Simple and Efficient Determination of the Tikhonov Regularization Parameter Chosen by the Generalized Discrepancy Principle for Discrete Ill-Posed Problems

Fermín S. Viloche Bazán

Abstract Discrete ill-posed problems where both the coefficient matrix and the right hand side are contaminated by noise appear in a variety of engineering applications. In this paper we consider Tikhonov regularized solutions where the regularization parameter is chosen by the generalized discrepancy principle (GDP). In contrast to Newton-based methods often used to compute such parameter, we propose a new algorithm referred to as GDP-FP, where derivatives are not required and where the regularization parameter is calculated efficiently by a fixed-point iteration procedure. The algorithm is globally and monotonically convergent. Additionally, a specialized version of GDP-FP based on a projection method, that is well-suited for large-scale Tikhonov problems, is also proposed and analyzed in detail. Numerical examples are presented to illustrate the effectiveness of the proposed algorithms on test problems from the literature.

Keywords Discrete ill-posed problems · Tikhonov regularization · Projection method · Generalized discrepancy principle · Noisy operator · Noisy right hand side

1 Introduction

We consider linear problems of the form

$$Af = g, \qquad A \in \mathbb{R}^{n \times n} \tag{1}$$

where the coefficient matrix A arises from the discretization of ill-posed problems such as first kind Fredholm integral equations with smooth kernel. Such matrices are severely ill-conditioned and may be singular. Linear problems possessing such a matrix are frequently

F. S. V. Bazán (✉)
Department of Mathematics, Federal University of Santa Catarina,
Florianópolis, SC CEP 88040-900, Brazil
e-mail: fermin@mtm.ufsc.br

referred to as *linear discrete ill-posed problems*. In these cases the solution of (1) is very sensitive to perturbations in the data such as measurement or approximation errors, and thus regularization methods are essential for computing meaningful approximate solutions. In this paper we consider the problem of computing stable approximations to the solution of (1) for the case where the data obey the following assumptions:

A1: Problem (1) is consistent, ie., g belongs to the column space of A, $\mathcal{R}(A)$,

A2: Instead of g and A we are given a noisy vector $\widetilde{g}$ and a noisy matrix $\widetilde{A}$ with available noise levels δ_g and δ_A, respectively, such that

$$\|\widetilde{g} - g\|_2 \le \delta_g, \quad \text{and} \quad \|\widetilde{A} - A\|_2 \le \delta_A. \tag{2}$$

Discrete ill-posed problems with data satisfying (2) appear in a number of applications such as potential theory, image processing and inverse scattering [1,8,9,17,22], among others. There are at least two well-distinguished approaches for these problems: (i) the classical Tikhonov regularization method, for which abundant literature exists, see, e.g., [11,23,30], and (ii) regularized total least squares methods [12,18,20,22,27–29]. Both approaches work well when the noise level in A and g are known accurately. However, since Tikhonov regularization is less expensive and simpler, we will concentrate only on this method.

The regularized solution determined by the method of Tikhonov is defined by

$$f_\lambda = \operatorname*{argmin}_{f \in \mathbb{R}^n} \left\{ \|\widetilde{A}f - \widetilde{g}\|_2^2 + \lambda^2 \|Lf\|_2^2 \right\}, \tag{3}$$

where $\lambda > 0$ is the regularization parameter and $L \in \mathbb{R}^{p \times n}$ is referred to as the regularization matrix. Notice that taking $\lambda = 0$ in (3) we obtain the ordinary least squares problem

$$f_{\text{LS}} = \operatorname*{argmin}_{f \in \mathbb{R}^n} \|\widetilde{A}f - \widetilde{g}\|_2^2. \tag{4}$$

A regularization Tikhonov problem is said to be in standard form if the matrix L is the identity matrix I_n, otherwise the Tikhonov regularization problem is said to be in general form. The regularization matrix L depends on the problem and is very often chosen as some discrete approximation to a derivative operator. The regularization parameter determines how close the regularized solution f_λ is to the solution of (1), and should be chosen carefully.

The goal of the present paper is to construct Tikhonov regularized solutions for problem (1) with data satisfying **A1**–**A2**, using the generalized discrepancy principle (GDP) [23] as parameter choice rule. GDP is very popular due to its rigorous theoretical foundation; for a detailed analysis of GDP in a Hilbert space setting, which gives rise to error bounds including convergence rates and algorithms, the reader is referred to [21]. GDP chooses the regularization parameter as the solution of the nonlinear equation, called the discrepancy equation,

$$G(\lambda) = \|\widetilde{A}f_\lambda - \widetilde{g}\|_2^2 - (\delta_g + \delta_A \|Lf_\lambda\|_2)^2 = 0. \tag{5}$$

The function G is increasing and has a unique root under appropriate conditions. Additionally, it is convex for small λ and concave for large λ. As a result, global and monotone convergence of Newton's method cannot be guaranteed [29]. To overcome this difficulty, Lu et al. [21] transform the Eq. (5) into an equivalent one that depends on two free parameters (η, ν) and perform an analysis to determine the set of parameters that ensure global convergence of Newton's method. Numerical examples reported in [21] suggest that among the pairs (η, ν) that guarantee convergence, there must be a pair for which the iteration converges faster, but the determination of such a pair is still lacking.

