Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

A sparse representation criterion: recovery conditions and implementation issues.

Jean Jacques Fuchs *

* *Irisa-Université de Rennes, Campus de Beaulieu, 35042 Rennes Cedex, France.(e-mail: fuchs@irisa.fr).*

Abstract:
Sparse representations techniques have become an active domain of research in signal processing with numerous applications in compression and coding, for instance. They are mostly 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. We replace the least-square-part by a ℓ_∞-part and investigate the recovery conditions of the so-obtained $\ell_\infty - \ell_1$ criterion. We then propose an algorithm, that minimizes the criterion, in a finite number of steps.

1. INTRODUCTION

There is currently a huge interest in sparse representations which is a technique that consists in decomposing a signal into a small number of components chosen from a user-designed over-complete set of vectors. It is mostly used to obtain a simple approximate model for a complex signal for compression or coding purposes in audio or video signal processing (Bertalmio [2003], Zibulewsky [2001]), 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 that the Nyquist rate with a controlled loss in information (Candes [2006], Donoho [2006], Candes [2006b]).

The current interest has been initiated in (Donoho [2001]) but earlier investigations had been proposed in different areas, (Mallat [1993], Gorodnitski [1997], Fuchs [1997], Sacchi [1998]).

In (Donoho [2001]), and later in e.g. (Gribonval [2003], Fuchs [2004]), the following problem is considered. 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 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

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

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 (Fuchs [2004])

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

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 probably 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 (Osborne [2000], Efron [2004], Maria [2006]). In the sequel, we propose to replace the criterion (2) by (Fuchs [1997])

$$\min_x \ \|x\|_1 \qquad \text{s.t.} \quad \|Ax - b\|_\infty \leq \rho.. \tag{4}$$

where the least-squares-part is replaced by an ℓ_∞-norm, i.e., a threshold on the maximal reconstruction error It is a convex criterion that 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) and to develop an optimization algorithm that converges in a finite number of steps.

2. THE CRITERION

2.1 Preliminary remarks

We consider the following problem:

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

where $\rho > 0$ has to be fixed by the user. If one splits the components x_i of x into $x_i^+ = \max(x_i, 0)$, $x_i^- = \max(-x_i, 0)$ and replaces x_i by $x_i^+ - x_i^-$ and $|x_i|$ by $x_i^+ + x_i^-$ and further introduces slack variables, one can transform this problem into a linear program in standard form and get its dual. Some interesting generic information about the optimum can be deduced as well as the optimality conditions, we will however obtain them in a different way below

978-1-1234-7890-2/08/$20.00 © 2008 IFAC **12425** **10.3182/20080706-5-KR-1001.0053**

2.2 Optimality conditions

Since both $\|x\|_1$ and $\|x\|_\infty$ are not continuously differentiable at all points, we introduce the sub-differential of these functions at x, it is a set of vectors, called the sub-gradients at x, denoted $\partial\|x\|_p$ with p=1 or ∞, one has Fletcher [1991]

$$\partial\|x\|_p = \{u | u^T x = \|x\|_p,\ \|u\|_q \leq 1\} \tag{5}$$

where q is such that $\frac{1}{p}+\frac{1}{q}=1$. Since for p=1, q=∞ and vice versa, one says that ℓ_1 and ℓ_∞ are dual norms. From (5), it follows that

$$\partial\ \|x\|_1 = \{u | u_i = \text{sign}(x_i) \text{ if } x_i \neq 0 \text{ and } |u_i| \leq 1 \text{ else}\}$$
$$\partial\ \|x\|_\infty = \{u |\ |x_i| = \|x\|_\infty \Rightarrow x_i u_i \geq 0,\ |x_i| < \|x\|_\infty$$
$$\Rightarrow u_i = 0;\ \|u\|_1 = 1 \text{ if } x \neq 0, \|u\|_1 \leq 1 \text{ else}\}$$

Note that if f is differentiable at x then $\partial f(x)$ reduces to the gradient. It is now possible to get optimality conditions for (4), by simply writing the first order necessary optimality conditions, that are also sufficient, since the criterion is convex. We will rather write the dual of (4) to obtain the optimality conditions in a form that is more convenient for latter use.

