Inverse Problems

PAPER

Elastic-net regularization versus ℓ^1-regularization for linear inverse problems with quasi-sparse solutions

To cite this article: De-Han Chen *et al* 2017 *Inverse Problems* **33** 015004

View the article online for updates and enhancements.

Related content

- Convergence rates in \ell{1}-regularization if the sparsity assumption fails
  Martin Burger, Jens Flemming and Bernd Hofmann
- On the interplay of basis smoothness and specific range conditions occurring in sparsity regularization
  Stephan W Anzengruber, Bernd Hofmann and Ronny Ramlau
- Parameter choice in Banach space regularization under variational inequalities
  Bernd Hofmann and Peter Mathé

This content was downloaded from IP address 137.189.49.36 on 22/03/2018 at 08:25

IOP Publishing Inverse Problems

Inverse Problems **33** (2017) 015004 (17pp) doi:10.1088/1361-6420/33/1/015004

Elastic-net regularization versus ℓ^1-regularization for linear inverse problems with quasi-sparse solutions

De-Han Chen[1], Bernd Hofmann[2,3] and Jun Zou[1,4,5]

[1] Department of Mathematics, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong
[2] Technische Universität Chemnitz, Fakultät für Mathematik, 09107 Chemnitz, Germany

E-mail: dhchen@math.cuhk.edu.hk, bernd.hofmann@mathematik.tu-chemnitz.de and zou@math.cuhk.edu.hk

Received 10 March 2016, revised 12 November 2016
Accepted for publication 21 November 2016
Published 9 December 2016

Abstract

We consider the ill-posed operator equation $Ax = y$ with an injective and bounded linear operator A mapping between ℓ^2 and a Hilbert space Y, possessing the unique solution $x^\dagger = \{x^\dagger{}_k\}_{k=1}^\infty$. For the cases that sparsity $x^\dagger \in \ell^0$ is expected but often slightly violated in practice, we investigate in comparison with the ℓ^1-regularization the elastic-net regularization, where the penalty is a weighted superposition of the ℓ^1-norm and the ℓ^2-norm square, under the assumption that $x^\dagger \in \ell^1$. There occur two positive parameters in this approach, the weight parameter η and the regularization parameter as the multiplier of the whole penalty in the Tikhonov functional, whereas only one regularization parameter arises in ℓ^1-regularization. Based on the variational inequality approach for the description of the solution smoothness with respect to the forward operator A and exploiting the method of approximate source conditions, we present some results to estimate the rate of convergence for the elastic-net regularization. The occurring rate function contains the rate of the decay $x^\dagger{}_k \to 0$ for $k \to \infty$ and the classical smoothness properties of $x^\dagger$ as an element in ℓ^2.

Keywords: linear ill-posed problems, sparsity constraints, elastic-net regularization, ℓ^1-regularization, convergence rates, source conditions

[3] The research work was initiated during the visit of this author to The Chinese University of Hong Kong in March 2014 and supported partially by a Direct Grant for Research from The Chinese University of Hong Kong. The work of this author was substantially supported by German Research Foundation (DFG) under grant HO 1454/8-2.
[4] The work of this author was substantially supported by Hong Kong RGC General Research Fund (Projects 405513 and 14306814).
[5] Author to whom any correspondence should be addressed.

 © 2016 IOP Publishing Ltd Printed in the UK

1. Introduction

In this paper, we are interested in studying the linear ill-posed problem

$$Ax = y, \quad x \in \ell^2,\ y \in Y, \tag{1.1}$$

where Y is an infinite dimensional real Hilbert space, and $A : \ell^2 \to Y$ an injective and bounded linear operator with a non-closed range. Since $A^{-1} : \text{range}(A) \subset Y \to \ell^2$ is unbounded in this case, the corresponding system (1.1) suffers from ill-posedness in the sense that solutions may not exist if the exact data $y = Ax^\dagger$ for $x^\dagger \in \ell^2$ comes with noise, namely only the noisy data y^δ of y is available, where $\delta > 0$ represents the noisy level in the data, i.e. $\|y^\delta - y\|_Y \leqslant \delta$; and when solutions exist they may still be far away from the exact solution $x^\dagger$ to (1.1), even if δ is small. In section 2, we will outline that general ill-posed linear operator equations in Hilbert spaces can be rewritten into the form (1.1).

The most widely adopted approach for regularizing the ill-posed system (1.1) is the Tikhonov regularization, which aims at finding the approximate solutions x_γ^δ to problem (1.1) as the minimizers of the variational problem

$$T_\gamma^\delta(x) := \frac{1}{2}\|Ax - y^\delta\|_Y^2 + \gamma\mathcal{R}(x) \to \min, \tag{1.2}$$

where $\gamma > 0$ is the regularization parameter, and $\mathcal{R}(x)$ is the penalty that may be chosen appropriately, with popular examples like $\|x\|_{\ell^2}^2$, $\|x\|_{TV}$, $\|x\|_H$ or $\|x\|_{\ell^q}^q := \sum_{k=1}^\infty |x_k|^q$ for $1 \leqslant q < +\infty$. In particular, it was shown in [19] and [30] that the penalty functional $\|x\|_{\ell^1}$ ensures that the ℓ^1-regularized solutions x_γ^δ to the variational problem

$$T_\gamma^\delta(x) := \frac{1}{2}\|Ax - y^\delta\|_Y^2 + \gamma\|x\|_{\ell^1} \to \min, \quad \text{subject to} \quad x \in \ell^1, \tag{1.3}$$

provide stable approximate solutions to equation (1.1) if the exact solution $x^\dagger = \{x^\dagger{}_k\}_{k=1}^\infty$ is sparse, i.e. $x_k \neq 0$ occurs only for a finite number of components. The sparsity has been recognized as an important structure in many fields, e.g. geophysics [37], imaging science [14], statistics [38] and signal processing [10], and hence has received considerable attention. In this work, motivated by the recent works on the multi-parameter Tikhonov functional [28, 29, 39], we consider the following multi-parameter variational problem

$$T_{\alpha,\beta}^\delta(x) := \frac{1}{2}\|Ax - y^\delta\|_Y^2 + \alpha\|x\|_{\ell^1} + \frac{\beta}{2}\|x\|_{\ell^2}^2 \to \min, \quad \text{subject to} \quad x \in \ell^1, \tag{1.4}$$

which is called *the elastic-net regularization*. The functional $T_{\alpha,\beta}^\delta$ was originally used in statistics [43]. The major motivation is the observation that the ℓ^1-regularization fails to identify group structure for problems with highly correlated features, and tends to select only one feature out of a relevant group. It was proposed and confirmed numerically in [43] that the elastic-net regularization may retrieve the whole relevant group correctly. For an application of the elastic-net regularization to learning theory, one may refer to [11]. Furthermore, the stability of the minimizer and its consistency have been studied, and convergence rates for both *a priori* and *a posteriori* parameter choice have been established under suitable source conditions (see [29]). Moreover, we would also like to emphasize that the elastic-net regularization can be viewed as a special case of the ℓ^1-regularization, due to the identity

$$\frac{1}{2}\|Ax-y^\delta\|_Y^2+\alpha\,\|x\|_{\ell^1}+\frac{\beta}{2}\|x\|_{\ell^2}^2=\frac{1}{2}\left\|\begin{bmatrix}A\\ \sqrt{\beta}I\end{bmatrix}x-\begin{bmatrix}y^\delta\\ 0\end{bmatrix}\right\|_{Y\times\ell^2}^2+\alpha\,\|x\|_{\ell^1}.$$

As it was done for the ℓ^1-*regularization* (1.3) in [8], we intend to enrich with the present work the analysis on elastic-net regularization by taking into account the case that the solution $x^\dagger$ is not truly sparse in many applications, but has infinitely many nonzero components $x^\dagger{}_k$ that decay sufficiently rapidly to zero as $k\to\infty$. We shall call this kind of solutions to be *quasi-sparse* in the sequel for convenience, which occur often in practice, e.g. when applying wavelets to audio signals or natural images (see [26, 40]), where the compression algorithms are usually constructed by making use of the fact that most coefficients are very small and can be ignored. We shall model the quasi-sparse solutions with the assumption $x^\dagger\in\ell^1$. Following [29], we consider for elastic-net regularization the pair (β,η) of positive regularization parameters instead of the pair (α,β) by setting $\eta:=\alpha/\beta$, then (1.4) is reformulated as

$$T^\delta_{\beta,\eta}(x):=\frac{1}{2}\|Ax-y^\delta\|_Y^2+\beta\,\mathcal{R}_\eta(x)\to\min,\quad\text{subject to}\quad x\in X,\tag{1.5}$$

with the penalty functional

$$\mathcal{R}_\eta(x):=\eta\,\|x\|_{\ell^1}+\frac{1}{2}\|x\|_{\ell^2}^2,\tag{1.6}$$

and we denote by $x^\delta_{\beta,\eta}$ the minimizers to (1.5). The degenerate form of ℓ^1-regularization, i.e. $\beta=0$ in (1.4), was studied intensively including convergence rates in [8] (see also the extensions in [17] and references therein) for the cases with quasi-sparse solutions. We will extend the results from [8], [22] and [29] to analyze the two-parameter situation of elastic-net regularization with respect to convergence rates when the sparsity assumption fails. It is worth mentioning that other modifications of (1.3) have already been discussed in literature, for instance, the term $\gamma\,\|x\|_{\ell^1}$ in the penalty functional may be replaced by some weighted or modified versions (see e.g. [30, 34]), or alternatively by non-convex sparsity-promoting terms like $\gamma\,\|x\|_{\ell^q}$ for $0<q<1$ (see e.g. [7, 42]) or $\gamma\,\|x\|_{\ell^0}:=\gamma\sum_{k=1}^\infty\operatorname{sgn}(|x_k|)$ (see [39]). However, the theory with respect to convergence rates for the cases with quasi-sparse solutions are still rather limited compared with the case of truly sparse solutions.

