Chin. Ann. Math.
35B(3), 2014, 383–398
DOI: 10.1007/s11401-014-0835-y

Chinese Annals of Mathematics, Series B
© The Editorial Office of CAM and Springer-Verlag Berlin Heidelberg 2014

Multi-parameter Tikhonov Regularization — An Augmented Approach*

Kazufumi ITO[1] Bangti JIN[2] Tomoya TAKEUCHI[3]

Abstract We study multi-parameter regularization (multiple penalties) for solving linear inverse problems to promote simultaneously distinct features of the sought-for objects. We revisit a balancing principle for choosing regularization parameters from the viewpoint of augmented Tikhonov regularization, and derive a new parameter choice strategy called the balanced discrepancy principle. A priori and a posteriori error estimates are provided to theoretically justify the principles, and numerical algorithms for efficiently implementing the principles are also provided. Numerical results on deblurring are presented to illustrate the feasibility of the balanced discrepancy principle.

Keywords Multi-parameter regularization, Augmented Tikhonov regularization, Balanced discrepancy principle
2000 MR Subject Classification 65J20, 65J22, 49N45

1 Introduction

We investigate a regularization technique for solving linear inverse problems modeled by

$$Ku^\dagger = g^\dagger, \tag{1.1}$$

where $g^\dagger$ is the (inaccessible) exact data and $u^\dagger \in X$ represents the unknown exact solution, and $K : X \to Y$ is a bounded linear operator. Here the spaces X and Y are general Banach spaces, and the operator K can be an embedding operator (image denoising), a convolution operator (deblurring, scattering) and the Radon transform (computed tomography). The objective is to find an approximation u to the solution $u^\dagger$ from noisy measurement $g^\delta \in Y$ of the exact data $g^\dagger$. The accuracy of the noisy data g^δ is measured by the standard L^2 fidelity functional $\phi(u^\dagger, g^\delta) = \frac{1}{2}\|Ku^\dagger - g^\delta\|^2 = \frac{1}{2}\delta^2$ with the noise level δ.

As is typical for many inverse problems, problem (1.1) suffers from ill-posedness or instability. This poses significant challenges to their accurate yet stable numerical solution in the presence of data noise, which is often the case in practical applications. Often, regularization is applied to find a stable approximate solution. One of the most widely used approaches is

Manuscript received June 9, 2013.
[1]Center for Research in Scientific Computation & Department of Mathematics, North Carolina State University, Raleigh, NC 27695, USA. E-mail: kito@unity.ncsu.edu
[2]Department of Mathematics, University of California, Riverside, 900 University Ave., Riverside, CA 92521, USA. E-mail: bangti.jin@gmail.com
[3]Collaborative Research Center for Innovative Mathematical Modelling, Institute of Industrial Science, The University of Tokyo 4-6-1-Cw601 Komaba, Meguro-ku, Tokyo 153-8505, Japan. E-mail: takeuchi@sat.t.u-tokyo.ac.jp
*This work was supported by the Army Research Office under DAAD19-02-1-0394, US-ARO grant 49308-MA, and US-AFSOR grant FA9550-06-1-0241.

known as Tikhonov regularization. It seeks to minimize the following functional

$$J_{\boldsymbol{\eta}}(u) = \phi(u, g^\delta) + \boldsymbol{\eta} \cdot \boldsymbol{\psi}(u), \tag{1.2}$$

over a closed convex feasible solution set $\mathcal{C}$. The solution to the minimization problem, denoted by $u_{\boldsymbol{\eta}}^\delta$ ($u_{\boldsymbol{\eta}}$ in case of the exact data $g^\dagger$), serves as an approximation to the exact solution $u^\dagger$. Here the (nonnegative) vector-valued penalty functional $\boldsymbol{\psi}$ encodes the a priori knowledge, and $\boldsymbol{\eta}\cdot\boldsymbol{\psi}(u)$ denotes the dot product between the regularization parameter vector $\boldsymbol{\eta} = (\eta_1, \eta_2)^{\mathrm{t}} \in \mathbb{R}_+^2$ and the penalty $\boldsymbol{\psi}(u) = (\psi_1(u), \psi_2(u))^{\mathrm{t}}$. The penalty $\boldsymbol{\psi}$ is selected to promote desirable features of the sought-for solution, e.g., edge, sparsity and texture; and often the optimization problem (1.2) is nonsmooth. The (vector) parameter $\boldsymbol{\eta}$ compromises the fidelity ϕ with the penalty $\boldsymbol{\psi}$, and its appropriate choice plays a crucial role in obtaining stable yet accurate solutions. Therefore, an automated selection rule and efficient algorithms for determining $\boldsymbol{\eta}$ are essential.

One distinct feature of the model (1.2) is that it includes multiple penalties (hence termed as multi-parameter regularization). This is motivated by the following empirical observations. In practice, many objects exhibit multiple distinct features/structures. However, one single penalty generally favors one feature over others, and thus unsuitable for promoting multiple distinct features. For example, total variation (TV) is well suited to reconstructing piecewise constant structures, however, it results in significant staircases in gray regions. One may improve TV-reconstruction by introducing an additional penalty, say L^1 norm of Δu where Δ is the Laplacian operator. Hence, a reliable recovery of several distinct features naturally calls for multiple penalties, and it is not surprising that the idea of multi-parameter regularization has been pursued earlier. For instance, in [9] the authors proposed a model to preserve both flat and gray regions in natural images by combining TV with Sobolev smooth penalty. We refer interested readers to [15, 17] (imaging), [19] (microarray data analysis), [18] (geodesy) and [13] (machine learning) for other interesting applications.

However, a general theory of multi-parameter regularization remains under development [1, 4, 13, 7]. In [1] the L-hypersurface was suggested for determining regularization parameters for finite-dimensional linear systems, but without any theoretical justification. In [4], a multi-resolution analysis for ill-posed linear operator equations was analyzed, and some convergence results were established. Lu et al. [13] discussed the discrepancy principle for Hilbert space scales, and derived some error estimates. However, the parameter selection is vastly nonunique due to lack of constraints and thus not directly applicable in practice, for which later a quasi-optimality criterion was suggested (see [14]). Recently, the authors [7] investigated the discrepancy principle and a balancing principle for general convex variational models. However, the nonuniqueness of the discrepancy principle remains unresolved, and further, there is still no theory for the balancing principle for multi-parameter regularization.

The present work extends our earlier work [7], and includes the following essential contributions. We first revisit the balancing principle in [7] from the viewpoint of augmented Tikhonov regularization (see [12]), and establish the equivalence. Then we derive a novel hybrid principle, the balanced discrepancy principle, by incorporating constraints into the augmented approach, which partially resolves the nonuniqueness issue. Further, a priori and a posterior error estimates are derived for both principles. The estimate in Theorem 2.4 was stated in [7] without a proof. Finally, we develop efficient algorithms for implementing these principles, and briefly discuss their properties.

The rest of the paper is organized as follows. In Section 2, we derive the balancing principle and the new hybrid principle, and develop relevant error estimates. In Section 3 we discuss efficient implementations of the two principles. Finally, we provide some numerical results to illustrate the hybrid principle in Section 4.

2 An Augmented Approach

The augmented Tikhonov (a-Tikhonov) regularization is one principled framework for choosing regularization parameters (see [12]). Here we describe the augmented approach for multi-parameter models, and derive the balancing principle and a novel balanced discrepancy principle.

