing the objective f. This can be done by a simple gradient step. The second task is when a new estimate, say $\hat{\mathbf{z}}_1^g$, is not k-sparse, it is straightforward to find the best/closest k-sparse signal $\hat{\mathbf{z}}_2$. This is achieved by a simple hard thresholding, leading to the IHT algorithm [2]. Of course, in the second projection step, we can be mindful of the objective function as well, and instead of simple hard thresholding, we may only take the support of $\hat{\mathbf{z}}_2$ obtained from the hard thresholding but optimize $f(\mathbf{z})$ over all $\mathbf{z}$'s with the same support in order to get a new estimate, let us call it again $\hat{\mathbf{z}}_2$. The HTP algorithm implements this idea [7]. Going further, Cevher [4] proposes to optimize the stepsize in order to minimize $f(\mathbf{z})$ while taking into account a possible change in the support selected by hard thresholding. We refer the readers to [2, 7, 4] for more details and turn our attention to analysis counterparts.

2.2 Overview of proposed cosparse recovery algorithms

The discussion above given the cosparsity ℓ of a sufficiently cosparse signal $\mathbf{x}_0$, in order to solve the ideal problem (1.6), we can look for a solution $\mathbf{x}_A$ that minimizes an objective $f(\mathbf{x}) := \|\mathbf{y} - \mathbf{M}\mathbf{x}\|_2^2$ under the constraint $p - \|\mathbf{\Omega}\mathbf{x}\|_0 = \ell$. From this, we are led to algorithms for the cosparse signal recovery in the spirit of IHT and HTP. We present a group of algorithms that share similar ideas:

- First, these are iterative algorithms: sequences of estimates $\hat{\mathbf{x}}_i$, $i = 1, 2, \ldots$ are obtained by iterating a basic rule;
- Second, they all share an intermediate gradient descent step: At iteration $i > 1$, having the estimate $\hat{\mathbf{x}}_{i-1}$ on our hand, we compute an intermediate estimate $\hat{\mathbf{x}}_i^g$ by

  $$\hat{\mathbf{x}}_i^g = \hat{\mathbf{x}}_{i-1} + \mu_i \mathbf{M}^T(\mathbf{y} - \mathbf{M}\hat{\mathbf{x}}_{i-1}) \tag{2.1}$$

  for some appropriate $\mu_i > 0$. This corresponds to a gradient descent with respect to the objective function $f(\mathbf{x})$.
- Third, by an appropriate 'projection,' the intermediate estimate $\hat{\mathbf{x}}_i^g$ is projected to an ℓ-cosparse element $\hat{\mathbf{x}}_i$.

2.3 Some notations

The 'projection' step is the most specific novelty of the proposed algorithms. To explain it, let us fix some notations. For an index set $\mathbf{\Lambda} \subset [\![1, p]\!]$, $\mathbf{\Omega}_{\mathbf{\Lambda}}$ is the submatrix obtained from $\mathbf{\Omega}$ by taking the rows indexed by $\mathbf{\Lambda}$. Hence, if $\mathbf{x}_0$ is an ℓ-cosparse signal, then there are ℓ rows $\mathbf{\Lambda}$ of $\mathbf{\Omega}$ such that $\mathbf{\Omega}_{\mathbf{\Lambda}}\mathbf{x}_0 = 0$. This means that $\mathbf{x}_0$ is orthogonal to range$(\mathbf{\Omega}_{\mathbf{\Lambda}}^T)$. In other words, $\mathbf{x}_0$ belongs to the 'cosparse subspace' $\mathscr{W}_{\mathbf{\Lambda}} = \mathbf{Null}(\mathbf{\Omega}_{\mathbf{\Lambda}})$, which is the orthogonal complement of range$(\mathbf{\Omega}_{\mathbf{\Lambda}}^T)$. We denote the orthogonal projection onto range$(\mathbf{\Omega}_{\mathbf{\Lambda}}^T)$ and $\mathscr{W}_{\mathbf{\Lambda}}$ by $\mathbf{P}_{\mathbf{\Lambda}}$ and $\mathbf{Q}_{\mathbf{\Lambda}}$, respectively.

2.4 Cosparse projections

From $\hat{\mathbf{x}}_i^g$ which is not necessarily cosparse, we wish to obtain a new estimate $\hat{\mathbf{x}}_i$ which *is* ℓ-cosparse. *How* we obtain such a cosparse estimate $\hat{\mathbf{x}}_i$ is interesting and important.

