IMA Journal of Numerical Analysis (2012) **32**, 1714–1732
doi:10.1093/imanum/drr051
Advance Access publication on February 30, 2012

Iterative parameter choice by discrepancy principle

BANGTI JIN
Department of Mathematics and Institute for Applied Mathematics and Computational Science, Texas A&M University, College Station, Texas 77843-3368, USA
btjin@math.tamu.edu

YUBO ZHAO
Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Xueyuan Avenue 1068, Shenzhen University Town, Shenzhen, People's Republic of China

AND

JUN ZOU*
Department of Mathematics, The Chinese University of Hong Kong, Shatin N.T., Hong Kong
*Corresponding author: zou@math.cuhk.edu.hk

[Received on 31 July 2011; revised on 31 July 2011]

This work is concerned with the numerical implementation of the discrepancy principle for nonsmooth Tikhonov regularization for linear inverse problems. First, some theoretical properties of the solutions to the discrepancy equation, i.e., uniqueness and upper bounds, are discussed. Then, the idea of Padé approximation is exploited for designing model functions with model parameters iteratively updated. Two algorithms are proposed for its efficient numerical realization, i.e., a two-parameter algorithm based on model functions and a quasi-Newton method, and their convergence properties are briefly discussed. Numerical results for four nonsmooth models are presented to demonstrate the accuracy of the principle and to illustrate the efficiency and robustness of the proposed algorithms.

Keywords: discrepancy principle; regularization parameter; uniqueness; model function approach; Tikhonov regularization.

1. Introduction

This work considers determining regularization parameters in nonsmooth Tikhonov regularization. We will focus on linear inverse problems that can be written

$$Kx = y^\delta, \tag{1.1}$$

where $x \in X$ and $y^\delta \in Y$ refer to the unknown and the noisy data, respectively. The spaces X and Y are Banach spaces with their respective norms denoted by $\|\cdot\|_X$ and $\|\cdot\|_Y$, with X being reflexive. The reflexivity of X is not essential. In the case of a Hilbert space, its norm will simply be denoted by $\|\cdot\|$. The operator $K : X \to Y$ is linear and bounded. The accuracy of the data y^δ relative to the exact data $y^\dagger = Kx^\dagger$ is measured by the noise level $\delta^2 = \phi(x^\dagger, y^\delta)$, where the fidelity $\phi(x, y^\delta) : X \times Y \mapsto \{0\} \cup \mathbb{R}_+$ measures the proximity of the model output Kx to the data y^δ. Some widely-used choices include $\frac{1}{2}\|Kx - y^\delta\|^2_{L^2}$, $\|Kx - y^\delta\|_{L^1}$ and $\int (Kx - y^\delta \log Kx) + C$, which are statistically suitable for Gaussian, (impulsive) Laplacian and Poissonian noise, respectively.

Inverse problems are numerically challenging due to their inherent ill-posedness in the sense that small perturbations of the data can cause large deviations in the solution. One standard procedure is regularization, which minimizes the Tikhonov functional

$$J_\eta(x) = \phi(x, y^\delta) + \eta\psi(x), \tag{1.2}$$

and takes its minimizer x_η as an approximation, where η is the regularization parameter. The penalty $\psi(x)$ encodes the *a priori* information, and some common choices include $\frac{1}{\nu}\|x\|_{L^\nu}^\nu$, $\frac{1}{2}\|x\|_{H^m}^2$ and $|x|_{TV}$. Nonsmooth models have attracted considerable attention; see Burger & Osher (2004); Resmerita (2005); Hofmann *et al.* (2007); Hein & Hofmann (2009); Hofmann & Yamamoto (2010), for theoretical studies.

The choice of η represents one of the most challenging issues; see Engl *et al.* (1996); Vogel (2002); Jin & Zou (2009). The discrepancy principle (Morozov, 1966) is very popular due to its rigorous foundation and *a posteriori* nature. Theoretically, existence and consistency of the rule were studied in Tikhonov *et al.* (1998). In Bonesky (2009), convergence rates were derived under appropriate source conditions. The principle determines the parameter $\eta(\delta)$ such that it solves

$$\phi(x_\eta, y^\delta) = \delta^2. \tag{1.3}$$

Since x_η is only implicitly defined via the functional J_η, in practice the principle amounts to solving a highly nonlinear and potentially nonsmooth equation in η.

The numerical realization of the principle has not received due attention despite its practical significance. This is especially important for nonsmooth models since their efficient minimization can be highly nontrivial and their repeated solutions can be computationally very demanding. Among existing approaches the model function approach (MFA) stands out prominently (Ito & Kunisch, 1992; Kunisch & Zou, 1998; Xie & Zou, 2002). Ito & Kunisch (1992) proposed a four-parameter model function for nonlinear parameter identification problems. Later, Kunisch & Zou (1998) derived a two-parameter model function for linear inverse problems, which was further improved in Xie & Zou (2002). Recently, it has also been applied to other choice rules (Heng *et al.*, 2010). Nonetheless, all these studies focus on the L^2–L^2 model, and a direct extension to nonsmooth models is not viable because these derivations hinge essentially on the inner product structure of Hilbert spaces.

In this paper we provide some light on the aforementioned numerical aspects of the principle. First, we provide some properties (uniqueness and upper bound) of the discrepancy equation, i.e., equation (1.3). Then, we discuss differentiability of the value function and generalize MFA in Kunisch & Zou (1998) and Xie & Zou (2002) via Padé approximants, which enables us to develop analogous strategies for general nonsmooth models. These represent our essential contributions. A hybrid algorithm combining the steady global convergence of the MFA and the fast local convergence of a quasi-Newton method is also proposed.

The rest of the paper is structured as follows. Section 2 discusses the discrepancy equation and derives some theoretical results. Section 3 studies the differentiability of the value function and the numerical solution of the discrepancy equation. Section 4 presents some numerical results for four nonsmooth models, i.e., L^2–TV, L^2–ℓ^1, L^1–TV and elastic net, to demonstrate the accuracy and optimality of the principle and to illustrate the efficiency and robustness of the algorithms.

2. Properties of discrepancy equation

This section discusses uniqueness and bounds of solutions to equation (1.3). We shall make the following assumptions: (a) and (b) are standard for the existence of minimizers to J_η, and $\tilde{x}$ in (c) can often be set to $\tilde{x} = 0$.

ASSUMPTIONS 2.1 The non-negative functionals ϕ and ψ in equation (1.2) satisfy

(a) for any $\eta > 0$ the functional $J_\eta(x)$ in (1.2) is coercive;
(b) the functionals $\phi(x, y)$ and $\psi(x)$ are weakly lower semicontinuous, i.e., $\phi(x, y) \leqslant \liminf_{n\to\infty} \phi(x_n, y)$ and $\psi(x) \leqslant \liminf_{n\to\infty} \psi(x_n)$ for any sequence $\{x_n\} \subset X$ converging weakly to x;
(c) there exists an $\tilde{x} \in X$ such that it minimizes $\psi(x)$ over X with $\psi(\tilde{x}) = 0$.

The minimizers x_η, with the set denoted by $S(\eta)$, of J_η may be nonunique. Hence, the functions $\eta \mapsto \phi(x_\eta, y^\delta)$ and $\eta \mapsto \psi(x_\eta)$ may be set-valued. However, they are monotone (Tikhonov & Arsenin, 1977; Tikhonov *et al.*, 1998).

LEMMA 2.2 The functions $\phi(x_\eta, y^\delta)$ and $\psi(x_\eta)$ are monotone in η in that for any $x_{\eta_1} \in S(\eta_1)$ and any $x_{\eta_2} \in S(\eta_2)$ there hold

$$(\eta_1 - \eta_2)(\psi(x_{\eta_1}) - \psi(x_{\eta_2})) \leqslant 0 \quad \text{and} \quad (\eta_1 - \eta_2)(\phi(x_{\eta_1}, y^\delta) - \phi(x_{\eta_2}, y^\delta)) \geqslant 0.$$

There exists at least a positive solution to equation (1.3) if $\lim_{\eta\to 0} \phi(x_\eta, y^\delta) < \delta^2$, $\lim_{\eta\to\infty} \phi(x_\eta, y^\delta) > \delta^2$ and $\phi(x_\eta, y^\delta)$ is continuous in η. The continuity of $\phi(x_\eta, y^\delta)$ in η is ensured by the uniqueness of the minimizer to J_η, and it holds for strictly convex models, e.g., L^2–L^p with $1 < p \leqslant 2$ and L^2–TV and L^2–ℓ^1 with injective K. Uniqueness is necessary for the continuity of $\phi(x_\eta, y^\delta)$; see Chan & Esedoḡlu (2005) for L^1–TV.

The solution to equation (1.3) may be nonunique, even if it does exist. Under the continuity condition, the set of solutions can form a closed interval. We shall examine the uniqueness issue, which has rarely been discussed. The next result reveals one simple uniqueness condition.

