

TIKHONOV REGULARIZATION WITH ℓ^0 -TERM COMPLEMENTING A CONVEX PENALTY: ℓ^1 -CONVERGENCE UNDER SPARSITY CONSTRAINTS

WEI WANG, SHUAI LU, BERND HOFMANN, AND JIN CHENG

ABSTRACT. Measuring the error by an ℓ^1 -norm, we analyze under sparsity assumptions an ℓ^0 -regularization approach, where the penalty in the Tikhonov functional is complemented by a general stabilizing convex functional. In this context, ill-posed operator equations $Ax = y$ with an injective and bounded linear operator A mapping between ℓ^2 and a Banach space Y are regularized. For sparse solutions, error estimates as well as linear and sublinear convergence rates are derived based on a variational inequality approach, where the regularization parameter can be chosen either a priori in an appropriate way or a posteriori by the sequential discrepancy principle. To further illustrate the balance between the ℓ^0 -term and the complementing convex penalty, the important special case of the ℓ^2 -norm square penalty is investigated showing explicit dependence between both terms. Finally, some numerical experiments verify and illustrate the sparsity promoting properties of corresponding regularized solutions.

1. INTRODUCTION

Variational sparsity regularization has gained significant attention in the past years, because of its wide field of applications in mathematical methods for imaging, natural sciences and finance. We refer for details to the corresponding chapters in [34, 35, 36] and diverse research papers like [3, 4, 5, 6, 27, 28, 30, 31, 32]. There are many papers on ℓ^1 -regularization of ill-posed operator equations under sparsity constraints using the ℓ^1 -norm as penalty term in a Tikhonov regularization approach (see [7] and references therein), but the variety of sparsity promoting penalties is much broader. In the elastic-net technique (cf., e.g., [9, 26, 38]) a multi-parameter version is implemented which uses a linear combination of ℓ^1 -norm and ℓ^2 -norm square. On the other hand, ℓ^p seminorms with $p < 1$ strongly support the sparsity selection of approximate solutions in variational regularization, but the penalties are no longer convex for such p . Moreover, the extremal case of using an ℓ^0 -term alone fails, because the required stabilizing effect of the penalty for overcoming the ill-posedness is completely lost in that case. Therefore, in [37] a linear combination of the ℓ^0 -term with the stabilizing ℓ^2 -norm square was suggested, which extends the idea of elastic-net regularization. Affirmative analytical and numerical results presented in [37] showed the utility of such an approach. For obtaining convergence rates a replacement of the common use of adapted source conditions in Banach space regularization (cf. [8, 33]) by the variational inequality approach [21, 12, 13, 24] was successful in this context. Unfortunately, such kind of regularization including an ℓ^0 -term is often connected with an enormous computational effort. Nevertheless, it is worthwhile to consider the error analysis of this approach in all facets, and we complement the analysis in this paper by deriving ℓ^1 -norm error results under the extended situation that general convex functionals are added to the ℓ^0 -term in the penalty.

Let $\tilde{A} \in \mathcal{L}(\tilde{X}, Y)$ be an *injective* and bounded linear operator mapping between the infinite dimensional separable Hilbert space \tilde{X} with inner product $\langle \cdot, \cdot \rangle_{\tilde{X}}$ and norm $\| \cdot \|_{\tilde{X}}$ and the infinite dimensional Banach space Y with norm $\| \cdot \|$ possessing the dual Y^* with norm $\| \cdot \|_{Y^*}$. The symbol $\langle \cdot, \cdot \rangle_{Y^* \times Y}$ denotes the corresponding dual pairing between Y and its dual. We assume that the range of \tilde{A} is not closed in Y , i.e. $\text{Range}(\tilde{A}) \neq \overline{\text{Range}(\tilde{A})}^Y$, which is equivalent to the fact that the Moore-Penrose inverse $\tilde{A}^\dagger : \text{Range}(\tilde{A}) \oplus \text{Range}(\tilde{A})^\perp \subset Y \rightarrow \tilde{X}$ is an unbounded linear operator. Hence the operator equation

$$(1) \quad \tilde{A}x = y, \quad x \in \tilde{X}, \quad y \in Y,$$

expressing the model of a linear inverse problem with the forward operator \tilde{A} , is ill-posed. Consequently, for noisy data $y^\delta \in Y$ replacing $y \in \text{Range}(\tilde{A})$ in (1) and satisfying the inequality

$$(2) \quad \|y - y^\delta\| \leq \delta,$$

with given noise level $\delta > 0$, solutions need not exist. When they exist the solution may be strongly perturbed even if δ is small.

For a prescribed orthonormal basis $\{u_i\}_{i=1}^\infty$ in \tilde{X} with elements $\tilde{x} = \sum_{i=1}^\infty x_i u_i$ we consider the infinite sequences $x = \{x_i\}_{i=1}^\infty$ collecting the Fourier coefficients $x_i := \langle \tilde{x}, u_i \rangle_{\tilde{X}}$. Our focus is on *sparse solutions* x^\dagger to equation (1) in the sense that $x^\dagger = \{x_i^\dagger\}_{i=1}^\infty := \{\langle \tilde{x}^\dagger, u_i \rangle_{\tilde{X}}\}_{i=1}^\infty \in \ell^0$, where as usual the symbol ℓ^0 denotes the set of infinite sequences with a finite number of nonzero components and we set in this context $\|x\|_{\ell^0} := \sum_{i=1}^\infty \text{sgn}(|x_i|)$, for

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

Because the sparsity can be lost when the basis changes, the choice of the basis $\{u_i\}_{i=1}^\infty$ must be done appropriately. Then we can consider instead of (1) the ill-posed equation

$$(3) \quad Ax = y, \quad x \in X := \ell^2, \quad y \in Y,$$

with uniquely determined solution x^\dagger , by exploiting the unitary synthesis operator $L : \ell^2 \rightarrow \tilde{X}$ defined as $\tilde{x} = Lx$ and the linear bounded and injective composition operator $A := \tilde{A} \circ L$ with non-closed range, i.e. $\text{Range}(A) \neq \overline{\text{Range}(A)}^Y$. For sequences $x = \{x_i\}_{i=1}^\infty$ we will denote by $\|x\|_{\ell^p} := \left(\sum_{i=1}^\infty |x_i|^p \right)^{1/p}$ the norms in the Banach spaces ℓ^p , $1 \leq p < \infty$, and by $\|x\|_{\ell^\infty} := \sup_{k \in \mathbb{N}} |x_k|$ the norm in ℓ^∞ . For simplicity we use the symbol $\langle \cdot, \cdot \rangle$ without subscript for the inner product in ℓ^2 .

To find stable approximate solutions to equation (3) based on noisy data $y^\delta \in Y$ we use the Tikhonov type regularization

$$(4) \quad \Phi_\alpha^\delta(x) := \frac{1}{q} \|Ax - y^\delta\|^q + \alpha \Omega(x) \rightarrow \min, \quad \text{subject to } x \in X = \ell^2,$$

with a regularization parameter $\alpha > 0$ and minimizers denoted by x_α^δ , for a penalty functional Ω of the form

$$(5) \quad \Omega(x) := \|x\|_{\ell^0} + \mathcal{R}(x),$$

where a proper, non-negative, *convex* and *stabilizing* functional \mathcal{R} complements the ℓ^0 -term in the penalty. For the exponent q in the misfit term we restrict to the interval $1 < q < \infty$ and exclude the case $q = 1$ in order to avoid exact penalization (cf. [1]). We note that Ω from (5) is a non-convex functional due to the non-convexity of the contained term $\|\cdot\|_{\ell^0}$. For extended discussion on non-convex minimization we refer to [16, 17].

As an important special case we have in mind $\mathcal{R}(x) := \frac{\eta}{2}\|x\|_{\ell^2}^2$, for a prescribed constant $\eta > 0$, such that

$$(6) \quad \Omega(x) := \|x\|_{\ell^0} + \frac{\eta}{2}\|x\|_{\ell^2}^2.$$

Regularized solutions for that penalty have been comprehensively discussed in [37]. It has been shown *ibid* that, for every $\eta > 0$, the functional Ω from (6) is *stabilizing* and *lower semi-continuous* on sublevel sets of ℓ^0 which belong to a fixed ball in X . Consequently, the minimizers $x_\alpha^\delta \in \ell^0 \subset X$ of (4) exist for all $\alpha > 0$ and are stable with respect to further perturbations of the data y^δ . Measuring the error by an ℓ^2 -norm, assertions on variational source conditions and convergence including rates of the regularized solutions x_α^δ to the exact solution $x^\dagger \in \ell^0$ have also been given in [37].

The goal of this paper is to complement the results presented in [37] from ℓ^2 -norm errors to ℓ^1 -norm errors in the case (6) and to extend assertions to penalties (5) with general convex functionals \mathcal{R} . For sparse solutions $x^\dagger \in \ell^0$ and wide classes of functionals \mathcal{R} we will verify the linear convergence rate

$$(7) \quad \|x_\alpha^\delta - x^\dagger\|_{\ell^1} = \mathcal{O}(\delta) \quad \text{as } \delta \rightarrow 0,$$

provided that the regularization parameters $\alpha > 0$ are chosen appropriately. Precisely, error estimates and linear convergence rates are derived based on a variational source condition approach, where the regularization parameter can be chosen either a priori in an appropriate way or a posteriori by the sequential discrepancy principle. If the convex functional \mathcal{R} is chosen inappropriate, then at least sublinear convergence rates can be proven based on the method of approximate source conditions [19, 20].