Going back to the synthesis case, recall that given a coefficient vector $\mathbf{z}$, we obtain the k-sparse vector $\mathbf{z}_k$ closest to $\mathbf{z}$ through the so-called hard-thresholding operation $\mathbf{z}_k = H_k(\mathbf{z})$: by retaining only the k largest coefficients of $\mathbf{z}$ –excluding possible ties.

This is not so in the cosparse analysis model: When we apply a hard thresholding to an analysis representation $\mathbf{\Omega}\hat{\mathbf{x}}_i^g$, we obtain a vector $\mathbf{z} := H_{p-\ell}(\boldsymbol{\Omega}\hat{\mathbf{x}}_i^g)$ with ℓ zero entries, but we can no longer assert that $\mathbf{z}$ is an admissible analysis representation in general, i.e., there is no $\mathbf{x}$ such that $\mathbf{z} = \boldsymbol{\Omega}\mathbf{x}$. To overcome this, we can instead estimate the cosupport as the set of ℓ smallest entries of $\boldsymbol{\Omega}\hat{\mathbf{x}}_i^g$,

$$\hat{\boldsymbol{\Lambda}}_i := \underset{\tilde{\boldsymbol{\Lambda}}:|\tilde{\boldsymbol{\Lambda}}|=\ell}{\operatorname{argmin}} \left\|\boldsymbol{\Omega}_{\tilde{\boldsymbol{\Lambda}}}\hat{\mathbf{x}}_i^g\right\|_2^2, \tag{2.2}$$

and then 'project' $\hat{\mathbf{x}}_i^g$ onto the associated cosparse subspace $\mathscr{W}_{\hat{\boldsymbol{\Lambda}}}$ to obtain a 'good' estimate. How we project on $\mathscr{W}_{\hat{\boldsymbol{\Lambda}}}$ is the main difference between the proposed algorithms.

Orthogonal projection. A straightforward solution is to simply perform an orthonormal projection of $\hat{\mathbf{x}}_i^g$. This is the approach we take for the algorithm of IHT type, which we name A-IHT. Hence, in A-IHT, we carry out the steps (2.2) and

$$\hat{\mathbf{x}}_i = \mathbf{Q}_{\hat{\boldsymbol{\Lambda}}_i}\hat{\mathbf{x}}_i^g. \tag{2.3}$$

Note that the quality of this projection heavily depends on the analysis operator $\boldsymbol{\Omega}$. The projection is in general not the optimal projection onto the union of analysis cosparse subspaces. Further implication of this is discussed in section 3.

Best data fidelity projection. From the description above, the cosparse subspace $\mathscr{W}_{\hat{\boldsymbol{\Lambda}}_i}$, plays the role of the support for the k largest coefficients of synthesis representations. In the algorithm of HTP type, which we call A-HTP, we replace the orthogonal projection onto $\mathscr{W}_{\hat{\boldsymbol{\Lambda}}_i}$, (2.3), with a best data fidelity projection: we look for an element of $\mathscr{W}_{\hat{\boldsymbol{\Lambda}}_i}$ that minimizes $\|\mathbf{y}-\mathbf{M}\mathbf{x}\|_2^2$. That is, $\hat{\boldsymbol{\Lambda}}_i$ is given by (2.2) and

$$\hat{\mathbf{x}}_i = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{y}-\mathbf{M}\mathbf{x}\|_2^2 \quad \text{subject to} \quad \boldsymbol{\Omega}_{\hat{\boldsymbol{\Lambda}}_i}\mathbf{x} = 0. \tag{2.4}$$

2.5 Choice of the stepsize

Note that the choice of gradient stepsize μ_i is crucial: If μ_i's are chosen too small, the algorithm gets stuck at a wrong solution. If too large, the algorithm diverges. For our algorithms, we consider two options for μ_i. In the first option, we can choose $\mu_i = \mu$ for some constant μ for all iterations. A theoretical discussion on how to choose μ properly is given in Section 3. Another way is to solve the following problem

$$\mu_i := \underset{\mu}{\operatorname{argmin}} \left\|\mathbf{y}-\mathbf{M}(\hat{\mathbf{x}}_{i-1}+\mu\mathbf{M}^T(\mathbf{y}-\mathbf{M}\hat{\mathbf{x}}_{i-1}))\right\|_2^2, \tag{2.5}$$

that has a simple closed form solution. Algorithm 1 summarizes the proposed family of algorithms.

