SIGNAL RECOVERY FROM INCOMPLETE MEASUREMENTS IN THE PRESENCE OF OUTLIERS

Björn Popilka, Simon Setzer and Gabriele Steidl

Institute of Mathematics
University of Mannheim
68131 Mannheim, Germany

Abstract. We study the restoration of a sparse signal or an image with a sparse gradient from a relatively small number of linear measurements which are additionally corrupted by a small amount of white Gaussian noise and outliers. We minimize $\ell_1 - \ell_1$ and $\ell_1 - TV$ regularization functionals using various algorithms and present numerical results for different measurement matrices as well as different sparsity levels of the initial signal/image and of the outlier vector.

1. Introduction

Recently, substantial progress was made in solving the fundamental problem of recovering a finite signal from a limited set of measurements [17, 8, 12, 29, 30, 19, 20]. Let $A \in \mathbb{C}^{n,N}$, $n \leq N$ be a matrix with some 'good' properties. Typical examples of such matrices are random matrices from the Gaussian ensemble or the symmetric Bernoulli ensemble or matrices whose rows are n random vectors from the unit sphere in $\mathbb{R}^N$. Let x_0 be a sparse vector. Usually, we measure sparsity in the ℓ_0–seminorm $\|x\|_0 := |\{j : x_j \neq 0\}|$. Moreover, for $1 \leq p < \infty$, we will use the ℓ_p–norms $\|x\|_p := (\sum_{j=0}^{N-1} |x_j|^p)^{1/p}$. In this paper, we are interested in recovering x_0 from measurements

$$b := Ax_0 + z_0 + e_0,$$

where z_0 denotes a vector with small ℓ_2–norm $\|z_0\|_2 \leq \varepsilon$, e.g., white Gaussian noise and e_0 is a sparse vector with large non-zero coefficients (outliers), e.g., due to missing data in the measurements. Gross sparse error vectors combined with errors with small ℓ_2–norm were recently considered in a setting different from the one considered in this paper in [6].

If we have no noise, i.e., $z_0 = e_0 = 0$, and n is sufficiently large, then, with high probability, the linear problem

$$(\mathrm{P}_1) \qquad \arg\min_x \|x\|_1 \qquad \text{s.t.} \quad Ax = b,$$

has a unique solution which is equal to x_0, cf. [9]. If we have Gaussian noise but no outliers, i.e., $e_0 = 0$, and n is sufficiently large, then with high probability, the solution of the quadratic problem

$$(\mathrm{P}_{2,1}^\varepsilon) \qquad \arg\min_x \|x\|_1 \qquad \text{s.t.} \quad \|b - Ax\|_2 \leq \varepsilon$$

or equivalently (for appropriate ε, λ) of

$$(\mathrm{P}_{2,1}) \qquad \arg\min_x \frac{1}{2}\|b - Ax\|_2^2 + \lambda\|x\|_1$$

is unique and provides a good approximation to x_0, cf. [10, 7].

As an alternative to solving $(\mathrm{P}_{2,1})$ one may use the Dantzig selector, cf. [11] or apply (orthogonal) matching pursuit instead of solving (P_1) , cf. [29, 23].

In the presence of outliers, method $(\mathrm{P}_{2,1})$ fails to work. However, in various papers, e.g., on image restoration [18, 3] the minimization of functionals with ℓ_1 data-fitting term has shown a good performance. A great amount of theoretical work in this direction was done by M. Nikolova in [24, 25, 26]. In this paper, we want to adopt these ideas to the incomplete measurement problem.

More precisely, in the case $z_0 = 0$, we examine the performance of

$$(\tilde{\mathrm{P}}_{1,1}) \qquad \arg\min_{x,e} \|e\|_1 + \lambda\|x\|_1 \qquad \text{s.t.} \quad Ax + e = b$$

or equivalently of

$$(\mathrm{P}_{1,1}) \qquad \arg\min_x \|b - Ax\|_1 + \lambda\|x\|_1.$$

If we have in addition Gaussian noise with $\|z_0\| \leq \varepsilon$ it makes sense to solve

$$(\mathrm{P}_{2,1,1}^{\varepsilon}) \qquad \arg\min_x \|e\|_1 + \lambda\|x\|_1 \qquad \text{s.t.} \quad \|b - Ax - e\|_2 \leq \varepsilon$$

or equivalently (for appropriate ε, α)

$$(\mathrm{P}_{2,1,1}) \qquad \arg\min_{x,e} \frac{1}{2\alpha}\|b - Ax - e\|_2^2 + \|e\|_1 + \lambda\|x\|_1.$$

It can be proved that problems $(\mathrm{P}_{2,1,1})$ and $(\mathrm{P}_{1,1})$ become equivalent if α, respectively ε tends to zero, cf. [1].

Finally, in order to recover images x_0 with sparse gradients we replace the ℓ_1-norm of x by a (discrete) total variation norm $\|\cdot\|_{\mathrm{TV}}$ and consider

$$(\mathrm{P}_{1,\mathrm{TV}}) \qquad \arg\min_x \|b - Ax\|_1 + \lambda\|x\|_{\mathrm{TV}}$$

and

$$(\mathrm{P}_{2,1,\mathrm{TV}}) \qquad \arg\min_{x,e} \frac{1}{2\alpha}\|b - Ax - e\|_2^2 + \|e\|_1 + \lambda\|x\|_{\mathrm{TV}}.$$

Alternatively, it is possible to replace the TV–norm by the Besov norm in $B_{1,1}^1$ and use more general wavelet frames.

Problems $(\mathrm{P}_{2,1,1})$, $(\mathrm{P}_{1,1})$ are convex with coercive functionals such that there exists a solution which however is in general not unique.

