Conditions on optimal support recovery in unmixing problems by means of multi-penalty regularization

Markus Grasmair

Department of Mathematical Sciences, Norwegian University of Science and Technology, 7491 Trondheim

E-mail: markug@math.ntnu.no

Valeriya Naumova

Simula Research Laboratory AS, Martin Linges vei 25, 1364 Fornebu, Oslo

E-mail: valeriya@simula.no

Abstract. Inspired by several real-life applications in audio processing and medical image analysis, where the quantity of interest is generated by several sources to be accurately modeled and separated, as well as by recent advances in regularization theory and optimization, we study the conditions on optimal support recovery in inverse problems of unmixing type by means of multi-penalty regularization.

We consider and analyze a regularization functional composed of a data-fidelity term, where signal and noise are additively mixed, a non-smooth, convex, sparsity promoting term, and a quadratic penalty term to model the noise. We prove not only that the well-established theory for sparse recovery in the single parameter case can be translated to the multi-penalty settings, but we also demonstrate the enhanced properties of multi-penalty regularization in terms of support identification compared to sole ℓ^1-minimization. We additionally confirm and support the theoretical results by extensive numerical simulations, which give a statistics of robustness of the multi-penalty regularization scheme with respect to the single-parameter counterpart. Eventually, we confirm a significant improvement in performance compared to standard ℓ^1-regularization for compressive sensing problems considered in our experiments.

1. Introduction

In many real-life applications such as audio processing or medical image analysis, one encounters the situation when given observations (most likely noisy) have been generated by several sources u_i that one wishes to reconstruct separately. In this case, the reconstruction problem can be understood as an inverse problem of unmixing type, where the solution $u^\dagger$ consists of several (two or more) components of different nature, which have to be identified and separated.

In mathematical terms, an unmixing problem can be stated as the solution of an equation

$$Au^\dagger = y,$$

where $u^\dagger = \sum_{i=1}^{L} u_i, u_i \in V_i$ and $V_1 + \ldots + V_L = \mathbb{R}^N$ but $\langle V_i, V_j \rangle \simeq 0$ for $i \neq j$ in the sense that

$$|\langle v_i, v_j \rangle| \simeq \delta_{ij} \tag{1}$$

for all $v_i \in V_i$ and $v_j \in V_j$, $\|v_i\|_{\ell_2} = \|v_j\|_{\ell_2} = 1, i \neq j$. In general, we are interested to acquire the minimal amount of information on u so that we can selectively reconstruct with the best accuracy one of the components $u_{\hat{\imath}}$, but not necessarily also the other components u_j for $j \neq \hat{\imath}$. In this settings, we further assume that A cannot be specifically tuned to recover $u_{\hat{\imath}}$ but should be suited to gain universal information to recover $u_{\hat{\imath}}$ by a specifically tuned decoder.

A concrete example of this setting is the *noise folding phenomenon* arising in compressed sensing, related to noise in the signal that is eventually amplified by the measurement procedure. In this setting, it is reasonable to consider a model problem of the type

$$A(u + v) = y, \tag{2}$$

where v is the random Gaussian noise with variance σ_v on the original signal $u \in \mathbb{R}^N$ and $A \in \mathbb{R}^{m \times N}$ is the linear measurement matrix. Several recent works (see, for instance, [1] and the references therein) illustrate how the measurement process actually causes the noise folding phenomenon. To be more specific, one can show that (2) is equivalent to solving

$$\hat{A}u + \omega = y, \tag{3}$$

where ω is composed by i.i.d. Gaussian entries with distribution $\mathbf{N}(0, \sigma_\omega)$, and the variance σ_ω is related to the variance of the original signal by $\sigma_\omega^2 = \frac{N}{m}\sigma_v^2$. This implies that the variance of the noise on the original signal is amplified by a factor of N/m.

Under the assumption that A satisfies the so-called restricted isometry property, it is known from the work on the Dantzig selector in [4] that one can reconstruct $u^\dagger$ from measurements y as in (3) such that

$$\|u - u^\dagger\|_2^2 \leq C^2 2\Big((1+k)\frac{N}{m}\sigma_v^2\Big)\log N, \tag{4}$$

where k denotes the number of nonzero elements of the solution u. The estimate (4) is considered (folklore) nearly-optimal in the sense that no other method can really improve the asymptotic error $\mathcal{O}(\frac{N}{m}\sigma_v^2)$. Therefore, the noise folding phenomenon may in practice significantly reduce the potential advantages of compressed sensing in terms of the trade-off between robustness and efficient compression (given by the factor N/m here) compared to other more traditional subsampling methods [5].

In [2] the authors present a two-step numerical method which allows not only to recover the large entries of the original signal u accurately, but also has

enhanced properties in terms of support identification over simple ℓ^1-minimization based algorithms. In particular, because of the lack of separation between noise and reconstructed signal components, the latter ones can easily fail to recover the support when the support is not given a priori. However, the computational cost of the second phase of the procedure presented in [2], being a non-smooth and non-convex optimization problem, is too demanding to be performed on problems with realistic dimensionalities. It was also shown that other methods based on a different penalization of the signal and noise can lead to higher support detection rate.

The follow up work [11], which addresses the noise folding scenario by means of multi-penalty regularization, provides the first numerical evidence of the superior performance of multi-penalty regularization compared to its single parameter counterparts for problem (2). In particular, the authors consider the functional

$$J_{p,q}(u,v) := \|A(u+v)-y\|_2^2 + \alpha\|u\|_p^p + \left(\beta\|v\|_q^q + \varepsilon\|v\|_2^2\right), \tag{5}$$

here $\alpha, \beta, \varepsilon \in \mathbb{R}_+$ may all be considered as regularization parameters of the problem. The parameter $\varepsilon > 0$ ensures the ℓ_2-coercivity of $J_{p,q}(u,\cdot)$ also with respect to the component v. In the infinite dimensional setting the authors presented a numerical approach to the minimization of (5) for $0 \le p < 2, 2 \le q < \infty$, based on simple iterative thresholding steps, and analyzed its convergence.

The results presented in this paper are very much inspired not only by the above-mentioned works in the signal processing and compressed sensing fields, but also by theoretical developments in sparsity-based regularization (see [8] and references therein) and multi-penalty regularization ([9, 10], just to mention a few). While the latter two directions are considered separately in most of the literature, there have also been some efforts to understand regularization and convergence behavior for multiple parameters and functionals, especially for image analysis [3, 12]. However, to the best of our knowledge, the present paper is the first one providing a theoretical analysis of the multi-penalty regularization with a non-smooth sparsity promoting regularization term, and an explicit comparison with the single-parameter counterpart.

