UTMS 2010–3 March 29, 2010

A regularization parameter for nonsmooth Tikhonov regularization

by

Kazufumi Ito, Bangti Jin and Tomoya Takeuchi

UNIVERSITY OF TOKYO

GRADUATE SCHOOL OF MATHEMATICAL SCIENCES

KOMABA, TOKYO, JAPAN

A Regularization Parameter for Nonsmooth Tikhonov Regularization

Kazufumi Ito* Bangti Jin† Tomoya Takeuchi‡

March 11, 2010

Abstract

In this paper we develop a novel criterion for choosing regularization parameters for nonsmooth Tikhonov functionals. The proposed criterion is solely based on the value function, and thus applicable to a broad range of functionals. It is analytically compared with the local minimum criterion, and a posteriori error estimates are derived. An efficient numerical algorithm for computing the minimizer is developed, and its convergence properties are also studied. Numerical results for several common nonsmooth functionals are presented.
keywords: Regularization parameter, nonsmooth functional, Tikhonov regularization, value function, error estimate

1 Introduction

We consider linear inverse problems of seeking an approximate solution $x \in X$ to

$$Kx = y^\delta, \tag{1}$$

when only a noisy version $y^\delta \in Y$ of the exact data $y^\dagger = Kx^\dagger$ with $x^\dagger$ being the exact solution is available, and the given data y^δ satisfies $\phi(x^\dagger, y^\delta) \le \delta$ for some metric ϕ. Here the spaces X and Y are Banach spaces, and the operator $K : X \to Y$ is bounded and linear.

As typical for inverse problems, it suffers from ill-posedness. In particular, small changes in the data y^δ can lead to huge deviations in the solution x. To restore the numerical stability, regularization has proved an effective approach. Amongst existing approaches, Tikhonov regularization, which amounts to minimizing the following functional

$$\mathcal{J}_\eta(x) = \phi(x, y^\delta) + \eta\psi(x),$$

is very popular and attractive. Here the functionals ϕ and ψ are known as data fitting functional and regularization functional, respectively. Some common choices of the data fitting functional $\phi(x, y^\delta)$ include $\frac{1}{2}\|Kx - y^\delta\|_{L^2}^2$, $\|Kx - y^\delta\|_{L^1}$ and $\int(Kx - y^\delta \ln Kx)$, which are statistically well suited to additive Gaussian noise, impulsive noise and Poisson noise, respectively. Typical regularization functionals include $\frac{1}{2}\|x\|_{L^2}^2$, $\frac{1}{p}\|x\|_{\ell^p}^p$, $\|x\|_{H^m}^2$ and $|x|_{TV}$. Moreover, the functional $\psi(x)$ is assumed to be nonnegative. The parameter η is called the regularization parameter, and it compromises data fitting with regularization.

The resulting functionals are often nonsmooth. Regularization of this form has attracted considerable interest in recent years, and found applications in a variety of disciplines, e.g. imaging science [30], signal processing [5, 12] and parameter identification [9, 11]. These functionals have demonstrated many

*Center for Research in Scientific Computation & Department of Mathematics, North Carolina State University, Raleigh, NC 27695, USA. (kito@math.ncsu.edu)

†Center for Industrial Mathematics, University of Bremen, D-28334 Bremen, Germany. (btjin@math.uni-bremen.de)

‡Center for Research in Scientific Computation, North Carolina State University, Raleigh, NC. (tntakeuc@ncsu.edu)

desirable properties, e.g. feature promoting/preserving like edge, sparsity and texture, compared to the more conventional Tikhonov regularization, i.e. L^2 data fitting with smoothness regularization. Because of their practical importance, nonsmooth functionals have been the subject of many recent investigations. Theoretically speaking, since the pioneering work [4], convergence and convergence rates under various conditions have been established [28, 29, 17, 25, 15, 26]. Numerically, several efficient algorithms have also been proposed, see e.g. [8, 6, 10, 33, 18, 14, 34] for a rather incomplete list.

But one of the most important problems in applying these regularization formulations, i.e. choosing an appropriate regularization parameter, remains largely unexplored. While the problem of parameter choice has been discussed in great length for quadratic regularization, see e.g. [13] for theoretical studies, and [16, 32] for details about of numerical implementation, the case of nonsmooth regularization has scarcely been addressed. This is attributed to the fact that there often exists only an implicit relation between the solution and the regularization parameter. As to existing studies in parameter choice for inverse problems in Banach spaces, we are aware of Morozov's discrepancy principle, which were recently investigated [2, 24]. Some theoretical results, e.g. convergence and convergence rates, were derived. In [24], two efficient algorithms for solving the discrepancy equation were also proposed. However, the discrepancy principle requires an estimate of the noise level, which is not always available in practice, and the existence of a solution to the discrepancy equation is not guaranteed for some nonsmooth functionals, e.g. L^1-TV and L^2-ℓ^1. Therefore, there is a significant interest in deriving rules which do not require a knowledge of the noise level. One such rule is due to the authors [19], which generalizes the work [23]. Existence of a solution and a posteriori error estimates are derived. Another is the model function approach, recently derived in [10], for the formulation L^2-ℓ^1. Finally, the Hanke-Raus rule and the quasi-optimality criterion have also been generalized [22], and their convergence behavior is discussed, in particular convergence is discussed under certain conditions on the exact solution and noise. However, there is no known efficient algorithm for these two rules.

In this paper, we propose a novel criterion for choosing regularization parameters for general nonsmooth Tikhonov regularization functionals. The proposed rule is solely based on the value function. Since the value function is always continuous, see the next section, the proposed rule is always well defined irrespective of the smoothness of the functionals. Therefore, it is especially attractive for functionals with nonunique minimizer, for which many existing rules are ill-defined. The new criterion is closely related to a known rule (local minimum criterion) due to Regińska [27], which however can be ill-defined for nonsmooth functionals. Moreover, the new criterion is theoretically better behaved in comparison with the former, and admits easier theoretical justifications in terms of a posteriori error estimates. Numerically, an efficient algorithm with practically very desirable monotone convergence is proposed.

The rest of the paper is organized as follows. In Section 2, we collect some important properties, e.g. continuity, concavity, monotonicity and differentiability, of the value function. The value function has been previously investigated in [31, 24, 21]. These properties are important in designing and analyzing the new criterion. In Section 3, we give the new criterion. It is then theoretically compared with the local minimum criterion, and justified in terms of a posteriori error estimates. We develop in Section 4 an efficient numerical algorithm to computing a minimizer of the proposed criterion, and discuss in detail the convergence property of the algorithm. Finally, numerical results for several examples are presented to illustrate the features of the proposed rule and to verify some theoretical results. These examples are of current interest, and include constrained Tikhonov regularization, total variation for deblurring images subjected impulsive noises, image reconstruction with sparsity and group sparsity.

Notation We shall denote a minimizer of the functional $\mathcal{J}_\eta$ by x_η^δ, the set of minimizers by $\mathcal{M}_\eta$, and the minimizer for exact data $y^\dagger$ by x_η, i.e.

$$x_\eta \in \arg\min_{x\in X} \left\{\phi(x, y^\dagger) + \eta\psi(x)\right\}.$$

The norm on a Hilbert space is generically denoted by $\|\cdot\|$.

2 Properties of the value function

We shall use extensively the value function $F(\eta)$ defined as below

$$F(\eta) = \inf_x \mathcal{J}_\eta(x).$$

In particular, the new criterion is derived based on the function F. This section collects some of its important properties, especially differentiability. These properties will play an important role in analyzing the proposed criterion as well as the fixed point iterative algorithm.

A first result shows the monotonicity and concavity of F.

Theorem 2.1. *The value function $F(\eta)$ is monotonically increasing and concave.*

Proof. Given a $\hat{\eta} < \eta$, for any $x \in X$, by the nonnegativity of $\psi(x)$, we have

$$F(\hat{\eta}) \leq \mathcal{J}_{\hat{\eta}}(x) = \phi(x, y^\delta) + \hat{\eta}\psi(x) \leq \phi(x, y^\delta) + \eta\psi(x).$$

Taking the infimum with respect to x yields $F(\hat{\eta}) \leq F(\eta)$.

Next we show the concavity of the function F. Let η_1 and η_2 be given. Set $\eta_t = (1-t)\eta_1 + t\eta_2$ for $t \in [0, 1]$, then

$$\begin{aligned} F((1-t)\eta_1 + t\eta_2) &= \inf_x \mathcal{J}_{\eta_t}(x) = \inf_x \left\{\phi(x, y^\delta) + ((1-t)\eta_1 + t\eta_2)\psi(x)\right\} \\ &\geq (1-t) \inf_x \left\{\phi(x, y^\delta) + \eta_1\psi(x)\right\} + t \inf_x \left\{\phi(x, y^\delta) + \eta_2\psi(x)\right\} \\ &= (1-t)F(\eta_1) + tF(\eta_2). \end{aligned}$$

Therefore, $F(\eta)$ is concave. □

A direct consequence of concavity is continuity.

Corollary 2.1. *$F(\eta)$ is continuous everywhere.*

Next we examine differentiability of the value function F. To this end, recall first the definition of one-sided derivatives (Dini derivatives) $D^\pm F$ of F

$$D^- F(\eta) = \lim_{h \to 0^+} \tfrac{F(\eta) - F(\eta - h)}{h}, \quad D^+ F(\eta) = \lim_{h \to 0^+} \tfrac{F(\eta + h) - F(\eta)}{h}.$$

The concavity and monotonicity of the function F in Theorem 2.1 ensures the existence of one-sided derivatives $D^\pm F(\eta)$.

Lemma 2.1. *The one-sided derivatives $D^- F$ and $D^+ F$ exist for all $\eta > 0$ and $D^\pm F \geq 0$.*

Proof. For a given η, take any $0 < h_1 < h_2 < \eta$ and set $t = 1 - \frac{h_1}{h_2} < 1$. Then $\eta - h_1 = t\eta + (1-t)(\eta - h_2)$. Now by the concavity of F, we have

$$F(\eta - h_1) \geq tF(\eta) + (1-t)F(\eta - h_2) = \left(1 - \tfrac{h_1}{h_2}\right) F(\eta) + \tfrac{h_1}{h_2} F(\eta - h_2).$$

Rearranging the terms gives

$$\frac{F(\eta) - F(\eta - h_1)}{h_1} \leq \frac{F(\eta) - F(\eta - h_2)}{h_2}.$$

Hence the sequence $\left\{\frac{F(\eta) - F(\eta - h)}{h}\right\}_h$ is monotonically decreasing as h tends to zero and bounded from below by zero, and the limit $\lim_{h \to 0^+} \frac{F(\eta) - F(\eta - h)}{h}$ exists. The existence of $D^+ F$ follows analogously. □

Remark 2.1. *The preceding results do not require the existence of a minimizer to the functional $\mathcal{J}_\eta$, and are valid for any space. All subsequent results remain true in the presence of constraints.*

As a consequence of the definition of one-sided derivatives, we have:

Corollary 2.2. *The one-sided derivatives D^-F and D^+F are left- and right continuous, respectively.*

The solution set $\mathcal{M}_\eta$ might contain multiple elements, i.e. there exist distinct $x_\eta^\delta, \hat{x}_\eta^\delta \in \mathcal{M}_\eta$ such that

$$F(\eta) = \phi(x_\eta^\delta, y^\delta) + \eta\psi(x_\eta^\delta) = \phi(\hat{x}_\eta^\delta, y^\delta) + \eta\psi(\hat{x}_\eta^\delta),$$

and

$$\phi(x_\eta^\delta, y^\delta) < \phi(\hat{x}_\eta^\delta, y^\delta), \quad \psi(x_\eta^\delta) > \psi(\hat{x}_\eta^\delta).$$

In other words, the functions $\phi(x_\eta^\delta, y^\delta)$ and $\psi(x_\eta)$ are potentially multi-valued. Nonetheless, the functions $\phi(x_\eta^\delta, y^\delta)$ and $\psi(x_\eta^\delta)$ are monotone with respect to the regularization parameter η.

Lemma 2.2. *Given $\eta_1, \eta_2 > 0$, if $\mathcal{M}_{\eta_1}$ and $\mathcal{M}_{\eta_2}$ are both nonempty, then for any $x_{\eta_1}^\delta \in \mathcal{M}_{\eta_1}$ and $x_{\eta_2}^\delta \in \mathcal{M}_{\eta_2}$, there hold*

$$(\psi(x_{\eta_1}^\delta) - \psi(x_{\eta_2}^\delta))(\eta_1 - \eta_2) \leq 0, \quad (\phi(x_{\eta_1}^\delta, y^\delta) - \phi(x_{\eta_2}^\delta, y^\delta))(\eta_1 - \eta_2) \geq 0.$$

Proof. The minimizing property of $x_{\eta_1}^\delta$ and $x_{\eta_2}^\delta$ gives

$$\begin{aligned}
\phi(x_{\eta_1}^\delta, y^\delta) + \eta_1\psi(x_{\eta_1}^\delta) &\leq \phi(x_{\eta_2}^\delta, y^\delta) + \eta_1\psi(x_{\eta_2}^\delta), \\
\phi(x_{\eta_2}^\delta, y^\delta) + \eta_2\psi(x_{\eta_2}^\delta) &\leq \phi(x_{\eta_1}^\delta, y^\delta) + \eta_2\psi(x_{\eta_1}^\delta).
\end{aligned}$$

Adding these two inequalities yields the first inequality. The second inequality follows analogously. $\square$

With the help of the function $\psi(x_\eta^\delta)$, the one-sided derivatives $D^\pm F(\eta)$ can be made more precise.