LEMMA 2.3 Let the functional J_η satisfy $S(\eta_1) \cap S(\eta_2) = \emptyset$ for distinct $\eta_1, \eta_2 > 0$. Then, equation (1.3) has a unique solution $\eta(\delta) > 0$.

Proof. We proceed by means of contradiction. Assume that there exist two distinct solutions η_1 and η_2 to equation (1.3), i.e., $\phi(x_{\eta_1}, y^\delta) = \phi(x_{\eta_2}, y^\delta) = \delta^2$. By the minimizing property of the minimizers $x_{\eta_1} \in S(\eta_1)$ and $x_{\eta_2} \in S(\eta_2)$ and the fact that $x_{\eta_2} \notin S(\eta_1)$ from the assumption $S(\eta_1) \cap S(\eta_2) = \emptyset$, we have

$$\phi(x_{\eta_1}, y^\delta) + \eta_1 \psi(x_{\eta_1}) < \phi(x_{\eta_2}, y^\delta) + \eta_1 \psi(x_{\eta_2}),$$

which consequently implies that $\psi(x_{\eta_1}) < \psi(x_{\eta_2})$. However, reversing the roles of η_1 and η_2 gives a contradictory inequality $\psi(x_{\eta_2}) < \psi(x_{\eta_1})$. □

To exploit Lemma 2.3, we first derive a result characterizing the solution x_η. It also gives an explicit formula for computing η^* from the data y^δ for Hilbertian Y and $\psi(x)$ being a norm on X (or its power). We denote by $\|\cdot\|_{X^*}$ the norm of the dual space X^* of the space X, and by $\langle\cdot,\cdot\rangle$ both the duality pairing between X^* and X and the inner product on a Hilbert space. Also we recall the subdifferential $\partial\psi(x)$ of ψ at x, i.e., $\partial\psi(x) = \{\zeta \in X^* : \psi(x') \geqslant \psi(x) + \langle\zeta, x' - x\rangle \text{ for all } x' \in X\}$.

THEOREM 2.4 Let Y be a Hilbert space, $\phi(x, y^\delta) = \frac{1}{2}\|Kx - y^\delta\|^2$, $\psi(x) = \frac{1}{p}\|x\|_X^p$ with $p \geqslant 1$, and $\|K^* y^\delta\|_{X^*} \neq 0$. Furthermore, let $\eta^* = \|K^* y^\delta\|_{X^*}$ if $p = 1$ and $\eta^* = +\infty$ if $p > 1$. Then, for $\eta \geqslant \eta^*$, $x_\eta = 0$ and for $\eta < \eta^*$, there hold

$$\langle K^*(Kx_\eta - y^\delta), x_\eta\rangle = -p\eta\psi(x_\eta) \quad \text{and} \quad \|K^*(y^\delta - Kx_\eta)\|_{X^*} = \eta\|x_\eta\|_X^{p-1}.$$

Conversely, if the above relations hold, then x_η is a minimizer of J_η.

Proof. First, we note that the optimality condition for x_η reads

$$K^*(Kx_\eta - y^\delta) + \eta\zeta_\eta = 0,$$

where $\zeta_\eta \in \partial\psi(x_\eta)$. We discuss the cases $p > 1$ and $p = 1$ separately.

For $p > 1$, $\partial\psi(x_\eta) = j_p(x_\eta)$, the duality mapping of weight $g(t) = t^{p-1}$ (Cioranescu, 1990, Theorem 4.4). It satisfies

$$\|j_p(x_\eta)\|_{X^*} = \|x_\eta\|_X^{p-1} \quad \text{and} \quad \langle j_p(x_\eta), x_\eta\rangle = \|x\|_X^p.$$

Consequently, the desired relations follow directly. Conversely, the two relations imply directly that (Cioranescu, 1990)

$$\eta^{-1}K^*(y^\delta - Kx_\eta) = j_p(x_\eta) \in \partial\psi(x_\eta),$$

which is a sufficient condition for the optimality of x_η by noting the convexity of J_η.

Meanwhile, for $p = 1$, by Cioranescu (1990, Proposition 3.4), we have $\partial\psi(x) = \{x^* \in X^* : \langle x^*, x\rangle = \|x\|_X, \|x^*\|_{X^*} = 1\}$ for $x \neq 0$. Therefore, at $x_\eta \neq 0$, the two relations obviously hold, and the converse follows as before. If $x_\eta = 0$, the minimizing property implies that for any $h \in X$ and $\varepsilon > 0$,

$$\tfrac{1}{2}\|K(\varepsilon h) - y^\delta\|^2 + \eta\psi(\varepsilon h) \geqslant \tfrac{1}{2}\|y^\delta\|^2.$$

Letting $\varepsilon \to 0^+$ yields $\eta\psi(h) - \langle Kh, y^\delta\rangle \geqslant 0$. Since the inequality holds for arbitrary $h \in X$ and $\|K^*y^\delta\|_{X^*} \neq 0$, we have $\eta \geqslant \eta^* = \|K^*y^\delta\|_{X^*}$. The converse follows by reversing the arguments. □

REMARK 2.5 In the case where $\psi(x)$ is a seminorm (or its power), i.e., $\psi(x) = \frac{1}{p}|x|_X^p$ with $p \geqslant 1$, then at a minimum x_η, the identity $\langle K^*(Kx_\eta - y^\delta), x_\eta\rangle = -p\eta\psi(x_\eta)$ remains valid. However, an explicit characterization of the critical value η^* is missing.

Now we can state the uniqueness for some nonsmooth models, e.g., $\frac{1}{p}\|\cdot\|_{L^p}^p$ and $|\cdot|_{TV}$.

COROLLARY 2.6 Assume that $\eta^* > 0$. Let $\phi(x, y^\delta) = \frac{1}{2}\|Kx - y^\delta\|^2$ and $\psi(x) = \frac{1}{p}\|x\|_X^p\big(\text{or } \frac{1}{p}|x|_X^p\big)$ with $p \geqslant 1$. If there exists an $\eta < \eta^*$ satisfying equation (1.3), then it is unique.

Proof. In the light of Remark 2.5 it suffices to show the norm case. Assume that there exist two distinct $\eta_1, \eta_2 < \eta^*$ both satisfying equation (1.3). Then, by Theorem 2.4, the solutions $x_{\eta_1} \in S(\eta_1)$ and $x_{\eta_2} \in S(\eta_2)$ satisfy

$$\langle K^*(y^\delta - Kx_{\eta_i}), x_{\eta_i}\rangle = \eta_i\|x_{\eta_i}\|_X^p, \quad i = 1, 2.$$

This implies $S(\eta_1) \cap S(\eta_2) = \emptyset$, and the assertion follows from Lemma 2.3. □

The next corollary is a direct consequence of Theorem 2.4.

COROLLARY 2.7 Assume that $\eta^* > 0$. Let $\phi(x, y^\delta) = \frac{1}{2}\|Kx - y^\delta\|^2$ and $\psi(x) = \frac{1}{p}\|x\|_X^p\big(\text{or } \frac{1}{p}|x|_X^p\big)$ with $p \geqslant 1$. Then, $\phi(x_\eta, y^\delta)$ and $\psi(x_\eta)$ are strictly monotone on the interval $]0, \eta^*]$.

Finally, we derive an upper bound for the solution to equation (1.3) for the case $\phi(x, y^\delta) = \frac{1}{2}\|Kx - y^\delta\|^2$ and $\psi(x) = \frac{1}{p}\|x\|_X^p$. It may be used as an initial guess in iterative algorithms for solving equation (1.3).

THEOREM 2.8 Let $\phi(x, y^\delta) = \frac{1}{2}\|Kx - y^\delta\|^2$ and $\psi(x) = \frac{1}{p}\|x\|_X^p (p \geqslant 1)$. Then, for the choice $\eta = \frac{\|K\|_{L(X,Y)}^p}{(\|y^\delta\| - \delta)^{p-1}}\delta$, there holds the estimate $\|Kx_\eta - y^\delta\|^2 \geqslant \delta^2$.

Proof. From Theorem 2.4 we have $\eta p \psi(x_\eta) = -\langle K^*(Kx_\eta - y^\delta), x_\eta\rangle = \langle y^\delta - Kx_\eta, Kx_\eta\rangle$. Applying the Cauchy–Schwarz inequality gives

$$\eta \leqslant \frac{\|Kx_\eta - y^\delta\|\|Kx_\eta\|}{p\psi(x_\eta)}.$$

Consequently, we have

$$\eta \leqslant \frac{\|Kx_\eta - y^\delta\|\|Kx_\eta\|}{p\psi(x_\eta)} = \frac{\|Kx_\eta - y^\delta\|\|Kx_\eta\|^p}{\|x_\eta\|_X^p\|Kx_\eta\|^{p-1}}.$$

Now the inequality $\|Kx_\eta\| \leqslant \|K\|_{L(X,Y)}\|x_\eta\|_X$ implies that

$$\eta \leqslant \frac{\|Kx_\eta - y^\delta\|\|K\|_{L(X,Y)}^p}{\|Kx_\eta\|^{p-1}} \leqslant \frac{\|Kx_\eta - y^\delta\|\|K\|_{L(X,Y)}^p}{(\|y^\delta\| - \|Kx_\eta - y^\delta\|)^{p-1}},$$

by noting that $\|Kx_\eta\| \geqslant \|y^\delta\| - \|Kx_\eta - y^\delta\|$. The desired estimate follows from the inequality. □

Hence, the value η by the principle is at most of order δ, i.e., $\mathcal{O}(\delta)$. This shows suboptimality of the principle in Hilbert spaces: it cannot achieve optimal convergence rates for the source parameter ν within the range $]\frac{1}{2}, 1]$ since then optimality can only be realized for η larger than $\mathcal{O}(\delta)$ (Engl *et al.*, 1996).