1.1. Content of the paper

In Section 2 we concisely recall the pertinent features and concepts of multi-penalty and single-penalty regularization. We further show that ℓ^1-regularization can be considered as the limiting case of the multi-penalty one, and thus the theory of ℓ^1-regularization can be applied to multi-penalty setting. In Section 3 we recall and discuss conditions for exact support recovery in the single-parameter case. The main contributions of the paper are presented in Sections 4 and 5, where we extend and generalize the results from the previous sections to the multi-penalty setting. In Section 5 we also open the discussion on the set of admissible parameters for the exact support recovery for unmixing problem in single-parameter as well as multi-penalty cases. In particular, we study the sensitivity of the multi-penalty scheme with respect to the parameter choice. The theoretical findings and discussion are illustrated and supported by extensive numerical validation

tests presented in Section 6. Finally, in Section 7 we compare the performance of the multi-penalty regularization and its single-parameter counterpart for compressive sensing problems.

2. Multi-Penalty and Single-Penalty Regularization

We first provide a short reminder and collect some definitions of the standard notation used in this paper.

The true solution $u^\dagger$ of the unmixing problem (2) is called k-sparse if it has at most k non-zero entries, i.e., $\#I = \#\operatorname{supp}(u^\dagger) \leq k$, where $I := \operatorname{supp}(u^\dagger) := \{i : u_i^\dagger \neq 0\}$ denotes the support of $u^\dagger$.

We propose to solve the unmixing problem (2) with k-sparse true solution using multi-penalty Tikhonov regularization of the form

$$T_{\alpha,\beta}(u,v) := \frac{1}{2}\|A(u+v)-y\|_2^2 + \alpha\|u\|_1 + \frac{\beta}{2}\|v\|_2^2 \to \min_{u,v}, \tag{6}$$

the solution of which we will denote by $(u_{\alpha,\beta}, v_{\alpha,\beta})$. We note that we can, formally, interpret standard ℓ^1-regularization as the limiting case $\beta = \infty$, setting

$$T_{\alpha,\infty}(u,v) := \begin{cases} \frac{1}{2}\|Au-y\|_2^2 + \alpha\|u\|_1 & \text{if } v = 0, \\ +\infty & \text{if } v \neq 0. \end{cases}$$

Obviously, the pair of minimizers of $T_{\alpha,\infty}$ will always be equal to $(u_\alpha, 0)$, where u_α minimizes $\frac{1}{2}\|Au-y\|_2^2 + \alpha\|u\|_1$.

Definition 1 *Let $\beta \in \mathbb{R}_+ \cup \{\infty\}$ be fixed. We say that a set $\mathcal{S} \subset \mathbb{R}^N \times \mathbb{R}^N$ is a* set of exact support recovery for the unmixing problem with operator A, *if there exists $\alpha > 0$, such that $\operatorname{supp}(u_{\alpha,\beta}) = \operatorname{supp}(u^\dagger)$ whenever the given data y has the form $y = A(u^\dagger + v)$ with $(u^\dagger, v) \in \mathcal{S}$.*

The parameters $\alpha > 0$ for which this property holds are called admissible for $\mathcal{S}$.

Specifically, we will study for $c > d > 0$ the sets

$$\mathcal{S}_{c,d,I} := \{(u,v) \in \mathbb{R}^N \times \mathbb{R}^N : \operatorname{supp}(u) = I,\ \inf_{i\in I}|u_i| > c,\ \|v\|_\infty < d\} \tag{7}$$

and the corresponding class

$$\mathcal{S}_{c,d,k} := \bigcup_{\#I \leq k} \mathcal{S}_{c,d,I}.$$

The set $\mathcal{S}_{c,d,I}$ is a set of exact support recovery, if there exists some regularization parameter $\alpha > 0$, such that we can apply multi-penalty regularization with parameters α and β (or single-parameter ℓ^1-regularization with parameter α in case $\beta = \infty$) to the unmixing problem (2) and obtain a result with the correct support, provided that the ℓ^∞-norm of the noise is smaller than d and the non-zero coefficients of $u^\dagger$ are larger than c. Typical examples of real-life signals that can be modeled by signals from the set $\mathcal{S}_{c,d,I}$ can be found in Asteroseismology, see for instance [2].

We note that this class of signals is very similar to the one studied in [2]. The main difference is that we focus on the case where the noise v is bounded only componentwise (that is, with respect to the ℓ^∞-norm), whereas [2] deals with noise that has a bounded ℓ^p-norm for some $1 \leq p \leq 2$. Additionally, we allow the noise also to mix with the signal $u^\dagger$ to be identified in the sense that the supports of v and $u^\dagger$ may have a non-empty intersection. In contrast, the signal and the noise are assumed to be strictly separated in [2].

Throughout the paper we will several times refer to the sign function sgn, which we always interpret as being the set valued function $\operatorname{sgn}(t) = 1$ if $t > 0$, $\operatorname{sgn}(t) = -1$ if $t < 0$, and $\operatorname{sgn}(t) = [-1, 1]$ if $t = 0$, applied componentwise to the entries of the vector u_α.

We use the notation A_I to denote the restriction of the operator A to the span of the support of $u^\dagger$. Additionally, we denote by

$$J := \{i : u_i^\dagger = 0\}$$

the complement of I, and by A_J the restriction of A to the span of J. We note that the adjoints A_I^* and A_J^* are simply the compositions of the adjoint A^* of A with the projections onto the spans of I and J, respectively.

As a first result, we show that the solution of the multi-penalty problem (6) simultaneously solves a related single-penalty problem.