The paper is organized as follows: In Section 2 we collect basic assumptions and provide a variational inequality with respect to the ℓ^1 -norm which does not appear in the penalty functional $\Omega(x)$. Error estimates for an *a priori* parameter choice rule and the sequential discrepancy principles are carried out in Section 3 for general stabilizing functionals $\mathcal{R}(x)$. To further illustrate the balance between the ℓ^0 -term and the complementing convex penalty, the important special case of the ℓ^2 -norm square penalty is investigated in Section 4 showing explicit dependence between both penalty terms. Finally some numerical examples and extended discussion are presented in Section 5 to verify the sparsity promoting properties of our proposed regularization schemes and other related aspects.

2. A VARIATIONAL SOURCE CONDITION APPROACH

We first present the main assumptions below.

Assumption 1.

- (a) Let $A : X \rightarrow Y$ in equation (3) be an injective and bounded linear operator mapping from the Hilbert space X to the Banach space Y with an

adjoint operator $A^* : Y^* \rightarrow X$. Moreover, let $\text{Range}(A) \neq \overline{\text{Range}(A)}^Y$ such that the equation (3) is ill-posed.

- (b) Let $y \in \text{Range}(A)$ be given such that the uniquely determined solution x^\dagger to equation (3) is sparse, i.e. $x^\dagger \in \ell^0$. Moreover, let

$$I := \{i \in \mathbb{N} : x_i^\dagger \neq 0\} \neq \emptyset$$

be the finite index set of non-zero components of x^\dagger . In this context, we set $|I| := \text{card}(I) = \|x^\dagger\|_{\ell^0} > 0$ and denote by $\Sigma(x^\dagger) := \text{span}(\{e^{(i)}\}_{i \in I})$ the corresponding $|I|$ -dimensional subspace of non-zero components, where $e^{(i)} := (0, 0, \dots, 0, 1, 0, \dots)$, with 1 at the i -th position for $i \in \mathbb{N}$, designate the elements of standard orthonormal basis in ℓ^2 .

- (c) Let $\mathcal{R} : X = \ell^2 \rightarrow [0, \infty]$ be a convex and lower semi-continuous functional, where we denote by $\mathcal{D}(\mathcal{R}) = \{x \in X : \mathcal{R}(x) < \infty\}$ the proper domain of \mathcal{R} and by $\partial\mathcal{R}(x) \subseteq X$ the subdifferential of the functional \mathcal{R} at the point $x \in \mathcal{D}(\mathcal{R})$.
- (d) The functional \mathcal{R} is assumed to be stabilizing in the sense that the sublevel sets

$$\mathcal{S}^{\mathcal{R}}(c) := \{x \in X : \mathcal{R}(x) \leq c\}$$

are weakly sequentially pre-compact in X for all $c \geq 0$. This means that every infinite sequence $\{x_n\}_{n=1}^\infty \subset \mathcal{S}^{\mathcal{R}}(c)$ possesses a subsequence which has a weak limit in X .

Remark 1. In the Hilbert space $X = \ell^2$ item (d) means that the functional \mathcal{R} is weakly coercive, i.e. $\|x\|_{\ell^2} \rightarrow \infty$ implies $\mathcal{R}(x) \rightarrow \infty$. For given $c \geq 0$ there is a radius $r(c) > 0$ such that $\sup_{x \in \mathcal{S}^{\mathcal{R}}(c)} \|x\|_{\ell^2} \leq r(c)$. \square

Assumption 2. For all $i \in \mathbb{N}$ we assume that there exist $\omega_i \in Y^*$ such that $e^{(i)} = A^*\omega_i$. Consequently, there holds $x_i = \langle e^{(i)}, x \rangle = \langle \omega_i, Ax \rangle_{Y^* \times Y}$ for all $x = (x_1, x_2, \dots) \in \ell^2$.

Remark 2. Assumption 2 was already suggested in [5, 15, 28]. In a recent paper [2] it was shown that the set of range conditions $e^{(i)} \in \text{Range}(A^*)$, $i \in \mathbb{N}$, which is equivalent to $u_i \in \text{Range}(\tilde{A}^*)$ for all elements of the orthonormal basis $\{u_i\}_{i=1}^\infty$ in \tilde{X} , applies for important classes of linear inverse problems. Precisely, Assumption 2 can be reinterpreted as a requirement on the ‘smoothness’ of the basis elements u_i with respect to the forward operator \tilde{A} , and we refer to [14] for consequence in the case of ‘non-smooth’ basis elements.

Under Assumption 2 we have for all $i \in I$

$$|x_i - x_i^\dagger| = |\langle e^{(i)}, x - x^\dagger \rangle| = |\langle \omega_i, A(x - x^\dagger) \rangle_{Y^* \times Y}| \leq \|\omega_i\|_{Y^*} \|A(x - x^\dagger)\|,$$

and hence

$$(8) \quad \sum_{i \in I} |x_i - x_i^\dagger| \leq K \|A(x - x^\dagger)\| \quad \text{for all } x \in \ell^2$$

with a constant $K = \sum_{i \in I} \|\omega_i\|_{Y^*}$. \square

Remark 3. If Assumption 2 fails, then the term $\sum_{i \in I} |x_i^\dagger - x_i|$ can be, as an alternative to (8), estimated from above by taking into the account that for any injective $A : \ell^2 \rightarrow Y$, which maps also continuously from ℓ^1 to Y , the modulus of injectivity (cf., e.g., [23, §3.2] and [29, §2.4]) $j(A, \Sigma(x^\dagger))$ defined as

$$j(A, \Sigma(x^\dagger)) := \inf_{0 \neq x \in \Sigma(x^\dagger)} \frac{\|Ax\|}{\|x\|_{\ell^1}} > 0$$

is a finite positive number. This is similar to ideas of the proofs of Theorems 14 and 15 in [18]. Precisely, we can estimate

$$\sum_{i \in I} |x_i^\dagger - x_i| \leq \frac{1}{j(A, \Sigma(x^\dagger))} \|A(P_I x - x^\dagger)\| \quad \text{for all } x \in \ell^1,$$

where P_I denotes the projection to the subspace $\Sigma(x^\dagger)$. Thus we obtain, for all $x \in \ell^1$,

$$\begin{aligned} \sum_{i \in I} |x_i^\dagger - x_i| &\leq \frac{1}{j(A, \Sigma(x^\dagger))} (\|A(x - x^\dagger)\| + \|A(I - P_I)x\|) \\ &\leq \frac{1}{j(A, \Sigma(x^\dagger))} (\|A(x - x^\dagger)\| + \|A\|_{\mathcal{L}(\ell^1, Y)} \|(I - P_I)x\|_{\ell^1}) \\ &\leq c_1 \|A(x - x^\dagger)\| + c_2 \sum_{i \notin I} |x_i| \end{aligned}$$

with positive constants c_1 and c_2 . As we will see, however, handling of the term $\sum_{i \notin I} |x_i|$ becomes a serious difficulty in the further estimation process. Hence, rates results are still missing when Assumption 2 is violated. \square

Assumption 3. Let $x^\dagger \in \mathcal{D}(\mathcal{R})$ and $\partial\mathcal{R}(x^\dagger) \neq \emptyset$. Assume that there exists some element $\xi^\dagger \in \partial\mathcal{R}(x^\dagger)$ satisfying a source condition

$$(9) \quad \xi^\dagger = A^*w, \quad \text{for some } w \in Y^*.$$

Remark 4. If we have, for some $\xi^\dagger \in \partial\mathcal{R}(x^\dagger)$, sparsity in the sense that $\xi^\dagger \in \ell^0$, then Assumption 2 implies that (9) and hence Assumption 3 hold, because with $x^\dagger \in \ell^0$

$$\xi^\dagger = \sum_{k \in \mathbb{N}: \xi_k^\dagger \neq 0} \xi_k^\dagger e^{(k)} = A^* \left(\sum_{k \in \mathbb{N}: \xi_k^\dagger \neq 0} \xi_k^\dagger \omega_k \right).$$

This is in particular under Assumption 1 the case whenever the implication

$$(10) \quad x^\dagger \in \ell^0 \implies \xi^\dagger \in \ell^0$$

holds true. \square

We provide two examples when Assumption 3 is satisfied or violated.

Example 1. Consider the family of functionals

$$\mathcal{R}(x) := \frac{\eta}{p} \|x\|_{\ell^p}^p, \quad \eta > 0,$$

with exponents $1 < p \leq 2$, where the subdifferential $\partial\mathcal{R}(x^\dagger)$ is always a singleton such that

$$\xi_k^\dagger = \eta |x_k^\dagger|^{p-1} \operatorname{sgn}(x_k^\dagger), \quad k \in \mathbb{N}.$$

For such exponents p the functionals \mathcal{R} are convex, lower semi-continuous and stabilizing in the sense of item (d) of Assumption 1. Consequently, Assumption 3 is applicable, because $x^\dagger \in \ell^0$ implies $\xi^\dagger \in \ell^0$ and hence (9) is satisfied. Note that therefore for the penalty functional

$$\Omega(x) := \|x\|_{\ell^0} + \frac{\eta}{p} \|x\|_{\ell^p}^p, \quad 1 < p \leq 2, \quad \eta > 0,$$

the main Theorem 1 below applies and provides us with linear convergence rates of the Tikhonov regularization when all other assumptions are fulfilled.