The outline of this paper is as follows: In order to get an idea of the number of outliers that can be handled, Section 2 starts by proving an ℓ_0 minimization result. In Section 3 we present the numerical algorithms for solving $(\mathrm{P}_{1,1})$, $(\mathrm{P}_{2,1,1})$ and their TV counterparts $(\mathrm{P}_{1,\mathrm{TV}})$ and $(\mathrm{P}_{2,1,\mathrm{TV}})$. Numerical examples examining the reconstruction capability of our algorithms w.r.t. the sparsity of x_0 and e_0 are presented in Section 4. In partucular, we numerically evaluate the probability that the solution of $(\mathrm{P}_{1,1})$ coincides with the original x_0 in dependence on the sparsity $\|x_0\|_0$ and $\|e_0\|_0$.

2. ℓ_0 Minimization

The starting point for examining the recovering of signals from incomplete data was the question if it is possible to reconstruct a sparse vector x_0 from its incomplete measurements $b := Ax_0$ by solving

$$(\text{P}_0) \qquad \|x\|_0 \rightarrow \min \qquad \text{s.t.} \quad Ax = b.$$

One result can be formulated in terms of the smallest number of linearly dependent columns of A, denoted by $\text{spark}(A)$, cf. [16].

Proposition 2.1. *If* $\text{spark}(A) > 2m$*, then, for every* $x_0 \in \mathbb{C}^N$ *with* $\|x_0\|_0 \leq m$ *and given* $b := Ax_0$*, the solution of* (P_0) *is unique and coincides with* x_0*. Conversely, if* $\text{spark}(A) \leq 2m$*, then there exist distinct vectors* x_0 *and* x_1 *with* $\|x_0\|_0, \|x_1\|_0 \leq m$ *such that* $Ax_0 = Ax_1$*.*

In particular, if every set of n columns of A is linearly independent, i.e., $\text{spark}(A) = n+1$, then perfect reconstruction is guaranteed for every $x_0 \in \mathbb{C}^N$ with $\|x_0\|_0 \leq m$ if and only if $n \geq 2m$.

To get a clue about the influence of the outliers, let us first assume that the positions $\Omega \subset \{1, \ldots, n\}$ of the K outliers are known. By $\bar{\Omega}$ we denote the complement of Ω in $\{1, \ldots, n\}$. Since the outliers carry no information about x_0 the best we can do is to solve

$$\arg\min_x \|x\|_0 \qquad \text{s.t.} \quad A|_{\bar{\Omega}} x = b|_{\bar{\Omega}},$$

where $b|_{\bar{\Omega}}$ denotes the restriction of b to those components with indices in $\bar{\Omega}$ and $A|_{\bar{\Omega}} \in \mathbb{C}^{n-K,N}$ contains the rows of A with indices in $\bar{\Omega}$. By Proposition 2.1 this problem has the unique solution x_0 if $2m < \text{spark}(A|_{\bar{\Omega}})$. Thus, if every set of $n - K$ columns of $A|_{\bar{\Omega}}$ is linearly independent, perfect reconstruction is guaranteed if $n - K \geq 2m$, i.e., $n \geq 2m + K$.

In general, we have no oracle that gives us the position of the outliers. Thus, given $b := Ax_0 + e_0$, we are looking for conditions such that

$$(\tilde{\text{P}}_{0,0}) \qquad \arg\min_{x,e} \|e\|_0 + \lambda\|x\|_0 \qquad \text{s.t.} \quad Ax + e = b$$

has the unique solution x_0. At least for matrices having only invertible quadratic submatrices a sufficient condition will be proved in the next proposition. Examples of such matrices are the N–th Fourier matrix $F_N := (e^{2\pi ijk/N})_{j,k=0}^{N-1}$ of prime size $N = p$, cf. [28] and the Toeplitz matrix with entries from the Gaussian radial basis function $(e^{-\sigma(j-k)^2})_{j,k=0}^{N-1}$, cf. [22].

Proposition 2.2. *Let* $x_0 \in \mathbb{C}^N$ *with* $\|x_0\|_0 = m$ *and* $e_0 \in \mathbb{C}^n$ *with* $\|e_0\|_0 = K$ *be given. Suppose that* $A \in \mathbb{C}^{n,N}$*,* $n \leq N$ *has only invertible submatrices. Let* $b := Ax_0 + e_0$*. If* $n \geq 2K + (\lambda+1)m$*, then, for* $\lambda \geq 1$*, problem* $(\tilde{\text{P}}_{0,0})$ *has the unique solution* $\hat{x} = x_0$*.*

Proof. By assumption we have that $\|e_0\|_0 + \lambda\|x_0\|_0 = K + \lambda m$. Assume that there exists a solution $(\hat{x}, \hat{e}) \neq (x_0, e_0)$ of $(\tilde{\text{P}}_{0,0})$. Then this solution must fulfill $A\hat{x} + \hat{e} = Ax_0 + e_0 = b$ and $\|\hat{e}\|_0 + \lambda\|\hat{x}\|_0 \leq K + \lambda m$. Let $\|\hat{x}\|_0 = s$ so that $\|\hat{e}\|_0 \leq K + \lambda m - \lambda s$. Then we have for $T = \text{supp}(\hat{x} - x_0)$ and $\Omega = \text{supp}(\hat{e} - e_0)$ that $|T| \leq m + s$ and $|\Omega| \leq 2K + \lambda m - \lambda s$. Moreover, since $A(\hat{x} - x_0) = e_0 - \hat{e}$ we obtain that $A|_{\bar{\Omega},T}(\hat{x} - x_0) = 0$, where $A|_{\bar{\Omega},T}$ denotes the restriction of A to the rows having indices in $\bar{\Omega}$ and to the columns with indices in T. Since $n \geq 2K + (\lambda+1)m$ and $\lambda \geq 1$ we have that $n \geq 2K + (\lambda+1)m + (1-\lambda)s$ and consequently $n - (2K + \lambda m - \lambda s) \geq$

$m + s$. Thus, $|\bar{\Omega}| \geq n - (2K + \lambda m - \lambda s) \geq m + s \geq |T|$ so that $A|_{\bar{\Omega}, T}$ is injective. Hence $\hat{x} - x_0 = 0$ and the assertion is proved.   □

3. NUMERICAL ALGORITHMS

Since the practical solution of (P_0) is too expansive (exponentially increasing in the problem size N) the ℓ_0–minimization was only discussed theoretically, while for practical computations the ℓ_0–seminorm was replaced by the ℓ_1–norm. Before providing the numerical algorithms for solving the corresponding minimization problems we prove the following simple, but interesting proposition which can be summarized as follows: The fact that x_0 is a solution of $(P_{1,1})$ does not depend on the magnitude of the outliers e_0, see also [26].