2.6 Targeted cosparsity level

Just as in the synthesis counterpart of the proposed algorithms, where a target sparsity level k must be selected before running the algorithms, we have to choose the targeted cosparsity level ℓ which will dictate the projection steps. In the synthesis case it is known that it may be interesting to over-estimate the sparsity k. Similarly in the analysis framework the question arises: In terms of recovery performance, does it help to under-estimate the cosparsity ℓ ? A tentative yes comes from the following heuristic: Let $\tilde{\boldsymbol{\Lambda}}$ be a subset of the cosupport $\boldsymbol{\Lambda}$ of signal $\mathbf{x}_0$ with $\tilde{\ell} := |\tilde{\boldsymbol{\Lambda}}| < \ell = |\boldsymbol{\Lambda}|$. Note that if the rank of $\boldsymbol{\Omega}_{\tilde{\boldsymbol{\Lambda}}}$ is greater than or equal to $d-m$, then in general it is sufficient to identify $\tilde{\boldsymbol{\Lambda}}$ in order to recover $\mathbf{x}_0$ from the relations $\mathbf{y} = \mathbf{M}\mathbf{x}_0$ and $\boldsymbol{\Omega}_{\tilde{\boldsymbol{\Lambda}}}\mathbf{x}_0 = 0$. Therefore, we can use $\tilde{\ell}$ as the effective cosparsity in the algorithms. The effect of varying the cosparsity level is shown in section 4.

Algorithm 1: Iterative Projection Algorithms: A-IHT, A-HTP

Input: $\mathbf{y}$, $\mathbf{M}$, $\boldsymbol{\Omega}$, $\tilde{\ell}$, $i_{\max}$, δ_{term}
Set $i = 0$, $\hat{\mathbf{x}}_0 = \mathbf{0}$
repeat
  $i := i+1$
  $\hat{\mathbf{x}}_i^g := \hat{\mathbf{x}}_{i-1} + \mu_i\mathbf{M}^T(\mathbf{y}-\mathbf{M}\hat{\mathbf{x}}_{i-1})$
  $\hat{\boldsymbol{\Lambda}}_i := \operatorname{argmin}_{\tilde{\boldsymbol{\Lambda}}:|\tilde{\boldsymbol{\Lambda}}|=\tilde{\ell}} \left\|\boldsymbol{\Omega}_{\tilde{\boldsymbol{\Lambda}}}\hat{\mathbf{x}}_i^g\right\|_2^2$
  case *A-IHT*
    $\hat{\mathbf{x}}_i := \mathbf{Q}_{\hat{\boldsymbol{\Lambda}}_i}\hat{\mathbf{x}}_i^g$
  case *A-HTP*
    $\hat{\mathbf{x}}_i := \operatorname{argmin}_{\mathbf{x}} \|\mathbf{y}-\mathbf{M}\mathbf{x}\|_2^2$ s.t. $\boldsymbol{\Omega}_{\hat{\boldsymbol{\Lambda}}_i}\mathbf{x} = 0$
until $i = i_{\max}$ or $\|\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_{i-1}\|_2 < \delta_{\text{term}}\|\hat{\mathbf{x}}_i\|_2$

---

3. THEORETICAL GUARANTEES

We now turn to theoretical performance guarantees for the proposed algorithms. For the general class of Iterative Projection Algorithms (IPA), which combine a gradient step and an optimal projection $P_{\mathscr{A}}$ onto a general union of subspaces, uniform guarantees were proved [1] assuming that $\mathbf{M}$ is bi-Lipschitz on the considered union of subspaces, denoted $\mathscr{A}$. In our case, $\mathscr{A} = \cup_{|\boldsymbol{\Lambda}|\geq\ell}\mathscr{W}_{\boldsymbol{\Lambda}}$, and the bi-Lipschitz constants of $\mathbf{M}$ are the largest α and smallest β where $0 < \alpha \leq \beta$ such that for all ℓ-cosparse vectors $\mathbf{x}_1, \mathbf{x}_2$:

$$\alpha\|\mathbf{x}_1+\mathbf{x}_2\|_2^2 \leq \|\mathbf{M}(\mathbf{x}_1+\mathbf{x}_2)\|_2^2 \leq \beta\|\mathbf{x}_1+\mathbf{x}_2\|_2^2. \tag{3.1}$$

Under this assumption[1], one can apply Theorem 2 from [1] to the idealized algorithm that performs an optimal projection $\hat{\mathbf{x}}_i = P_{\mathscr{A}}(\hat{\mathbf{x}}_i^g)$ as a third step (instead of the suboptimal projections described in section 2.4).