Example 2. Consider alternatively the family of functionals

$$\mathcal{R}(x) := \frac{1}{2} \|Bx\|_{\ell^2}^2$$

with a bounded linear operator $B : \ell^2 \rightarrow \ell^2$ such that $Be^{(1)} := \sum_{i=1}^{\infty} \lambda_i e^{(i)}$ with $\sum_{i=1}^{\infty} \lambda_i^2 < \infty$, $\lambda_i > 0$ ($i = 1, 2, \dots$), and $Be^{(k)} := e^{(k)}$ for all $k \geq 2$. Then we have $\xi^\dagger = B^* B x^\dagger$, where always $\xi^\dagger \notin \ell^0$, although the solution $0 \neq x^\dagger \in \ell^0$ is sparse.

The basic Lemma 1 below follows directly from Assumptions 1 – 3 and yields a variational inequality which, acting as a variational source condition, allows for convergence rates if an ℓ^0 -term complemented by a convex term forms the penalty functional of Tikhonov regularization.

Lemma 1. *Let*

$$(11) \quad \mathcal{M} := \{x \in \ell^0 : \|x\|_{\ell^2} \leq \rho\}$$

for a prescribed sufficiently large real value $\rho \geq 1$. Then under Assumptions 1, 2 and 3 there exist constants $0 < \beta \leq 1$ and $c_0 \geq 0$ depending on A , \mathcal{R} and $x^\dagger \in \ell^0$ such that the variational inequality

$$(12) \quad \beta \|x - x^\dagger\|_{\ell^1} \leq \|x\|_{\ell^0} - \|x^\dagger\|_{\ell^0} + \mathcal{R}(x) - \mathcal{R}(x^\dagger) + c_0 \|A(x - x^\dagger)\| \quad \text{for all } x \in \mathcal{M}$$

is valid.

Proof. Since $x^\dagger \in \ell^0$ and I is a finite subset of \mathbb{N} , we derive for all $x \in \ell^0$

$$\begin{aligned} & \|x\|_{\ell^0} - \|x^\dagger\|_{\ell^0} \\ &= \sum_{i \in I} \left[\operatorname{sgn}(|x_i|) - \operatorname{sgn}(|x_i^\dagger|) \right] + \sum_{i \notin I} \operatorname{sgn}(|x_i|). \end{aligned}$$

Now,

$$\begin{aligned} \|x - x^\dagger\|_{\ell^1} &= \sum_{i \in I} |x_i^\dagger - x_i| + \sum_{i \notin I} |x_i| \\ &\leq \sum_{i \in I} |x_i^\dagger - x_i| + \max_{i \notin I} |x_i| \sum_{i \notin I} \operatorname{sgn}(|x_i|), \end{aligned}$$

and because of $\max_{i \notin I} |x_i| \leq \|x\|_{\ell^\infty} \leq \|x\|_{\ell^2} \leq \rho$ we have

$$\begin{aligned} \|x - x^\dagger\|_{\ell^1} &\leq \sum_{i \in I} |x_i^\dagger - x_i| + \rho \sum_{i \notin I} \operatorname{sgn}(|x_i|) \\ &\leq \sum_{i \in I} |x_i^\dagger - x_i| + \rho \left[\|x\|_{\ell^0} - \|x^\dagger\|_{\ell^0} - \sum_{i \in I} \left[\operatorname{sgn}(|x_i|) - \operatorname{sgn}(|x_i^\dagger|) \right] \right]. \end{aligned}$$

To estimate the right-hand side of the last inequality from above we consider $J(x) := \{i \in I : |x_i| = 0\}$ and set $\underline{\kappa} := \min_{i \in I} |x_i^\dagger| > 0$. Then we can further estimate as

$$(13) \quad \sum_{i \in I} \left[\operatorname{sgn}(|x_i^\dagger|) - \operatorname{sgn}(|x_i|) \right] = \sum_{i \in J(x)} \left[\operatorname{sgn}(|x_i^\dagger|) \right] \leq \frac{1}{\underline{\kappa}} \sum_{i \in J(x)} |x_i^\dagger - x_i|.$$

If $J(x)$ is an empty set, the inequality (13) still holds true by interpreting the sum on the right-hand side as zero. We thus conclude that

$$\|x - x^\dagger\|_{\ell^1} \leq \sum_{i \in I} |x_i^\dagger - x_i| + \rho \left[\|x\|_{\ell^0} - \|x^\dagger\|_{\ell^0} + \frac{1}{\underline{\kappa}} \sum_{i \in I} |x_i^\dagger - x_i| \right].$$

As a consequence of (8) this yields for all $x \in \mathcal{M}$ the inequality

$$(14) \quad \frac{1}{\rho} \|x - x^\dagger\|_{\ell^1} \leq \|x\|_{\ell^0} - \|x^\dagger\|_{\ell^0} + K \left(\frac{1}{\rho} + \frac{1}{\underline{\kappa}} \right) \|A(x - x^\dagger)\|$$

with the constant $K = \sum_{i \in I} \|\omega_i\|_{Y^*}$.

For the convex functional \mathcal{R} and $\xi^\dagger \in \partial\mathcal{R}(x^\dagger)$ we have

$$0 \leq \mathcal{R}(x) - \mathcal{R}(x^\dagger) + \langle \xi^\dagger, x^\dagger - x \rangle$$

and hence

$$(15) \quad 0 \leq \mathcal{R}(x) - \mathcal{R}(x^\dagger) + |\langle \xi^\dagger, x - x^\dagger \rangle| \quad \text{for all } x \in X = \ell^2,$$

therefore with (9)

$$0 \leq \mathcal{R}(x) - \mathcal{R}(x^\dagger) + \|w\|_{Y^*} \|A(x - x^\dagger)\| \quad \text{for all } x \in \ell^2.$$

In combination with inequality (14) we thus obtain the variational source condition (12) with the constants $0 < \beta = \frac{1}{\rho} \leq 1$ and

$$c_0 = K \left(\frac{1}{\rho} + \frac{1}{\underline{\kappa}} \right) + \|w\|_{Y^*}.$$

This completes the proof. \square

Remark 5. From [35, Lemma 8.21] we know that for the existence of a constant $C > 0$ satisfying $|\langle \xi^\dagger, x - x^\dagger \rangle| \leq C \|A(x - x^\dagger)\|$ for all $x \in X$ the condition (9) is even necessary. \square

3. CONVERGENCE RATES RESULTS

As next step we are going to apply Lemma 1 to prove convergence rates for the variant (4) of Tikhonov regularization with the mixed penalty functional Ω from (5) and regularized solutions x_α^δ as minimizers of Φ_α^δ . This, however, requires that for all $\alpha > 0$ such regularized solutions exist. This will be shown by the following lemma.

Lemma 2. *For all $\alpha > 0$ and $y^\delta \in Y$ there are minimizers x_α^δ to the Tikhonov functional Φ_α^δ from (4).*

Proof. In [37, Lemma 2.1] it was shown that the functional $\|\cdot\|_{\ell^0}$ is weakly lower semi-continuous with respect to ℓ^0 -norm bounded sequences which are weakly convergent in ℓ^2 . Based on that fact the proof of [37, Theorem 2.3] can be extended from the specific penalty Ω from (6) used there to our general convex stabilizing penalty Ω from (5). This proves the lemma. \square

Now we are ready to formulate the first theorem based on the variational source condition (12).

Theorem 1. *Under Assumptions 1 – 3 the linear convergence rate*

$$(16) \quad \|x_{\alpha_*}^\delta - x^\dagger\|_{\ell^1} = \mathcal{O}(\delta) \quad \text{as } \delta \rightarrow 0,$$

holds true for the Tikhonov regularization (4) with $1 < q < \infty$ and the penalty functional Ω from (5) whenever the regularization parameter $\alpha_ > 0$ is chosen either a priori as $\alpha_* = \alpha_*(\delta) \sim \delta^{q-1}$ or a posteriori as $\alpha_* = \alpha_*(\delta, y^\delta)$ by applying the sequential discrepancy principle (see Remark 6 below).*

Proof. If we can prove that $x_{\alpha_*}^\delta \in \mathcal{M}$, the convergence rate result (16) of the theorem is an immediate consequence of the Theorems 1 and 2 from [22] based on the variational source condition (12). Namely, an inspection of the proofs of both theorems shows that no convexity assumption is imposed on the occurring penalties, and hence the non-convex penalty Ω from (5) is applicable. Now it remains to show that $x_{\alpha_*}^\delta \in \mathcal{M}$ is true for a sufficiently large value ρ .