2.1 Derivation of the principles

2.1.1 Balancing principle

First we sketch the augmented approach. For the multi-parameter model (1.2), it can be derived analogously from hierarchical Bayesian inference as in [12], and the resulting augmented functional $J(u,\tau,\boldsymbol{\lambda})$ reads

$$J(u,\tau,\boldsymbol{\lambda}) = \tau\phi(u,g^\delta) + \boldsymbol{\lambda}\cdot\boldsymbol{\psi}(u) + \boldsymbol{e}\cdot(\beta\boldsymbol{\lambda} - \alpha\ln\boldsymbol{\lambda}) + \beta_0\tau - \alpha_0\ln\tau,$$

where the vector $\boldsymbol{e}$ is given by $\boldsymbol{e} = (1,1)^{\mathrm{t}}$. The functional $J(u,\tau,\boldsymbol{\lambda})$ maximizes the posterior probability density function

$$p(u,\tau,\boldsymbol{\lambda}|g^\delta) \propto p(g^\delta|u,\tau,\boldsymbol{\lambda})\,p(u,\tau,\boldsymbol{\lambda}).$$

The functional $J(u,\tau,\boldsymbol{\lambda})$ is derived under the assumption that the scalars λ_i and τ have Gamma distributions with known parameter pairs. The parameter pairs (α,β) and (α_0,β_0) are related to the shape parameters in the statistical priors on the prior precision λ_i and noise precision τ, respectively. The special case $\beta_0 = \beta = 0$ is known as noninformative prior and customarily adopted in practice. Hence we focus our derivation on this case. Upon letting $\eta_i = \frac{\lambda_i}{\tau}$, the necessary optimality condition of any minimizer $(u_{\boldsymbol{\eta}}^\delta,\lambda_i,\tau)$ to the a-Tikhonov functional $J(u,\tau,\boldsymbol{\lambda})$ is given by

$$\begin{cases} u_{\boldsymbol{\eta}}^\delta = \arg\min\limits_{u\in\mathcal{C}} \{\phi(u,g^\delta) + \boldsymbol{\eta}\cdot\boldsymbol{\psi}(u)\}, \\ \lambda_i = \dfrac{\alpha}{\psi_i(u_{\boldsymbol{\eta}}^\delta)}, \quad i=1,2, \\ \tau = \dfrac{\alpha_0}{\phi(u_{\boldsymbol{\eta}}^\delta,g^\delta)}. \end{cases} \tag{2.1}$$

Now by rewriting the system with $\gamma = \frac{\alpha_0}{\alpha}$, we arrive at the following system for $(u_{\boldsymbol{\eta}}^\delta,\boldsymbol{\eta})$:

$$\begin{cases} u_{\boldsymbol{\eta}}^\delta = \arg\min\limits_{u\in\mathcal{C}} \{\phi(u,g^\delta) + \boldsymbol{\eta}\cdot\boldsymbol{\psi}(u)\}, \\ \eta_i = \dfrac{1}{\gamma}\dfrac{\phi(u_{\boldsymbol{\eta}}^\delta,g^\delta)}{\psi_i(u_{\boldsymbol{\eta}}^\delta)}, \quad i=1,2. \end{cases} \tag{2.2}$$

The optimality system (2.2) reveals the mechanism of the augmented approach: It selects an optimal regularization parameter $\boldsymbol{\eta}$ in the model (1.2) by balancing the penalty $\boldsymbol{\psi}$ with the fidelity ϕ, from which the term balancing principle follows. We note the term balancing principle here should not be confused with Lepskii's principle, which is also sometimes called a balancing principle (see [16]). The Lepskii's principle does require a knowledge of noise level.

Next we characterize (2.2) using the value function $F(\boldsymbol{\eta})$ (see [8]) defined by

$$F(\boldsymbol{\eta}) = \inf_{u\in\mathcal{C}} J_{\boldsymbol{\eta}}(u).$$

The function $F(\boldsymbol{\eta})$ is continuous, and it is almost everywhere differentiable, cf. Lemma 2.1. We denote by F_{η_i} the partial derivative of $F(\boldsymbol{\eta})$ with respect to η_i. The proof is analogous to [8], and hence omitted.

Lemma 2.1 *The function $F(\boldsymbol{\eta})$ is monotone and concave, and hence almost everywhere differentiable. Further, if it is differentiable, then there holds*

$$F_{\eta_i}(\boldsymbol{\eta}) = \psi_i(u_{\boldsymbol{\eta}}^\delta).$$

Next we provide an alternative characterization of (2.2). First we define the function $\Phi_\gamma(\boldsymbol{\eta})$ by

$$\Phi_\gamma(\boldsymbol{\eta}) = \frac{F(\boldsymbol{\eta})^{\gamma+2}}{\eta_1\eta_2}. \tag{2.3}$$

The necessary optimality condition for $\Phi_\gamma(\boldsymbol{\eta})$, provided that $F(\boldsymbol{\eta})$ is differentiable, reads

$$\frac{\partial\Phi_\gamma}{\partial\eta_i} = \frac{F(\boldsymbol{\eta})^{\gamma+1}}{\eta_1\eta_2}\frac{(-F(\boldsymbol{\eta}) + (2+\gamma)\eta_i F_{\eta_i}(\boldsymbol{\eta}))}{\eta_i} = 0, \quad i = 1, 2,$$

which, upon noting Lemma 2.1, is equivalent to

$$\begin{cases} -\phi(u_{\boldsymbol{\eta}}^\delta, g^\delta) + (1+\gamma)\eta_1\psi_1(u_{\boldsymbol{\eta}}^\delta) - \eta_2\psi_2(u_{\boldsymbol{\eta}}^\delta) = 0, \\ -\phi(u_{\boldsymbol{\eta}}^\delta, g^\delta) - \eta_1\psi_1(u_{\boldsymbol{\eta}}^\delta) + (1+\gamma)\eta_2\psi_2(u_{\boldsymbol{\eta}}^\delta) = 0. \end{cases}$$

Solving the system with respect to η_i yields $\eta_i = \frac{1}{\gamma}\frac{\phi(u_{\boldsymbol{\eta}}^\delta, g^\delta)}{\psi_i(u_{\boldsymbol{\eta}}^\delta)}$. Hence, the optimality system of the function Φ_γ coincides with that of the functional $J(u, \tau, \boldsymbol{\lambda})$. In summary, we have shown our first main result.

Proposition 2.1 *Let the value function $F(\boldsymbol{\eta})$ be differentiable. Then all critical points of the function Φ_γ are solutions to system* (2.2).

Remark 2.1 Two remarks on Φ_γ are in order. First, it is very flexible in that the parameter γ may be calibrated to achieve specific desirable properties. Second, by the concavity in Lemma 2.1, $F(\boldsymbol{\eta})$ is continuous and thus the problem of minimizing Φ_γ over any bounded and closed region in $\mathbb{R}^2_+$ is well defined. These observations are valid for a general fidelity.

2.1.2 Balanced discrepancy principle

To solve stably and accurately problem (1.1), one should use all prior information, e.g., the noise level $\phi(u^\dagger, g^\delta) = c := \frac{1}{2}c_m^2\delta^2$ for some $c_m \geq 1$, and other relevant knowledge, whenever it is

available. This can be realized by incorporating constraints into the augmented approach, and then deriving the corresponding optimal system. For instance, for the constraint $\phi(u, g^\delta) \le c$, the Lagrangian approach gives the following a-Tikhonov functional:

$$J(u, \tau, \boldsymbol{\lambda}, \mu) = \tau\phi(u, g^\delta) + \boldsymbol{\lambda} \cdot \boldsymbol{\psi}(u) - \alpha \boldsymbol{e} \cdot \ln \boldsymbol{\lambda} - \alpha_0 \ln \tau + \tau\langle \phi(u, g^\delta) - c, \mu\rangle,$$

where the unknown scalar $\mu \ge 0$ is the Lagrange multiplier for the inequality constraint $\phi(u, g^\delta) \le c$. Its optimality system reads

$$\begin{cases} u_{\boldsymbol{\eta}}^\delta = \arg\min\limits_u \{\phi(u, g^\delta) + \boldsymbol{\eta} \cdot \boldsymbol{\psi}(u) + \langle \phi(u, g^\delta) - c, \mu\rangle\}, \\ \lambda_i = \dfrac{\alpha}{\psi_i(u_{\boldsymbol{\eta}}^\delta)}, \quad i = 1, 2, \\ \tau = \dfrac{\alpha_0}{(1+\mu)\phi(u_{\boldsymbol{\eta}}^\delta, g^\delta)}, \\ c \ge \phi(u_{\boldsymbol{\eta}}^\delta, g^\delta), \quad \mu \ge 0. \end{cases}$$

Hence the constraint $\phi(u_{\boldsymbol{\eta}}^\delta, g^\delta) \le c$ and the balancing principle are both fulfilled:

$$\gamma\eta_i\psi_i(u_{\boldsymbol{\eta}}^\delta) = (1+\mu)\phi(u_{\boldsymbol{\eta}}^\delta, g^\delta), \quad i = 1, 2. \tag{2.4}$$

In the case of one single penalty, identity (2.4) does not provide any additional constraint since the multiplier μ is also unknown. We observe that the active constraint, i.e., $\|Ku_{\boldsymbol{\eta}}^\delta - g^\delta\| = c_m\delta$, is exactly the discrepancy principle (see [5]). The constraint is active under certain conditions (see [10]). Nonetheless, in case of multiple penalties, the discrepancy principle alone cannot uniquely determine $\boldsymbol{\eta}$. Hence we include also system (2.4), which might help resolve the nonuniqueness issue. Upon simplification, this yields a new hybrid principle

$$\begin{cases} \phi(u_{\boldsymbol{\eta}}^\delta, g^\delta) = \dfrac{1}{2} c_m^2 \delta^2, \\ \eta_1\psi_1(u_{\boldsymbol{\eta}}^\delta) = \eta_2\psi_2(u_{\boldsymbol{\eta}}^\delta). \end{cases} \tag{2.5}$$

The principle can be interpreted as the augmented approach with the constraint $\{u : \|Ku - g^\delta\| = c_m\delta\}$, $c_m \ge 1$. Hence it integrates the classical discrepancy principle $\|Ku_{\boldsymbol{\eta}}^\delta - g^\delta\| = c_m\delta$ with the balancing principle, and we shall name the new rule (2.5) balanced discrepancy principle. One noteworthy feature of (2.5) is that it does not involve the free parameter γ.