Lemma 1 *The pair $(u_{\alpha,\beta}, v_{\alpha,\beta})$ solves (6) if and only if*

$$v_{\alpha,\beta} = (\beta + A^*A)^{-1}(A^*y - A^*Au_{\alpha,\beta})$$

and $u_{\alpha,\beta}$ solves the optimization problem

$$\frac{1}{2}\|B_\beta u - y_\beta\|_2^2 + \alpha\|u\|_1 \to \min \tag{8}$$

with

$$B_\beta = \left(\mathrm{Id} + \frac{AA^*}{\beta}\right)^{-1/2} A$$

and

$$y_\beta = \left(\mathrm{Id} + \frac{AA^*}{\beta}\right)^{-1/2} y.$$

Proof: We can solve the optimization problem in (6) in two steps, first with respect to v and then with respect to u. Assuming that u is fixed, the optimality condition for v in (6) reads

$$A^*(A(u + v) - y) + \beta v = 0. \tag{9}$$

That is, for fixed u, the optimum in (6) with respect to v is obtained at

$$v(u) := (\beta + A^*A)^{-1}(A^*y - A^*Au).$$

Inserting this into the Tikhonov functional, we obtain the optimization problem

$$\frac{1}{2}\|A(u+v(u))-y\|_2^2+\alpha\|u\|_1+\frac{\beta}{2}\|v(u)\|_2^2\to\min_u.$$

Using (9), we can write

$$\begin{aligned}\frac{1}{2}\|A(u+v(u))-y\|_2^2&=\frac{1}{2}\langle A(u+v(u))-y,Au-y\rangle\\&\qquad+\frac{1}{2}\langle A^*(A(u+v(u))-y),v(u)\rangle\\&=\frac{1}{2}\langle A(u+v(u))-y,Au-y\rangle-\frac{\beta}{2}\|v(u)\|_2^2.\end{aligned}$$

Thus the optimization problem for u simplifies to

$$\frac{1}{2}\langle A(u+v(u))-y,Au-y\rangle+\alpha\|u\|_1\to\min_u. \tag{10}$$

Now note that

$$\begin{aligned}A(u+v(u))-y&=A(\mathrm{Id}-(\beta+A^*A)^{-1}A^*A)u-(\mathrm{Id}-A(\beta+A^*A)^{-1}A^*)y\\&=A(\mathrm{Id}+A^*A/\beta)^{-1}u-(\mathrm{Id}+AA^*/\beta)^{-1}y\\&=(\mathrm{Id}+AA^*/\beta)^{-1}(Au-y).\end{aligned}$$

Inserting this equality in (10), we obtain the optimization problem

$$\frac{1}{2}\langle(\mathrm{Id}+AA^*/\beta)^{-1}(Au-y),Au-y\rangle+\alpha\|u\|_1\to\min_u,$$

which is the same as (8). □

Remark 1 *As a consequence of Lemma 1, we can apply the theory of ℓ^1-regularization also to the multi-penalty setting we consider here. In particular, this yields, for fixed $\beta>0$, estimates of the form*

$$\|u^\dagger-u_{\alpha,\beta}\|_1\le C_{1,\beta}\alpha+C_{2,\beta}\frac{\|y_\beta-B_\beta u^\dagger\|_2^2}{\alpha}$$

provided that $u^\dagger$ satisfies a source condition of the form $B_\beta^*\eta\in\partial(\|u^\dagger\|_1)$ with $|(B_\beta^*\eta)_i|<1$ for every $i\notin\operatorname{supp}(u^\dagger)$, and the restriction of the mapping B_β to the span of the support of $u^\dagger$ is injective (see [8]). Additionally, it is easy to show that these conditions hold for B_β provided that they hold for A and β is sufficiently large.

3. Sets of Exact Support Recovery—Single-penalty Setting

The main focus of this paper is the question whether multi-penalty regularization allows for the exact recovery of the support of the true solution $u^\dagger$ and how it compares to single-penalty regularization. Because, as we have seen in Lemma 1, multi-penalty regularization can be rewritten as single-parameter regularization for the regularized operator B_β and right hand side y_β, we will first discuss recovery conditions in the single-parameter setting.

In order to find conditions for exact support recovery, we first recall the necessary and sufficient optimality condition for ℓ^1-regularization:

Lemma 2 *The vector u_α minimizes*

$$T_\alpha(u) := \frac{1}{2}\|Au - y\|_2^2 + \alpha\|u\|_1,$$

if and only if

$$A^*(Au_\alpha - y) \in -\alpha\,\mathrm{sgn}(u_\alpha).$$

Using this result, we obtain a condition that guarantees exact support recovery for the single-penalty method:

Lemma 3 *We have* $\mathrm{supp}(u_\alpha) = I$*, if and only if there exists* $w_\alpha \in (\mathbb{R}\setminus\{0\})^I$ *such that*

$$A_I^*(A_I w_\alpha - y) = -\alpha\,\mathrm{sgn}(w_\alpha),$$
$$\|A_J^*(A_I w_\alpha - y)\|_\infty \le \alpha,$$

Proof: This immediately follows from Lemma 2 by testing the optimality conditions on the vector u_α given by $(u_\alpha)_i = (w_\alpha)_i$ for $i \in I$ and $(u_\alpha)_i = 0$ else. □

Our main result concerning support recovery for single-parameter regularization is the following:

Proposition 1 *Assume that A_I is injective and that*

$$\|A_J^* A_I (A_I^* A_I)^{-1}\|_\infty < 1. \tag{11}$$

Then the set $\mathcal{S}_{c,d,I}$ defined in (7) is a set of exact support recovery for the unmixing problem whenever

$$\frac{c}{d} > \frac{\|A_J^*(A_I(A_I^*A_I)^{-1}A_I^* - \mathrm{Id})A\|_\infty \|(A_I^*A_I)^{-1}\|_\infty}{1 - \|A_J^*A_I(A_I^*A_I)^{-1}\|_\infty} + \|(A_I^*A_I)^{-1}A_I^*A\|_\infty. \tag{12}$$

Moreover, every parameter $\alpha > 0$ satisfying

$$\frac{d\|A_J^*(A_I(A_I^*A_I)^{-1}A_I^* - \mathrm{Id})A\|_\infty}{1 - \|A_J^*A_I(A_I^*A_I)^{-1}\|_\infty} \le \alpha < \frac{c - d\|(A_I^*A_I)^{-1}A_I^*A\|_\infty}{\|(A_I^*A_I)^{-1}\|_\infty} \tag{13}$$

is admissible on $S_{c,d,I}$.

Proof: First we note that the injectivity of A_I implies that the mapping $A_I^*A_I$ is invertible. Thus the condition (11) actually makes sense. Moreover, the inequality (12) is necessary and sufficient for the existence of α satisfying (13).