Lemma 2.3. *Suppose that for a given $\eta > 0$ the set $\mathcal{M}_\eta$ is nonempty. Then there hold*

$$D^+F(\eta) \leq \psi(x_\eta^\delta) \leq D^-F(\eta), \quad F(\eta) - \eta D^-F(\eta) \leq \phi(x_\eta^\delta, y^\delta) \leq F(\eta) - \eta D^+F(\eta).$$

Proof. For any $\hat{\eta}$ such that $0 < \hat{\eta} < \eta$, we have

$$F(\hat{\eta}) = \inf_x \mathcal{J}_{\hat{\eta}}(x) \leq \mathcal{J}_{\hat{\eta}}(x_\eta^\delta) = \phi(x_\eta^\delta, y^\delta) + \hat{\eta}\psi(x_\eta^\delta).$$

Therefore, we have

$$F(\eta) - F(\hat{\eta}) \geq \phi(x_\eta^\delta, y^\delta) + \eta\psi(x_\eta^\delta) - \phi(x_\eta^\delta, y^\delta) - \hat{\eta}\psi(x_\eta^\delta) = (\eta - \hat{\eta})\psi(x_\eta^\delta).$$

Thus we obtain

$$\tfrac{F(\eta)-F(\hat{\eta})}{\eta-\hat{\eta}} \geq \psi(x_\eta^\delta).$$

By passing to limit $\hat{\eta} \to \eta$, it follows that $D^-F(\eta) \geq \psi(x_\eta^\delta)$. The inequality $D^+F(\eta) \leq \psi(x_\eta^\delta)$ follows analogously. The remaining assertion follows from these two inequalities and the definition of the value function F. $\square$

The next two results follow directly from the above two lemmas.

Lemma 2.4. *(a) If F' exists at $\eta > 0$, then $\psi(x_\eta^\delta) = F'(\eta)$ and $\phi(x_\eta^\delta, y^\delta) = F(\eta) - \eta F'(\eta)$.*

(b) There exists a countable set $\mathbb{C} \subset \mathbb{R}^+$ such that for any $\eta \in \mathbb{R}^+ \backslash \mathbb{C}$, F is differentiable, $\phi(x_\eta^\delta, y^\delta)$ and $\psi(x_\eta^\delta)$ are continuous and

$$\psi(x_\eta^\delta) = F'(\eta), \quad \phi(x_\eta^\delta, y^\delta) = F(\eta) - \eta F'(\eta).$$

Proof. The assertion (a) is a direct consequence of Lemma 2.3. Assertion (b) follows from the fact that $\psi(x_\eta^\delta)$ is monotone, see Lemma 2.2, and thus there exist at most countable discontinuity points. $\square$

We shall need the next assumption for establishing refined properties of F.

Assumption 2.1. *The functionals $\phi(x, y^\delta)$ and $\psi(x)$ satisfy:*

(a) The functional $\mathcal{J}_\eta$ is coercive for any $\eta > 0$, and any sequence bounded with respect to both ϕ and ψ contains a subsequence converging weakly $*$ in the topology of X.

(b) The functionals ϕ and ψ are weak $*$ lower semicontinuous.

These two assumptions are needed for the existence of a minimizer to $\mathcal{J}_\eta$, and thus not restrictive. Under Assumption 2.1, the attainability of one-sided derivatives $D^\pm F$ can be assured.

Lemma 2.5. *Under Assumption 2.1, there exist $x_\eta^{\delta-}, x_\eta^{\delta+} \in \mathcal{M}_\eta$ such that $D^-F(\eta) = \psi(x_\eta^{\delta-})$ and $D^+F(\eta) = \psi(x_\eta^{\delta+})$ for all $\eta > 0$.*

Proof. Fix $\eta > 0$, and let $h > 0$ be a parameter such that $h \ll \eta$ and $h \to 0$. We shall show that the sequence $\{x_{\eta-h}^\delta\}_h$ contains a minimizing subsequence for the functional $\mathcal{J}_\eta$. By the monotonicity of $F(\eta)$ in Theorem 2.1, we have

$$\mathcal{J}_{\eta-h}(x_{\eta-h}^\delta) = F(\eta - h) \leq F(\eta).$$

Thus $\phi(x_{\eta-h}^\delta, y^\delta) \leq F(\eta)$ and $\psi(x_{\eta-h}^\delta) < \frac{F(\eta)}{\eta-h}$ and thus the sequence $\{x_{\eta-h}^\delta\}_h$ is uniformly bounded by the coercivity of the functional $\mathcal{J}_\eta$. By Assumption 2.1, there exists a subsequence of $\{x_{\eta-h}^\delta\}$, also denoted by $\{x_{\eta-h}^\delta\}_h$, that converges weak $*$ to some $x^* \in X$. Then by the continuity of $F(\eta)$ and weak $*$ lower semicontinuity of ϕ and ψ, it follows that

$$\begin{aligned} F(\eta) = \lim_{h\to 0^+} F(\eta - h) &\geq \liminf_{h\to 0^+} \phi(x_{\eta-h}^\delta, y^\delta) + \eta \liminf_{h\to 0^+} \psi(x_{\eta-h}^\delta) \\ &\geq \phi(x^*, y^\delta) + \eta \liminf_{h\to 0^+} \psi(x_{\eta-h}^\delta) \\ &\geq \phi(x^*, y^\delta) + \eta\psi(x^*) = \mathcal{J}_\eta(x^*) \geq F(\eta). \end{aligned}$$

Consequently, $\psi(x^*) = \liminf_{h\to 0^+} \psi(x_{\eta-h}^\delta) = \lim_{h\to 0^+} \psi(x_{\eta-h}^\delta)$ by Lemma 2.2, and $x^* \in \mathcal{M}_\eta$. It suffices to show that $\psi(x^*) = D^-F(\eta)$. We observe from Lemma 2.3 that

$$\liminf_{h\to 0^+} D^+F(\eta - h) \leq \liminf_{h\to 0^+} \psi(x_{\eta-h}^\delta) \leq \liminf_{h\to 0^+} D^-F(\eta - h) = D^-F(\eta).$$

where we have utilized the left continuity of D^-F, see Corollary 2.2. Using the inequality

$$D^-F(\eta) \leq D^+F(\eta - h),$$

we deduce that $\liminf_{h\to 0^+} D^+F(\eta - h) = D^-F(\eta)$. Consequently, we obtain $\liminf_{h\to 0^+} \psi(x_{\eta-h}^\delta) = D^-F(\eta)$ and $\psi(x^*) = D^-F(\eta)$. Similarly, we can show the existence of a minimizer $x_\eta^{\delta+}$ such that $\psi(x_\eta^{\delta+}) = D^+F(\eta)$. $\square$

Remark 2.2. *Lemma 2.3 implies that $\psi(x_\eta^{\delta-}) = \max_{x\in\mathcal{M}_\eta} \psi(x)$ and $\psi(x_\eta^{\delta+}) = \min_{x\in\mathcal{M}_\eta} \psi(x)$.*

If $\psi(x_\eta) = \psi(\hat{x}_\eta^\delta)$ for all $x_\eta^\delta, \hat{x}_\eta^\delta \in \mathcal{M}_\eta$, then $D^-F(\eta) = D^+F(\eta)$. Consequently, we have

Theorem 2.2. *If Assumption 2.1 holds and the minimizer to $\mathcal{J}_\eta$ is unique for all η, then F is continuously differentiable and for all $\eta > 0$*

$$\psi(x_\eta^\delta) = F'(\eta) \quad \text{and} \quad \phi(x_\eta^\delta, y^\delta) = F(\eta) - \eta F'(\eta).$$

We shall also use the following two results.

Lemma 2.6. *Suppose $F''(\eta)$ exists. Then we have $F''(\eta) \le 0$.*

Proof. It follows from

$$\begin{aligned} F''(\eta) &= \lim_{h\to 0} \frac{F(\eta+h) - 2F(\eta) + F(\eta-h)}{h^2} \\ &= \lim_{h\to 0} \frac{1}{h}\left(\frac{F(\eta+h)-F(\eta)}{h} - \frac{F(\eta)-F(\eta-h)}{h}\right) \\ &= \lim_{h\to 0} \frac{1}{h}\left(D^+F(\eta) - D^-F(\eta)\right) \le 0, \end{aligned}$$

where we have used the fact that $D^+F(\eta) \le D^-F(\eta)$, see the proof of Lemma 2.1. □

Lemma 2.7. *For any η, if both $\phi'(x_\eta^\delta, y^\delta)$ and $\psi'(x_\eta^\delta)$ exist, then there holds*

$$\phi'(x_\eta^\delta, y^\delta) + \eta\psi'(x_\eta^\delta) = 0.$$

Proof. For any $\hat{\eta} < \eta$, the minimizing property of x_η^δ gives

$$\phi(x_\eta^\delta, y^\delta) + \eta\psi(x_\eta^\delta) \le \phi(x_{\hat{\eta}}^\delta, y^\delta) + \eta\psi(x_{\hat{\eta}}^\delta).$$

Consequently, we have

$$\frac{\phi(x_\eta^\delta, y^\delta) - \phi(x_{\hat{\eta}}^\delta, y^\delta)}{\psi(x_\eta^\delta) - \psi(x_{\hat{\eta}}^\delta)} \ge -\eta.$$

Similarly, we can derive

$$\frac{\phi(x_\eta^\delta, y^\delta) - \phi(x_{\hat{\eta}}^\delta, y^\delta)}{\psi(x_\eta^\delta) - \psi(x_{\hat{\eta}}^\delta)} \le -\hat{\eta}.$$

Combining these two inequalities and letting $\hat{\eta} \to \eta$ concludes the proof. □

3 A new criterion

We are now in a position to propose a new criterion for choosing regularization parameters in Tikhonov regularization. The new criterion consists of minimizing the function $\Phi_\gamma(\eta)$ defined as

$$\Phi_\gamma(\eta) = \frac{\gamma^\gamma}{(1+\gamma)^{1+\gamma}} \frac{F^{1+\gamma}(\eta)}{\eta},$$

where γ is a positive constant. The new criterion takes a local minimizer η_γ of the function $\Phi_\gamma(\eta)$ over a certain closed interval in the positive semi-axis $\mathbb{R}^+$ as the regularization parameter. The new criterion is closely related to a criterion due to Regińska [27] (also known as the local minimum criterion), which in our notation (see Theorem 2.2) consists of minimizing the function Ψ defined by

$$\Psi_\gamma(\eta) = (F(\eta) - \eta F'(\eta))^\gamma F'(\eta).$$

The close relation between the two criteria is revealed in the next result.

Proposition 3.1. *Assume $F''(\eta)$ exists and does not vanish. Then the functions Φ_γ and Ψ_γ share the same set of critical points.*

Proof. Observe that

$$\Phi'_\gamma = \frac{\gamma^\gamma}{(1+\gamma)^{1+\gamma}\eta^2} F^\gamma(\eta)[(1+\gamma)\eta F'(\eta) - F(\eta)].$$

Since $F(\eta) > 0$ for all $\eta > 0$, the minimizer η_γ solves the equation

$$(1+\gamma)\eta F'(\eta) - F(\eta) = 0. \tag{2}$$

Next we note that

$$\Psi'_\gamma = -(F(\eta) - \eta F'(\eta))^{\gamma-1} F''(\eta)((1+\gamma)\eta F'(\eta) - F(\eta)),$$

i.e. the regularization parameter determined by the criterion solves also equation (2). □

Note that the equivalence between the proposed criterion Φ_γ and the criterion Ψ_γ requires the existence of a nonvanishing F''. A sufficient condition for the existence of the second derivative and the negativity can be found in [20]. Equation (2) is known as the balancing principle, and has been derived previously in the context of L^1-L^2 formulation using the model function approach [10].

Remark 3.1. *The new criterion Φ_γ makes only use of the value function $F(\eta)$, not of the function $F'(\eta)$, which can be potentially multi-valued in case that the functional $\mathcal{J}_\eta$ has multiple minimizers. In contrast, the value function $F(\eta)$ is always continuous, see Theorem 2.1, and thus the optimization problem of minimizing $\Psi_\gamma(\eta)$ over any bounded closed intervals is always well-defined. The local minimum criterion and balancing principle, i.e. equation* (2)*, are ill-defined, and they are problematic in practical use, for formulations with potentially nonunique minimizers, e.g. L^1-TV and L^2-ℓ^1 with noninjective operator K. Therefore, the proposed criterion Φ_γ is advantageous then.*

Remark 3.2. *Obviously, $\eta = +\infty$ is a global minimizer of the criteria Φ_γ and Ψ_γ. The existence of a finite minimizer to the criterion Ψ_γ is not always guaranteed. Similarly, this is also the case for the proposed criterion Φ_γ. However, this can be remedied by a simple modification*

$$\tilde{\Phi}_\gamma = \frac{\gamma^\gamma}{(1+\gamma)^{1+\gamma}} \frac{(F(\eta) + \beta_0 \eta)^{1+\gamma}}{\eta},$$

where β_0 is a small number. Under the condition $\lim_{\eta\to 0^+} \phi(x_\eta, y^\delta) > 0$, a finite positive minimizer is guaranteed, which follows from

$$\lim_{\eta\to 0^+} \tilde{\Phi}_\gamma(\eta) = +\infty \quad \text{and} \quad \lim_{\eta\to +\infty} \tilde{\Phi}_\gamma(\eta) = +\infty,$$

and the continuity of the value function $F(\eta)$. A similar modification can be made for the local minimum criterion Ψ_γ and the balancing principle. This kind of modification has been previously derived in the Bayesian framework [23].

The remaining parts of this section are devoted to the analysis of the criterion Φ_γ, in particular analytical comparison with Ψ_γ, and a posteriori error estimates.

3.1 Comparison of Φ_γ with Ψ_γ

This part is devoted to the comparison of the proposed criterion Φ_γ with the criterion Ψ_γ, and some properties of the minimizer.