Proposition 3.1. *Let x_0 be a solution of $(P_{1,1})$, where $b := Ax_0 + e_0$. Then we have for any $\tilde{x}_0$ with $\mathrm{sgn}(\tilde{x}_0) = \mathrm{sgn}(x_0)$ and any $\tilde{e}_0$ with $\mathrm{sgn}(\tilde{e}_0) = \mathrm{sgn}(e_0)$ (componentwise) that $\tilde{x}_0$ is a solution of $(P_{1,1})$, where $b := A\tilde{x}_0 + \tilde{e}_0$.*

Proof. Let $\hat{x} = x_0 + \hat{f}$ be the solution of $(P_{1,1})$ with $b = Ax_0 + e_0$. Setting $x = x_0 + f$, we see that this is the case if and only if $\hat{f}$ is a solution of

$$(1) \qquad \arg\min_f \|e_0 - Af\|_1 + \lambda\|x_0 + f\|_1.$$

Since the functional is convex we know that $\hat{f}$ is a solution of (1) if and only if the zero vector is in the subdifferential of the functional at $\hat{f}$, i.e.,

$$0 \in A^{\mathrm{T}} \frac{A\hat{f} - e_0}{|A\hat{f} - e_0|} + \lambda \frac{x_0 + \hat{f}}{|x_0 + \hat{f}|},$$

where the quotient is meant componentwise and as usual $x/|x| := [-1, 1]$ if $x = 0$. In particular, we have that x_0 is a solution of $(P_{1,1})$ if and only if $\hat{f} = 0$ is a solution of (1) if and only if

$$0 \in A^{\mathrm{T}} \frac{e_0}{|e_0|} + \lambda \frac{x_0}{|x_0|}.$$

However, the right–hand side does only depend on $\mathrm{sgn}(x_0)$ and $\mathrm{sgn}(e_0)$ but not on their sizes. Hence we are done.   □

Problem $(P_{1,1})$. If the matrix A is real–valued, we can compute a minimizer of $(P_{1,1})$ by applying the algorithm proposed in [18]. To this end, we use the decomposition $x_{j+} := \max\{x_j, 0\}$, $x_{j-} := -\min\{x_j, 0\}$. Then, $x_j = x_{j+} - x_{j-}$ and $|x_j| = x_{j+} + x_{j-}$. Similarly, we decompose $\frac{1}{\lambda}(Ax - b) = y_+ - y_-$ and set $X := (x_+^{\mathrm{T}}, x_-^{\mathrm{T}}, y_+^{\mathrm{T}}, y_-^{\mathrm{T}})^{\mathrm{T}}$. Hence, we can solve the following linear program using MATLAB resp. CPLEX LINPROG

$$(2) \qquad \arg\min_x 1_{2N+2n}^{\mathrm{T}} X \quad \text{s.t.} \quad \left\{ \begin{array}{rcl} \left[\frac{1}{\lambda} A, -\frac{1}{\lambda} A, -I_n, I_n \right] X & = & \frac{1}{\lambda} b, \\ X & \geq & 0, \end{array} \right.$$

where 1_N denotes the vector consisting of N components 1.

In case of a complex–valued matrix A, we can compute the minimizer by second-order cone programming (SOCP), cf. [21], [31]. It is easily seen that a solution to $(P_{1,1})$ can be found by solving

$$(3) \quad \arg\min \ 1_N^{\mathrm{T}} u + \lambda \, 1_N^{\mathrm{T}} v \quad \text{s.t.} \quad \begin{cases} \tilde{A} \begin{pmatrix} x^{\mathrm{real}} \\ x^{\mathrm{imag}} \end{pmatrix} - \tilde{b} = \begin{pmatrix} s^{\mathrm{real}} \\ s^{\mathrm{imag}} \end{pmatrix} \\ (u_j, s_j^{\mathrm{real}}, s_j^{\mathrm{imag}}) \in \mathcal{K}^3, \quad \forall j = 1, \ldots, N \\ (v_j, x_j^{\mathrm{real}}, x_j^{\mathrm{imag}}) \in \mathcal{K}^3, \quad \forall j = 1, \ldots, N \end{cases}$$

where

$$\tilde{A} = \begin{pmatrix} \mathrm{real}(A) & -\mathrm{imag}(A) \\ \mathrm{imag}(A) & \mathrm{real}(A) \end{pmatrix}, \quad \tilde{b} = \begin{pmatrix} \mathrm{real}(b) \\ \mathrm{imag}(b) \end{pmatrix}.$$

The cone $\mathcal{K}^p$ is defined by

$$\mathcal{K}^p = \{(t_0, \ldots, t_{p-1}) \in \mathbb{R}^p \ : \ \|(t_1, \ldots, t_{p-1})\|_2 \le t_0\}.$$

We used MOSEK to solve this SOCP.

Problem $(\mathrm{P}_{2,1,1})$. We solve $(\mathrm{P}_{2,1,1})$ with the following alternating minimization algorithm suggested in [2] in the context of image decomposition. As initial value we use $e^{(0)} := 0$ but any other initialization is possible.

