On sparse representations in arbitrary redundant bases

Jean-Jacques FUCHS
IRISA/Université de Rennes I
Campus de Beaulieu - 35042 Rennes Cedex - France
e-mail: fuchs@irisa.fr.

submitted to IEEE-I-IT nov. 2002, accepted for publication.

Abstract

The purpose of this contribution is to generalize some recent results on sparse representations of signals in redundant bases. The question that is considered is the following: given a matrix A of dimension (n, m) with $m > n$ and a vector $b = Ax$, find a sufficient condition for b to have a unique sparsest representation x as a linear combination of columns of A.

Answers to this question are known when A is the concatenation of two unitary matrices and either an extensive combinatorial search is performed or a linear program is solved. We consider arbitrary A matrices and give a sufficient condition for the unique sparsest solution to be the unique solution to both a linear program or a parametrized quadratic program. The proof is elementary and the possibility of using a quadratic program opens perspectives to the case where $b = Ax + e$ with e a vector of noise or modeling errors.

Index terms—Basis pursuit, global matched filter, linear program, quadratic program, redundant dictionaries, sparse representations.

1 Introduction

Let us consider a set of m n-dimensional vectors a_j with $m > n$ and let us denote A the (n, m) matrix having these vectors as columns. Any linear combination b of these m vectors can then be written as $b = Ax$ with x an m-dimensional vector of weights.

If x has just a few non-zero components it may well be the unique and sparsest representation and to recover it from the knowledge of b one could then seek the sparsest among all the solutions to $Ax = b$. The aim of this contribution is to give conditions under which this is feasible in reasonable time.

Finding the solution having the smallest possible number of non-zero components, i.e., solving the following optimization problem

$$\min_x \ \|x\|_0 \quad \text{subject to} \quad Ax = b\,, \tag{P_0}$$

where $\|x\|_0$ denotes the number of non-zero components in x, is a difficult problem that can only be solved using a combinatorial approach, i.e., testing systematically all the potential combinations of columns. This approach is thus unfeasible and it is usual to consider instead the following much simpler optimization problem

$$\min_x \ \|x\|_1 \quad \text{subject to} \quad Ax = b\,, \tag{LP}$$

where $\|x\|_1 = \sum |x_i|$ denotes the ℓ_1 norm of x. The problem (LP) is easily transformed into a linear program whose solution is straightforward to obtain.

The case where A is the concatenation of two square orthogonal matrices U_1 and U_2: $A = [U_1\ U_2]$, was investigated by Donoho and Huo [1], and Elad and Bruckstein [2]. It was shown that if a solution x satisfies

$$\|x\|_0 < \frac{1}{M}, \qquad \text{with} \ \ M = \sup_{1 \le i \ne j \le m} |a_i^T a_j|, \tag{1}$$

then it is the unique sparsest solution and that under the stronger condition

$$\|x\|_0 < \frac{1}{2}(1 + \frac{1}{M}) \quad \text{with} \ \ M \ \ \text{as above}, \tag{2}$$

the unique sparsest solution is also the unique minimum point of (LP). Stronger results are established in [2] for this kind of A matrices.

Below we prove that if condition (2) is satisfied by a solution x_o of $Ax = b$, with A a matrix built upon an arbitrary number of vectors with unit Euclidean norm, then it is the unique minimum point of both a parametrized quadratic program (QP) in a sense to be defined later (Theorems 2 and 3), and the linear program (LP) (Theorems 3 and 4). But this implies that x_o is also the the unique sparsest solution of $Ax = b$ since if a sparser representation existed, it would satisfy (2), the same reasoning would hold and one would arrive at a contradiction since both (QP) and (LP) are convex programs that have a unique optimum attained at a single point under (2). A similar result has independently been obtained in [3] for (LP) using a completely different approach.

To establish this result we merely apply more general but non-explicit results presented in [4, 5] to this very specific problem where the emphasis is on sparsity. The proof we present goes through if the vectors a_j are not normalized but its last part (see Section 5) would be more intricate.

