Applied Mathematics and Computation 219 (2012) 2100–2113

Contents lists available at SciVerse ScienceDirect

Applied Mathematics and Computation

journal homepage: www.elsevier.com/locate/amc

On a generalization of Regińska's parameter choice rule and its numerical realization in large-scale multi-parameter Tikhonov regularization

Fermín S. Viloche Bazán [a,*,1], Leonardo S. Borges [b,2], Juliano B. Francisco [a,3]

[a] Department of Mathematics, Federal University of Santa Catarina, 88040-900 Florianópolis, SC, Brazil
[b] Department of Applied Mathematics, IMECC-UNICAMP, University of Campinas, CP 6065, 13081-970 Campinas SP, Brazil

ARTICLE INFO

Keywords:
Parameter choice rules
Multi-parameter Tikhonov regularization
Large-scale discrete ill-posed problems

ABSTRACT

A crucial problem concerning Tikhonov regularization is the proper choice of the regularization parameter. This paper deals with a generalization of a parameter choice rule due to Regińska (1996) [31], analyzed and algorithmically realized through a fast fixed-point method in Bazán (2008) [3], which results in a fixed-point method for multi-parameter Tikhonov regularization called MFP. Like the single-parameter case, the algorithm does not require any information on the noise level. Further, combining projection over the Krylov subspace generated by the Golub–Kahan bidiagonalization (GKB) algorithm and the MFP method at each iteration, we derive a new algorithm for large-scale multi-parameter Tikhonov regularization problems. The performance of MFP when applied to well known discrete ill-posed problems is evaluated and compared with results obtained by the discrepancy principle. The results indicate that MFP is efficient and competitive. The efficiency of the new algorithm on a super-resolution problem is also illustrated.

1. Introduction

Linear least squares problems of the form

$$\min \|g - Af\|_2^2, \quad A \in \mathbb{R}^{m\times n}, \ m \geqslant n, \ g \in \mathbb{R}^m, \ f \in \mathbb{R}^n \tag{1.1}$$

with A large and ill-conditioned arise in a number of areas in science and engineering. They are commonly referred to as discrete ill-posed problems and arise, for example, when discretizing first kind integral equations with smooth kernel as in signal processing and image restoration, or when seeking to determine the internal structure of a system by external measurements, e.g., computerized tomography. In practical applications g represents data obtained experimentally and it is of the form $g = g^{\text{exact}} + e$, where e represents noise, g^{exact} denotes the unknown error-free data and $Af^{\text{exact}} = g^{\text{exact}}$. Note that due to the noise and the ill-conditioning of A, the naive least squares solution of (1.1), $f_{LS} = A^{\dagger}g$ (where $A^{\dagger}$ denotes the Moore–Penrose pseudoinverse of A) is dominated by noise and thus some form of regularization is needed in order to obtain a useful approximation to f^{exact}. The earliest and most known and well established regularization method is that due to Tikhonov [35] where f^{exact} is approximated by regularized solutions defined as

* Corresponding author.
E-mail addresses: fermin@mtm.ufsc.br (F.S. Viloche Bazán), lsbplsb@yahoo.com (L.S. Borges), juliano@mtm.ufsc.br (J.B. Francisco).
1 The work of this author is supported by CNPq, Brazil, Grant 308154/2008-8.
2 This research is supported by FAPESP, Brazil, Grant 2009/52193-1.
3 The work of this author is supported by CNPq, Brazil, Grant 479729/2011-5.

0096-3003/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.amc.2012.08.054

$$f_\lambda = \underset{f\in\mathbb{R}^n}{\operatorname{argmin}}\left\{\|g - Af\|_2^2 + \lambda^2\|Lf\|_2^2\right\}, \quad L \in \mathbb{R}^{p\times n}, \tag{1.2}$$

where L, referred to as the regularization matrix, is introduced to incorporate desirable properties on the solution such as smoothness, and $\lambda > 0$ is the regularization parameter. Common choices of L include the identity matrix, in which the problem is said to be in standard form, and discrete differential operators. The proper choice of the regularization parameter is a nontrivial problem for which several parameter choice methods exist. These include discrepancy principle (DP) [26], which requires a priori knowledge of the noise level, and a number of methods that do not require this information such as L-Curve criterion [17], Generalized Cross-Validation (GCV) [13] and Regińska's parameter choice rule [31]. For recent contributions which exploit the discrepancy principle the reader is referred to [30,38]. Apart from the above classical approaches, several works based on other techniques have been proposed including preconditioning, see, e.g., [37], as well as optimization tools [10,34]. Most of the above methods can be readily implemented using the generalized singular value decomposition (GSVD) of the matrix pair (A, L) when A and L are small or of moderate size. However, for large-scale problems the GSVD is computationally demanding and thus iterative or projection methods are preferable; these include [11,16,20,22,27,32], and a method called GKB-FP [5], which combines projection over the Krylov subspace generated by the Golub–Kahan bidiagonalization algorithm [14] and fixed-point iterations at each step.

Although Tikhonov regularization has been widely applied to solve ill-posed problems, it has been mostly confined to a single constraint. However, there are situations where the noise-free solution exhibits several distinct features and a natural question is how to incorporate them into the regularization approach. In this paper we are concerned with multi-parameter Tikhonov regularization problems where the minimization problem (1.1) is replaced by

$$f_\lambda = \underset{f\in\mathbb{R}^n}{\operatorname{argmin}}\left\{\|g - Af\|_2^2 + \sum_{i=1}^{q}\lambda_i^2\|L_i f\|_2^2\right\}, \quad L_i \in \mathbb{R}^{p_i\times n},\ \ i = 1,\ldots,q, \tag{1.3}$$

where $\lambda = [\lambda_1,\ldots,\lambda_q]^T$, $\lambda_i > 0$, is a vector of regularization parameters. Note that solving (1.3) amounts to solve the regularized normal equations

$$\left(A^TA + \sum_{i=1}^{q}\lambda_i^2 L_i^T L_i\right)f = A^Tg,$$

whose solution f_λ is unique when

$$\mathcal{N}(A)\cap\mathcal{N}(L_1)\cap\cdots\cap\mathcal{N}(L_q) = 0, \tag{1.4}$$

where $\mathcal{N}(A)$ denotes the null space of A. Condition (1.4) is met, e.g., when one regularization matrix is the identity or when A has full column rank; throughout the paper (1.4) is always assumed to be true. Applications of formulation (1.3) have appeared in a number of problems such as the determination of geopotentials from precise satellite orbits [36], high-resolution image reconstruction with displacement errors [25], image super-resolution [39], and estimation of parameters in jump diffusion processes [12]. Like the one-parameter case, the parameter choice rules for the multi-parameter case can be separated into two classes: rules that exploit a priori knowledge about the noise level and rules that do not exploit this information. As an example of parameter choice rule that exploit the knowledge of the norm $\|e\|$ we cite Lu et al. [23], Lu and Pereverzev [24], and the papers [1,2,9] where the choice of the parameters depends on the structure of the noise. The second class include a generalization of L-Curve method [6], a multivariate GCV [8], an approach due to Brezinski et al. [8] where the regularized solution is taken to be a constrained linear combination of one-parameter regularized solutions, and the minimum distance function approach of Belge et al. [7].

However, to the best of our knowledge, very little is known about efficient algorithms for solving large-scale multi-parameter Tikhonov regularization problems of the form (1.3). For a first attempt see [32] where an algorithm based on generalized Arnoldi iterations is proposed. The algorithm looks promising but its efficiency remains to be verified.

The main purpose of this paper is to present a generalization of Regińska's parameter choice rule to the multi-parameter case, which results in a fixed-point method for multi-parameter Tikhonov regularization called MFP, as well as to propose a GKB-FP type algorithm that is well suited for large-scale multi-parameter problems. Throughout the paper we assume that no estimate of $\|e\|_2$ is available. The rest of the paper is organized as follows. In Section 2 we review the original Regińska's parameter choice rule and provide theoretical results which support its generalization to the multi-parameter case. Our algorithm for large-scale Tikhonov regularization is described in Section 3. In Section 4 the efficiency of the proposed algorithm is illustrated by comparing our results with the results obtained by other methods. Conclusions are in Section 5.

2. Generalization of Regińska's parameter choice rule to multiple parameters and its algorithmic realization

The purpose here is to generalize the parameter choice rule due to Regińska [31] to multi-parameter Tikhonov regularization and to introduce a corresponding algorithmic realization.

2.1. Brief review of original Regińska's rule

Let $x(\lambda) = \|g - Af_\lambda\|_2^2$ and $y(\lambda) = \|Lf_\lambda\|_2^2$ where f_λ solves the one-parameter Tikhonov problem (1.2). Based on the fact that $y(\lambda)$ decreases with λ while $x(\lambda)$ increases, Regińska [31] proposed as Tikhonov regularization parameter a local minimum of the function

$$\Psi(\lambda) = x(\lambda)y(\lambda)^\mu, \quad \mu > 0. \tag{2.1}$$

Bazán [3] investigated the properties of Ψ and concluded that its minimizers are fixed-points of

$$\phi(\lambda;\mu) = \sqrt{\mu}\frac{\|g - Af_\lambda\|_2}{\|Lf_\lambda\|_2}, \quad \mu > 0, \tag{2.2}$$

which gave rise to the FP-algorithm. Practically, the FP-algorithm proceeds as follows

- Given an initial guess $\lambda^{(0)}$, set $\mu = 1$ and calculate the iterations

$$\lambda^{(k+1)} = \phi(\lambda^{(k)};1), \quad k \geqslant 0, \tag{2.3}$$

until the largest convex fixed-point of $\phi(\lambda, 1)$ is captured.
- If $\mu = 1$ does not work, μ is adjusted as explained in [3,4] and the iterations restart.