The paper is organized as follows. In section 2 we will fix the basic problem setup, notations and assumptions, and then proceed to an overview of the smoothness conditions for proving convergence rates of single or multi-parameter Tikhonov regularization, where we shall show that the source conditions do not hold when the sparsity is violated. In section 3 we derive the convergence rates of regularized solutions for general linear ill-posed problems under variational inequalities, in which the regularization parameter is chosen according to three varieties of *a posteriori* parameter choices, i.e. two-sided discrepancy principle, sequential discrepancy principle and adapted Lepski principle. These results are then applied to ℓ^1-regularization (1.2) directly in section 4. In section 5, by deriving an appropriate variational inequality, we establish the convergence rates for the regularized solutions $x^\delta_{\beta,\eta}$ of elastic-net regularization (1.4) for a fixed $\eta>0$ and the case with quasi-sparse solutions.

2. Problem setting and basic assumptions

Let $\widetilde{X}$ (resp. Y) be an infinite dimensional real Hilbert space, endowed with an inner product $\langle\cdot,\cdot\rangle_{\widetilde{X}}$ (resp. $\langle\cdot,\cdot\rangle_Y$) and a norm $\|\cdot\|_{\widetilde{X}}$ (resp. $\|\cdot\|_Y$), $\widetilde{X}$ be separable, and $\widetilde{A}\in\mathcal{L}(\widetilde{X},Y)$ an injective

and bounded linear operator mapping between $\widetilde{X}$ and Y. In addition, we assume that $\mathrm{range}(\widetilde{A})$ of $\widetilde{A}$ is not closed, which is equivalent to that the inverse $\widetilde{A}^{-1} : \mathrm{range}(\widetilde{A}) \subset Y \to \widetilde{X}$ is unbounded. Thus the operator equation

$$\widetilde{A}\,\tilde{x} \;=\; y, \qquad \tilde{x} \in \widetilde{X}, \quad y \in Y, \tag{2.1}$$

with uniquely determined solution $\widetilde{x^{\dagger}} \in \widetilde{X}$ is ill-posed. This means that for noisy data $y^{\delta} \in Y$ replacing $y \in \mathrm{range}(\widetilde{A})$ in (2.1), solutions may not exist, and even when they exist the solutions may be still far from $\widetilde{x^{\dagger}}$ under the deterministic noise model

$$\| y - y^{\delta} \|_{Y} \leqslant \delta, \tag{2.2}$$

with small noise level $\delta > 0$.

With the setting $A := \widetilde{A} \circ U$, where $U : \ell^2 \to \widetilde{X}$ is the unitary synthesis operator characterizing the isometric isomorphy between the separable Hilbert spaces $\widetilde{X}$ and ℓ^2, the operator equation (2.1) can be rewritten in the form

$$A\,x \;=\; y, \qquad x \in \ell^2, \quad y \in Y. \tag{2.3}$$

This transforms (2.1) to the desired structure (1.1). We note that A is also injective, so this linear operator equation is ill-posed, i.e. $\mathrm{range}(A) \neq \overline{\mathrm{range}(A)}^{\,Y}$.

For any sequence $x = \{x_k\}_{k=1}^{\infty}$, we will denote by $\|x\|_{\ell^q} := \left(\sum_{k=1}^{\infty} |x_k|^q\right)^{1/q}$ the norm in the Banach spaces ℓ^q for $1 \leqslant q < \infty$, and by $\|x\|_{\ell^\infty} := \sup_{k\in\mathbb{N}} |x_k|$ the norm in ℓ^∞. The same norm $\|x\|_{c_0} := \sup_{k\in\mathbb{N}} |x_k|$ is used for the Banach space c_0 of infinite sequences tending to zero. On the other hand, the symbol ℓ^0 will stand for the set of all sparse sequences x, where $x_k \neq 0$ occurs only for a finite number of components. In the sequel we also set for short

$$X := \ell^1,$$

and consequently X^* for the dual space ℓ^∞ of X.

In the sequel, let $\langle \cdot, \cdot \rangle_{B^* \times B}$ denote the dual pairing between a Banach space B and its dual space B^*, and $v_n \rightharpoonup v_0$ stand for the weak convergence in B, i.e. $\lim_{n\to\infty} \langle w, v_n \rangle_{B^*\times B} = \langle w, v_0 \rangle_{B^*\times B}$ for all $w \in B^*$. For a Hilbert space B we identify B and B^* such that weak convergence takes the form $\lim_{n\to\infty} \langle w, v_n \rangle_B = \langle w, v_0 \rangle_B$ for all $w \in B$. Furthermore, we denote by $e^{(k)}$, with 1 at the kth position for $k = 1, 2, \ldots$, the elements of the standard orthonormal basis in ℓ^2, which also is the normalized canonical Schauder basis in c_0 and ℓ^q $(1 \leqslant q < \infty)$. That is, we find $\lim_{n\to\infty} \| x - \sum_{k=1}^{n} x_k e^{(k)} \|_{c_0} = 0$ for all $x \in c_0$ and $\lim_{n\to\infty} \| x - \sum_{k=1}^{n} x_k e^{(k)} \|_{\ell^q} = 0$ for all $x \in \ell^q$, $1 \leqslant q < \infty$. For the operator $A : \ell^2 \to Y$ we can consider its adjoint operator $A^* : Y \to \ell^2$ by the condition

$$\langle v, Ax \rangle_Y \;=\; \langle A^* v, x \rangle_{\ell^2} \qquad \text{for all} \quad x \in \ell^2,\ v \in Y.$$

Now we are stating a set of assumptions for the further consideration of equation (1.1) with a uniquely determined solution $x^{\dagger}$ and of the regularized solutions $x^{\delta}_{\alpha,\beta}$ and x^{δ}_{γ} solving the extremal problems (1.3) and (1.4), respectively.

Assumption 2.1.

(a) *The operator A in equation (1.1) is an injective and bounded linear operator mapping ℓ^2 to the Hilbert space Y with a non-closed range. i.e.* $\mathrm{range}(A) \neq \overline{\mathrm{range}(A)}^{\,Y}$.

(b) *Element $x^{\dagger} \in \ell^1$ solves equation (1.1).*

(c) *For each* $k \in \mathbb{N}$, *there exists* $f^{(k)} \in Y$ *such that* $e^{(k)} = A^* f^{(k)}$, *i.e. it holds that* $x_k = \langle e^{(k)}, x\rangle_{\ell^2} = \langle f^{(k)}, Ax\rangle_Y$ *for all* $x = \{x_k\}_{k=1}^{\infty} \in \ell^2$.

Remark 2.2. Item (c) above seems to be only a technical condition. If one considers the general operator equation (2.1), then it is equivalent to that $u^{(k)} = \widetilde{A}^* f^{(k)}$, $k \in \mathbb{N}$, for all elements of the orthonormal basis $\{u^{(k)}\}_{k=1}^{\infty}$ in $\widetilde{X}$ characterizing the unitary operator U. However, this condition was motivated for a wide class of linear inverse problems by using the Gelfand triple [2]. This series of independent source conditions for all $e^{(k)}$ is nothing but a requirement on the choice of the basis elements $u^{(k)}$ in $\widetilde{X}$. Roughly speaking, the basis elements must be in some sense 'smooth enough' under the auspices of the operator $\widetilde{A}$.