2.2 Error estimates

Now we derive error estimates for (2.3) and (2.5), capitalizing on [3, 5–6]. We discuss the following three scenarios separately: hybrid principle (2.5), purely balancing principle (2.3) in Hilbert and Banach spaces. These theoretical results partially justify their practical usages.

2.2.1 Balanced discrepancy principle

In this part, we discuss the consistency and an a priori error estimate for the hybrid principle (2.5). To this end, we make the following assumption.

Assumption 2.1 *There exists a τ-topology such that for any $\boldsymbol{\eta} > 0$, the functional $J_{\boldsymbol{\eta}}(u)$ is coercive and its level set $\{u \in \mathcal{C} : J_{\boldsymbol{\eta}}(u) \le c\}$ for any $c > 0$ is compact in τ-topology, and the functionals ϕ and ψ_i are τ lower semi-continuous.*

Remark 2.2 The τ-topology is naturally induced by the penalty functional $\boldsymbol{\psi}$, and it is not arbitrarily in order to ensure the lower semicontinuity.

Now we can state a consistency result. The line of proof is standard (see [7]), and thus omitted.

Theorem 2.1 *Let Assumption* 2.1 *be fulfilled, and* $t(\boldsymbol{\eta}) = \frac{\eta_1(\delta)}{\eta_1(\delta)+\eta_2(\delta)}$. *Let the sequence* $\{\boldsymbol{\eta}(\delta)\}_\delta$ *be selected by* (2.5). *If a subsequence of* $\{\boldsymbol{\eta}(\delta)\}_\delta$ *converges and* $\widetilde{t} := \lim_{\delta\to 0} t(\delta) \in (0,1)$, *then the subsequence* $\{u^\delta_{\boldsymbol{\eta}(\delta)}\}_\delta$ *contains a subsequence* τ*-converging to a* $[\widetilde{t}, 1-\widetilde{t}]^{\mathrm{t}} \cdot \boldsymbol{\psi}$*-minimizing solution of* $Ku = g^\dagger$ *and*

$$\lim_{\delta\to 0}[t(\delta), 1-t(\delta)]^{\mathrm{t}} \cdot \boldsymbol{\psi}(u^\delta_{\boldsymbol{\eta}}) = [\widetilde{t}, 1-\widetilde{t}]^{\mathrm{t}} \cdot \boldsymbol{\psi}(u^\dagger).$$

Remark 2.3 The condition $\widetilde{t} \in (0,1)$ in Theorem 2.1 is tantamount to the uniform boundedness of the penalties $\psi_i(u^\delta_{\boldsymbol{\eta}})$.

Next we have the following convergence rate, i.e., the distance between the approximation $u^\delta_{\boldsymbol{\eta}}$ and the true solution $u^\dagger$ (in Bregman distance (see [3])) in terms of the noise level δ. We denote the subdifferential of a convex functional $\psi(u)$ at $u^\dagger$ by $\partial\psi(u^\dagger)$, i.e.,

$$\partial\psi(u^\dagger) = \{\xi \in X^* : \psi(u) \geq \psi(u^\dagger) + \langle \xi, u-u^\dagger\rangle,\ \forall u \in X\},$$

and the Bregman distance $d_\xi(u, u^\dagger)$ for any $\xi \in \partial\psi(u^\dagger)$ is defined as

$$d_\xi(u, u^\dagger) := \psi(u) - \psi(u^\dagger) - \langle \xi, u-u^\dagger\rangle.$$

Now we can state a convergence rates result.

Theorem 2.2 *Let the exact solution* $u^\dagger$ *satisfy the source condition: For any* $t \in [0,1]$, *there exists a* $w_t \in Y$ *such that*

$$K^* w_t = \xi_t \in \partial([t, 1-t]^{\mathrm{t}} \cdot \boldsymbol{\psi}(u^\dagger)).$$

Then for any $\boldsymbol{\eta}^*$ *determined by the principle* (2.5) *and with*

$$t^* = t(\boldsymbol{\eta}^*) = \frac{\eta_1^*(\delta)}{\eta_1^*(\delta) + \eta_2^*(\delta)} \in [0,1],$$

the following estimate holds:

$$d_{\xi_{t^*}}(u^\delta_{\boldsymbol{\eta}^*}, u^\dagger) \leq (1+c_m)\|w_{t^*}\|\delta.$$

Proof The line of proof is again well known, but we include a sketch for completeness. In view of the minimizing property of the approximation $u^\delta_{\boldsymbol{\eta}^*}$ and the constraint $\|Ku^\delta_{\boldsymbol{\eta}^*} - g^\delta\| = c_m\delta$, we have

$$[t^*, 1-t^*]^{\mathrm{t}} \cdot \boldsymbol{\psi}(u^\delta_{\boldsymbol{\eta}^*}) \leq [t^*, 1-t^*]^{\mathrm{t}} \cdot \boldsymbol{\psi}(u^\dagger).$$

The source condition implies that there exists a $\xi_{t^*} \in \partial([t^*, 1-t^*]^{\mathrm{t}} \cdot \boldsymbol{\psi}(u^\dagger))$ and $w_{t^*} \in Y$ such that $\xi_{t^*} = K^* w_{t^*}$. From this and the Cauchy-Schwarz inequality, we deduce

$$\begin{aligned} d_{\xi_{t^*}}(u^\delta_{\boldsymbol{\eta}^*}, u^\dagger) &= [t^*, 1-t^*]^{\mathrm{t}} \cdot \boldsymbol{\psi}(u^\delta_{\boldsymbol{\eta}^*}) - [t^*, 1-t^*]^{\mathrm{t}} \cdot \boldsymbol{\psi}(u^\dagger) - \langle \xi_{t^*}, u^\delta_{\boldsymbol{\eta}^*} - u^\dagger\rangle \\ &\leq -\langle \xi_{t^*}, u^\delta_{\boldsymbol{\eta}^*} - u^\dagger\rangle = -\langle K^* w_{t^*}, u^\delta_{\boldsymbol{\eta}^*} - u^\dagger\rangle = -\langle w_{t^*}, K(u^\delta_{\boldsymbol{\eta}^*} - u^\dagger)\rangle \\ &\leq \|w_{t^*}\|\|K(u^\delta_{\boldsymbol{\eta}^*} - u^\dagger)\| \leq \|w_{t^*}\|(\|Ku^\delta_{\boldsymbol{\eta}^*} - g^\delta\| + \|g^\delta - Ku^\dagger\|) \leq (1+c_m)\|w_{t^*}\|\delta. \end{aligned}$$

This shows the desired estimate.

Remark 2.4 In Theorem 2.2, the order of convergence relies solely on the constraint $\|Ku_{\boldsymbol{\eta}}^{\delta}-g^{\delta}\|=c_m\delta$, while the weight t^* is determined by the balancing principle. Hence the reduced system (2.4) does help resolve the vast nonuniqueness issue in the discrepancy principle.

2.2.2 Balancing principle in Hilbert spaces

We derive a posteriori estimates for the balancing principle Φ_γ (2.3), i.e., the distance between the approximation $u_{\boldsymbol{\eta}^*}^{\delta}$ and the exact solution $u^\dagger$ in terms of the noise level $\delta=\|g^\delta-g^\dagger\|$ and the realized residual $\delta_*=\|Ku_{\boldsymbol{\eta}^*}^{\delta}-g^{\delta}\|$ etc. We first treat quadratic regularizations $\psi_i(u)=\frac{1}{2}\|L_iu\|^2$ with linear operators L_i fulfilling $\ker(L_i)\cap\ker(K)=\{0\}$, $i=1,2$, and each induces a semi-norm. One typical choice is that ψ_1 and ψ_2 impose the L^2-norm and higher-order Sobolev smoothness, e.g., $\psi_1(u)=\frac{1}{2}\|u\|_{\mathrm{L}^2}^2$ and $\psi_2(u)=\frac{1}{2}\|u\|_{\mathrm{H}^1}^2$. We shall utilize a weighted (semi-)norm $\|\cdot\|_t$ defined by

$$\|u\|_t^2=t\|L_1u\|^2+(1-t)\|L_2u\|^2,$$

where the weight $t\equiv t(\boldsymbol{\eta})\in[0,1]$ is defined as before, and by $Q_t=tL_1^*L_1+(1-t)L_2^*L_2$ and $L_t=Q_t^{\frac{1}{2}}$ and $\widetilde{K}_t=KL_t^{-1}$. Clearly, $\|u\|_t=\|L_tu\|$. We note that the adjoint K^* (and hence $\widetilde{K}_t^*$) depends on the value t.

