A monotone scheme for sparsity optimization in ℓ^p with $p \in (0, 1]$

Daria Ghilli * **Karl Kunisch** **

* *Institute of Mathematic and Scientific Computing, University of Graz, Austria (email: daria.ghilli@uni-graz.at)*
** *Institute of Mathematic and Scientific Computing, University of Graz, Austria (email: karl.kunisch@uni-graz.at)*

Abstract: Nonsmooth nonconvex optimization problems are considered in infinite dimensional sequence spaces ℓ^p with $p \in (0, 1]$. Our starting points are necessary optimality conditions in the form of a complementary system and a monotonically convergent algorithm for a regularized version of the original problem. We propose an algorithm for solving the necessary optimality condition based on a combination of the monotone scheme and an active-set strategy. Numerical results for different test cases are provided, e.g. for optimal control problems and microscopy image reconstruction.

Keywords: nonsmooth nonconvex optimization, sparsity optimization, active-set method, monotone algorithm, optimal control problems, image reconstruction

1. INTRODUCTION

We consider the following nonconvex nonsmooth optimization problem

$$\min_{x \in \ell^p} J(x) = \frac{1}{2}|Ax - b|_2^2 + \beta |x|_p^p, \tag{1}$$

$$\ell^p = \{x \in \ell^2 : \textstyle\sum_{k=1}^{\infty} |x_k|^p < \infty\}, p \in (0, 1] ,$$

$$|x|_p = \left(\sum_{k=1}^{\infty} |x_k|^p \right)^{\frac{1}{p}} ,$$

which is a norm if $p = 1$ and a quasi-norm for $0 < p < 1$. Optimization of ℓ^p-functionals as in (1) arises frequently in sparse recovery as an efficient way to extract the essential features of generalized solutions. In signal processing, especially compressed sensing (see Chartrand (2009), Nikolova et al. (2008)), sparsity can be effectively utilized for data acquisition, signal trasmission, storage and processing (see e.g. Candes and Tao (2005)). Sparse recovery has attracted increasing attention also in statistics as an efficient variable selection tool (see Tibshirani (1996)). In image analysis ℓ^p-functionals with $p \in (0, 1)$ have recently been proposed as nonconvex and nonsmooth extensions of the total generalized variation regularizer (see Hintermüller and Wu (2013), Ochs et al. (2015) and the references therein). Also, the use of ℓ^p-functionals with $p \in (0, 1)$ is of particular importance in fracture mechanics (see Pietro (2013)). The literature on sparsity optimization problems as (1) is rapidly increasing, here we mention also Bredies and Lorentz (2015), Fornasier and Ward (2010), Li and Pong (2014), Jiao et al. (Preprint 2013), Ramlau and Zarzer (2012).

In Ito and Kunisch (2014) existence for (1) is proven and convenient necessary optimality conditions are given (see Section 2). These conditions are of complementary type, such that they do not require the a-priori knowledge whether a specific coordinate of an optimal solution is different from zero or not. Rather this distinction is built into the optimality condition itself. Here we focus on the development of a numerical scheme for solving the above mentioned necessary optimality conditions. This is an important issue in the study of problems as (1), since, due to the behaviour of $s \in \mathbb{R} \to |s|^p \in \mathbb{R}^+$ at $s = 0$, problems (1) are nonsmooth and nonconvex, and hence standard algorithms are not readily available. Our scheme relies deeply on the particular form of the optimality conditions, which allows us to develop an active-set strategy. Then, the nonlinear equation satisfied by the non zero components is solved through a monotonically convergent iteration procedure. The monotone scheme and the active-set monotone algorithm are presented in Section 3 and 4, respectively.

The performance of the monotone active-set algorithm is successfully tested in three different cases, the M-matrix problem, a time-dependent optimal control example and an application to microscopy image reconstruction (see Section 5). The interest in the M-matrix problem is mainly academical, since in this case convergence of the proposed algorithm can readily be analyzed.

In the recent work Jiao et al. (Preprint 2013) a primal dual active set strategy is proposed for problems similar to (1), but, differently from our approach, no iteration procedure is used to solve the nonlinear problem arising at each iteration level of the active set scheme and no monotonicity is investigated.

