16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP

SPARSE REPRESENTATIONS: RECOVERY CONDITIONS AND FAST ALGORITHM FOR A NEW CRITERION

Jean-Jacques Fuchs

IRISA / Université de Rennes 1
Campus de Beaulieu - 35042 Rennes Cedex - France
phone: + (33) 299847223, email: fuchs@irisa.fr
web: www.irisa.fr/temics/staff/fuchs/

ABSTRACT

Most applications of sparse representations are based on a combined $\ell_2-\ell_1$ criterion, where the least-squares-part ensures closeness to the observations and the ℓ_1-part sparsity. This choice leads to quite efficient algorithms and has a clear connection to maximum likelihood approaches in case of additive Gaussian noise. We replace the least-squares-part by a ℓ_1-part and investigate the recovery conditions of the so-obtained $\ell_1 - \ell_1$ criterion. We then propose an algorithm, that minimizes the criterion, in a finite number of steps.

1. INTRODUCTION

"Sparse representation" is a technique that consists in decomposing a signal (a finite dimensional vector) into a small number of components (vectors) chosen from a user-designed over-complete set called the dictionary. It is mostly used to obtain a simple approximate model of a complex signal for compression or coding purposes in audio or video signal processing [1, 2], but theoretical investigations tend to extend its applicability to a variety of new domains, as for instance, compressed sensing or compressed sampling, in which one investigates the possibility to sample a signal at a rate much lower than the Nyquist rate with a controlled loss in information [3, 4].

The current interest has been initiated in [5] but earlier investigations had been proposed in different areas, [6, 7, 8].

The following problem is considered in [5], and later generalized, in e.g. [9, 10]. Given a $n\times m$ matrix A with $m \gg n$ and a vector b that indeed admits an exact sparse representation, say $b = Ax_o$, with x_o having just a few non-zero components, when is it possible to recover x_o? It is shown that, if the number of non-zero entries in x_o is smaller than a given bound, then x_o is indeed the unique sparsest representation. It is also established that one can replace the exhaustive search for the sparsest solution by the easy to solve linear program

$$\min_x \ \|x\|_1 \qquad \text{s.t.} \qquad Ax = b, \tag{1}$$

while keeping similar bounds on the number of non-zero entries in x_o. But seeking the sparsest exact representation may be useless, either because there is none, or there is one, but observed in additive noise. An approximate reconstruction is often preferable and it then makes sense to replace (1) by [10]

$$\min_x \ \|x\|_1 \quad \text{s.t.} \quad \|Ax - b\|_2^2 \le \rho^2, \tag{2}$$

with ρ^2 the tolerance to be defined, or, somehow equivalently, by the following criterion

$$\min_x \ \frac{1}{2}\|Ax - b\|_2^2 + h\|x\|_1, \quad h > 0 \ . \tag{3}$$

One seeks the representation with smallest ℓ_1-norm, that yields an approximation error smaller than a specified threshold. This criterion is the most often considered currently and fast dedicated algorithms that are quite efficient and thus allow to handle problems of large dimensions, have been developed [12, 15, 11]. In the sequel, we propose to replace the criterion (2) by

$$\min_x \ \|x\|_1 \qquad \text{s.t.} \qquad \|Ax - b\|_1 \le \rho, \tag{4}$$

where in the constraint that bounds the reconstruction error the ℓ_2-norm is replaced by an ℓ_1-norm [14]. While the ℓ_2-norm on the residual vector $r = Ax - b$ is clearly associated with Gaussian noise, the ℓ_1-norm is associated with Laplace or double-exponential noise. This optimization problem is convex and can be transformed into a linear program, but we will handle it in a different way, to get the recovery conditions that will tell us under which conditions is it possible to recover x_o from the optimum of (4) with $b = Ax_o$ and to develop an optimization algorithm that converges in a number of steps, possibly much smaller than the number of steps required by, say, the simplex algorithm for linear programs.