Algorithm 3.2. *Initialization:* $e^{(0)} \in \mathbb{R}^n$
For $k = 0, \ldots$ repeat until a stopping criterion is reached

   1. For given $e^{(k)}$ and $y := b - e^{(k)}$ compute the solution $x^{(k)}$ of

$$(4) \qquad \arg\min_x \ \frac{1}{2\alpha} \|y - Ax\|_2^2 + \lambda \|x\|_1.$$

   2. For given $x^{(k)}$ and $\tilde{y} := b - Ax^{(k)}$ compute the solution $e^{(k+1)}$ of

$$(5) \qquad \arg\min_e \ \frac{1}{2\alpha} \|\tilde{y} - e\|_2^2 + \|e\|_1.$$

Let us have a look at the two subproblems. In the second step, the functional is strictly convex and has a unique solution $\hat{e}$ which can be obtained by soft shrinkage of $\tilde{y}$ with threshold α cf. [3], i.e., $\hat{e} = S_\alpha(\tilde{y})$, where

$$S_\alpha(x) := \begin{cases} x - \alpha & \text{for } x > \alpha, \\ x + \alpha & \text{for } x < -\alpha, \\ 0 & \text{for } x \in [-\alpha, \alpha]. \end{cases}$$

In the first step, there exists a minimizer due to convexity and coercivity of the functional but uniqueness is not guaranteed. Since the computation of the minimizer requires similar consideration as the approach to $(\mathrm{P}_{2,1,\mathrm{TV}})$ in the next paragraph, we consider more generally

$$(6) \qquad \arg\min_x \ \frac{1}{2} \|y - Ax\|_2^2 + \beta \| \, |Lx| \, \|_1,$$

where $L \in \mathbb{R}^{pN, N}$ and $|X| := \big((\sum_{k=0}^{p-1} X_{j+kp}^2)^{1/2} \big)_{j=0}^{N-1}$. The functional $\mathcal{J}(x) := \| \, |Lx| \, \|_1$ is convex and one-homogeneous such that its dual $\mathcal{J}^*$ is the indicator function of the set

$$\begin{aligned} \mathcal{C} \quad &:= \quad \{x \in \mathcal{R}(L^{\mathrm{T}}) : \langle x, y \rangle \le \mathcal{J}(y) \quad \forall y\} \\ (7) \qquad &= \quad \{x \in \mathcal{R}(L^{\mathrm{T}}) : \min_{x = L^{\mathrm{T}} X} \| \, |X| \, \|_\infty \le 1\}, \end{aligned}$$

see [13],[27]. Then the following proposition holds true, cf. [4].

Proposition 3.3. *Let $\mu > 0$ be chosen such that $\mu\|A^*A\|_2 < 1$. Then any solution of the fixed point equation*

$$(8) \qquad\qquad x = (I - \Pi_{\mu\beta\mathcal{C}})(x + \mu A^*(y - Ax))$$

is a solution of (6). *Assuming the existence of a solution of* (6), *the sequence*

$$(9) \qquad\qquad x^{(k+1)} = (I - \Pi_{\mu\beta\mathcal{C}})(x^{(k)} + \mu A^*(y - Ax^{(k)}))$$

converges for any initial vector $x^{(0)}$ to a fixed point of (8).

For our problem $(P_{2,1,1})$ we have that $\beta = \alpha\lambda$, $L = I$ and $p = 1$. Consequently, $\mathcal{C} = \{x \in \mathbb{R}^N : \|x\|_\infty \le 1\}$ and

$$\Pi_{\mu\beta\mathcal{C}}(y) = \arg\min_u \|y - u\|_2^2 \qquad \text{s.t.} \quad \|u\|_\infty \le \beta\mu.$$

This can be solved componentwise, i.e., for $j = 0, \dots, N-1$ we have to compute

$$(10) \qquad\qquad \arg\min_{u_j} |y_j - u_j|_2 \qquad \text{s.t.} \quad |u_j| \le \beta\mu.$$

Obviously, the solution of (10) is given by $\beta\mu$ if $y_j \ge \beta\mu$, by $-\beta\mu$ if $y_j \le -\beta\mu$ and by y_j otherwise. Hence, $I - \Pi_{\mu\beta\mathcal{C}}$ is the soft shrinkage operator with threshold $\beta\mu$. In summary, we realize step 1 of Algorithm 3.2 by

$$(11) \qquad\qquad x^{(k+1)} = S_{\beta\mu}\left(x^{(k)} + \mu A^*(y - Ax^{(k)})\right).$$

For other approaches to obtain (11), e.g., via surrogate techniques, see [15, 5]. Concerning the convergence of the alternating minimization algorithm we have the following proposition.

Proposition 3.4. *Let $F(x, e) := \frac{1}{2\alpha}\|b - Ax - e\|_2^2 + \|e\|_1 + \lambda\|x\|_1$. Then, for every sequence $(x^{(k)}, e^{(k)})_k$ obtained by the alternating minimization Algorithm 3.2 it holds that*

$$(12) \qquad\qquad \lim_{k\to\infty} F(x^{(k)}, e^{(k)}) = \gamma,$$

where $\gamma := \min_{x,e} F(x, e)$.

The proof applies the ideas of [1, 2] to our setting in a straightforward way and is left to the reader.