Evidently, we have $x_\alpha^\delta \in \ell^0$ for all $\alpha > 0$, because $\|\cdot\|_{\ell^0}$ is part of the penalty functional Ω . Moreover, from $\Phi_{\alpha_*}^\delta(x_{\alpha_*}^\delta) \leq \Phi_{\alpha_*}^\delta(x^\dagger)$ it follows that $\mathcal{R}(x_{\alpha_*}^\delta) \leq \frac{\delta^q}{q\alpha_*} + \|x^\dagger\|_{\ell^0} + \mathcal{R}(x^\dagger)$. For both choices of the regularization parameter α_* under consideration we have that $\frac{\delta^q}{\alpha_*}$ tends to zero as $\delta \rightarrow 0$ (cf. [1, Proposition 8]) and is consequently bounded by a constant $\hat{K} > 0$ whenever $0 < \delta \leq \bar{\delta}$. For those sufficiently small $\delta > 0$ we have $\mathcal{R}(x_{\alpha_*}^\delta) \leq \bar{K}$ for $\bar{K} := \frac{\hat{K}}{q} + \|x^\dagger\|_{\ell^0} + \mathcal{R}(x^\dagger) < \infty$. The stabilizing property of the functional \mathcal{R} from item (d) of Assumption 1 ensures that there is a radius $r(\bar{K})$ such that $\|x_{\alpha_*}^\delta\|_{\ell^2} \leq r(\bar{K})$. Prescribing ρ such that the inequality $\rho \geq r(\bar{K})$ is valid, we finally find $x_{\alpha_*}^\delta \in \mathcal{M}$ for sufficiently small $\delta > 0$. This completes the proof. \square

Remark 6. We say that $\alpha_* = \alpha(\delta, y^\delta)$ is chosen according to the *sequential discrepancy principle* if the regularization parameter is taken for sufficiently large $\alpha_0 > 0$ and prescribed constants $0 < \nu < 1$, $\tau > 1$, from a geometric sequence

$$\Delta_\nu := \{\alpha_j : \alpha_j := \nu^j \alpha_0, \quad j = 1, 2, \dots\},$$

such that $\alpha_* = \alpha_{j_*} \in \Delta_\nu$ satisfies

$$\|Ax_{\alpha_{j_*}}^\delta - y^\delta\| \leq \tau \delta < \|Ax_{\alpha_{(j_*-1)}}^\delta - y^\delta\|.$$

\square

Now let us consider the situation that the source condition (9) and hence Assumption 3 are violated. Then the estimate (15) is the starting point for applying the *method of approximate source conditions* (cf. [20] and [19]) for obtaining convergence rates which are lower than (7). In this context, we call $\varphi : (0, \infty) \rightarrow (0, \infty)$ an index function if it is continuous and strictly increasing with $\lim_{t \rightarrow +0} \varphi(t) = 0$. A basic tool is the positive, concave and strictly decreasing distance function

$$(17) \quad d_{\xi^\dagger}(R) := \min\{\|\xi^\dagger - A^*w\|_{\ell^2} : w \in X, \|w\|_{Y^*} \leq R\}, \quad R > 0,$$

satisfying $\lim_{R \rightarrow \infty} d_{\xi^\dagger}(R) = 0$. When the source condition (9) fails, then we have for all $R > 0$ elements $w_R \in Y^*$ with $\|w_R\|_{Y^*} = R$ and $\zeta_R \in \ell^2$ with $\|\zeta_R\|_{\ell^2} = d_{\xi^\dagger}(R)$ such that $\xi^\dagger = A^*w_R + \zeta_R$.

Theorem 2. *Under Assumptions 1 and 2 and in the case that Assumption 3 is violated, i.e. the source condition (9) fails, define the strictly concave index function*

$$\varphi(t) := d_{\xi^\dagger}(\Psi^{-1}(t)),$$

including both the distance function d_{ξ^\dagger} from (17) and the auxiliary function $\Psi(R) := \frac{d_{\xi^\dagger}(R)}{R}$, then the sublinear convergence rate

$$\|x_{\alpha_*}^\delta - x^\dagger\|_{\ell^1} = \mathcal{O}(\varphi(\delta)) \quad \text{as } \delta \rightarrow 0,$$

holds true for the Tikhonov regularization (4) with $1 < q < \infty$ and the penalty functional Ω from (5) whenever the regularization parameter $\alpha_ > 0$ is chosen either a priori as $\alpha_* = \alpha_*(\delta) = \frac{\delta^p}{\varphi(\delta)}$ or a posteriori as $\alpha_* = \alpha_*(\delta, y^\delta)$ by applying the sequential discrepancy principle.*

Proof. The proof is analogous to the proof of the above Theorem 1 and based on Theorems 1 and 2 from [22]. The new aspect is to derive an inequality of the form

$$0 \leq \mathcal{R}(x) - \mathcal{R}(x^\dagger) + C \varphi(\|A(x - x^\dagger)\|), \quad 0 < C < \infty,$$

for all x from a closed ball $B_{\bar{r}}(x^\dagger)$ of $X = \ell^2$ with sufficiently large radius \bar{r} . To show such an inequality we use (15) and estimate $|\langle \xi^\dagger, x - x^\dagger \rangle|$ for the $x \in X$ under consideration from above as

$$\begin{aligned} |\langle \xi^\dagger, x - x^\dagger \rangle| &\leq |\langle w_R, A(x - x^\dagger) \rangle_{Y^* \times Y} + \langle \zeta_R, x - x^\dagger \rangle| \\ &\leq \|w_R\|_{Y^*} \|A(x - x^\dagger)\| + d_{\xi^\dagger}(R) \bar{r} \leq R \|A(x - x^\dagger)\| + d_{\xi^\dagger}(R) \bar{r}. \end{aligned}$$

After some calculation this implies the estimate

$$|\langle \xi^\dagger, x - x^\dagger \rangle| \leq (\bar{r} + 1) d_{\xi^\dagger}(\Psi^{-1}(\|A(x - x^\dagger)\|))$$

and completes the proof taking into the account that

$$K\delta + C\varphi(\delta) = \mathcal{O}(\varphi(\delta)) \quad \text{as } \delta \rightarrow 0.$$

□

4. THE SPECIAL CASE $\mathcal{R}(x) = \frac{\eta}{2} \|x\|_{\ell^2}^2$

In this section, we consider the special case

$$\mathcal{R}(x) = \frac{\eta}{2} \|x\|_{\ell^2}^2, \quad \eta > 0,$$

which yields the specific penalty Ω in (6). Our focus is on the impact of the coefficient η on the error estimates. Evidently, Assumption 3 is always satisfied with the source element $\omega = \sum_{k \in \mathbb{N}; \xi_k^\dagger \neq 0} \xi_k^\dagger \omega_k$, which implies that $\|\omega\|_{Y^*} \leq K\eta \|x^\dagger\|_{\ell^\infty}$ with K defined

in the context of formula (8). In Theorem 1 we have derived linear error estimates for Tikhonov regularization with the combined penalty functional $\Omega(x)$, but we have ignored factors weighting the ℓ^0 -term relative to the convex stabilizing part in the penalty Ω as it is the factor η in (6).

In the sequel we exclude the singular case $x^\dagger = 0$ and estimate the error $\|x_{\alpha, \eta}^\delta - x^\dagger\|_{\ell^1}$, denoting $x_{\alpha, \eta}^\delta$ be the minimizer, under appropriate parameter choice rules. For any fixed $\eta > 0$, from $\Phi_\alpha(x_{\alpha, \eta}^\delta) \leq \Phi_\alpha(x^\dagger)$ the following inequality hold true

$$\|x_{\alpha, \eta}^\delta\|_{\ell^2}^2 \leq \frac{2\delta^q}{q\alpha\eta} + \frac{2}{\eta} \|x^\dagger\|_{\ell^0} + \|x^\dagger\|_{\ell^2}^2.$$

Consequently, we can find $x_{\alpha, \eta}^\delta \in \mathcal{M}$ for sufficiently small $\delta > 0$ and

$$(18) \quad \rho > \sqrt{\frac{2}{\eta} \|x^\dagger\|_{\ell^0} + \|x^\dagger\|_{\ell^2}^2}.$$

More precisely, a priori parameter choices $\alpha_* = \alpha_*(\delta)$ and a posteriori parameter choices $\alpha_* = \alpha_*(\delta, y^\delta)$ with some $\bar{\delta} > 0$ yield $\|x_{\alpha_*, \eta}^\delta\|_{\ell^2} \leq \rho$ and hence $x_{\alpha_*, \eta}^\delta \in \mathcal{M}$ for $0 < \delta \leq \bar{\delta}$ provided that the limit condition $\delta^q/\alpha_* \rightarrow 0$ as $\delta \rightarrow 0$ holds true. Thus, the choice of a sufficiently large $\rho > 0$ allows us to satisfy the premise of the following theorem for $0 < \delta \leq \bar{\delta}$.

Theorem 3. *For the specific penalty Ω from (6) let items (a) and (b) of Assumption 1 and Assumption 2 be satisfied. Fix a sufficiently large bound $\rho > 0$ for the set \mathcal{M} from*

(11). Then for all triples (η, α, δ) of positive real numbers with $x_{\alpha, \eta}^\delta \in \mathcal{M}$ there holds, denoting $c_0(\eta)$ by

$$c_0(\eta) := K \left(\frac{1}{\rho} + \frac{1}{\underline{\kappa}} + \eta \|x^\dagger\|_{\ell^\infty} \right),$$

the error estimate

$$(19) \quad \|x_{\alpha, \eta}^\delta - x^\dagger\|_{\ell^1} \leq \rho \left[\frac{\delta^q}{q\alpha} + c_0(\eta)\delta + (c_0(\eta))^{\frac{q}{q-1}} \alpha^{\frac{1}{q-1}} \right].$$

For the standard a priori parameter choice $\alpha_* = \alpha(\delta) := \delta^{q-1}$, the error estimate

$$(20) \quad \|x_{\alpha_*, \eta}^\delta - x^\dagger\|_{\ell^1} \leq C \delta \quad \text{with} \quad C = \rho \left[\frac{1}{q} + c_0(\eta) + (c_0(\eta))^{\frac{q}{q-1}} \right],$$

holds true for $\eta > 0$ and $\rho > 0$ satisfying the condition (18) and for sufficiently small $\delta > 0$.