Theorem 3.1 *If $\beta \leq \frac{1}{\mu} < (1+\frac{1}{2})\alpha$ then, given $\mathbf{y} = \mathbf{M}\mathbf{x}+\mathbf{e}$ where $\mathbf{x}$ is ℓ-cosparse, after a controlled number of iterations, i^*, we have*

$$\|\mathbf{x}-\hat{\mathbf{x}}_{i^*}\|_2^2 \leq c_1\|\mathbf{e}\|_2^2, \tag{3.2}$$

where c_1 is a function of α, μ and an accuracy factor.

Near optimal projections. Unfortunately, calculating the optimal cosparse projection seems to be a hard problem due to the *a priori* combinatorial search for the best cosparse subspace $\mathscr{W}_{\boldsymbol{\Lambda}}$. Can we do it more efficiently ? This remains an open question, and in practice, we can merely hope to achieve a sub-optimal projection. Denoting $\mathbf{x}^* = P_{\mathscr{A}}\mathbf{x}$ the vector resulting from the optimal projection and $\boldsymbol{\Lambda}^*$ the cosupport of $\mathbf{z}^* = \boldsymbol{\Omega}\mathbf{x}^*$, we say that P is near optimal with constant C if for any vector $\mathbf{x}$ it obeys:

$$\|P\mathbf{x}-\mathbf{x}\|_2^2 \leq C \inf_{\tilde{\mathbf{x}}\in\mathscr{A}} \|\tilde{\mathbf{x}}-\mathbf{x}\|_2^2 = C\|\mathbf{x}^*-\mathbf{x}\|_2^2. \tag{3.3}$$

[1] This is the analysis counterpart to the $\mathbf{D}$-RIP property for the matrix $\mathbf{M}$ introduced in [3], where recovery guarantees have been developed for the solution of (1.5) *under the sparse synthesis model* with the transposed dictionary $\mathbf{D} = \boldsymbol{\Omega}^T$, in the case that $\boldsymbol{\Omega}^T$ is a tight frame. Hence, (3.1) will be called the $\boldsymbol{\Omega}$-RIP for $\mathbf{M}$.

Near optimality of proposed projections. The practical projection used in the first proposed algorithm is near optimal when $\mathbf{\Omega}$ obeys yet another RIP-like property:

$$L_\ell \|v\|_2^2 \leq \|\mathbf{\Omega}_\mathbf{\Lambda} v\|_2^2 \leq U_\ell \|v\|_2^2 \tag{3.4}$$

for all $v \in \text{range}(\mathbf{\Omega}_\mathbf{\Lambda}^T)$ and every $\mathbf{\Lambda}$ of size ℓ.

Lemma 3.2 *For $\mathbf{\Omega}$ that satisfies Condition* (3.4)*, the selection of the cosupport $\hat{\mathbf{\Lambda}}$ according to* (2.2)*, followed by a projection on the subspace $\mathscr{W}_{\hat{\mathbf{\Lambda}}}$ according to* (2.3) *provides a projection which is near optimal with constant $C = U_\ell/L_\ell$.*

Proof: Since $\|\mathbf{\Omega}_{\hat{\mathbf{\Lambda}}}\mathbf{x}\|^2 = \|\mathbf{\Omega}_{\hat{\mathbf{\Lambda}}}\mathbf{\Omega}_{\hat{\mathbf{\Lambda}}}^\dagger\mathbf{\Omega}_{\hat{\mathbf{\Lambda}}}\mathbf{x}\|_2^2 = \|\mathbf{\Omega}_{\hat{\mathbf{\Lambda}}}\mathbf{P}_{\hat{\mathbf{\Lambda}}}\mathbf{x}\|_2^2 = \|\mathbf{\Omega}_{\hat{\mathbf{\Lambda}}}(\mathbf{Id}-\mathbf{Q}_{\hat{\mathbf{\Lambda}}})\mathbf{x}\|_2^2$, by the RIP (3.4) and the definition of $\hat{\mathbf{\Lambda}}$, $L_\ell\|\mathbf{Q}_{\hat{\mathbf{\Lambda}}}\mathbf{x}-\mathbf{x}\|_2^2 \leq \|\mathbf{\Omega}_{\hat{\mathbf{\Lambda}}}\mathbf{x}\|_2^2 \leq \|\mathbf{\Omega}_{\mathbf{\Lambda}^*}\mathbf{x}\|_2^2$. In a similar way, $\|\mathbf{\Omega}_{\mathbf{\Lambda}^*}\mathbf{x}\|^2 = \|\mathbf{\Omega}_{\mathbf{\Lambda}^*}(P_{\mathscr{A}}\mathbf{x}-\mathbf{x})\|_2^2 \leq U_\ell\|P_{\mathscr{A}}\mathbf{x}-\mathbf{x}\|_2^2$. By combining the lower and upper bounds we get that $\|\mathbf{Q}_{\hat{\mathbf{\Lambda}}}\mathbf{x}-\mathbf{x}\|_2^2 \leq \frac{U_\ell}{L_\ell}\|P_{\mathscr{A}}\mathbf{x}-\mathbf{x}\|_2^2$,