2 The criterion

Instead of (LP), let us consider the following optimization problem

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

where $\|x\|_2^2 = (\sum x_i^2)^{1/2}$ denotes the ℓ_2 norm of x. This is a convex unconstrained non-smooth optimization problem that can be converted into a quadratic program [6]. Introducing the new variables $x_i^+ = \max(x_i, 0)$, $x_i^- = \max(-x_i, 0)$ and constraining them to be greater than or equal to zero, one can then replace x_i by $x_i^+ - x_i^-$ and $|x_i|$ by $x_i^+ + x_i^-$ to arrive at a standard quadratic program.

The unique global minimum of this convex optimization problem possibly attained on a convex set can thus be obtained using standard algorithms available from any scientific program library. This criterion and similar ones have been considered for a while now [5, 8, 9].

To assess the role played by h in the criterion (QP), it is useful to present the following dual of (QP)

$$\min_x \|Ax\|_2^2 \quad \text{subject to} \quad \|A^T(Ax - b)\|_\infty \leq h \tag{DQP}$$

where $\|x\|_\infty = \max_i \ |x_i|$ [10]. The constraint of the dual says that at a feasible point, the residuals or reconstruction errors defined as $r = b - Ax$ are such that their correlations with the columns of A never exceed h. The parameter h allows to tune the maximal magnitude of these correlations. If the ℓ_2 norm of the columns of A is equal to one, this criterion allows thus for reconstruction errors that are of order h.

One can make the following remarks about the optimum x^* of (QP) or (DQP) as h varies from 0 to $+\infty$. They are either obvious or can be deduced from the results presented in Sections 3 and 6.

$\diamond$ For $h = 0$, one is left with $\min_x \ \|b - Ax\|_2^2$, and, provided A is full rank, the value of the minimum is zero and it is attained for all points in a convex set (a linear manifold).

$\diamond$ For $h = 0^+$, i.e., for h positive and arbitrarily close to zero, the solution is attained at the point(s) in the previous set having least ℓ_1-norm.

$\diamond$ For $h \geq \ \| A^T b \|_\infty$, the optimum is attained at $x^* = 0$, see (DQP).

As will be shown below for $h = 0^+$, (QP) has the same solution(s) as (LP). As h increases, the optimum of (QP) drifts away from the optimum of (LP), $\|Ax - b\|_2^2$ which was equal to zero, increases and $\|x\|_1$ decreases. More generally, the interval $]0^+, \ \| A^T b \|_\infty[$ can be divided into sub-intervals characterized by the fact that within each such sub-interval the number of non-zero components of the optimum of (QP) is constant. As h increases within one of the sub-intervals the non-zero components of the optimum mostly decrease in absolute value until one of them becomes zero and h hits the boundary of the sub-interval. For

$h \geq \| A^T b \|_\infty$ all the non-zero components have disappeared and the optimum remains at zero.

3 Optimality conditions for (QP)

In this section, we specify the necessary and sufficient conditions (NSC) satisfied at a minimum point of (QP) together with a sufficient condition for the minimum point to be unique.

The optimality conditions for (QP) can be obtained in a quite traditional way by first transforming the problem into a quadratic program as indicated above and writing the first order necessary (Kuhn-Tucker) conditions that are also sufficient since the problem is convex. We adopt a more direct path. Since the criterion is non-smooth at zero because of the presence of $\|x\|_1$, we introduce the sub-differential of $\|x\|_1$, a set of vector called the sub-gradients, denoted $\partial\|x\|_1$,

$$\begin{aligned}\partial \|x\|_1 &= \{u | u^T x = \|x\|_1,\ \|u\|_\infty \leq 1\}\\ &= \{u | u_i = \text{sign}(x_i) \text{ if } x_i \neq 0 \text{ and } |u_i| \leq 1 \text{ otherwise}\},\end{aligned}$$

where $\text{sign}(x_i) = 1$ when $x_i > 0$ and $\text{sign}(x_i) = -1$ when $x_i < 0$. An NSC for x^* to be a global minimum of (QP) is that the vector zero is a sub-gradient of the criterion at x^* [7]

$$\exists\, u \in \partial \|x\|_1 \quad \text{such that} \quad A^T(Ax^* - b) + hu = 0. \tag{NSC}$$