The next result shows a first interesting relation between the proposed and local minimum criteria.

Theorem 3.1. *Let γ be a positive number. Then*

$$\Phi_\gamma(\eta) \le \Psi_\gamma(\eta) \quad \textit{for all } \eta.$$

The equality is achieved if and only if η solves equation (2).

Proof. Recall that for any $a, b \geq 0$ and $p, q > 1$ with $\frac{1}{p} + \frac{1}{q} = 1$, there holds the inequality

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q},$$

with equality holds if and only if $a^p = b^q$. Let $p = \frac{1+\gamma}{\gamma}$ and $q = 1 + \gamma$. For notational simplicity, we denote $\phi(x_\eta, y^\delta)$ and $\psi(x_\eta)$ by ϕ and ψ, respectively. Applying the inequality with $a = \phi^{\frac{\gamma}{1+\gamma}} \eta^{-\frac{\gamma}{2(1+\gamma)}}$ and $b = (\gamma\psi)^{\frac{1}{1+\gamma}} \eta^{\frac{1}{2(1+\gamma)}}$ gives

$$\phi^{\frac{\gamma}{1+\gamma}} (\gamma\psi)^{\frac{1}{1+\gamma}} \eta^{\frac{1-\gamma}{2(1+\gamma)}} \leq \frac{\gamma}{1+\gamma} \frac{\phi + \eta\psi}{\eta^{\frac{1}{2}}} = \frac{\gamma}{1+\gamma} \frac{F(\eta)}{\eta^{\frac{1}{2}}}.$$

Collecting terms in the inequality yields

$$\phi^{\frac{\gamma}{1+\gamma}} \psi^{\frac{1}{1+\gamma}} \leq \gamma^{-\frac{1}{1+\gamma}} \frac{\gamma}{1+\gamma} \frac{F(\eta)}{\eta^{\frac{1}{1+\gamma}}}.$$

Hence, we have

$$\Psi_\gamma(\eta) \leq \frac{\gamma^\gamma}{(1+\gamma)^{1+\gamma}} \frac{F^{1+\gamma}(\eta)}{\eta} = \Phi_\gamma(\eta).$$

The equality holds if and only if $a^p = b^q$, i.e.

$$[\phi^{\frac{\gamma}{1+\gamma}} \eta^{-\frac{\gamma}{2(1+\gamma)}}]^{\frac{1+\gamma}{\gamma}} = [(\gamma\psi)^{\frac{1}{1+\gamma}} \eta^{\frac{1}{2(1+\gamma)}}]^{1+\gamma}.$$

Simplifying this equation yields $\phi - \gamma\eta\psi = 0$, i.e. equation (2). This concludes the proof. $\square$

Theorem 3.1 indicates that the new criterion Φ_γ is sharper than the criterion Ψ_γ, which lends itself to easier numerical implementation as a sharper local minimum is easier to locate numerically. However, it does not give a quantitative measure of the sharpness. To this end, we first observe from Lemma 2.6, the quantity $\frac{\Psi'_\gamma}{\Phi'_\gamma}$ is nonnegative. The next result reveals more clearly the sharpness.

Theorem 3.2. *Assume that F'' exists and is continuous. If a local minimizer η_γ of the function Φ_γ verifies the second order condition, i.e. $\Phi''_\gamma > 0$, then in its neighborhood there holds*

$$\frac{\Psi'_\gamma}{\Phi'_\gamma} < 1.$$

Proof. Direct computation gives

$$\begin{aligned}
\Phi'_\gamma &= \tfrac{\gamma^\gamma}{(1+\gamma)^{1+\gamma}\eta^2} F^\gamma [(1+\gamma)\eta F' - F], \\
\Psi'_\gamma &= -(F - \eta F')^{\gamma-1} F''((1+\gamma)\eta F' - F),
\end{aligned}$$

and

$$\Phi''_\gamma = \tfrac{\gamma^\gamma}{(1+\gamma)^{1+\gamma}} \tfrac{F^{\gamma-1}}{\eta^3} [\gamma(1+\gamma)(\eta F')^2 + (1+\gamma)\eta^2 F'' F - 2(1+\gamma)\eta F' F + 2F^2].$$

Consequently,

$$\begin{aligned}
\tfrac{\eta^2 \Phi''_\gamma}{\Phi_\gamma} &= \gamma(1+\gamma)\tfrac{(\eta F')^2}{F^2} + (1+\gamma)\tfrac{\eta^2 F''}{F} - 2(1+\gamma)\tfrac{\eta F'}{F} + 2 \\
&= \gamma(1+\gamma)\theta^2 + (1+\gamma)\tfrac{\eta^2 F''}{F} - 2(1+\gamma)\theta + 2,
\end{aligned}$$

where $\theta = \frac{\eta F'}{F}$, i.e.

$$-(1+\gamma)\tfrac{\eta^2 F''}{F} = \gamma(1+\gamma)\theta^2 - 2(1+\gamma)\theta + 2 - \tfrac{\eta^2 \Phi''_\gamma}{\Phi_\gamma}.$$

Assisted with this identity, we deduce

$$\begin{aligned}
\frac{\Psi'_\gamma}{\Phi'_\gamma} &= [-F'']\phi^{\gamma-1}\eta^2(\gamma+1)^{\gamma+1}\gamma^{-\gamma}F^{-\gamma}\\
&= \left(1+\tfrac{1}{\gamma}\right)^{\gamma}\cdot(1+\gamma)\tfrac{-\eta^2F''}{F}\cdot\left(\tfrac{\phi}{F}\right)^{\gamma-1}\\
&= \left(1+\tfrac{1}{\gamma}\right)^{\gamma}\cdot\left[\gamma(1+\gamma)\theta^2-2(1+\gamma)\theta+2-\tfrac{\eta^2\Phi''_\gamma}{\Phi_\gamma}\right]\cdot(1-\theta)^{\gamma-1}.
\end{aligned}$$

At a local minimizer η_γ, we have $\theta=\frac{1}{1+\gamma}$, and thus

$$\frac{\Psi'_\gamma}{\Phi'_\gamma}=1-\left(1+\tfrac{1}{\gamma}\right)\frac{\eta^2\Phi''_\gamma}{\Phi_\gamma}<1,$$

by noting the assumption $\Phi''_\gamma(\eta)>0$ at η_γ. The remaining assertion follows from the continuity of F''. ☐

Theorem 3.2 indicates that the criterion Φ_γ is indeed sharper than Ψ_γ in a neighborhood of a local minimum point of Γ_γ. In case that the continuity of F'' does not hold, the inequality $\frac{\Psi'_\gamma}{\Phi'_\gamma}<1$ remains true at the minimizer η_γ.

We can give a more refined sharpness result under further assumptions.

Proposition 3.2. *If both $\phi'(x_\eta^\delta,y^\delta)$ and $\psi'(x_\eta^\delta)$ exist and are continuous, and $\lim_{\eta\to0^+}\phi'(x_\eta^\delta,y^\delta)$ is finite and does not vanish. Then for any $\gamma\geq1$, there holds*

$$\frac{\Psi'_\gamma(\eta)}{\Phi'_\gamma(\eta)}<1,$$

for all η sufficiently small.

Proof. Because $\lim_{\eta\to0^+}\phi'(x_\eta^\delta,y^\delta)$ is finite and does not vanish, and the function $\phi'(x_\eta^\delta,y^\delta)$ is continuous, we deduce that there exists a neighborhood $\mathcal{N}$ of $\eta=0$ and two positive constants c_1 and c_2, such that

$$c_1\leq\phi'(x_\eta^\delta,y^\delta)\leq c_2.$$

Now recall the identity $\phi'(x_\eta^\delta,y^\delta)+\eta\psi'(x_\eta^\delta)=0$, see Lemma 2.7. Consequently, there holds

$$\psi'(x_\eta^\delta)=-\frac{\phi'(x_\eta^\delta,y^\delta)}{\eta},$$

and then integrating the identity from η_0 to η gives

$$\psi(x_\eta^\delta)-\psi(x_{\eta_0}^\delta)\geq c_1\ln\frac{\eta_0}{\eta},$$

where $\eta_0,\eta\in\mathcal{N}$. In particular, this implies that $\psi(x_\eta^\delta)\to+\infty$ as $\eta\to0$. Therefore, for η sufficiently small, we have

$$c_\gamma\psi(x_\eta^\delta)\geq\phi'(x_\eta^\delta,y^\delta),$$

with $c_\gamma=\frac{\gamma^\gamma}{(1+\gamma)^{1+\gamma}}$. Noting the inequality

$$F^{\gamma-1}\geq(F-\eta F')^{\gamma-1}$$

for $\gamma\geq1$, observing that the desired assertion is equivalent to

$$c_\gamma\frac{F^\gamma}{\eta^2}>-(F-\eta F')^{\gamma-1}F'',$$

and the identity $\phi'(x_\eta^\delta,y^\delta)+\eta\psi'(x_\eta^\delta)=0$, we conclude the proof of the proposition. ☐

Proposition 3.3. *Assume that F'' exists and is continuous, and η_γ is a local minimizer of Φ_γ with $\Phi_\gamma''(\eta_\gamma) > 0$. Then there exists a neighborhood $\mathcal{N}(\eta_\gamma)$ of η_γ such that for any $\eta_0, \eta \in \mathcal{N}(\eta_\gamma)$, the following estimate holds*

$$\frac{\phi(x_\eta^\delta, y^\delta)}{\phi(x_{\eta_0}^\delta, y^\delta)} \le \left(\frac{\eta}{\eta_0}\right)^{c(\gamma)},$$

where the constant $c(\gamma) < 1$ depends on γ and $c(\gamma) \to \frac{1}{1+\gamma}$ as the neighborhood shrinks to η_γ.

Proof. From Theorem 3.2 we have

$$\tfrac{\gamma^\gamma}{(1+\gamma)^{1+\gamma}} \tfrac{F^\gamma}{\eta^2} \ge -(F - \eta F')^{\gamma-1} F''.$$

First we recall the identity $\eta\psi'(x_\eta^\delta) + \phi'(x_\eta^\delta, y^\delta) = 0$. At a local minimizer η_γ, we have the relation $\gamma\eta_\gamma\psi(x_{\eta_\gamma}^\delta) = \phi(x_{\eta_\gamma}^\delta, y^\delta)$, see equation (2), and for any $\epsilon > 0$, there exists a neighborhood of η_γ, in which there holds $|\gamma\eta\psi(x_\eta^\delta) - \phi(x_\eta^\delta, y^\delta)| \le \epsilon\phi(x_\eta^\delta, y^\delta)$ by the continuity assumption. Thus the above inequality simplifies to

$$\tfrac{\gamma^\gamma}{(1+\gamma)^{1+\gamma}} \tfrac{((\frac{1+\epsilon}{\gamma}+1)\phi)^\gamma}{\eta} \ge \phi^{\gamma-1}\phi',$$

i.e.

$$c(\gamma)\tfrac{\phi}{\eta} \ge \phi',$$

where $c(\gamma) = \frac{(1+\gamma+\epsilon)^\gamma}{(1+\gamma)^{1+\gamma}} < 1$ for $\epsilon < (1+\gamma)[(1+\gamma)^{\frac{1}{\gamma}} - 1]$. Integrating the inequality from η_0 to η, we have

$$\int_{\eta_0}^{\eta} (\ln(\phi))' d\eta \le c(\gamma) \int_{\eta_0}^{\eta} \tfrac{1}{\eta} d\eta.$$

The desired estimate follows directly from this inequality. □

Next we turn to the differentiability of the criterion Φ_γ. Since the existence of $D^\pm F(\eta)$ are guaranteed at all $\eta > 0$, $D^\pm\Phi_\gamma(\eta)$ also exist for all $\eta > 0$. Moreover we have the following result.

Theorem 3.3. *The inequality $D^-\Phi_\gamma(\eta) \le D^+\Phi_\gamma(\eta)$ holds if and only if $F'(\eta)$ exists.*

Proof. Note that $D^\pm\Phi_\gamma(\eta) = \frac{\gamma^\gamma}{(1+\gamma)^{1+\gamma}\eta^2} F^\gamma[(1+\gamma)\eta D^\pm F - F]$. Then, it is easy to see $D^-\Phi_\gamma(\eta) \le D^+\Phi_\gamma(\eta)$ is equivalent to $D^-F(\eta) \le D^+F(\eta)$. We also know that $D^+F(\eta) \le D^-F(\eta)$ holds for all η, see Lemma 2.3. Therefore the assertion is true. □

The following corollary shows an interesting property of the minimizer to the function Φ_γ.

Corollary 3.1. *At a local minimizer $\eta_\gamma > 0$ to the function Φ_γ, F' exists.*

Proof. Note that $D^-\Phi_\gamma \le D^+\Phi_\gamma$ at a local minimum point. □

Let η_γ be a local minimizer of Φ_γ, and consider a set $\Gamma := \{\eta_\gamma \mid 0 \le \gamma\}$. Corollary 3.1 indicates that F' exists on Γ.

3.2 A posteriori error estimate

In this part, we first recall the so-called Bakushinskiĭ's veto for general heuristic parameter choice rules, i.e. nonconvergence in the worst-case scenario analysis. Then we provide justifications in terms of a posteriori error estimates for the special case of quadratic data fitting, i.e. Y is a Hilbert space, in conjunction with convex variational regularization, i.e. L^2-ψ with ψ being a general convex functional. In particular, we will discuss three cases separately: conventional smoothness regularization; general convex ψ regularization; and sparsity reconstruction $\psi(x) = \|x\|_{\ell^1}$.

Bakushinskiĭ's veto refers to the fact that, theoretically speaking, all heuristic parameter choice rules, which do not make use of the knowledge about the exact noise level δ, suffer from nonconvergence phenomena in the framework of worst-case scenario analysis. This is a consequence of a theorem due to Bakushinskiĭ [1]. To state the theorem, we let G be a mapping from a metric space Y to X.