Recovery guarantees with near optimal projections. While the existing result (Theorem 2 from [1]) is valid for optimal projections, it can be extended to near optimal projections. The detailed proof of the following result is too long to fit in this conference paper and will appear in another place, hence we only state it as a conjecture.

Conjecture 3.3 *Consider $\mathbf{y} = \mathbf{M}\mathbf{x} + \mathbf{e}$, where $\mathbf{x}$ is ℓ-cosparse and the projection is near optimal with constant C. Let $\sigma_\mathbf{M}$ be the maximal singular value of $\mathbf{M}$, α and β the constant of the $\mathbf{\Omega}$-RIP for $\mathbf{M}$. For the noiseless case $\mathbf{e} = 0$, if $\beta < \frac{1}{\mu} < (1+\sqrt{1-b})\alpha$, where $b = \frac{(C-1)\sigma_M^2}{\alpha C}$, then the i-th iteration of A-IHT and A-HTP with a constant step size μ satisfies*

$$\|\mathbf{x}-\hat{\mathbf{x}}_{i+1}\|^2 \leq c_3\|\mathbf{x}-\hat{\mathbf{x}}_i\|_2^2, \tag{3.5}$$

where $c_3 < 1$, hence the iterates converge geometrically.

The existence of a stepsize leading to convergence is guaranteed if $\beta < (1+\sqrt{1-b})\alpha$. This is only possible if $\beta < 2\alpha$. Vice-versa, if $\beta < 2\alpha$, this holds true provided a near optimal projection is achievable with a constant $C \geq 1$ sufficiently close to one. Note that for an orthogonal $\mathbf{\Omega}$, the analysis model coincides with the synthesis one. In terms of standard RIP for $\mathbf{M}\mathbf{\Omega}^T$, the condition $\beta < 2\alpha$ reads $\delta_{2k}(\mathbf{M}\mathbf{\Omega}^T) < 1/3$, and coincides with the results of IHT with a constant step size [7].

4. EXPERIMENTAL RESULTS

To assess their performance, the proposed algorithms were tested on synthetic cosparse signal recovery problems. For a fixed $\mathbf{\Omega} \in \mathbb{R}^{p\times d}$, $\mathbf{M} \in \mathbb{R}^{m\times d}$, a noise level ε, and a cosparsity ℓ, a cosparse signal recovery problem is constructed as follows: A random index set $\mathbf{\Lambda}$ of size ℓ of $[\![1,p]\!]$ is generated. An orthonormal basis $\mathbf{B}$ for $\mathbf{Null}(\mathbf{\Omega}_\mathbf{\Lambda})$ is computed. An i.i.d. Gaussian random vector $\mathbf{c}$, of the same length as the number of columns of $\mathbf{B}$, is generated. Finally, a target cosparse signal $\mathbf{x}_0$ is formed as $\mathbf{Bc}$. The observation vector is given by $\mathbf{y} = \mathbf{M}\mathbf{x}_0 + \mathbf{e}$ where $\mathbf{e}$ is an i.i.d. random Gaussian vector with norm $\varepsilon\|\mathbf{M}\mathbf{x}_0\|_2$. The quintet $(\mathbf{\Omega}, \mathbf{M}, \mathbf{y}, \varepsilon, \ell)$ constitutes an instance of the cosparse signal recovery problem and is given as an input to the algorithms.