To write this NSC in a more usable way, we distinguish between the non-zero components and the zero components of x^*. We denote $\bar{x}^*$ the reduced dimensional vector built upon the non-zero components of x^*. Similarly $\bar{A}$ denotes the associated columns in A. One then has, e.g., $Ax^* = \bar{A}\bar{x}^*$. For the rows in (NSC) associated with the non-zero entries $\bar{x}^*$, the sub-gradient is unique and equal to the gradient, for the other rows the sub-gradient takes any value in $[-1,\ 1]$. The necessary and sufficient conditions become

$$\bar{A}^T(b - \bar{A}\bar{x}^*) = h\ \text{sign}(\bar{x}^*), \tag{NSC$_1$}$$

$$|a_j^T(b - \bar{A}\bar{x}^*)| \leq h \ \text{ for } \ a_j \notin \bar{A}. \tag{NSC$_2$}$$

One can further establish that if $\bar{A}$ is full rank and if the inequalities in (NSC$_2$) are strictly satisfied, then x^* is the unique or strict minimum point of (QP). The following theorem thus holds.

Theorem 1: Sufficient conditions for x^* to be a strict minimum point of (QP) are:

1) $\bar{A}^T(b - \bar{A}\bar{x}^*) = h\ \text{sign}(\bar{x}^*)$, (3)

2) $|a_j^T(b - \bar{A}\bar{x}^*)| < h$ for $a_j \notin \bar{A}$, (4)

3) $\bar{A}$ full rank,

where $\bar{x}^*$ denotes the reduced dimensional vector built upon the non-zero components of x^* and $\bar{A}$ is such that $Ax^* = \bar{A}\bar{x}^*$. □

If the conditions of Theorem 1 are satisfied, relation (3) leads to

$$\bar{x}^* = \bar{A}^+ \, b - h \, (\bar{A}^T \bar{A})^{-1} \, \mathrm{sign}(\bar{x}^*) \tag{5}$$

with $\bar{A}^+ = (\bar{A}^T \bar{A})^{-1} \bar{A}^T$, the pseudo-inverse of $\bar{A}$. The last term in the right hand side of (5) is a bias term induced by the regularization term in (QP) which can be removed a posteriori if desired.

Relation (5) is an implicit equation since $\bar{x}^*$ appears on both sides. It is valid for a fixed h. However, provided there are no zero components in $\bar{A}^+ \, b$, there exists an $h_m > 0$ such that for $h \in \;]0, \; h_m[$ taking $\mathrm{sign}(\bar{x}^*) = \mathrm{sign}(\bar{A}^+ \, b)$ in the right side of (5) leads to an $\bar{x}^*$ that satisfies (3). For h sufficiently small, the choice of $\bar{A}$ somehow fully characterizes $\bar{x}^*$.

4 Separability condition

We are now ready to develop a sufficient condition under which a solution of $Ax = b$ can be recovered from the unique minimum point of (QP). Let us denote x_o the solution one wants to recover, let $\bar{x}_o$ be the p-dimensional vector built upon the non-zero component in x_o and $\bar{A}_o$ the (n, p)-dimensional matrix built with the associated columns of A so that $Ax_o = \bar{A}_o \bar{x}_o = b$. We assume $\bar{A}_o$ to be full rank.

For a non-zero h, the optimum of (QP) is attained at a point, we denote x^*, that achieves a compromise between the two parts of the criterion $\|Ax - b\|_2^2$ and $\|x\|_1$. One thus has $Ax^* \neq b$ and x^* is never exactly equal to a solution of $Ax = b$. We will therefore consider that (QP) allows to recover x_o if its optimum x^* and x_o have their non-zero components at the same locations and with the same signs.