Definition 3.1. *The function G is termed regularizable in some subset $D \subset Y$ if it is defined in this subset D and the mapping $R(y,\delta) = R_\delta(y)$ exists, acting from Y to X and such that*

$$\lim_{\delta\to 0} \Delta(R_\delta, \delta, y) = 0, \quad \forall y \in D,$$

where the quantity $\Delta(R_\delta, \delta, y)$ is defined as

$$\Delta(R_\delta, \delta, y) = \sup_{y^\delta \in Y, \rho_Y(y^\delta, y) \le \delta} \rho_X(R_\delta(y^\delta), G(y)).$$

The theorem reads as follows. For completeness, we include a short proof.

Theorem 3.4 ([1])**.** *The mapping G is regularizable by the mapping $R(\cdot)$, not dependent explicitly on δ, if and only if G can be extended to all Y and this extension is continuous in D as a mapping defined on all Y.*

Proof. For the sufficiency, we can take $R(\cdot) = G$ on the extension of G on Y and continuous on $D \subset Y$. For necessity, we need only to show that G is continuous on the domain D. This follows from

$$\lim_{\delta\to 0} \rho_X(R_\delta(y), G(y)) = \lim_{\delta\to 0} \rho_X(R(y), G(y)) = 0,$$

by taking $y^\delta = y$ in Definition 3.1 by observing that $y \in \{\tilde{y} \in Y : \rho_Y(\tilde{y}, y) \le \delta\}$ and that R_δ is independent of δ. The theorem is finished by the continuity of $R(\cdot)$. $\square$

Remark 3.3. *Note that in the proof, the condition $y \in \{\tilde{y} \in Y : \rho_Y(\tilde{y}, y) \le \delta\}$ plays an essential role. This stems from the definition $\Delta(R_\delta, \delta, y) = \sup_{y^\delta \in Y, \rho_Y(y^\delta, y) \le \delta} \rho_X(R_\delta(y^\delta), G(y))$ in Definition 3.1. It would be interesting to relax it to $\tilde{\Delta}(R_\delta, \delta, y) = \sup_{y^\delta \in Y, \rho_Y(y^\delta, y) = \delta} \rho_X(R_\delta(y^\delta), G(y))$. It remains unclear whether Theorem 3.4 still holds under the less restrictive assumption.*

Despite the above result, we can still partially justify the proposed criterion by establishing a posteriori error estimate for the case Y is a Hilbert space, i.e., $\phi(x, y^\delta) = \frac{1}{2}\|Kx - y^\delta\|^2$. By a posteriori error estimate, we mean the distance between the approximate solution $x^\delta_{\eta^*}$ and the exact solution $x^\dagger$ in terms of the exact noise level δ, the computable residual $\delta_* = \|Kx^\delta_{\eta^*} - y^\delta\|$ and other relevant quantities. We shall treat the following three cases separately: conventional quadratic regularization, general convex regularization and sparsity regularization due to their distinct features.

We first specialize to the Hilbert space setting. More precisely, we consider $\phi(x, y^\delta) = \frac{1}{2}\|Kx - y^\delta\|^2$ and $\psi(x) = \frac{1}{2}\|x\|^2$ and with η^* chosen by the proposed criterion. To this end, we adopt the general framework of reference [13]. Let $g_\eta(t) = \frac{1}{\eta+t}$ and $r_\eta(t) = 1 - t g_\eta(t) = \frac{\eta}{\eta+t}$, and let $\omega_\mu : (0, \|K\|^2) \to \mathbb{R}$ be such that for all $\gamma \in (0, \gamma_0)$ and $t \in [0, \|K\|^2]$, $t^\mu |r_\gamma(t)| \le \omega_\mu(\gamma)$ holds. Then for $0 < \mu \le 1$, we have $\omega_\mu(\eta) = \eta^\mu$. Moreover, define the source sets $X_{\mu,\rho}$ by $X_{\mu,\rho} := \{x \in X : x = (K^*K)^\mu w,\ \|w\| \le \rho\}$. With these preliminaries, we are ready to state our first result on *a posteriori* error estimates.

Theorem 3.5. *Assume that the exact solution $x^\dagger \in X_{\mu,\rho}$ for some $0 < \mu \le 1$. Let η^* be determined by criterion Φ_γ, and $\delta_* := \|y^\delta - Kx^\delta_{\eta^*}\|$. Then we have*

$$\|x^\dagger - x^\delta_{\eta^*}\| \le C \left(\rho^{\frac{1}{1+2\mu}} + \frac{F(\delta^{\frac{2}{1+2\mu}})^{\frac{1+\gamma}{2}}}{F(\eta^*)^{\frac{1+\gamma}{2}}} \right) \max\{\delta, \delta_*\}^{\frac{2\mu}{1+2\mu}}. \tag{3}$$

Proof. We decompose the error $x^\dagger - x^\delta_\eta$ into

$$x^\dagger - x^\delta_\eta = r_\eta(K^*K)x^\dagger + g_\eta(K^*K)K^*(y^\dagger - y^\delta).$$

Introducing the source representer w with $x^\dagger = (K^*K)^\mu w$, the interpolation inequality gives

$$\begin{aligned}
\|r_\eta(K^*K)x^\dagger\| &= \|r_\eta(K^*K)(K^*K)^\mu w\| \\
&\leq \|(K^*K)^{\frac{1}{2}+\mu} r_\eta(K^*K)w\|^{\frac{2\mu}{2\mu+1}} \|r_\eta(K^*K)w\|^{\frac{1}{2\mu+1}} \\
&= \|r_\eta(KK^*)Kx^\dagger\|^{\frac{2\mu}{2\mu+1}} \|r_\eta(K^*K)w\|^{\frac{1}{2\mu+1}} \\
&\leq c\left(\|r_\eta(KK^*)y^\delta\| + \|r_\eta(KK^*)(y^\delta - y^\dagger)\|\right)^{\frac{2\mu}{2\mu+1}} \|w\|^{\frac{1}{2\mu+1}},
\end{aligned}$$

where the constant c depends only on the maximum of r_η over $[0, \|K\|^2]$. By noting the relation

$$r_{\eta^*}(KK^*)y^\delta = y^\delta - Kx^\delta_{\eta^*},$$

we obtain

$$\|r_{\eta^*}(K^*K)x^\dagger\| \leq c(\delta_* + c\delta)^{\frac{2\mu}{2\mu+1}} \rho^{\frac{1}{2\mu+1}} \leq c_1 \max\{\delta, \delta_*\}^{\frac{2\mu}{2\mu+1}} \rho^{\frac{1}{2\mu+1}}.$$

It remains to estimate the term $\|g_{\eta^*}(K^*K)K^*(y^\delta - y^\dagger)\|$. The standard estimate [13, Theorem 4.2] yields

$$\|g_{\eta^*}(K^*K)K^*(y^\delta - y^\dagger)\| \leq c\frac{\delta}{\sqrt{\eta^*}},$$

However, by the minimizing property of η^*, we have

$$\tfrac{F(\eta^*)^{1+\gamma}}{\eta^*} \leq \tfrac{F(\hat{\eta})^{1+\gamma}}{\hat{\eta}}$$

We may take $\hat{\eta} = \delta^{\frac{2}{1+2\mu}}$, and then we have

$$\frac{1}{\eta^*} \leq \frac{F(\delta^{\frac{2}{1+2\mu}})^{1+\gamma}}{F(\eta^*)^{1+\gamma}} \frac{1}{\delta^{\frac{2}{1+2\mu}}}.$$

Combining the preceding estimates, we arrive at

$$\begin{aligned}
\|x^\dagger - x^\delta_{\eta^*}\| &\leq c_1 \max\{\delta, \delta_*\}^{\frac{2\mu}{2\mu+1}} \rho^{\frac{1}{2\mu+1}} + c\frac{F(\delta^{\frac{2}{1+2\mu}})^{\frac{1+\gamma}{2}}}{F(\eta^*)^{\frac{1+\gamma}{2}}} \delta^{\frac{2\mu}{1+2\mu}} \\
&\leq C\left(\rho^{\frac{1}{1+2\mu}} + \frac{F(\delta^{\frac{2}{1+2\mu}})^{\frac{1+\gamma}{2}}}{F(\eta^*)^{\frac{1+\gamma}{2}}}\right) \max\{\delta, \delta_*\}^{\frac{2\mu}{1+2\mu}}.
\end{aligned}$$

This shows the desired *a posteriori* error estimate. □

Next we present an a posteriori error estimate for L^2-ψ with $\psi(x)$ being convex. We will use the Bregman distance to measure the error. We shall need the concept of a ψ-minimizing solution.

Definition 3.2. *An element $x^\dagger \in X$ is called a ψ-minimizing solution of $Kx^\dagger = y^\dagger$ if*

$$\psi(x^\dagger) \leq \psi(x), \ \forall x \in X \text{ such that } Kx = y^\dagger.$$

Denote the subdifferential of $\psi(x)$ at $x^\dagger$ by $\partial\psi(x^\dagger)$, i.e. $\partial\psi(x^\dagger) = \{\xi \in X^* : \psi(x) \geq \psi(x^\dagger) + \langle \xi, x - x^\dagger\rangle, \forall x \in X\}$, and the Bregman distance $D_\xi(x, x^\dagger)$ by

$$D_\xi(x, x^\dagger) := \left\{\psi(x) - \psi(x^\dagger) - \langle \xi, x - x^\dagger\rangle : \xi \in \partial\psi(x^\dagger)\right\}.$$

We shall often omit the subscript ξ in the Bregman distance $D_\xi(x, x^\dagger)$.

Recall that the optimality condition of x_α reads

$$-\frac{K^*(Kx_\eta - y^\dagger)}{\eta} \in \partial\psi(x_\eta).$$

Let $x_\eta \to x^\dagger$ weakly as $\eta \to 0$. Then we have

$$K^* w \in \partial\psi(x^\dagger),$$

if the limit $\lim_{\eta\to 0} \frac{y^\dagger - Kx_\eta}{\eta}$ exits in weak sense. The latter condition does not hold a priori, and it represents the source condition.

We shall need the following error estimates, which plays the role of a triangle inequality for Bregman distance.

Lemma 3.1. *Let the exact solution $x^\dagger$ fulfill the source condition: there exists an $w \in Y$ such that $\xi = K^* w$ with $\xi \in \partial\psi(x^\dagger)$, and let $\xi_\eta = K^*(y^\dagger - Kx_\eta)/\eta$. Then there holds*

$$\left|D_\xi(x_\eta^\delta, x^\dagger) - (D_{\xi_\eta}(x_\eta^\delta, x_\eta) + D_\xi(x_\eta, x^\dagger))\right| \le 6\|w\|\delta.$$

Proof. The complete proof of the lemma can be found in [22]. The definition of Bregman distance gives

$$\begin{aligned} D_\xi(x_\eta^\delta, x^\dagger) &= D_{\xi_\eta}(x_\eta^\delta, x_\eta) + D_\xi(x_\eta, x^\dagger) + \langle \xi - \xi_\eta, x_\eta - x_\eta^\delta \rangle \\ &= D_{\xi_\eta}(x_\eta^\delta, x_\eta) + D_\xi(x_\eta, x^\dagger) + \langle w + (Kx_\eta - y^\dagger)/\eta, K(x_\eta - x_\eta^\delta)\rangle. \end{aligned}$$

Then it follows that

$$\left|D_\xi(x_\eta^\delta, x^\dagger) - (D_{\xi_\eta}(x_\eta^\delta, x_\eta) + D_\xi(x_\eta, x^\dagger))\right| \le (\|w\| + \|Kx_\eta - y^\dagger\|/\eta)\|K(x_\eta - x_\eta^\delta)\|.$$

The proof is completed by observing the bound for $\|Kx_\eta - y^\dagger\|/\eta$ and $\|K(x_\eta - x_\eta^\delta)\|$ [22]. □

Now we are ready to present another *a posteriori* error estimate.