Concerning the general importance of ℓ^p-functionals with $p \in (0, 1)$, numerical experiments have shown that their use can promote sparsity better than the ℓ^1-norm, allowing possibly a smaller number of measurements in feature selection and compressed sensing (see e.g. Chartrand (2009), Nikolova et al. (2010)) and providing better edge preservation in total variation-based image (see e.g. Nikolova et al.

⋆ Supported by the ERC advanced grant 668998 (OCLOC) under the EU's H2020 research programme.

(2010), Roth and Black (2009)). Also, the use of nonconvex optimization can be considered from natural image statistics (Huang and Mumford (1999)) and it appears to be more robust with respect to heavy-tailed distributed noise.
Finally, we remark that problems as (1) with an operator inside the ℓ^p-term will be subject of future work.

2. NECESSARY OPTIMALITY CONDITIONS

We denote $A_i = (A, e_i)$, where e_i is the sequence with 1 in the i-th coordinate and 0 otherwise. In Ito and Kunisch (2014), Theorem 2.2, the following necessary optimality conditions are proven.

Theorem 1. (*Necessary optimality conditions*). If $\bar{x}$ is a global minimizer of (1), then we have

$$\begin{cases} \bar{x}_i = 0 & \text{if } |(A_i, f_i)| < \mu_i \\ (A_i, A\bar{x} - b) + \dfrac{\beta p \bar{x}_i}{|\bar{x}_i|^{2-p}} = 0 & \text{if } |(A_i, f_i)| > \mu_i, \end{cases} \tag{2}$$

where $f_i = b - A\bar{x} + A_i\bar{x}_i$, $\mu_i = \beta^{\frac{1}{2-p}}(2-p)(2(1-p))^{-\frac{1-p}{2-p}}|A_i|_2^{1-\frac{p}{2-p}}$.

If $|(A_i, f_i)| = \mu_i, \bar{x}_i = 0$ or $\bar{x}_i = \left(\frac{2\beta(1-p)}{|A_i|_2^2}\right)^{\frac{1}{2-p}} \operatorname{sgn}((A_i, f_i))$.

Remark 2. (i) From Theorem 1 it follows that a minimizer is not necessarily unique;

(ii) Condition (2) is of complementary type, separating the active components of $\bar{x}$ from the inactive by the sign of $|(A_i, f_i)| - \mu_i$. In order to obtain (2) the quadratic nature of the smooth term of the cost is used.

Corollary 3. Let $\bar{x}$ be a global minimizer of (1). Then

(i) *lower bound on the inactive components:*
If $\bar{x}_i \neq 0$, then $|\bar{x}_i| \geq \left(\frac{2\beta(1-p)}{|A_i|_2^2}\right)^{\frac{1}{2-p}}$;

(ii) *sparsity of the solution*:

$$|\{i : \bar{x}_i \neq 0\}| \leq |b|_2^2 \left(\frac{(2\beta)^{\frac{2}{p}}}{\sup_{i\in\mathbb{N}} |A_i|_2^2}(1-p)\right)^{\frac{p}{p-2}}.$$

3. MONOTONE CONVERGENT SCHEME FOR A REGULARIZED PROBLEM

Following Ito and Kunisch (2014), in order to overcome the singularity of $(|s|^p)' = \frac{ps}{|s|^{2-p}}$ near $s = 0$, we consider for $\varepsilon > 0$ the following regularized version of (1)

$$J_\varepsilon(x) = \frac{1}{2}|Ax - b|_2^2 + \beta\Psi_\varepsilon(|x|^2), \tag{3}$$

where $\varepsilon > 0$ and for $t \geq 0$

$$\Psi_\varepsilon(t) = \begin{cases} \dfrac{p}{2}\dfrac{t}{\varepsilon^{2-p}} + (1 - \dfrac{p}{2})\varepsilon^p & \text{for } 0 \leq t \leq \varepsilon^2 \\ t^{\frac{p}{2}} & \text{for } t \geq \varepsilon^2, \end{cases}$$

and $\Psi_\varepsilon(|x|^2)$ is short for $\sum_{i=1}^{\infty}\Psi_\varepsilon(|x_i|^2)$. Note that

$$\Psi'_\varepsilon(t) = \frac{p}{2\max(\varepsilon^{2-p}, t^{\frac{2-p}{2}})}, \quad \text{for } t \geq 0,$$

hence $\Psi \in C^1([0,\infty), \mathbb{R})$.