3. Numerical algorithms

In this section we develop efficient algorithms for solving equation (1.3). Generally, equation (1.3) is highly nonlinear, nonconvex and potentially nonsmooth, and thus, its numerical treatment is nontrivial. We will exploit the idea of model functions (Ito & Kunisch, 1992; Kunisch & Zou, 1998; Xie & Zou, 2002). Our essential contributions here include a differentiability result and interpreting model functions via Padé approximation. First, we introduce the value function

$$F(\eta) := J_\eta(x_\eta) = \min_{x \in X} J_\eta(x).$$

3.1 *Value function*

Here we collect some important analytical properties of the function $F(\eta)$.

LEMMA 3.1 The function $F(\eta)$ is continuous and monotonically increasing.

Proof. The assertion follows directly from the minimizing property of x_η. □

LEMMA 3.2 The following limits hold for the function $F(\eta)$:

$$\lim_{\eta \to 0^+} F(\eta) = \lim_{\eta \to 0^+} \phi(x_\eta, y^\delta) \equiv \phi^0 \quad \text{and} \quad \lim_{\eta \to +\infty} F(\eta) = \lim_{\eta \to +\infty} \phi(x_\eta, y^\delta) \equiv \phi^\infty. \tag{3.1}$$

Proof. Assumption 2.1(c) and the minimizing property of x_η imply

$$F(\eta) \equiv \phi(x_\eta, y^\delta) + \eta\psi(x_\eta) \leqslant \phi(\tilde{x}, y^\delta) + \eta\psi(\tilde{x}) = \phi(\tilde{x}, y^\delta) < +\infty, \tag{3.2}$$

and a trivial lower bound for $F(\eta)$ is 0, i.e., $F(\eta)$ is uniformly bounded. Hence, both limits in (3.1) make sense. There exists an $x^\varepsilon \in X$ such that $\phi^0 \leqslant \phi(x^\varepsilon, y^\delta) \leqslant \phi^0 + \varepsilon$ and by the extremal property of x_η,

$$\phi^0 \leqslant \phi(x_\eta, y^\delta) + \eta\psi(x_\eta) \leqslant \phi(x^\varepsilon, y^\delta) + \eta\psi(x^\varepsilon) \leqslant \phi^0 + \varepsilon + \eta\psi(x^\varepsilon).$$

Letting $\eta \to 0^+$ gives $\phi^0 \leqslant \lim_{\eta\to 0^+} F(\eta) \leqslant \phi^0 + \varepsilon$. This shows the first assertion.

Let $\{\eta_n\}_n$ be a sequence tending to infinity. By Lemma 2.2 and inequality (3.2), $\psi(x_{\eta_n}) \to 0$ and $\phi(x_{\eta_n}, y^\delta) \to \phi^\infty$. Hence, by Assumption 2.1(a), $\{x_{\eta_n}\}_n$ is uniformly bounded, and there exists a subsequence, also denoted $\{x_{\eta_n}\}_n$, and some $x^* \in X$, such that $x_{\eta_n} \to x^*$ weakly and $\psi(x^*) = 0$. Now Assumption 2.1(b) and the minimizing property of x_{η_n} yield

$$\begin{aligned}\phi(x^*, y^\delta) &\leqslant \liminf_{n\to\infty} \phi(x_{\eta_n}, y^\delta) \leqslant \limsup_{n\to\infty} \phi(x_{\eta_n}, y^\delta)\\ &\leqslant \limsup_{n\to\infty}\{\phi(x_{\eta_n}, y^\delta) + \eta_n\psi(x_{\eta_n}) - \eta_n\psi(x^*)\}\\ &\leqslant \limsup_{n\to\infty}\{\phi(x^*, y^\delta) + \eta_n\psi(x^*) - \eta_n\psi(x^*)\} = \phi(x^*, y^\delta),\end{aligned}$$

i.e., $\lim_{n\to\infty} \phi(x_{\eta_n}, y^\delta) = \phi(x^*, y^\delta)$. Letting $n \longrightarrow \infty$ in the inequality $\phi(x_{\eta_n}, y^\delta) \leqslant \phi(x_{\eta_n}, y^\delta) + \eta_n\psi(x_{\eta_n}) \leqslant \phi(x^*, y^\delta)$ establishes the second assertion. □

In practice, ϕ^0 and ϕ^∞ are often straightforward to obtain: ϕ^0 is the minimal value of $\phi(x, y^\delta)$ over the space X, while ϕ^∞ is the minimal value of $\phi(x, y^\delta)$ over the subset of X which annihilates ψ.

Finally, we show a differentiability result for $F(\eta)$.

THEOREM 3.3 If the function $\psi(x_\eta)$ is continuous at $\tilde{\eta} > 0$, then $F(\eta)$ is differentiable at $\tilde{\eta}$ and

$$F'(\tilde{\eta}) = \psi(x_{\tilde{\eta}}).$$

Proof. For $\eta > \tilde{\eta}$, from the minimizing property of $x_{\tilde{\eta}}$, the relation

$$F(\eta) - F(\tilde{\eta}) = [\phi(x_\eta, y^\delta) - \phi(x_{\tilde{\eta}}, y^\delta) + \tilde{\eta}(\psi(x_\eta) - \psi(x_{\tilde{\eta}}))] + (\eta - \tilde{\eta})\psi(x_\eta) \geqslant (\eta - \tilde{\eta})\psi(x_\eta) \tag{3.3}$$

follows. Similarly, by the minimizing property of x_η, we deduce

$$F(\eta) - F(\tilde{\eta}) = \phi(x_\eta, y^\delta) - \phi(x_{\tilde{\eta}}, y^\delta) + \eta(\psi(x_\eta) - \psi(x_{\tilde{\eta}})) + (\eta - \tilde{\eta})\psi(x_{\tilde{\eta}}) \leqslant (\eta - \tilde{\eta})\psi(x_{\tilde{\eta}}). \tag{3.4}$$

Hence, inequalities (3.3)–(3.4) yield

$$\psi(x_\eta) \leqslant \tfrac{F(\eta) - F(\tilde{\eta})}{\eta - \tilde{\eta}} \leqslant \psi(x_{\tilde{\eta}}).$$

The claim now follows from the continuity of $\psi(x_\eta)$ at $\tilde{\eta}$. □

Throughout, we shall assume that $\psi(x_\eta)$ is continuous in η.

3.2 *Two numerical algorithms*

We shall develop two algorithms, i.e., MFA and a quasi-Newton method.

The idea of the MFA is as follows (Kunisch & Zou, 1998; Xie & Zou, 2002). First, by Theorem 3.3, one rewrites equation (1.3) equivalently as

$$F(\eta) - \eta F'(\eta) = \delta^2, \tag{3.5}$$

and approximates $F(\eta)$ with a model function $m_k(\eta)$ at the kth iteration given by

$$m_k(\eta) = b + \tfrac{c_k}{t_k+\eta},$$

where b, c_k and t_k are constants to be determined. In existing MFA implementations the constant b is fixed by imposing further assumptions on the asymptotes of $m_k(\eta)$: in Kunisch & Zou (1998) the condition $m_k(0) = 0$ was enforced, whereas in Xie & Zou (2002), b was set to $\frac{1}{2}\|y^\delta\|^2$, i.e., $b = \phi^\infty$.