Now let $(u^\dagger, v) \in \mathcal{S}_{c,d,I}$ and assume that α satisfies (13). We denote

$$s_i^\dagger := \mathrm{sgn}(u_i^\dagger), \qquad i \in I,$$

and define

$$w_\alpha := u_I^\dagger + (A_I^*A_I)^{-1}A_I^*Av - \alpha(A_I^*A_I)^{-1}s^\dagger.$$

Because $\|s^\dagger\|_\infty = 1$, it follows from the second inequality in (13) that

$$|(w_\alpha)_i - u_i^\dagger| \le \|(A_I^*A_I)^{-1}A_I^*A\|_\infty\|v\|_\infty + \alpha\|(A_I^*A_I)^{-1}\|_\infty < c \le |u_i^\dagger|,$$

and therefore

$$\mathrm{sgn}(w_\alpha) = \mathrm{sgn}(u_I^\dagger) = s^\dagger.$$

Thus w_α actually satisfies the equation

$$\begin{aligned} w_\alpha &= u_I^\dagger + (A_I^* A_I)^{-1} A_I^* A v - \alpha (A_I^* A_I)^{-1} \operatorname{sgn}(w_\alpha) \\ &= (A_I^* A_I)^{-1} A_I^* (A_I u_I^\dagger + Av) - \alpha (A_I^* A_I)^{-1} \operatorname{sgn}(w_\alpha) \\ &= (A_I^* A_I)^{-1} A_I^* y - \alpha (A_I^* A_I)^{-1} \operatorname{sgn}(w_\alpha), \end{aligned}$$

and thus

$$A_I^*(A_I w_\alpha - y) = -\alpha \operatorname{sgn}(w_\alpha),$$

which is the first condition in Lemma 3.

It remains to show that

$$\|A_J^*(A_I w_\alpha - y)\|_\infty \leq \alpha.$$

However,

$$\begin{aligned} A_J^*(A_I w_\alpha - y) &= A_J^*(A_I w_\alpha - A_I u_I^\dagger - Av) \\ &= A_J^*(A_I (A_I^* A_I)^{-1} A_I^* A v - Av - \alpha A_I (A_I^* A_I)^{-1} s^\dagger), \end{aligned}$$

and thus

$$\begin{aligned} \|A_J^*(A_I w_\alpha - y)\|_\infty \leq d \|A_J^*(A_I (A_I^* A_I)^{-1} A_I^* - \mathrm{Id}) A\|_\infty \\ + \alpha \|A_J^* A_I (A_I^* A_I)^{-1}\|_\infty. \end{aligned}$$

Now the first inequality in (13) implies that this term is smaller than α. Thus w_α satisfies the conditions of Lemma 3, and thus $\operatorname{supp}(u_\alpha) = I$. □

Remark 2 *In the case where $A = \mathrm{Id}$ is the identity operator, the conditions above reduce to the conditions that $c > 2d$ and $d \leq \alpha < c - d$. Since ℓ^1-regularization in this setting reduces to soft thresholding, these conditions are very natural and are actually both sufficient and necessary: Since the noise may componentwise reach the value of d, it is necessary to choose a regularization parameter of at least d in order to remove it. However, on the support I of the signal, the smallest values of the noisy signal value are at least of size $c - d$. Thus they are retained as long as the regularization parameter does not exceed this value.*

For more complicated operators A, the situation is similar, i.e., a too small regularization parameter α is not able to remove all the noise, while a too large one destroys part of the signal as well. The exact bounds for the admissible regularization parameters, however, are much more complicated.

4. Sets of Exact Support Recovery—Multi-Penalty Setting

We now consider the setting of multi-penalty regularization for the solution of the unmixing problem. Applying Lemma 1, we can treat multi-penalty regularization with the same methods as single-penalty regularization. To that end, we introduce the regularized operator

$$A_\beta := \left(I + \frac{AA^*}{\beta}\right)^{-1} A. \tag{14}$$

In particular, we have with the notation of Lemma 3 that $B_\beta^* B = A_\beta^* A$ and $B_\beta^* y_\beta = A_\beta^* y$.

As a first result, we obtain the following analogon to Lemma 3:

Lemma 4 *We have* $\operatorname{supp}(u_{\alpha,\beta}) = I$*, if and only if there exists* $w_\alpha \in (\mathbb{R} \setminus \{0\})^I$ *such that*

$$\begin{aligned} &A_{\beta,I}^*(A_I w_{\alpha,\beta} - y) = -\alpha \operatorname{sgn}(w_{\alpha,\beta}), \\ &\|A_{\beta,J}^*(A_I w_\alpha - y)\|_\infty \leq \alpha, \end{aligned}$$

Proof: Applying Lemma 2 to the single-penalty problem (8), we obtain the conditions

$$\begin{aligned} &B_{\beta,I}^*(B_{\beta,I} w_{\alpha,\beta} - y_\beta) = -\alpha \operatorname{sgn}(w_{\alpha,\beta}), \\ &\|B_{\beta,J}^*(B_{\beta,I} w_{\alpha,\beta} - y_\beta)\|_\infty \leq \alpha. \end{aligned}$$

Now the claim follows from the equalities

$$\begin{aligned} B_{\beta,I}^* B_{\beta,I} &= A_{\beta,I}^* A_I, & B_{\beta,I}^* y_\beta &= A_{\beta,I}^* y, \\ B_{\beta,I}^* B_{\beta,I} &= A_{\beta,I}^* A_I, & B_{\beta,I}^* y_\beta &= A_{\beta,I}^* y. \end{aligned}$$

□

Since the proof of Proposition 1 only depends on the optimality conditions and the representation of the data as $y = Au^\dagger + Av$, we immediately obtain a generalization of Proposition 1 to the multi-penalty setting.

Proposition 2 *Assume that* $0 < \beta < \infty$ *is such that*

$$\|A_{\beta,J}^* A_I (A_{\beta,I}^* A_I)^{-1}\|_\infty < 1. \tag{15}$$

Then the set $\mathcal{S}_{c,d,I}$ *is a set of exact support recovery for the unmixing problem in the multi-penalty setting whenever*

$$\begin{aligned} \frac{c}{d} > \|(A_{\beta,I}^* A_I)^{-1} A_{\beta,I}^* A\|_\infty \\ + \frac{\|A_{\beta,J}^*(A_I(A_{\beta,I}^* A_I)^{-1} A_{\beta,I}^* - \mathrm{Id})A\|_\infty \|(A_{\beta,I}^* A_I)^{-1}\|_\infty}{1 - \|A_{\beta,J}^* A_I (A_{\beta,I}^* A_I)^{-1}\|_\infty}. \end{aligned} \tag{16}$$