Lemma 1. The dual of (4) is

$$\max_d\ b^T d - \rho\|d\|_1 \quad s.t.\ \ \|A^T d\|_\infty \leq 1 \qquad \square \tag{6}$$

Proof: We first rewrite (4) as

$$\min_{x,\ c}\ \|x\|_1 \quad \text{s.t.}\ \ \|c\|_\infty \leq \rho \ \text{ and } \ Ax - b = c.$$

Introducing Lagrange multipliers $\lambda \in R^+$ and $d \in R^n$, the Lagrangian of this problem is

$$\ell(..) = \ \|x\|_1 + \lambda(\|c\|_\infty - \rho) - d^T(Ax - b - c),$$

and defining $\phi(\lambda,\ d) = \min_{x,\ c}\ \ell(x, c, \lambda, d)$, the dual problem is $\max_{\lambda\geq 0,\ d}\phi(\lambda,\ d)$.

In order to evaluate $\phi(\lambda,\ d)$, we first take the minimum of $\ell(\ .\ )$ with respect to x

$$\min_x \|x\|_1 - d^T Ax + ... = \min_x x^T u - x^T A^T d + ...$$

This minimum may not be finite for all d, but, since we latter take the maximum in d, theses cases can be ignored. The minimum is finite if and only if $A^T d = u$ for some $u \in \partial\|x\|_1$. From (5), it follows that such a point exists only if $\|A^T d\|_\infty \leq 1$ and the contribution of the terms in x to $\phi(\ .\ )$ is then zero.

Similarly, the minimum in c may not be finite for all d. It is finite if and only if $\lambda v + d = 0$ for some $v \in \partial\|c\|_\infty$. Such a point exists only if $\|d\|_1 \leq \lambda$ and the contribution of the terms in c to ϕ is then zero. The dual problem is thus

$$\max_{\lambda\geq 0,\ d}\ d^T b - \lambda\rho \quad \text{s.t.} \quad \|A^T d\|_\infty \leq 1, \quad \|d\|_1 \leq \lambda$$

and taking the maximum with respect to $\lambda \geq 0$ leads to the announced result. □

Using the primal and the dual, which are both convex programs, one has the following result.

Theorem 2. The optima of (4) and (6) are respectively x and d, if and only

$$Ax - b = -\rho v \quad \text{and} \quad A^T d = u \tag{7}$$

for some $u \in \partial\|x\|_1$ and $v \in \partial\|d\|_1$ □

Proof: The proof is immediate. Both points x and d are feasible and lead to equal costs. □

We will use the two relations in (7) 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 (5), that the sub-gradient $u \in \partial\|x\|_1$ is such that $\bar{u}$ =sign($\bar{x}$) and $\|\bar{\bar{u}}\|_\infty \leq 1$.

We will also need d-induced partitions of the rows of A. We partition (the optimal) d into its non-zero components $\underline{d}$ and $\underline{\underline{d}} = 0$, and accordingly the rows of A into $\underline{A}$ and $\underline{\underline{A}}$. Similar d-induced partitions apply to v, $\bar{A}$ and $\bar{\bar{A}}$, for instance. Again since $v \in \partial\|d\|_1$, one has $\underline{v}$ =sign($\underline{d}$) and $\|\underline{\underline{v}}\|_\infty \leq 1$.

We may thus partition A, either in block-columns $\bar{A}$ and $\bar{\bar{A}}$, or in block-rows $\underline{A}$ and $\underline{\underline{A}}$, or, combining the two, into four blocks.

3. RECOVERY CONDITIONS

We now establish the following *recovery* conditions

Theorem 3. 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 of (4), for ρ sufficiently small, if there exists a

$$d = \arg\min_d \|d\|_1 \ \text{ s.t. } \ \bar{A}_o^T d = \text{sign}(\bar{x_o}),$$
$$\text{that satisfies} \qquad \|\bar{\bar{A}}_o^T d\|_\infty < 1. \qquad \square \tag{8}$$

Comment: The optimum x of (4) will not be equal to x_o for $\rho > 0$. What one asks for, is that the sub-gradient u of $\|x\|_1$ at the optimum of (4) satisfies $\bar{u}$ =sign($\bar{x}_o$) and $\|\bar{\bar{u}}\|_\infty$<1. We will prove, that 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.