Proof. Since $x_{\alpha, \eta}^\delta$ is a minimizer of functional Φ_α , inserting the inequality (12), we have

$$(21) \quad \begin{aligned} \frac{1}{q} \|Ax_{\alpha, \eta}^\delta - y^\delta\|^q &\leq \frac{\delta^q}{q} + \alpha(\Omega(x^\dagger) - \Omega(x_{\alpha, \eta}^\delta)) \\ &\leq \frac{\delta^q}{q} - \frac{\alpha}{\rho} \|x_{\alpha, \eta}^\delta - x^\dagger\|_{\ell^1} + \alpha c_0(\eta) \|Ax_{\alpha, \eta}^\delta - x^\dagger\| \\ &\leq \frac{\delta^q}{q} - \frac{\alpha}{\rho} \|x_{\alpha, \eta}^\delta - x^\dagger\|_{\ell^1} + \alpha c_0(\eta) (\|Ax_{\alpha, \eta}^\delta - y^\delta\| + \delta). \end{aligned}$$

As $q > 1$ we use Young's inequality with the dual exponent $q' = q/(q-1)$ as $ab \leq a^q/q + b^{q'}/q'$, $a, b > 0$, to obtain

$$\alpha c_0(\eta) \|Ax_{\alpha, \eta}^\delta - y^\delta\| \leq \frac{(q-1)}{q} (\alpha c_0(\eta))^{\frac{q}{q-1}} + \frac{1}{q} \|Ax_{\alpha, \eta}^\delta - y^\delta\|^q.$$

Combining this with (21) we derive (19) and obviously obtain further (20) for the a priori parameter choice $\alpha_* := \delta^{q-1}$. This proves the theorem. \square

To keep the constant C in (20) small, it seems that the bound $\rho > 0$ would have to be prescribed as small as possible, which implies that η should be large enough. On the other hand, the value $c_0(\eta) > 0$ would have to be prescribed as small as possible as well, which implies that η should be small. As will be shown below in Section 5, the numerical error estimates first decreases and increases later when η decreases from large to small values.

For the a posteriori parameter choice based on the sequential discrepancy principle, we can also establish an error estimate with clear dependence of the coefficient η .

Theorem 4. *Under the conditions of Theorem 3 consider the regularization parameter choice based on the sequential discrepancy principle. Suppose $\alpha_* = \alpha_{j_*} \in \Delta_\nu$ is chosen according to*

$$\|Ax_{\alpha_{j_*}, \eta}^\delta - y^\delta\| \leq \tau \delta < \|Ax_{\alpha_{(j_*-1)}, \eta}^\delta - y^\delta\|.$$

Then the ℓ^1 error estimate

$$\|x_{\alpha_{j_*}, \eta}^\delta - x^\dagger\|_{\ell^1} \leq C_2 \delta$$

holds true with the constant

$$C_2 = \rho c_0(\eta) \left[\left(\frac{2}{\tau-1} \right)^{q-1} \frac{\tau^q + 1}{\tau^q - 1} \frac{1}{\nu} + \tau + 1 \right].$$

Proof. The inequality (12) provides us with

$$(22) \quad \|x - x^\dagger\|_{\ell^1} \leq \rho \left[\Omega(x) - \Omega(x^\dagger) + c_0(\eta) \|Ax - Ax^\dagger\| \right]$$

and

$$(23) \quad \Omega(x^\dagger) - \Omega(x) \leq c_0(\eta) \|Ax - Ax^\dagger\|.$$

Moreover, by fixing any parameter set (η, α, δ) we have for the minimizer $x_{\alpha, \eta}^\delta$ that

$$\frac{1}{q} \|Ax_{\alpha, \eta}^\delta - y^\delta\|^q + \alpha \Omega(x_{\alpha, \eta}^\delta) \leq \frac{\delta^q}{q} + \alpha \Omega(x^\dagger),$$

and consequently

$$(24) \quad \Omega(x_{\alpha, \eta}^\delta) - \Omega_\alpha(x^\dagger) \leq \frac{\delta^q}{q\alpha}.$$

Choosing another parameter α_{j^*-1} , we derive

$$(25) \quad \begin{aligned} \frac{1}{q} \|Ax_{\alpha_{j^*-1}, \eta}^\delta - y^\delta\|^q &\leq \frac{1}{q} \|Ax^\dagger - y^\delta\|^q + \alpha_{j^*-1} \left(\Omega(x^\dagger) - \Omega(x_{\alpha_{j^*-1}, \eta}^\delta) \right) \\ &\stackrel{(23)}{\leq} \frac{\delta^q}{q} + \frac{c_0(\eta)\alpha_{j^*}}{\nu} \|Ax_{\alpha_{j^*-1}, \eta}^\delta - Ax^\dagger\|. \end{aligned}$$

Noticing the fact that $\|Ax_{\alpha_{j^*-1}, \eta}^\delta - y^\delta\| > \tau\delta$, we thus obtain

$$(26) \quad \frac{\delta^q}{q} \leq \frac{c_0(\eta)\alpha_{j^*}}{(\tau^q - 1)\nu} \|Ax_{\alpha_{j^*-1}, \eta}^\delta - Ax^\dagger\|.$$

On the other hand, by $(a + b)^q \leq 2^{q-1}(a^q + b^q)$, $a, b \geq 0$, $q \geq 1$ and the inequalities (25)-(26), we can derive

$$\begin{aligned} \frac{\|Ax_{\alpha_{j^*-1}, \eta}^\delta - Ax^\dagger\|^q}{q} &\leq 2^{q-1} \left(\frac{\delta^q}{q} + \frac{1}{q} \|Ax_{\alpha_{j^*-1}, \eta}^\delta - y^\delta\|^q \right) \\ &\stackrel{(25)}{\leq} 2^q \frac{\delta^q}{q} + 2^{q-1} \frac{c_0(\eta)\alpha_{j^*}}{\nu} \|Ax_{\alpha_{j^*-1}, \eta}^\delta - Ax^\dagger\| \\ &\stackrel{(26)}{\leq} 2^{q-1} \frac{\tau^q + 1}{\tau^q - 1} \frac{c_0(\eta)\alpha_{j^*}}{\nu} \|Ax_{\alpha_{j^*-1}, \eta}^\delta - Ax^\dagger\|, \end{aligned}$$

that is,

$$\frac{c_0(\eta)\alpha_{j^*}}{\nu} \geq \frac{\tau^q - 1}{\tau^q + 1} \frac{1}{2^{q-1}q} \|Ax_{\alpha_{j^*-1}, \eta}^\delta - Ax^\dagger\|^{q-1}.$$

Substituting $(\tau - 1)\delta \leq \|Ax_{\alpha_{j^*-1}, \eta}^\delta - Ax^\dagger\|$ into previous inequality, we can derive

$$(27) \quad \frac{c_0(\eta)\alpha_{j^*}}{\nu} \geq \frac{1}{q} \frac{\tau^q - 1}{\tau^q + 1} \left(\frac{\tau - 1}{2} \right)^{q-1} \delta^{q-1}.$$

Finally, combining (22), (24) and (27), we obtain

$$\begin{aligned} \|x_{\alpha_{j^*}, \eta}^\delta - x^\dagger\|_{\ell^1} &\leq \rho \left[\frac{\delta^q}{q\alpha_{j^*}} + c_0(\eta)(\tau + 1)\delta \right] \\ &\leq \rho c_0(\eta) \left[\left(\frac{2}{\tau - 1} \right)^{q-1} \frac{\tau^q + 1}{(\tau^q - 1)\nu} \delta + (\tau + 1)\delta \right] \end{aligned}$$

which completes the proof. \square

As one can observe, constants near error estimates of Theorems 3 and 4 highly depend on the coefficient η and the exact solution x^\dagger . Even if one cannot know the exact solution, the subsequent section will show that a more sparse approximate solution is expected when the factor η becomes small.

5. NUMERICAL EXAMPLES AND EXTENDED DISCUSSION

5.1. Numerical examples. In this section, we provide some numerical examples verifying and illustrating the theoretical predictions of the previous sections, for linear forward operators in the first subsection, and for a nonlinear test example in the second subsection.

5.1.1. The discretized linear case. The following finite-dimensional ill-posed problem in the sequence space is considered such that

$$A\mathbf{x} = \mathbf{y} + \omega = \mathbf{y}^\delta,$$

where ω is a Gaussian random variable with zero mean and a variance matrix $\delta^2 I$.

We use the Tikhonov type functional

$$(28) \quad \Phi_\alpha^\delta(\mathbf{x}) := \frac{1}{2} \|A\mathbf{x} - \mathbf{y}^\delta\|^2 + \alpha (\|\mathbf{x}\|_{\ell^0} + \mathcal{R}(\mathbf{x})),$$

with $\mathcal{R}(x) := \frac{\eta}{2} \|\mathbf{x}\|_{\ell^2}^2$ or $\mathcal{R}(x) := \frac{\eta}{p} \|\mathbf{x}\|_{\ell^p}^p$ ($1 < p < 2$) respectively. Noticing that the difficulty of obtaining the minimizer of Φ_α^δ remains, we adopt the Tikhonov type functional (28) into Bayesian interfaces as in [37]. More precisely, we introduce the prior density, in the finite-dimensional regime,

$$p(\mathbf{x}) \propto \exp(-\tilde{\alpha} (\|\mathbf{x}\|_{\ell^0} + \mathcal{R}(\mathbf{x}))).$$

The posterior distribution then has a density

$$(29) \quad p(\mathbf{x}|\mathbf{y}^\delta) \propto \exp\left(-\frac{1}{2\delta^2} \|A\mathbf{x} - \mathbf{y}^\delta\|^2 - \tilde{\alpha} (\|\mathbf{x}\|_{\ell^0} + \mathcal{R}(\mathbf{x}))\right).$$

Then the maximum a posteriori (MAP) estimate $\hat{\mathbf{x}}$ of (29) coincides with the minimizer of (28) with $\alpha = \tilde{\alpha}\delta^2$. To better illustrate the influence of the convex/stabilizing functional $\mathcal{R}(x)$, as mentioned above, we consider the following regularization schemes

$$(30) \quad \begin{aligned} \Phi_{\alpha,2}^\delta(\mathbf{x}) &:= \frac{1}{2} \|A\mathbf{x} - \mathbf{y}^\delta\|^2 + \alpha \left(\|\mathbf{x}\|_{\ell^0} + \frac{\eta}{2} \|\mathbf{x}\|_{\ell^2}^2 \right); \\ \Phi_{\alpha,1.1}^\delta(\mathbf{x}) &:= \frac{1}{2} \|A\mathbf{x} - \mathbf{y}^\delta\|^2 + \alpha \left(\|\mathbf{x}\|_{\ell^0} + \frac{\eta}{1.1} \|\mathbf{x}\|_{\ell^{1.1}}^{1.1} \right). \end{aligned}$$

We shall emphasize that the latter penalty functional $\Omega(x) = \|\mathbf{x}\|_{\ell^0} + \frac{\eta}{1.1} \|\mathbf{x}\|_{\ell^{1.1}}^{1.1}$ has a better sparsity promoting properties compared with the former one.