Problems $(P_{1,\mathrm{TV}})$ and $(P_{2,1,\mathrm{TV}})$. Finally, we deal with images $X \in \mathbb{R}^{\tilde{N},\tilde{N}}$ having a sparse gradient rather than being sparse themself. A typical example of such an image is the Shepp–Logan phantom in Fig. 2. We reshape these images columnwise into a vector x of length $N = \tilde{N}^2$. For defining a discrete version of the TV–norm we introduce the directional forward difference matrix

$$\mathcal{D} := \begin{pmatrix} I_{\tilde{N}} \otimes D \\ D \otimes I_{\tilde{N}} \end{pmatrix} \in \mathbb{R}^{2N,N} \quad \text{with} \quad D := \begin{pmatrix} -1 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 & 0 \\ & & & \ddots & \ddots & \ddots & \\ 0 & 0 & 0 & \dots & -1 & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & -1 & 1 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \end{pmatrix}$$

and set

$$\|x\|_{\mathrm{TV}} := \| \, |\mathcal{D}x| \, \|_1.$$

Problem $(P_{2,1,\mathrm{TV}})$ can be solved by an alternating minimization algorithm similar to Algorithm 3.2. We only have to replace $\|x\|_1$ in (4) by $\|x\|_{\mathrm{TV}}$. Then $L = \mathcal{D}$ and $p = 2$ in (6). Of course this results in another indicator set $\mathcal{C}$ and consequently in

another projection operator $\Pi_{\mu\beta\mathcal{C}}$. The resulting projections can be computed, e.g., by using Chambolle's algorithm [13], see also [27].

For solving problem (P$_{1,\mathrm{TV}}$) one can try to follow the lines of [14] or [26]. Both algorithms introduce an additional parameter ε in the data-fitting and/or the regularization term to cope with the singularity of the absolute value function at zero and converge rather slowly in general. Another possibility to solve (P$_{1,\mathrm{TV}}$) is SOCP. This was recently proposed for several image processing problems, e.g. in [31] and [21]. To this end, we rewrite (P$_{1,\mathrm{TV}}$) as follows:

$$
\begin{aligned}
&\arg\min \ 1_N^{\mathrm{T}} u + \lambda\, 1_N^{\mathrm{T}} v \\
&\text{s.t.} \quad Ax - b = s \\
&\qquad\quad \begin{pmatrix} d_x \\ d_y \end{pmatrix} = \mathcal{D}x \\
&\qquad\quad (u_j, s_j) \in \mathcal{K}^2 \\
&\qquad\quad (t_j, (d_x)_j, (d_y)_j) \in \mathcal{K}^3
\end{aligned}
$$

(13)

We restrict our attention to real matrices A. However, the corresponding problem for complex matrices can also be solved by SOCP since we can separate the real and complex parts of A and b. This results in cones of 'dimension' 3 and 5.

4. Numerical Examples

In our numerical examples we consider two types of matrices: First, real matrices $A \in \mathbb{R}^{n,N}$ which are constructed by randomly selecting n columns from the orthogonal discrete cosine transform matrix of typ II:

$$
C_N := \left(\frac{2}{N} \right)^{1/2} \left(\varepsilon_j \cos \frac{j(2k+1)\pi}{2N} \right)_{j,k=0}^{N-1},
$$

where $\varepsilon_0 = 1/\sqrt{2}$ and $\varepsilon_j = 1$ for $j \neq 0$. We mention that we have obtained similar results for the solution of (P$_{1,1}$) and (P$_{2,1,1}$) with matrices having Gaussian random numbers with mean zero and variance $1/n$ as entries. In our tests we choose $N = 64$. The second class of matrices we considered consists of randomly chosen columns of the discrete Fourier transform matrix of length N

$$
F_N := \frac{1}{\sqrt{N}} \left(e^{-\frac{2\pi ijk}{N}} \right)_{j,k=0}^{N-1}.
$$

We are interested in the probability that the solution $\hat{x}$ of (P$_{1,1}$) coincides with x_0 for various values of $m = \|x_0\|_0$ and $K = \|e_0\|_0$. We say that we have recovered x_0 exactly if $\|\hat{x} - x_0\|_\infty < 10^{-4}$. Note that we obtained similar results for the thresholds 10^{-6} and 10^{-8}. The values for the m non-zero components of x_0 were drawn from a uniform distribution on $[-1, -0.1] \cup [0.1, 1]$. The K outliers were randomly chosen from the set $\{e_{\min} = \min(Ax_0),\ e_{\max} = \max(Ax_0)\}$ in case of the cosine matrix. For complex matrices A we use the two values

$$
e_{\min} = \min(\mathrm{real}(Ax_0)) + i\,\min(\mathrm{imag}(Ax_0)),
$$
$$
e_{\max} = \max(\mathrm{real}(Ax_0)) + i\,\max(\mathrm{imag}(Ax_0)).
$$

The parameter λ was set to 1. The experiment was repeated 5000 times for every m, K.

In our first example we considered the matrices $A = C_N$ and $A = F_N$ for $N = n = 64$ as well as matrices which were constructed by randomly choosing $n = 40$