Proof. If d satisfies (8), it is also the optimum of

$$\min_d \|d\|_1, \ \ \text{s.t. } \bar{A}_o^T d = \text{sign}(\bar{x_o}), \ \ \|\bar{\bar{A}}_o^T d\|_\infty \leq 1. \tag{9}$$

The dual of this optimization problem is

$$\max_z \text{sign}(\bar{x_o})\bar{z} - \|\bar{\bar{z}}\|_1, \ \ \text{s.t. } \|\bar{A}_o\bar{z} + \bar{\bar{A}}_o\bar{\bar{z}}\|_\infty \leq 1 \tag{10}$$

where we have imposed on z the x_o-induced partition. But, since the optimum z of (10) is equal to the optimal Lagrange multipliers of (9), it follows (see the strict inequality in the condition of (8)), that indeed $\bar{\bar{z}} = 0$, which validates this partition a posteriori.

Using the optimum d of (9, 8) and the associated optimum z of (10 for which $\bar{\bar{z}} = 0$, we define $u = A^T d$, $v = Az$ and $x = x_o - \rho z$ and check that the so-obtained quadruple x, u, d, and v satisfies (7), this permits to complete the proof.

Indeed, we have already seen that x_o and z have the same partition, which is thus also valid for x. One gets $Ax = Ax_o - \rho Az = b - \rho v$. From $v = Az$, it follows $v = Az = \bar{A}_o \bar{z}$ and this vector has all the properties required to belong to $\partial\|d\|_1$. The same holds for $u = A^T d$ which has all the required properties for a vector in $\partial\|x\|_1$ □

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 (10) 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 (8), admits a explicit solution (Fuchs [2004]), that can be further transformed into explicit conditions on the sparsity of x_o, no such miracle happens for the ℓ_∞-norm, since the optimum of (8) has no explicit analytical expression.

Anyway, these conditions are purely theoretical, quite conservative, though sharp and in practice they are essentially un-usable.

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. One can also use the linear programming theory to establish that if, for the ρ of interest, the optimum d of the dual has $p \leq n$ non zero components then the same holds generically for the optimal x of the primal which is thus sparse.

We will not use the linear programming approach but the two relations in (7) to develop an algorithm that solves (4) in a finite number of steps and should thus be more efficient than the standard linear program solvers.

Due to the presence of u and v, which belong to sets, the two relations in (7) 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 optima x and d vary with ρ.

The idea of the algorithm is to deduce from the two relations in (7) how the optimal x and d vary with ρ, to observe that this is indeed feasible as long as ρ belongs to an interval whose boundaries are easy to obtain from (7) and that indeed one can also propagate the optima to the neighboring interval. One then, simply, starts with ρ large, $(\rho > \rho_0 = \|b\|_\infty)$, for which the optimum x is at zero and follow the optimum $x(\rho)$ for diminishing ρ. The number of nonzero components in $x(\rho)$ essentially increases and never exceeds n. Though we are only interested in the optimum for a given value of ρ, we will build it, for decreasing ρ, and stop when we attain the ρ of interest.

More precisely, as ρ decreases, there is a first interval $]\ \rho_1,\ \rho_0]$, in which the optimum has just one non-zero component, then a second interval for which it has two non-zero components, and so on. The conditions in (7) tell us how to get the boundaries of the intervals, how to propagate the optima within the intervals, and, it remains to find out how to cross the boundaries.

4.2 Development

Assume we have the quadruple $x,\ u,\ d,\ v$, that satisfies the optimality conditions (7) 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 change. We have already indicated, that x and d have generically the same number, we denote p, of nonzero components. It is the dimension of $\bar{x}$ and $\underline{d}$

From the second condition in (7), $A^T d = u$, one essentially deduces that $\bar{\underline{A}}^T \underline{d} = \bar{u}$, with $\bar{\underline{A}}$ a square order-p matrix, we assume invertible. Since $\bar{u}$ =sign($\bar{x}$) is a constant vector within the current interval, this tells us that $\underline{d}$, and thus d is invariant, within the interval.