2. THE CRITERION

2.1 Preliminary remarks

We consider the following problem:

$$\min_x \ \|x\|_1 \quad \text{subject to} \quad \|Ax - b\|_1 \le \rho,$$

where $\rho > 0$ has to be fixed by the user. It is instructing to transform the problem into a linear program by introducing slack variables. The standard idea is to split the component x_i of x into $x_i^+ = \max(x_i, 0)$, $x_i^- = \max(-x_i, 0)$ and to replace x_i by $x_i^+ - x_i^-$ and $|x_i|$ by $x_i^+ + x_i^-$ and more generally x by $x^+ - x^-$ and $\|x\|_1$ by $\mathbf{1}^T(x^+ + x^-)$ with $\mathbf{1}$ a vector of 1's of adequate dimension. Applying the same kind of operation to the constraint, one gets the following program

$$\min_{x^+, x^-, t^+, t^-, w} \qquad \mathbf{1}^T(x^+ + x^-) \tag{5}$$

$$\begin{aligned} \text{s.t.} \qquad & A(x^+ - x^-) - b = t^+ - t^-, \\ & \mathbf{1}^T(t^+ + t^-) + w = \rho, \\ & x^+,\ x^-,\ t^+,\ t^-, w\ \geq 0. \end{aligned}$$

Since this linear program in standard form has $2m+2n+1$ variables and $n+1$ equality constraints, its optimum is generically attained at a point having $n+1$ nonzero components and if p of these components belong to x, the remaining $n+1-p$ components belong to t, since w is obviously equal to zero at the optimum. This is turn tells us that Ax is equal to b at $p-1$ locations, i.e., for $p-1$ indexes.

The dual of this linear program is now straightforward to obtain and rewriting it in a compact way one gets

$$\max_{d,\delta}\ b^T d + \rho\delta \quad \text{s.t.}\ \ \|A^T d\|_\infty \leq 1, \quad \delta \leq -\|d\|_\infty$$

which leads to

$$\max_{d} b^T d - \rho_1 \|d\|_\infty \quad s.t.\ \ \|A^T d\|_\infty \leq 1 \qquad (6)$$

It is not a surprise that ℓ_1-norms in the primal lead to ℓ_∞-norms in the dual. Remember that in the ℓ_p-norms context, dual norms are such that $\frac{1}{p} + \frac{1}{q} = 1$, thus ℓ_1 and ℓ_∞ are dual norms.

2.2 Optimality conditions

In order to be able to characterize easily the conditions satisfied by the optimum of (4), we introduce the sub-differential of a convex function f at a point x, denoted $\partial f(x)$ [18]. It is useful for functions that are not continuously differentiable at all points, as for instance $\|x\|_1$ when x has zero components. It is a set of vectors called the sub-gradients of f at x and the set reduces to the gradient at points where f is differentiable. For $f(x) = \|x\|_p$ one has

$$\begin{aligned} \partial\ \|x\|_1 &= \{u|\ u_i = \text{sign}(x_i) \text{ if } x_i \neq 0, \\ &\qquad\qquad |u_i| \leq 1 \text{ otherwise}\} \\ \partial\ \|x\|_\infty &= \{v|\ |x_i| = \|x\|_\infty \Rightarrow x_i v_i \geq 0, \\ &\qquad |x_i| < \|x\|_\infty \Rightarrow v_i = 0, \\ &\qquad \|v\|_1 = 1 \text{ if } x \neq 0, \|v\|_1 \leq 1 \text{ else}\} \end{aligned} \qquad (7)$$

where $\text{sign}(x_i) = 1$ if $x_i > 0$ and $\text{sign}(x_i) = -1$ if $x_i < 0$.

It is now easy to obtain the optimality conditions for (4). Since this problem is convex, the first order necessary optimality conditions are also sufficient. The Lagrangian of the primal is

$$\ell(x, \mu) = \|x\|_1 + \mu(\|Ax - b\|_1 - \rho), \quad \mu \geq 0$$

and the optimality conditions are thus

$$u' + \mu A^T w = 0, \text{ with } u' \in \partial\|x\|_1,\ w \in \partial\|Ax - b\|_1,\ \mu \geq 0.$$

One gets equivalent conditions that are in a form that is more suited for our later use when one considers both the primal (4) and the dual (6). One has the following theorem.

Theorem 1. The optima of (4) and (6) are respectively x and d if and only if

$$\begin{aligned} & Ax - b = -\rho v \quad \text{and} \quad A^T d = u \\ & \text{for some}\ \ u \in \partial\|x\|_1 \text{ and } \ v \in \partial\|d\|_\infty \quad \diamond \end{aligned} \qquad (8)$$

Proof: The proof is immediate. Both points x and d are feasible and lead to identical costs. $\diamond$

One can of course verify that these conditions are equivalent to those developed just above using the Lagrangian.