In this work, we propose a fixed-point type derivative-free algorithm for computing the regularization parameter chosen by GDP which we refer to as GDP-FP. The algorithm is shown to be globally and monotonically convergent and has convergence properties that do not depend on additional parameters as required by the algorithms proposed in [21]. GDP-FP can be readily implemented using the generalized singular value decomposition (GSVD) of the matrix pair $(\widetilde{A}, L)$ when $\widetilde{A}$ and L are small or of moderate size. For large scale-problems, a specialized version of GDP-FP based on a projection method is also proposed and discussed in detail. The main idea of the method introduced here is to approximate the regularized solution f_λ by using a sequence of regularized solutions $f_\lambda^{(k)}$ in appropriate Krylov subspaces of increasing dimension, in such a way that the regularization parameter for the projected problem is chosen as the solution of "the projected discrepancy equation":

$$G^{(k)}(\lambda) = \|\widetilde{A} f_\lambda^{(k)} - \widetilde{g}\|_2^2 - (\delta_b + \delta_A \|L f_\lambda^{(k)}\|_2)^2 = 0. \tag{6}$$

Analogous to $G(\lambda)$, $G^{(k)}(\lambda)$ increases as λ increases and has a unique solution under appropriate conditions.

The rest of the paper is organized as follows. Section 2 summarizes properties regarding Tikhonov regularization and some technical results to be used later on. The theoretical and computational aspects of GDP-FP are developed and discussed in Sect. 3. Section 4 starts with a review of the Golub-Kahan bidiagonalization (GKB) algorithm [4,6], and proceeds with our specialized version of GDP-FP for large-scale problems. The proposed approach combines GKB with Tikhonov regularization in the generated Krylov subspace, where the regularization parameter for the projected problem is chosen by GDP-FP, following the ideas of the algorithm GKB-FP proposed in [4]. Again, the convergence properties of the proposed approach do not depend on user specified parameters. In Sect. 5 we provide numerical results that illustrate the effectiveness of our algorithm on test problems from the literature, including an inverse scattering problem. The paper ends with some concluding remarks.

2 Preliminaries

This section describes some technical results and the notation used throughout the paper. We start by noting that the regularized solution f_λ can be equivalently defined as the solution of the linear least squares problem

$$\min_{f \in \mathbb{R}^n} \left\| \begin{pmatrix} \widetilde{A} \\ \lambda L \end{pmatrix} f - \begin{pmatrix} \widetilde{g} \\ 0 \end{pmatrix} \right\|_2. \tag{7}$$

Hence it follows that the regularized solution f_λ is unique when the coefficient matrix in (7) has full column rank, which means the null spaces $\mathcal{N}(\widetilde{A})$ and $\mathcal{N}(L)$ intersect trivially, i.e.,

$$\mathcal{N}(\widetilde{A}) \cap \mathcal{N}(L) = \{0\}, \tag{8}$$

and the regularized solution f_λ solves the normal equations for (7)

$$(\widetilde{A}^T \widetilde{A} + \lambda^2 L^T L) f = \widetilde{A}^T \widetilde{g}. \tag{9}$$

To analyze properties of f_λ and related quantities, we need to introduce the generalized singular value decomposition (GSVD) of the matrix pair $(\widetilde{A}, L)$ [15]. For this, as usual, we assume that L is $p \times n$, $\mathrm{rank}(L) = p \leq n$, and that the pair $(\widetilde{A}, L)$ have a GSVD

$$\widetilde{A} = U \begin{bmatrix} \Sigma & 0 \\ 0 & I_{n-p} \end{bmatrix} X^{-1}, \quad L = V [M; 0] X^{-1}. \tag{10}$$

Here both $U = [u_1, \ldots, u_n] \in \mathbb{R}^{m \times n}$ and $V = [v_1, \ldots, v_p] \in \mathbb{R}^{p \times p}$ have orthonormal columns, and $X \in \mathbb{R}^{n \times n}$ is nonsingular. Moreover, Σ and M are diagonal matrices defined by $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_p)$, $M = \text{diag}(\mu_1, \ldots, \mu_p)$, whose entries are nonnegative and normalized such that

$$\sigma_i^2 + \mu_i^2 = 1, \quad i = 1, \ldots, p.$$

Then the generalized singular values of the matrix pair $(\widetilde{A}, L)$ are defined by

$$\gamma_i = \sigma_i / \mu_i, \quad i = 1, \ldots, p \tag{11}$$

and we shall assume them to appear in nondecreasing order. With the GSVD at hand, the regularized solution can be expressed as

$$f_\lambda = \sum_{i=1}^p \frac{\gamma_i^2}{\gamma_i^2 + \lambda^2} \frac{u_i^T \widetilde{g}}{\sigma_i} x_i + \sum_{i=p+1}^n u_i^T \widetilde{g} \, x_i. \tag{12}$$