To obtain a single approximant from the posterior distribution density $p(\mathbf{x}|\mathbf{y}^\delta)$ given a known observation \mathbf{y}^δ , we can either choose an MAP estimate or a posterior mean (PM) estimate where the latter one needs to integrate the posterior probability densities numerically. In current section, we implement the Gibbs sampler to construct transition kernels in the MCMC method. For extended discussion on statistical inverse problems and MCMC methods, we refer to the monograph [25] and references therein.

The first example considers the numerical differentiation solving the Volterra integral equation of the first kind

$$Ax(s) = \int_0^s x(t)dt, \quad s \in (0, 1)$$

where the solution x is equally discretized with 128 mesh points in the interval $(0, 1)$ and the exact solution vector \mathbf{x} contains 128 coefficients with 6 non-zero ones. The noise ω is added to the exact data \mathbf{y} with a noise covariance $\delta^2 = 10^{-6}$. The regularization parameter $\tilde{\alpha}$ in the posterior distribution density $p(\mathbf{x}|\mathbf{y}^\delta)$ is chosen in a regularization parameter set $\{2^{-k}/\delta^2, k = 0, 1, \dots\}$ where the chosen $\tilde{\alpha}$ provides a slightly larger residual with respect to the noise level. Recalling Theorem 4, such a parameter choice rule obeys the sequential discrepancy principle there. In order to keep the numerical realization consistent to each other, we use the Gibbs sampler (see e.g. [25, Section 3.6.3]) to obtain both minimizers of the two regularization schemes in (30). The sample size of the MCMC methods is chosen by 10^5 . Except the MAP estimate \hat{x}_{MAP} , we also generate the PM estimate \hat{x}_{PM} by choosing a suitable burn-in period $[0, 5000]$. We summarize the quantity information for both MAP and PM estimates in Table 1 where the ℓ^2 and ℓ^1 relative errors take the standard form, for instance, with $\|\hat{x} - x^\dagger\|_{\ell^2}/\|x^\dagger\|_{\ell^2}$ and $\|\hat{x} - x^\dagger\|_{\ell^1}/\|x^\dagger\|_{\ell^1}$. Here \hat{x} represents either the MAP or PM estimate. As Table 1 shows, the smaller η the more sparse estimates one can obtain. On the other hand, Theorem 4 has asserted that if one chooses the prescribed constant η too large or too small, the constant near the ℓ^1 -norm error estimate becomes large for the penalty functional $\Omega(x) = \|\mathbf{x}\|_{\ell^0} + \frac{\eta}{2}\|\mathbf{x}\|_{\ell^2}^2$. Such observation is mostly verified by the quantity information in Table 1. For instance, the PM estimate has a minimal ℓ^1 relative error by choosing $\eta = 1/2^4$. Though it is not easy to prove the same assertion for the stabilizing functional $\frac{\eta}{1.1}\|\mathbf{x}\|_{\ell^{1.1}}^{1.1}$, the numerical performance is analogous to that of the quadratic stabilizing one by varying η .

TABLE 1. Numerical results of the first example with respect to the PM and MAP estimates.

The quantity information for the PM estimate.			
	Relative ℓ^2 error	Relative ℓ^1 error	$\ \hat{x}_{PM}\ _{\ell^0}$
$\ell^0 + \ell^2$ scheme, $\eta/2 = 1$	1.76e-1	3.64e-1	60
$\ell^0 + \ell^2$ scheme, $\eta/2 = 1/2^4$	5.11e-2	4.98e-2	6
$\ell^0 + \ell^2$ scheme, $\eta/2 = 1/2^8$	6.03e-2	5.76e-2	6
$\ell^0 + \ell^{1.1}$ scheme, $\eta/1.1 = 5$	1.10e-1	9.75e-2	7
$\ell^0 + \ell^{1.1}$ scheme, $\eta/1.1 = 1$	5.45e-2	5.08e-2	6
$\ell^0 + \ell^{1.1}$ scheme, $\eta/1.1 = 1/2^4$	5.71e-2	5.27e-2	6
$\ell^0 + \ell^{1.1}$ scheme, $\eta/1.1 = 1/2^8$	6.28e-2	5.59e-2	6

The quantity information for the MAP estimate.			
	Relative ℓ^2 error	Relative ℓ^1 error	$\ \hat{x}_{MAP}\ _{\ell^0}$
$\ell^0 + \ell^2$ scheme, $\eta/2 = 1$	1.76e-1	2.86e-1	15
$\ell^0 + \ell^2$ scheme, $\eta/2 = 1/2^4$	5.62e-2	5.68e-2	6
$\ell^0 + \ell^2$ scheme, $\eta/2 = 1/2^8$	5.97e-2	5.64e-2	6
$\ell^0 + \ell^{1.1}$ scheme, $\eta/1.1 = 5$	1.13e-1	1.10e-1	7
$\ell^0 + \ell^{1.1}$ scheme, $\eta/1.1 = 1$	6.72e-2	6.40e-2	6
$\ell^0 + \ell^{1.1}$ scheme, $\eta/1.1 = 1/2^4$	5.51e-2	5.00e-2	6
$\ell^0 + \ell^{1.1}$ scheme, $\eta/1.1 = 1/2^8$	5.85e-2	5.54e-2	6

We select two particular choices of $\eta/p = 1$, $\eta/p = 1/2^4$ with $p = 2$, $p = 1.1$ and plot a logarithmic scale of the ℓ^2 (and ℓ^1) absolute error with respect to the absolute noise level δ in Figure 1. The reference black dashed line of order $O(\delta)$ is provided as

well to verify the decaying speed of the numerical error estimates. Both choices of η provide similar decaying speeds for small noise level but slightly vary when the noise level becomes large.

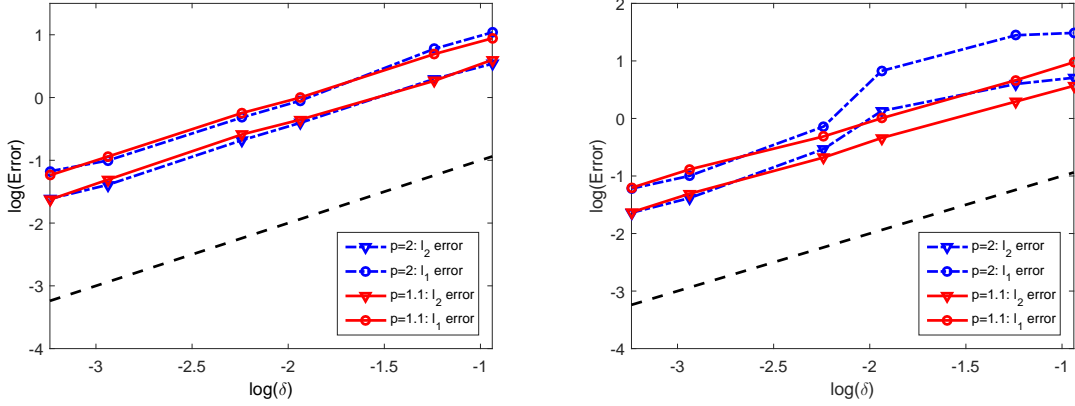


FIGURE 1. First example (PM): Absolute error versus noise level in logarithmic scales. Black dashed line is the reference line of order δ , Left: $\eta/p = 1/2^4$, right: $\eta/p = 1$.

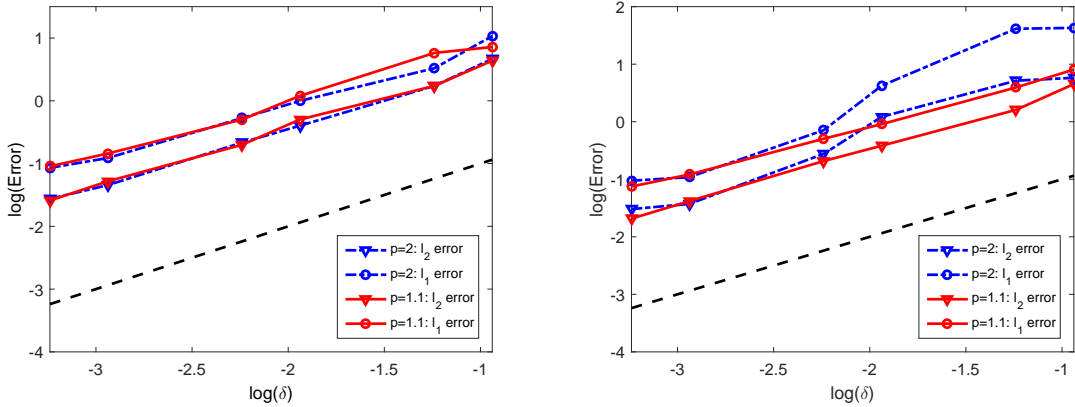


FIGURE 2. First example (MAP): Absolute error versus noise level in logarithmic scales. Black dashed line is the reference line of order δ , Left: $\eta/p = 1/2^4$, right: $\eta/p = 1$.