Moreover, all the pairs of parameter (α, β) *satisfying (16) and*

$$\frac{d\|A_{\beta,J}^*(A_I(A_{\beta,I}^* A_I)^{-1} A_{\beta,I}^* - \mathrm{Id})A\|_\infty}{1 - \|A_{\beta,J}^* A_I (A_{\beta,I}^* A_I)^{-1}\|_\infty} \leq \alpha < \frac{c - d\|(A_{\beta,I}^* A_I)^{-1} A_{\beta,I}^* A\|_\infty}{\|(A_{\beta,I}^* A_I)^{-1}\|_\infty} \tag{17}$$

are admissible on $\mathcal{S}_{c,d,I}$*.*

Proof: The proof is analogous to the proof of Proposition 1. □

Remark 3 *We note that the condition*

$$\|A_J^* A_I (A_I^* A_I)^{-1}\|_\infty < 1$$

implies the analogous inequality for A_β *provided that* β *is sufficiently large. Similarly, if* α *satisfies the conditions in Proposition 1 that guarantee admissibility on* $S_{c,d,I}$*, the pair* (α, β) *will satisfy the conditions for admissibility in Proposition 2 provided that* β *is sufficiently large. The converse, however, need not be true: If the pair* (α, β) *is*

admissible for exact support recovery on $\mathcal{S}_{c,d,k}$ with multi-penalty regularization, it need not be true that the single parameter α is admissble for the single-penalty setting as well. Examples where this actually happens can be found in Section 6 (see in particular Table 1).

5. Admissible parameters

As a consequence of Propositions 1 and 2, we obtain that the condition

$$\sup_{|I|\leq k} \|A^*_{\beta,J} A_I (A^*_{\beta,I} A_I)^{-1}\|_\infty < 1 \tag{18}$$

is sufficient for $S_{c,d,k}$ to be a set of exact support recovery for the unmixing problem, provided that the ratio c/d is sufficiently large; the condition for the single-parameter case can be extracted from (18) by setting $\beta = \infty$, in which case A_β reduces to A.

Now define the signal-to-noise ratio of a pair (u, v) as

$$\rho(u,v) := \frac{\inf\{|u_i| : i \in \operatorname{supp}(u)\}}{\|v\|_\infty}.$$

That is, $\rho(u, v)$ is the ratio of the smallest significant value of the signal u, and the largest value of the noise v. Denote moreover

$$R(\beta,k) := \max_{|I|\leq k} \Bigg\{ \frac{\|A^*_{\beta,J}(A_I(A^*_{\beta,I}A_I)^{-1}A^*_{\beta,I} - \mathrm{Id})A\|_\infty \|(A^*_{\beta,I}A_I)^{-1}\|_\infty}{1 - \|A^*_{\beta,J}A_I(A^*_{\beta,I}A_I)^{-1}\|_\infty} + \|(A^*_{\beta,I}A_I)^{-1}A^*_{\beta,I}A\|_\infty \Bigg\}.$$

Then the inequality (16) implies that multi-penalty regularization with parameter β allows for the recovery of the support of k-sparse vectors u from data $A(u+v)$ provided the signal-to-noise ratio of the pair (u, v) satisfies

$$\rho(u,v) > R_{\beta,k}.$$

Whenever the signal-to-noise ratio is larger than $R_{\beta,k}$, we can recover the support of the vector u with *some* regularization parameter α. There are, however, upper and lower limits for the admissible parameters α, given by inequality (17). In order to visualize them, we consider instead the ratio

$$\theta(\alpha, v) := \frac{\alpha}{\|v\|_\infty}.$$

Defining

$$\Theta^{\min}_{\beta,k} := \max_{|I|\leq k} \frac{\|A^*_{\beta,J}(A_I(A^*_{\beta,I}A_I)^{-1}A^*_{\beta,I} - \mathrm{Id})A\|_\infty}{1 - \|A^*_{\beta,J}A_I(A^*_{\beta,I}A_I)^{-1}\|_\infty}$$

and

$$\Theta^{\max}_{\beta,k}(\vartheta) := \min_{|I|\leq k} \frac{\vartheta - \|(A^*_{\beta,I}A_I)^{-1}A^*_{\beta,I}A\|_\infty}{\|(A^*_{\beta,I}A_I)^{-1}\|_\infty},$$

we then obtain the condition

$$\Theta^{\min}_{\beta,k} \leq \theta(\alpha, v) < \Theta^{\max}_{\beta,k}(\rho(u,v)) \tag{19}$$

for exact support recovery. If the ratio $\theta(\alpha, v)$ is smaller than $\Theta^{\min}_{\beta,k}$, then it can happen that some of the noise v is not filtered out by the regularization method. On the other hand, if $\theta(\alpha, v)$ is larger than $\Theta^{\max}_{\beta,k}(\rho(u, v))$, then some parts of the signal $u^\dagger$ might actually be lost because of the regularization.

We note that the function $\Theta^{\max}_{\beta,k}$ is piecewise linear and concave, and $\lim_{\vartheta\to\infty} \Theta^{\max}_{\beta,k}(\vartheta) = +\infty$. Thus the region of admissible parameters defined by (19) is a convex and unbounded polyhedron. Moreover, we have that $\Theta^{\max}_{\beta,k}(R_{\beta,k}) = \Theta^{\min}_{\beta,k}$. Additionally, we note that the behaviour of the function $\Theta^{\max}_{\beta,k}$ near infinity is determined by the term

$$\Sigma_{\beta,k} := \max_{|I|\leq k} \|(A^*_{\beta,I} A_I)^{-1}\|_\infty.$$

If this value is small, then the slope of the function $\Theta^{\max}_{\beta,k}(\vartheta)$ for large values of ϑ is large, and thus the set of admissible parameter grows fast with increasing signal-to-noise ratio. If, on the other hand, $\Sigma_{\beta,k}$ is large, then the set of admissible parameters is relatively small even for large signal-to-noise ratio. Thus $\Sigma_{\beta,k}$ can be reasonably interpreted as the sensitivity of multi-parameter regularization with respect to parameter choice. The larger $\Sigma_{\beta,k}$ is, the more precise the parameter α has to be chosen in order to guarantee exact support recovery.

6. Numerical Validation

The main motive behind the study and application of multi-penalty regularization is the problem that ℓ^1-regularization is often not capable to identify the support of signal correctly (see [2] and references therein). Including the additional ℓ^2-regularization term, however, might lead to an improved performance in terms of support recovery, because we can expect that the ℓ^2-term takes care of all the small noise components.

In order to verify this observation, a series of numerical experiments was performed, in which we illustrate for which parameters and Gaussian matrices the conditions for support recovery derived in the previous section were satisfied. In addition, we studied whether the inclusion of the ℓ^2-term indeed increases the performance.