The other condition $Ax - b = -\rho v$, first becomes $\bar{A}\bar{x} = b - \rho v$, and, further, yields

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

Solving the first relation for $\bar{x}$, and substituting in the second, one gets

$$\bar{x}(\rho) = \bar{\underline{A}}^{-1}\underline{b} - \rho \bar{\underline{A}}^{-1}\underline{v} \tag{11}$$

$$\text{and} \qquad \underline{\underline{v}}(\rho) = \frac{1}{\rho}\underline{\underline{b}} - \frac{1}{\rho}\bar{\underline{\underline{A}}}\ \bar{\underline{A}}^{-1}\underline{b} + \bar{\underline{\underline{A}}}\ \bar{\underline{A}}^{-1}\underline{v}. \tag{12}$$

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

As ρ varies, two remarkable events can happen: a component in $\underline{\underline{v}}(\rho)$ becomes equal to ± 1 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 increased (decreased). We will only consider decreasing values of ρ but both events can happen.

$\diamond$ If component i_{p+1} in $\underline{\underline{v}}(\rho)$ becomes, say, $+1$ as ρ=ρ_l, this means that the corresponding component in $\underline{\underline{d}}$, which is zero, will become positive if ρ further decreases and will thus need to be moved from $\underline{\underline{d}}$ to $\underline{d}$.

Quite generally, for $\rho = \rho_l$, there are two valid expressions for $\bar{x}(\rho)$ the one above valid locally for $\rho \geq \rho_l$ and another one, of dimension p+1, valid locally for $\rho \leq \rho_l$ and where $\bar{x}(\rho_l)$ has a zero component. It is this last expression with its associated new partitions that we need to find.

We know already that it is row i_{p+1}, that changes in the d-induced partition. For instance, row i_{p+1} is removed from $\bar{\underline{A}}$ and added to $\bar{\underline{\underline{A}}}$ which becomes, say $\bar{\underline{A}}_t$ (the $_{-t}$ to notify that we are in a transition phase), a $(p+1) \times p$ matrix. It remains then to identify the component j_{p+1} of $\bar{\bar{x}}$ that becomes nonzero as ρ becomes slightly smaller than ρ_l.

The index j_{p+1} cannot be deduced from (7). Indeed, since $\underline{d}$, which represents the Lagrange multipliers of the primal (4), has p+1 nonzero components, this means that p+1 constraints will be active in the primal, locally for $\rho \leq \rho_l$. These are precisely those associated with $\underline{\bar{A}}_t$ and the primal (4) for, say, $\rho = \rho_l - \epsilon$ with small positive ϵ, reduces to

$$\min_x \|x\|_1 \quad \text{s.t.} \quad \underline{\bar{A}}_t x = \underline{b}_t - \rho \underline{v}_t \tag{13}$$

By continuity, one knows that the components in $\bar{x}(\rho_l)$ will remain nonzero in $\bar{x}(\rho_l - \epsilon)$, and, one simply has to identify which component $x_j \in \bar{\bar{x}}$ has to be added to $\bar{x}$ to yield the optimum of (13). If we denote, locally, α_j the j-th column of $\underline{\bar{\bar{A}}}_t$ and B_j the square order p+1 matrix $B_j = [\ \underline{\bar{A}}_t \ \ \alpha_j]$, the j-th potential solution of (13) is

$$\bar{x}_j(\rho_l - \epsilon) = B_j^{-1}(\underline{b}_t - \rho_l \underline{v}_t \ + \epsilon \underline{v}_t) = \bar{x}_j(\rho_l) \ + \epsilon B_j^{-1} \underline{v}_t$$

and the sought-for index j is

$$j_{p+1} = \arg\min_j \|B_j^{-1} \underline{v}_t\|_1,$$

since $\|B_j^{-1} \underline{v}_t\|_1$ is the marginal additional cost in (13). One can, indeed, obtain an explicit expression of the additional marginal cost for each of the $m - p$ potential solutions, using the explicit expression of the inverse of the block-columns matrix $B_.$. The computational complexity of the optimization is thus quite limited, though augmenting with p.