Proposition 2.3. *The range of* $A^* : Y \to \ell^2$ *is a nonclosed subset of* ℓ^2 *but dense in the sense of the* ℓ^2*-norm, i.e.* $\overline{\mathrm{range}(A^*)}^{\,\ell^2} = \ell^2$. *On the other hand,* $\mathrm{range}(A^*)$ *is always a subset of* c_0 *and hence not dense in* ℓ^∞ *in the sense of the supremum norm, i.e.* $\overline{\mathrm{range}(A^*)}^{\,\ell^\infty} \neq \ell^\infty$.

Proof. The proof is based on the properties of A from assumption 2.1(a). Noting that $\ell^2 \subset c_0$, we know $\mathrm{range}(A^*) \subset c_0$ for the adjoint operator $A^* : Y \to \ell^2$. On the other hand, the condition $\overline{\mathrm{range}(A^*)}^{\,\ell^2} = \ell^2$ is a consequence of the injectivity of A, while the non-closedness of $\mathrm{range}(A^*)$ in Y follows from the closed range theorem (see [41]). □

The smoothness of the solution to the ill-posed operator equation (1.1) with respect to the forward operator A plays an important role for obtaining error estimates and convergence rates in Tikhonov-type regularization, e.g. see [35, 36]. Such smoothness can be expressed by source conditions. In particular, for the most prominent form of the ℓ^2-regularization

$$\frac{1}{2}\|Ax - y^\delta\|_Y^2 + \gamma\,\|x\|_{\ell^2}^2 \to \min, \quad \text{subject to} \quad x \in \ell^2, \tag{2.4}$$

the classical theory of the Tikhonov regularization in Hilbert spaces applies (see [13, 20]). By making use of the purely quadratic penalty, the minimizers x_γ^δ achieve the convergence rate

$$\|x_\gamma^\delta - x^\dagger\|_{\ell^2} = \mathcal{O}(\sqrt{\delta}) \quad \text{as} \quad \delta \to 0 \tag{2.5}$$

under the source condition that

$$x^\dagger = A^* v, \quad v \in Y, \tag{2.6}$$

when the regularization parameter is chosen *a priori* as $\gamma = \gamma(\delta) \sim \delta$ or *a posteriori* as $\gamma = \gamma(\delta, y^\delta)$ based on the discrepancy principle $\|Ax_\gamma^\delta - y^\delta\|_Y = \tau\,\delta$ for some prescribed $\tau \geqslant 1$.

If $x^\dagger$ is not smooth enough to satisfy (2.6), then the method of *approximate source conditions* may help to bridge this gap when the concave and nonincreasing distance function

$$d_{x^\dagger}^{A^*}(R) := \inf_{v \in Y:\, \|v\|_Y \leqslant R} \|x^\dagger - A^* v\|_{\ell^2}, \qquad R > 0, \tag{2.7}$$

tends to zero as $R \to \infty$. In such case the decay rate of $d_{x^\dagger}^{A^*}(R) \to 0$ as $R \to \infty$ characterizes the degree of violation with respect to (2.6). Then the convergence rate is slower than (2.5) and the rate function depends on $d_{x^\dagger}^{A^*}$ (see [22, 24], and also [4]). From proposition 2.3 we have for all elements $x^\dagger \notin \ell^2 \setminus \mathrm{range}(A^*)$ that $\lim_{R\to\infty} d_{x^\dagger}^{A^*}(R) = 0$, because $\overline{\mathrm{range}(A^*)}^{\,\ell^2} = \ell^2$.

If the element $x^\dagger \in \ell^2$ fails to satisfy the source condition (2.6), however, we know that there is an *index function* g and a source element $w \in \ell^2$ such that (see [33])

$$x^\dagger = g(A^*A)\, w \tag{2.8}$$

Here we say that $g : (0, \infty) \to (0, \infty)$ is an index function if it is continuous, strictly increasing and satisfies the condition $\lim_{t\to 0^+} g(t) = 0$. For $g(t) = \sqrt{t}$, the source conditions (2.6) and (2.8) are equivalent. For any $x^\dagger \in \ell^2$, the connection between g in (2.8) and the distance functions in (2.7) were outlined in [24, section 5.3]. In particular, we can find from [12, theorem 3.1] that, for any exponents $0 < \theta < 1$,

$$d_{x^\dagger}^{A^*}(R) \leqslant \frac{K}{R^{\frac{\theta}{1-\theta}}} \quad \forall\ R \geqslant \underline{R} \quad \text{if} \quad x^\dagger \in \text{range}[(A^*A)^{\theta/2}], \tag{2.9}$$

with some positive constants K and $\underline{R}$. In this case we say that $x^\dagger$ satisfies a Hölder source condition with exponent θ.

For a general convex but not purely quadratic penalty functional $\mathcal{R}$, the benchmark source condition is given by

$$\xi^\dagger = A^* v, \quad v \in Y, \tag{2.10}$$

for some subgradient $\xi^\dagger \in \partial\mathcal{R}(x^\dagger)$ and source element $v \in Y$ (see [9]). As a result of the following proposition, the source condition (2.10) for $\mathcal{R} = \mathcal{R}_\eta$ (see (1.6)) can only hold if the solution is sparse, i.e. $x^\dagger \in \ell^0$. Hence this condition, which is also important for the convergence rates of elastic-net regularization in [29], completely fails for the quasi-sparsity case of our interest in this work. The approximate source condition approach was extended to the general Banach space situation with convex penalties $\mathcal{R}$ in [21]. As an analog to (2.7), the corresponding distance functions for $\mathcal{R} = \mathcal{R}_\eta$ attain the form

$$d_{\xi^\dagger}(R) := \inf_{v\in Y:\, \|v\|_Y \leqslant R} \|\xi^\dagger - A^* v\|_{\ell^\infty}, \qquad R > 0. \tag{2.11}$$

However, the proposition below also implies that this approximate source condition also fails, i.e. $d_{\xi^\dagger}(R) \to 0$ as $R \to \infty$ cannot hold if $x^\dagger \in \ell^1 \setminus \ell^0$, which makes it impossible to verify convergence rates based on this approach.

On the other hand, explicit convergence rates of regularized solutions $x^\delta_{\alpha,\beta}$ for elastic-net Tikhonov regularization (1.4) may require smoothness properties of $x^\dagger$ (see [27]). More precisely, if for any $t \in [0, 1]$ there exists w_t such that

$$\xi_t := A^* w_t \in \partial\psi_t(x^\dagger), \tag{2.12}$$

where $\psi_t(x) := t\|x\|_{\ell^1} + (1-t)\|x\|^2_{\ell^2}$, then the convergence rate

$$D_{\xi_t}(x^\delta_{\alpha^*(\delta),\beta^*(\delta)}, x^\dagger) = \mathcal{O}(\delta)$$

can be established, for Bregman distance $D_{\xi_t}(x, x^\dagger) := \psi_t(x) - \psi_t(x^\dagger) - \langle \xi_t, x - x^\dagger\rangle_{\ell^2}$, and parameter choice $(\alpha^*(\delta), \beta^*(\delta))$ based on the multi-parameter discrepancy principle, i.e. $(\alpha^*(\delta), \beta^*(\delta))$ satisfies

$$\|x_{\alpha^*(\delta),\beta^*(\delta)} - x^\dagger\|_{l^2} = c_m\delta^2$$

with some prescribed constant $c_m \geqslant 1$. The proposition below also implies that the condition (2.12) fails if $x^\dagger$ is not truly sparse, because $\psi_t(x) = 2(1-t)\mathcal{R}_{t/(2-2t)}(x)$ for all $t \in [0, 1)$.