The necessary optimality condition for J_ε is given by

$$A^*Ax + \frac{\beta p}{max(\varepsilon^{2-p}, |x|^{2-p})}x = A^*b, \tag{4}$$

where the max-operation is interpreted coordinate wise. In order to solve (4) we use the following iteration procedure:

$$A^*Ax^{k+1} + \frac{\beta p}{max(\varepsilon^{2-p}, |x^k|^{2-p})}x^{k+1} = A^*b. \tag{5}$$

In Ito and Kunisch (2014), Theorem 4.1, the following convergence result is proven:

Theorem 4. For $\varepsilon > 0$ let $\{x^k\}$ be generated by (5). Then, $J_\varepsilon(x^k)$ is strictly monotonically decreasing unless there exists some k such that $x^k = x^{k+1}$ and x^k satisfies the necessary optimality condition (4). Moreover every weakly convergent subsequence of x^k, of which there exists at least one, converges in ℓ^2 to a solution of (4).

Remark 5. We successfully tested the monotone convergent algorithm in two different examples. The numerical strategy combines the iteration procedure (5) with a continuation strategy with respect to ε, that is, ε is incrementally decreased until the stopping critera to find a solution of (4) are satisfied. The stopping critera depend on the problem considered, in particular, in the M-matrix example the algorithm stops when two consecutive iterates coincide, in the control problem when the ℓ^∞ norm of the residue of (4) is less than 10^{-8}.
Our results show that the monotone algorithm is efficient in computing a solution of the ε-reguralized optimality condition (4). We remark that the value of the functional for each iteration was checked to be monotonically decreasing accordingly to Theorem 4.

4. THE ACTIVE-SET MONOTONE ALGORITHM

Note that the mononote algorithm presented in Section 3 depends on the regularization parameter ε. We developed a second algorithm with the aim of taking into account the ε-dependence and finding a solution of the original necessary optimality condition (2). We remark that our proposed algorithm, which we present in the following, uses the monotone scheme inside.
For $x \in \mathbb{R}^n$, A, f, μ defined as in Section 2, we denote by

$$\mathcal{A} = \{i \text{ s. t. } |(A_i, f_i)| < \mu_i\},$$

the active indexes and by $\mathcal{I} = \mathcal{A}^c$ the inactive ones. For $\varepsilon > 0, x \in \mathbb{R}^n$, set $S = \{i : |x_i| < \varepsilon\}$. We say that x is a singular point if $S \neq \emptyset$.
The algorithm has an outer main cycle where the necessary optimality condition (2) is solved by combining an active-set strategy and the monotone algorithm. In particular, the monotone scheme is used to solve in an inner cycle the ε-approximate necessary optimality condition (4). The stopping criteria depend on the specific problem considered, for more details we refer to Section 5.

Algorithm 1 Active set monotone scheme

1: Set *jmax, kmax*, $\varepsilon > 0, j = 0$. Initialize x;
2: **while** $j \leq jmax$ **do**
3: $x_i = 0$ for $i \in \mathcal{A}_j$; $k = 0, x^0 = x_{\mathcal{I}_j}$;
4: **while** $k \leq kmax$ **do**
5: solve for $x^{k+1}_{\mathcal{I}_j}$

$$A^*_{\mathcal{I}_j} A_{\mathcal{I}_j} x^{k+1}_{\mathcal{I}_j} + \frac{\beta p}{\max(\varepsilon^{2-p}, |x^k_{\mathcal{I}_j}|^{2-p})} x^{k+1}_{\mathcal{I}_j} = A^*_{\mathcal{I}_j} b;$$