Theorem 2.3 *Let $\mu\in(0,1]$ be fixed, and the exact solution $u^\dagger$ satisfy the source condition: For any $t\in[0,1]$, there exists a $w_t\in Y$ such that $L_tu^\dagger=(\widetilde{K}_t^*\widetilde{K}_t)^\mu w_t$. Then for any parameter $\boldsymbol{\eta}^*$ selected by (2.3) with $t^*=t(\boldsymbol{\eta}^*)=\frac{\eta_1^*(\delta)}{\eta_1^*(\delta)+\eta_2^*(\delta)}$, the following estimate holds:*

$$\|u_{\boldsymbol{\eta}^*}^{\delta}-u^\dagger\|_{t^*}\le C\Big(\|w_{t^*}\|^{\frac{1}{2\mu+1}}+\frac{F^{\frac{2+\gamma}{4}}(\delta^{\frac{2}{2\mu+1}}\boldsymbol{e})}{F^{\frac{2+\gamma}{4}}(\boldsymbol{\eta}^*)}\Big)\max\{\delta_*,\delta\}^{\frac{2\mu}{2\mu+1}}.$$

Proof We decompose the error $u_{\boldsymbol{\eta}}^{\delta}-u^\dagger$ into $u_{\boldsymbol{\eta}}^{\delta}-u^\dagger=(u_{\boldsymbol{\eta}}^{\delta}-u_{\boldsymbol{\eta}})+(u_{\boldsymbol{\eta}}-u^\dagger)$, and bound the two terms separately. First we estimate the error $u_{\boldsymbol{\eta}}^{\delta}-u_{\boldsymbol{\eta}}$. It follows from the optimality conditions for $u_{\boldsymbol{\eta}}$ and $u_{\boldsymbol{\eta}}^{\delta}$ that

$$(K^*K+\eta_1L_1^*L_1+\eta_2L_2^*L_2)(u_{\boldsymbol{\eta}}-u_{\boldsymbol{\eta}}^{\delta})=K^*(g^\dagger-g^\delta).$$

Multiplying the identity with $u_{\boldsymbol{\eta}}-u_{\boldsymbol{\eta}}^{\delta}$ and using the Cauchy-Schwarz and Young's inequalities give

$$\begin{aligned}&\|K(u_{\boldsymbol{\eta}}^{\delta}-u_{\boldsymbol{\eta}})\|^2+\eta_1\|L_1(u_{\boldsymbol{\eta}}^{\delta}-u_{\boldsymbol{\eta}})\|^2+\eta_2\|L_2(u_{\boldsymbol{\eta}}^{\delta}-u_{\boldsymbol{\eta}})\|^2\\ =\;&\langle K(u_{\boldsymbol{\eta}}^{\delta}-u_{\boldsymbol{\eta}}),g^\dagger-g^\delta\rangle\\ \le\;&\|K(u_{\boldsymbol{\eta}}^{\delta}-u_{\boldsymbol{\eta}})\|^2+\frac{1}{4}\|g^\dagger-g^\delta\|^2.\end{aligned}$$

Next let $s=\eta_1+\eta_2$. Then we get

$$\|u_{\boldsymbol{\eta}}^{\delta}-u_{\boldsymbol{\eta}}\|_t\le\frac{\|g^\delta-g^\dagger\|}{2\sqrt{s}}\le\frac{\delta}{2\sqrt{s}}\le\frac{\delta}{2\sqrt{\max_i\eta_i}}.$$

Meanwhile, the minimizing property of $\boldsymbol{\eta}^*$ to the rule Φ_γ implies that for any $\widehat{\boldsymbol{\eta}}$,

$$\frac{F^{2+\gamma}(\boldsymbol{\eta}^*)}{\max(\eta_i^*)^2}\le\frac{F^{2+\gamma}(\boldsymbol{\eta}^*)}{\eta_1^*\eta_2^*}\le\frac{F^{2+\gamma}(\widehat{\boldsymbol{\eta}})}{\widehat{\eta}_1\widehat{\eta}_2}.$$

In particular, we may take $\widehat{\boldsymbol{\eta}} = \delta^{\frac{2}{2\mu+1}}\boldsymbol{e}$ and arrive at

$$\|u^\delta_{\boldsymbol{\eta}^*} - u_{\boldsymbol{\eta}}\|_{t^*} \le \frac{F^{\frac{2+\gamma}{4}}(\delta^{\frac{2}{2\mu+1}}\boldsymbol{e})}{F^{\frac{2+\gamma}{4}}(\boldsymbol{\eta}^*)}\delta^{\frac{2\mu}{2\mu+1}}.$$

Next we estimate the approximation error $u_{\boldsymbol{\eta}} - u^\dagger$. To this end, we observe

$$\begin{aligned} u_{\boldsymbol{\eta}} - u^\dagger &= (K^*K + \eta_1 L_1^*L_1 + \eta_2 L_2^*L_2)^{-1}(\eta_1 L_1^*L_1 + \eta_2 L_2^*L_2)u^\dagger \\ &= s(K^*K + sQ_t)^{-1}Q_t u^\dagger = sL_t^{-1}(L_t^{-1}K^*KL_t^{-1} + sI)^{-1}L_t u^\dagger. \end{aligned}$$

Hence,

$$L_t(u_{\boldsymbol{\eta}} - u^\dagger) = s(\widetilde{K}_t^*\widetilde{K}_t + sI)L_t u^\dagger.$$

Consequently, we deduce from the source condition and the moment inequality (see [5])

$$\begin{aligned} \|u_{\boldsymbol{\eta}} - u^\dagger\|_t &= \|L_t(u_{\boldsymbol{\eta}} - u^\dagger)\| = \|s(\widetilde{K}_t^*\widetilde{K}_t + sI)^{-1}L_t u^\dagger\| \\ &= \|s(\widetilde{K}_t^*\widetilde{K}_t + sI)^{-1}(\widetilde{K}_t^*\widetilde{K}_t)^\mu w_t\| \\ &\le \|s(\widetilde{K}_t^*\widetilde{K}_t + sI)^{-1}(\widetilde{K}_t^*\widetilde{K}_t)^{\frac{1}{2}+\mu}w_t\|^{\frac{2\mu}{2\mu+1}}\|s(\widetilde{K}_t^*\widetilde{K}_t + sI)^{-1}w_t\|^{\frac{1}{2\mu+1}} \\ &= \|s(\widetilde{K}_t^*\widetilde{K}_t + sI)^{-1}\widetilde{K}_t L_t u^\dagger\|^{\frac{2\mu}{2\mu+1}}\|s(\widetilde{K}_t^*\widetilde{K}_t + sI)^{-1}w_t\| \\ &\le c(\|s(\widetilde{K}_t\widetilde{K}_t^* + sI)^{-1}g^\delta\| + \|s(\widetilde{K}_t\widetilde{K}_t^* + sI)^{-1}(g^\delta - g^\dagger)\|)^{\frac{2\mu}{2\mu+1}}\|w_t\|^{\frac{1}{2\mu+1}}, \end{aligned}$$

where the constant c depends only on the maximum of $r_s(t) = \frac{s}{s+t}$ over $[0, \|\widetilde{K}_t\|^2]$. Further, we note the relation

$$\begin{aligned} s(\widetilde{K}_t\widetilde{K}_t^* + sI)^{-1}g^\delta &= g^\delta - (\widetilde{K}_t K_t^* + sI)^{-1}\widetilde{K}_t\widetilde{K}_t^* g^\delta \\ &= g^\delta - \widetilde{K}(\widetilde{K}_t^*\widetilde{K}_t + sI)^{-1}\widetilde{K}_t^* g^\delta \\ &= g^\delta - K(K^*K + sQ_t)^{-1}K^* g^\delta = g^\delta - Ku^\delta_{\boldsymbol{\eta}}. \end{aligned}$$

Hence, we deduce

$$\|u_{\boldsymbol{\eta}^*} - u^\dagger\|_{t^*} \le c(\delta_* + c\delta)^{\frac{2\mu}{2\mu+1}}\|w_t\|^{\frac{1}{2\mu+1}} \le c_1 \max\{\delta_*, \delta\}^{\frac{2\mu}{2\mu+1}}.$$

By combining these two estimates, we arrive at the desired inequality.