Theorem 3.6. *Assume that the exact solution $x^\dagger$ satisfies the source condition: there exists a $w \in Y$ such that $K^* w \in \partial\psi(x^\dagger)$. Let $\delta_* = \|Kx_{\eta^*} - y^\delta\|$. Then for each η^* given by the proposed criterion, there holds*

$$D(x_{\eta^*}^\delta, x^\dagger) \le C\left(1 + \frac{F(\delta)^{1+\gamma}}{F(\eta^*)^{1+\gamma}}\right)\max\{\delta, \delta_*\}.$$

Proof. By Lemma 3.1, we have for any η

$$D(x_\eta^\delta, x^\dagger) \le D(x_\eta^\delta, x_\eta) + D(x_\eta, x^\dagger) + 6\|w\|\delta.$$

It suffices to estimate the two terms involving Bregman distance. We first estimate the term $D(x_\eta, x^\dagger)$. To this end, observe by the minimizing property of the x_η, i.e.,

$$\tfrac{1}{2}\|Kx_\eta - y\|^2 + \eta\psi(x_\eta) \le \tfrac{1}{2}\|Kx^\dagger - y^\dagger\|^2 + \eta\psi(x^\dagger) = \eta\psi(x^\dagger),$$

Collecting the terms and noting the definition of $D(x_\eta, x^\dagger)$, we have

$$\tfrac{1}{2}\|Kx_\eta - y^\dagger\|^2 + \eta D(x_\eta, x^\dagger) \le \eta\langle \xi, x - x^\dagger\rangle,$$

which upon utilizing the source condition gives

$$\tfrac{1}{2}\|Kx_\eta - y^\dagger\|^2 + \eta D(x_\eta, x^\dagger) \le -\eta\langle w, K(x_\eta - x^\dagger)\rangle \le \eta\|w\|\|K(x_\eta - x^\dagger)\|.$$

Similarly, by the minimizing property of x_η^δ, we have

$$\tfrac{1}{2}\|Kx_\eta^\delta - y^\delta\|^2 + \eta\psi(x_\eta^\delta) \le \tfrac{1}{2}\|Kx_\eta - y^\delta\|^2 + \eta\psi(x_\eta).$$

From the optimality condition of x_η to $\mathcal{J}_\eta$, we have $\xi' = -\frac{1}{\eta}K^*(Kx_\eta - y^\dagger) \in \partial\psi(x_\eta)$. Therefore, we have

$$\begin{aligned} \tfrac{1}{2}\|Kx_\eta^\delta - y^\delta\|^2 + \eta D(x_\eta^\delta, x_\eta) &\le \tfrac{1}{2}\|Kx_\eta - y^\delta\|^2 - \eta\langle \xi', x_\eta^\delta - x_\eta\rangle \\ &= \tfrac{1}{2}\|Kx_\eta - y^\delta\|^2 + \langle Kx_\eta - y^\dagger, K(x_\eta^\delta - x_\eta)\rangle \\ &= \tfrac{1}{2}\|Kx_\eta^\delta - y^\delta\|^2 - \tfrac{1}{2}\|K(x_\eta^\delta - x_\eta)\|^2 - \langle y^\dagger - y^\delta, K(x_\eta - x_\eta^\delta)\rangle. \end{aligned}$$

Consequently, we have

$$\begin{aligned}\tfrac{1}{2}\|K(x_\eta^\delta - x_\eta)\|^2 + \eta D(x_\eta^\delta, x_\eta) &\leq -\langle y^\dagger - y^\delta, K(x_\eta - x_\eta^\delta)\rangle \\ &\leq \|y^\dagger - y^\delta\|\|K(x_\eta - x_\eta^\delta)\|.\end{aligned} \tag{4}$$

With the help of these two estimates, we have

$$\begin{aligned}D(x_{\eta^*}, x^\dagger) &\leq \|w\|\|Kx_{\eta^*} - y^\dagger\| \\ &\leq \|w\|(\|K(x_{\eta^*} - x_{\eta^*}^\delta)\| + \|Kx_{\eta^*}^\delta - y^\delta\| + \|y^\delta - y^\dagger\|) \\ &\leq \|w\|(2\delta + \delta_* + \delta) \leq 4\|w\| \max(\delta, \delta_*).\end{aligned}$$

Next we estimate the term $D(x_\eta^\delta, x_\eta)$. By inequality (4), we have

$$D(x_{\eta^*}^\delta, x_{\eta^*}) \leq \frac{\delta^2}{2\eta^*}.$$

Recalling the minimizing property of η^*, we have for $\hat{\eta}$

$$\frac{F(\eta^*)^{1+\gamma}}{\eta^*} \leq \frac{F(\hat{\eta})^{1+\gamma}}{\hat{\eta}}, \quad \text{i.e.} \quad \frac{1}{\eta^*} \leq \frac{F(\hat{\eta})^{1+\gamma}}{F(\eta^*)^{1+\gamma}} \frac{1}{\hat{\eta}}.$$

We may take $\hat{\eta} = \delta$ and combine the above two inequalities to get

$$D(x_{\eta^*}^\delta, x_{\eta^*}) \leq \frac{F(\delta)^{1+\gamma}}{F(\eta^*)^{1+\gamma}} \frac{\delta}{2}.$$

Now summarizing the three estimates gives

$$D(x_{\eta^*}^\delta, x^\dagger) \leq C\left(1 + \frac{F(\delta)^{1+\gamma}}{F(\eta^*)^{1+\gamma}}\right) \max\{\delta, \delta_*\},$$

with $C = \max\{10\|w\|, \frac{1}{2}\}$. This concludes the proof of the theorem. □

In case that the regularization functional ψ is uniformly convex functionals, e.g. $\frac{1}{p}\|\cdot\|_{\ell^p}^p$ and $\|\cdot\|_{W^{k,p}}(k \geq 0, p > 1)$ [3], the above result directly yields a convergence rate in norms. However, the interesting case of ℓ^1 regularization, i.e. $X = \ell^2$ and $\psi(x) = \|x\|_{\ell^1}$ is not covered, as the Bregman distance can vanish for distinct x and x'. This can be remedied as below. To this end, we recall the following result [15].

Lemma 3.2. *Let $X = \ell^2$ and $\psi(x) = \|x\|_{\ell^1}$. Assume that the solution $x^\dagger$ is sparse and satisfies the source condition: there exists a $w \in Y$ such that $K^*w \in \partial\psi(x^\dagger)$, and the operator K satisfies for any finitely supported x_1 and x_2, there holds $Kx_1 = Kx_2$ implies $x_1 = x_2$. Then there exist two positive constants c_1 and c_2 such that*

$$\|x - x^\dagger\|_{\ell^2} \leq c_1[\psi(x) - \psi(x^\dagger)] + c_2\|K(x - x^\dagger)\|.$$

We can now state a third a posteriori error estimate.

Theorem 3.7. *Assume that the conditions in Lemma 3.2 are satisfied, and η^* is determined by the criterion Φ_γ. Then there exists a constant $C > 0$ such that*

$$\|x_{\eta^*}^\delta - x^\dagger\| \leq C\left(1 + \frac{F(\delta)^{1+\gamma}}{F(\eta^*)^{1+\gamma}}\right) \max\{\delta, \delta_*\}.$$

Proof. By Lemma 3.2, the definition of Bregman distance $D(x, x^\dagger)$, the source condition and the Cauchy-Schwarz inequality, we have

$$\begin{aligned}\|x - x^\dagger\| &\le c_1[\psi(x) - \psi(x^\dagger)] + c_2\|K(x - x^\dagger)\| \\ &= c_1 D(x, x^\dagger) + c_1\langle \xi, x - x^\dagger\rangle + c_2\|K(x - x^\dagger)\| \\ &= c_1 D(x, x^\dagger) + c_1\langle w, K(x - x^\dagger)\rangle + c_2\|K(x - x^\dagger)\| \\ &\le c_1 D(x, x^\dagger) + (c_1\|w\| + c_2)\|K(x - x^\dagger)\|.\end{aligned}$$

Now by virtue of Lemma 3.1, we have

$$\|x_{\eta^*}^\delta - x^\dagger\| \le c_1(D(x_{\eta^*}^\delta, x_{\eta^*}) + D(x_{\eta^*}, y^\delta) + 6\|w\|\delta) + (c_1\|w\| + c_2)\|K(x_{\eta^*}^\delta - x^\dagger)\|.$$

Next we bound each term on the right hand side. First observe

$$\|K(x_{\eta^*}^\delta - x^\dagger)\| \le \|Kx_{\eta^*}^\delta - y^\delta\| + \|y^\delta - Kx^\dagger\| \le \delta_* + \delta \le 2\max\{\delta, \delta_*\}.$$

Then for the term $D(x_{\eta^*}, x^\dagger)$, we obtain as before

$$\begin{aligned}D(x_{\eta^*}, x^\dagger) &\le \|w\|\|K(x_{\eta^*} - x^\dagger)\| \\ &\le \|w\|(\|K(x_{\eta^*} - x_{\eta^*}^\delta)\| + \|Kx_{\eta^*}^\delta - y^\delta\| + \delta) \\ &\le \|w\|(2\delta + \delta_* + \delta) \le 4\|w\|\max\{\delta, \delta_*\}.\end{aligned}$$

Finally, for the term $D(x_{\eta^*}^\delta, x_{\eta^*})$, we proceed as before to obtain

$$D(x_{\eta^*}^\delta, x_{\eta^*}) \le \frac{F(\delta)^{1+\gamma}}{F(\eta^*)^{1+\gamma}}\frac{\delta}{2}.$$

Now combining the above three estimates gives

$$D(x_\eta^\delta, x^\dagger) \le C\left(1 + \frac{F(\delta)^{1+\gamma}}{F(\eta^*)^{1+\gamma}}\right)\max\{\delta, \delta_*\},$$

with $C = \max\{12c_1\|w\| + c_2, \frac{c_1}{2}\}$. This completes the proof of the theorem. □

The preceding three theorems indicate that the approximation $x_{\eta^*}^\delta$ with η^* chosen by the rule Φ_γ converges to the exact solution $x^\dagger$ at an identical rate for a priori rule and discrepancy principle under the same source condition [4, 23], so long as the discrepancy δ^* is of order δ.

4 Numerical algorithm

We propose to compute a minimizer of the function Φ_γ by the following simple fixed point iteration algorithm.

`Fixed point algorithm`

(i) Set $k = 0$ and choose η_0.

(ii) Solve for x_{k+1} by the Tikhonov regularization method

$$x_{k+1} = \arg\min_x \left\{\phi(x, y^\delta) + \eta_k\psi(x)\right\}.$$

(iii) Update the regularization parameter η_{k+1} by

$$\eta_{k+1} = \frac{1}{\gamma} \frac{\phi(x_{k+1}, y^\delta)}{\psi(x_{k+1})}.$$

(iv) Check the stopping criterion. If not converged, set $k = k + 1$ and repeat from Step (ii).

We would like to point out that we have not specified the solver for Tikhonov regularization problems arising in Step (ii). The problem *per se* may be approximately solved with an iterative algorithm, which was adopted in our numerical experiments. Numerically, we have observed that it will not jeopardize the steady convergence of the fixed point algorithm so long as the the regularization problem is solved with reasonable accuracy.

The following lemma provides an interesting and practically very important observation on the monotonicity of the sequence $\{\eta_k\}$ generated by fixed point algorithm, and the monotonicity is key to the demonstration of the convergence of the algorithm. We introduce $r(\eta) = \phi(x_\eta, y^\delta) - \gamma\eta\psi(x_\eta)$.

Lemma 4.1. *For any initial guess η_0, the sequence $\{\eta_k\}_k$ generated by the algorithm is monotone. Moreover, the sequence is monotonically decreasing (increasing) if $r(\eta_0) < 0$ $(r(\eta_0) > 0)$.*

Proof. By the definition of η_k, we have

$$\begin{aligned} \eta_k - \eta_{k-1} &= \gamma^{-1} \tfrac{\phi(x_k, y^\delta)}{\psi(x_k)} - \gamma^{-1} \tfrac{\phi(x_{k-1}, y^\delta)}{\psi(x_{k-1})} \\ &= \tfrac{\phi(x_k)\psi(x_{k-1}) - \phi(x_{k-1})\psi(x_k)}{\gamma\psi(x_{k-1})\psi(x_k)} \\ &= \tfrac{[\phi(x_k) - \phi(x_{k-1})]\psi(x_{k-1}) + \phi(x_{k-1})[\psi(x_{k-1}) - \psi(x_k)]}{\gamma\psi(x_{k-1})\psi(x_k)} \end{aligned}$$

We assume that $\eta_k \neq \eta_{k-1}$, otherwise it is trivial. Lemma 2.2 indicates that each term is of the same sign with $\eta_{k-1} - \eta_{k-2}$, and the sequence $\{\eta_k\}_k$ is monotone. Next observe that if $r(\eta_0) > 0$, there holds

$$\eta_1 = \tfrac{\phi(x_{\eta_0}, y^\delta)}{\gamma\psi(x_{\eta_0})} > \eta_0.$$

The remaining assertion follows from this directly. □

Remark 4.1. *In the lemma, the uniqueness of the solution x_η is not required. The lemma holds for any x_η belongs to the set $\mathcal{M}_\eta$ of minimizers of $\mathcal{J}_\eta$.*

We note that in Lemma 4.4, the sequence can diverge to $+\infty$, i.e. the convergence in general can only be understood in a generalized sense. Nonetheless, the convergence can be ensured if the initial guess η_0 satisfies $r(\eta_0) < 0$.

Theorem 4.1. *Assume that the initial guess η_0 satisfies $r(\eta_0) < 0$. The sequence $\{\eta_k\}_k$ converges.*

Proof. By Lemma 4.1, the sequence $\{\eta_k\}_k$ is monotonically decreasing, and it is bounded from below by zero. Therefore, it converges. □

Remark 4.2. *Assume that F'' exists. Then the asymptotic convergence rate r^* of the algorithm is dictated by*

$$\begin{aligned} r^* &:= \lim_{k\to\infty} \tfrac{\eta^* - \eta_{k+1}}{\eta^* - \eta_k} = \tfrac{d}{d\eta}\gamma^{-1} \tfrac{\phi(x_\eta, y^\delta)}{\psi(x_\eta)}|_{\eta=\eta^*} = \gamma^{-1} \tfrac{d}{d\eta} \tfrac{[F(\eta) - \eta F'(\eta)]}{F'(\eta)}|_{\eta=\eta^*} \\ &= \gamma^{-1} \tfrac{-F''(\eta^*)[\eta^* F'(\eta^*) + F(\eta^*) - \eta^* F'(\eta^*)]}{(F'(\eta^*))^2} \\ &= -\gamma^{-1} \tfrac{F''(\eta^*)}{\psi(x_{\eta^*})^2} \left[\eta^* \psi(x_{\eta^*}) + \phi(x_{\eta^*}, y^\delta)\right] = -\tfrac{F''(x_{\eta^*}) F(x_{\eta^*})}{\gamma\psi(x_{\eta^*})^2}. \end{aligned}$$

By noting the relation $\gamma\eta^*\psi(x_{\eta^*}) = \phi(x_{\eta^*}, y^\delta)$ at a critical point η^*, we derive

$$r^* = (1 + \gamma^{-1}) \tfrac{-\eta^* F''(\eta^*)}{\psi(x_{\eta^*})}.$$

The monotonicity of the sequence $\{\eta_k\}_k$ implies that $r^* \leq 1$, however, a precise estimate of the rate r^* is still missing. Nonetheless, a fast convergence of the algorithm is always numerically observed.