rows of the above two matrices. Fig. 1 shows the probability that the solution $\hat{x}$ of $(P_{1,1})$ coincides with x_0 in dependence of m and K for these four matrices.
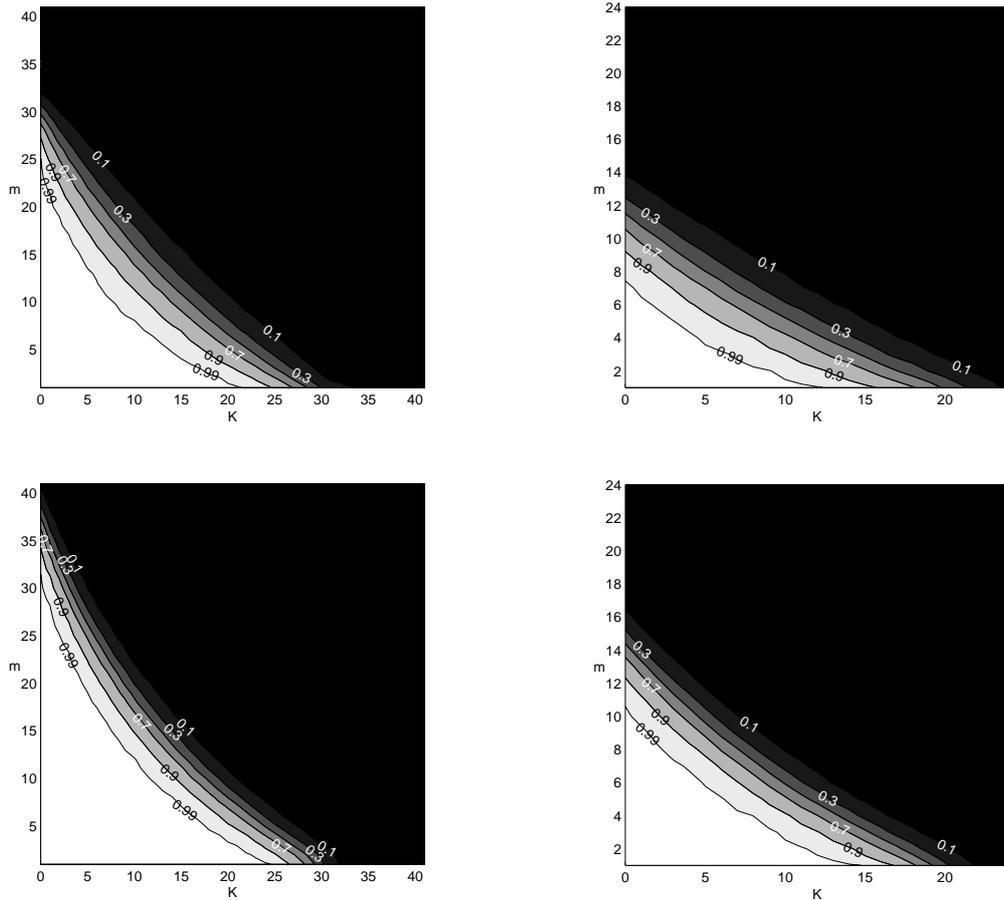
FIGURE 1. Probability that the solution $\hat{x}$ of $(P_{1,1})$ coincides with x_0 in dependence on m and K for different matrices A. Top left: $A = C_N$, i.e. the full DCT-II matrix ($n = N = 64$). Top right: A was constructed by randomly selecting $n = 40$ rows of C_N. Bottom left: $A = F_N$ for $n = N = 64$. Bottom right: $n = 40$ rows randomly chosen rows of F_N.

Finally, we are interested in images having a sparse gradient. We use the Shepp-Logan phantom image ($N = 64$). The measurement operator A is constructed as follows. First, we construct two matrices $A_1, A_2 \in \mathbb{R}^{n,N}$ by randomly selecting $n = 40$ rows of the matrix C_N, then we set $A = A_1 \otimes A_2$.

As before, we corrupt the measured data by $K = 16$ min-max outliers. Furthermore, we add a small amount of Gaussian noise. Solving $(P_{2,1,TV})$ and $(P_{1,TV})$ yields the results shown in Fig. 3 and Fig. 4. The solution to problem $(P_{2,1,TV})$ was computed by applying Algorithm 3.2 to the two-dimensional case (parameters: $\mu = 0.9, \lambda = 0.003, \alpha = 0.04$ for Gaussian noise with standard deviation 0.001 and $\alpha = 0.03$ for Gaussian noise with standard deviation 0.01, respectively). The algorithm was stopped when the relative distance between two consecutive images produced by the algorithm, measured in the Frobenius norm, was smaller than 10^{-8}. A solution to problem $(P_{1,TV})$ was found by solving the SOCP (13) with MOSEK (parameters: $\lambda = 5$ for Gaussian noise with standard deviation 0.001 and $\lambda = 6$ for Gaussian noise with standard deviation 0.01, respectively). Fig. 3 and Fig. 4 show that the proposed methods perform remarkably well.
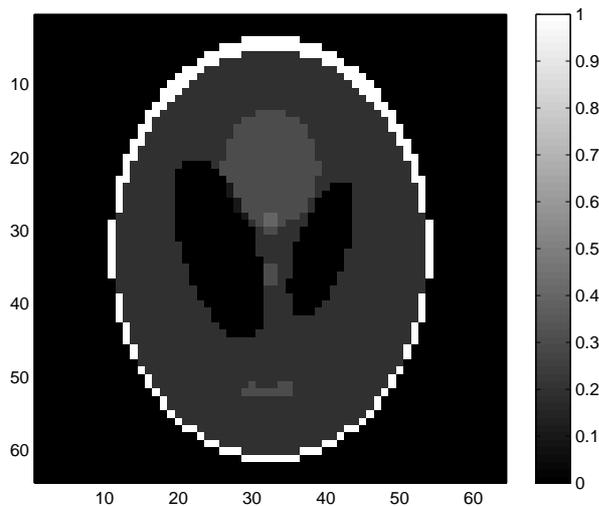
FIGURE 2. The Shepp-Logan phantom test image.

References

[1] J. F. Aujol, G. Aubert, L. Blanc-Féraud, and A. Chambolle, *Image decomposition into a bounded variation component and an oscillating component*, Journal of Mathematical Imaging and Vision, **22** (2005), 71–88.

[2] J. F. Aujol and A. Chambolle, *Dual norms and image decomposition models*, Technical Report 5130, INRIA Report, Sophia Antipolis Cedex, France, 2004.