The success of this choice (hence of the FP-algorithm) is supported by the observation that the minimizer of Ψ corresponds to a good balance between the size of the solution norm and the size of the residual norm, in which case the error in f_λ with respect to f^{exact} tends to be minimized. This justifies the excellent performance of the FP-algorithm when compared to other well respected methods, as reported in [3,4,21]. We recall from [4] that a fixed-point of ϕ is said to be *convex* when the associated L-Curve is locally convex at that point. Typical behavior of curves $\Psi(\lambda)$ and $\phi(\lambda)$ can be seen in Fig. 1 where small circles are used to highlight the location of the minimizer of $\Psi(\lambda)$ and the corresponding fixed-point of $\phi(\lambda, 1)$.

2.2. Parameter choice rule for multi-parameter case

The previous single-parameter choice rule can be extended to the multi-parameter case by selecting as regularization parameter a local minimum of the function

$$\Psi(\lambda) = x(\lambda)y_1(\lambda)^{\mu_1}y_2(\lambda)^{\mu_2}\cdots y_q(\lambda)^{\mu_q}, \quad \mu_i > 0, \tag{2.4}$$

where f_λ solves the multi-parameter Tikhonov problem (1.3) and

$$x(\lambda) = \|g - Af_\lambda\|_2^2, \quad y_i(\lambda) = \|L_if_\lambda\|_2^2, \; i = 1,\ldots,q. \tag{2.5}$$

We shall now discuss conditions for a point $\lambda = [\lambda_1,\ldots,\lambda_q]^T$ to be a local minimizer of function Ψ. Note that the gradient $\nabla\Psi(\lambda)$ can be demonstrated to be

$$\nabla\Psi = y_1^{\mu_1}\cdots y_q^{\mu_q}(J\Omega + \nabla x), \tag{2.6}$$

where

$$J = \begin{bmatrix} \frac{\partial y_1}{\partial\lambda_1} & \cdots & \frac{\partial y_q}{\partial\lambda_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_1}{\partial\lambda_q} & \cdots & \frac{\partial y_q}{\partial\lambda_q} \end{bmatrix}, \quad \text{and} \quad \Omega = \begin{pmatrix} \mu_1\frac{x}{y_1} \\ \vdots \\ \mu_q\frac{x}{y_q} \end{pmatrix}.$$

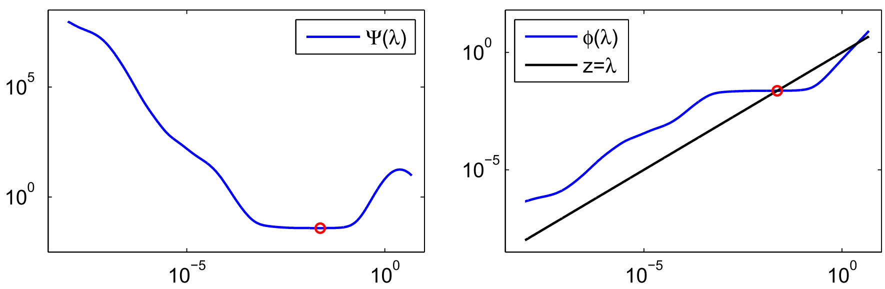

Fig. 1. Functions $\Psi(\lambda)$ and $\phi(\lambda, 1)$ for i_laplace test problem from [18], $n = 256$, and data with relative noise level 1%. For this test problem the regularization matrix is the identity, the selected parameter is $\lambda^* = 0.0233$ and the relative error in f_{λ^*} is 18.44% (the optimal one is $\lambda_{\text{optimal}} = 0.0105$ and the relative error in the associated solution is 18.27%).

Since each partial derivative of $y_i(\lambda)$ with respect to λ_j depends on λ, sometimes we will denote $J(\lambda)$ to highlight the dependence of J on λ; the same observation applies for Ω. For future reference, note that, as

$$x(\lambda) = \|g - Af_\lambda\|_2^2 = g^T g - 2g^T Af_\lambda + f_\lambda^T A^T Af_\lambda, \quad \text{and} \quad y_i(\lambda) = \|L_i f_\lambda\|_2^2 = f_\lambda^T L_i^T L_i f_\lambda,$$

then

$$\frac{\partial x}{\partial \lambda_j} = 2\left(\frac{\partial f_\lambda}{\partial \lambda_j}\right)^T A^T (Af_\lambda - g), \quad \frac{\partial y_i}{\partial \lambda_j} = 2\left(\frac{\partial f_\lambda}{\partial \lambda_j}\right)^T L_i^T L_i f_\lambda. \tag{2.7}$$

The following technical results will be needed.

Lemma 2.1. *Under assumption (1.4), the following properties hold*

(i) $$\frac{\partial x}{\partial \lambda_j} = -\lambda_1^2 \frac{\partial y_1}{\partial \lambda_j} - \cdots - \lambda_q^2 \frac{\partial y_q}{\partial \lambda_j}. \tag{2.8}$$

(ii) *the vectors* $\frac{\partial f_\lambda}{\partial \lambda_j}$, $j = 1, \ldots, q$ *are linearly independent if, and only if, the vectors* $L_j^T L_j f_\lambda$, $j = 1, \ldots, q$ *are linearly independent.*

(iii) J *is nonsingular provided that* $\frac{\partial f_\lambda}{\partial \lambda_j}$, $j = 1, \ldots, q$ *are linearly independent.*

Proof. Assumption (1.4) implies that (1.3) has a unique solution f_λ such that

$$(A^T A + \lambda_1^2 L_1^T L_1 + \cdots + \lambda_q^2 L_q^T L_q) f_\lambda = A^T g. \tag{2.9}$$

This implies $A^T(Af_\lambda - g) = -\lambda_1^2 L_1^T L_1 f_\lambda - \cdots - \lambda_q^2 L_q^T L_q f_\lambda$. Now it suffices to multiply this equation by $\left(\frac{\partial f_\lambda}{\partial \lambda_j}\right)^T$, $j = 1, \ldots, q$, and then use (2.7); this proves (i). On the other hand, differentiation of (2.9) with respect to λ_j leads to

$$2\lambda_j L_j^T L_j f_\lambda + (A^T A + \lambda_1^2 L_1^T L_1 + \cdots + \lambda_q^2 L_q^T L_q)\left(\frac{\partial f_\lambda}{\partial \lambda_j}\right) = 0. \tag{2.10}$$

Let $B = A^T A + \lambda_1^2 L_1^T L_1 + \cdots + \lambda_q^2 L_q^T L_q$. Then (2.10) can be rewritten as

$$B\left[\frac{\partial f_\lambda}{\partial \lambda_1} \ \frac{\partial f_\lambda}{\partial \lambda_2} \ \cdots \ \frac{\partial f_\lambda}{\partial \lambda_q}\right] = -[2\lambda_1 L_1^T L_1 f_\lambda \ \ 2\lambda_2 L_2^T L_2 f_\lambda \ \ \cdots \ \ 2\lambda_q L_q^T L_q f_\lambda].$$

This proves (ii) since by assumption (1.4) B is definite positive. On the other hand, an immediate consequence of (2.10) is

$$\left(\frac{\partial f_\lambda}{\partial \lambda_i}\right)^T (A^T A + \lambda_1^2 L_1^T L_1 + \cdots + \lambda_q^2 L_q^T L_q)\left(\frac{\partial f_\lambda}{\partial \lambda_j}\right) = -2\lambda_j \left(\frac{\partial f_\lambda}{\partial \lambda_i}\right)^T L_j^T L_j f_\lambda.$$

Using (2.7) we have

$$\left(\frac{\partial f_\lambda}{\partial \lambda_i}\right)^T (A^T A + \lambda_1^2 L_1^T L_1 + \cdots + \lambda_q^2 L_q^T L_q)\left(\frac{\partial f_\lambda}{\partial \lambda_j}\right) = -\lambda_j \frac{\partial y_j}{\partial \lambda_i}, \tag{2.11}$$

which in matrix form is $F^T(A^T A + \lambda_1^2 L_1^T L_1 + \cdots + \lambda_q^2 L_q^T L_q)F = -J\,\mathrm{diag}(\lambda_1, \ldots, \lambda_q)$, where $F = \left[\frac{\partial f_\lambda}{\partial \lambda_1} \ \frac{\partial f_\lambda}{\partial \lambda_2} \ \cdots \ \frac{\partial f_\lambda}{\partial \lambda_q}\right]$; this proves (iii) as F has full column rank. $\square$

In the single-parameter case $\|Lf_\lambda\|_2$ is a decreasing function of λ. For the multi-parameter case we can use (2.11) to prove the following similar result.