Theorem 2: The solution x_o of $Ax = b$ with $\bar{A}_o$ and $\bar{x}_o$ as defined above and $\bar{A}_o$ a full rank matrix, can be recovered from the unique optimum point of (QP) if

$$1) \qquad |a_j^T d_o| < 1, \;\; \forall a_j \notin \bar{A}_o \quad \text{with } d_o = \bar{A}_o^{+T} \mathrm{sign}(\bar{x}_o), \tag{6}$$

$$2) \qquad h \in \;]0, \; h_m[\;\; \text{the domain in which } \mathrm{sign}\{\bar{x}_o - h \, (\bar{A}_o^T \bar{A}_o)^{-1} \, \mathrm{sign}(\bar{x}_o)\} = \mathrm{sign}(\bar{x}_o), \tag{7}$$

with $\bar{A}^+ = (\bar{A}^T \bar{A})^{-1} \bar{A}^T$, the pseudo-inverse of $\bar{A}$. $\Box$

Proof: We associate with $\bar{x}_o$ a vector $\bar{x}^*$ that, together with $\bar{A}_o$, satisfies the sufficient conditions of Theorem 1. Since $\bar{x}_o = \bar{A}_o^+ \, b$, the parameter h in (7) is such that

$$\mathrm{sign}(\bar{A}_o^+ \, b) = \mathrm{sign}\{\bar{A}_o^+ \, b - h \, (\bar{A}_o^T \bar{A}_o)^{-1} \, \mathrm{sign}(\bar{A}_o^+ \, b)\}$$

and the vector $\bar{x}^*$ defined as

$$\bar{x}^* = \bar{A}_o^+ \, b - h \, (\bar{A}_o^T \bar{A}_o)^{-1} \, \mathrm{sign}(\bar{x}^*) \qquad \text{with} \quad \mathrm{sign}(\bar{x}^*) = \mathrm{sign}(\bar{x}_o)$$

then satisfies (3) with $\bar{A}$ replaced by $\bar{A}_o$. But for this vector $\bar{x}^*$ one has

$$|a_j^T (b - \bar{A}_o \bar{x}^*)| = h \, |a_j^T \bar{A}_o^{+T} \mathrm{sign}(\bar{x}_o)| = h \, |a_j^T d_o|$$

and it follows that (6) is equivalent to (4). This completes the proof since the vector $\bar{x}^*$ defined above, together with $\bar{A}_o$, satisfies all the conditions of Theorem 1 and allows to recover x_o. □

While the size h_m of the allowed domain for h depends upon the magnitudes of the components of $\bar{x}_o$, condition (6) is independent of these magnitudes. Condition (6) defines two parallel *separating* hyperplanes $H_\pm$, associated with a single vector d_o $H_\pm = \{a | a^T d_o = \pm 1\}$ which must be such that the columns that are present in the decomposition lie in these hyperplanes: $|a_j^T d_0| = \pm 1, \ \forall a_j \in \bar{A}_o$ (see the definition of the vector d_o), while the other columns (not in $\bar{A}_o$) must lie strictly in between (see (6)).

This is why in [4, 5] we called Eq. (6) the separability condition.

5 Sparsity condition

In this section, we specialize the separability condition (6) to the specific problem considered in [1, 2] and assume the columns a_j of A to have unit Euclidean norm. We transform the separability condition (6) into the more usable but probably more conservative condition (2). We prove

Theorem 3: If the columns in A are normalized to one in ℓ_2 norm, then (2) $\Rightarrow$ (6), i.e.,

$$\|x_o\|_0 < \frac{1}{2}(1 + \frac{1}{M}) \quad \Rightarrow \quad |a_j^T \bar{A}_o ({\bar{A}_o}^T \bar{A}_o)^{-1} \mathrm{sign}(\bar{x_o})| < 1, \ \forall a_j \notin \bar{A}_o. \tag{8}$$

Proof: For simplicity we set $p = \|x_o\|_0$. Remember that

$$M = \sup_{1 \le i \ne j \le m} |a_i^T a_j|.$$

The diagonal part of ${\bar{A}_o}^T \bar{A}_o$ is equal to the identity matrix I since we assume $a_j^T a_j = 1, \ \forall j$. We write ${\bar{A}_o}^T \bar{A}_o = I - H$ where H is built with the non-diagonal elements of $-{\bar{A}_o}^T \bar{A}_o$.