For both A-IHT and A-HTP, we employed a constant step size[2] μ chosen based on the ideas from section 3. The stopping criteria (cf Algorithm 1) was $i_{\max} = 500$ and $\delta_{\text{term}} = 10^{-6}$. The target cosparsity[3] $\tilde{\ell}$ is specific to each experiment.

4.1 Robustness against noise

To assess the behavior of the algorithms against different levels of noise, we chose the setting $d = 200$, $p = 240$, $m = 160$, and $\ell = 180$ and varied the noise level in the range from (almost) 0 to 0.5. We generated 20 instances of cosparsity recovery problems as described at the beginning of this section, and fed the algorithms. The average relative error in estimating $\mathbf{x}$ was computed. For both A-IHT and A-HTP, the target cosparsity level was set to $\tilde{\ell} = 0.9\ell$. For comparison with existing cosparse recovery algorithms, we run the same experiment for the GAP [11] and the analysis ℓ_1-minimization (1.5) with $\varepsilon = 0$. For solving the analysis ℓ_1 optimization problem, we used the Matlab `cvx` package [8] with highest precision and the final solution was debiased.

Figure 1 shows the results. Note that neither GAP nor ℓ_1-minimization (in this experiment) are noise aware, which may explain their poorer performance compared to the proposed algorithms when the input SNR is below 25dB. Still, the comparatively better robustness of the proposed algorithms against noise remains remarkable, given that they do not use any explicit knowledge of the noise level.

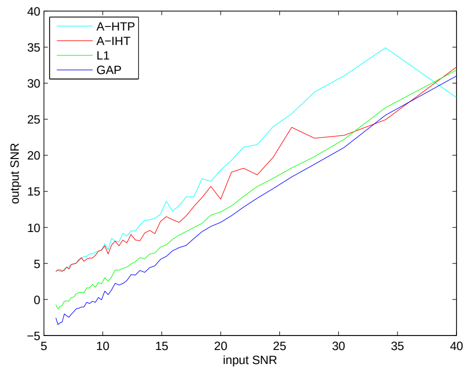

Figure 1: Output SNR $20\log_{10}(\|\mathbf{x}_0\|_2 / \|\hat{\mathbf{x}} - \mathbf{x}_0\|_2)$ as a function of the input SNR $-20\log_{10}(\varepsilon)$ for the tested algorithms.

4.2 Performance for cosparse signal recovery

Next we assessed the algorithm performance in the task of cosparse signal recovery, when the number of measurements and the cosparsity vary. We fixed $d = 200$ and $p = 240$, and varied the number m of measurements and the cosparsity ℓ according to the formulae: $\delta = m/d$, $\rho = (d-\ell)/m$ in order to obtain phase transition diagrams for successful recovery. For each pair of parameters $0 < \delta, \rho < 1$ on a grid, the following was repeated 50 times: $\mathbf{\Omega}$ was drawn as the transpose of a random tight frame of size $d \times p$; $\mathbf{M}$ was

[2] For μ_i given by (2.5), the results (not shown in this paper) are similar.

[3] we actually use $\max(\tilde{\ell}, d-m)$ as the target cosparsity to ensure uniqueness.

drawn from the Gaussian ensemble; a cosparse signal recovery problem with $\varepsilon = 0$ was constructed and fed to the algorithms. A relative error less than 10^{-4} was counted as a perfect recovery. For both A-IHT and A-HTP, two variants were implemented, one with a target cosparsity $\tilde{\ell} = \ell$, the second one with $\tilde{\ell} = 0.75\ell$. The GAP and the analysis ℓ_1-minimization were again used for comparison.

Figures 2-3 show the resulting phase transition diagrams. Each pixel indicates the percentage of the 50 cosparse signals perfectly recovered at the corresponding setting (δ, ρ), from black (100% failure) to white (100% success). The GAP algorithm and analysis ℓ_1-minimization yield the largest white areas, showing the best recovery performance of all tested algorithms. However, A-HTP was informally observed to perform the recovery faster than the other algorithms *when* it succeeded.

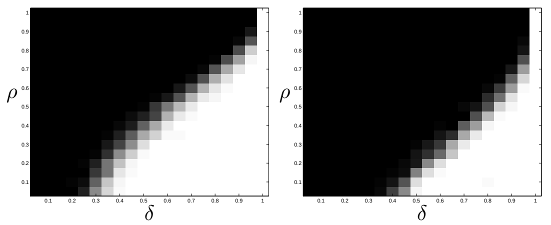

Figure 2: Recovery rate of GAP (left) and analysis ℓ_1-minimization (right).