Corollary 2.2. *The partial derivative of* $y_i(\lambda)$ *with respect to* λ_i *is always negative.*

We can now substitute (2.8) in (2.6) to conclude that the gradient of Ψ can be written as

$$\nabla\Psi(\lambda) = y_1^{\mu_1} \cdots y_q^{\mu_q} J\left(\Omega - \begin{pmatrix} \lambda_1^2 \\ \vdots \\ \lambda_q^2 \end{pmatrix}\right).$$

Therefore, the necessary condition for the function $\Psi(\lambda)$ be minimized at $\lambda^* = [\lambda_1^*, \ldots, \lambda_q^*]^T \neq 0$, $\nabla\Psi(\lambda^*) = 0$, requires that $J(\lambda^*)$ be not singular and

$$\Omega(\lambda^*) = \begin{pmatrix} {\lambda_1^*}^2 \\ \vdots \\ {\lambda_q^*}^2 \end{pmatrix} \iff {\lambda_i^*}^2 = \mu_i \frac{x(\lambda^*)}{y_i(\lambda^*)}, \quad i = 1, \ldots, q. \tag{2.12}$$

Therefore, if $\Psi(\lambda)$ reaches a maximum/minimum at λ^*, this λ^* must be a fixed-point of the vector-valued function $\Phi : \mathbb{R}^q \to \mathbb{R}^q$ defined by

$$\Phi(\lambda) = \begin{pmatrix} \phi_1(\lambda, \mu_1) \\ \vdots \\ \phi_q(\lambda, \mu_q) \end{pmatrix}, \quad \phi_i(\lambda; \mu_i) = \sqrt{\mu_i}\frac{\|g - Af_\lambda\|_2}{\|L_i f_\lambda\|_2}, \ \ i = 1, \ldots, q. \tag{2.13}$$

The question about existence of fixed-points will be addressed geometrically at the end of the section. We now give a theorem stating conditions for minimizing Ψ.

Theorem 2.3. *A sufficient condition for a fixed-point λ^* of $\Phi(\lambda)$ be a local minimizer of $\Psi(\lambda)$ is that $2J(\lambda^*)\mathrm{diag}(\lambda^*) + J(\lambda^*)H(\lambda^*)J(\lambda^*)^T$ be negative definite, where*

$$H(\lambda) = \begin{bmatrix} \frac{\lambda_1^2}{y_1}(1+\mu_1) & \lambda_2^2 \frac{\mu_1}{y_1} & \cdots & \lambda_q^2 \frac{\mu_1}{y_1} \\ \lambda_1^2 \frac{\mu_2}{y_2} & \frac{\lambda_2^2}{y_2}(1+\mu_2) & \cdots & \lambda_q^2 \frac{\mu_2}{y_2} \\ \vdots & \ddots & \cdots & \vdots \\ \lambda_1^2 \frac{\mu_q}{y_q} & \lambda_2^2 \frac{\mu_q}{y_q} & \cdots & \frac{\lambda_q^2}{y_q}(1+\mu_q) \end{bmatrix}.$$

Proof. If λ^* is a fixed-point of $\Phi(\lambda)$, the gradient and the Hessian of $\Psi(\lambda)$ at λ^* can be shown to be

$$\nabla\Psi(\lambda^*) = y_1^{\mu_1}(\lambda^*) \cdots y_q^{\mu_q}(\lambda^*) J(\lambda^*) \begin{pmatrix} -\lambda_1^{*^2} + \phi_1(\lambda^*)^2 \\ \vdots \\ -\lambda_q^{*^2} + \phi_q(\lambda^*)^2 \end{pmatrix} = 0,$$

$$\nabla^2\Psi(\lambda^*) = -2y_1^{\mu_1}(\lambda^*) \cdots y_q^{\mu_q}(\lambda^*) J(\lambda^*)\mathrm{diag}(\lambda_1^*, \ldots, \lambda_q^*)(I_q - J_\phi(\lambda^*)),$$

where $J_\phi(\lambda)$ denotes the Jacobian of the function $\Phi(\lambda)$. Using Eq. (2.8) at λ^* and the fact that $y_i(\lambda^*)\lambda_i^{*2} = \mu_i x(\lambda^*)$ the partial derivative of $\phi_i(\lambda)$ with respect to λ_j can be written as

$$\frac{\partial\phi_i}{\partial\lambda_j} = \frac{1}{2y_i}\left(-\lambda_1^2 \frac{\mu_i}{\lambda_i}\frac{\partial y_1}{\partial\lambda_j} - \cdots - \lambda_i(1+\mu_i)\frac{\partial y_i}{\partial\lambda_j} - \cdots - \lambda_q^2 \frac{\mu_i}{\lambda_i}\frac{\partial y_q}{\partial\lambda_j}\right).$$

Hence, $J_\phi(\lambda^*)$ takes the form

$$J_\phi(\lambda^*) = -\frac{1}{2}\begin{bmatrix} \frac{\lambda_1}{y_1}(1+\mu_1) & \lambda_2^2 \frac{\mu_1}{y_1\lambda_1} & \cdots & \lambda_q^2 \frac{\mu_1}{y_1\lambda_1} \\ \lambda_1^2 \frac{\mu_2}{y_2\lambda_2} & \frac{\lambda_2}{y_2}(1+\mu_2) & \cdots & \lambda_q^2 \frac{\mu_2}{y_2\lambda_2} \\ \vdots & \ddots & \cdots & \vdots \\ \lambda_1^2 \frac{\mu_q}{y_q\lambda_q} & \lambda_2^2 \frac{\mu_q}{y_q\lambda_q} & \cdots & \frac{\lambda_q}{y_q}(1+\mu_q) \end{bmatrix} \begin{bmatrix} \frac{\partial y_1}{\partial\lambda_1} & \cdots & \frac{\partial y_1}{\partial\lambda_q} \\ \frac{\partial y_2}{\partial\lambda_1} & \cdots & \frac{\partial y_2}{\partial\lambda_q} \\ \vdots & \cdots & \vdots \\ \frac{\partial y_q}{\partial\lambda_1} & \cdots & \frac{\partial y_q}{\partial\lambda_q} \end{bmatrix}.$$

Thus, the expression for the Hessian of $\Psi(\lambda)$ at λ^* can be rewritten as

$$\nabla^2\Psi(\lambda^*) = -y_1^{\mu_1}(\lambda^*) \cdots y_q^{\mu_q}(\lambda^*)\left[2J(\lambda^*)\mathrm{diag}(\lambda_1^*, \ldots, \lambda_q^*) + J(\lambda^*)H(\lambda^*)J(\lambda^*)^T\right].$$

By hypothesis $\left[2J(\lambda^*)\mathrm{diag}(\lambda^*) + J(\lambda^*)H(\lambda^*)J(\lambda^*)^T\right]$ is negative definite, so $\nabla^2\Psi(\lambda^*)$ is positive definite and λ^* is a local minimizer of Ψ. $\square$

From [3, Lemma 1], where the original rule of Regińska was investigated, we learned that for $\mu = 1$ and λ larger than the largest generalized singular value of the matrix pair (A, L), it holds $\phi(\lambda, 1) \geqslant \lambda$. This is an useful result that allows us to locate fixed-points when they exist. A similar result can be provided for the multi-parameter case under certain conditions.

Theorem 2.4. *Assume that one regularization matrix, say L_1, has full column rank. Let $\bar{\gamma}_1$ be the largest generalized singular value of the matrix pair (A, L_1). Assume also that $f_\lambda \notin \mathcal{N}(L_i)$, $i = 2, \ldots, q$. Then for all $\lambda = (\lambda_1, \ldots, \lambda_q)$, with $\lambda_1 > \bar{\gamma}_1$ and $\lambda_i > 0$, it holds that*

$$\phi_i(\lambda; 1) > \lambda_i, \quad i = 1, \ldots, q. \tag{2.14}$$

Proof. Consider the auxiliary Tikhonov problem

$$\bar{f}_\xi = \underset{\bar{f}\in\mathbb{R}^n}{\mathrm{argmin}}\{\|g - A\bar{f}\|_2^2 + \xi^2\|L_i\bar{f}\|_2^2\}, \quad \xi > 0, \tag{2.15}$$

where

$$L_\lambda = \begin{bmatrix} \lambda_1 L_1 \\ \vdots \\ \lambda_q L_q \end{bmatrix}, \quad \lambda_i > 0, \;\; i = 1, \ldots, q.$$

It turns out that $\bar{f}_\xi$ is unique and $\bar{f}_1 = f_\lambda$. Let $\varphi(\xi) = \|g - A\bar{f}_\xi\|_2 / \|L_\lambda \bar{f}_\xi\|_2$, $\xi > 0$. Then for all $\lambda_i > 0$ we have

$$\frac{\phi_i(\lambda;1)^2}{\lambda_i^2} = \frac{\|g - Af_\lambda\|_2^2}{\lambda_i^2 \|L_i f_\lambda\|_2^2} > \frac{\|g - Af_\lambda\|_2^2}{\lambda_1^2 \|L_1 f_\lambda\|_2^2 + \cdots + \lambda_q^2 \|L_q f_\lambda\|_2^2} = \varphi^2(1), \tag{2.16}$$

so it suffices to prove that $\varphi(1) \geqslant 1$. Let $\bar{\gamma}_\lambda$ and $\bar{x}$ be the largest generalized singular value and corresponding eigenvector of the matrix pair (A, L_λ), respectively. Since due to [3, Lemma 1], it holds that $\varphi(\xi) \geqslant \xi$ whenever $\xi \geqslant \bar{\gamma}_\lambda$, it suffices to prove that $\bar{\gamma}_\lambda \leqslant 1$. To this end note that the largest generalized singular value of the matrix pair (A, L_1) can be characterized as

$$\bar{\gamma}_1^2 = \max_{x \neq 0} \frac{x^T A^T A x}{x^T L_1^T L_1 x}. \tag{2.17}$$