Since all the non-zero components in the square matrix H of order p are smaller than M, its spectral radius $\rho(H)$ satisfies $\rho(H) \le (p-1)M$ by Gershgorin theorem [11]. ¿From (2) it thus follows that $\rho(H) < 1$ and one can apply Neumann's lemma [11] to get

$$({\bar{A}_o}^T \bar{A}_o)^{-1} = (I - H)^{-1} = I + \sum_{k>0} H^k.$$

For a matrix B, we denote $|B|$ the matrix with entries $|b_{i,j}|$ and write $B < C$ if $b_{i,j} < c_{i,j}, \forall i, j$, one then has

$$\begin{aligned} |({\bar{A}_o}^T \bar{A}_o)^{-1}| &= |I + \sum_{k>0} H^k| \le I + \sum_{k>0} |H^k| \\ \le I + \sum_{k>0} |H|^k &\le I + \sum_{k>0} (ME)^k = (I - ME)^{-1} \end{aligned}$$

where E is a square matrix whose entries are equal to one except those on the diagonal which are equal to zero. One thus has $E = \mathbf{1} - \mathbf{I}$, where $\mathbf{1}$ denotes here and below a matrix of ones with adequate dimensions. It is then easy to check that

$$\begin{aligned} &I - ME = (1+M)I - M\mathbf{1} \quad && \Rightarrow (I - ME)^{-1} = \alpha I + \beta \mathbf{1}, \\ &\text{with } \alpha = \frac{1}{1+M} \qquad \text{and} && \beta = \frac{M}{(1+M)(1+M-Mp)}. \end{aligned}$$

Combining this result with, e.g., $|a_j^T \bar{A}_o| \leq M\mathbf{1}^T$ lead to the following inequalities

$$\begin{aligned} |a_j^T \bar{A}_o ({\bar{A}_o}^T \bar{A}_o)^{-1} \text{sign}(\bar{x_o})| \quad & \leq |a_j^T \bar{A}_o| |({\bar{A}_o}^T \bar{A}_o)^{-1}| |\text{sign}(\bar{x_o})| \\ & \leq M \mathbf{1}^T |({\bar{A}_o}^T \bar{A}_o)^{-1}| \mathbf{1} \\ & \leq M \; (\alpha p + \beta p^2) \\ & \leq \frac{Mp}{1 + M - Mp} < 1, \end{aligned}$$

where the very last step follows from (2). □

Combining Theorems 2 and 3, we have established that

Corollary: If a solution x_o of $Ax = b$ satisfies (2) then (QP) has a unique solution point from which x_o can be recovered, provided h in (QP) has been taken small enough. □

Indeed taking $h = 0^+$ guarantees that $x^* = x_o$ as is the case for the (LP) criterion.

Remark : Specializing the proof above to the case where A is the concatenation of 2 unitary matrices yields the condition $\|x\|_0 < (\sqrt{2} - 0.5)/M$, obtained in [2], that is slightly weaker than (2). In this case ${\bar{A}_o}^T \bar{A}_o$ has a block diagonal form and $|a_j^T \bar{A}_o|$ has some zero entries leading quite easily to this tighter bound.

6 Special case of the linear program

We claimed in Section 2 that the optimum of (QP) converges to the optimum of (LP) when h in (QP) decreases to zero. For simplicity we assume both optima to be strict and we rewrite (5) as

$$\bar{x}^*(h) = \bar{A}^+ \, b - h \; (\bar{A}^T \bar{A})^{-1} \; \text{sign}(\bar{x}^*(h)),$$

where we emphasize the dependence of $\bar{x}^*$ on h. We assume that $h \in \;]0, \; h_m[$, the non-empty domain in which $\text{sign}(\bar{x}^*(h)) = \text{sign}(\bar{A}^+ \, b)$. In this domain, (QP) is equivalent to

$$\min_x \; \|x\|_1 \qquad \text{subject to} \quad \|Ax - b\|_2^2 \leq B(h),$$

with $B(h) = \|Ax^*(h) - b\|_2^2 = 0(h^2)$. But the optimum point of this new problem converges to the optimum of (LP) as $h \; \downarrow 0$ since $B(h) \; \downarrow 0$. This establishes the claimed equivalence and actually establishes that condition (2) is also sufficient for (LP) to retrieve the unique sparsest solution.