[3] J.-F. Aujol, G. Gilboa, T. Chan, and S. Osher, *Structure–texture image decomposition – modelling, algorithms and parameter selection*, International Journal of Computer Vision, **67(1)** (2006), 111–136.
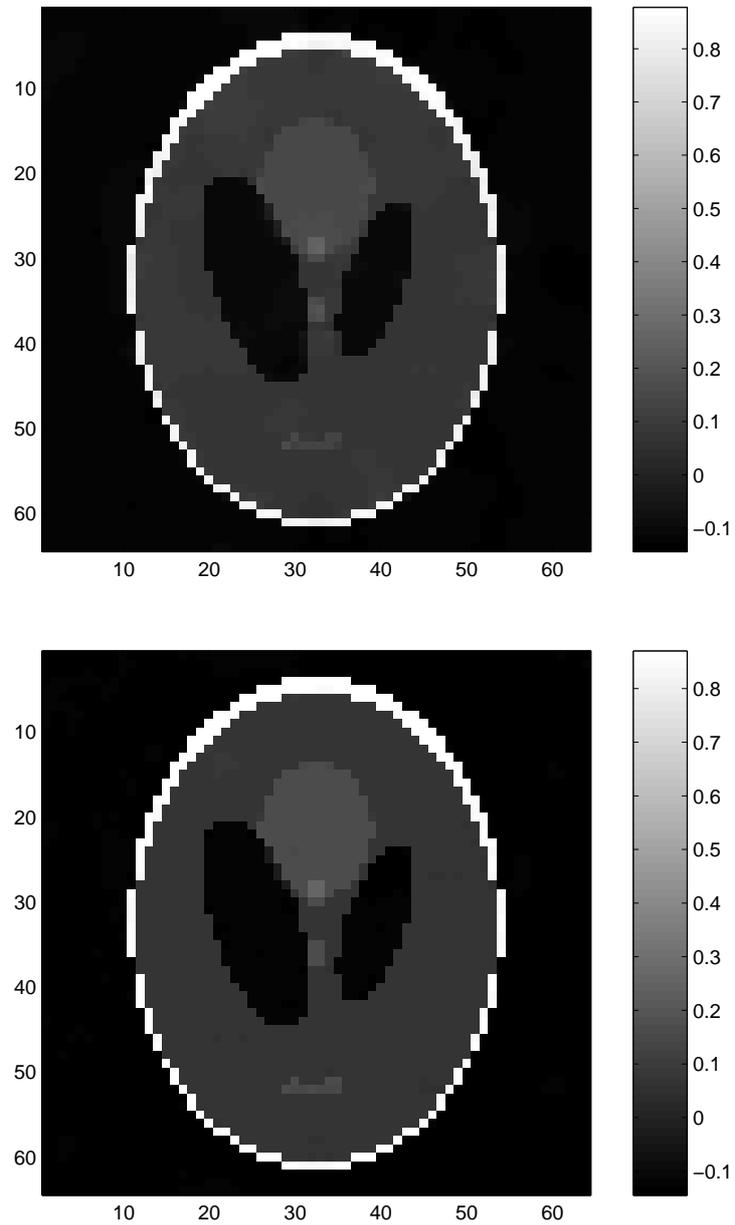
FIGURE 3. Top: Reconstruction of the Shepp-Logan phantom test image via $(P_{2,1,TV})$. Noise: $K = 16$ min-max outliers and Gaussian white noise with standard deviation 0.001. The difference between the reconstructed and the original image measured in terms of the Frobenius norm is 7.8671. Bottom: Recovered image using $(P_{1,TV})$. Frobenius norm difference to the original image: 8.4960.
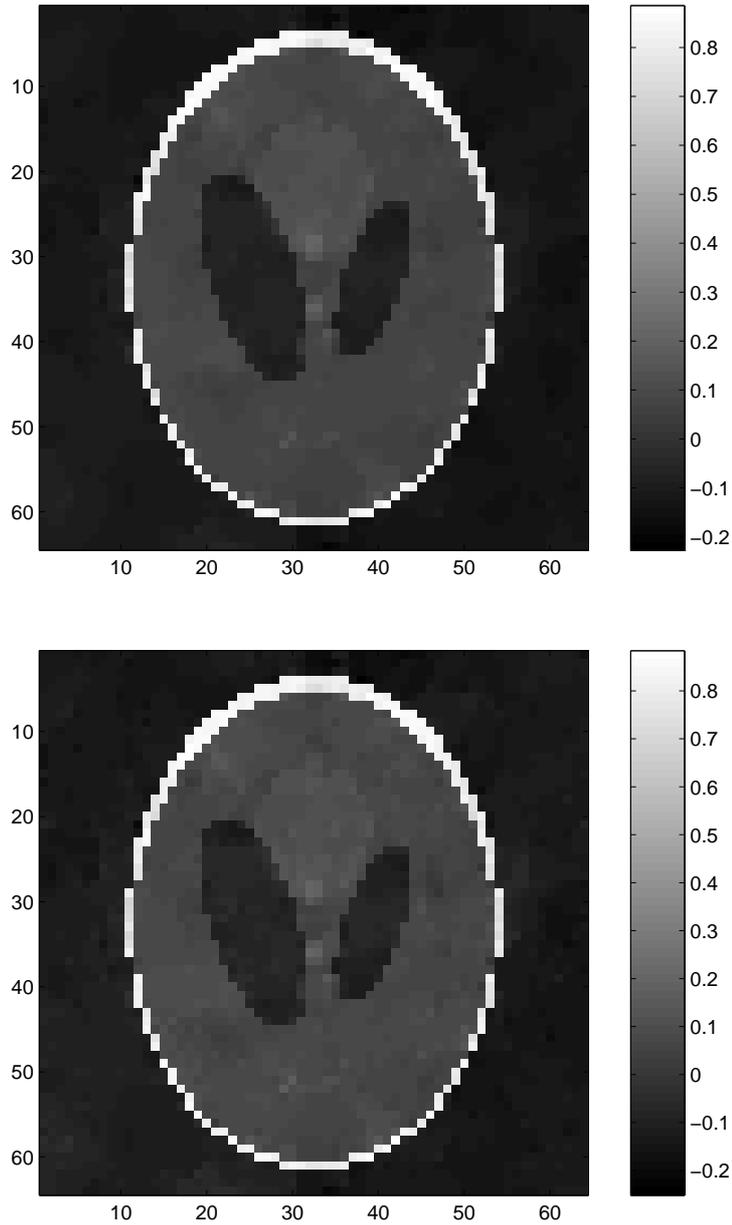
Figure 4. Top: Reconstruction using $(\mathrm{P}_{2,1,\mathrm{TV}})$. Noise: $K = 16$ min-max outliers and Gaussian white noise with standard deviation 0.01. Frobenius norm difference to the original image: 8.0470. Bottom: Recovered image using $(\mathrm{P}_{1,\mathrm{TV}})$. Difference between the reconstructed and the original image: 8.6429.