For each $\eta > 0$ the convex functional $\mathcal{R}_\eta$ defined in (1.6) attains finite values on X. Moreover for each $x \in X$, by using the subgradients $\zeta = \{\zeta_k\}_{k=1}^\infty \subset X^*$ of $\|x\|_{\ell^1}$, the subdifferential $\partial\mathcal{R}_\eta(x)$ collects all subgradients $\xi = \{\xi_k\}_{k=1}^\infty \subset X^*$ of the form

$$\xi_k = \eta\,\zeta_k + x_k, \quad \text{where} \quad \zeta_k \begin{cases} = 1 & \text{if} \quad x_k > 0, \\ \in [-1,1] & \text{if} \quad x_k = 0, \\ = -1 & \text{if} \quad x_k < 0, \end{cases} \qquad k \in \mathbb{N}. \tag{2.13}$$

Proposition 2.4. *If $x^\dagger \in \ell^1 \setminus \ell^0$ and $v \in Y$, then for any $\eta > 0$ the condition $A^*v \in \partial\mathcal{R}_\eta(x^\dagger)$ cannot hold. Also it does not hold that $d_{\xi^\dagger}(R) \to 0$ as $R \to \infty$ (see (2.11)) in this case.*

Proof. Assume that $A^*v \in \partial\mathcal{R}_\eta(x^\dagger)$ holds for some $x^\dagger \in \ell^1 \setminus \ell^0$ and $v \in Y$. Then, by formula (2.13) we have for every $\xi^\dagger := A^*v \in \partial\mathcal{R}_\eta(x^\dagger)$ that $\xi_k^\dagger = \pm\eta\zeta_k + x^\dagger{}_k$ for all $k \in \mathbb{N}$. Then using the signum function

$$\operatorname{sgn}(z) := \begin{cases} 1 & \text{if } z > 0, \\ 0 & \text{if } z = 0, \\ -1 & \text{if } z < 0, \end{cases}$$

there is a subsequence $\{x^\dagger{}_{k_l}\}_{l=1}^\infty$ of $\{x^\dagger{}_k\}_{k=1}^\infty$ such that $|\operatorname{sgn}(x^\dagger{}_{k_l})| = 1$ for all $l \in \mathbb{N}$ and $\lim_{l\to\infty}|x^\dagger{}_{k_l}| = 0$. Therefore we have

$$\left|\frac{[A^*v]_{k_l} - x^\dagger{}_{k_l}}{\eta}\right| = 1,$$

which gives a contradiction, because proposition 2.3 implies $A^*v \in c_0$, and hence that the left-hand side of this equation tends to zero for $l \to \infty$. The second assertion is a simple consequence of the fact that $\|\xi^\dagger - w\|_{\ell^\infty} \geqslant \eta$ holds for all elements $w \in c_0$ in this case. □

From the reasoning above we know that the source conditions always fail for the quasi-sparse solutions. To overcome the difficulty, we shall use the variational inequalities (variational source conditions) instead. For more discussions about smoothness of solutions and their influences on convergence rates, we refer to [23, 25] and [15, 16, 18] for further details.

3. Convergence rates under variational inequalities

Throughout this section we extend our consideration to the more general situation of an ill-posed operator equation

$$A\,x \;=\; y, \qquad x \in Z, \quad y \in Y, \tag{3.1}$$

and regularized solutions x_γ^δ to (3.1) with regularization parameter $\gamma > 0$, which are minimizers of the functional

$$T_\gamma^\delta(x) := \frac{1}{2}\|Ax - y^\delta\|_Y^2 + \gamma\,\mathcal{R}(x) \to \min, \quad \text{subject to} \quad x \in Z, \tag{3.2}$$

where the nonnegative penalty functional $\mathcal{R}$ is convex, lower semi-continuous and stabilizing, and Z is a Hilbert space. Here we call $\mathcal{R}$ stabilizing if the sublevel sets $\mathcal{M}_c := \{x \in Z :\ \mathcal{R}(x) \leqslant c\}$ are weakly sequentially compact subsets of Z for all $c \geqslant 0$. In this context, the linear forward operator $A : Z \to Y$ is assumed to be injective and bounded, and the uniquely determined solution $x^\dagger$ of (3.1) is required to satisfy the condition that $x^\dagger \in \mathcal{D}(\mathcal{R})$ with

$$\mathcal{D}(\mathcal{R}) := \{x \in Z :\ \mathcal{R}(x) < \infty\}.$$

Under the aforementioned assumptions on $\mathcal{R}$, we know that regularized solutions x_γ^δ exist for all $\gamma > 0$ and $y^\delta \in Y$, and are stable with respect to perturbations in the data y^δ (see e.g. [23, 35, 36]).

For an index function g and an error measure $E : \mathcal{D}(\mathcal{R}) \times \mathcal{D}(\mathcal{R}) \to \mathbb{R}_+$, we need an appropriate choice $\gamma_* = \gamma_*(\delta, y^\delta)$ of the regularization parameter and certain smoothness of $x^\dagger$ with respect to the forward operator A to obtain a convergence rate of the form

$$E(x_{\gamma_*}^\delta, x^\dagger) = \mathcal{O}(g(\delta)) \quad \text{as} \quad \delta \to 0. \tag{3.3}$$

In particular, the variational inequalities of the form

$$\lambda\, E(x, x^\dagger) \leqslant \mathcal{R}(x) - \mathcal{R}(x^\dagger) + C\, g(\|A(x - x^\dagger)\|_Y) \quad \text{for all} \quad x \in \mathcal{D}(\mathcal{R}) \tag{3.4}$$

become popular for the description of solution smoothness, which were developed independently in [15] and [18], where λ and C are constants satisfying $0 < \lambda \leqslant 1$ and $C > 0$, and the index function g was assumed to be concave.

Concerning appropriate selection strategies for regularization parameters, we list the following three principles for *a posteriori* parameter choice $\gamma_* = \gamma(\delta, y^\delta)$, for which proposition 3.2 below will apply and yield corresponding reasonable convergence rates.

TDP: For the prescribed τ_1 and τ_2 satisfying $1 \leqslant \tau_1 \leqslant \tau_2 < \infty$, the two-sided discrepancy principle (TDP) suggests to choose the regularization parameter $\gamma_* = \gamma_{\text{TDP}}$ such that

$$\tau_1\, \delta \leqslant \|Ax_{\gamma_{\text{TDP}}}^\delta - y^\delta\|_Y \leqslant \tau_2\, \delta. \tag{3.5}$$

Since the discrepancy functional $\|Ax_\gamma^\delta - y^\delta\|_Y$ is continuous and increasing with respect to $\gamma \in (0, \infty)$, the regularization parameter $\gamma_{\text{TDP}} > 0$ exists for all $y^\delta \in Y$ whenever $\delta > 0$ is sufficiently small. We refer to, e.g. [3] for more details.

SDP: For the prescribed τ, q and γ_0 satisfying $\tau > 1, 0 < q < 1$ and $\gamma_0 > 0$, and the decreasing geometric sequence

$$\Delta_q := \{\gamma_j :\ \gamma_j = q^j \gamma_0,\ j \in \mathbb{N}\},$$

the sequential discrepancy principle (SDP) suggests to choose the regularization parameter $\gamma_* = \gamma_{\text{SDP}}$ such that $\gamma_{\text{SDP}} \in \Delta_q$ satisfies

$$\|Ax_{\gamma_{\text{SDP}}}^\delta - y^\delta\|_Y \leqslant \tau\delta < \|Ax_{\gamma_{\text{SDP}}/q}^\delta - y^\delta\|_Y. \tag{3.6}$$

When using the SDP, we are interested in finding the largest value γ from the sequence Δ_q such that $\|Ax_\gamma^\delta - y^\delta\|_Y \leqslant \tau\delta$. For the well-definedness of γ_{SDP} from SDP, its properties and convergence of regularized solutions $x_{\gamma_{\text{SDP}}}^\delta$ as $\delta \to 0$, we refer to [1]. In principle, one can say that γ_{SDP} is uniquely determined for all $0 < q < 1$ and $y^\delta \in Y$ whenever $\gamma_0 > 0$ is large enough.

LEP: To apply the Lepski principle (LEP) for choosing the regularization parameter $\gamma > 0$ under (3.4), we restrict our consideration to the symmetric error measures E satisfying the triangle inequality up to some constant $1 \leqslant C_E < \infty$, i.e. for all $x^{(i)} \in \mathcal{D}(\mathcal{R}),\ i = 1, 2, 3,$

$$E(x^{(1)}, x^{(2)}) = E(x^{(2)}, x^{(1)}) \tag{3.7}$$

and

$$E(x^{(1)}, x^{(2)}) \leqslant C_E\,(E(x^{(1)}, x^{(3)}) + E(x^{(3)}, x^{(2)})). \tag{3.8}$$

For such symmetric error measures E, the prescribed q and γ_0 satisfying $q \in (0,1)$ and $\gamma_0 > 0$, the increasing geometric sequence

$$\widetilde{\Delta}_q := \{\gamma_j :\ \gamma_j = \gamma_0/q^j,\ j \in \mathbb{N}\},$$

and the strictly decreasing function

$$\Theta(\gamma) := \frac{17\delta^2}{2\lambda\gamma}$$

with a fixed $\delta > 0$ and λ from the variational inequality (3.4), which characterizes an upper bound of $E(x_\gamma^\delta, x^\dagger)$ for all $\gamma_0 \leqslant \gamma \leqslant \gamma_{\text{apri}}$ and *a priori* parameter choice $\gamma_{\text{apri}} = \frac{\delta^2}{C\,g(\delta)}$, the adapted Lepski principle (LEP) suggests to choose the regularization parameter $\gamma_* = \gamma_{\text{LEP}}$ such that γ_* is the largest value γ in $\widetilde{\Delta}_q$ satisfying

$$E(x_{\gamma'}^\delta, x_\gamma^\delta) \leqslant 2\,C_E\,\Theta(\gamma') \qquad \text{for all} \quad \gamma' \in \widetilde{\Delta}_q \quad \text{with} \quad \gamma_0 \leqslant \gamma' < \gamma. \tag{3.9}$$

We like to mention that the LEP is based on *a priori* parameter choice (see [25, section 4.2.1] and [31, section 1.1.5]). Under the variational inequality (3.4) with an error measure E, *a priori* parameter choice γ_{apri} yields the convergence rate (3.3) for $\gamma_* = \gamma_{\text{apri}}$ (see [25, section 4.1]).