6: if $x^{k+1}_{\mathcal{I}_j}$ is a singular point, break;
7: if stopping criteria is satisfied, break;
8: otherwise $k = k + 1$;
9: **end while**
10: $x_{\mathcal{I}_j} = x^{k+1}_{\mathcal{I}_j}$. If stopping criteria is satisfied, break;
11: otherwise $j = j + 1$.
12: **end while**

Remark 6. $\varepsilon > 0$ in **Algorithm** 1 is fixed and chosen as $\varepsilon = \min_i \left(\frac{2\beta(1-p)}{|A_i|_2^2}\right)^{\frac{1}{2-p}}$. This choice is made accordingly to the lower bound on the inactive components established in Corollary 3 (i).

Remark 7. Roughly speaking, singular points are those points x where $\max\{\varepsilon, |x_i|\} = \varepsilon$ for some i. The implementation of **Algorithm** 1 without step 6 shows that, even though the number of these points is always low, the existence of only a few of them could considerably increase the residue of the necessary optimality condition (2).

5. NUMERICAL RESULTS

In this section we report the numerical results obtained by testing **Algorithm** 1 in two different situations, the M-matrix example and a control problem (subsections 5.1 and 5.2, respectively).

For both examples we tested the algorithm for different values of p and by incrementally increasing β for each value of p. Consistent with our expectation, in both examples we find more sparsity for bigger values of β.

In subsection 5.3, we present an application of the algorithm to an image reconstrunction problem in microscopy imaging.

In the tables, *outer it, inner it*, $|\cdot|_0$, $|\cdot|_2$, *Res* are the number of iterations in the outer and inner cycle, the number of components bigger than 10^{-10}, the euclidean norm and the residue in the optimality condition (2), respectively.

5.1 M-matrix example

We consider

$$\min_{x \in \mathbb{R}^{n \times n}} \frac{1}{2}|Ax - b|_2^2 + \beta |x|_p, \tag{6}$$

A is the forward finite difference gradient

$$A = \begin{pmatrix} G_1 \\ G2 \end{pmatrix},$$

with $G_1 \in \mathbb{R}^{n(n+1)\times n^2}, G_2 \in \mathbb{R}^{n(n+1)\times n^2}$ as

$$G_1 = I \otimes D, \quad G_2 = D \otimes I,$$

I is the $n \times n$ identity matrix, $\otimes$ the tensor product, $D = (n+1)\tilde{D}$, $\tilde{D} \in \mathbb{R}^{(n+1)\times n}$ is

$$\begin{pmatrix} 1 & 0 & 0 \cdots & 0 \\ -1 & 1 & 0 \cdots & 0 \\ & & & \\ 0 & \cdots & 0 & -1 \quad 1 \\ 0 & \cdots & 0 & 0 \quad -1 \end{pmatrix}.$$

Then $A^T A$ is an M matrix coinciding with the 5-point star discretization on a uniform mesh on a square of the Laplacian with Dirichlet boundary conditions. Moreover (6) can be equivalently expressed as

$$\min_{x \in \mathbb{R}^{n \times n}} \frac{1}{2}|Ax|_2^2 - (x, f) + \beta |x|_p, \tag{7}$$

where $f = A^T b$. If $\beta = 0$ this is the discretized variational form of the elliptic equation

$$-\Delta y = f \text{ in } \Omega, \quad y = 0 \text{ on } \partial\Omega. \tag{8}$$

For $\beta > 0$ the variational problem (7) gives a sparsity enhancing solution for the elliptic equation (8), that is, the displacement y will be 0 when the forcing f is small. Our tests are conducted with f chosen as discretization of $f = 10x_1 \sin(5x_2)\cos(7x_1)$. The inizialization is chosen as the solution of the corresponding non-sparse optimal control problem.

We compute an approximate solution of the optimality condition (2) of the original problem, and stop the algorithm when two consecutive iterates coincide, i.e. a solution to the discrete optimality system is obtained.