Then

$$\frac{\bar{\gamma}_1^2}{\lambda_1^2} = \frac{\bar{x}^T A^T A \bar{x}}{\bar{x}^T (\lambda_1 L_1)^T (\lambda_1 L_1) \bar{x}} \geqslant \frac{x^T A^T A x}{x^T (\lambda_1 L_1)^T (\lambda_1 L_1) x} \quad \forall x \neq 0 \geqslant \frac{x^T A^T A x}{x^T (\lambda_1 L_1)^T (\lambda_1 L_1) x + \cdots + x^T (\lambda_q L_q)^T (\lambda_q L_1) x} \quad \forall x \neq 0.$$

But this implies that

$$\frac{\bar{\gamma}_1^2}{\lambda_1^2} \geqslant \max_{x \neq 0} \frac{x^T A^T A x}{x^T (\lambda_1 L_1)^T (\lambda_1 L_1) x + \cdots + x^T (\lambda_q L_q)^T (\lambda_q L_1) x} = \bar{\gamma}_\lambda^2.$$

Therefore, provided that $\lambda_1 > \bar{\gamma}_1$, the largest generalized singular value of the matrix pair (A, L_λ) satisfies $\bar{\gamma}_\lambda < 1$, and the theorem is proved. □

We stress that though in many problems the regularization matrices are rank deficient, full rank regularization matrices are also used in several areas, see, e.g., [33]. A practical consequence of Theorem 2.4 is that if all regularization matrices are of full column rank, then there exists a box $B = [0, \bar{\gamma}_1] \times [0, \bar{\gamma}_2] \times \cdots [0, \bar{\gamma}_q]$, where $\bar{\gamma}_i$ denotes the largest generalized singular value of the matrix pair (A, L_i), such that

$$\phi_i(\lambda, 1) > \lambda_i \quad \text{with } \lambda_i > \bar{\gamma}_i, \;\; \lambda_j > 0, \;\; j \neq i, \;\; i = 1, \ldots, q. \tag{2.18}$$

This result generalizes Lemma 1 in [3]. Note that, while (2.14) inform us that fixed-points of Φ cannot fall outside the region $[0, \bar{\gamma}_1] \times Z$, where $Z = \{(\lambda_2, \ldots, \lambda_q) \in \mathbb{R}^{q-1} \;/\; \lambda_i > 0, \;\; i = 2, \ldots, q\}$, (2.18) asserts that fixed-points of Φ cannot fall outside the box $[0, \bar{\gamma}_1] \times [0, \bar{\gamma}_2] \times \cdots \times [0, \bar{\gamma}_q]$. Both of these conclusions are important since they help us to detect divergence of our fixed-point algorithm to be presented in the next section.

We now give a brief discussion about existence of fixed-points. For the sake of simplicity we shall discuss the two-parameter case which has a simple and nice geometric interpretation. This can be explained as follows. Consider the planes Π_i and surfaces $\mathcal{S}_i$ in $\mathbb{R}^3$, defined respectively by:

$$\Pi_i = \{(\lambda_1, \lambda_2, z)/z = \lambda_i\}, \quad \mathcal{S}_i = \{(\lambda_1, \lambda_2, z)/z = \phi_i(\lambda, 1)\}, \qquad i = 1, 2.$$

Let $C_i = \Pi_i \cap \mathcal{S}_i$. A typical surface $z = \Psi(\lambda)$ as well as the projections of C_i onto the plane $z = 0$ are displayed in Fig. 2; surfaces S_i and planes Π_i are shown in Fig. 3.

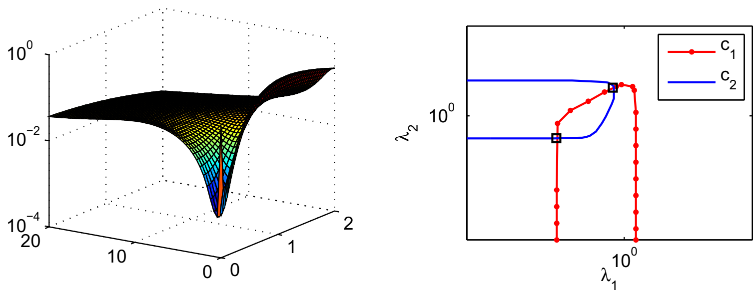

Fig. 2. Surface $z = \Psi(\lambda)$ (left) and projections of C_i onto the plane $z = 0$ (right) for i_laplace test problem from [18], $n = 256$, data with 1% noise, $L_1 = I$ and L_2 being a discrete first order differential operator.

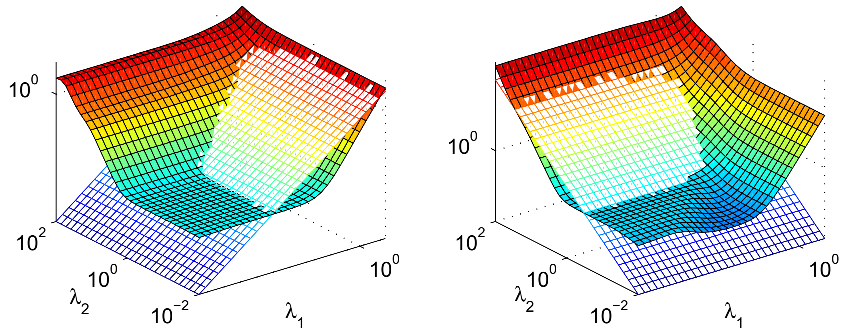

Fig. 3. Surfaces S_i and planes Π_i for i_laplace test problem.

Then, based on (2.12), it is apparent that Φ will have fixed-points, provided that $C_1 \cap C_2$ is not empty. If this is the case, every point in the intersection is a fixed-point of Φ. Note that the intersection $C_1 \cap C_2$ in Fig. 2 provides two fixed-points of Φ.

2.3. FP-algorithm for multi-parameter Tikhonov regularization

Our algorithm for the multi-parameter case, denoted hereafter by MFP, follows exactly the same steps as the fixed-point algorithm for the single-parameter case; it can be roughly described as follows.

- Given an initial guess $\lambda^{(0)} = [\lambda_1^{(0)}, \ldots, \lambda_q^{(0)}]^T$, set $\mu_i = 1$, $i = 1 : q$, and calculate the sequences

$$\lambda_i^{(k+1)} = \phi_i(\lambda^{(k)}; 1), \quad i = 1, \ldots, q \text{ and } k \geqslant 0 \tag{2.19}$$

  until a stopping criterion is satisfied
- If $\lambda_i^{(k)}$ diverges for some i, μ_i is adjusted and the iterations restart. Adjustment of μ_i is done similarly as in the single-parameter case; see [3] for details.

The choice of the initial guess is always a crucial point for iterative methods and there are probably many alternatives for the initial guess of MFP. Two of these are as follows: (a) choose as initial guess a set of small parameters, say $\lambda_i = 10^{-4}$, and then proceed as described above, and (b) apply the single-parameter fixed-point algorithm [3,4] to q single-parameter Tikhonov subproblems, one for each L_i, and take the found fixed-points as initial guess. The main advantage of (b) is that the FP-algorithm for the single-parameter case provides a parameter $\mu_i \neq 1$ when adjustment is required. Our numerical experiments are carried out following the second option, where we choose to stop the iterations when the relative change of consecutive iterates is small, i.e., when

$$\|\lambda^{(k+1)} - \lambda^{(k)}\|_2 < \varepsilon_1 \|\lambda^{(k)}\|_2,$$

where ε_1 is a small tolerance parameter or, to prevent slow convergence, when

$$\|\lambda^{(k+1)} - \lambda^{(k)}\|_2 < \varepsilon_2 \|\lambda^{(1)}\|_2,$$

where ε_2 is another small tolerance parameter. Finally, since the computation of the sequence (2.19) requires the function Φ to be evaluated repeatedly for several values of λ, the regularized solution f_λ must be calculated efficiently. This can be done by first noting that problem (1.3) can be rewritten as

$$f_\lambda = \underset{f \in \mathbb{R}^n}{\operatorname{argmin}} \|\bar{g} - \overline{A}_\lambda f\|_2^2, \tag{2.20}$$

where $\overline{A}_\lambda = [A^T \; \lambda_1 L_1^T \; \cdots \; \lambda_q L_q^T]^T$, and $\bar{g} = [g^T \; 0^T]^T$. Proceeding this way, problem (2.20) can be solved efficiently by using the QR decomposition or the SVD of $\overline{A}_\lambda$.