For the adapted Lepski principle, the following error estimate holds.

Lemma 3.1 *Assume the variational inequality* (3.4) *holds for a nonnegative error measure* $E(\cdot)$ *satisfying* (3.7) *and* (3.8). *If* γ_0 *is sufficiently small such that* $E(x_{\gamma_0}^\delta, x^\dagger) \leqslant \Theta(\gamma_0)$, *then* γ_{LEP} *from LEP is uniquely determined for all* $0 < q < 1$ *and* $y^\delta \in Y$ *and meets the error estimate*

$$E(x_{\gamma_{\text{LEP}}}^\delta, x^\dagger) \leqslant C\,C_E(2 + C_E)\,\frac{17}{2q\lambda}\,g(\delta), \tag{3.10}$$

where the positive constants C *and* λ *are from* (3.4), *and* C_E *from* (3.8).

Proof. The proof goes along the same line as the one for theorem 3 in [25], in combination with proposition 1 from [32]. First we introduce

$$\gamma_+ := \max\{\gamma \in \widetilde{\Delta} :\ E(x_{\gamma'}^\delta, x^\dagger) \leqslant \Theta(\gamma') \quad \text{for all} \quad \gamma' \in \widetilde{\Delta}_q,\ \gamma_0 \leqslant \gamma' \leqslant \gamma\}.$$

Then we will show that $\gamma_k < \gamma_{\text{apri}} \leqslant \gamma_k/q$ for some $k \in \mathbb{N}$ and

$$E(x_{\gamma_{\text{LEP}}}^\delta, x^\dagger) \leqslant C_E(2 + C_E)\Theta(\gamma_+). \tag{3.11}$$

From [25, lemma 3] we derive that $E(x_\gamma^\delta, x^\dagger) \leqslant \hat{C}\,\delta^2/\gamma$ for all $\gamma_0 \leqslant \gamma \leqslant \gamma_{\text{apri}}$ with $\hat{C} = \frac{17}{2\lambda}$, which yields $E(x_\gamma^\delta, x^\dagger) \leqslant \Theta(\gamma)$ for all $\gamma_0 \leqslant \gamma \leqslant \gamma_{\text{apri}}$ and hence $\gamma_+ \geqslant \gamma_{\text{apri}}$. Therefore, we have for all $\gamma_0 \leqslant \gamma \leqslant \gamma_+$ the estimate

$$E(x_\gamma^\delta, x_{\gamma_+}^\delta) \leqslant C_E\Big(E(x_\gamma^\delta, x^\dagger) + E(x^\dagger, x_{\gamma_+}^\delta)\Big) \leqslant C_E(\Theta(\gamma) + \Theta(\gamma_+)) \leqslant 2C_E\Theta(\gamma).$$

This ensures the inequality $\gamma_{\mathrm{LEP}} \geqslant \gamma_+$. Then we can find

$$E(x^\delta_{\gamma_{\mathrm{LEP}}}, x^\dagger) \leqslant C_E\Big(E(x^\delta_{\gamma_{\mathrm{LEP}}}, x^\delta_{\gamma_+}) + E(x^\delta_{\gamma_+}, x^\dagger)\Big) \leqslant 2C_E^2\Theta(\gamma_+) + C_E\Theta(\gamma_+)$$
$$= C_E(2 + C_E)\Theta(\gamma_+),$$

which gives (3.11). By (3.11) we obtain in analogy to the proof of theorem 3 in [25] that

$$E(x^\delta_{\gamma_{\mathrm{LEP}}}, x^\dagger) \leqslant C_E(2 + C_E)\Theta(\gamma_+) \leqslant C_E(2 + C_E)\,\Theta(\gamma_k) = \frac{C_E(2 + C_E)}{q}\,\Theta\left(\frac{\gamma_k}{q}\right)$$
$$\leqslant \frac{C_E(2 + C_E)}{q}\,\Theta\Big(\gamma_{\mathrm{apri}}\Big) \leqslant C\,C_E\,(2 + C_E)\,\frac{17}{2\lambda q}\,g(\delta),$$

which completes the proof. □

The following convergence rate estimate follows directly from [15, theorem 4.24], [25, theorems 2 and 3] and lemma 3.1 for the three aforementioned *a posteriori* choices of regularization parameters.

Proposition 3.2. *If the variational inequality* (3.4) *is valid for a nonnegative error measure* $E : \mathcal{D}(\mathcal{R}) \times \mathcal{D}(\mathcal{R}) \to \mathbb{R}_+$ *with constants* $0 < \lambda \leqslant 1$, $C > 0$, *and a concave index function* g, *then we have the estimate* (3.3) *of the convergence rate for the regularized solutions* $x^\delta_{\gamma_*}$ *if the regularization parameter* γ_* *is chosen as* $\gamma_* = \gamma_{\mathrm{TDP}}$ *from the two-sided discrepancy principle, as* $\gamma_* = \gamma_{\mathrm{SDP}}$ *from the sequential discrepancy, or as* $\gamma_* = \gamma_{\mathrm{LEP}}$ *from the Lepski principle provided that* E *satisfies* (3.7) *and* (3.8).

In the subsequent two sections we will apply proposition 3.2 for the penalty functionals $\mathcal{R}(x) = \|x\|_{\ell^1}$ and $\mathcal{R}(x) = \eta\|x\|_{\ell^1} + \frac{1}{2}\|x\|^2_{\ell^2}$ (with an arbitrarily fixed $\eta > 0$), respectively.

4. Application to ℓ^1-regularization

For $Z := \ell^2$, $\mathcal{R}(x) := \|x\|_{\ell^1}$ and $\mathcal{D}(\mathcal{R}) = X = \ell^1$, proposition 3.2 applies with $\lambda = 1$, $C = 2$ and $g = \varphi$ defined by

$$\varphi(t) := \inf_{n\in\mathbb{N}}\left(\sum_{k=n+1}^{\infty} |x^\dagger_k| + t\sum_{k=1}^{n}\|f^{(k)}\|_Y\right), \tag{4.1}$$

as a consequence of [8, theorem 5.2], in which it was proven that the variational inequality (3.4) holds with $E(x, x^\dagger) = \|x - x^\dagger\|_{\ell^1}$. The error measure E is a metric in $\mathcal{D}(\mathcal{R})$, and hence (3.7) and (3.8) are valid with $C_E = 1$. Thus, we have

$$\|x - x^\dagger\|_{\ell^1} \leqslant \|x\|_{\ell^1} - \|x^\dagger\|_{\ell^1} + 2\,\varphi(\|A(x - x^\dagger)\|_Y) \quad \text{for all} \quad x \in X = \ell^1, \tag{4.2}$$

where φ is a concave index function. Here, the rate function g in (3.3) depends on the decay rate of the remaining components $x^\dagger{}_k \to 0$ as $k \to \infty$ and the behaviour of $\|f^{(k)}\|_Y$ (see assumption 2.1 (c)), which is mostly a growth to infinity as $k \to \infty$. The studies in [25] ensured that the same rate result is valid for all three regularization parameter choices $\gamma_* = \gamma_{\mathrm{TDP}}$, $\gamma_* = \gamma_{\mathrm{SDP}}$ and $\gamma_* = \gamma_{\mathrm{LEP}}$.

Example 5.3 in [8] makes the convergence rate g in ℓ^1-regularization explicit as a Hölder rate:

$$\|x_{\gamma_*}^{\delta} - x^{\dagger}\|_{\ell^1} = \mathcal{O}\left(\delta^{\frac{\mu}{\mu+\nu}}\right) \qquad \text{as} \qquad \delta \to 0 \tag{4.3}$$

for the case when monomials

$$\sum_{k=n+1}^{\infty} |x^{\dagger}{}_k| \leqslant K_1\, n^{-\mu}, \qquad \sum_{k=1}^{n} \|f^{(k)}\|_Y \leqslant K_2\, n^{\nu}, \tag{4.4}$$

with exponents $\mu, \nu > 0$ and some constants $K_1, K_2 > 0$, characterize the decay of the solution components and the growth of the $f^{(k)}$-norms, respectively.