We interpret the model $m_k(\eta)$ as a Padé approximation of $F(\eta)$ with suitable constraint on the asymptotes. Then, existing algorithms (Kunisch & Zou, 1998; Xie & Zou, 2002) are reproduced by enforcing the asymptotes $m_k(0) = 0$ and $m_k(+\infty) = \frac{1}{2}\|y^\delta\|^2$, respectively. This interpretation is new and crucial for subsequent developments since the derivations in Kunisch & Zou (1998) and Xie & Zou (2002) hinge essentially on the inner product structure. Due to its better theoretical underpinnings and numerical performance, we focus on the variant in Xie & Zou (2002), i.e., fixing b at ϕ^∞. So, we arrive at the model

$$m_k(\eta) = \phi^\infty + \tfrac{c_k}{t_k+\eta}.$$

Then, by enforcing interpolation conditions at η_k, i.e.,

$$m_k(\eta_k) = \phi^\infty + \tfrac{c_k}{t_k+\eta_k} = F(\eta_k) \quad \text{and} \quad m_k'(\eta_k) = -\tfrac{c_k}{(t_k+\eta_k)^2} = F'(\eta_k),$$

we derive

$$c_k = -\tfrac{\left(F(\eta_k)-\phi^\infty\right)^2}{F'(\eta_k)} \quad \text{and} \quad t_k = \tfrac{\phi^\infty - \phi(x_{\eta_k}, y^\delta)}{\psi(x_{\eta_k})} - 2\eta_k.$$

Hence, c_k is negative for $\eta_k < \eta^*$, and t_k is generally nondeterminate, and positive if and only if

$$\phi^\infty - \phi(x_{\eta_k}, y^\delta) - 2\eta_k \psi(x_{\eta_k}) > 0, \tag{3.6}$$

which, by Lemma 3.2, holds if η_k is sufficiently small.

If the condition $t_k > 0$ holds, the definition of m_k gives

$$\lim_{\eta\to+\infty} m_k(\eta) = \phi^\infty = \lim_{\eta\to+\infty} F(\eta),$$
$$\lim_{\eta\to+\infty} m_k'(\eta) = 0 = \lim_{\eta\to+\infty} F'(\eta).$$

Moreover, by its definition,

$$m_k'(\eta) = -\tfrac{c_k}{(t_k+\eta)^2} > 0 \quad \text{and} \quad m_k''(\eta) = \tfrac{2c_k}{(t_k+\eta)^3} < 0.$$

In view of Lemmas 2.2 and 3.1 and Theorem 3.3 the model $m_k(\eta)$ preserves the asymptotic behaviour, monotonicity and concavity of $F(\eta)$. Hence, we arrive at an approximation to equation (3.5),

$$G_k(\eta) \equiv m_k(\eta) - \eta m_k'(\eta) = \delta^2, \tag{3.7}$$

which is computationally more tractable than equation (1.3) because of the simplicity of $m_k(\eta)$. Note that $G_k'(\eta) = \frac{-2c_k\eta}{(t_k+\eta)^3} > 0$, i.e., $G_k(\eta)$ preserves the monotonicity of $\phi(x_\eta, y^\delta)$.

The next result shows the local solvability of equation (3.7).

LEMMA 3.4 If $t_k > 0$ and $\phi(x_{\eta_k}, y^\delta)$ is close to δ^2, then there exists a unique positive solution to equation (3.7).

Proof. The approximate equation (3.7) at the kth step reads $\phi^\infty + \frac{c_k}{t_k+\eta} + \eta\frac{c_k}{(t_k+\eta)^2} = \delta^2$. We discuss two cases separately, i.e., $F(\eta_k) - \eta_k F'(\eta_k) > \delta^2$ and $F(\eta_k) - \eta_k F'(\eta_k) < \delta^2$. In the former case, by the monotonicity, it suffices to show

$$\lim_{\eta\to 0^+}(m_k(\eta) - \eta m_k'(\eta)) = \phi^\infty + \frac{c_k}{t_k} < \delta^2.$$

Substituting the formulas for c_k and t_k and observing Theorem 3.3 gives

$$\left(\phi(x_{\eta_k}, y^\delta) - \delta^2\right)\left(\phi^\infty - \phi(x_{\eta_k}, y^\delta) - 2\eta_k\psi(x_{\eta_k})\right) < \eta_k^2\psi(x_{\eta_k})^2,$$

which together with the assumption yields the assertion. In the latter case the assertion follows directly from $\lim_{\eta\to+\infty}(\phi^\infty + \frac{c_k}{t_k+\eta} + \eta\frac{c_k}{(t_k+\eta)^2}) = \phi^\infty > \delta^2$. □

A direct implementation of the procedure is at best locally convergent because equation (3.7) is only locally solvable. To construct a globally convergent algorithm we replace $G_k(\eta)$ by its relaxation (Xie & Zou, 2002)

$$\hat{G}_k(\eta) = G_k(\eta) + \alpha_k\left(G_k(\eta) - \phi(x_{\eta_k}, y^\delta)\right),$$

where the relaxation constant α_k is chosen such that

$$\hat{G}_k(\eta) = \delta^2 \tag{3.8}$$

always has a unique solution. If $\phi(x_{\eta_k}, y^\delta) > \delta^2$, this can be achieved by prescribing $\hat{G}_k(0) = \hat{\alpha}\delta^2$ for all $\hat{\alpha} \in [0, 1)$, which consequently suggests $\alpha_k = \frac{G_k(0)-\hat{\alpha}\delta^2}{\phi(x_{\eta_k},y^\delta)-G_k(0)}$. It follows from the monotonicity of $G_k(\eta)$ that $1 + \alpha_k > 0$ if $\phi(x_{\eta_k}, y^\delta) > \delta^2$, and thus, $\hat{G}_k(\eta)$ preserves monotonicity. Now, we can describe a two-parameter algorithm (TPA for short); see Algorithm 1. The algorithm terminates if the relative change $|\eta_{k+1} - \eta_k|/|\eta_k|$ falls below a given tolerance.

Algorithm 1 Two-parameter algorithm (TPA)

1: Choose η_0, $\hat{\alpha}$ and K, and set $k = 1$.
2: **for** $k = 1, \ldots, K$ **do**
3: Solve (1.2) for x_{η_k}, compute $F(\eta_k)$ and $F'(\eta_k)$, and update t_k and c_k;
4: set the kth model function $m_k(\eta) = \phi^\infty + \frac{c_k}{t_k+\eta}$, and set $G_k = m_k(\eta) - \eta m_k'(\eta)$;
5: solve for η_{k+1} from the relaxed discrepancy equation

$$\left(1 + \frac{G_k(0)-\hat{\alpha}\delta^2}{\phi(x_{\eta_k},y^\delta)-G_k(0)}\right)G_k(\eta) = \delta^2 + \frac{G_k(0)-\hat{\alpha}\delta^2}{\phi(x_{\eta_k},y^\delta)-G_k(0)}\phi(x_{\eta_k}, y^\delta);$$

6: check the stopping criterion.
7: **end for**
8: **output** approximations η_K and x_{η_K}.

The next theorem shows the monotone convergence of the algorithm for 'large' initial guesses. Its proof parallels that in Xie & Zou (2002), and hence, it is omitted.

THEOREM 3.5 Let the solution $\eta^\dagger$ to equation (1.3) be unique, and let η_0 satisfy $\phi(x_{\eta_0}, y^\delta) > \delta^2$. Then, the sequence $\{\eta_k\}_k$ generated by the TPA is well defined. Moreover, the sequence is either finite, i.e., it terminates at some η_k satisfying $\phi(x_{\eta_k}, y^\delta) \leqslant \delta^2$, or it is infinite and converges to $\eta^\dagger$ strictly monotonically decreasingly.

Numerically, the TPA converges robustly, but the local convergence is slow. To improve the overall performance we hybridize it with the secant method. The variant of secant method adopted is shown in Algorithm 2, where the stopping criterion is identical to Algorithm 1. It computes η_{k+1} in the η^{-1} coordinate when η_k over-regularizes and in the η coordinate when η_k under-regularizes. Despite the nonlinearity of equation (1.3), Algorithm 2 has a large convergence basin. The next result sheds some light on the L^2–L^2 model.

Algorithm 2 Secant method

1: Choose η_0, η_1 and K, and set $k = 1$.
2: **for** $k = 1, \ldots, K$ **do**
3: Compute the residual $\phi(x_{\eta_k}, y^\delta)$;
4: if $\phi(x_{\eta_k}, y^\delta) > \delta^2$, compute η_{k+1} by

$$\tilde{\eta}_{k+1} = \frac{1}{\eta_k} - \frac{(\phi(x_{\eta_k}, y^\delta) - \delta^2)(\eta_{k-1} - \eta_k)}{\eta_{k-1}\eta_k(\phi(x_{\eta_k}, y^\delta) - \phi(x_{\eta_{k-1}}, y^\delta))}, \quad \eta_{k+1} = \frac{1}{\tilde{\eta}_{k+1}};$$

otherwise, compute η_{k+1} by

$$\eta_{k+1} = \eta_k - \frac{(\phi(x_{\eta_k}, y^\delta) - \delta^2)(\eta_k - \eta_{k-1})}{\phi(\eta_k, y^\delta) - \phi(\eta_{k-1}, y^\delta)};$$

5: solve (1.2) for the Tikhonov solution $x_{\eta_{k+1}}$;
6: check the stopping criterion.
7: **end for**
8: **output** approximations η_K and x_{η_K}.

LEMMA 3.6 In the case of $\phi(x, y^\delta) = \frac{1}{2}\|Kx - y^\delta\|^2$ and $\psi(x) = \frac{1}{2}\|Lx\|^2$, then $\phi\left(x_{\eta^{-1}}, y^\delta\right)$ is convex in η.