We end this section with a brief discussion on the convergence properties of the iterates (2.19), concentrating, in particular, on the two-parameter case; the general case can be handled similarly. Obviously, being (2.19) a sequence generated by fixed-point iteration, no more than local convergence results can be obtained. The key idea of the convergence analysis is to recognize that the dynamics of the iterates of the multi-parameter case is essentially the same as the dynamics of the iterates of the one-parameter case. We thus start by considering the one-parameter iterates defined in (2.3), $\lambda^{(k+1)} = \phi(\lambda^{(k)}, 1)$, $k \geqslant 0$. As usual, we shall assume that the residual norm $\|g - Af_\lambda\|_2$ does not vanish at $\lambda = 0$, and function $\phi(\lambda^{(k)}, 1)$ has a unique fixed-point that minimizes $\Psi(\lambda)$ and a unique fixed-point that maximizes $\Psi(\lambda)$. Let $\bar{\lambda}$ and $\tilde{\lambda}$ denote the minimizer of $\Psi(\lambda)$ and the maximizer of $\Psi(\lambda)$, respectively. Such a situation is illustrated in Fig. 4 (left) for i_laplace test problem. Then, based on the property that $\phi(\lambda, 1)$ is always an increasing function, the analysis in [3] (see Theorem 2) leads us to the following conclusions:

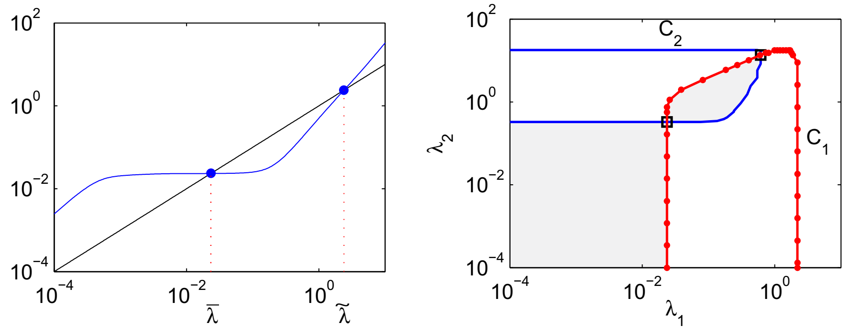

Fig. 4. Left: function $\phi(\lambda, 1)$ for i_laplace test problem with the same data as in Fig. 1. Right: convergence regions of iterates (2.19) (shaded area) for the two-parameter Tikhonov problem. In this case, the test problem is i_laplace with $n = 256$, the right hand side has 1% noise, $L_1 = I$, and L_2 is a discrete first order differential operator. Curves C_1 and C_2 are the same as in Fig. 2.

- the convergence region of the iterates $\lambda^{(k)}$ is the open interval $]0, \widetilde{\lambda}[$.
- $\lambda^{(k)}$ converges to $\bar{\lambda}$ as far as $\lambda^{(0)}$ belongs to the convergence region. If this is the case, either $\lambda^{(k)}$ is a decreasing sequence if $\lambda^{(0)} \in]\bar{\lambda}, \widetilde{\lambda}[$ (which means $\phi(\lambda^{(0)}, 1) < \lambda^{(0)}$) or $\lambda^{(k)}$ is an increasing sequence if $\lambda^{(0)} \in]0, \bar{\lambda}[$ (which means $\phi(\lambda^{(0)}, 1) > \lambda^{(0)}$).

We now turn to the two-parameter case. Let $\mathcal{R}_1$ be the region between the curves C_1 and C_2 and let $\mathcal{R}_2$ be the region bounded by the curves C_1 and C_2 and by the lines $\lambda_1 = 0$, $\lambda_2 = 0$, see Fig. 4. Then it turns out that the sequence $\lambda_i^{(k)}$, $i = 1, 2$, will converge to $\bar{\lambda}_i$ as far as the initial guess $\lambda_i^{(0)}$, $i = 1, 2$ lies inside $\mathcal{R}_1 \cup \mathcal{R}_2$, which means either $\phi_i(\lambda_i^{(0)}) < \lambda_i^{(0)}$, $i = 1, 2$, or $\phi_i(\lambda_i^{(0)}) > \lambda_i^{(0)}$, $i = 1, 2$. In either case, similar to the one-parameter case, convergence is assured because $\phi_i(\lambda_1, \lambda_2)$ is an increasing function of λ_i if the other parameter is kept fixed, as follows from Lemma 2.1 and Corollary 2.2.

3. Extending GKB-FP to large-scale multi-parameter Tikhonov regularization

We have seen that MFP requires the problem (2.20) to be solved repeatedly in order to calculate solution and residual norms for distinct values of the parameters $\lambda_1, \ldots, \lambda_q$. The approach is simple and can be implemented efficiently via QR decomposition or SVD of $\overline{A}_\lambda$ for small to medium sized problems. However, this does not work for large problems and therefore alternative approaches are needed. In this section such an approach is proposed. Specifically, the purpose of this section is to extend the GKB-FP algorithm to large-scale multi-parameter Tikhonov regularization. GKB-FP combines the Golub–Kahan bidiagonalization (GKB) algorithm with Tikhonov regularization in the generated Krylov subspace, with the regularization parameter for the projected problem chosen by the one-parameter FP method [3] at each iteration.

Recall that after $k < n$ steps, the GKB algorithm applied to A with initial vector $g/\|g\|_2$ yields two matrices $U_{k+1} = [u_1, \ldots, u_{k+1}] \in \mathbb{R}^{m\times(k+1)}$ and $V_k = [v_1, \ldots, v_k] \in \mathbb{R}^{n\times k}$ with orthonormal columns, and a lower bidiagonal matrix $B_k \in \mathbb{R}^{(k+1)\times k}$, such that

$$\beta_1 U_{k+1} e_1 = g = \beta_1 u_1, \tag{3.1}$$
$$A V_k = U_{k+1} B_k, \tag{3.2}$$
$$A^T U_{k+1} = V_k B_k^T + \alpha_{k+1} v_{k+1} e_{k+1}^T, \tag{3.3}$$

where e_i denotes the i-th unit vector in $\mathbb{R}^{k+1}$. The columns of V_k provide an orthonormal basis for the generated Krylov subspace $\mathcal{K}_k(A^T A, A^T g)$, which is an excellent choice for use when solving discrete ill-posed problems [19].

The main idea behind our extension of GKB-FP for large-scale multi-parameter Tikhonov problems is to produce a finite sequence of approximate solutions $f_\lambda^{(k)}$ obtained by minimizing the Tikhonov functional (1.3) over the subspace $\mathcal{K}_k(A^T A, A^T g)$. Therefore the approximate solution $f_\lambda^{(k)}$ is determined as

$$f_\lambda^{(k)} = \underset{f\in\mathcal{K}_k}{\operatorname{argmin}} \left\{ \|g - Af\|_2^2 + \sum_{i=1}^{q} \lambda_i^2 \|L_i f\|_2^2 \right\},$$

which in turn, can be calculated as

$$f_\lambda^{(k)} = V_k z_\lambda^{(k)}, \quad z_\lambda^{(k)} = \underset{z\in\mathbb{R}^k}{\operatorname{argmin}} \left\{ \|\beta_1 e_1 - B_k z\|_2^2 + \sum_{i=1}^{q} \lambda_i^2 \|L_i V_k z\|_2^2 \right\}, \tag{3.4}$$

where the last minimization problem is referred to as the projected problem. When $L_1 = I$ and $L_i = 0$, $i \neq 1$, the method of the present paper reduces to GKB-FP. Note that if we calculate the QR decomposition of $L_i V_k$, $L_i V_k = Q_i^{(k)} R_i^{(k)}$, the norm $\|L_i V_k z\|_2$ in (3.4) can be replaced by $\|R_i^{(k)} z\|_2$. Proceeding this way, the minimization problem becomes simpler and computation of $f_\lambda^{(k)}$ via (3.4) reduces to

$$f_\lambda^{(k)} = V_k z_\lambda^{(k)}, \quad z_\lambda^{(k)} = \underset{z \in \mathbb{R}^k}{\operatorname{argmin}} \|B_\lambda^{(k)} z - \bar{g}\|_2,$$

where $B_\lambda^{(k)} = \begin{bmatrix} B_k \\ \lambda_1 R_1^{(k)} \\ \vdots \\ \lambda_q R_q^{(k)} \end{bmatrix}$, and $\bar{g} = \begin{bmatrix} \beta_1 e_1 \\ 0 \end{bmatrix}$; so $z_\lambda^{(k)}$ can be efficiently computed in several ways, e.g., by a direct method or by first transforming the matrix $B_\lambda^{(k)}$ to upper triangular form, as done when implementing GKB-FP [5]. In addition, the approximate solution $f_\lambda^{(k)}$ and the corresponding residual $r_\lambda^{(k)} = g - Af_\lambda^{(k)}$ satisfy

$$\|L_i f_\lambda^{(k)}\|_2 = \|R_k z_\lambda^{(k)}\|_2, \quad \|r_\lambda^{(k)}\|_2 = \|B_k z_\lambda^{(k)} - \beta_1 e_1\|. \tag{3.5}$$

For each $k \geqslant 1$ consider the function

$$\Phi^{(k)}(\lambda) = [\phi_1^{(k)}(\lambda, \mu_1), \ldots, \phi_q^{(k)}(\lambda, \mu_q)]^T,$$

with

$$\phi_i^{(k)}(\lambda, \mu_i) = \sqrt{\mu_i} \frac{\|B_k z_\lambda^{(k)} - \beta_1 e_1\|_2}{\|R_i^{(k)} z_\lambda^{(k)}\|_2}, \quad i = 1, \ldots, q. \tag{3.6}$$

Our proposal for large-scale multi-parameter Tikhonov regularization is to follow the same steps as GKB-FP. That is, for chosen $p > 1$ and $k \geqslant p$, our projection algorithm computes the fixed-point $\lambda^{(k)^*}$ of $\Phi^{(k)}(\lambda)$ that minimizes $\Psi^{(k)}(\lambda) = \|r_\lambda^{(k)}\|_2^2 \prod_{i=1}^q \|L_i f_\lambda^{(k)}\|_2^{2\mu_i}$, following, e.g., MFP, and proceeds by repeating the process until a stopping criterion is satisfied. For algorithmic details of GKB-FP, the reader is referred to [5]. Numerical examples have shown that the minimizer of $\Psi(\lambda)$ associated with the large-scale problem is captured in a relatively small number of GKB steps.