Next we examine the descent property of sequence $\{\eta_k\}_k$ for the function Φ_γ. This is not evident as the fixed point algorithm involves only the necessary optimality condition, see equation (2). To simplify notation, we shall denote by T the operator

$$T(\eta) = \gamma^{-1} \tfrac{\phi(x_\eta^\delta, y^\delta)}{\psi(x_\eta^\delta)}.$$

The next result shows the monotonicity of the operator T.

Lemma 4.2. *The operator T is monotone in the sense that if $0 < \eta_0 < \eta_1$, then*

$$T(\eta_0) \le T(\eta_1). \tag{5}$$

Proof. By the monotonicity of ϕ and ψ with respect to η, see Lemma 2.2, we have

$$\phi(x_{\eta_0}^\delta, y^\delta) \le \phi(x_{\eta_1}^\delta, y^\delta), \quad \psi(x_{\eta_0}^\delta) \ge \psi(x_{\eta_1}^\delta). \tag{6}$$

The result now follows directly from the definition of the operator T. $\square$

Lemma 4.3. *For any η_0, the sequence $\{T^k\eta_0\}_k$ is either strictly monotone or there exists some k_0 such that $\{T^k\eta_0\}_{k=0}^{k_0}$ is strictly monotone and $T^k\eta_0 = T^{k_0}\eta_0$ for all $k \ge k_0$.*

Proof. By Lemma 2.2, the sequence $\{T^k\eta_0\}_k$ is always monotone. Without loss of generality, we may assume that it is monotonically increasing. If the sequence is not strictly monotonically increasing, then there exists a smallest positive integer $k_0 \in \mathbb{N}$ such that

$$T^{k_0}\eta_0 = T^{k_0+1}\eta_0.$$

Applying the operator T repeated on the identity concludes the proof of the assertion. $\square$

We shall also need a "sign-preserving" property of the operator T: the function $r(\eta)$ cannot vanish or change sign on the open interval between η_0 and the limit η^* of the sequence $\{T^k(\eta_0)\}_k$.

Lemma 4.4. *Let η_0 be such that $\{T^k(\eta_0)\}$ converges to η^*. Then the function $r(\eta)$ can't vanish or change sign over the open interval $(\min(\eta_0, \eta^*), \max(\eta_0, \eta^*))$.*

Proof. Without loss of generality we assume that $\eta_0 < T(\eta_0)$, or equivalently $r(\eta_0) > 0$, as the other case can be treated similarly. Assume that the assertion is false, i.e. the function $r(\eta)$ vanishes or changes sign over the open interval $(\min(\eta_0, \eta^*), \max(\eta_0, \eta^*))$. Therefore, there exists a $\hat{\eta} \in (\eta_0, \eta^*)$ such that $r(\hat{\eta}) \le 0$, i.e. $T(\hat{\eta}) \le \hat{\eta}$, by Lemma 4.1. Appealing again to Lemmas 4.1 and 4.3, there exists some $k \in \mathbb{N}$ such that

$$T^k(\eta_0) \le \hat{\eta} < T^{k+1}(\eta_0).$$

However, by Lemma 4.2, we have

$$\hat{\eta} < T^{k+1}(\eta_0) \le T(\hat{\eta}) \le T^{k+2}(\eta_0),$$

which is a contradiction to $T(\hat{\eta}) \le \hat{\eta}$. $\square$

Now we are in a position to the descent property of the fixed point algorithm.

Theorem 4.2. *The sequence $\{\eta_k\}_k$ is descent for the function Φ_γ.*

Proof. By Theorem 2.1, the function F is monotone. Thus it is almost everywhere differentiable. Therefore, for almost every $\eta \in \mathbb{R}^+$, we have

$$\Phi'_\gamma(\eta) = \tfrac{\gamma^\gamma}{(1+\gamma)^{1+\gamma}} \tfrac{F^\gamma(\eta)}{\eta^2}(-r(\eta)).$$

By Lemma 4.4, the function $r(\eta)$ remains the same sign over the interval $(\min(\eta_0, \eta^*), \max(\eta_0, \eta^*))$. Without loss of generality we may assume the sequence is increasing, i.e. $r(\eta) > 0$. By the monotonicity of the functions ϕ and ψ, there are at most countable discontinuity points, where Φ'_γ has only bounded jumps, over the interval. Therefore,

$$\Phi_\gamma(\eta_{k+1}) = \Phi(\eta_k) + \int_{\eta_k}^{\eta_{k+1}} \Phi'_\gamma(\eta) < \Phi(\eta_k).$$

This shows the desired the descent property. □

Remark 4.3. *Under further conditions on the value function F, e.g. the second order derivative F'' exists and continuous, then an analogous descent property holds also for the function Ψ.*

By the strict descent property of the sequence $\{\eta_k\}$ for the function Φ_γ, we have the next result.

Proposition 4.1. *The limit η^* of a sequence $\{\eta_k\}$ by the fixed point algorithm is a critical point of the function Φ_γ.*

Interestingly, the asymptotic convergence rate r^* in Remark 4.2 is intimately connected with the minimizing property of η^* to $\Phi_\gamma(\eta)$, as revealed in the following proposition.

Proposition 4.2. *Assume that F'' exists. Let η^* be a critical point to the function Φ_γ such that Φ_γ verifies the second-order derivative test and $F''(\eta^*) \neq 0$. Then $r^* < 1$ if and only if η^* is a local minimizer.*

Proof. The second-order derivative of $\Phi_\gamma(\eta)$ at η^* is computed as

$$\Phi''_\gamma(\eta^*) = \tfrac{\gamma^\gamma}{(1+\gamma)^{1+\gamma}} \tfrac{F^\gamma(\eta^*)}{(\eta^*)^2} [\gamma F'(\eta^*) + (1+\gamma)\eta^* F''(\eta^*)] > 0.$$

Lemma 2.2 indicates that $F'(\eta) = \psi(x_\eta^\delta)$ is non-increasing with respect to η, thus the function $F''(\eta)$ is always non-positive. The first term in the square bracket is always positive. Under the assumption that Φ_γ verifies the second-order derivative test at extrema and $F''(\eta^*) \neq 0$, we deduce that a critical point η^* is a local minimizer of $\Phi_\gamma(\eta)$ if

$$\gamma F'(\eta^*) + (1+\gamma)\eta^* F''(\eta^*) > 0.$$

By rearranging the terms, we arrive at

$$r^* = (1 + \gamma^{-1}) \tfrac{-\eta^* F''(\eta^*)}{F'(\eta^*)} < 1.$$

The converse follows easily by reversing all the steps. □

5 Numerical results

This section presents numerical results for several benchmark linear inverse problems, which are adapted from Hansen's package `Regularization Tool` [16] and range from mild to severe ill-posedness, by nonsmooth Tikhonov formulations, to illustrate features of the new criterion. These are Fredholm integral equations of the first kind with kernel $k(s,t)$ and solution $x(t)$. The discretized linear system takes the form $\mathbf{Kx} = \mathbf{y}^\delta$, and is of size 300×300. The regularizing functional is referred to as ϕ-ψ type, e.g. L^1-TV denotes the one with L^1 data-fitting and TV regularization.

Unless otherwise specified, the data is contaminated by additive Gaussian noise, i.e. $y_i^\delta = y_i + \max_i\{|y_i|\}\varepsilon\xi_i$, where ξ_i are standard normal random variables, and ε refers to the relative noise level. Six noise levels, i.e. $\varepsilon \in \{5\times10^{-2}, 1\times10^{-2}, 1\times10^{-3}, 1\times10^{-4}, 1\times10^{-5}, 1\times10^{-6}\}$, are considered. In the fixed point algorithm, the initial guess η_0 is taken to be 1×10^{-3}, and it is stopped if the relative change of η is smaller than 1×10^{-3}. The nonsmooth minimization problems arising from L^2-TV, L^2-ℓ^1

Table 1: Numerical results for Example 1 with different noise levels.

ε	δ_t	δ_b	γ	η_b	e_b	η_o	e_o
5e-2	3.12e-1	3.15e-1	5.46e0	1.11e-4	8.74e-2	4.04e-6	2.58e-2
1e-2	6.24e-2	6.23e-2	2.44e0	7.16e-6	1.60e-2	6.25e-7	7.71e-3
1e-3	6.24e-3	6.21e-3	7.68e-1	2.13e-7	4.15e-3	7.20e-8	3.35e-3
1e-4	6.24e-4	6.19e-4	2.41e-1	6.65e-9	1.68e-3	1.68e-9	1.44e-3
1e-5	6.24e-5	6.15e-5	7.58e-2	2.08e-10	5.25e-4	8.29e-11	5.88e-4
1e-6	6.24e-6	1.03e-5	2.50e-2	1.57e-11	2.77e-3	3.91e-11	4.74e-4

and L^1-TV formulations are solved by the iteratively reweighted least-squares method. The accuracy of the reconstruction x_η is measured by the relative error $e = \|x_\eta^\delta - x^\dagger\|_{L^2}/\|x^\dagger\|_{L^2}$.

Note that as a consequence of Theorem 3.4, heuristic choice rules suffer from nonconvergence. In particular, the parameter γ in the criteria Φ_γ and Ψ_γ should depend on the noise level δ, even though the dependence can be rather weak. To the best of our knowledge, there is no known rule for choosing an appropriate value for the parameter γ in the criterion Ψ_γ since its appearance. To remedy this difficulty, we propose the following two-step procedure for determining γ.

`Two-step procedure`

(a) Give γ_0 and η_0.

(b) Run the fixed point algorithm with γ_0 and η_0 until convergence, with limit $\tilde{\eta}$.

(c) Calculate the functional value $\phi(x_{\tilde{\eta}}^\delta, y^\delta)$.

(d) Set γ as follows

$$\gamma = \gamma_0 \left(\frac{\phi(x_{\tilde{\eta}}^\delta, y^\delta)}{0.05\phi(0, y^\delta)} \right)^d,$$

with $d = \frac{1}{4}$ and $d = \frac{1}{2}$ for L^2 and L^1 data fitting, respectively. Unless otherwise stated, $\gamma_0 = 10$.

5.1 L^2-H^1 with constraint

Example 1 (gravity surveying [16] with nonnegativity constraint)**.** *The functions k and x are given by $k(s,t) = \frac{1}{4}\left(\frac{1}{16} + (s-t)^2\right)^{-\frac{3}{2}}$ and $x(t) = \sin(\pi t) + \frac{1}{2}\sin(2\pi t)$, respectively, and the integration interval is* $[0,1]$.

The function x stands for material density, and thus it is natural to enforce nonnegativity constraint. The constrained optimization problems are solved by built-in `MATLAB` function `quadprog`. The problem is severely ill-posed with a condition number 4.54×10^{19}.

The numerical results for Example 1 with various noise levels are shown in Table 1. In the table, δ_t and δ_b stand for the true noise level and the one estimated by computing $\phi(x_{\eta^*}^\delta, y^\delta)$. The value for γ is determined by the two-step procedure. The subscripts b and o respectively denote the result by the proposed criterion and the optimal one, which is determined by calculating the solution for a range of regularization parameters and then selecting the one with smallest relative error. Firstly, we observe that the estimated noise level δ_b is in excellent agreements with the exact noise level. Secondly, the parameter γ adapts itself automatically to noise level: its value decreases as the noise level δ_t decreases. With this automatic adaption of γ, the regularization parameter determined by the proposed criterion, i.e. η_b, is close to the optimal one η_o, for all six noise levels under consideration. Consequently, the respective reconstructions are accurate, and within a factor of three compared to the optimal one. Fig. 1(c) shows the reconstruction given by the proposed criterion for 5% noise, which is fairly close to the exact one. We

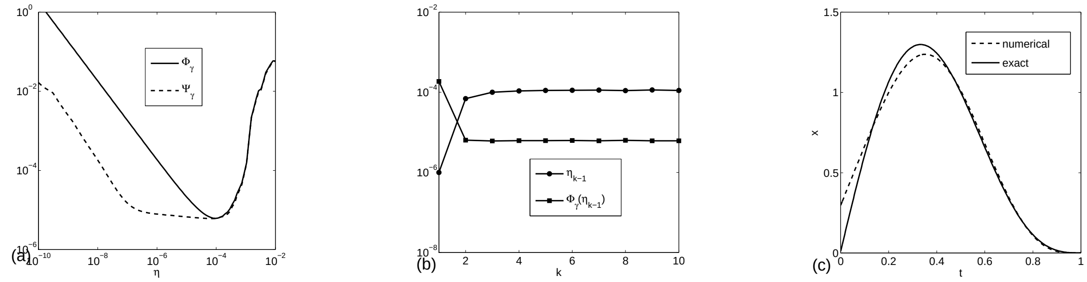


Figure 1: Numerical results for Example 1 with 5% noise: (a) Φ_γ v.s. Ψ_γ with $\gamma = 5.46$, (b) convergence of algorithm, and (c) reconstruction.

Table 2: Numerical results for Example 2 with different noise levels.

ε	δ_t	δ_b	γ	η_b	e_b	η_o	e_o
5e-2	2.52e-1	2.51e-1	5.38e0	3.00e-3	2.41e-1	1.75e-4	2.16e-1
1e-2	5.05e-2	5.02e-2	2.40e0	2.64e-4	1.28e-1	7.20e-5	1.12e-1
1e-3	5.05e-3	5.02e-3	7.59e-1	8.31e-6	9.39e-2	8.68e-6	9.39e-2
1e-4	5.05e-4	5.02e-4	2.39e-1	2.63e-7	7.97e-2	1.84e-7	7.95e-2
1e-5	5.05e-5	5.00e-5	7.57e-2	8.27e-9	7.53e-2	5.96e-8	7.36e-2
1e-6	5.05e-6	4.98e-6	2.39e-2	2.61e-10	6.60e-2	1.84e-11	6.32e-2

would like to point out that the reconstruction for fixed γ without automatic adapting tends to oscillate as the noise level decreases to zero. Therefore, the two-step procedure is necessary.