$\diamond$ If, as ρ decreases, a component in $\bar{x}(\rho)$ becomes zero, for $\rho = \rho_l$, this means that a component in $\bar{x}$ has to be removed from $\bar{x}$, all the x-induced partitions wil change and enter a transition phase. The square order p matrix $\underline{\bar{A}}$ will become $\underline{\bar{A}}_t$ of dimension $p \times (p-1)$. From an optimization point of view, since the dual has still p nonzero components for ρ slightly smaller than ρ_l, a new component x_{j_p} selected in $\bar{\bar{x}}$ will replace the exiting one. The problem to solve is the same as above (13).

4.3 Initialization step

We now know how to propagate the optimal quadruple within an interval (11, 12), how to get the boundaries of the intervals, how to cross the boundaries, it remains to indicate how to initialize the procedure. It follows easily from (4), that the first boundary point is at $\rho_0 = \|b\|_\infty$. For $\rho > \rho_0$, the optimum of (4) is at zero. If $i_1 = \arg\max_i |b_i|$, as ρ decreases, $|b_{i_1}|$ has to be decreased by introducing a nonzero component in x. The most efficient, as far as the ℓ_1-norm of x is concerned, component of x, is x_{j_1} with $j_1 = \arg\max_j |a_{i_1,j}|$. One thus has $\underline{\bar{A}} = a_{i_1,j_1}$ and this completes the initialization step. The different parts of the optimal quadruple are straightforward to obtain, (11), for instance, is

$$x_{j_1} = \frac{b_{i_1}}{a_{i_1,j_1}} - \rho \frac{b_{i_1}}{a_{i_1,j_1}},$$

within the first interval.

5. RELATIONS TO PREVIOUS WORKS

Several recent papers have proposed similar path-following methods for solving (3) (Osborne [2000], Efron [2004], Maria [2006]). All these methods are related to continuation techniques, which have also been studied in the optimization literature (Allgower [1993]). 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).

We proposed a preliminary sketch of the algorithm described above, in (Fuchs [2005]). It is recognized in image coding, for instance, that the ℓ_∞ norm should be preferred to the ubiquitous ℓ_2-norm that spreads and averages the errors on the whole image and may thus potentially yield poor edges coding.

REFERENCES

M. Bertalmio et al. Simultaneous structure and texture image impainting *IEEE Trans. on Image Processing*, 12, 882–889, 2003.

M. Zibulewsky and B. Pearlmutter Blind source separation by sparse decomposition on a signal dictionary *Neural Computation* , 13, 863–882, 2001.

E. Candes, J. Romberg and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information." *IEEE T. on I.T.*, 52, 489-509, 2006.

D. Donoho, "Compressed sensing." *IEEE Trans. on I.T.*, 52, 1289-1306, 2006.

E. Candes and T. Tao, "Near optimal signal recovery from random projections: Universal coding strategies ?" *IEEE Trans. on I.T.*, 52, 5406-5425, 2006.

D.L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Trans. on I.T.*, 47, 11, 2845-2862.

S. Mallat and Z. Zhang, "Matching Pursuit with time-frequency dictionaries," *IEEE Trans. on S.P.*, 41, 3397-3415, 1993.

I.F. Gorodnitski and B.D. Rao "Sparse signal reconstruction from limited data using FOCUSS: a reweighted minimum-norm algorithm" *IEEE Trans. on S.P.*, 45, 3, 600-616, Mar. 1997.

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

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.

R. Gribonval and M. Nielsen. Sparse representations in unions of bases. *IEEE Trans. on I.T.* 49, 12, 3320-3325, Dec. 2003.

J.J. Fuchs. More on sparse representations in arbitrary bases. *IEEE Trans. on I.T.* 50, 6, 1341-1344, June 2004,

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.

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

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.

R. Fletcher. Practical methods of optimization. *Wiley*, 1991.

E. Allgower and K. Georg. "Continuation and path following" *Acta Numerica*, 2,31–64, 1993.

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