To make our proposal computationally feasible, the following aspects must be considered

- the initial guess of the fixed-point method on the projected problem at step $k+1$ is taken to be the fixed-point $\lambda^{(k)^*}$ and
- the QR factorization $L_i V_p = Q_i^{(p)} R_i^{(p)}$ is calculated only once at step p, and is updated in subsequent steps.

Algorithms for updating the QR factorization can be found in [15, Chapter 12].

4. Numerical results

We illustrate the effectiveness of our algorithm by considering two-parameter and three-parameter regularization cases. Two problems are taken from the Regularization Toolbox [18] and one problem comes from image super-resolution. For each problem we ran 20 instances with distinct noisy vectors $g = g^{\text{exact}} + e$ where e is a random vector generated by the Matlab randn function, scaled so that $\text{NL} = \|e\|/\|g^{\text{exact}}\| = 0.001,\ 0.01$ and 0.025. For comparison, we also report results obtained by the method of Brezinski et al. (CLC for short) [8], and the method of Lu and Pereverzev [24] which is based on the discrepancy principle (DP). CLC proposes as approximate solution a linear combination of one-parameter solutions f_{λ_i}:

$$f_\lambda(\eta) = \sum_{i=1}^q \eta_i f_{\lambda_i}, \quad \eta = (\eta_1, \ldots, \eta_q), \tag{4.1}$$

where

$$f_{\lambda_i} = \underset{f \in \mathbb{R}^n}{\operatorname{argmin}} \{\|g - Af\|_2^2 + q\lambda_i^2 \|Lf\|_2^2,\ i = 1, \ldots q\}, \tag{4.2}$$

with the parameters η_i being constrained to $\eta_1 + \cdots + \eta_q = 1$. In our implementation of CLC, one-parameter solutions f_{λ_i} are computed by the FP method while the parameters η_i are computed following the criterion described in [8]. As for the discrepancy principle, the idea is to choose regularization parameters $(\lambda_1, \ldots, \lambda_q)$ such that the regularized solution f_λ satisfies

$$\|g - Af_\lambda\|_2 = c\delta, \quad c \geqslant 1, \tag{4.3}$$

which is a non linear equation with infinitely many solutions. The main difficulty with DP is that it can produce useless solutions if one of the computed parameters is very small, as we will illustrate later.

The starting values for MFP were taken to be the parameters used by CLC in (4.1) and the iterations terminated when $\|\lambda^{(k+1)} - \lambda^{(k)}\| \leqslant \lambda^{(k)} 10^{-6}$. The initial guess and other parameters required by DP are the same as in [24], i.e., we take $\lambda_1^{(0)} = \sqrt{0.2}$, $\lambda_2^{(0)} = \sqrt{0.1}$, $c = 1$, $\gamma = 0.5$ and $\delta = \|e\|_2$. All computations were carried out on a Core I7 with 3.3 GHz and 8 GB RAM using Matlab. To describe the results we use the following notation:

- $\bar{E}_f$, $\bar{\lambda}_1$, $\bar{\lambda}_2$: average values of relative error in f_λ and average values of computed parameters.
- k_M: maximum number of iterations required by MFP to converge or maximum subspace dimension of $\mathcal{K}_k(A^TA, A^Tb)$ after convergence of the extended GKB-FP algorithm.

4.1. Fredholm integral equations of first kind

We consider two discrete ill-posed problems arising from the discretization of two Fredholm integral equations of the form

$$\int_a^b K(s,t)f(t)\,dt = g(s), \quad c \leqslant s \leqslant d, \tag{4.4}$$

generated by the functions i_laplace and phillips from [18]. Therefore, for each problem there is a triple $\{A, f^{\text{exact}}, g^{\text{exact}}\}$ such that $Af^{\text{exact}} = g^{\text{exact}}$. As regularization matrices we choose combinations of the identity matrix I and discrete differential operators $\mathcal{L}_{1,n} \in \mathbb{R}^{(n-1)\times n}$ and $\mathcal{L}_{2,n} \in \mathbb{R}^{(n-2)\times n}$ defined by

$$\mathcal{L}_{1,n} = \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{bmatrix}, \quad \mathcal{L}_{2,n} = \begin{bmatrix} 1 & -2 & 1 & & & \\ & 1 & -2 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & 1 & -2 & 1 \end{bmatrix}. \tag{4.5}$$

4.1.1. Inversion of the Laplace transformation

In this case the kernel in (4.4) is $K(s,t) = e^{-st}$, $s, t \in [0, \infty)$ while the right-hand side and corresponding solution are $g(s) = 1/(s + 1/2)$ and $f(t) = e^{-t/2}$, respectively, and the associated discrete ill-posed problem is generated by the function i_laplace (Example 1). For the selected size, $n = 256$, the numerical rank of A is 36 and the condition number $\kappa(A) \approx 1.4 \times 10^{33}$. As for the regularization matrices, we consider the cases: (i) $L_1 = I$ and $L_2 = \mathcal{L}_{1,n}$, and (ii) $L_1 = I$ and $L_2 = \mathcal{L}_{2,n}$. Numerical results for the case (i) are summarized in Table 1. Note that all algorithms produced results with approximately the same order of accuracy and that the number of iterations required by MFP to converge are quite small (less than 5); concerning DP, it is also fast but we note that the constant $c = 1$ may prevent convergence, as occurs for NL = 0.001 where the maximum number of iterations (set to 100) is reached for a few runs. The reason is that the stopping rule for DP with $c = 1$, $\|g - Af\|_2 < c\delta$, may take too many steps to be satisfied, hence taking $c \gtrapprox 1$ can be more reliable; repeating the experiment with $c = 1.1$ the maximum number of iterations is 9. The results for case (ii) (Table 2) show that the quality of the MFP-based solutions is significantly better than the quality produced by both DP and CLC, the worst results being obtained by CLC. The reason of the poor quality of CLC-based solutions for NL = 0.01 and NL = 0.025 is because the rule used by CLC was not able to appropriately balance the one-parameter solutions. In this case both MFP and DP take approximately the same number of iterations to converge.

In the previous section we remarked that DP may lead to useless approximate solutions if one of the parameters $\lambda_1^{(k)}$ is very small. To illustrate such a situation we take the initial guess to be the same as that used by CLC and keep the other

Table 1
Numerical results for i_laplace test problem with $L_1 = I$ and $L_2 = \mathcal{L}_{1,n}$.

	NL = 0.001			NL = 0.01			NL = 0.025		
	MFP	CLC	DP	MFP	CLC	DP	MFP	CLC	DP
$\bar{E}_f$	0.0170	0.0632	0.0104	0.0200	0.0597	0.0292	0.0277	0.0754	0.0579
$\bar{\lambda}_1$	0.0023	0.0023	0.0002	0.0229	0.0231	0.0482	0.0580	0.0584	0.0859
$\bar{\lambda}_2$	0.0322	0.0322	0.1980	0.3262	0.3257	0.2106	0.8315	0.8265	0.2318
k_M	3	–	100	4	–	8	5	–	6

Table 2
Numerical results for i_laplace test problem with $L_1 = I$ and $L_2 = \mathcal{L}_{2,n}$.

	NL = 0.001			NL = 0.01			NL = 0.025		
	MFP	CLC	DP	MFP	CLC	DP	MFP	CLC	DP
$\bar{E}_f$	0.0176	0.5148	0.0615	0.0792	1.4898	0.1469	0.0809	1.9383	0.1627
$\bar{\lambda}_1$	0.0023	0.0023	0.0075	0.0255	0.0231	0.0429	0.0615	0.0584	0.0774
$\bar{\lambda}_2$	0.3864	0.3868	0.0115	6.7183	6.8705	0.0323	18.0429	17.6902	0.0658
k_M	6	–	13	10	–	8	10	–	7

Table 3
Numerical results for i_laplace test problem with $L_1 = I$ and $L_2 = \mathcal{L}_{2,n}$ and the same initial guess as CLC.

	NL = 0.001		NL = 0.01		NL = 0.025	
	MFP	DP	MFP	DP	MFP	DP
$\bar{E}_f$	0.0176	0.4883	0.0792	**1.0634**	0.0809	**1.2596**
$\bar{\lambda}_1$	0.0023	0.0003	0.0255	0.00000001	0.0615	0.00000002
$\bar{\lambda}_2$	0.3864	0.3376	6.7183	1.8030	18.0429	5.1006
k_M	6	35	10	42	10	44

Table 4
Average error in one-parameter regularized solutions for i_laplace test problem.