In Section 3.1, we have analytically compared the proposed criterion with Regińska's criterion, and concluded that the former is sharper than the latter. We illustrate this numerically, see Fig. 1(a). The curve Ψ_γ is relatively flat in the region $[1 \times 10^{-7}, 2 \times 10^{-3}]$, although there is a local minimum. In contrast, the function Φ_γ has a distinct minimum at 1.11×10^{-4}. The descent property of sequence $\{\eta_k\}$ for the function Φ_γ is shown in Fig. 1(b). A rapid and steady convergence of the fixed point algorithm is also observed: the iterate $\{\eta_k\}$ is monotonic and the convergence is numerically achieved within four iterations. Similar convergence behavior is observed for other noise levels.

5.2 L^2-TV

Example 2 (TV reconstruction, adapted from `shaw` [16]). *The functions k and x are given by $k(s,t) = (\cos s + \cos t)\left(\frac{\sin u}{u}\right)^2$ with $u(s,t) = \pi(\sin s + \sin t)$ and $x(t) = \chi_{[-\frac{\pi}{6}, \frac{\pi}{6}]}$ with χ being the indicator function, respectively and the integration interval is $[-\frac{\pi}{2}, \frac{\pi}{2}]$.*

The problem is severely ill-posed with a condition number 4.19×10^{20}. The exact solution x is piecewise constant, and thus TV regularization is deemed suitable. The numerical results are summarized in Table 2. Again the estimated noise level δ_b represents an excellent approximation to the exact one δ_t. The accuracy e_b of the reconstructions improves as the noise level δ_t decreases, and it is comparable with the optimal one. The reconstructed profile remains accurate and stable for ε up to 5%, see Fig. 2(c). Note that it exhibits typical stair-casing effect of TV regularization.

For the L^2-TV formulation, the criterion Ψ_γ has only an ambiguous local minimum, see Fig. 2(a). The curve is very flat over a wide range, which can potentially cause numerical problems to some algorithms for locating the minimum with gradient-type algorithm. In sharp contrast, the proposed criterion still has a very distinct minimum. This numerically validates the theoretical results in Section 3.1. A fast and steady convergence of the fixed point algorithm is again observed, and the convergence is reached within four iterations. Therefore, it is computationally very efficient. Numerically, we found that at Step (b) of

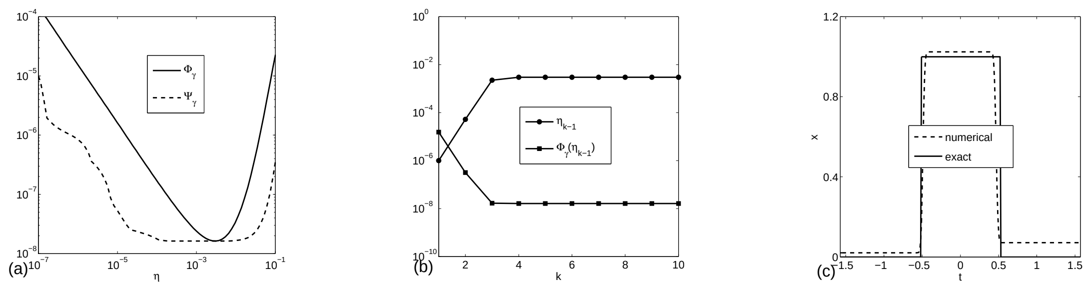


Figure 2: Numerical results for Example 2 with 5% noise: (a) Φ_γ v.s. Ψ_γ with $\gamma = 5.38$, (b) convergence of algorithm, and (c) reconstruction.

Table 3: Numerical results for Example 3.

ε	δ_t	δ_b	γ	η_b	e_b	η_o	e_o
5e-2	2.74e-2	2.69e-2	5.70e-1	2.63e-3	6.39e-1	8.11e-4	5.99e-1
1e-2	5.49e-3	5.36e-3	2.55e-1	2.33e-4	4.61e-1	3.39e-4	4.59e-1
1e-3	5.49e-4	5.33e-4	8.01e-2	7.38e-6	3.94e-1	1.00e-5	3.90e-1
1e-4	5.49e-5	5.31e-5	2.49e-2	2.36e-7	1.98e-1	2.95e-9	1.32e-1
1e-5	5.49e-6	5.27e-6	7.30e-3	7.92e-9	3.37e-2	3.09e-10	1.32e-2
1e-6	5.49e-7	4.75e-7	3.66e-2	1.28e-11	2.91e-3	1.76e-10	2.18e-3

the two-step procedure, there is no need to let the fixed point algorithm run to convergence. One or two fixed point iterations would suffice the goal of adapting the parameter γ.

5.3 L^2-ℓ^1

Example 3 (Sparse reconstruction, adapted from `phillips` [16]). *Let* $\varphi(t) = \left[1 + \cos\frac{\pi t}{3}\right]\chi_{|t-s|<3}$, *and* $S = [-3, -2.96] \cup [0.6, 0.64] \cup [3, 3.04]$. *The functions* k *and* x *are given by* $k(s,t) = \varphi(s-t)$ *and* $x(t) = \chi_S$, *respectively, and the integration interval is* $[-6, 6]$.

The problem is mildly ill-posed with a condition number 2.14×10^8. The exact solution consists of three small spikes and has a sparse representation with respect to pixel basis, and thus ℓ^1 regularization is suitable. For this problem, γ_0 is set to 1. The numerical results are show in Table 3 and Fig 3. The numerical solution shows the feature of sparsity-promoting ℓ^1 regularization: the locations of the all three small spikes are perfectly detected, and the retrieved magnitudes are acceptable taking into account the large amount of noise.

Again, the criterion Ψ_γ is very flat over some region, whereas the criterion Φ_γ exhibits a clear local minimum over there. The algorithm converges steadily and quickly within six iterations for all five noise levels, and function Φ_γ decreases monotonically as the iteration proceeds.

5.4 L^1-TV

Example 4 (Deblurring 1D image). *The functions* k *and* x *are given by* $k(s,t) = \frac{1}{4\sqrt{2\pi}} e^{-\frac{(s-t)^2}{0.0032}} \chi_{|s-t|<0.15}$ *and* $x(t) = \chi_{[.34,.66]}$, *respectively, and the integration interval is* $[0, 1]$.

The problem is mildly ill-posed with a condition number 7.35×10^7. The data is corrupted by additive random valued impulsive noise as follows

$$y_i^\delta = \begin{cases} y_i + \sigma_m \xi_i, & \text{with probability } q, \\ y_i, & \text{with probability } 1-q, \end{cases}$$

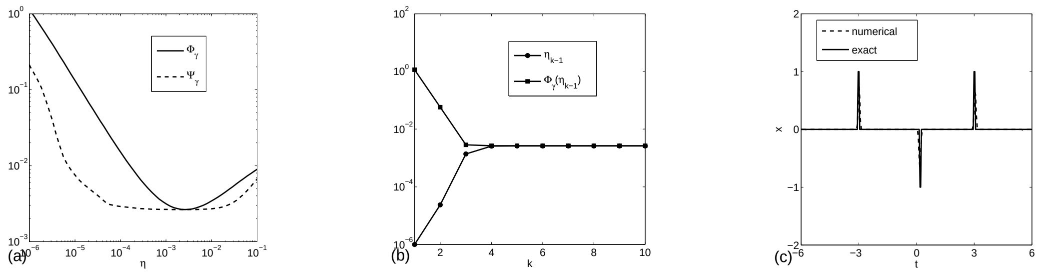


Figure 3: Numerical results for Example 3 with 5% noise: (a) Φ_γ v.s. Ψ_γ with $\gamma = 5.70 \times 10^{-1}$, (b) convergence of algorithm, and (c) reconstruction.

Table 4: Numerical results for Example 4.

(ε, q)	δ_t	δ_b	γ	η_b	e_b	η_o	e_o
(0.1,0.3)	1.74e-2	1.74e-2	1.06e1	8.25e-4	4.37e-7	1.84e-3	7.66e-8
(0.3,0.3)	5.22e-2	5.22e-2	1.78e1	1.47e-3	9.25e-8	1.84e-3	7.68e-8
(0.5,0.3)	8.69e-2	8.69e-2	2.24e1	1.95e-3	7.59e-8	1.84e-3	7.68e-8
(0.7,0.3)	1.22e-1	1.22e-1	2.57e1	2.37e-3	8.84e-8	1.84e-3	7.68e-8
(0.9,0.3)	1.56e-1	1.56e-1	2.81e1	2.79e-3	1.14e-7	1.84e-3	7.68e-8
(1.0,0.1)	5.86e-2	5.86e-2	1.83e1	1.61e-3	6.10e-8	5.96e-4	3.95e-8
(1.0,0.2)	1.32e-1	1.32e-1	2.57e1	2.58e-3	1.20e-7	9.10e-4	8.72e-8
(1.0,0.3)	1.74e-1	1.74e-1	2.91e1	3.00e-3	1.28e-7	1.84e-3	7.68e-8
(1.0,0.4)	2.35e-1	2.34e-1	3.17e1	3.66e-3	1.57e-1	1.00e-2	1.48e-1
(1.0,0.5)	3.27e-1	3.27e-1	3.50e1	4.69e-3	1.98e-2	2.81e-3	1.26e-2
(1.0,0.6)	3.77e-1	3.77e-1	3.68e1	5.12e-3	8.26e-2	2.72e-2	7.96e-2

where ξ_i is the standard Gaussian random variable, the constant $\sigma_m := \varepsilon \max_{1\leq i\leq 100} |y_i|$. The L^1 data fidelity and TV regularization are adopted to cope with the impulsive nature of the noise and to reconstruct piecewise constant solutions, respectively.

The numerical results for the problem are summarized in Table 4. The two parameters in the noise model, i.e. the magnitude ε and the corruption percentage q, exert dramatically different effects on the reconstruction. The corruption percentage q has a profound effect on the reconstruction accuracy: the error e is reduced by an order of two as q decreases from 0.6 to 0.1. With q fixed at 0.3, the reconstruction accuracy remains almost unchanged as the corruption magnitude ε increases from 0.1 to 0.9. This is in stark contrast with preceding L^2 data fitting. The exceedingly high accuracy is attributed to the fact that L^1 data fitting functional can detects the noise sites and encourages exact data fitting at remaining ones. This has been previously observed, see e.g. reference [10] for detailed numerical demonstrations in the context of L^1-L^2 functional and [7] and references therein for theoretical investigations. For the L^1-TV functional [7], η plays the role of a characteristic parameter, and at some specific values the profile of the solution undergoes sudden transition. This explains the observation that even though there is some discrepancy between the regularization parameters by the proposed criterion and the optimal ones, the accuracy of the reconstructions remains very close to each other. Moreover, the estimated noise level is always very accurate. The numerical reconstruction for $(\varepsilon, q) = (1.0, 0.6)$ is shown in Fig. 4(c), which approximates excellently the exact one.

The criterion Ψ_γ for the problem is nonsmooth, see Fig. 4(a). This is attributed to the fact that the Tikhonov functional is not strictly continuous and there can exist multiple minimizers, and then the functions $\phi(x_\eta, y^\delta)$ and $\psi(x_\eta)$ can have discontinuity points. Consequently, the criterion Ψ_γ can be discontinuous, and thus ill-defined. In addition, it is very flat over a range, which causes numerical inconveniences. However, the proposed criterion Φ_γ is based on the value function $F(\eta)$, which is always

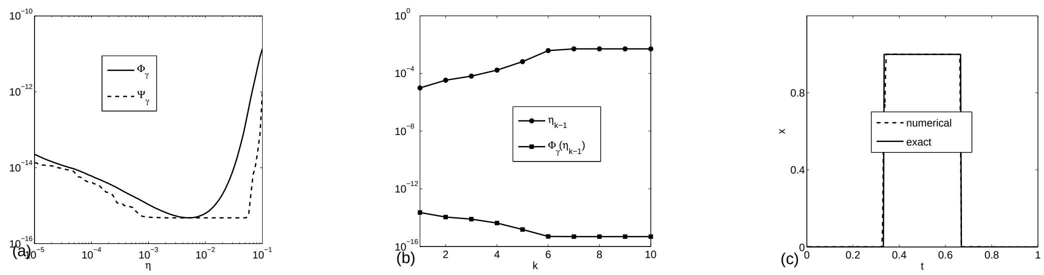


Figure 4: Numerical results for Example 4 with (1.0, .6): (a) Φ_γ v.s. Ψ_γ with $\gamma = 3.68 \times 10^1$, (b) convergence of algorithm, and (c) reconstruction.

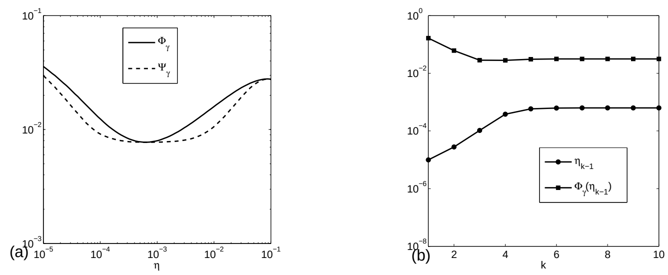


Figure 5: Numerical results for Example 5 with $\delta = 2.00 \times 10^{-2}$: (a) Φ_γ v.s. Ψ_γ with $\gamma = 4.79 \times 10^{-1}$ and (b) convergence of algorithm.