Our tests were conducted for a $n = 64$ mesh for different values of p, and β incrementally increasing from 10^{-4} to 10, see Table 1 and 2 for $p = .5$, $p = .1$ respectively. The results show significant differences with respect to different values of β for any fixed p, in particular, consistent with our expectations, the sparsity of the solution increases with β (see the third row of Tables 1, 2 and Figure 1). For example, for $\beta = 1$ the solution to (7) is 0. Moreover, for different p we find different levels of sparsity and values for the ℓ^p-norm of the solution, thus showing sensitivity with respect to p.

The second and third row show the number of iterations of the outer and inner cycle, respectively. Note that the number of iterations for the outer cycle is small. For the inner cycle, we expect a bigger number of iterations, since it requires to solve a nonlinear system by an iteration procedure. However, we see that the total number of iterations decreases singificantly for smaller values of p and bigger values of β and in these cases the total number of iterations is small.

Finally, we remark that if we modify the inizialization, the method converges to the same solution with no remarkable modifications in the number of iterations.

Table 1. M-matrix example, $p = .5$, $n = 64$

β	0.0001	.01	.1	.5	1
outer it.	1	9	28	44	22
inner it.	51	107	218	136	1
$\|x\|_0$	3969	3904	3264	761	0
$\|x\|_p^p$	464.17	457.96	379.53	92.69	0
Res.	10^{-13}	10^{-13}	10^{-13}	10^{-13}	0

Table 2. M-matrix example, $p = .1$, $n = 64$

β	0.0001	.01	.1	.5	1
outer it.	2	24	31	14	7
inner it.	51	51	51	1	1
$\|x\|_0$	3948	3470	948	0	0
$\|x\|_p^p$	2500	2250	642.07	0	0
Res.	10^{-13}	10^{-13}	10^{-13}	0	0

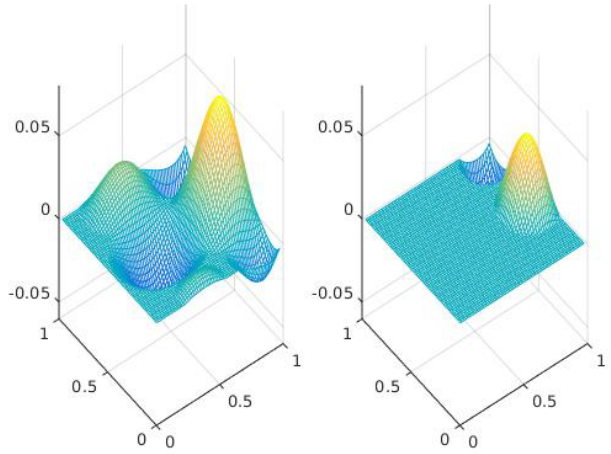

Fig. 1. M-matrix example, solution for $p = .5, \beta = .1$ (left), $p = .5, \beta = .5$ (right)

5.2 Sparsity in a control problem

We consider the linear control system

$$\frac{d}{dt}y(t) = \mathcal{A}y(t) + Bu(t), \quad y(0) = 0,$$

that is,

$$y(T) = \int_0^T e^{\mathcal{A}(T-s)}Bu(s)ds,$$

where the linear closed operator $\mathcal{A}$ generates a C_0-semigroup $e^{\mathcal{A}t}$, $t \geq 0$ on the state space X. More specifically, we consider the one dimensional controlled heat equation for $y = y(t, x)$:

$$y_t = y_{xx} + b_1(x)u_1(t) + b_2(x)u_2(t), \quad x \in (0, 1),$$

with homogeneous boundary conditions $y(t, 0) = y(t, 1) = 0$. The differential operator $\mathcal{A}y = y_{xx}$ is discretized in space by the second order finite difference approximation with $n = 49$ interior spatial nodes ($\Delta x = \frac{1}{50}$). We use two time dependent controls $\overrightarrow{u} = (u_1, u_2)$ with corresponding spatial control distributions b_i chosen as step functions:

$$b_1(x) = \chi_{(.2,.3)}, \quad b_2(x) = \chi_{(.6,.7)}.$$

The control problem consists in finding the control function $\overrightarrow{u}$ that steers the state $y(0) = 0$ to a neighborhood of the desired state y_d at the terminal time $T = 1$. We discretize the problem in time by the mid-point rule, i.e.,