A direct analysis of this result can indeed be achieved quite easily along the lines used for the (QP) criterion. Let us do so for completeness in this section using the notations introduced above.

Theorem 4: 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, is the unique optimum point of (LP) if

$$|a_j^T d_0| < 1 \;\; \forall a_j \notin \; \bar{A}_0 \text{ for some } d_0 \text{ such that } \bar{A}_0^T d_0 = \text{sign}(\bar{x}_o). \tag{9}$$

Proof : Let us recall the (LP) criterion

$$\min_x \; \|x\|_1 \quad \text{subject to} \quad Ax = b. \tag{LP}$$

A slight difficulty arises from the fact that the expected solution x_o of (LP) is degenerate, i.e., has less than n non-zero components. In order to characterize the optimality of the solution, it is then convenient to introduce the dual linear program [6]

$$\max_d d^T b \;\; \text{subject to} \quad \|d^T A\|_\infty \leq 1 \tag{DLP}$$

which, as a linear program, is not in standard form either.

For a vector x_o satisfying $Ax_o = \bar{A}_o\bar{x}_o = b$ to be an optimum of (LP), one needs to be able to associate with it a solution, say d_o, of (DLP) that is feasible and has an identical cost

$$\|d_o^T A\|_\infty \leq 1 \;\; \text{and} \;\; \|x_o\|_1 = d_o^T b.$$

Rewriting the primal cost $x_o^T \text{sign}(x_o) = \bar{x}_o^T \text{sign}(\bar{x}_o)$ and the dual cost $d_o^T b = d_o^T A x_o = d_o^T \bar{A}_o \bar{x}_o$ shows that they are equal for all $\bar{x}_o$ if $d_o^T \bar{A}_0 = \text{sign}(\bar{x}_o^T)$. For d_o to be a feasible point of (DLP) it remains then to check

$$|a_j^T d_o| \leq 1 \;\; \forall a_j \notin \bar{A}_o. \tag{10}$$

For x_o to be the unique solution of (LP) one has to ask in addition that the dual problem be non-degenerate in the sense that all constraints associated with zero Lagrange multipliers be strictly satisfied. Since the Lagrange multipliers of (DLP) are the components of x_o, this means that the inequalities in (10) have to be strict, transforming (10) into (9). □

¿From Theorems 3 and 4 it thus follows that

Corollary: If (2) holds then x_o is the unique solution of (LP). □

Condition (9) is weaker than (6) that needs to hold for (QP), since the vector d_o is now less constrained. It belongs to an (n-p)-dimensional linear manifold defined in (9) while the vector d_o defined in (6) for the (QP) criterion is a specific vector of this manifold, the one of least ℓ_2 norm. It is indeed possible to construct toy examples where (QP) allows to recover a solution (not satisfying the sufficient condition (2)) while (LP) does not.

7 Conclusions

Given any (n, m) matrix A with $m > n$ and a solution x_o of the set of under-determined linear equations $Ax = b$, we have given sufficient conditions (6) under which it is possible to recover x_o from the optimum of a quadratic program [5] and slightly weaker sufficient conditions (9) under which this is possible using a linear program.

If the columns in A are further taken to have unit Euclidean norm and one is primarily interested in sparse solutions then we have shown that both (6) and (9) are satisfied if

$$\|x_o\|_0 < \tfrac{1}{2}(1 + \tfrac{1}{M}) \quad \text{with} \quad M = \sup_{1 \le i \neq j \le m} |a_i^T a_j|, \tag{2}$$

where $\|x\|_0$ is the number of non-zero components in x. This also establishes that this condition guarantees that x_o is the unique sparsest solution to $Ax = b$. This result on sparse solutions extends those previously published which considered $(n, 2n)$ matrices A obtained by concatenating two unitary matrices and is similar to the result obtained in [3] using a completely different approach.

The result holds both for the linear program (LP) seeking the minimal ℓ_1 norm solution to $Ax = b$ and for the following quadratic program parametrized by an hyper-parameter h

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

For the criterion (QP), under an additional condition on h that is always satisfied for $h = 0^+$, its unique optimum x^* and the sought-for solution x_o have their non-zero components at the same locations and with identical signs. Since this criterion allows for non-zero residuals $r = Ax^* - b \neq 0$, with magnitudes of the order of h, x^* is never a solution of $Ax = b$ and is the best one can expect.