On the other hand, example 5.3 in [6] outlines the situation that the $f^{(k)}$-norm growth is of power type, but instead of (4.4) the decay of $x^{\dagger}{}_k \to 0$ is much faster, expressed by an exponential decay rate. Precisely, if

$$\sum_{k=n+1}^{\infty} |x^{\dagger}{}_k| \leqslant K_1 \exp(-n^{\sigma}) \quad \text{and} \quad \sum_{k=1}^{n} \|f^{(k)}\|_Y \leqslant K_2\, n^{\nu} \tag{4.5}$$

hold with exponents $\sigma, \nu > 0$ and some constants $K_1, K_2 > 0$, then we have the convergence rate

$$\|x_{\gamma_*}^{\delta} - x^{\dagger}\|_{\ell^1} = \mathcal{O}\left(\delta\left(\log(\frac{1}{\delta})\right)^{\frac{\nu}{\sigma}}\right) \qquad \text{as} \qquad \delta \to 0, \tag{4.6}$$

instead of (4.3).

The rate (4.6) is not far from the best rate

$$\|x_{\gamma_*}^{\delta} - x^{\dagger}\|_{\ell^1} = \mathcal{O}(\delta) \qquad \text{as} \qquad \delta \to 0, \tag{4.7}$$

which was already established for truly sparse solutions $x^{\dagger} \in \ell^0$ in [19]. We note that, for all $x^{\dagger} \in \ell^0$ with some $k_{\max}$ as the largest index $k \in \mathbb{N}$ such that $x^{\dagger}{}_k \neq 0$, there is a uniform constant $K = \sum_{k=1}^{k_{\max}} \|f^{(k)}\|_Y$ such that the function $\varphi(t)$ in (4.1) can be estimated from above by Kt. Since mostly $\|f^{(k)}\|_Y$ grows rapidly to infinity as $k \to \infty$, the constant K may be large even if $k_{\max}$ is not big. As already mentioned in [29], the super-rate (4.7) of the ℓ^1-regularization may not be expected when the corresponding constant K explodes.

5. Application to elastic-net regularization

In this section, we fix the parameter $\eta > 0$ arbitrarily and consider the case with $Z := \ell^2$, $\mathcal{R}(x) := \mathcal{R}_\eta(x) = \eta\|x\|_{\ell^1} + \frac{1}{2}\|x\|_{\ell^2}^2$ and $\mathcal{D}(\mathcal{R}_\eta) = X = \ell^1$. Evidently, the penalty functional $\mathcal{R}_\eta$ is convex, lower semi-continuous and stabilizing in Z, which ensures the existence and stability of regularized solutions $x_{\beta_*,\eta}^{\delta}$ and its convergence $\lim_{\delta\to 0}\|x_{\beta_*,\eta}^{\delta} - x^{\dagger}\|_{\ell^2} = 0$ for any $\eta > 0$ if $\beta_* = \beta_{\text{TDP}}$ or $\beta_* = \beta_{\text{SDP}}$. We refer to [1, 8, 29] for further discussions. On the other hand, the application of proposition 3.2 to the elastic-net regularization (1.5), where β is chosen by TDP and SDP, or LEP, requires us to construct an appropriate variational inequality. Below (see theorem 5.2) we will perform this construction by a weighted superposition of the variational inequality (4.2) used in section 4 and a corresponding one for the ℓ^2-term in the penalty $\mathcal{R}_\eta$, which will be the main purpose of lemma 5.1.

Let us recall the distance function $d_{x^{\dagger}}^{A^*}(\cdot)$ defined in (2.7), and introduce a continuous and strictly decreasing auxiliary function for $R > 0$ and $x^{\dagger} \in \ell^2 \setminus \text{range}(A^*)$:

$$\Phi(R) := [d_{x^{\dagger}}^{A^*}(R)]^2/R. \tag{5.1}$$

Using the limit conditions

$$\lim_{R\to 0}\Phi(R)=+\infty \quad \text{and} \quad \lim_{R\to\infty}\Phi(R)=0, \tag{5.2}$$

it is not difficult to see that the function

$$\widehat{\psi}(t):=[d^{A^*}_{x^\dagger}(\Phi^{-1}(t))]^2, \quad t>0, \tag{5.3}$$

is an index function. On the other hand, for $x^\dagger\in\mathrm{range}(A^*)$ we can always find $R_0>0$ such that $d^*_{x^\dagger}(R_0)=0$.

Lemma 5.1. *The variational inequality* (4.2) *holds true for* $Z:=\ell^2$, $\mathcal{R}(x):=\frac{1}{2}\|x\|_{\ell^2}$ *and* $\mathcal{D}(\mathcal{R})=X=\ell^2$, $E(x,x^\dagger):=\|x-x^\dagger\|^2_{\ell^2}$ *and some index functions* $g:=\psi$, *and positive constants* λ *and* C. *More precisely, if* $x^\dagger\in\ell^2\setminus\mathrm{range}(A^*)$, *we can take* ψ *as a concave index function* $\psi:[0,\infty)\to\mathbb{R}$ *such that* $\psi(t)\geqslant\widehat{\psi}(t)$ *for all* $t>0$, $\lambda:=\frac{1}{4}$ *and* $C:=2$; *if* $x^\dagger\in\mathrm{range}(A^*)$, *we can set* $\psi(t):=t$, $\lambda:=\frac{1}{2}$ *and* $C:=R_0$ *with* $R_0>0$ *satisfying* $d^{A^*}_{x^\dagger}(R_0)=0$.

Proof. It is readily checked that

$$\frac{1}{2}\|x-x^\dagger\|^2_{\ell^2}=\frac{1}{2}\|x\|^2_{\ell^2}-\frac{1}{2}\|x^\dagger\|^2_{\ell^2}-\langle x^\dagger,x-x^\dagger\rangle_{\ell^2}. \tag{5.4}$$

Now we separate two cases. In the first case, characterized by $x^\dagger\notin\mathrm{range}(A^*)$, we decompose for arbitrary fixed $R>0$ the element $x^\dagger$ as $x^\dagger=A^*v_R+u_R$, where $\|v_R\|_Y=R$ and $\|u_R\|_{\ell^2}=d^{A^*}_{x^\dagger}(R)>0$. It is known (see [5, pp 377–8]) that the infimum in (2.7) is a minimum in this case and for all $R>0$, and that such elements $v_R\in Y$ and $u_R\in\ell^2$ always exist. Using this fact, we can conclude that

$$-\langle x^\dagger,x-x^\dagger\rangle_{\ell^2}=\langle A^*v_R,x-x^\dagger\rangle_{\ell^2}+\langle u_R,x-x^\dagger\rangle_{\ell^2}=\langle v_R,A(x-x^\dagger)\rangle_Y+\langle u_R,x-x^\dagger\rangle_{\ell^2},$$

which yields the estimate

$$-\langle x^\dagger,x-x^\dagger\rangle_{\ell^2}\leqslant R\|A(x-x^\dagger)\|_Y+d^{A^*}_{x^\dagger}(R)\|x-x^\dagger\|_{\ell^2}$$

for the third term in the right-hand side of the identity (5.4). Hence, we may employ for $x^\dagger\in\ell^2\setminus\mathrm{range}(A^*)$ the auxiliary function Φ, defined by (5.1). Thanks to the limit conditions (5.2), we can choose $R:=\Phi^{-1}(\|A(x-x^\dagger)\|_Y)$. Then we obtain upon Young's inequality that

$$\begin{aligned}-\langle x^\dagger,x-x^\dagger\rangle_{\ell^2}&\leqslant R\|A(x-x^\dagger)\|_Y+[d^{A^*}_{x^\dagger}(R)]^2+\frac{\|x-x^\dagger\|^2_{\ell^2}}{4}\\&=2\,[d^{A^*}_{x^\dagger}(\Phi^{-1}(\|A(x-x^\dagger)\|_Y))]^2+\frac{\|x-x^\dagger\|^2_{\ell^2}}{4}.\end{aligned}$$

Consequently, recalling (5.4) we know that for $x^\dagger\in\ell^2\setminus\mathrm{range}(A^*)$ the variational inequality

$$0\leqslant\frac{1}{4}\|x-x^\dagger\|^2_{\ell^2}\leqslant\frac{1}{2}\|x\|^2_{\ell^2}-\frac{1}{2}\|x^\dagger\|^2_{\ell^2}+2\,[d^{A^*}_{x^\dagger}(\Phi^{-1}(\|A(x-x^\dagger)\|_Y))]^2 \tag{5.5}$$

is valid for all $x\in\ell^2$. Recalling the definition of $\widehat{\psi}$ and ψ, we know that the last term in the right hand-side of the inequality (5.5) is exactly $2\widehat{\psi}(\|A(x-x^\dagger)\|_Y)$, and this inequality holds still true if this term is replaced by $2\psi(\|A(x-x^\dagger)\|_Y)$.