$$Au = \sum_{k=1}^{m} e^{\mathcal{A}\left(T - t_k - \frac{\Delta t}{2}\right)}B\overrightarrow{u_k}\Delta t, \tag{9}$$

where $\overrightarrow{u} \in \mathbb{R}^{2m}$ is a discretized control vector with coordinates $\overrightarrow{u_k} \in \mathbb{R}^2$ representing the values at the mid-point of the intervals (t_k, t_{k+1}). A uniform step-size $\Delta t = \frac{1}{50}$ ($m = 50$) is utilized. The solution of the control problem is based on the sparsity formulation (1), where b is the discretized target function chosen as the Gaussian distribution $y_d(x) = 10\exp(-314(x - .7)^2))$ centered at $x = .7$. That is, we apply our algorithm for the discretized optimal control problem in time and space where x from (1) is the discretized control vector $\overrightarrow{u} \in \mathbb{R}^{2m}$ which is mapped by A to the discretized output y at time 1 by means of (9). Moreover b from (1) is the discretized state y_d with respect to the spatial grid Δx.

Since the second control distribution is well within the support of the desired state y_d, we expect the authority of this control to be stronger than that of the first one, which is away from the target.

We compute an approximate solution of the optimality condition (2) of the original problem, where by approximate we mean that the algorithm stops when the ℓ^∞ norm of the residue in the optimality condition (2) is less than 10^{-9}.

We refer to Table 3, 4, where we show the results for $p = .1$, $p = .01$, respectively. Again the sparsity of the solution increases with β, and as we expected, it increases much faster on the first control u_1 than the second. Note that the number of iterations is small for each value of β and considerably smaller than the number of iterations needed in the M-matrix example.

Table 3. Sparsity in a control problem, $p = .1$

β	10^{-7}	10^{-5}	10^{-3}	10^{-1}	10
outer it.	7	7	8	3	2
inner it.	8	7	9	10	6
$\|x\|_0$	17	17	7	4	1
$\|x\|_p^p$	153	133	30.1	8.5	1.87
Res.	10^{-9}	10^{-10}	10^{-12}	10^{-9}	10^{-10}

Table 4. Sparsity in a control problem, $p = .01$

β	10^{-7}	10^{-5}	10^{-3}	10^{-1}	10
outer it.	4	6	6	5	2
inner it.	6	7	9	8	5
$\|x\|_0$	35	23	9	7	2
$\|x\|_p^p$	42	28.2	10.6	8	2.1
Res.	10^{-9}	10^{-10}	10^{-11}	10^{-11}	10^{-10}

5.3 Compressive sensing approach for microscopy image reconstruction

Recently in microscopy imaging, single molecule detection-based techniques have been more and more investigated in order to go beyond the diffraction limit. We focus on the STORM (stochastic optical reconstruction microscopy) method, see Zhu et al. (2012), Babcock et al. (2013). In STORM, the image reconstruction process consists in a series of imaging cycles. In each of them only a fraction of the fluorophores in the field of view are switched on, such that each of the active fluorophores is optically resolvable from the rest. Despite the advantage of obtaining sub-diffraction-limit spatial resolution, the time to acquire a super-resolution image is limited by the maximum density of fluorescent emitters that can be accurately localized per imaging frame. In order to get at the same time better resolution and higher emitter density per imaging frame, compressive sensing methods based on l^1 techniques have been recently applied, see for example in Zhu et al.

(2012), Babcock et al. (2013). In the following we test our algorithm for $l^p, p < 1$ functionals.
To be more specific, each single frame reconstruction can be achieved by solving the following constrained-minimization problem:

$$\min_{x \in \mathbb{R}^n} |x|_p^p \quad \text{such that} \quad |Ax - b|_2 \leq \varepsilon, \tag{10}$$

where $p \in (0, 1)$, x is the up-sampled, reconstructed image, b is the experimentally observed image, and A is the impulse reponse (of size $m \times n$, where m and n are the numbers of pixels in b and x, respectively). A is usually called the point spread function (PSF) and describes the response of an imaging system to a point source or point object.