In a first set of experiments, we have generated a set of 20 Gaussian random matrices of different sizes and have tested for each three-dimensional subspace spanned by the basis elements whether the condition (15) is satisfied, first for the single-penalty case, and then for the multi-penalty case with different values of β. The results for matrices of dimensions 30 times 60 and 40 times 80, respectively, are summarized in Table 1 and Figure 1.

As to be expected from the bad numerical performance of ℓ^1-regularization in terms of support recovery, the inequality (11) fails in a relatively large number of cases, especially when the discrepancy between the dimension N of the vectors to be recovered and the number of measurements m is quite large. For instance, in the case $N = 60$ and $m = 30$, the condition most of the time failed for more than half of the

three-dimensional subspaces. In contrast, the corresponding condition (15) for multi-parameter regularization fails in the same situation only for about an eighth of the subspaces if $\beta = 1$, and in even fewer cases for $\beta = 0.1$.

For other combinations of dimensionality of the problem and number of measurements, the situation is similar. Introducing the additional ℓ^2-penalty term always allows for the exact support reconstruction on a larger number of subspaces than single-penalty regularization. Additionally, the results indicate that the number of recoverable subspaces increases with decreasing β.

$m = 30$	Single-penalty	Multi-penalty		
$N = 60$		$\beta = 10$	$\beta = 1$	$\beta = 0.1$
Median	0.5425	0.3814	0.1214	0.0623
Mean	0.5559	0.3922	0.1225	0.0635
Standard deviation	0.05652	0.04142	0.01518	0.01083
$m = 40$	Single-penalty	Multi-penalty		
$N = 80$		$\beta = 10$	$\beta = 1$	$\beta = 0.1$
Median	0.2696	0.1523	0.0396	0.0256
Mean	0.2746	0.1547	0.0413	0.0262
Standard deviation	0.03060	0.01848	0.00659	0.00447

Table 1. Percentage of 3-sparse subspaces for which the condition (15) failed. The condition was tested on samples of 20 Gaussian random matrices of dimensions 30 times 60 (upper table) and 40 times 80 (lower table). Other combinations of dimensionality and number of measurements showed qualitatively similar results.

In the case where $N = 80$ and $m = 60$ (that is, we want to reconstruct 80-dimensional vectors from 60 measurements), the sufficient condition (15) for multi-penalty regularization was satisfied in our numerical experiments for all 3-sparse subspaces for parameters β smaller than 5. In this situation, we have therefore additionally computed the significant values $R_{\beta,3}$ and $\Sigma_{\beta,3}$, that is, the minimal recoverable signal-to-noise ratio and the parameter sensitivity for three-dimensional subspaces. The results of these numerical experiments are shown in Table 2 and Figure 2.

The results indicate that decreasing the regularization parameter β tends to decrease the necessary signal-to-noise ratio $R_{\beta,k}$ as well. While a regularization parameter $\beta = 5$ required always an unreasonably large signal-to-noise ratio in order to guarantee the recoverability of the support, the ratios for $\beta = 0.5$ or $\beta = 0.1$ turned out to be much more reasonable. However, the results also showed that decreasing the regularization parameter need not necessarily have a beneficial effect on the recoverability. For some matrices it happened that the necessary signal-to-noise ratio $R_{\beta,k}$ increased while the regularization parameter β was decreased (see Figure 2, left).

Additionally, the results indicate that the parameter sensitivity $\Sigma_{\beta,k}$ increases considerably as β decreases (see Figure 2, right). As a consequence, the range of

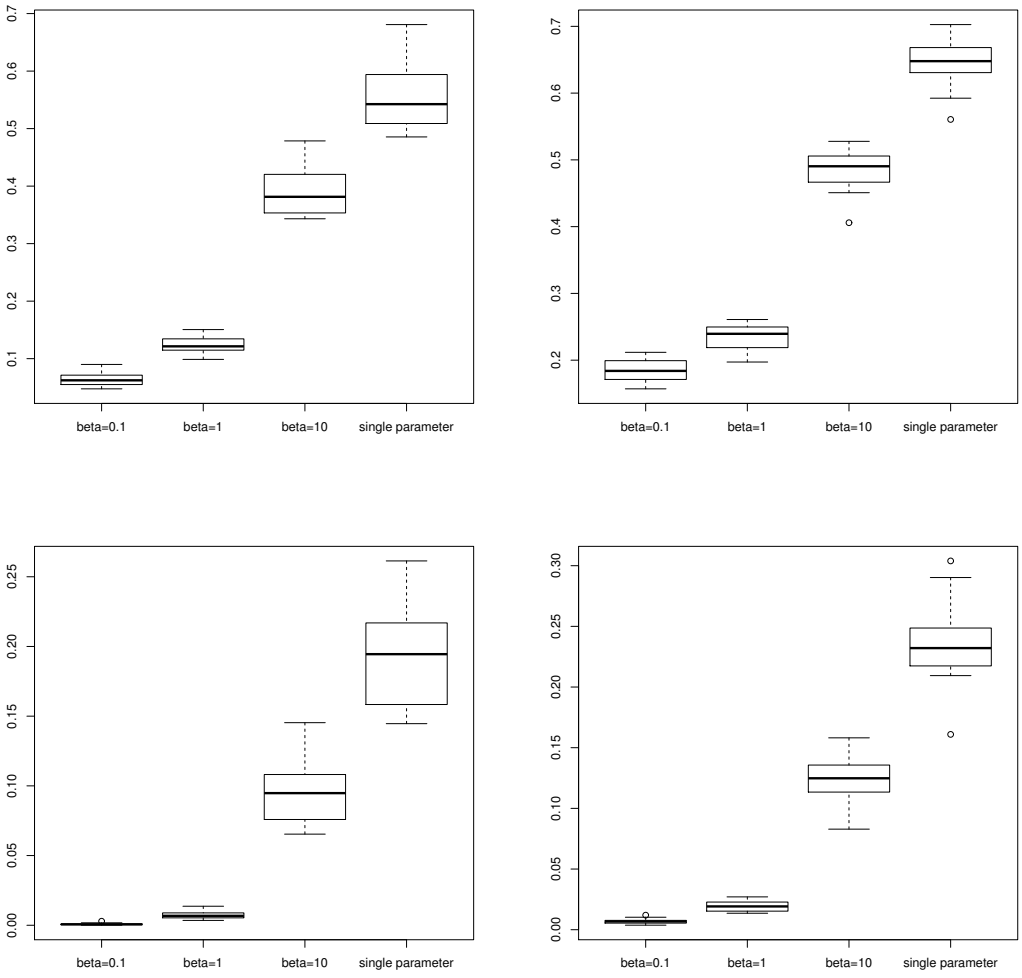

Figure 1. Percentage of 3-sparse subspaces for which the condition (15) failed. For each of the different settings the condition has been tested on 20 Gaussian random matrices for multi-parameter regularization with $\beta = 0.1$, $\beta = 1$ and $\beta = 10$, and for single parameter regularization. *Upper left:* Dimension 30 times 60. *Upper right:* Dimension 30 times 80. *Lower left:* Dimension 40 times 60. *Lower right:* Dimension 40 times 80.

admissible parameters α tends to be significantly smaller for smaller β, and it can happen much more easily that the combined effect of ℓ^2 and ℓ^1-regularization leads to a classification of signals as noise. While multi-penalty regularization with a small parameter β might therefore lead to a better necessary signal-to-noise ratio for recovery, it requires at the same time a better balance between the two involved parameters α and β.