Proof. Let $\alpha = \eta^{-1}$, then x_α solves $(\alpha K^*K + L^*L)x_\alpha = \alpha K^* y^\delta$. Letting $r_\alpha = Kx_\alpha - y^\delta$ we can check that $\frac{\mathrm{d}x_\alpha}{\mathrm{d}\alpha}$ and $\frac{\mathrm{d}^2 x_\alpha}{\mathrm{d}\alpha^2}$ satisfy

$$(\alpha K^*K + L^*L)\frac{\mathrm{d}x_\alpha}{\mathrm{d}\alpha} = -K^* r_\alpha \quad \text{and} \quad (\alpha K^*K + L^*L)\frac{\mathrm{d}^2 x_\alpha}{\mathrm{d}\alpha^2} = -2K^*K\frac{\mathrm{d}x_\alpha}{\mathrm{d}\alpha}.$$

Denoting by Q_α the bounded inverse operator $(\alpha K^*K + L^*L)^{-1}$ gives

$$\frac{\mathrm{d}x_\alpha}{\mathrm{d}\alpha} = -Q_\alpha K^* r_\alpha \quad \text{and} \quad \frac{\mathrm{d}^2 x_\alpha}{\mathrm{d}\alpha^2} = 2Q_\alpha K^*KQ_\alpha K^* r_\alpha.$$

Consequently,

$$\frac{\mathrm{d}^2}{\mathrm{d}\alpha^2}\frac{1}{2}\|Kx_\alpha - y^\delta\|^2 = \left\langle Kx_\alpha - y^\delta, K\frac{\mathrm{d}^2 x_\alpha}{\mathrm{d}\alpha^2}\right\rangle + \left\langle K\frac{\mathrm{d}x_\alpha}{\mathrm{d}\alpha}, K\frac{\mathrm{d}x_\alpha}{\mathrm{d}\alpha}\right\rangle = 3\|KQ_\alpha K^* r_\alpha\|^2 \geqslant 0.$$

This concludes the proof of the lemma. □

Hence, the function $\phi(x_{\eta^{-1}}, y^\delta)$ is convex in η for the L^2–L^2 model, and Newton type methods converge monotonically in exact arithmetic. For nonsmooth models the convexity may not hold, but nonetheless numerical experiments are very encouraging.

4. Numerical experiments and discussions

This section presents numerical results for four benchmark inverse problems adapted from the package Regularization Tool (Hansen, 1998) to illustrate the accuracy and optimality of the principle and the robustness and efficiency of the algorithms. These are Fredholm integral equations of the first kind with kernel $k(s,t)$ and solution $x^\dagger(t)$. The discretized linear system is of the form $\mathbf{Kx} = \mathbf{y}^\delta$ with size 300×300. The model is referred to as ϕ–ψ type, e.g., L^1–TV denotes L^1 fidelity and TV penalty. Table 1 summarizes their main features: the degree of ill-posedness (Cond is the condition number) and continuity of $\phi(x_\eta, y^\delta)$. The initial guess for η is $\eta_0 = 1.0$, and the hybrid algorithm is initialized with four iterations of the TPA. The constant $\hat{\alpha}$ is fixed at $\frac{1}{4}$ as in Xie & Zou (2002). The minimization problems arising from the L^2–TV, L^2–ℓ^1 and L^1–TV models are solved by an iteratively reweighted least squares method (Clason *et al.*, 2010) and that from the L^2–elastic net model by a Newton-type method (Jin *et al.*, 2009).

Unless otherwise specified the data $\mathbf{y}^\delta$ is generated as follows: $y_i^\delta = y_i + \max_i\{|y_i|\}\varepsilon\xi_i$, where the ξ_i follow the standard normal distribution and ε is the relative noise level. The accuracy of the solution x_η is measured by the relative error $e = \|x_\eta - x^\dagger\|_{L^2}/\|x^\dagger\|_{L^2}$. For each example we also present the optimal choice $\eta_{\rm opt}$, i.e., the one achieving the minimal error e, by solving equation (1.2) for 100 logarithmically-uniformly-distributed points over the interval $[1\times10^{-15}, 1\times10^{-1}]$. The subscripts opt and dp denote the optimal solution and that by the discrepancy principle, respectively, whereas tpa and sm denote the results by the TPA and the hybrid algorithm, respectively. The number in the parentheses denotes the number of iterations and that for the hybrid algorithm includes initialization with four iterations of the TPA.

4.1 *Example 1:* L^2–TV (Numerical differentiation, adapted from deriv2)

The functions k and $x^\dagger$ are given by

$$k(s,t) = \begin{cases} s(t-1), & s < t, \\ t(s-1), & s \geqslant t, \end{cases} \quad \text{and} \quad x^\dagger(t) = \begin{cases} 1, & \frac{1}{3} < t \leqslant \frac{2}{3}, \\ 0, & \text{otherwise}, \end{cases}$$

and the integration interval is $[0, 1]$. The exact solution $x^\dagger$ is piecewise constant, and thus, TV penalty is suitable. The initial guess η_0 is taken to be 1×10^{-2}.

TABLE 1 *Numerical examples*

Example	Description	$\phi(x_\eta, y^\delta)$	Cond($\mathbf{K}$)	Noise	Program	ϕ–ψ
1	Differentiation	Continuous	1.09×10^5	Gaussian	deriv2	L^2–TV
2	Phillips	Continuous	2.14×10^8	Gaussian	phillips	L^2–ℓ^1
3	Deblurring	Discontinuous	7.35×10^8	Impulsive	blur	L^1–TV
4	2D deblurring	Continuous	3.26×10^5	Gaussian	blur	L^2–elastic net

We observe from Table 2 that the parameter η_{dp} (by the discrepancy principle) agrees very well with the optimal one η_{opt}, albeit its value is slightly larger. Furthermore, the accuracy errors e_{dp} and e_{opt} do not differ much. The TPA converges steadily and fast, see Fig. 1(a,b), and it yields an acceptable and even very accurate approximation, in terms of the error, within eight iterations. The algorithm tends to accelerate as the noise level decreases. However, its local convergence is slow, and it requires many iterations to reach a very accurate approximation of the solution to equation (1.3) (see Fig. 1(c)) where the symbol G refers to $|\phi(x_\eta, y^\delta) - \delta^2|$. The hybrid algorithm converges locally very fast. The initial guesses given by the TPA are smaller than but very close to η^* beyond which x_η does not change, and $\phi(x_{k-1}, y^\delta)$ and $\phi(x_k, y^\delta)$ are close to $\phi(x_\infty, y^\delta)$. Nonetheless, the iterates are rather steady, and the convergence is swift. The secant method is very efficient in decreasing G, which is useful when an accurate estimate of the true noise level is available. The solution is accurate and stable for ε up to 5%; see Fig. 2 and Table 2. The TV solution exhibits a typical staircasing effect: it is piecewise constant and its magnitude is slightly distorted.

Now we compare the algorithms with a direct method based on sampling at $\eta_k = \eta_0 q^{k-1}, q \in (0, 1)$. Clearly, the efficiency of this method depends crucially on the shrinkage factor q: the smaller is the factor q, the more accurate is the selected parameter, but at the expense of increased computing efforts. Also, the attainable accuracy is inherently limited by discrete sampling. We evaluate the method also on the data set with $\varepsilon = 5\%$, and set $q = 0.9$ and $\eta_0 = 1\times10^{-2}$. The error e after eight iterations of the method is 8.68×10^{-1}, which is far inferior to those in Table 2. The best possible result is achieved after 35 iterations: the selected parameter η and the error e are 2.78×10^{-4} and 2.53×10^{-1}, respectively, which represents a good approximation to η_{dp} and e_{dp}. However, it achieves G at 4.83×10^{-7}, and this can be realized by the hybrid algorithm within 10 iterations; cf. Fig. 2. Hence, the hybrid algorithm is preferred to the direct method. Although not presented, we would like to note that similar observations can be made for other examples.

4.2 *Example 2: L^2–ℓ^1* (Sparse reconstruction, adapted from Phillips' problem)

Let $\phi(t) = \left[1+\cos\frac{\pi t}{3}\right]\chi_{|t-s|<3}$ with χ being the indicator function, and $S = [-3,-2.96]\cup[0.6, 0.64]\cup[3, 3.04]$. The functions k and $x^\dagger$ are given by $k(s,t) = \phi(s-t)$ and $x^\dagger(t) = \chi_S$, and the integration interval is $[-6, 6]$. The exact solution $x^\dagger$ has a sparse representation in the pixel basis, and thus, the ℓ^1 penalty is suitable.