Remark 8. Solving problems as (10) is denoted as compressive sensing in the literature of miscroscopy imaging. Indeed, in the compressive sensing approach to microscopy image reconstruction, the sparse basis is a high resolution grid, in which fluorophore locations are presented-effectively an up-sampled, reconstructed image-while the noisy measurement basis is the lower resolution camera pixels, on which fluorescence signal are detected experimentally.

We reformulate problem (10) as:

$$\min_{x \in \mathbb{R}^n} \frac{1}{2}|Ax - b|_2^2 + \beta |x|_p^p \tag{11}$$

and we solve problem (11) by applying our algorithm.

Remark 9. We may consider (11) arising from (10) with β the reciprocal of the Lagrange multiplier associated to the inequality constraint $|Ax - b|_2 \leq \varepsilon$.

The algorithm was tested for different resolution images, that is, with a 16×16 pixel conventional image and a 128×128 true image. The values for the impulse reponse A and the measured data b were chosen accordingly to the literature, in particular A was taken as the Gaussian PSF matrix with variance $\sigma = 8$ and size $3 \times \sigma = 24$, and b was simulated by convolving the impulse reponse A with a random 0-1 mask over the image adding a white random noise so that the signal to noise ratio is .01.
In order to measure the performance of our algorithm, we plot a graphic of the average over six recoveries of the support recovery against the noise (see for example Duval and Peyré (2015), Candes et al. (2006)). More specifically, we show a plot of the location recovery, the exact recovery (up to a certain tolerance) and the number of surplus and missed emitters recovered (Error+, Error- respectively).
We show the graphics for a sparse 0-1 cross image of size 64×64 in Figure 2 (location and exact recovery) and Figure 3 (Error+, Error-). Note that the values of the Error+, Error- have to be compared to the total number of emitters, which is 197. The algorithm also performed well on non-sparse standard phantom images.
We remark that the locations and intensity of the emitters decay linearly with respect to the noise level, in line with the result of Duval and Peyré (2015). Note that for small noise both the location recovery is near to $n^2 = 4096$, e.g. the (averaged) location recovery is 4087.3.
We compare our results with the one obtained by the FISTA algorithm for the same problem (see Figure 4). We remark that the levels of the location and exact recoveries are lower than the ones recovered by the active-set monotone scheme, at least for values of the noise near .01. In particular, the Error+ (surplus emettitors) is always above 190, whereas with the active-set monotone scheme is zero for small value of the noise. On the other hand, FISTA is faster than our algorithm (as expected, since the active-set monotone scheme solves a nonlinear equation for each minimization problem.)

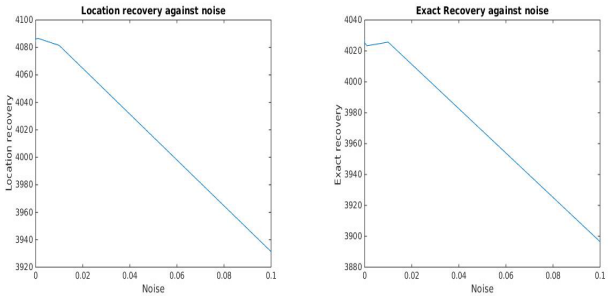

Fig. 2. Location rec. (left), Exact rec. (right), active-set monot., $p = .1, \beta = 10^{-6}$

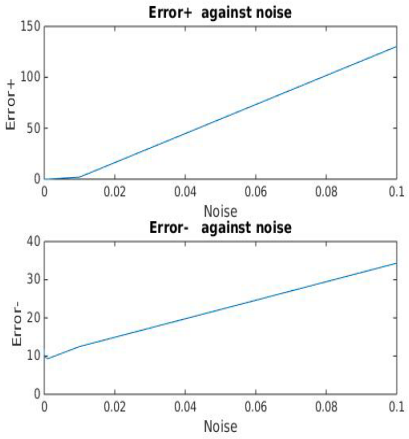

Fig. 3. Error+ (surplus of emettitors), Error- (missed emettitors), active-set monot., $p = .1, \beta = 10^{-6}$