2.2.3 Balancing principle in Banach space

Lastly, we turn to the balancing principle for general convex regularization $\boldsymbol{\psi}$. We first recall the following technical lemma (see [11]) for single convex regularization ψ. The first estimates the propagation error, and the second plays the role of a triangle inequality.

Lemma 2.2 (see [11]) *Let the exact solution $u^\dagger$ satisfy the following source condition: There exists a $w \in Y$ such that $K^*w = \xi \in \partial\psi(u^\dagger)$, and let $\xi_\eta = K^*(g^\dagger - Ku_\eta)\frac{1}{\eta}$. Then there hold*

$$\begin{aligned} &d_{\xi_\eta}(u^\delta_\eta, u_\eta) \le \frac{\delta^2}{2\eta} \quad \textit{and} \quad \|K(u^\delta_\eta - u_\eta)\| \le 2\delta, \\ &|d_\xi(u^\delta_\eta, u^\dagger) - (d_{\xi_\eta}(u^\delta_\eta, u_\eta) + d_\xi(u_\eta, u^\dagger))| \le 6\|w\|\delta. \end{aligned}$$

Now we can state an estimate for the balancing principle (2.3) in Banach spaces. The estimate has been stated in [7] but without a proof.

Theorem 2.4 *Let the exact solution $u^\dagger$ satisfy the source condition: For any $t \in [0,1]$ there exists a $w_t \in Y$ such that*

$$K^* w_t = \xi_t \in \partial([t, 1-t]^{\mathrm{t}} \cdot \boldsymbol{\psi}(u^\dagger)).$$

Then for every $\boldsymbol{\eta}^*$ selected by (2.3) and with $t^* = t(\boldsymbol{\eta}^*) = \frac{\eta_1^*(\delta)}{\eta_1^*(\delta)+\eta_2^*(\delta)}$, the following estimate holds:

$$d_{\xi_{t^*}}(u_{\boldsymbol{\eta}^*}^\delta, u^\dagger) \le C\Big(\|w_{t^*}\| + \frac{F^{1+\frac{\gamma}{2}}(\delta \boldsymbol{e})}{F^{1+\frac{\gamma}{2}}(\boldsymbol{\eta}^*)}\Big) \max(\delta, \delta_*).$$

Proof For any $t \in [0,1]$, let $\psi_t(u^\dagger) = [t, 1-t]^{\mathrm{t}} \cdot \boldsymbol{\psi}(u^\dagger)$ and $\xi_t \in \partial \psi_t(u^\dagger)$, with ξ_t and w_t being the subgradient and the representer in the source condition, respectively. By Lemma 2.2, we have that for $\boldsymbol{\eta}$

$$d_{\xi_t}(u_{\boldsymbol{\eta}}^\delta, u^\dagger) \le d_{\xi_{\boldsymbol{\eta}}}(u_{\boldsymbol{\eta}}^\delta, u_{\boldsymbol{\eta}}) + d_{\xi_t}(u_{\boldsymbol{\eta}}, u^\dagger) + 6\|w_t\|\delta,$$

where $\xi_{\boldsymbol{\eta}} = -K^*(K(u_{\boldsymbol{\eta}}) - g^\dagger)\frac{1}{s} \in \partial \psi_t(u_{\boldsymbol{\eta}})$ and $s = \eta_1 + \eta_2$. It suffices to bound the terms involving Bregman distance. We first estimate the approximation error $d_{\xi_t}(u_{\boldsymbol{\eta}}, u^\dagger)$. To this end, observe by the minimizing property of the element $u_{\boldsymbol{\eta}}$, i.e.,

$$\frac{1}{2}\|Ku_{\boldsymbol{\eta}} - g^\dagger\|^2 + s\psi_t(u_{\boldsymbol{\eta}}) \le \frac{1}{2}\|Ku^\dagger - g^\dagger\|^2 + s\psi_t(u^\dagger) = s\psi_t(u^\dagger).$$

This inequality, the definition of $d_{\xi_t}(u_{\boldsymbol{\eta}}, u^\dagger)$, the source condition and Lemma 2.2 imply

$$\begin{aligned} d_{\xi_t}(u_{\boldsymbol{\eta}}, u^\dagger) &\le -\langle w_t, K(u_{\boldsymbol{\eta}} - u^\dagger)\rangle \\ &\le \|w_t\|\|Ku_{\boldsymbol{\eta}} - g^\dagger\| \\ &\le \|w_t\|(\|K(u_{\boldsymbol{\eta}} - u_{\boldsymbol{\eta}}^\delta)\| + \|Ku_{\boldsymbol{\eta}}^\delta - g^\delta\| + \|g^\delta - g^\dagger\|) \\ &\le \|w_t\|(2\delta + \delta_* + \delta) \le 4\|w_t\| \max(\delta, \delta_*). \end{aligned}$$

Next we estimate the term $d_{\xi_{\boldsymbol{\eta}}}(u_{\boldsymbol{\eta}}^\delta, u_{\boldsymbol{\eta}})$. In view of Lemma 2.2, we have

$$d_{\xi_{\boldsymbol{\eta}}}(u_{\boldsymbol{\eta}}^\delta, u_{\boldsymbol{\eta}}) \le \frac{\delta^2}{2s} \le \frac{\delta^2}{2\max(\eta_i)}.$$

Meanwhile, the minimizing property of $\boldsymbol{\eta}$ to the rule Φ_γ gives that for any $\widehat{\boldsymbol{\eta}}$,

$$\frac{F^{2+\gamma}(\boldsymbol{\eta})}{\max(\eta_i)^2} \le \frac{F^{2+\gamma}(\boldsymbol{\eta})}{\eta_1\eta_2} \le \frac{F^{2+\gamma}(\widehat{\boldsymbol{\eta}})}{\widehat{\eta}_1\widehat{\eta}_2}.$$

Upon letting $\widehat{\boldsymbol{\eta}} = \delta \boldsymbol{e}$ and combining the preceding two inequalities, we get

$$d_{\xi_{\boldsymbol{\eta}}}(u_{\boldsymbol{\eta}}^\delta, u_{\boldsymbol{\eta}}) \le \frac{F(\delta \boldsymbol{e})^{1+\frac{\gamma}{2}}}{F(\boldsymbol{\eta})^{1+\frac{\gamma}{2}}}\frac{\delta}{2}.$$

Now combining these three estimates gives the desired assertion.

The a posteriori error estimate in Theorem 2.4 coincides with that for a priori choice, e.g., $\boldsymbol{\eta} \sim \delta \boldsymbol{e}$, if the realized discrepancy δ_* is of the same order with the exact noise level δ.

3 Numerical Algorithms

Now we describe the algorithms for numerically realizing the hybrid principle and the balancing principle, i.e., Broyden's method and fixed-point algorithm, and discuss their properties.

3.1 Broyden's method

In practice, the application of the hybrid principle invokes solving the nonlinear system (2.5), which is nontrivial due to its potential nonsmoothness and high degree of nonlinearity. We propose using Broyden's method (see [2]) for its efficient solution (see Algorithm 1 for a complete description).

For the numerical treatment, we reformulate system (2.5) equivalently as

$$\mathbf{T}(\boldsymbol{\eta}) \equiv \begin{pmatrix} \phi(u_{\boldsymbol{\eta}}^{\delta}, g^{\delta}) - \dfrac{1}{2}\delta^2 + \eta_2\psi_2(u_{\boldsymbol{\eta}}^{\delta}) - \eta_1\psi_1(u_{\boldsymbol{\eta}}^{\delta}) \\ \phi(u_{\boldsymbol{\eta}}^{\delta}, g^{\delta}) - \dfrac{1}{2}\delta^2 + \eta_1\psi_1(u_{\boldsymbol{\eta}}^{\delta}) - \eta_2\psi_2(u_{\boldsymbol{\eta}}^{\delta}) \end{pmatrix} = 0.$$

The system is numerically more amenable than (2.4). In Algorithm 1, the Jacobian $\mathbf{J}_0$ can be approximated by finite difference. Step 7 represents the celebrated Broyden update. The stopping criterion is based on monitoring the residual norm $\|\mathbf{T}(\boldsymbol{\eta})\|$. Note that each iteration involves evaluating $\mathbf{T}(\boldsymbol{\eta})$, which in turn incurs solving one optimization problem of minimizing $J_{\boldsymbol{\eta}}$. Our experiences indicate that it converges fast and steadily, however, a convergence analysis is still missing.