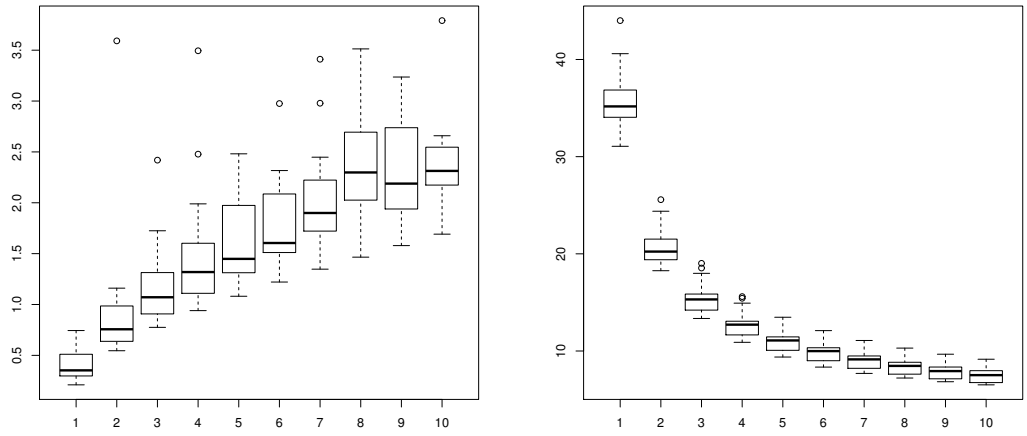

Figure 2. Influence of the regularization parameter β on the set of exact support recovery and the admissible parameters. *Left:* Logarithm of base 10 of the recoverable signal-to-noise ratio $R_{\beta,3}$ for different values of β. *Right:* The sensitivity $\Sigma_{\beta,3}$ for different values of β.
The dimension of the matrix is in all cases 60 times 80, and 20 Gaussian random matrices have been used for each parameter β.

		β				
		0.1	0.3	0.5	1	5
$R_{\beta,3}$	minimum	1.623	5.961	12.05	49.14	3599.3
	median	2.252	11.84	28.17	205.7	15041.4
	maximum	5.546	262.5	302.85	6178.4	203621.6
$\Sigma_{\beta,3}$	minimum	31.07	13.35	9.379	6.517	3.793
	median	35.17	15.31	11.086	7.517	4.278
	maximum	44.00	19.02	13.474	9.154	4.977

Table 2. Influence of the parameter β on the values of $R_{\beta,k}$ and $\Sigma_{\beta,k}$. The values have been computed for 20 Gaussian random matrices of dimension 60 times 80 with $k = 3$.

7. Applications. Numerical experiments

In order to support our theoretical findings even further, we present in this section some statistical data obtained by solving series of compressive sensing problems by means of multi-penalty and ℓ^1-regularization. Similarly to [11, 2] we consider in our numerical experiments the model problem of the type

$$y = T(u^\dagger + v),$$

where $T \in \mathbb{R}^{m\times N}$ is an i.i.d. Gaussian matrix, $u^\dagger$ is a sparse vector and v is a noise vector. The choice of T corresponds to compressed sensing measurements [7].

In the experiments, we consider 30 problems of this type with $u^\dagger$ randomly generated, $\inf_{i\in I}|u^\dagger| > 1.5$ and $\#\operatorname{supp}(u^\dagger) = 7$, and v is a random vector whose components are uniformly distributed on $[-1,1]$, and normalized such that $\|v\|_\infty = 0.3$. We also consider the Gaussian matrices of the size $m = 50$, $N = 100$.

In order to approximate minimizers of the multi-penalty (6) and the corresponding single-penalty functional we use the iterative soft-thresholding algorithm [6]. The regularization parameters α and β were chosen from the grid $Q^k_{\alpha_0} \times Q^k_{\beta_0}$, where $Q^k_{\alpha_0} := \{\alpha = \alpha_i = \alpha_0 k^i, \alpha_0 = 0.0002, k = 1.25, i = 0,\dots,50\}$, and $Q^k_{\beta_0} := \{\beta = \beta_i = \beta_0 k^i, \beta_0 = 0.01, q = 1.15, i = 0,\dots,30\}$. For all possible combinations of (α,β) we run the iterative soft-thresholding algorithm with fifty inner loop iterations and starting values $u^{(0)} = v^{(0)} = 0$.

In order to assess the obtained results, we compare the performance of the considered regularization schemes. We measure the approximation error (AE) by $\|u - u^\dagger\|_2$, as well as by the number of elements in the symmetric difference (SD) $\#(\operatorname{supp}(u)\Delta\operatorname{supp}(u^\dagger))$. The SD is defined as follows: $i \in \operatorname{supp}(u)\Delta\operatorname{supp}(u^\dagger)$ if and only if either $i \notin \operatorname{supp}(u)$ and $i \in \operatorname{supp}(u^\dagger)$ or $i \in \operatorname{supp}(u)$ and $i \notin \operatorname{supp}(u^\dagger)$.

For each problem we compute the best multi-penalty solution $u^\dagger = u^\dagger(\alpha,\beta)$ meaning that no other pairs $(\alpha,\beta) \in Q^k_{\alpha_0} \times Q^k_{\beta_0}$ can improve the accuracy of the algorithm. Simultaneously, for each problem from our data set we compute the best mono-penalty solution $u^\dagger = u^\dagger(\alpha)$. Then, we compute the mean value of the AE and SD. The respective results are shown in table 3. Additionally, figure 3 shows an example of the typical results obtained for single- and for multi-penalty regularization.