In the second case, where a source condition $x^\dagger \in \text{range}(A^*)$ is satisfied, we know that there exists some $R_0 > 0$ such that $d^{A^*}_{x^\dagger}(R_0) = 0$, and we can simply estimate the third term in the right-hand side of (5.4) as

$$-\langle x^\dagger, x - x^\dagger\rangle_{\ell^2} \leqslant R_0 \|A(x - x^\dagger)\|_Y .$$

Then the variational inequality attains the simpler form

$$0 \leqslant \frac{1}{2}\|x - x^\dagger\|^2_{\ell^2} \leqslant \frac{1}{2}\|x\|^2_{\ell^2} - \frac{1}{2}\|x^\dagger\|^2_{\ell^2} + R_0 \, \|A(x - x^\dagger)\|_Y , \tag{5.6}$$

thus we can set $\psi(t) := t$. □

Theorem 5.2. *Let $E_\eta(x, x^\dagger) := \eta \, \|x - x^\dagger\|_{\ell^1} + \frac{1}{4}\|x - x^\dagger\|^2_{\ell^2}$ be an error functional, and g_η be a concave index function given by*

$$g_\eta(t) := 2\eta \, \varphi(t) + K\psi(t), \quad t > 0, \tag{5.7}$$

where φ is from (4.1) and ψ is defined in lemma 5.1, and the constants $K = 2$ for $x^\dagger \in \ell^1 \setminus \text{range}(A^*)$ and $K = R_0$ for $x^\dagger \in \ell^1 \cap \text{range}(A^*)$ satisfying $d^*_{x^\dagger}(R_0) = 0$. Then the following variational inequality

$$E_\eta(x, x^\dagger) \leqslant \mathcal{R}_\eta(x) - \mathcal{R}_\eta(x^\dagger) + g_\eta(\|A(x - x^\dagger)\|_Y) \qquad \forall \; x \in \ell^1 \tag{5.8}$$

holds for the elastic-net regularization (1.5)–(1.6) and an arbitrary weight parameter $\eta > 0$.

Moreover, as an immediate consequence of the inequality (5.8), the convergence rate

$$E_\eta(x^\delta_{\beta_*,\eta}, x^\dagger) = \mathcal{O}(g_\eta(\delta)) \quad \text{as} \quad \delta \to 0 \tag{5.9}$$

holds for the elastic-net regularized solutions $x^\delta_{\beta_*,\eta}$ when the regularization parameter β_* is chosen from TDP, SDP or LEP. Thus the following alternative rate estimate follows

$$\|x^\delta_{\beta_*,\eta} - x^\dagger\|_{\ell^2} \leqslant \|x^\delta_{\beta_*,\eta} - x^\dagger\|_{\ell^1} = \mathcal{O}(\varphi(\delta) + \psi(\delta)) \quad \text{as} \quad \delta \to 0. \tag{5.10}$$

Proof. Taking into account (5.5) and (5.6), we can deduce from (4.2) the variational inequality

$$E_\eta(x, x^\dagger) \leqslant \mathcal{R}_\eta(x) - \mathcal{R}_\eta(x^\dagger) + [2\,\eta\,\varphi + K\,\psi]\,(\|A(x - x^\dagger)\|_Y) \tag{5.11}$$

of type (3.4) for all $x \in \ell^1$, with the constants $K = 2$ for $x^\dagger \in \ell^1 \setminus \text{range}(A^*)$ and $K = R_0$ for $x^\dagger \in \ell^1 \cap \text{range}(A^*)$ satisfying $d^*_{x^\dagger}(R_0) = 0$. We also note that the error functional E_η satisfies the conditions (3.7) and (3.8) with $C_E = 2$, and that the function $g_\eta(t)$ is a concave index function.

Then we can apply proposition 3.2 with $\lambda = 1$, $E = E_\eta$, $g = g_\eta$ and $C = 1$ for all three *a posteriori* parameter choices of β under consideration to obtain the desired results. □

Example 5.3. We discuss now the convergence rate in (5.10) under the assumptions that $x^\dagger$ satisfies the Hölder source condition with exponent $\theta > 0$, implying a power-type decay of the distance function (2.9), and that the power-type decay of solution components and power-type growth of the $f^{(k)}$-norms (4.4) hold. Note that $x^\dagger \in \ell^0$ implies $x^\dagger \in \text{range}(A^*)$, which is a direct consequence of assumption 2.1 (c). Then (4.3) and (4.7) show that $\varphi(\delta) \sim \delta^{\frac{\mu}{\mu+\nu}}$ if $x^\dagger \in \ell^1 \setminus \ell^0$ and

$\varphi(\delta) \sim \delta$ if $x^\dagger \in \ell^0$. On the other hand, the concave index function $\widehat{\psi}(\delta) = \psi(\delta) \sim \delta^{\frac{2\theta}{\theta+1}}$ can be seen from formula (5.3) if $x^\dagger \in \ell^1 \setminus \mathrm{range}(A^*)$, while it occurs that $\psi(\delta) \sim \delta$ if $x^\dagger \in \ell^1 \cap \mathrm{range}(A^*)$. In summary, we have by theorem 5.2 the Hölder convergence rate

$$\|x^\delta_{\beta_*,\eta} - x^\dagger\|_{\ell^1} = \mathcal{O}(\delta^\kappa), \tag{5.12}$$

where κ is given by

$$\kappa = \begin{cases} 1 & \text{if} \quad x^\dagger \in \ell^0, \\ \dfrac{\mu}{\mu+\nu} & \text{if} \quad x^\dagger \in (\ell^1 \cap \mathrm{range}(A^*)) \setminus \ell^0, \\ \min\left(\dfrac{\mu}{\mu+\nu}, \dfrac{2\theta}{\theta+1}\right) & \text{if} \quad x^\dagger \in \ell^1 \setminus \mathrm{range}(A^*). \end{cases} \tag{5.13}$$

In the case $x^\dagger \in \ell^1 \setminus \ell^0$, we observe the behaviour of the exponent (5.13) for the Hölder convergence rate (5.12) that fast convergence rates occur only for almost sparse solutions. There is a trade-off here in the sense that the closer the exponent κ is to one the more drastic must be the decay of the components $x^\dagger_k$ of the solution $x^\dagger$ if k tends to infinity. More precisely, for κ close to one, the decay exponent μ from (4.4) for the solution components has to be sufficiently large and the exponent θ of the range-type source condition occurring in (2.9) has to be sufficiently close to one.

Remark 5.4. Taking into account the estimates from [25], we can distinguish upper bounds of $E_\eta(x^\delta_{\beta_*,\eta}, x^\dagger)$ in theorem 5.2 for $\beta_* = \beta_{\mathrm{TDP}}$ and $\beta_* = \beta_{\mathrm{SDP}}$, which, however, yield the same convergence rate (5.9). To be more precise, we find for sufficiently small $\delta > 0$ that

$$\eta\|x^\delta_{\beta_*,\eta} - x^\dagger\|_{\ell^1} + \frac{1}{4}\|x^\delta_{\beta_*,\eta} - x^\dagger\|^2_{\ell^2} \leqslant C_* \left(2\eta\, \varphi(\delta) + K\psi(\delta)\right) \tag{5.14}$$

holds with

$$C_* = \begin{cases} \tau_2 + 1 & \text{for} \quad \beta_* = \beta_{\mathrm{TDP}}, \\ (\tau+1)\max\left\{\dfrac{2(\tau^2+1)}{q(\tau-1)^2(\tau+1)}, 1\right\} & \text{for} \quad \beta_* = \beta_{\mathrm{SDP}}. \end{cases} \tag{5.15}$$

From (5.14) we derive the ℓ^1-norm estimate

$$\|x^\delta_{\beta_*,\eta} - x^\dagger\|_{\ell^1} \leqslant 2C_*\, \varphi(\delta) + \frac{C_*}{\eta} K\psi(\delta). \tag{5.16}$$

In contrast to the approach for (5.10), we may derive directly from (5.14) an ℓ^2-norm estimate of the form

$$\|x^\delta_{\beta_*,\eta} - x^\dagger\|_{\ell^2} \leqslant 2\sqrt{C_* \left(2\eta\, \varphi(\delta) + K\psi(\delta)\right)} \tag{5.17}$$

with a lower (square root) rate as $\delta \to 0$. On the other hand, it follows from the formula (3.10) that the estimate (5.14) holds true for $\beta_* = \beta_{\mathrm{LEP}}$ with constant $C_* = 34/q$.