We will use the two relations in (8) to both obtain recovery conditions and develop the announced iterative algorithm.

2.3 Some specific notations

Partitioning will play an important role in the sequel and we now introduce the, somehow awkward, notations that we will use. We will split or partition the optimum x, of dimension m, into its non-zero components, we denote $\bar{x}$, and its zero components $\bar{\bar{x}}$, and partition accordingly (the columns in) A into $\bar{A}$ and $\bar{\bar{A}}$. It then follows that, for instance, $Ax = \bar{A}\bar{x}$ or from (7), that the sub-gradient $u \in \partial\|x\|_1$ is such that $\bar{u} = \text{sign}(\bar{x})$ and $\|\bar{\bar{u}}\|_\infty \leq 1$.

We will also introduce d-induced partitions of the rows of d, v, A and b. We have seen above that, if the optimal x has p nonzero components, then $Ax - b$ has p - 1 zero components which tells us that v in (8) has $p-1$ zero components. This in turn, see (7), implies that generically the optimal d has n - $p+1$ components equal to $\pm\|d\|_\infty$ and the $p-1$ remaining components that are smaller in absolute value. We thus partition the optimal d into $\underline{d}$ of dimension $p-1$, which satisfies $\|\underline{d}\|_\infty \leq \|d\|_\infty$ and $\underline{\bar{d}} = \|d\|_\infty \text{sign}(\underline{d})$. We partition accordingly v into $\underline{v} = 0$ and $\underline{\bar{v}}$ which satisfies $\|\underline{\bar{v}}\|_1 = 1$ and $\underline{\bar{v}}^T \underline{\bar{d}} = \|d\|_\infty$.

The matrix A will thus be partitioned into four blocks with, e.g., $\underline{A}$ of dimension $(p-1, p)$.

3. RECOVERY CONDITIONS

We are interested in conditions under which the optimum of (4) with $b = Ax_o$ and ρ small enough allows to recover x_o.

The optimum x of (4) will not be equal to x_o for $\rho > 0$. What one asks for, is that its sub-gradient u satisfies $\bar{u} = \text{sign}(\bar{x}_o)$ and $\|\bar{\bar{u}}\|_\infty < 1$, this guaranties that the optimum x and x_o have their nonzero components at the same locations and with the same signs. We will prove that, if the conditions in Theorem 2 below are satisfied then, for ρ sufficiently small, the optimum x of (4) is of the form $x = x_o - \rho z$ with $\bar{x}(\rho) = \bar{x}_o - \rho\bar{z}$ and $\bar{\bar{z}} = 0$, for some vector z to be defined below.

Theorem 2. The solution x_o of $Ax = b$ with $b = Ax_o = \bar{A}_o\bar{x}_o$ and $\bar{A}_o$ a full-rank matrix, can be recovered from the unique optimum point $x(\rho)$ of (4),

for ρ sufficiently small, if there exists

$$\delta = \arg\min_{\delta} \|\delta\|_\infty \;\text{ s.t. }\; \bar{A}_o^T\delta = \text{sign}(\bar{x_o})$$
$$\text{that satisfies} \quad \|\bar{\bar{A}}_o^T\delta\|_\infty < 1 \qquad \diamond (9)$$

Notice the x_o-induced partition of the columns of A that leads to $\bar{A}_o$, for instance, to be distinguished from $\bar{A}$.

Proof: If there is a vector δ that satisfies (9), this same δ is an optimum of

$$\min_{\delta} \|\delta\|_\infty \;\text{ s.t. }\; \bar{A}_o^T\delta = \text{sign}(\bar{x_o}),\; \|\bar{\bar{A}}_o^T\delta\|_\infty \leq 1. \quad (10)$$

One can then establish that the dual of (10) is

$$\max_{z} \text{sign}(\bar{x_o})\bar{z} - \|\bar{\bar{z}}\|_1 \quad \text{s.t. } \; \|\bar{A}_o\bar{z} + \bar{\bar{A}}_o\bar{\bar{z}}\|_1 \leq 1 \quad (11)$$

where we have partitioned the vector z of dimension m as $z = [\bar{z}^T \;\; \bar{\bar{z}}^T]^T$ using the x_o-induced partition. We shall see later that $\bar{\bar{z}} = 0$ which justifies a posteriori this notation.