[4] J. Bect, L. Blanc-Féraud, G. Aubert, and A. Chambolle, *A l^1-unified variational framework for image restauration*, in "Proc. European Conference on Computer Vision" (eds. T. Pajdla and J. Matas), volume 3024 of *Lecture Notes in Computer Science*, Prague, Springer (2004), 1–13.

[5] K. Bredies, D. A. Lorenz, and P. Maass, *Equivalence of a generalized conditional gradient method and the method of surrogate functionals*, Technical report, Preprint, University of Bremen, 2005.

[6] E. J. Candes and P. A. Randall, *Highly robust error correction against gross and small errors*, Technical report, Preprint, Caltech, 2006.

[7] E. J. Candes, J. Romberg, and T. Tao, *Robust uncertainty principles: exact signal recovery from highly incomplete frequency information*, IEEE Transactions on Information Theory, **52(2)** (2006), 489–509.

[8] E. J. Candes, J. Romberg, and T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, Comm. Pure Appl. Math., **59(8)** (2006), 1207–1223.

[9] E. J. Candes and T. Tao, *Decoding by linear programming*, IEEE Transactions on Information Theory, **51(12)** (2005), 4203–4215.

[10] E. J. Candes and T. Tao, *Near optimal signal recovery from random projections: universal encoding strategies?*, IEEE Transactions on Information Theory, **52(12)** (2006), 5406–5425.

[11] E. J. Candes and T. Tao, *The dantzig selector: statistical estimation when p is much larger than n*, Annals of Statistics, to appear, 2005.

[12] E. J. Candes and T. Tao, *Error correction via linear programming*, in "Proc. 46th Annual IEEE Symposium on Foundations of Computer Science", IEEE, 2005, 295–308.

[13] A. Chambolle, *An algorithm for total variation minimization and applications*, Journal of Mathematical Imaging and Vision, **20(1-2)** (2004), 89–97.

[14] T. Chan and S. Esedoglu, *Aspects of total variation regularized L^1 function approximation*, SIAM Journal on Applied Mathematics, **65(5)** (2005), 1817–1837.

[15] I. Daubechies, M. Defrise, and C. De Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Comm. Pure Appl. Math., **51** (2004), 1413–1541.

[16] D. Donoho and M. Elad, *Optimally sparse representation in general (non–orthogonal) dictionaries via ℓ_1 minimization*, Proceedings of the National Academy of Sciences, **100(5)** (2003), 2197–2202.

[17] D. L. Donoho, M. Stark, and V. Temlyakov, *Stable recovery of sparse overcomplete representations in the presence of noise*, IEEE Transactions on Information Theory, **52(1)** (2006), 6–18.

[18] H. Fu, M. K. Ng, M. Nikolova, and J. L. Barlow, *Efficient minimization methods of mixed ℓ_2-ℓ_1 and ℓ_1-ℓ_1 norms for image restauration*, SIAM Journal on Scientific Computing, **27(6)** (2006), 1881–1902.

[19] J. J. Fuchs, *On sparse representations in arbitrary redundant bases*, IEEE Transactions on Information Theory, **50(6)** (2004), 1341–1344.

[20] J. J. Fuchs, *Recovery of exact sparse representations in the presence of bounded noise*, IEEE Transactions on Information Theory, **51(10)** 2005, 3601–3608.

[21] D. Goldfarb and W. Yin, *Second-order cone programming methods for total variation-based image restoration*, SIAM Journal on Scientific Computing, **27(2)** (2005), 622–645.

[22] S. Karlin, "Total Positivity," Stanford University Press, Stanford, 1968.

[23] J. Kunis and H. Rauhut, *Random sampling of sparse trigonometric polynomials II – orthogonal matching pursuit versus basis pursuit*, Technical report, University of Chemnitz, Chemnitz, 2006.

[24] M. Nikolova, *Weakly constrained minimization. Application to the estimation of images and signals involving constant regions*, Technical Report 2001-23, Centre de Mathématiques et de Leurs Applications, ENS de Cachan, France, 2001.

[25] M. Nikolova, *Minimizers of cost-functions involving nonsmooth data-fidelity terms. Application to the processing of outliers*, SIAM Journal on Numerical Analysis, **40(3)** (2002), 965–994.

[26] M. Nikolova, *A variational approach to remove outliers and impulse noise*, Journal of Mathematical Imaging and Vision, **20** (2004), 99–120.

[27] G. Steidl, *A note on the dual treatment of higher–order regularization functionals*, Computing, **76** (2006), 135–148.

[28] T. Tao, *An uncertainty principle for cyclic groups of prime order*, Math. Res. Letters, **12** (2005), 121–127.

[29] J. A. Tropp, *Greed is good: algorithmic results for sparse approximation*, IEEE Transactions on Information Theory, **50(10)** (2004), 2231–2242.

[30] J. A. Tropp, *Just relax: convex programming methods for identifying sparse signals in noise*, IEEE Transactions on Information Theory, **52(3)** (2006), 1030–1051.

[31] W. Yin, D. Goldfarb, and S. Osher, *Image cartoon-texture decomposition and feature selection using the total variation regularized L^1 functional*, Technical Report 05-47, UCLA-Reports, August 2005.

Received September 2004; revised February 2005.

E-mail address: steidl@math.uni-mannheim.de