	NL = 0.001		NL = 0.01		NL = 0.025	
	FP	DP	FP	DP	FP	DP
$L = I$	0.1552	0.1612	0.1793	0.1866	0.1930	0.1995
$L = \mathcal{L}_{1,n}$	0.0636	0.0127	0.0600	0.0475	0.0758	0.0702
$L = \mathcal{L}_{2,n}$	0.5146	0.6402	1.4897	**1.1710**	1.9383	**1.2348**

starting parameters unchanged. The results displayed in Table 3 are apparent. Note that for NL = 0.01 and NL = 0.025 the computed solution are dominated by noise (with relative error exceeding 100%). The reason is that $\bar{\lambda}_1^2$ practically vanishes (see Table 3), in which case the solutions determined by DP behave similarly as those determined by DP applied to a one-parameter problem with the choice $L = \mathcal{L}_{2,n}$, see Table 4. Note that the errors of the solution determined by DP (in boldface) for the one-parameter case in Table 4 are close to those shown in Table 3 (in boldface too).

4.1.2. Phillips test problem

This test problem was first studied by Phillips [29] and then analyzed in several places. The kernel is of the form $K(s,t) = \varphi(s-t)$, $s,t \in [-6,6]$, with

$$\varphi(t) = \begin{cases} 1+\cos(\pi t/3), & |t| < 3 \\ 0, & |t| \geqslant 3 \end{cases},$$

while the solution and the right-hand side are $f(t) = \varphi(t)$ and

$$g(s) = (6-|s|)\left[1+\frac{1}{2}\cos(\pi s/3)\right] + \frac{9}{2\pi}\sin(\pi|s|/3).$$

As before we take $n = 256$ and consider the same regularization matrices as in cases (i) and (ii) of the inverse Laplace transformation test problem. The data set is generated by phillips function. The results are summarized in Tables 5 and 6. For this test problem, both MFP and DP converged in a few iterations (as illustrated by k_M), and all algorithms performed quite well for all noise levels.

4.2. Super-resolution image reconstruction problem

High-resolution (HR) images are important in a number of areas such as medical imaging and video surveillance. We consider the problem of estimating an HR image from observed multiple Low-resolution (LR) images. Let the original HR image of size $M = M_1 \times M_2$ in vector form be denoted by $f \in \mathbb{R}^M$, and let the k-th LR image of size $N = N_1 \times N_2$ in vector form be denoted by $g_k \in \mathbb{R}^N$, $k = 1,2,\ldots,q$, with $M_1 = N_1 \times D_1$, $M_2 = N_2 \times D_2$, where D_1 and D_2 represent down-sampling factors for the horizontal and vertical directions, respectively. Assuming that the acquisition process of the LR sequence involves blurring, motion, subsampling and additive noise, an observation model that relates f to g_k is written as [28]

$$g_k = A_k f + \epsilon_k, \tag{4.6}$$

where A_k is $N \times M$, and ϵ_k stands for noise. To estimate the HR image f from all LR images g_k via single-parameter Tikhonov regularization we solve the problem

$$f_\lambda = \underset{f\in\mathbb{R}^M}{\operatorname{argmin}}\{\|g - Af\|_2^2 + \lambda^2\|Lf\|_2^2\}, \tag{4.7}$$

where $g = [g_1^T \cdots g_q^T]^T$, $A = [A_1^T \cdots A_q^T]^T$, and L is an appropriate regularization matrix. Regularization is needed as A is severely ill-conditioned. Here we estimate the 96×96 image *tree* from a sequence of five noisy LR images with $D_1 = D_2 = 2$, hence $A \in \mathbb{R}^{11520\times 9216}$. We consider two and three parameter regularization cases involving as regularization matrices the $M \times M$ identity matrix, I_M, and the discrete 2D differential operators defined by

Table 5
Numerical results for phillips test problem with $L_1 = I$ and $L_2 = \mathcal{L}_{1,n}$.

	NL = 0.001			NL = 0.01			NL = 0.025		
	MFP	CLC	DP	MFP	CLC	DP	MFP	CLC	DP
$\overline{E}_f$	0.0138	0.0143	0.0100	0.0218	0.0248	0.0238	0.0309	0.0328	0.0339
$\bar{\lambda}_1$	0.0049	0.0048	0.0534	0.0502	0.0494	0.1614	0.1274	0.1247	0.2468
$\bar{\lambda}_2$	0.1742	0.1741	0.0836	1.7829	1.7806	0.2155	4.5766	4.5476	0.3807
k_M	4	–	7	4	–	3	5	–	2

Table 6
Numerical results for phillips test problem with $L_1 = I$ and $L_2 = \mathcal{L}_{2,n}$.

	NL = 0.001			NL = 0.01			NL = 0.025		
	MFP	CLC	DP	MFP	CLC	DP	MFP	CLC	DP
$\overline{E}_f$	0.0097	0.0097	0.0174	0.0262	0.0262	0.0240	0.0477	0.0461	0.0344
$\bar{\lambda}_1$	0.0050	0.0048	0.1581	0.0507	0.0494	0.1614	0.1326	0.1247	0.2468
$\bar{\lambda}_2$	3.7597	3.7591	0.00003	39.4760	39.3964	0.0009	110.507	109.297	0.1383
k_M	4	–	3	5	–	3	9	–	2

Table 7
Numerical results for image super-resolution *tree* test problem.

	NL = 0.001			NL = 0.01			NL = 0.025		
	A	B	C	A	B	C	A	B	C
$\overline{E}_f$	0.0361	0.0274	0.0310	0.0533	0.0493	0.0516	0.0709	0.0720	0.0703
$\bar{\lambda}_1$	0.0007	0.0007	0.0007	0.0082	0.0086	0.0089	0.0222	0.0238	0.0257
$\bar{\lambda}_2$	0.0025	0.0026	0.0027	0.0301	0.0335	0.0340	0.0827	0.0970	0.1019
$\bar{\lambda}_3$	–	–	0.0027	–	–	0.0351	–	–	0.1104
k_M	313	207	177	51	47	46	38	41	54

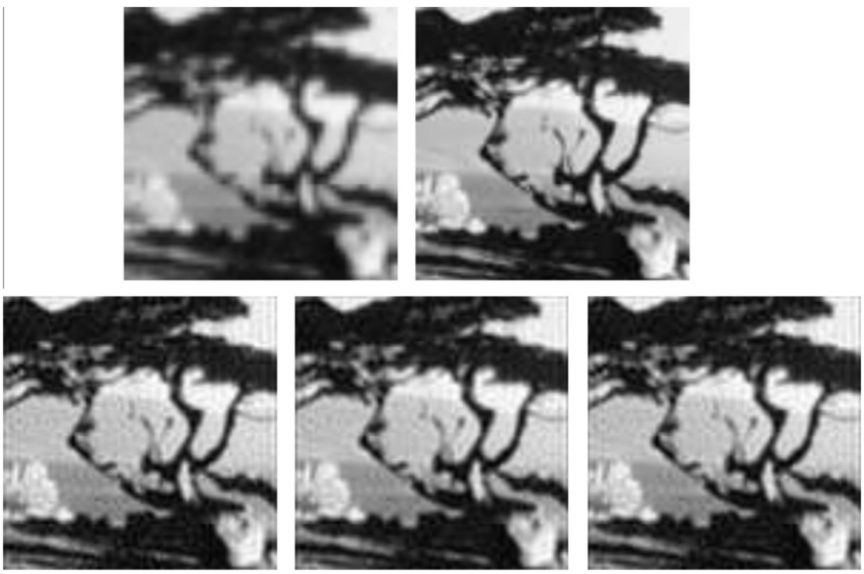

Fig. 5. Top: blurred+noise LR image (left), HR image (right). Bottom: estimated images for the cases A (left), B (center) and C (right), respectively. $N = 0.001$.

$$\overline{\mathcal{L}}_1 = \begin{bmatrix} I_{M_1} \otimes \mathcal{L}_{1,M_1} \\ \mathcal{L}_{1,M_1} \otimes I_{M_1} \end{bmatrix}, \quad \overline{\mathcal{L}}_2 = \begin{bmatrix} I_{M_1} \otimes \mathcal{L}_{2,M_1} \\ \mathcal{L}_{2,M_1} \otimes I_{M_1} \end{bmatrix},$$

where $\mathcal{L}_{1,M_1}$ and $\mathcal{L}_{2,M_1}$ are as in (4.5). Note that in this case $M_1 = 96$ implies $\overline{\mathcal{L}}_1 \in \mathbb{R}^{18240\times 9216}$ and $\overline{\mathcal{L}}_2 \in \mathbb{R}^{18048\times 9216}$. Hence matrix $\overline{A}_\lambda$ in (2.20) becomes too large and the multi-parameter Tikhonov problem (1.3) may not be handled efficiently via QR decomposition as done in the previous examples. It is precisely for large-scale problems like this that our projection algorithm is useful. For the two-parameter case we consider the choices A: $L_1 = I_M$, $L_2 = \overline{\mathcal{L}}_1$ and B: $L_1 = I_M$, $L_2 = \overline{\mathcal{L}}_2$. For the three-parameter case (labeled by C) we choose $L_1 = I_M$, $L_2 = \overline{\mathcal{L}}_1$ and $L_3 = \overline{\mathcal{L}}_2$. Average results summarized in Table 7 show that the quality of the computed solutions at each noise level is about the same for the three choices A, B or C. Note that in accordance with regularizing properties of Krylov methods, the number of iterations tend to decrease as the noise level grows. Fig. 5 displays one blurred+noise LR image, the true HR image, and three restored HR images, corresponding to respectively the cases A, B and C. Since the restored images corresponding to the cases A, B or C are nearly indistinguishable independently of the chosen noise level, we conclude that for this test problem there is no need to apply three-parameter regularization with regularization matrices as in case C.