5.1.2. *A nonlinear test example.* We also tested comparable approximate solutions for a discretized nonlinear operator equation $F(x) = y$, with a nonlinear Hammerstein forward operator

$$F(x)(s) = \int_0^s x^2(t) dt, \quad s \in (0, 1).$$

The regularized solutions are minimizers to a discretized version of the nonlinear analog

$$\Phi_\alpha^\delta(x) := \frac{1}{q} \|F(x) - y^\delta\|^q + \alpha \Omega(x) \rightarrow \min, \quad \text{subject to } x \in X = \ell^2,$$

to the extremal problem (4) with penalties $\Omega(x) = \|x\|_{\ell^0} + \frac{q}{p} \|x\|_{\ell^p}^p$ for situations $p = 2$ and $1 < p < 2$.

This time, we implement a wavelet expansion to transform the exact solution x into a sequential space. More precisely, given the fixed orthonormal wavelet system $\{\varphi, \psi\}$, the exact solution x can be expanded in the following form

$$x = \sum_{k \in \mathbb{Z}} \langle x, \phi_{0,k} \rangle \phi_{0,k} + \sum_j \sum_{k \in \mathbb{Z}} \langle x, \psi_{j,k} \rangle \psi_{j,k} =: \sum_{\lambda \in \Lambda} \langle x, \Phi_\lambda \rangle \Phi_\lambda,$$

with an appropriately chosen index set Λ . The sequential solution \mathbf{x} is presented by

$$\mathbf{x} = \{\langle x, \Phi_\lambda \rangle\}_{\lambda \in \Lambda}.$$

Choosing the Haar wavelets, we discretize the exact solution with 64 coefficients and only 6 of them are non-zero. Similar to the linear example, we summarize the quantity information in Table 2. The logarithmic-scale fitting curve between the ℓ^2 (and ℓ^1) absolute error and the absolute noise level δ are displayed in Figure 3. Different from the linear cases, the smaller η , the better approximants towards the exact solution. One may guess that in the nonlinear cases, the balance between the stabilizing functional \mathcal{R} and the ℓ^0 -term in the penalty is more sophisticated and shall be analyzed more carefully.

TABLE 2. Numerical results of the second example with respect to the PM and MAP estimates.

The quantity information for the PM estimate.			
	Relative ℓ^2 error	Relative ℓ^1 error	$\ \hat{x}_{PM}\ _{\ell^0}$
$\ell^0 + \ell^2$ scheme, $\eta/2 = 5$	5.93e-2	4.49e-2	28
$\ell^0 + \ell^2$ scheme, $\eta/2 = 1$	3.72e-2	2.97e-2	6
$\ell^0 + \ell^2$ scheme, $\eta/2 = 1/2^4$	2.55e-2	2.00e-2	6
$\ell^0 + \ell^{1.1}$ scheme, $\eta/1.1 = 5$	3.17e-2	2.20e-2	6
$\ell^0 + \ell^{1.1}$ scheme, $\eta/1.1 = 1$	2.56e-2	2.02e-2	6
$\ell^0 + \ell^{1.1}$ scheme, $\eta/1.1 = 1/2^4$	1.44e-2	1.06e-2	6

The quantity information for the MAP estimate.			
	Relative ℓ^2 error	Relative ℓ^1 error	$\ \hat{x}_{MAP}\ _{\ell^0}$
$\ell^0 + \ell^2$ scheme, $\eta/2 = 5$	5.87e-3	3.82e-2	8
$\ell^0 + \ell^2$ scheme, $\eta/2 = 1$	3.74e-2	2.94e-2	6
$\ell^0 + \ell^2$ scheme, $\eta/2 = 1/2^4$	2.57e-2	2.00e-2	6
$\ell^0 + \ell^{1.1}$ scheme, $\eta/1.1 = 5$	2.89e-2	2.18e-2	6
$\ell^0 + \ell^{1.1}$ scheme, $\eta/1.1 = 1$	2.61e-2	2.03e-2	6
$\ell^0 + \ell^{1.1}$ scheme, $\eta/1.1 = 1/2^4$	1.56e-2	1.18e-2	6

5.2. Extended discussion. We establish ℓ^1 -norm error estimates for a Tikhonov type regularization with an ℓ^0 -term and a stabilizing convex functional. The error estimates are obtained by the variational inequality in Lemma 1. On the other hand, sparsity of x^\dagger expresses, in some sense, a well-posed situation for solving the operator equation (3), although the range of A is assumed to be non-closed indicating ill-posedness. This effect becomes clear if one considers the *conditional stability estimate*

$$\|x - x^\dagger\|_{\ell^1} \leq \frac{1}{j(A, \Sigma(x^\dagger))} \|A(x - x^\dagger)\| \quad \text{for all } x \in \Sigma(x^\dagger).$$

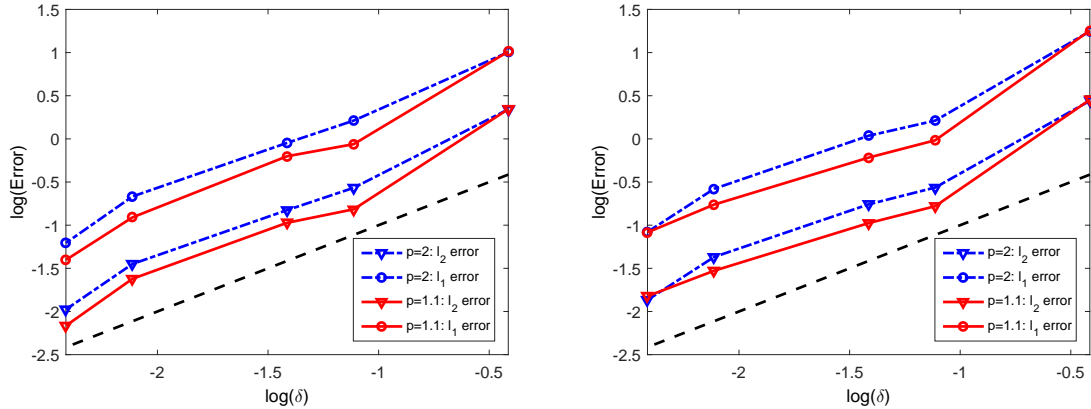


FIGURE 3. Second example: Absolute error versus noise level in logarithmic scales when $\eta/p = 1/2^4$. Black dashed line is the reference line of order δ . Left: PM estimate; Right: MAP estimate.

If it would be possible to ensure that the regularized solutions x_α^δ belong to the stability set $\Sigma(x^\dagger)$ for sufficiently small $\delta > 0$, then the combination of conditional stability and regularization suggested originally in the article [10] (see also the more recent papers [11, 24]) could be used for obtaining linear convergence rates when the regularization parameter $\alpha_* > 0$ is chosen a priori as $\alpha_* = \alpha_*(\delta)$ satisfying the inequalities

$$(31) \quad \underline{c}\delta^q \leq \alpha_*(\delta) \leq \bar{c}\delta^q$$

for constants $0 < \underline{c} \leq \bar{c} < \infty$. Evidently in contrast to Theorems 1, 3 above, where $\alpha_*(\delta) \sim \delta^{p-1}$ is chosen, in the case (31) the quotient $\frac{\delta^q}{\alpha_*(\delta)}$ does not tend to zero as $\delta \rightarrow 0$, but remains in the interval $[\underline{c}, \bar{c}]$. Unfortunately, the capability of our variety of ℓ^0 -regularization is only to ensure that $\|x_{\alpha_*}^\delta\|_{\ell^0} \leq \bar{K} < \infty$ for all regularized solutions whenever $\delta > 0$ is small enough. This, however, is far away from the requirement $x_{\alpha_*}^\delta \in \Sigma(x^\dagger)$ of the conditional stability approach, since the distribution of the $|I|$ nonzero components in x^\dagger is a priori completely unknown. We take some numerical evidence of the parameter choice rule (31) by choosing $\alpha_* = c\delta^p$ with $c = 0.1$ and $c = 0.2$ respectively. The numerical results are displayed in Figure 4. As one can observe, though the error in the logarithmic scale does not form a straight line as in Figures 1-2, the error is still acceptable. Generally, both constants $c = 0.1$ and $c = 0.2$ provide comparably the same accuracy.

ACKNOWLEDGMENT

W. Wang is supported by NSFC (No.11401257). S. Lu is supported by NSFC (No.11522108, 91730304), Shanghai Municipal Education Commission (No.16SG01) and Special Funds for Major State Basic Research Projects of China (2015CB856003). B. Hofmann is supported by German Research Foundation under grant HO 1454/12-1. J. Cheng is supported by NSFC (key projects no.11331004, no.11421110002) and the Programme of Introducing Talents of Discipline to Universities (number B08018).

REFERENCES

- [1] S. W. Anzengruber, B. Hofmann, and P. Mathé. Regularization properties of the discrepancy principle for Tikhonov regularization in Banach spaces. *Applicable Analysis*, 93(7):1382–1400, 2014.