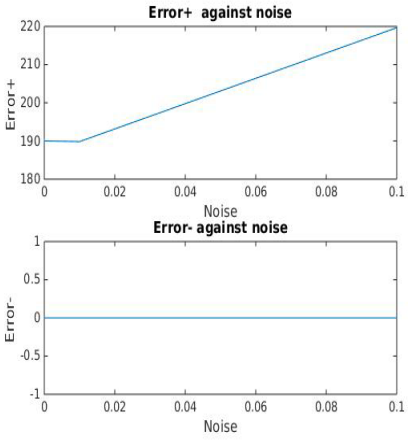

Fig. 4. Error+ (surplus of emettitors), Error- (missed emettitors), FISTA $\beta = 10^{-6}$

REFERENCES

Babcock, H.P., Moffitt, J.R., Cao, Y., and Zhuang, X. (2013). Fast compressed sensing analysis for superresolution imaging using l1-homotopy. *Opt Express*, 21, 28583–28596.

Bredies, K. and Lorentz, D. (2015). Minimization of nonsmooth, nonconvex functionals by iterative thresholding. *Journal of Optimization Theory and Applications*, 165, 78–112.

Candes, E., Romberg, J., and Tao, T. (2006). Stable signal recovery from incomplete and inaccurate measurements. *Communication in Pure and Applied Mathematics*, 59, 1207–1223.

Candes, E.J. and Tao, T. (2005). Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51.

Chartrand, R. (2009). Fast algorithms for nonconvex compressive sensing: mri reconstruction from very few data. *IEEE Interantional Symposium on Biomedical Imaging (ISBI)*.

Duval, V. and Peyré, G. (2015). Exact support recovery for sparse spikes dconvolution. *Foundations of Computational Mathemadics*, 15, 1315–1355.

Fornasier, M. and Ward, R. (2010). Iterative thresholing meets free-discontinuity problems. *Foundations of Computational Mathematics*, 10, 527–567.

Hintermüller, M. and Wu, T. (2013). Nonconvex tv^q-models in image restoration: analysis and a trust-region regularization-based superlinearly convergent solver. *SIAM J. Imaging Sciences*, 6, 1385–1415.

Huang, J. and Mumford, D. (1999). Proceedings of the international conference on computer vision and pattern recognition (cvpr). *Fort Collins, C0*, 541–547.

Ito, K. and Kunisch, K. (2014). A variational approach to sparsity optimization based on lagrange multiplier theory. *IOP Science Inverse Problems*, 30.

Jiao, Y., Jin, B., Lu, X., and Ren, W. (Preprint 2013). A primal dual active set algorithm for a class of nonconvex sparsity optimization. *unpublished*.

Li, G. and Pong, T.K. (2014). Global convergence of splitting methods for nonconvex composite optimization. *SIAM J. on Optimization*, 25, 2434–2460.

Nikolova, M., Ng, M.K., and C.-P.Tam (2010). Fast nonconvex nonsmooth minimization methods for image restoration and reconstruction. *IEEE Trans. Image Process.*, 19, 3073–3088.

Nikolova, M., Ng, M.K., Zhang, S., and Ching, W.K. (2008). Efficient reconstruction of piecewise constant images using nonsmooth nonconvex minimization. *SIAM J. Imaging Sciences*, 1, 2–25.

Ochs, P., Dosovitskiy, A., Brox, T., and Pock, T. (2015). On iteratively reweighted algorithms for nonsmooth nonconvex optimization in computer vision. *SIAM J. Imaging Sciences*, 8, 331–372.

Pietro, G.D. (2013). A variational approach to fracture and other inelastic phenomena. *Journal of Elasticity*, 112, 3–77.

Ramlau, R. and Zarzer, C. (2012). On the optimization of a tikhonov functional with non-convex sparsity constraints. *ETNA*, 39, 476–507.

Roth, S. and Black, M. (2009). Fields of experts. *Internat. J. Comput. Vision*, 82, 205–229.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B*, 58, 267–288.

Zhu, L., Zhang, W., Elnatan, D., and Huang, B. (2012). Faster storm using compressed sensing. *Nature Methods*, 9, 721–723.