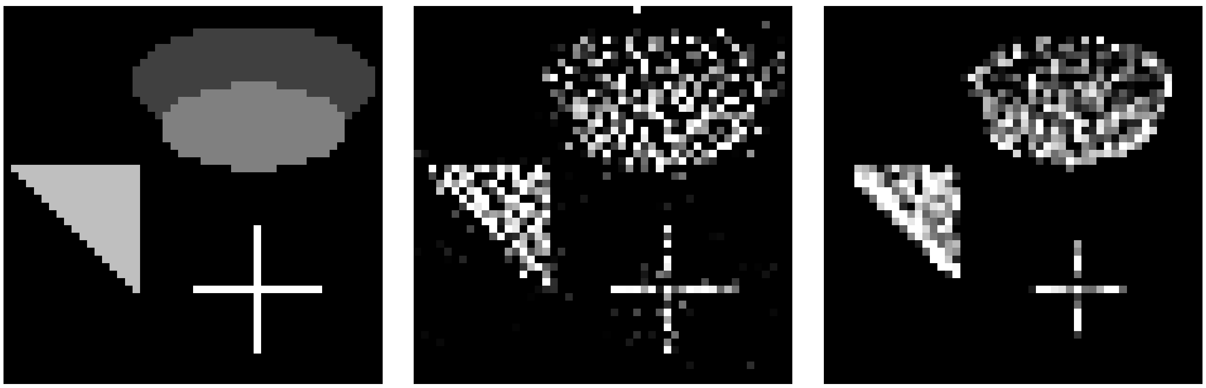
continuous irrespective of the uniqueness of minimizer to the functional $\mathcal{J}_\eta$. Thus the function Φ_γ is always smoother than Ψ_γ. Moreover, it still exhibits a distinct minimum. A steady and fast convergence of the fixed point algorithm is also observed, and the descent property of the iteration for the criterion Φ_γ remains true, see Fig. 4(b), which corroborates Theorem 4.2.

5.5 L^2-elastic net

Example 5 (2D image deblurring)**.** *This is the blur example from Regularization toolbox [16]. We set the parameters `N=50`, `band=5` and `sigma=5`. The true solution is shown in Fig. 6(a).*

The problem is mildly ill-posed with a condition number 5.83×10^{12}. We employed the so-called elastic-net regularization [35], i.e. $\psi(x) = \|x\|_{\ell^1} + \frac{\rho}{2}\|x\|_{\ell^2}^2$, with a fixed $\rho = 1 \times 10^{-3}$. Elastic net regularization stabilizes the ℓ^1-norm regularization, and statistically favors a grouping effect, see [35], and thus it is deemed suitable for identifying the group structure. For this example, γ_0 is set to 1. The nonsmooth optimization problem is solved by a Newton type method.

The numerical results are shown in Figs. 5 and 6. The value of γ determined by the two step procedure is 4.79×10^{-1}. Firstly, we observe that the criterion Ψ_γ varies little over the range $[1.0 \times 10^{-4}, 1.0 \times 10^{-2}]$, whereas the criterion Φ_γ has a distinct minimum. Secondly, the proposed algorithm converges very steadily and rapidly, and three to four iterations of the algorithm can yield a very good approximation. The estimated noise level is $\delta_b = 1.76 \times 10^{-2}$, which slightly under-estimates the exact one $\delta_t = 2.00 \times 10^{-2}$. The regularization parameter determined by the proposed criterion is $\eta_b = 6.30 \times 10^{-4}$, and it is smaller than the optimal value $\eta_o = 2.15 \times 10^{-2}$. In terms of visual inspection, these two reconstructions can identity the group structure clearly and are close to each other and thus represent reasonable reconstructions. However, the reconstruction of the cross by the proposed criterion is clearly better than the optimal one, i.e. x_o. The comparison is also explicitly indicated by the respective reconstruction errors, i.e. $e_b = 1.05$ and $e_o = 7.68 \times 10^{-1}$. This example clearly illustrates the optimality



$x^\dagger$ $\qquad$ x_b $\qquad$ x_o

Figure 6: (a) The exact solution $x^\dagger$, (b) the reconstruction x_b by the proposed criterion and (c) the reconstruction x_o with the smallest error e_o, obtained by sampling 50 values of η.

of the proposed criterion and the fixed point algorithm for high-dimensional problems.

6 Conclusions

We have proposed, analyzed and implemented a new criterion for choosing the regularization parameter in nonsmooth Tikhonov regularization. Firstly, we established some properties of the minimum value function, especially differentiability properties. The new criterion is solely defined in terms of the value function. The proposed criterion is analytically compared with an existing criterion due to Regińska. In particular, it is shown that the former is numerically more amenable in case both are well-defined as the minimizers are more sharply located. A posteriori error estimates were derived to partially justify the algorithm. An effective fixed point algorithm is suggested, and its monotonically convergence is investigated. In particular, the monotonicity and descent property of the iterates for the proposed criterion were established, and a local linear convergence of algorithm was also shown. Numerical results of five regularizing formulations, i.e. L^2-H^1 with constraint, L^2-TV, L^2-ℓ^1, L^1-TV and elastic net regularization, for several benchmark examples are presented to illustrate its salient features. The numerical results indicate that the proposed rule is effective in that it can deliver acceptable results, which are comparable with the optimal choice, and the algorithm merits a fast and steady convergence.

Acknowledgements

This work was partly carried out during the visit of Bangti Jin at Department of Mathematics, North Carolina State University. He would like to thank Professor Kazufumi Ito for hospitality. Bangti Jin is supported by Alexander von Humboldt foundation through a postdoctoral research fellowship.

References

[1] A. B. Bakushinskiĭ, *Remarks on the choice of regularization parameter from quasioptimality and relation tests*, USSR Comput. Math. Math. Phys., 24 (1984), pp. 181–182.

[2] T. Bonesky, *Morozov's discrepancy principle and Tikhonov-type functionals*, Inverse Problems, 25 (2009), pp. 015015, 11.

[3] T. Bonesky, K. S. Kazimierski, P. Maass, F. Schöpfer, and T. Schuster, *Minimization of Tikhonov functionals in Banach spaces*, Abstr. Appl. Anal., (2008), pp. Art. ID 192679, 19.

[4] M. Burger and S. Osher, *Convergence rates of convex variational regularization*, Inverse Problems, 20 (2004), pp. 1411–1421.

[5] E. J. Candès and T. Tao, *Decoding by linear programming*, IEEE Trans. Inform. Theory, 51 (2005), pp. 4203–4215.

[6] A. Chambolle, *An algorithm for total variation minimization and applications*, J. Math. Imaging Vision, 20 (2004), pp. 89–97.

[7] T. F. Chan and S. Esedoḡlu, *Aspects of total variation regularized L^1 function approximation*, SIAM J. Appl. Math., 65 (2005), pp. 1817–1837.

[8] T. F. Chan, G. H. Golub, and P. Mulet, *A nonlinear primal-dual method for total variation-based image restoration*, SIAM J. Sci. Comput., 20 (1999), pp. 1964–1977.

[9] Z. Chen and J. Zou, *An augmented Lagrangian method for identifying discontinuous parameters in elliptic systems*, SIAM J. Control Optim., 37 (1999), pp. 892–910.

[10] C. Clason, B. Jin, and K. Kunisch, *A semismooth newton method for L^1 data fitting with automatic choice of regularization parameters and noise calibration*, SIAM J. Imag. Sci., (2010), p. in press.

[11] C. Clason and K. Kunisch, *A duality-based approach to elliptic control problems in non-reflexive Banach spaces*, ESAIM Control Optim. Calc. Var., (2009), p. to appear.

[12] D. Donoho, *Compressed sensing*, IEEE Trans. Inform. Theory, 52 (2006), pp. 1289–1306.

[13] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, vol. 375 of Mathematics and its Applications, Kluwer Academic Publishers Group, Dordrecht, 1996.

[14] T. Goldstein and S. Osher, *The split Bregman method for $L1$-regularized problems*, SIAM J. Imaging Sci., 2 (2009), pp. 323–343.

[15] M. Grasmair, M. Haltmeier, and O. Scherzer, *Sparse regularization with lq penalty term*, Inverse Problems, 24 (2008), p. 055020 (13pp).

[16] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems*, SIAM Monographs on Mathematical Modeling and Computation, SIAM, Philadelphia, PA, 1998.

[17] 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 (2007), pp. 987–1010.

[18] Y. Huang, M. K. Ng, and Y.-W. Wen, *A fast total variation minimization method for image restoration*, Multiscale Model. Simul., 7 (2008), pp. 774–795.

[19] K. Ito, B. Jin, and J. Zou, *A new choice rule for regularization parameters in Tikhonov regularization*. techical report 362, Department of Mathematics, Chinese University of Hong Kong, 2008.

[20] K. Ito and K. Kunisch, *On the choice of the regularization parameter in nonlinear inverse problems*, SIAM J. Optim., 2 (1992), pp. 376–404.

[21] K. Ito and T. Takeuchi, *The minimum value function for the tikhonov regularization and its applications*. preprint, arXiv:0910.4351.

[22] B. Jin and D. A. Lorenz, *Heuristic parameter-choice rules for convex variational regularization based on error estimates*. Preprint, arXiv.org/abs/1001.5346.

[23] B. Jin and J. Zou, *Augmented Tikhonov regularization*, Inverse Problems, 25 (2009), pp. 025001, 25.

[24] ———, *Iterative schemes for Morozov's discrepancy principle in optimizations arising from inverse problems*. submitted, 2009.

[25] D. A. Lorenz, *Convergence rates and source conditions for Tikhonov regularization with sparsity constraints*, J. Inverse Ill-Posed Probl., 16 (2008), pp. 463–478.

[26] A. Neubauer, *On enhanced convergence rates for Tikhonov regularization of nonlinear ill-posed problems in Banach spaces*, Inverse Problems, 25 (2009), pp. 065009, 10.

[27] T. Regińska, *A regularization parameter in discrete ill-posed problems*, SIAM J. Sci. Comput., 17 (1996), pp. 740–749.

[28] E. Resmerita, *Regularization of ill-posed problems in Banach spaces: convergence rates*, Inverse Problems, 21 (2005), pp. 1303–1314.

[29] E. Resmerita and O. Scherzer, *Error estimates for non-quadratic regularization and the relation to enhancement*, Inverse Problems, 22 (2006), pp. 801–814.

[30] L. I. Rudin, S. J. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D, 60 (1992), pp. 259–268.

[31] A. N. Tikhonov, A. S. Leonov, and A. G. Yagola, *Nonlinear Ill-Posed Problems. Vol. 1, 2*, vol. 14 of Applied Mathematics and Mathematical Computation, Chapman & Hall, London, 1998. Translated from the Russian.

[32] C. R. Vogel, *Computational Methods for Inverse Problems*, vol. 23 of Frontiers in Applied Mathematics, SIAM, Philadelphia, PA, 2002.

[33] P. Weiss, L. Blanc-Féraud, and G. Aubert, *Efficient schemes for total variation minimization under constraints in image processing*, SIAM J. Sci. Comput., 31 (2009), pp. 2047–2080.

[34] S. Wright, R. Nowak, and M. Figueiredo, *Sparse reconstruction by separable approximation*, IEEE Trans. Signal Process., 57 (2009), pp. 2479–2493.

[35] H. Zou and T. Hastie, *Regularization and variable selection via the elastic net*, J. Royal Stat. Soc. Ser. B, 67 (2005), pp. 301–320.

UTMS

2009–18 Hitoshi Kitada: *Asymptotically outgoing and incoming spaces and quantum scattering.*

2009–19 V. G. Romanov and M. Yamamoto : *Recovering a Lamé Kernel in a viscoelastic equation by a single boundary measurement.*

2009–20 Hermann Brunner, Leevan Ling and Masahiro Yamamoto: *Numerical simulations of two-dimensional fractional subdiffusion problems.*

2009–21 Hajime Fujita, Mikio Furuta and Takahiko Yoshida: *Torus fibrations and localization of index II - Local index for acyclic compatible system -.*

2009–22 Oleg Yu. Imanuvilov, Gunther Uhlmann, and Masahiro Yamamoto: *Partial Cauchy data for general second order elliptic operators in two dimensions.*

2009–23 Yukihiro Seki: *On exact dead-core rates for a semilinear heat equation with strong absorption.*

2009–24 Yohsuke Takaoka: *On existence of models for the logical system MPCL.*

2009–25 Takefumi Igarashi and Noriaki Umeda: *Existence of global solutions in time for Reaction-Diffusion systems with inhomogeneous terms in cones.*

2010–1 Norikazu Saito: *Error analysis of a conservative finite-element approximation for the Keller-Segel system of chemotaxis.*

2010–2 Mourad Bellassoued and Masahiro Yamamoto: *Carleman estimate with second large parameter for a second order hyperbolic operators in a Riemannian manifold.*

2010–3 Kazufumi Ito, Bangti Jin and Tomoya Takeuchi: *A regularization parameter for nonsmooth Tikhonov regularization.*

The Graduate School of Mathematical Sciences was established in the University of Tokyo in April, 1992. Formerly there were two departments of mathematics in the University of Tokyo: one in the Faculty of Science and the other in the College of Arts and Sciences. All faculty members of these two departments have moved to the new graduate school, as well as several members of the Department of Pure and Applied Sciences in the College of Arts and Sciences. In January, 1993, the preprint series of the former two departments of mathematics were unified as the Preprint Series of the Graduate School of Mathematical Sciences, The University of Tokyo. For the information about the preprint series, please write to the preprint series office.

ADDRESS:
Graduate School of Mathematical Sciences, The University of Tokyo
3–8–1 Komaba Meguro-ku, Tokyo 153-8914, JAPAN
TEL +81-3-5465-7001 FAX +81-3-5465-7012