Algorithm 1 Broyden's method for system (2.5)

1 Set $k = 0$ and choose $\boldsymbol{\eta}^0$.
2 Compute the Jacobian $\mathbf{J}_0 = \nabla\mathbf{T}(\boldsymbol{\eta}^0)$ and equation residual $\mathbf{T}(\boldsymbol{\eta}^0)$.
3 **for** $k = 1, \cdots, K$ **do**
4 Calculate the quasi-Newton update $\Delta\boldsymbol{\eta} = -\mathbf{J}_{k-1}^{-1}\mathbf{T}(\boldsymbol{\eta}^{k-1})$.
5 Update the regularization parameter $\boldsymbol{\eta}$ by $\boldsymbol{\eta}^k = \boldsymbol{\eta}^{k-1} + \Delta\boldsymbol{\eta}$.
6 Evaluate the equation residual $\mathbf{T}(\boldsymbol{\eta}^k)$ and set $\Delta\mathbf{T} = \mathbf{T}(\boldsymbol{\eta}^k) - \mathbf{T}(\boldsymbol{\eta}^{k-1})$.
7 Compute Jacobian update

$$\mathbf{J}_k = \mathbf{J}_{k-1} + \frac{1}{\|\Delta\boldsymbol{\eta}\|^2}[\Delta\mathbf{T} - \mathbf{J}_k\Delta\boldsymbol{\eta}] \cdot \Delta\boldsymbol{\eta}^{\mathrm{t}}.$$

8 Check the stopping criterion.
9 **end for**
10 Output the solution

3.2 Fixed point algorithm

In this part, we describe a fixed point algorithm for computing the minimizer of the rule Φ_γ. The algorithm was originally introduced in [7], but without any analysis. One basic version is listed in Algorithm 2, where the subscript $-i$ refers to the index different from i. The

stopping criterion at Step 4 can be based on monitoring the relative change of the regularization parameter $\boldsymbol{\eta}$ or the inverse solution $u_{\boldsymbol{\eta}}^{\delta}$.

Algorithm 2 Fixed point algorithm for minimizing (2.3)

1 Set $k = 0$ and choose $\boldsymbol{\eta}^0$.
2 Solve for u^{k+1} by the Tikhonov regularization
$$u^{k+1} = \arg\min_u \{\phi(u, g^\delta) + \boldsymbol{\eta}^k \cdot \boldsymbol{\psi}(u)\}.$$
3 Update the regularization parameter $\boldsymbol{\eta}^{k+1}$ by
$$\eta_i^{k+1} = \frac{1}{1+\gamma} \frac{\phi(u^{k+1}, g^\delta) + \eta_{-i}^k \psi_{-i}(u^{k+1})}{\psi_i(u^{k+1})}, \quad i = 1, 2.$$
4 Check the stopping criterion.

We shall analyze Algorithm 2. First, we introduce a fixed point operator $\mathbf{T}$ by

$$\mathbf{T}(\boldsymbol{\eta}) = (1+\gamma)^{-1} \begin{pmatrix} \dfrac{\phi(u_{\boldsymbol{\eta}}^\delta, g^\delta) + \eta_2 \psi_2(u_{\boldsymbol{\eta}}^\delta)}{\psi_1(u_{\boldsymbol{\eta}}^\delta)} \\ \dfrac{\phi(u_{\boldsymbol{\eta}}^\delta, g^\delta) + \eta_1 \psi_1(u_{\boldsymbol{\eta}}^\delta)}{\psi_2(u_{\boldsymbol{\eta}}^\delta)} \end{pmatrix}.$$

We shall also need the next result (see [8, Lemma 2.1 and Corrollary 2.3]).

Lemma 3.1 *The function $\psi_i(u_{\boldsymbol{\eta}}^\delta)$ is monotonically decreasing in η_i, and the following relations hold:*

$$\frac{\partial}{\partial \eta_i}(\phi(u_{\boldsymbol{\eta}}^\delta, g^\delta) + \eta_{-i}\psi_{-i}(u_{\boldsymbol{\eta}}^\delta)) + \eta_i \frac{\partial}{\partial \eta_i}\psi_i(u_{\boldsymbol{\eta}}^\delta) = 0, \quad i = 1, 2.$$

We have the next monotone result for the fixed point operator $\mathbf{T}$.

Proposition 3.1 *Let the function $F(\boldsymbol{\eta})$ be twice differentiable. Then the map $\mathbf{T}(\boldsymbol{\eta})$ is monotone if $F^2(\boldsymbol{\eta})F_{\eta_1\eta_1}(\boldsymbol{\eta})F_{\eta_2\eta_2}(\boldsymbol{\eta}) > (F_{\eta_1}(\boldsymbol{\eta})F_{\eta_2}(\boldsymbol{\eta}) - F(\boldsymbol{\eta})F_{\eta_1\eta_2}(\boldsymbol{\eta}))^2$.*

Proof Let $A(\boldsymbol{\eta}) = \phi + \eta_2\psi_2$ and $B(\boldsymbol{\eta}) = \phi + \eta_1\psi_1$. By Lemma 3.1, there hold

$$\frac{\partial A}{\partial \eta_1} + \eta_1 \frac{\partial \psi_1}{\partial \eta_1} = 0 \quad \text{and} \quad \frac{\partial B}{\partial \eta_2} + \eta_2 \frac{\partial \psi_2}{\partial \eta_2} = 0. \tag{3.1}$$

With the help of these two relations, we deduce

$$\begin{aligned} \frac{\partial}{\partial \eta_1}\frac{A}{\psi_1} &= \frac{\psi_1 \frac{\partial A}{\partial \eta_1} - A\frac{\partial \psi_1}{\partial \eta_1}}{\psi_1^2} = \frac{\psi_1(-\eta_1 \frac{\partial \psi_1}{\partial \eta_1}) - A\frac{\partial \psi_1}{\partial \eta_1}}{\psi_1^2} \\ &= -\frac{1}{\psi_1^2}\frac{\partial \psi_1}{\partial \eta_1}(\eta_1\psi_1 + A) = -\frac{F}{\psi_1^2}\frac{\partial \psi_1}{\partial \eta_1} \end{aligned}$$

and

$$\begin{aligned} \frac{\partial}{\partial \eta_2}\frac{A}{\psi_1} &= \frac{\psi_1 \frac{\partial A}{\partial \eta_2} - A\frac{\partial \psi_1}{\partial \eta_2}}{\psi_1^2} = \frac{\psi_1 \frac{\partial}{\partial \eta_2}(F - \eta_1\psi_1) - (F - \eta_1\psi_1)\frac{\partial \psi_1}{\partial \eta_2}}{\psi_1^2} \\ &= \frac{1}{\psi_1^2}\Big[\psi_1\psi_2 - F\frac{\partial \psi_1}{\partial \eta_2}\Big], \end{aligned}$$

where we have used the relation $\frac{\partial F}{\partial \eta_2} = \psi_2$ from Lemma 2.1. Similarly, we have

$$\frac{\partial}{\partial \eta_2}\frac{B}{\psi_2} = -\frac{F}{\psi_2^2}\frac{\partial \psi_2}{\partial \eta_2} \quad \text{and} \quad \frac{\partial}{\partial \eta_1}\frac{B}{\psi_2} = \frac{1}{\psi_2^2}\Big[\psi_1\psi_2 - F\frac{\partial \psi_2}{\partial \eta_1}\Big].$$

Therefore, the Jacobian $\nabla \mathbf{T}$ of the operator $\mathbf{T}$ is given by

$$\nabla \mathbf{T} = (1+\gamma)^{-1}\begin{pmatrix} -\frac{F}{\psi_1^2}\frac{\partial \psi_1}{\partial \eta_1} & \frac{1}{\psi_1^2}\Big[\psi_1\psi_2 - F\frac{\partial \psi_1}{\partial \eta_2}\Big] \\ \frac{1}{\psi_2^2}\Big[\psi_1\psi_2 - F\frac{\partial \psi_2}{\partial \eta_1}\Big] & -\frac{F}{\psi_2^2}\frac{\partial \psi_2}{\partial \eta_2} \end{pmatrix}.$$