The Lagrange dual (11) of (10) can be obtained by transforming these problems into linear programs, for instance.

Remember now that the optimum z of the dual are the Lagrange multipliers associated with the constraints of the primal. At the optimum, $\bar{z}$ is thus to be associated with $\bar{A}_o^T\delta = \text{sign}(\bar{x_o})$ and $\bar{\bar{z}}$ with $\|\bar{\bar{A}}_o^T\delta\|_\infty \leq 1$.

With the optimum δ of (10) that satisfies (9) is then associated an optimum z of the dual that is such that $\bar{\bar{z}} = 0$ because the second set of constraints in the primal are strictly satisfied.

Choosing then these optimal δ and z that satisfy (9) and lead to identical costs in (10,11), we construct x, d and their associated sub-gradients that satisfy (8).

We take d in (8) equal to δ in (10) and propose to define u as $u = A^T\delta$. From the constraints in (10) it follows that this u is indeed such that $\bar{u} = \text{sign}(\bar{x_o})$ and $\|\bar{\bar{u}}\|_\infty < 1$ which are the properties required for a sub-gradient of $\|x\|_1$ for an x that has the same structure as x_o .

We further define $v = Az = \bar{A}_o\bar{z}$ and $x = x_o - \rho z$ and establish that they satisfy (8). We know already that z and x_o have the same partition which is thus also valid for x.

Premultiplying $\underline{x} = x_o - \rho z$ by A we get $Ax = b - \rho v$ with $v = Az = \bar{A}_o\bar{z}$ which has then the properties required for a sub-gradient in $\partial\|d\|_\infty$. One can, for instance, observe that the dual can a posteriori be rewritten $\max \|\bar{z}\|_1$ subject to $\|\bar{A}_o\bar{z}\|_1 \leq 1$ which attains generically its optimum at a point where $\bar{A}_o\bar{z}$ is zero at $p-1$ locations. $\diamond$

We are reached our goal, we have obtained the conditions (9) under which x_o can be recovered from the solution of (4).

To summarize, we have shown that if Theorem 2. is satisfied, the optimum x of (4) can be written $x = x_o - \rho z$ with z the optimum of (11) that admits the same partition as x_o. It follows that x_o can be *recovered* from x, for ρ sufficiently small.

As opposed to the recovery conditions one gets for the ℓ_2-norm (2,3), for which the equivalent of (9), admits a explicit solution [10], that can be further transformed into explicit conditions on the sparsity of x_o, no such miracle happens for the ℓ_1-norm, since the optimum of (9) has no explicit analytical expression.

For completeness let us add that it has been shown in [10] that (1) allows to recover x_o provided there is a vector d that satisfies $\bar{A}_o^T d = \text{sign}(\bar{x_o})$ and $\|\bar{\bar{A}}_o^T d\|_\infty < 1$. For the present criterion these conditions have to be satisfied by a d that has minimal ℓ_∞-norm.

4. OPTIMIZATION ALGORITHM

4.1 Introduction

The solution of (4) can be obtained, for instance, applying the simplex algorithm to (4) rewritten as a linear program. We propose to obtain it using an algorithm based on the two relations in (8) that solves (4) in a finite number of steps and is more efficient than the standard linear program solvers.

Due to the presence of u and v, which belong to sets, the two relations in (8) are far from defining the optimal x and d. They nevertheless carry a lot of information, that is helpful if one is interested in the way the optimal x and d vary locally with ρ. More precisely, if the optima are known for a given value of ρ, they implicitly define how they vary in the neighborhood of this ρ and also carry enough information to precisely locate the boundaries of the neighborhood, i.e. the interval in ρ, on which the optima can be extended.