Allowing for non-zero residuals is of practical importance since it allows for the presence of noise or errors due to mis-modeling. It is also the main reason for considering (QP) rather then (LP). In the presence of noise, the true model is

$$b = Ax_o + e$$

with e a vector of white Gaussian noise, for instance. Asking for an exact reconstruction of b as a linear combination of columns of A will yield a non-sparse estimate x^* of x_o. One possibility is then to threshold the components of this estimate in order to recover x_o by removing the small components induced by the presence of the noise. But a more sensible approach is probably to take into account from the beginning the presence of the noise and to solve the quadratic program (QP) with h of the order of the standard deviation of the noise [9, 10] to get a sparse x^* that does not require pruning.

Another similar situation arises when the true model of the observed vector b is

$$b = \sum_{i=1}^{p} \alpha_i a(\tau_i)$$

with $a(\tau)$ a known family of vectors parametrized by a scalar τ and α_i the scalar weights. In order to apply the previous setting to this estimation problem, one way to proceed is to uniformly discretize the values of τ over its compact domain to get the m columns $a_j = a(\tau)|_{\tau=t_j}$ of the A matrix. Since the τ_i will generically not belong to the sampling points set $\{t_j\}$, there is again no sparse representation of b as a linear combination of the columns of A. In the simplest case of a single component $b = a(\tau_1)$, the components of x^* can be shown to be samples from an interpolating function [13].

Tentative analyses of these more complex scenarios where both noise and mis-modeling errors are present can be found in [12, 13] for (QP) or the following modified (LP) criterion :

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

Acknowledgments The author wants to thank one of the reviewers and the Associate Editor for the numerous comments they made that helped improve the readability of the submission.

References

[1] D.L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inform. Theory*, vol. 47, no. 7, pp. 2845–2862, Nov. 2001.

[2] M. Elad and A.M. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases," *IEEE Trans. Inform. Theory*, vol. 48, no. 9, pp. 2558–2567, Sep. 2002.

[3] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Trans. Inform. Theory*, vol. 49, no. 12, pp. 3320–3325, Dec. 2003.

[4] J.J. Fuchs, "Une approche à l'estimation et l'identification simultanées," *XVIe Colloque GRETSI*, vol. 2, pp. 1273–1276, Grenoble, France, 1997.

[5] J.J. Fuchs, Detection and estimation of superimposed signals. *IEEE ICASSP*, vol. III, pp. 1649–1652, Seattle, U.S.A., 1998.

[6] D. G. Luenberger, *Introduction to linear and nonlinear programming.* Boston, MA: Addison Wesley, 1973.

[7] R. Fletcher, *Practical methods of optimization.* New York, N.Y.: John Wiley and Sons, 1987.

[8] S. Chen, D. Donoho and M. Saunders, "Atomic Decomposition by basis pursuit," *SIAM J. Scientific Computing*, vol. 20, no. 1, pp. 33–61, Jan. 1999.

[9] J.J. Fuchs, "Multipath time-delay detection and estimation," *IEEE Trans. Signal Processing*, vol. 47, no. 1, pp. 237–243, Jan. 1999.

[10] J.J. Fuchs, "On the application of the global matched filter to DOA estimation with uniform circular arrays," *IEEE Trans. Signal Processing*, vol. 49, pp. 702–709, Apr. 2001. *IEEE Signal Processing Society 2003 Best Paper Award.*

[11] A.S. Householder, *The theory of matrices in numerical analysis.* New York, N.Y.: Blaisdell, 1964.

[12] J.J. Fuchs, "Extension of the Pisarenko method to sparse linear arrays," *IEEE Trans. Signal Processing*, vol. 45, pp. 2413–2421, Oct. 1997.

[13] J.J. Fuchs and B. Delyon, "Minimum L_1-norm reconstruction function for oversampled signals: Application to time-delay estimation," *IEEE Trans. Inform. Theory*, vol. 46, no. 4, 1666–1673, Jul. 2000.