Now Lemma 3.1 implies that

$$-\frac{\partial \psi_i}{\partial \eta_i} \geq 0.$$

Hence, it suffices to show that the determinant $|\nabla \mathbf{T}| > 0$. By Lemma 2.1, the identity

$$\frac{\partial \psi_1}{\partial \eta_2} = F_{\eta_1\eta_2} = \frac{\partial \psi_2}{\partial \eta_1}$$

holds, and thus $|\nabla \mathbf{T}|$ is given by

$$|\nabla \mathbf{T}| = (1+\gamma)^{-1}\frac{1}{\psi_1^2\psi_2^2}\Big[F^2\frac{\partial \psi_1}{\partial \eta_1}\frac{\partial \psi_2}{\partial \eta_2} - \Big(\psi_1\psi_2 - F\frac{\partial \psi_2}{\partial \eta_1}\Big)^2\Big].$$

Hence, the nonnegativity of $|\nabla \mathbf{T}|$ follows from the assumption

$$F^2(\boldsymbol{\eta})F_{\eta_1\eta_1}(\boldsymbol{\eta})F_{\eta_2\eta_2}(\boldsymbol{\eta}) - (F_{\eta_1}(\boldsymbol{\eta})F_{\eta_2}(\boldsymbol{\eta}) - F(\boldsymbol{\eta})F_{\eta_1\eta_2}(\boldsymbol{\eta}))^2 > 0.$$

This concludes the proof.

4 Numerical Experiments

We now provide some numerical results for the hybrid principle (2.5); and the balancing principle (2.3) has been numerically exemplified in [7] and will not be addressed here. The examples are integral equations of the first kind with kernel $k(s,t)$ and solution $u(t)$. All the examples are taken from [7]. The discretized linear system takes the form $\mathbf{K}\mathbf{u}^\dagger = \mathbf{g}^\dagger$. The data $\mathbf{g}^\dagger$ is then corrupted by noises, i.e., $g_i^\delta = g_i^\dagger + \max_i\{|g_i^\dagger|\}\varepsilon\zeta_i$, where ζ_i are standard Gaussian variables, and ε is the relative noise level.

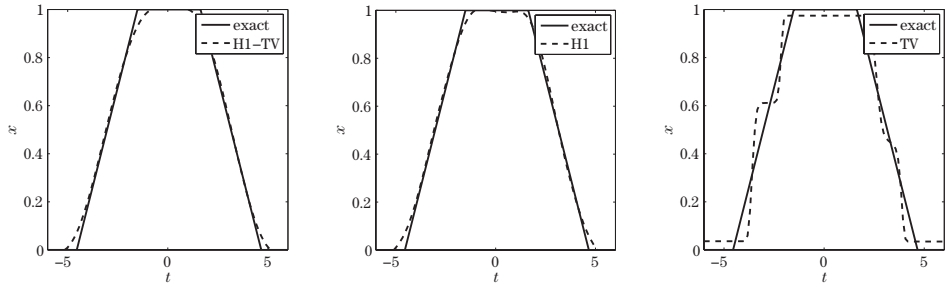
4.1 $\mathbf{H^1}$-TV model

Example 4.1 Let $\xi(t) = \chi_{|t|\leq 3}(1+\cos\frac{\pi t}{3})$, and the kernel $k(s,t)$ is given by $\xi(s-t)$. The true solution $u^\dagger$ exhibits both flat and smoothly varying regions and it is shown in Figure 1, and the integration interval is $[-6,6]$. We adopt two penalties $\psi_1(u) = |u|_{\mathrm{H}^1}^2$ and $\psi_2(u) = |u|_{\mathrm{TV}}$.

The numerical results are summarized in Table 1. In the table, the subscripts bdp and opt respectively refer to the hybrid principle and the optimal choice, i.e., the value giving the smallest error. The single-parameter models are indicated by subscripts h1 and tv, and the regularization parameter shown in Table 1 is the optimal one. The accuracy of the results is measured by the relative L^2 error $e = \frac{\|u-u^\dagger\|_{\mathrm{L}^2}}{\|u^\dagger\|_{\mathrm{L}^2}}$. We observe that the H^1-TV model in

Table 1 Numerical results for Example 4.1

ϵ	$\boldsymbol{\eta}_{\rm bdp}$	$\boldsymbol{\eta}_{\rm opt}$	$\eta_{\rm h1}$	$\eta_{\rm tv}$	$e_{\rm bdp}$	$e_{\rm opt}$	$e_{\rm h1}$	$e_{\rm tv}$
5e-2	(5.89e-3,9.67e-3)	(2.30e-4,2.05e-3)	6.17e-4	9.67e-3	3.50e-2	2.65e-2	3.96e-2	1.07e-1
5e-3	(3.41e-4,5.98e-4)	(2.34e-5,3.92e-4)	8.34e-5	4.51e-4	2.45e-2	1.09e-2	2.70e-2	9.49e-2
5e-4	(2.93e-6,5.41e-6)	(2.55e-6,4.48e-5)	1.26e-6	5.16e-5	1.22e-2	8.86e-3	1.38e-2	4.49e-2
5e-5	(1.19e-7,2.26e-7)	(5.88e-8,4.36e-6)	8.98e-8	3.79e-6	6.91e-3	5.53e-3	9.40e-3	1.68e-2
5e-6	(4.94e-9,9.50e-9)	(1.93e-10,6.22e-9)	5.18e-10	2.80e-7	4.64e-3	2.90e-3	5.29e-3	5.13e-3



$u_{\rm bdp}$ $u_{\rm h1}$ $u_{\rm tv}$

Figure 1 Numerical results for Example 4.1 with $\varepsilon = 5\%$ noise.

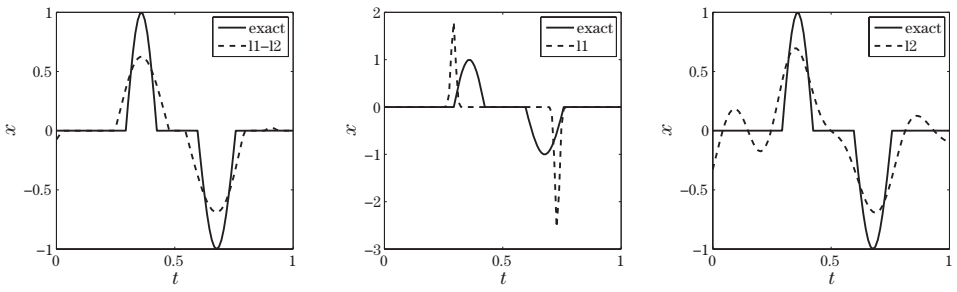
conjunction with the hybrid principle achieves a smaller error than either H^1 or TV with the optimal choice, thereby showing the advantages of the H^1-TV model. Further, the hybrid principle gives an error fairly close to the optimal one, within a factor of two, and the error decreases as the noise level decreases.

Let us briefly comment on the performance of the multi-parameter model. The classical H^1 model recovers the flat region unsatisfactorily, whereas the TV approach clearly suffers from staircasing effect in the gray region and reduced magnitude in the flat region, cf. Figure 1. In contrast, the H^1-TV model preserves the magnitude of the flat region while recovering the gray region excellently. Therefore, the H^1-TV model does combine the strengths of both H^1 and TV models. Finally, we would like to remark that Broyden's method converges rapidly with the convergence achieved in five iterations, and the convergence behavior is not sensitive to the initial guess.

4.2 Elastic-net model

Example 4.2 The kernel $k(s,t)$ is given by $\frac{1}{4}(\frac{1}{16} + (s-t)^2)^{-\frac{3}{2}}$, the exact solution $u^\dagger$ consists of two bumps and it is shown in Figure 2. The penalties are $\psi_1(u) = \|u\|_{\ell^1}$ and $\psi_2(u) = \frac{1}{2}\|u\|_{\ell^2}^2$ to retrieve the groupwise sparsity structure, which is known as elastic-net in statistics (see [19]). The integration interval is $[0,1]$. The size of the problem is 100.

It is observed from Table 2 that the hybrid principle gives slightly too small but otherwise reasonable estimate for the optimal choice. A close look at Figure 2 indicates that the solution $u_{\rm l2}$ has almost no zero entries, and thus it fails to distinguish between relevant and irrelevant factors. Meanwhile, many entries of the ℓ^1 solution are zero, and thus some relevant factors are correctly identified. However, it tends to select only a part instead of all relevant factors. The elastic-net combines the best of both ℓ^1 and ℓ^2 models, and it achieves the desired goal of



u_{bdp} u_{l1} u_{l2}

Figure 2 Numerical results for Example 4.2 with $\varepsilon = 5\%$ noise.