5. Conclusions

We presented a generalization of Regińska's choice rule that resulted in a fixed-point algorithm for multi-parameter Tikhonov regularization called MFP. The method does not require a priori knowledge of the noise level in the data. The analysis and the resulting algorithm presented here can therefore be regarded as a natural generalization of the results and algorithm for the one-parameter Tikhonov problem published in [3]. The numerical results show that multi-parameter Tikhonov regularization can improve significantly the quality of the standard one-parameter Tikhonov formulation. In particular, the results show that MFP performs in general better than CLC, and that MFP can produce solutions with accuracy comparable to that of the discrepancy principle, with the observation that the drawbacks of the discrepancy principle, namely, the need for a priori knowledge of the error norm $\|e\|_2$ and non-uniqueness of solutions of the discrepancy equation (4.3), are not present when using MFP. Further, following the main ideas of GKB-FP (which realizes the one-parameter fixed-point method for large-scale problems), we proposed a GKB-FP type algorithm for large-scale multi-parameter Tikhonov regularization. Further investigation of the proposed algorithm is necessary in order to assess its potential; efficient ways to solve the multi-parameter Tikhonov regularization are the subject of ongoing research. A rigorous analysis about existence of fixed-points for the multi-parameter case and their classification as done in [4] for one-parameter problems is rather involved and therefore postponed to a future work.

Acknowledgments

The authors are grateful to the reviewers for their valuable comments which have significantly improved the presentation of this work.

References

[1] F. Bauer, O. Ivanyshyn, Optimal regularization with two interdependent regularization parameters, Inverse Prob. 23 (2007) 331–342.
[2] F. Bauer, S.V. Pereverzev, An utilization of a rough approximation of a noise covariance within the framework of multi-parameter regularization, Int. J. Tomogr. Stat. 4 (2006) 1–12.
[3] F.S.V. Bazán, Fixed-point iterations in determining the Tikhonov regularization parameter, Inverse Prob. 24 (2008) 035001.
[4] F.S.V. Bazán, J.B. Francisco, An improved fixed-point algorithm for determining a Tikhonov regularization parameter, Inverse Prob. 25 (2009) 045007.
[5] F.S.V. Bazán, L.S. Borges, GKB-FP: an algorithm for large-scale discrete ill-posed problems, BIT 50 (2010) 481–507.
[6] M. Belge, M.E. Kilmer, E.L. Miller, Simultaneous multiple regularization parameter selection by means of the L-hypersurface with applications to linear inverse problems posed in the wavelet domain, in: Proceedings of the SPIE' 98 – Bayesian Inference for, Inverse Problems, 1998, p. 3459.
[7] M. Belge, M.E. Kilmer, E.L. Miller, Efficient determination of multiple regularization parameters in a generalized L-curve framework, Inverse Prob. 18 (2002) 1161–1183.
[8] C. Brezinski, M. Redivo-Zaglia, G. Rodriguez, S. Seatzu, Multi-parameter regularization techniques for ill-conditioned linear systems, Numer. Math. 94 (2003) 203–228.
[9] Z. Chen, Y. Lu, Y. Xu, H. Yang, Multi-parameter Tikhonov regularization for linear ill-posed operator equations, J. Comput. Math. 26 (2008) 37–55.
[10] Z. Chen, C. Xiang, K. Zhao, X. Liu, Convergence analysis of Tikhonov-type regularization algorithms for multiobjective optimization problems, Appl. Math. Comput. 211 (2009) 167–172.
[11] J. Chung, J.G. Nagy, D.P. O'Leary, A weighted-GCV method for Lanczos-hybrid regularization, Electron. Trans. Numer. Anal. 28 (2008) 149–167.
[12] D. Düvelmeyer, B. Hofmann, A multi-parameter regularization approach for estimating parameters in jump diffusion processes, J. Inverse Ill-Posed Prob. 14 (2006) 861–880.
[13] G.H. Golub, M. Heath, G. Wahba, Generalized cross-validation as a method for choosing a good ridge parameter, Technometrics 21 (1979) 215–222.
[14] G.H. Golub, W. Kahan, Calculating the singular values and pseudo-inverse of a matrix, SIAM J. Numer. Anal. Ser. B 2 (1965) 205–224.
[15] G.H. Golub, C.F. Van Loan, Matrix Computations, third ed., The Johns Hopkins University Press, London, 1996.
[16] L. Gongsheng, L. Jinqing, F. Xiaoping, M. Yu, A new gradient regularization algorithm for source term inversion in 1D solute transportation with final observations, Appl. Math. Comput. 196 (2008) 646–660.
[17] P.C. Hansen, D.P. O'Leary, The use of the L-curve in the regularization of discrete ill-posed problems, SIAM J. Sci. Comput. 14 (1993) 1487–1503.
[18] P.C. Hansen, Regularization tools: a MATLAB package for analysis and solution of discrete ill-posed problems, Numer. Algor. 6 (1994) 1–35.
[19] P.C. Hansen, Rank-Deficient and Discrete Ill-Posed Problems, SIAM, Philadelphia, PA, 1998.
[20] P.C. Hansen, T.K. Jensen, Smoothing-norm preconditioning for regularizing minimum-residual methods, SIAM. J. Matrix Anal. Appl. 29 (2006) 1–14.
[21] K.H. Leem, G. Pelekanos, F.S.V. Bazán, Fixed-point iterations in determining a Tikhonov regularization parameter in Kirsch's factorization method, Appl. Math. Comput. 21 (12) (2010) 3747–3753.

[22] J. Lampe, L. Reichel, H. Voss, Large-scale Tikhonov regularization via reduction by orthogonal projection, Linear Algebra Appl. 436 (8) (2012) 2845–2865.
[23] S. Lu, S.V. Pereverzev, Y. Shao, U. Tautenhahn, Discrepancy curves for multi-parameter regularization, J. Inverse Ill-Posed Prob. 18 (2010) 655–676.
[24] S. Lu, S.V. Pereverzev, Multi-parameter regularization and its numerical realization, Numer. Math 118 (2011) 1–31.
[25] Y. Lu, L. Shen, Y. Xu, Multi-parameter regularization methods for high-resolution image reconstruction with displacement errors, IEEE Trans. Circuits Syst. 54 (2007) 1788–1799.
[26] V.A. Morozov, Regularization Methods for Solving Incorrectly Posed Problems, Springer, New York, 1984.
[27] C.C. Paige, M.A. Saunders, LSQR: an algorithm for sparse linear equations and sparse least squares, ACM Trans. Math. Softw. 8 (1982) 43–71.
[28] S.C. Park, M.K. Park, M.G. Kang, Super-resolution image reconstruction: a technical overview, IEEE Signal Process. Mag. 20 (1996) 535–547.
[29] D.L. Phillips, A technique for the numerical solution of certain integral equations of the first kind, J. ACM 9 (1962) 84–97.
[30] M.P. Rajan, A posteriori parameter choice with an efficient discretization scheme for solving ill-posed problems, Appl. Math. Comput. 204 (2008) 891–904.
[31] T. Regińska, A regularization parameter in discrete ill-posed problems, SIAM J. Sci. Comput. 3 (1996) 740–749.
[32] L. Reichel, F. Sgallari, Q. Ye, Tikhonov regularization based on generalized Krylov subspace methods, Appl. Numer. Math. 62 (9) (2012) 1215–1228.
[33] L. Reichel, Q. Ye, Simple square smoothing regularization operators, Electron. Trans. Numer. Anal. 33 (2009) 63–83.
[34] M. Salahi, Regularization tools and robust optimization for ill-conditioned least squares problem: a computational comparison, Appl. Math. Comput. 217 (2011) 7985–7990.
[35] A.N. Tikhonov, Solution of incorrectly formulated problems and the regularization method, Soviet Math. Dokl. 4 (1963) 1035–1038.
[36] P. Xu, Y. Fukuda, Y. Liu, Multiple parameter regularization: numerical solutions and applications to the determination of geopotential from precise satellite orbits, J. Geod. 80 (2006) 17–27.
[37] J-F. Yin, Preconditioner based on the Sherman–Morrison formula for regularized least squares problems, Appl. Math. Comput. 215 (2009) 3007–3016.
[38] J. Zhang, M. Mammadov, A new method for solving linear ill-posed problems, Appl. Math. Comput. 218 (2012) 10180–10187.
[39] M.V.W. Zibetti, F.S.V. Bazán, J. Mayer, Determining the regularization parameters for super-resolution problems, Signal Process. 88 (2008) 2890–2901.