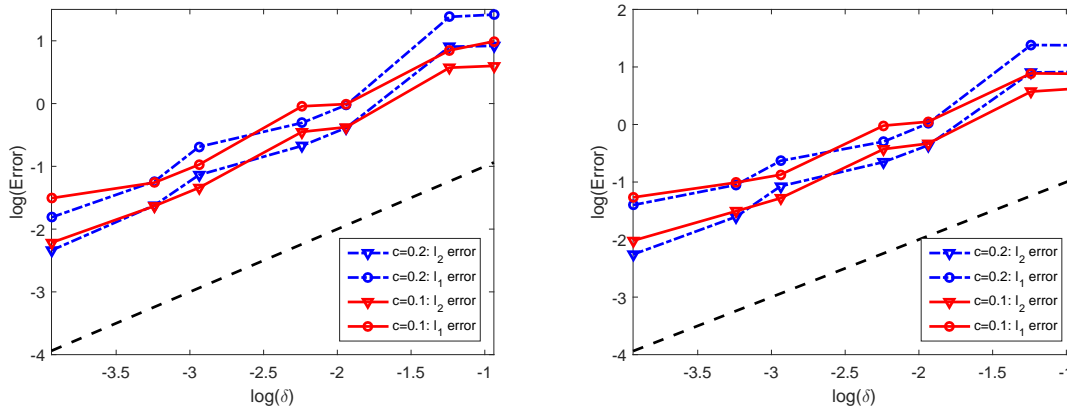


FIGURE 4. First example with the parameter choice rule $\alpha_* = c\delta^p$ and $c = 0.1, 0.2$. Absolute error versus noise level in logarithmic scales when $p = 2$ and $\eta/p = 1/2^4$. Black dashed line is the reference line of order δ . Left: PM estimate, Right: MAP estimate.

- [2] S. W. Anzengruber, B. Hofmann, and R. Ramlau. On the interplay of basis smoothness and specific range conditions occurring in sparsity regularization. *Inverse Problems*, 29(12):125002 (21pp), 2013.
- [3] R. I. Boĭ and B. Hofmann. An extension of the variational inequality approach for obtaining convergence rates in regularization of nonlinear ill-posed problems. *Journal of Integral Equations and Applications*, 22(3):369–392, 2010.
- [4] R. I. Boĭ and B. Hofmann. The impact of a curious type of smoothness conditions on convergence rates in ℓ^1 -regularization. *Eurasian Journal of Mathematical and Computer Applications*, 1(1):29–40, 2013.
- [5] K. Bredies and D. A. Lorenz. Regularization with non-convex separable constraints. *Inverse Problems*, 25(8):085011, 14pp., 2009.
- [6] J. Brodie, I. Daubechies, C. De Mol, D. Giannone, and I. Loris. Sparse and stable Markowitz portfolios. *PNAS*, 106:12267–12272, 2009.
- [7] M. Burger, J. Flemming, and B. Hofmann. Convergence rates in ℓ^1 -regularization if the sparsity assumption fails. *Inverse Problems*, 29(2):025013 (16pp), 2013.
- [8] M. Burger and S. Osher. Convergence rates of convex variational regularization. *Inverse Problems*, 20(5):1411–1421, 2004.
- [9] D. Chen, B. Hofmann, and J. Zou. Elastic-net regularization versus ℓ^1 -regularization for linear inverse problems with quasi-sparse solutions. *Inverse Problems*, 33(1):015004, 17pp, 2017.
- [10] J. Cheng and M. Yamamoto. One new strategy for a priori choice of regularizing parameters in Tikhonov’s regularization. *Inverse Problems*, 16(4):L31–L38, 2000.
- [11] J. Cheng, B. Hofmann, and S. Lu. The index function and Tikhonov regularization for ill-posed problems. *J. Comput. Appl. Math.*, 265:110–119, 2014.
- [12] J. Flemming. *Generalized Tikhonov Regularization and Modern Convergence Rate Theory in Banach Spaces*. Shaker Verlag, Aachen, 2012.
- [13] J. Flemming. Variational smoothness assumptions in convergence rate theory – an overview. *Journal of Inverse and Ill-Posed Problems*, 21(3):395–409, 2013.
- [14] J. Flemming and M. Hegland. Convergence rates in ℓ^1 -regularization when the basis is not smooth enough. *Applicable Analysis*, 94(3):464–476, 2015.
- [15] M. Grasmair. Well-posedness and convergence rates for sparse regularization with sublinear ℓ^q penalty term. *Inverse Probl. Imaging*, 3(3):383–387, 2009.
- [16] M. Grasmair. Generalized Bregman distances and convergence rates for non-convex regularization methods. *Inverse Problems*, 26(11):115014, 16pp., 2010.
- [17] M. Grasmair. Non-convex sparse regularisation. *J. Math. Anal. Appl.*, 365(1):19–28, 2010.
- [18] M. Grasmair, M. Haltmeier, and O. Scherzer. Sparse regularization with ℓ^q penalty term. *Inverse Problems*, 24(5):055020, 13pp., 2008.
- [19] T. Hein and B. Hofmann. Approximate source conditions for nonlinear ill-posed problems—chances and limitations. *Inverse Problems*, 25(3):035003, 16pp., 2009.

- [20] B. Hofmann. Approximate source conditions in Tikhonov-Phillips regularization and consequences for inverse problems with multiplication operators. *Math. Methods Appl. Sci.*, 29(3):351–371, 2006.
- [21] B. Hofmann, B. Kaltenbacher, C. Pöschl, and O. Scherzer. A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators. *Inverse Problems*, 23(3):987–1010, 2007.
- [22] B. Hofmann and P. Mathé. Parameter choice in Banach space regularization under variational inequalities. *Inverse Problems*, 28(10):104006, 17pp, 2012.
- [23] B. Hofmann, P. Mathé and M. Schieck. Modulus of continuity for conditionally stable ill-posed problems in Hilbert space. *J. Inv. Ill-posed Problems* 16(6):567–585, 2008.
- [24] B. Hofmann and M. Yamamoto. On the interplay of source conditions and variational inequalities for nonlinear ill-posed problems. *Applicable Analysis*, 89(11):1705–1727, 2010.
- [25] J. Kaipio J and E. Somersalo. *Statistical and Computational Inverse Problems*, volume 160 of *Applied Mathematical Sciences*. Springer, New York, 2005.
- [26] B. Jin, D. Lorenz and S. Schiffler. Elastic-net regularization: error estimates and active set methods. *Inverse Problems*, 25(11):115022 (26pp), 2009.
- [27] S. Kindermann. Convex Tikhonov regularization in Banach spaces: new results on convergence rates. *J. Inverse Ill-Posed Probl.*, 24(3):341–350, 2016.
- [28] D. A. Lorenz. Convergence rates and source conditions for Tikhonov regularization with sparsity constraints. *J. Inverse Ill-Posed Probl.*, 16(5):463–478, 2008.
- [29] P. Mathé and B. Hofmann. Direct and inverse results in variable Hilbert scales. *Journal of Approximation Theory* 154(2):77–89, 2008.
- [30] R. Ramlau. Regularization properties of Tikhonov regularization with sparsity constraints. *Electron. Trans. Numer. Anal.*, 30:54–74, 2008.
- [31] R. Ramlau and E. Resmerita. Convergence rates for regularization with sparsity constraints. *Electron. Trans. Numer. Anal.*, 37:87–104, 2010.
- [32] R. Ramlau and G. Teschke. Sparse recovery in inverse problems. In *Theoretical foundations and numerical methods for sparse recovery*, volume 9 of *Radon Ser. Comput. Appl. Math.*, pages 201–262. Walter de Gruyter, Berlin, 2010.
- [33] E. Resmerita. Regularization of ill-posed problems in Banach spaces: convergence rates. *Inverse Problems*, 21(4):1303–1314, 2005.
- [34] O. Scherzer (ed.), *Handbook of Mathematical Methods in Imaging*, in 3 volumes, Springer, New York, 2011.
- [35] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. *Variational Methods in Imaging*, volume 167 of *Applied Mathematical Sciences*. Springer, New York, 2009.
- [36] T. Schuster, B. Kaltenbacher, B. Hofmann, and K.S. Kazimierski. *Regularization Methods in Banach Spaces*, volume 10 of *Radon Ser. Comput. Appl. Math.* Walter de Gruyter, Berlin/Boston, 2012.
- [37] W. Wang, S. Lu, H. Mao, and J. Cheng. Multi-parameter Tikhonov regularization with the ℓ^0 sparsity constraint. *Inverse Problems*, 29(6):065018 (18pp), 2013.
- [38] H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 67(2):301–320, 2005.

COLLEGE OF MATHEMATICS, PHYSICS AND INFORMATION ENGINEERING, JIAXING UNIVERSITY, ZHEJIANG, CHINA 314001

E-mail address: weiwang@mail.zjxu.edu.cn

SCHOOL OF MATHEMATICAL SCIENCES, FUDAN UNIVERSITY, SHANGHAI, CHINA 200433

E-mail address: slu@fudan.edu.cn

FACULTY OF MATHEMATICS, CHEMNITZ UNIVERSITY OF TECHNOLOGY, 09107 CHEMNITZ, GERMANY

E-mail address: bernd.hofmann@mathematik.tu-chemnitz.de

SCHOOL OF MATHEMATICAL SCIENCES, FUDAN UNIVERSITY, SHANGHAI, CHINA 200433

E-mail address: jcheng@fudan.edu.cn