Remark 5.5. By introducing the weight parameter $\eta > 0$ in section 1, the natural two-parameter regularization (1.4) of the elastic-net approach reduces to the one-parameter

regularization (1.5). If the weight η is fixed for all $\delta > 0$, then the convergence rate in (5.10), also the Hölder rate expressed by the exponents κ in Example 5.3, is the same for all $0 < \eta < \infty$, but the upper bounds on the right-hand side of (5.16) and (5.17) depend on η. A very illustrative situation occurs when we consider TDP with $\tau := \tau_1 = \tau_2$. At least for sufficiently small $\delta > 0$, the regularization parameter $\beta_*(\eta)$ is well-defined for all $\eta > 0$ and the pairs $(\beta_*(\eta), \eta)$ form a 'discrepancy curve' with $\|Ax^{\delta}_{\beta_*(\eta),\eta} - y^{\delta}\|_Y = \tau\delta$ and the uniform convergence rates of all associated regularized solutions. Then we can select one pair from the curve with the goal to implement additional solution features; see more discussions in [31, p 166].

6. Conclusions

In this work we have derived some variational inequalities for both ℓ^1- and elastic-net regularizations. Then we have applied these variational inequalities to obtain some explicit convergence rates, and compared the results with the ones from the classical source conditions. This increases significantly the range of the regularized solutions, for which the convergence rates can be achieved. Three different principles of *a posteriori* parameter choices are also discussed, and their influences on convergence rates are analyzed. The basic principles, analysis tools, and the selection strategies for the choice of regularization parameters can be equally applied to general multi-parameter Tikhonov-type regularizations.

Acknowledgments

The authors would like to thank the anonymous referees for their many insightful and constructive suggestions and comments, which have helped us to improve the presentation and the results of the paper significantly.

References

[1] Anzengruber S W, Hofmann B and Mathé P 2014 Regularization properties of the discrepancy principle for Tikhonov regularization in Banach spaces *Appl. Anal.* **93** 1382–400
[2] Anzengruber S W, Hofmann B and Ramlau R 2013 On the interplay of basis smoothness and specific range conditions occurring in sparsity regularization *Inverse Problems* **29** 125002
[3] Anzengruber S W and Ramlau R 2010 Morozov's discrepancy principle for Tikhonov-type functionals with nonlinear operators *Inverse Problems* **26** 025001
[4] Bakushinsky A B and Kokurin M Y 2004 *Iterative Methods for Approximate Solution of Inverse Problems* (Dordrecht: Springer)
[5] Boţ R I and Hofmann B 2010 An extension of the variational inequality approach for obtaining convergence rates in regularization of nonlinear ill-posed problems *J. Integral Equ. Appl.* **22** 369–92
[6] Boţ R I and Hofmann B 2013 The impact of a curious type of smoothness conditions on convergence rates in ℓ^1-regularization *Eurasian J. Math. Comput. Appl.* **1** 29–40
[7] Bredies K and Lorenz D A 2009 Regularization with non-convex separable constraints *Inverse Problems* **25** 085011
[8] Burger M, Flemming J and Hofmann B 2013 Convergence rates in ℓ^1-regularization if the sparsity assumption fails *Inverse Problems* **29** 025013
[9] Burger M and Osher S 2004 Convergence rates of convex variational regularization *Inverse Problems* **20** 1411–21
[10] Candés E J and Tao T C 2016 Near optimal singal recovery from random projections: universial encoding strategies *IEEE Trans. Inf. Theory* **52** 5406–25

[11] De Mol C, De Vito E and Rosasco L 2009 Elastic-net regularization in learning theory *J. Complexity* **25** 201–30
[12] Düvelmeyer D, Hofmann B and Yamamoto M 2007 Range inclusions and approximate source conditions with general benchmark functions *Numer. Funct. Anal. Optim.* **28** 1245–61
[13] Engl H W, Hanke M and Neubauer A 1996 *Regularization of Inverse Problems* (Dordrecht: Kluwer)
[14] Figueiredo M A T and Neubauer A 2003 An EM algorithm for wavelet-based image restoration *IEEE Trans. Image Process* **12** 906–16
[15] Flemming J 2012 *Generalized Tikhonov Regularization and Modern Convergence Rate Theory in Banach Spaces* (Aachen: Shaker Verlag)
[16] Flemming J 2012 Solution smoothness of ill-posed equations in Hilbert spaces: four concepts and their cross connections *Appl. Anal.* **91** 1029–44
[17] Flemming J, Hofmann B and Veselić I 2016 A unified approach to convergence rates for ℓ^1-regularization and lacking sparsity *J. Inverse Ill-Posed Probl.* **24** 139–48
[18] Grasmair M 2010 Generalized Bregman distances and convergence rates for non-convex regularization methods *Inverse Problems* **26** 115014
[19] Grasmair M, Haltmeier M and Scherzer O 2008 Sparse regularization with ℓ^q penalty term *Inverse Problems* **24** 055020
[20] Groetsch C W 1984 *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind* (Boston, MA: Pitman)
[21] Hein T and Hofmann B 2009 Approximate source conditions for nonlinear ill-posed problems—chances and limitations *Inverse Problems* **25** 035003
[22] Hofmann B 2006 Approximate source conditions in Tikhonov–Phillips regularization and consequences for inverse problems with multiplication operators *Math. Methods Appl. Sci.* **29** 351–71
[23] Hofmann B, Kaltenbacher B, Pöschl C and Scherzer O 2007 A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators *Inverse Problems* **23** 987–1010
[24] Hofmann B and Mathé P 2007 Analysis of profile functions for general linear regularization methods *SIAM J. Numer. Anal.* **45** 1122–41
[25] Hofmann B and Mathé P 2012 Parameter choice in Banach space regularization under variational inequalities *Inverse Problems* **28** 104006
[26] Hurri J, Hyvärinen A and Oja E 1997 Wavelets and natural image statistics *Proc. of the 10th Scandinavian Conf. on Image Analysis (Lappenranta, Finland)* vol 1, ed M Frydrych *et al* (Lappenranta: Pattern Recognition Society of Finland) pp 13–8
[27] Ito K, Jin B and Takeuchi T 2011 Multi-parameter Tikhonov regularization *Methods Appl. Anal.* **18** 31–46
[28] Jiang D, Feng H and Zou J 2012 Convergence rates of Tikhonov regularizations for parameter identification in a parabolic–elliptic system *Inverse Problems* **28** 104002
[29] Jin B, Lorenz D and Schiffler S 2009 Elastic-net regularization: error estimates and active set methods *Inverse Problems* **25** 115022
[30] Lorenz D A 2008 Convergence rates and source conditions for Tikhonov regularization with sparsity constraints *J. Inverse Ill-Posed Probl.* **16** 463–78
[31] Lu S and Pereverzev S V 2013 *Regularization Theory for Ill-Posed Problems* (*Inverse and Ill-Posed Problems Series* vol 58) (Berlin: Walter de Gruyter)
[32] Mathé P 2006 The Lepski principle *Inverse Problems* **22** 11–5
[33] Mathé P and Hofmann B 2008 How general are general source conditions? *Inverse Problems* **24** 015009
[34] Ramlau R and Resmerita E 2010 Convergence rates for regularization with sparsity constraints *Electron. Trans. Numer. Anal.* **37** 87–104
[35] Scherzer O, Grasmair M, Grossauer H, Haltmeier M and Lenzen F 2009 *Variational Methods in Imaging* (*Applied Mathematical Sciences* vol 167) (New York: Springer)
[36] Schuster T, Kaltenbacher B, Hofmann B and Kazimierski K S 2012 *Regularization Methods in Banach Spaces* (*Radon Series on Computational and Applied Mathematics* vol 10) (Berlin: Walter de Gruyter)
[37] Taylor H, Bank S and McCoy J 1979 Deconvolution with the ℓ_1 norm *Geophysics* **44** 39–52
[38] Tibshirani R 1996 Regression shrinkage and selection via the lasso *J. R. Stat. Soc.* **58** 267–88
[39] Wang W, Lu S, Mao H and Cheng J 2013 Multi-parameter Tikhonov regularization with the ℓ^0 sparsity constraint *Inverse Problems* **29** 065018

[40] Wilson R, Calway A D and Pearson E R S 1992 A generalized wavelet transform for Fourier analysis: the multiresolution Fourier transform and its application to image and audio signal analysis *IEEE Trans. Inf. Theory* **38** 674–90
[41] Yosida K 1995 *Functional Analysis* (*Classics in Mathematics*) (Berlin: Springer) (reprint of the 6th edn of 1980)
[42] Zarzer C A 2009 On Tikhonov regularization with non-convex sparsity constraints *Inverse Problems* **25** 025006
[43] Zou H and Hastie T 2005 Regularization and variable selection via the elastic net *J. R. Stat. Soc.* B **67** 301–20