The matrix $\mathbf{K}$ is of full rank, and by Corollary 2.6, the solution to equation (1.3) is unique. The parameter η_{dp} is practically identical to the optimal one η_{opt} for all five noise levels; see Table 3. Interestingly, the value of η in both cases decreases at the rate $\eta \sim \delta$, and the error e apparently decreases also as $e \sim \delta$. Hence, the ℓ^1 penalty circumvents the well-known saturation phenomenon for Tikhonov regularization in Hilbert spaces. The algorithms achieve a fast and steady convergence; cf. Fig. 3(a). The approximations after eight iterations are already very close to the exact one, and the hybrid algorithm yields more accurate results. Figure 4 shows the prominent feature of the sparsity-promoting ℓ^1 penalty: the locations of all three small spikes are perfectly detected, and their retrieved magnitudes are reasonably accurate.

4.3 *Example 3: L^1–TV* (Deblurring one-dimensional image)

The functions k and $x^\dagger$ are given by $k(s,t) = \frac{1}{4\sqrt{2\pi}}e^{-\frac{(s-t)^2}{32}}\chi_{|s-t|<15}$ and $x^\dagger(t) = \chi_{[101,200]}$, and the integration interval is $[0, 300]$. The data are corrupted by additive random-valued impulsive

TABLE 2 *Numerical results for Example* 1

ε	δ^2	$\eta_{\rm opt}$	$\eta_{\rm dp}$	$\eta_{\rm tpa}$	$\eta_{\rm sm}$	$e_{\rm opt}$	$e_{\rm dp}$	$e_{\rm tpa}$	$e_{\rm sm}$
5×10^{-6}	1.52×10^{-11}	7.74×10^{-11}	2.50×10^{-9}	$6.26\times10^{-9}(8)$	$5.10\times10^{-9}(8)$	9.46×10^{-4}	9.69×10^{-3}	1.78×10^{-2}	1.14×10^{-2}
5×10^{-5}	1.52×10^{-9}	1.29×10^{-9}	5.98×10^{-8}	$3.88\times10^{-7}(8)$	$1.80\times10^{-7}(8)$	9.80×10^{-3}	4.87×10^{-2}	6.68×10^{-2}	5.20×10^{-2}
5×10^{-4}	1.52×10^{-7}	2.78×10^{-8}	2.44×10^{-6}	$6.10\times10^{-6}(8)$	$3.81\times10^{-6}(8)$	5.96×10^{-2}	8.31×10^{-2}	8.82×10^{-2}	8.38×10^{-2}
5×10^{-3}	1.52×10^{-5}	1.67×10^{-5}	2.59×10^{-5}	$6.48\times10^{-5}(8)$	$3.98\times10^{-5}(8)$	1.06×10^{-1}	1.08×10^{-1}	1.29×10^{-1}	1.09×10^{-1}
5×10^{-2}	1.52×10^{-3}	1.67×10^{-4}	2.81×10^{-4}	$7.59\times10^{-4}(8)$	$5.41\times10^{-4}(8)$	2.42×10^{-1}	2.53×10^{-1}	3.23×10^{-1}	2.65×10^{-1}

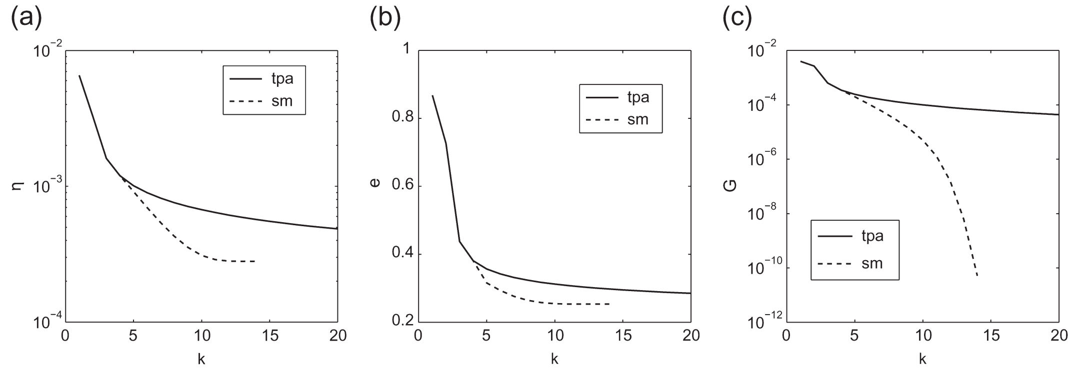

FIG. 1. The convergence of (a) η, (b) e and (c) G for Example 1 with 5% noise.

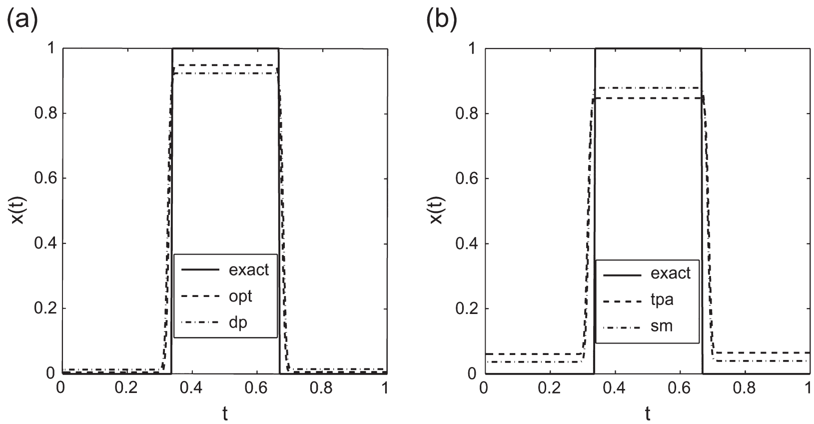

FIG. 2. The numerical solutions for Example 1 with 5% noise: (a) discrepancy principle versus optimal choice and (b) TPA versus hybrid algorithm.

noise as follows:

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

where ξ_i follows the standard Gaussian distribution, the constant $\sigma_m := \varepsilon \max_{1 \leqslant i \leqslant 300} |y_i|$. The exact noise level δ^2 is dictated by the corruption percentage r and the standard deviation σ_m. The L^1 fidelity and TV penalty are adopted to cope with impulsive noises and to reconstruct piecewise constant solutions (Chan & Esedoḡlu, 2005), respectively. The initial guess η_0 is 10. We note that the L^1–TV model is not covered by the theory herein because of the nonuniqueness of minimizers to the functional J_η (Chan & Esedoḡlu, 2005).

The value η_{dp} is close to the optimal one η_{opt} for all noise levels (cf. Table 4), with η_{opt} determined by sampling η over the interval $[1\times10^{-5}, 1\times10^{1}]$ at 100 points. Interestingly, η_{opt} remains unchanged as the noise level varies by 3 orders of magnitude and also the variation of η_{dp} is negligible. Here, the parameter η is characteristic: at some values the solution profile undergoes sudden transition and the profile might change little when η varies without crossing these values (Chan & Esedoḡlu, 2005). The

TABLE 3 *Numerical results for Example* 2

ε	δ^2	η_{opt}	η_{dp}	η_{tpa}	η_{sm}	e_{opt}	e_{dp}	e_{tpa}	e_{sm}
5×10^{-6}	8.19×10^{-11}	1.87×10^{-7}	3.38×10^{-7}	$8.12\times10^{-7}(8)$	$5.13\times10^{-7}(8)$	4.99×10^{-3}	5.84×10^{-3}	7.25×10^{-3}	6.01×10^{-3}
5×10^{-5}	8.19×10^{-9}	2.31×10^{-6}	4.26×10^{-6}	$1.04\times10^{-5}(8)$	$6.54\times10^{-6}(8)$	1.17×10^{-2}	1.23×10^{-2}	1.61×10^{-2}	1.26×10^{-2}
5×10^{-4}	8.19×10^{-7}	3.51×10^{-5}	4.43×10^{-5}	$1.06\times10^{-4}(8)$	$6.69\times10^{-5}(8)$	3.19×10^{-2}	3.27×10^{-2}	4.12×10^{-2}	3.33×10^{-2}
5×10^{-3}	8.19×10^{-5}	5.34×10^{-4}	4.39×10^{-4}	$1.06\times10^{-3}(8)$	$6.69\times10^{-4}(8)$	1.24×10^{-1}	1.25×10^{-1}	1.40×10^{-1}	1.24×10^{-1}
5×10^{-2}	8.19×10^{-3}	4.33×10^{-3}	4.32×10^{-3}	$1.07\times10^{-2}(8)$	$6.65\times10^{-3}(8)$	2.83×10^{-1}	2.83×10^{-1}	5.32×10^{-1}	2.80×10^{-1}

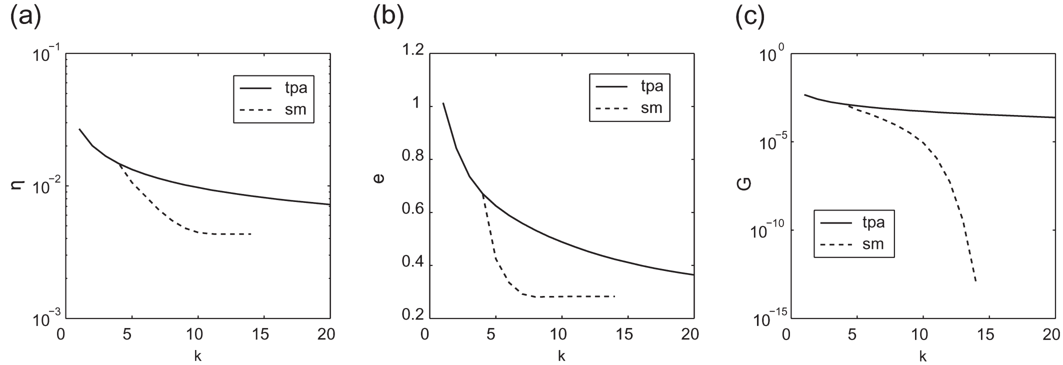

FIG. 3. The convergence of (a) η, (b) e and (c) G for Example 2 with 5% noise.