As a matter of fact, it is then also possible to cross these boundary, i.e., to propagate the optima to the next interval. This is the idea that is used to develop the algorithm. It remains only to initialize the procedure and this is easily done by starting with $\rho > \rho_0 = \|b\|_1$ for which the optimal x and d are both at zero. As ρ decreases, there is a first interval $]\ \rho_1,\ \rho_0]$, in which the optimal x has just one non-zero component, then a second interval for which it has two non-zero components, and so on. Note that the number of nonzero components in $x(\rho)$ does not necessarily increase as ρ decreases. Though we are only interested in the optimum for a given value of ρ, we will build it, for decreasing ρ, and stop when the ρ of interest is in the current interval.

4.2 Development: the standard step

Assume we have the quadruple $x,\ u,\ d,\ v$, that satisfies the optimality conditions (8) for a given ρ, we will extend it within an interval in ρ. We partition the four vectors, using the notations introduced in Section 2.3. The boundaries of the intervals are precisely the values of ρ for which these partitions need to be changed.

We denote p the number of nonzero components in the optimal x that are in $\bar{x}$, this implies that $v \in \partial\|d\|_\infty$ has $p-1$ zero components in $\underline{v}$. These two observations fully characterize the partitions described in Section 2.3.

From the second condition in (8), $A^T d = u$, it follows

$$\bar{A}^T d = \;\bar{u}\; \text{ and } \;\; \bar{\bar{A}}^T d = \;\bar{\bar{u}}$$

but since, see (7), d can be partitioned into $\underline{d}$ and

$\underline{\underline{d}}$ =sign($\underline{\underline{v}}$) $\|d\|_\infty$, one has

$$\bar{A}^T d = \bar{u} \quad \Rightarrow \underline{\underline{A}}^T \underline{d} + \|\underline{\underline{d}}\|_\infty \underline{\underline{A}}^T \text{ sign } (\underline{\underline{v}}) = \bar{u}.$$

This can be seen as a set of p equations in p unknowns in $\underline{d}$ and $\|\underline{\underline{d}}\|_\infty$ and tells us that d remains constant as ρ varies within an interval. The same thus holds for $\bar{\bar{u}} = \bar{\bar{A}}^T d$

The first condition in (8) $Ax - b = -\rho\ v$ similarly yields

$$\bar{A}\bar{x} - b = -\rho\ v \quad \Rightarrow \quad \underline{\bar{A}}\bar{x} = \underline{b} \quad \text{and} \quad \underline{\underline{\bar{A}}}\bar{x} = \underline{\underline{b}} - \rho\ \underline{\underline{v}}$$

The last of these relations premultiplied by sign($\underline{\underline{v}}$)^T leads to

$$\text{sign}(\underline{\underline{v}})^T \underline{\underline{\bar{A}}}\bar{x} = \text{sign}(\underline{\underline{v}})^T \underline{\underline{b}} - \rho$$

a linear equation which together with $\underline{\bar{A}}\bar{x} = \underline{b}$ forms a system of p equations in the p unknowns $\bar{x}$. Provided the matrix of this system is invertible, which we will assume, we get an expression of the optimal $\bar{x}$ of the form, say, $\bar{x}(\rho) = X_1 - \rho X_2$. Substituting this expression in $\underline{\underline{\bar{A}}}\bar{x} = \underline{\underline{b}} - \rho\ \underline{\underline{v}}$, one gets in turn an expression of the *optimal* $\underline{\underline{v}}$ of the form, say, $\underline{\underline{v}}(\rho) = V_1 + V_2/\rho$.

It follows that, as ρ varies, within the current interval, only $\bar{x}$ and $\underline{\underline{v}}$ are varying, the remaining (six) parts in the optimal quadruple $x,\ u,\ d,\ v$ are invariant.

As ρ varies, two remarkable events can happen: a component in $\underline{\underline{v}}(\rho)$ becomes equal to 0 or a component in $\bar{x}(\rho)$ becomes zero. The upper bound ρ_u (lower bound ρ_l) of the current interval is the ρ associated with the event that happens first when ρ is increases (decreases). We will only consider decreasing values of ρ but both events can nevertheless happen.

$\diamond$ If component i_{p+1} in $\underline{\underline{v}}(\rho)$ becomes zero first, as ρ=ρ_l. Three different changes can happen as ρ further decreases. i) The corresponding component in $\underline{\underline{d}}$ changes its sign and the partitions remain the same. ii) The corresponding component in d becomes *smaller* is moved from $\underline{\underline{d}}$ to $\underline{d}$ and another component of d moves the other way. iii) The corresponding component in d becomes *smaller* is moved from $\underline{\underline{d}}$ to $\underline{d}$ and a component of x, or more precisely $\bar{\bar{x}}$ becomes nonzero. In case i) and ii) p remains constant while in case iii) it increases by one.