The effect of the target cosparsity level $\tilde{\ell}$ can also be seen in Figure 3. Namely, we clearly observe the benefit of targeting cosparsity 0.75ℓ rather than ℓ. Let us point out, however, that the usage of too low $\tilde{\ell}$ leads to complete failure in terms of recovery (result not shown). Therefore, $\tilde{\ell}$ is best seen to be chosen slightly smaller than ℓ.

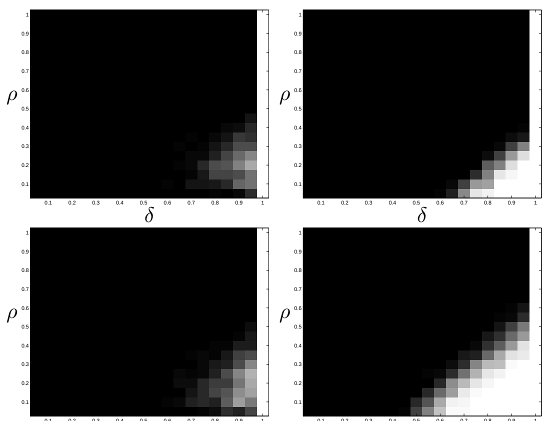

Figure 3: Recovery rate of proposed algorithms. Top: A-IHT; Bottom: A-HTP; Left: $\tilde{\ell} = \ell$; Right: $\tilde{\ell} = 0.75\ell$.

5. CONCLUSION

This paper introduced a family of new iterative algorithms for stably recovering/approximating cosparse signals from low-dimensional projections. These algorithms are the analogous, in a cosparse modeling framework, of the IHT and HTP algorithms for sparse signal recovery. This work extends the result of [1] in the cosparse analysis model framework by allowing the use of sub-optimal projections for the general class of Iterative Projection Algorithms (IPA) described in [1]. We have provided a tentative theoretical foundation for such sub-optimal projections, hence enabling the *implementation* of two classes of algorithms that can be *proven* to converge and recover sufficiently cosparse signals in the noiseless setting. Moreover, the robustness of these algorithms against noise and their effectiveness in recovering cosparse signals was demonstrated by experimental evidence. Further investigation of performance guarantees of the algorithms in the presence of noise, their computational complexity, and the effect of cosparsity levels and gradient step size are ongoing work.

REFERENCES

[1] T. Blumensath. Sampling and reconstructing signals from a union of linear subspaces. *CoRR*, http://arxiv.org/abs/0911.3514, 2009.

[2] T. Blumensath and M. E. Davies. Iterative hard thresholding for compressed sensing. *Appl. Comput. Harm. Anal.*, 2009.

[3] E. Candès, Y. Eldar, D. Needell, and P. Randall. Compressed sensing with coherent and redundant dictionaries. *Appl. Comput. Harm. Anal.*, In Press, 2010.

[4] V. Cevher. On accelerated hard thresholding methods for sparse approximation. Technical report, EPFL, 2011.

[5] D. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization. *Proc. Nat. Aca. Sci.*, 100(5):2197–2202, Mar. 2003.

[6] M. Elad, P. Milanfar, and R. Rubinstein. Analysis versus synthesis in signal priors. *Inverse Problems*, 23(3):947–968, June 2007.

[7] S. Foucart. Hard thresholding pursuit: an algorithm for compressive sensing. submitted, 2010.

[8] M. Grant, S. Boyd, and Y. Ye. CVX: Matlab Software for Disciplined Convex Programming, August 2008.

[9] R. Gribonval and M. Nielsen. Sparse representations in unions of bases. *IEEE Trans. Inform. Theory*, 49(12):3320–3325, Dec. 2003.

[10] Y. Lu and M. Do. A theory for sampling signals from a union of subspaces. *Signal Processing, IEEE Transactions on*, 56(6):2334 –2345, June 2008.

[11] S. Nam, M. Davies, M. Elad, and R. Gribonval. Cosparse analysis modeling - Uniqueness and algorithms. In *Proc. ICASSP 2011.*, Prague, Czech Republic, May 2011.

[12] Y. Pati, R. Rezaiifar, and P. Krishnaprasad. Orthonormal matching pursuit : recursive function approximation with applications to wavelet decomposition. In *Proceedings of the* 27^{th} *Annual Asilomar Conf. on Signals, Systems and Computers*, Nov. 1993.