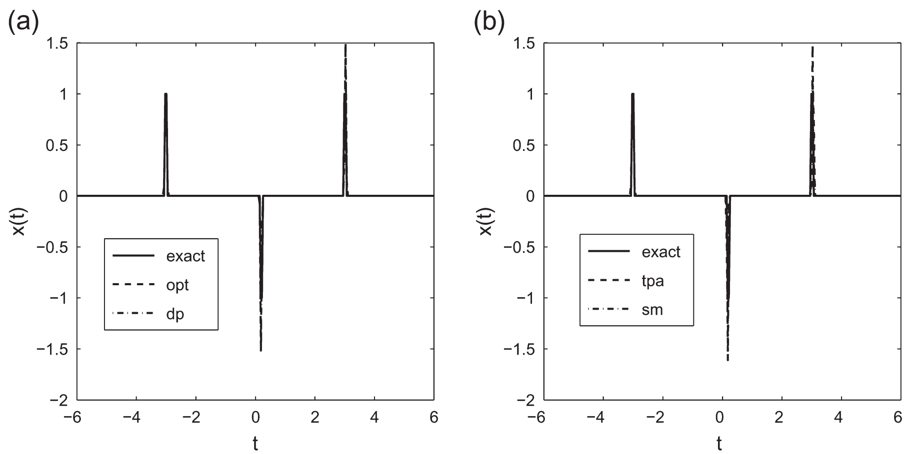

FIG. 4. The numerical solutions for Example 2 with 5% noise: (a) discrepancy principle versus optimal choice and (b) TPA versus hybrid algorithm.

numerical results for $r = 0.5$ are far more accurate than those for $r = 0.3$, which shows the profound effect of the percentage r.

The TPA converges steadily (cf. Fig. 5), but very slowly: the 20th iterate represents only a rough approximation. The blame is attributed to clustering points around a solution to equation (1.3), which might render a rational approximation of $F(\eta)$ nearly invalid. A steady and fast convergence of Algorithm 2 is again observed, and the convergence is achieved within 12 iterations. Observing the large amount of noise, the reconstruction agrees well with the exact one; see Fig. 6. Hence, the TPA can provide a rough but visually acceptable estimate, despite the fact that the value η and error grossly overestimate the optimal ones.

4.4 *Example 4: L^2-elastic net* (Two-dimensional image deblurring)

This is the blur example from Hansen (1998). We set the parameters N=50, band=5 and sigma=1.2. The true solution is shown in Fig. 8. We used elastic net, i.e., $\psi(x) = \|x\|_{\ell^1} + \frac{\gamma}{2}\|x\|_{\ell^2}^2$, with a fixed $\gamma = 1\times10^{-3}$, which statistically favours a grouping effect (Zou & Hastie, 2005).

TABLE 4 *Numerical results for Example* 3

(r, ε)	δ^2	$\eta_{\rm opt}$	$\eta_{\rm dp}$	$\eta_{\rm tpa}$	$\eta_{\rm sm}$	$e_{\rm opt}$	$e_{\rm dp}$	$e_{\rm tpa}$	$e_{\rm sm}$
$(0.3, 5\times10^{-3})$	1.30×10^{-1}	5.34×10^{-1}	1.01×10^{0}	$4.72\times10^{0}(20)$	$1.48\times10^{0}(15)$	3.74×10^{-7}	5.72×10^{-7}	5.88×10^{-3}	3.81×10^{-6}
$(0.3, 5\times10^{-2})$	1.30×10^{0}	5.34×10^{-1}	1.19×10^{0}	$8.22\times10^{0}(20)$	$1.61\times10^{0}(15)$	3.72×10^{-7}	4.96×10^{-7}	2.35×10^{-2}	8.76×10^{-6}
$(0.3, 5\times10^{-1})$	1.30×10^{1}	5.34×10^{-1}	1.15×10^{0}	$9.75\times10^{0}(20)$	$1.69\times10^{0}(15)$	3.78×10^{-7}	3.89×10^{-7}	2.95×10^{-2}	1.40×10^{-5}
$(0.5, 5\times10^{-3})$	2.45×10^{-1}	6.13×10^{-1}	7.14×10^{-1}	$6.85\times10^{0}(20)$	$7.14\times10^{-1}(15)$	2.85×10^{-3}	2.94×10^{-3}	2.45×10^{-2}	2.92×10^{-3}
$(0.5, 5\times10^{-2})$	2.45×10^{0}	7.05×10^{-1}	7.81×10^{-1}	$9.41\times10^{0}(20)$	$7.89\times10^{-1}(15)$	1.89×10^{-2}	1.91×10^{-2}	5.44×10^{-2}	1.91×10^{-2}
$(0.5, 5\times10^{-1})$	2.45×10^{1}	8.11×10^{-1}	8.24×10^{-1}	$9.88\times10^{0}(20)$	$8.07\times10^{-1}(15)$	2.89×10^{-2}	2.86×10^{-5}	7.50×10^{-2}	2.91×10^{-2}

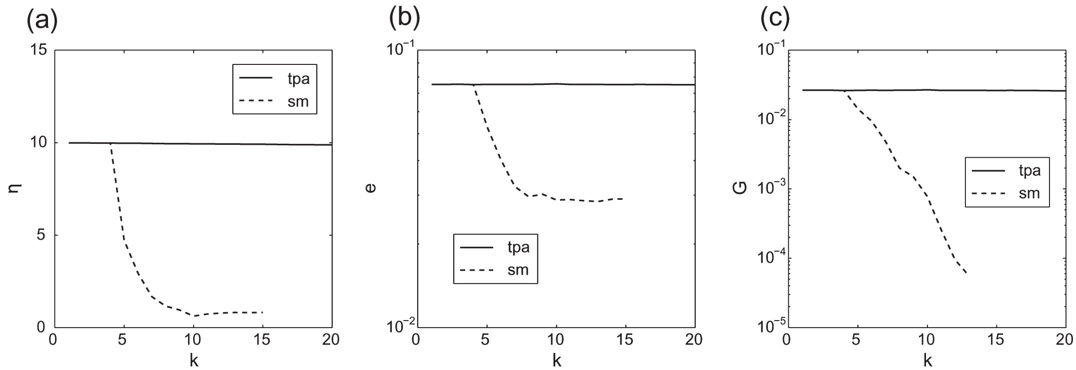

FIG. 5. The convergence of (a) η, (b) e and (c) G for Example 3 with $r = 0.5$ and $\varepsilon = 0.5$ noise.

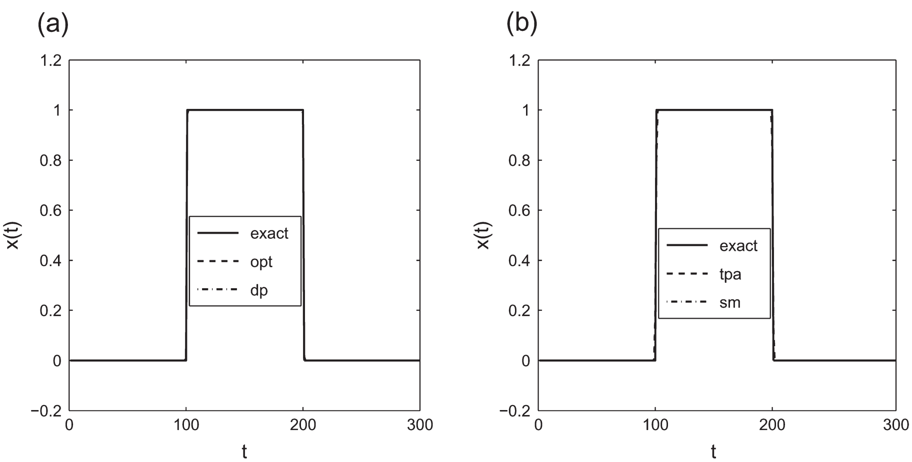

FIG. 6. The numerical solutions for Example 3 with $r = 0.5$ and $\varepsilon = 0.5$ noise: (a) discrepancy principle versus optimal choice and (b) TPA versus hybrid algorithm.