Table 2 Numerical results for Example 4.2

ϵ	$\boldsymbol{\eta}_{\mathrm{bdp}}$	$\boldsymbol{\eta}_{\mathrm{opt}}$	η_{l1}	η_{l2}	e_{bdp}	e_{opt}	e_{l1}	e_{l2}
5e-2	(2.44e-3,9.60e-3)	(2.81e-3,1.16e-3)	1.16e0	3.11e-3	4.09e-1	8.57e-2	1.29e0	4.58e-1
5e-3	(7.30e-5,2.25e-4)	(2.59e-4,1.11e-4)	9.67e-5	3.13e-5	1.96e-1	1.20e-2	9.00e-1	2.90e-1
5e-4	(4.73e-6,1.27e-5)	(2.23e-5,1.11e-5)	1.27e-5	4.13e-6	7.50e-2	8.18e-3	6.18e-1	2.17e-1
5e-5	(3.29e-7,8.42e-7)	(2.73e-6,1.28e-6)	1.12e-6	3.79e-8	2.01e-2	4.69e-3	4.85e-1	1.66e-1
5e-6	(2.56e-8,6.50e-8)	(1.60e-7,9.92e-8)	5.14e-9	1.25e-9	1.16e-2	2.27e-3	2.62e-1	9.55e-2

identifying the group structure.

4.3 Image deblurring

Example 4.3 The kernel $k(s,t)$ performs standard Gaussian blur with standard deviation 1 and blurring width 5. The exact solution $u^\dagger$ is shown in Figure 3. The size of the image is 50×50. The penalties are $\psi_1(u) = \|u\|_{\ell^1}$ and $\psi_2(u) = \frac{1}{2}\|u\|^2_{\ell^2}$.

This example represents a more realistic problem of image deblurring. Here one half of the data points are retained, which renders the problem far more ill-posed. The ℓ^1 solution is very spiky, cf. Figure 3, and neighboring pixels act independently of each other. In particular, many pixels in the blocks and the cross are missing. In contrast, the solution u_{l2} is smooth, but there are many small spurious oscillations in the background. The elastic-net model achieves the best of the two: Retaining the block structure with only few spurious nonzero coefficients. The numbers are also very telling: $e_{\mathrm{bdp}} = 2.96\text{e-}1$, $e_{\mathrm{o}} = 2.44\text{e-}1$, $e_{l1} = 9.21\text{e-}1$, and $e_{l2} = 3.42\text{e-}1$. Hence, the error e_{bdp} agrees well with the optimal choice, and it is smaller than that with the optimal choice for either ℓ^1 or ℓ^2 models.

5 Conclusions

We have studied multi-parameter regularization from the viewpoint of augmented Tikhonov regularization, and shown a unified way to derive the balancing principle and balanced discrepancy principle. A priori and a posteriori error estimates for the principles were provided, and efficient numerical algorithms (Broyden's method and fixed point algorithm) were presented and discussed. Numerical results were presented to illustrate the feasibility of the balanced discrepancy principle.

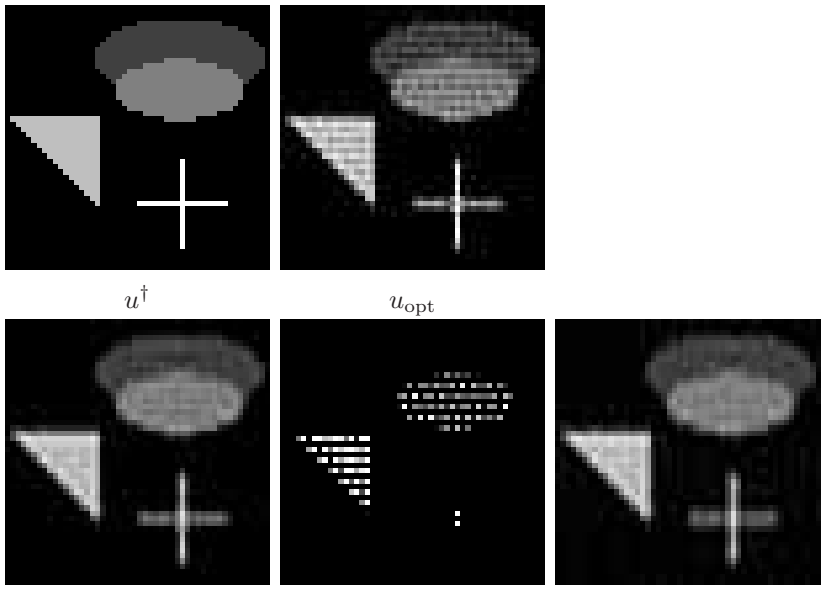


Figure 3 Numerical results for Example 4.3 with $\varepsilon = 1\%$ noise. The selected regularization parameters are $\eta_{\rm bdp} = (4.70\text{e-}3, 4.65\text{e-}3)$, $\eta_{\rm opt} = (1.26\text{e-}2, 1.31\text{e-}3)$, $\eta_{\rm l1} = 5.67\text{e-}1$, and $\eta_{\rm l2} = 3.51\text{e-}3$.

Acknowledgement This work was partially carried out during the visit of the first author at Institute for Applied Mathematics and Computational Science of Texas A&M University. He would like to thank the institute for the hospitality.

References

[1] Belge, M., Kilmer, M. E. and Miller, E. L., Efficient determination of multiple regularization parameters in a generalized L-curve framework, *Inverse Problems*, **18**(4), 2002, 1161–1183.

[2] Broyden, C. G., A class of methods for solving nonlinear simultaneous equations, *Math. Comp.*, **19**(92), 1965, 577–593.

[3] Burger, M. and Osher, S., Convergence rates of convex variational regularization, *Inverse Problems*, **20**(5), 2004, 1411–1420.

[4] Chen, Z., Lu, Y., Xu, Y. and Yang, H., Multi-parameter Tikhonov regularization for linear ill-posed operator equations, *J. Comput. Math.*, **26**(1), 2008, 37–55.

[5] Engl, H. W., Hanke, M. and Neubauer, A., Regularization of Inverse Problems, Kluwer, Dordrecht, 1996.

[6] Hofmann, B., Kaltenbacher, B., Poeschl, C. and Scherzer, O., A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators, *Inverse Problems*, **23**(3), 2007, 987–1010.

[7] Ito, K., Jin, B. and Takeuchi, T., Multi-parameter Tikhonov regularization, *Methods Appl. Anal.*, **18**(1), 2011, 31–46.

[8] Ito, K., Jin, B. and Takeuchi, T., A regularization parameter for nonsmooth Tikhonov regularization, *SIAM J. Sci. Comput.*, **33**(3), 2011, 1415–1438.

[9] Ito, K. and Kunisch, K., BV-type regularization methods for convoluted objects with edge, flat and grey scales, *Inverse Problems*, **16**(4), 2000, 909–928.

[10] Ivanov, V. K., Vasin, V. V. and Tanana, V. P., Theory of Linear Ill-Posed Problems and Its Applications, VSP, Utrecht, 2nd edition, 2002.

[11] Jin, B. and Lorenz, D. A., Heuristic parameter-choice rules for convex variational regularization based on error estimates, *SIAM J. Numer. Anal.*, **48**(3), 2010, 1208–1229.

[12] Jin, B. and Zou, J., Augmented Tikhonov regularization, *Inverse Problems*, **25**(2), 2009, 025001, 25 pages.

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

[14] Lu, S., Pereverzev, S. V., Shao, Y. and Tautenhahn, U., Discrepancy curves for multi-parameter regularization, *J. Inv. Ill-Posed Probl.*, **18**(6), 2010, 655–676.

[15] Lu, Y., Shen, L. and Xu, Y., Multi-parameter regularization methods for high-resolution image reconstruction with displacement errors, *IEEE Trans. Circuits Syst. I. Regul. Pap.*, **54**(8), 2007, 1788–1799.

[16] Mathé, P., The Lepskii principle revisited, *Inverse Problems*, **22**(3), 2006, L11–L15.

[17] Stephanakis, I. M., Regularized image restoration in multiresolution spaces, *Opt. Eng.*, **36**(6), 1997, 1738–1744.

[18] Xu, P., Fukuda, Y. and Liu, Y., Multiple parameter regularization: numerical solutions and applications to the determination of geopotential from precise satellite orbits, *J. Geod.*, **80**(1), 2006, 17–27.

[19] Zou, H. and Hastie, T., Regularization and variable selection via the elastic net, *J. R. Stat. Soc. Ser. B*, **67**(2), 2005, 301–320.