$\diamond$ If, as ρ decreases, a component in $\bar{x}(\rho)$ becomes zero first for $\rho = \rho_l$, again three different changes can happen as ρ further decreases. i) One allows the component to change its sign and the partitions remain the same. ii) One sets this component to zero, moves it from $\bar{x}$ to $\bar{\bar{x}}$ and another component in x moves the other way. iii) One sets this component to zero, moves it from $\bar{x}$ to $\bar{\bar{x}}$ and moves similarly a component from $\underline{d}$ to $\underline{\underline{d}}$. In case i) and ii) p remains constant while in case iii) it decreases by one.

Whether the new boundary value is attained in $\underline{\underline{v}}(\rho)$ or in $\bar{x}(\rho)$, to test the three types of potential modifications and find the optimal one, one uses the condition $A^T d = u$. In cases i) and ii) p remains constant and only $\bar{A}^T d = \bar{u}$ intervenes, while in case iii) both $\bar{A}^T d = \bar{u}$ and $\bar{\bar{A}}^T d = \bar{\bar{u}}$ are to be used. In each case one computes the new d vector (with its characteristic structure), checks that it satisfies $\|\underline{d}\|_\infty \leq \|\underline{\underline{d}}\|_\infty$ and retains the optimal modifications which is the one for which $\|d\|_\infty$ undergoes the smallest increase. Remember that d and thus also $\|d\|_\infty$ remain constant over the whole interval and thus jumps to a new value when crossing a boundary, the optimal modification can be shown to be the one associated with the smallest increase in $\|d\|_\infty$.

We do not detail how all these potential new d-vectors are computed, the computational complexity is quite low and just three order-p matrices have to be inverted, one in each case.

There is of course, generically, only one optimal solution for any ρ and thus one optimal modification when crossing a boundary. Instead of selecting among all the valid modifications (those for which $\|\underline{d}\|_\infty \leq \|\underline{\underline{d}}\|_\infty$) the one with the smallest jump, one can also for each valid modification complement d with the associated v, x and u and stop the search as soon as the current modification satisfies (8).

It remains to detail the initialization step.

4.3 The initialization step

One can use both the primal (4) or dual form (6) of the criterion to justify the initialization step but it can also be deduced from the optimality conditions (8).

From the primal, it follows that for $\rho > \rho_0 = \|b\|_1$, the first boundary value, the optimum is at $x = 0$. For ρ slightly smaller the most efficient component in x has index $j_1 = \arg\max_j |a_j^T \text{sign}(b)|$ with a_j the j-th column of A, and has sign the sign of $a_{j_1}^T \text{sign}(b)$

In the dual, the optimum is at zero for $\rho > \rho_0$ and for ρ slightly smaller, the most efficient d is of the form $d = \alpha\ \text{sign}(b)$, where α is fixed by the most *constraining* constraint. It is constraint j_1 and the associated α is $\alpha = 1/|a_{j_1}^T \text{sign}(b)|$.

At this point one can commute to the standard procedure with p the number of nonzero components in the optimal x equal to one. More formally we have among others: $\underline{\underline{A}} = a_{j_1}$, $b = \underline{\underline{b}}$, $v = \underline{\underline{v}}$, sign($\underline{\underline{v}}$) =sign($b$). $d = \underline{\underline{d}} = \|d\|_\infty \text{sign}(\underline{\underline{v}})$ and the next event to happen, that will define ρ_1 , the next boundary value, is a component in $\underline{\underline{v}}$ that becomes zero as $\bar{x}(\rho) = x_{j_1}(\rho)$ *increases* since the other event is not possible when $p = 1$.

4.4 Relations to previous works

Several recent papers have proposed similar path-following methods for solving (3), [12, 15, 11]. All these methods are related to continuation techniques, which have also been studied in the optimization literature [13]. When the solution is sparse, i.e. when the (unknown) optimum has just a few non-zero components, they are indeed very fast but their computational complexity increases more than linearly in the number of non zero components in the optimum. To our knowledge, however, no such algorithms have been proposed for the criterion (4) except for some preliminary remarks in [14].