		AE	*SD*	α	β
	minimum	5.00	3	4.59	
SP	mean	11.21	6	5.74	
	maximum	13.74	8	11.21	
	minimum	1.05	1	3.67	0.76
MP	mean	5.92	3	5.74	7.09
	maximum	8.55	5	8.97	9.42

Table 3. For 30 problems for the solution of the single-penalty (upper panel) as well as multi-penalty regularization (lower panel) the minimum / maximum AE, SD and optimal values of the regularization parameters are provided. The mean values are also provided.

[1] E. Arias-Castro and Y. Eldar. Noise folding in compressed sensing. *IEEE Signal Process. Lett*, pages 478–481, 2011.

[2] M. Artina, M. Fornasier, and S. Peter. Damping noise-folding and enhanced support recovery in compressed sensing. *IEEE Trans. Signal Proc.*, 63:5990–6002, 2015.

[3] K. Bredies and M. Holler. Regularization of linear inverse problems with total generalized variation. *Journal of Inverse and Ill-posed Problems*, 1569-3945 (online):1–38, 2014.

[4] E. Candes and T. Tao. The dantzig selector: Statistical estimation when p is much larger than n. *Ann. Statist.*, 35:2313–2351, 2007.

[5] M.A. Davenport, J.N. Laska, J. Treichler, and R.G. Baraniuk. The pros and cons of compressive

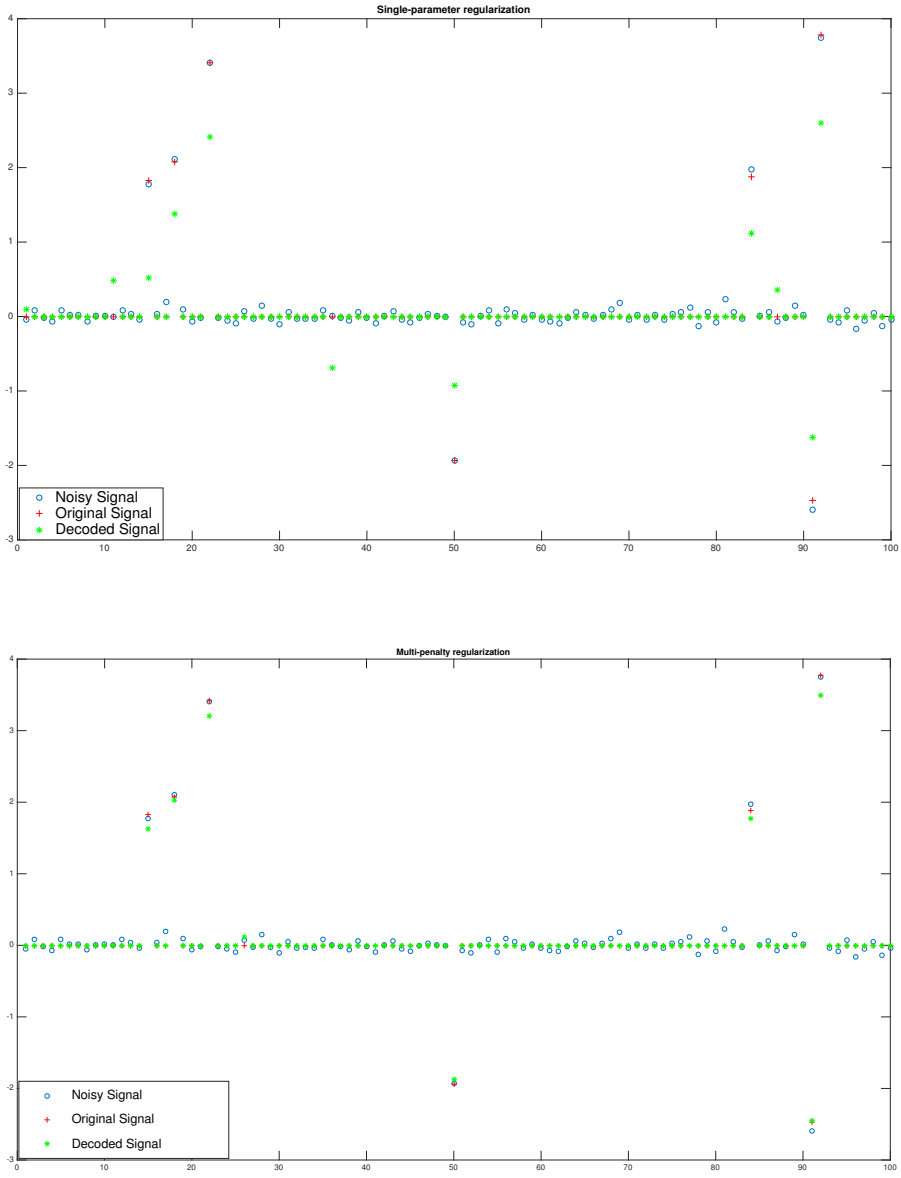

Figure 3. The figure reports the results of two different decoding procedures of the same problem, where the circles represent the noisy signal and the crosses represent the original signal. *Upper figure:* results with single-penalty regularization. *Lower figure:* results with multy-penalty regularization. Note that multi-penalty regularization allows for a better reconstruction of the support of the true signal $u^\dagger$.

sensing for wideband signal acquisition: Noise folding versus dynamic range. *Signal Processing, IEEE Transactions on*, 60(9):4628–4642, Sept 2012.

[6] M. Fornasier. Domain decomposition methods for linear inverse problems with sparsity constraints. *Inverse Problems*, 23(6):2505–2526, 2007.

[7] S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing*. Springer, New York, 2013.

[8] M. Grasmair, O. Scherzer, and M. Haltmeier. Necessary and sufficient conditions for linear convergence of l1-regularization. *Communications on Pure and Applied Mathematics*, 64(2):161–182, 2011.

[9] S. Lu and S. V. Pereverzev. Multi-parameter regularization and its numerical realization. *Numer. Math.*, 118(1):1–31, 2011.

[10] V. Naumova and S. V. Pereverzyev. Multi-penalty regularization with a component-wise penalization. *Inverse Problems*, 29(7):075002, 2013.

[11] V. Naumova and S. Peter. Minimization of multi-penalty functionals by alternating iterative thresholding and optimal parameter choices. *Inverse Problems*, 30:125003, 1–34, 2014.

[12] W. Wang, S. Lu, H Mao, and J. Cheng. Multi-parameter Tikhonov regularization with ℓ^0 sparsity constraint. *Inverse Problems*, 29:065018, 2013.