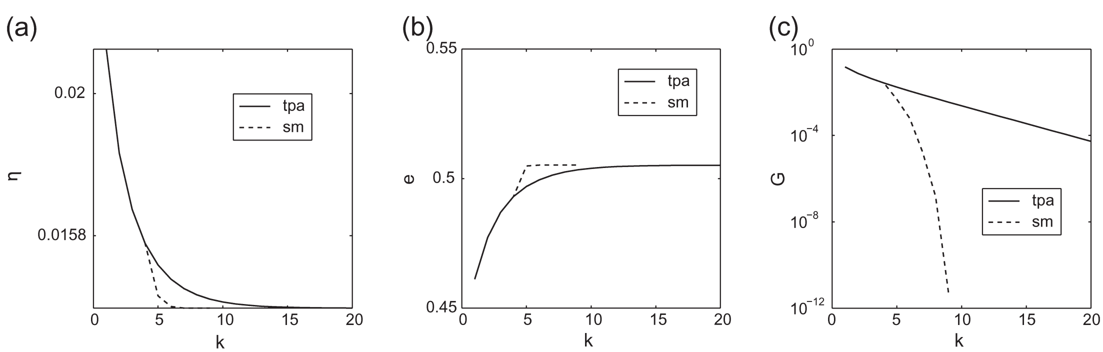

FIG. 7. The convergence of (a) η, (b) e, and (c) G for Example 4 with $\delta^2 = 1.25\times 10^{-1}$.

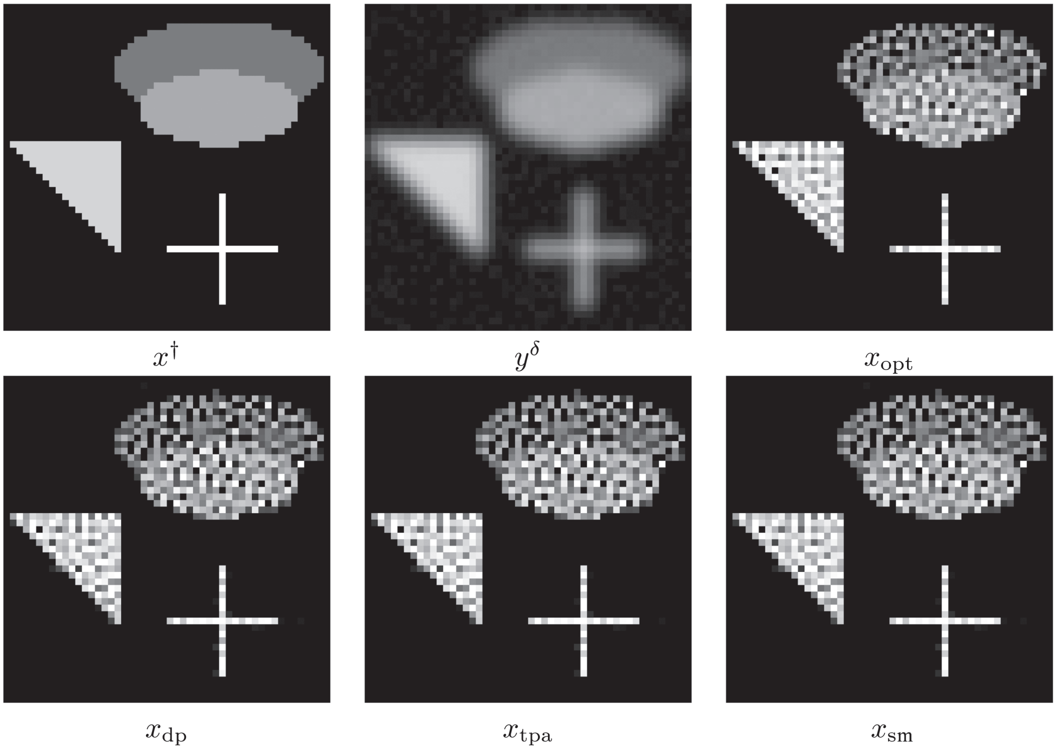

FIG. 8. The numerical solutions for Example 4 with $\delta^2 = 1.25\times10^{-1}$.

The TPA and the secant method both converge very steadily and rapidly, but the local convergence of the TPA is far slower; cf. Fig. 7. Nonetheless, three to four iterations of the TPA already agree well with the exact solution. The parameter $\eta_{\rm dp}$ is 1.41×10^{-2}, and it is well approximated by eight iterations of either algorithm ($\eta_{\rm tpa} = 1.44\times10^{-2}$, $\eta_{\rm sm} = 1.41\times10^{-2}$). It is smaller than the optimal choice, i.e., 3.03×10^{-2}. Visually, all four solutions are reasonable and fairly close to each other; cf. Fig. 8. This is also indicated by the errors, i.e., $e_{\rm opt} = 4.46\times10^{-1}$, $e_{\rm dp} = 5.05\times10^{-1}$, $e_{\rm tpa} = 5.03\times10^{-1}$ and $e_{\rm sm} = 5.05\times10^{-1}$. This again shows the feasibility of the principle and the efficiency of the algorithms.

5. Concluding remarks

We have studied several numerical aspects of the discrepancy principle. The uniqueness of the solution to the discrepancy equation was shown. This was achieved by deriving an explicit characterization of the solution sets. Two numerical algorithms for its efficient realization were proposed, and their convergence properties were also briefly discussed. The TPA was based on the idea of Padé approximation, generalizing the idea of model functions. The hybrid algorithm drastically enhances its local convergence by the secant method. Numerical results show that the algorithms achieve a fast and steady convergence, and the principle achieves almost optimal convergence rates. Theoretically, it is of much interest to further investigate the convergence of the algorithms in the case of inexact minimization the functional J_η, as often occurs in practical scenarios.

Acknowledgements

The work of B. J. was performed during a visit to the Institute of Mathematical Sciences, CUHK, and that of J. Z. substantially supported by Hong Kong RGC grants (Projects 405110 and 404611).

References

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

Burger, M. & Osher, S. (2004) Convergence rates of convex variational regularization. *Inverse Problems*, **20**, 1411–1421.

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

Cioranescu, I. (1990) *Geometry of Banach Spaces, Duality Mappings and Nonlinear Problems*. Dordrecht: Kluwer Academic Publishers Group.

Clason, C., Jin, B. & Kunisch, K. (2010) A semismooth Newton method for L^1 data fitting with automatic choice of regularization parameters and noise calibration. *SIAM J. Imaging Sci.*, **3**, 199–231.

Engl, H. W., Hanke, M. & Neubauer, A. (1996) *Regularization of Inverse Problems*. Dordrecht: Kluwer.

Hansen, P. C. (1998) *Rank-Deficient and Discrete Ill-Posed Problems*. Philadelphia, PA: SIAM.

Hein, T. & Hofmann, B. (2009) Approximate source conditions for nonlinear ill-posed problems—chances and limitations. *Inverse Problems*, **25**:035003, 16.

Heng, Y., Lu, S., Mhamdi, A. & Pereverzev, S. V. (2010) Model functions in the modified L-curve method—case study: the heat flux reconstruction in pool boiling. *Inverse Problems*, **26**:055006, 13.

Hofmann, B., Kaltenbacher, B., Pöschl, C. & Scherzer, O. (2007) A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators. *Inverse Problems*, **23**, 987–1010.

Hofmann, B. & Yamamoto, M. (2010) On the interplay of source conditions and variational inequalities for nonlinear ill-posed problems. *Appl. Anal.*, **89**, 1705–1727.

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

Jin, B., Lorenz, D. A. & Schiffler, S. (2009) Elastic-net regularization: error estimates and active set methods. *Inverse Problems*, **25**:115022, 26.

Jin, B. & Zou, J. (2009) Augmented Tikhonov regularization. *Inverse Problems*, **25**:025001, 25.

Kunisch, K. & Zou, J. (1998) Iterative choices of regularization parameters in linear inverse problems. *Inverse Problems*, **14**, 1247–1264.

Morozov, V. A. (1966) On the solution of functional equations by the method of regularization. *Soviet Math. Dokl.*, **7**, 414–417.

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

Tikhonov, A. N. & Arsenin, V. Y. (1977) *Solutions of Ill-Posed Problems*. New York: John Wiley & Sons.

Tikhonov, A. N., Leonov, A. S. & Yagola, A. G. (1998) *Nonlinear Ill-Posed Problems,* vol. 1, 2. London: Chapman & Hall.

Vogel, C. R. (2002) *Computational Methods for Inverse Problems*. Philadelphia, PA: SIAM.

Xie, J. & Zou, J. (2002) An improved model function method for choosing regularization parameters in linear inverse problems. *Inverse Problems*, **18**, 631–643.

Zou, H. & Hastie, T. (2005) Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **67**, 301–320.