5. CONCLUDING REMARKS

We have considered the criterion

$$\min_x \|x\|_1 \quad \text{subject to} \quad \|Ax - b\|_1 \leq \rho,$$

that is known as a ℓ_1- ℓ_1 penalized criterion. We have indicated how it is possible to use it in a sparse representation context where the A-matrix has far more columns than rows and presented so-called recovery conditions. We have also shown how to construct a fast algorithm that minimizes it. While the criterion can be transformed into a linear program and thus solved using standard subroutines, our approach is indeed faster if the number of non-zero component in the optimum is small, corresponding to so-called degenerate linear programs. So far we have applied this criterion to image denoising and coding and in both cases, it appears that, at least for the preliminary investigations we performed, the results are less good than those obtained using the standard ℓ_1- ℓ_2 penalized criterion

$$\min_x \frac{1}{2}\|Ax - b\|_2^2 + h\|x\|_1, \quad h > 0 \ .$$

We now plan to investigate its applicability in decoding linear codes. Linear programming and sparse representations approaches have already been considered in this context [16, 17]. In this application it is more the recovery conditions obtained for the criterion that might be of interest.

REFERENCES

[1] P. Frossard, P. Vandergheynst, R. M. Figueras i Ventura, and M. Kunt, "A Posteriori Quantization of Progressive Matching Pursuit Streams" *IEEE-T-SP*, vol. 52, 2, pp. 525-535, 2004.

[2] R.Neff and A. Zakhor, "Very low bit-rate video coding based on matching pursuit video coder", *IEEE-T-CSVT*, vol. 7, 1, Feb. 1997.

[3] D. Donoho, "Compressed sensing." *IEEE-T-IT*, 52, 1289-1306, 2006.

[4] E. Candes and T. Tao, "Near optimal signal recovery from random projections: Universal coding strategies ?" *IEEE-T-IT*, 52, 5406-5425, 2006.

[5] D.L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. *IEEE-T-IT*, 47, 11, 2845-2862, 2001.

[6] S. Mallat and Z. Zhang, "Matching Pursuit with time-frequency dictionaries," *IEEE-T-SP*, 41, 3397-3415, 1993.

[7] J.J. Fuchs. Extension of the Pisarenko method to sparse linear arrays *IEEE-T-SP*, 45, 2413-2421, Oct. 1997.

[8] M.D.Sacchi, T.J. Ulrych and C.J. Walker. Interpolation and extrapolation using a high-resolution discrete Fourier transform. *IEEE-T-SP*, 46, 1, 31-38, Jan. 1998.

[9] R. Gribonval and M. Nielsen. Sparse representations in unions of bases. *IEEE-T-IT* 49, 12, 3320-3325, Dec. 2003.

[10] J.J. Fuchs. More on sparse representations in arbitrary bases. *IEEE-T-IT* 50, 6, 1341-1344, June 2004.

[11] S. Maria and J.J. Fuchs, "Application of the Global Matched Filter to STAP data: an efficient algorithmic approach." In *Proceedings ICASSP*, Toulouse, may 2006.

[12] M. Osborne, B. Presnell and B. Turlach. "A new approach to variable selection in least squares problems" *IMA J. of Numerical Analysis*, 20, 3,389–403, 2000.

[13] E. Allgower and K. Georg. "Continuation and path following" *Acta Numerica*, 2, 31-64, 1993.

[14] J.J. Fuchs. "Some further results on the recovery algorithms." In *Proceedings of SPARS'05*, Rennes, France, Nov. 2005.

[15] B. Efron, T. Hastie, I. Johnstone and R.Tibshirani, "Least angle regression," *Annals of Statistics*, 32, 407-499, Apr. 2004.

[16] E. Candes and T. Tao, "Decoding by linear programming" *IEEE-T-IT*, 51, 4203-4215, 2005.

[17] J. Feldman, M.J. Wainwright and D.R. Karger, "Using linear programming to decode binary linear codes" *IEEE-T-IT*, 51, 954-972, 2005.

[18] R. Fletcher, Practical methods of optimization. *Wiley*, 1991.