Here x_i denotes the i-th column of matrix X. Define also $\alpha_i = |u_i^T \widetilde{g}|^2$ (the squared Fourier coefficient of $\widetilde{g}$) and $\delta_0 = \|(I - UU^T)\widetilde{g}\|_2$ (the size of the incompatible component of $\widetilde{g}$ that lies outside the column space of $\widetilde{A}$). Further, let

$$\mathbf{y}(\lambda) = \eta^2(\lambda) \doteq \|L f_\lambda\|_2^2, \tag{13}$$
$$\mathbf{x}(\lambda) = \rho^2(\lambda) \doteq \|b - \widetilde{A} f_\lambda\|_2^2. \tag{14}$$

In terms of the GSVD these functions reduce to

$$\mathbf{x}(\lambda) = \sum_{i=1}^p \frac{\lambda^4 \alpha_i}{(\gamma_i^2 + \lambda^2)^2} + \delta_0^2, \quad \mathbf{y}(\lambda) = \sum_{i=1}^p \frac{\gamma_i^2 \alpha_i}{(\gamma_i^2 + \lambda^2)^2}. \tag{15}$$

Hence, for $\lambda > 0$, the derivatives with respect to λ of $\mathbf{x}(\lambda)$ and $\mathbf{y}(\lambda)$, respectively, become

$$\mathbf{x}'(\lambda) = 4\lambda^3 \sum_{i=1}^p \frac{\gamma_i^2 \alpha_i}{(\gamma_i^2 + \lambda^2)^3} > 0, \quad \mathbf{y}'(\lambda) = -4\lambda \sum_{i=1}^p \frac{\gamma_i^2 \alpha_i}{(\gamma_i^2 + \lambda^2)^3} < 0. \tag{16}$$

Thus, the squared residual norm increases with λ and the squared seminorm decreases. Additionally, we can exploit (16) to obtain

$$d\mathbf{y}/d\mathbf{x} = -1/\lambda^2 \tag{17}$$

which implies $\frac{d^2 \mathbf{y}}{d\mathbf{x}^2} > 0$. So $\mathbf{y} = \mathbf{y}(\mathbf{x})$ (i.e., $\|Lf_\lambda\|_2^2$ as a function of $\|\widetilde{A} f_\lambda - \widetilde{g}\|_2^2$) is a monotonically decreasing convex function of $\mathbf{x}$. The following theorem shows that a similar result holds for the norms themselves.

Theorem 2.1 *Let f_λ denote the Tikhonov regularized solution of problem (3). Then $\|Lf_\lambda\|_2$ is a monotonically decreasing convex function of $\|\widetilde{A} f_\lambda - \widetilde{g}\|_2$. Moreover, the extreme residual norms (resp. extreme solution seminorms) corresponding to zero and infinite regularization satisfy the following properties*

$$\delta_0 \le \|\widetilde{A} f_\lambda - \widetilde{g}\| \le \|\widetilde{g}\|_2, \quad 0 \le \|Lf_\lambda\| \le \|Lf_{LS}\|_2, \tag{18}$$

with

$$\lim_{\lambda \to 0} \|\widetilde{A} f_\lambda - \widetilde{g}\| = \delta_0, \quad \lim_{\lambda \to \infty} \|\widetilde{A} f_\lambda - \widetilde{g}\| = \|\widetilde{g}\|_2 \tag{19}$$
$$\lim_{\lambda \to 0} \|Lf_\lambda\| = \|Lf_{LS}\|, \quad \lim_{\lambda \to \infty} \|Lf_\lambda\| = 0 \tag{20}$$

where f_{LS} is the unregularized solution of (3).

Proof See [15, Theorem 4.6.1].

Lemma 2.2 *Let* $\Phi(\lambda) = \frac{\|\widetilde{A}f_\lambda - \widetilde{g}\|_2}{\|Lf_\lambda\|_2}$. *Then* Φ *is a monotonically nondecreasing function, and* $\forall \lambda > 0$ *we have*

$$\Phi'(\lambda) < 2\frac{\Phi(\lambda)}{\lambda}. \tag{21}$$

Proof To prove that Φ increases as λ increases it suffices to use (16), (17) and the derivative of Φ. For the second part, using the notation introduced in (13), (14) note that

$$\mathbf{x}'(\lambda) = 2\rho(\lambda)\rho'(\lambda), \quad \text{and,} \quad \mathbf{y}'(\lambda) = 2\eta(\lambda)\eta'(\lambda).$$

This implies

$$\frac{d\mathbf{y}}{d\mathbf{x}} = \frac{1}{\Phi}\frac{d\eta}{d\rho} \Leftrightarrow \frac{d\eta}{d\rho} = -\frac{\Phi(\lambda)}{\lambda^2},$$

where the last equality is because of (17). Taking derivative with respect to λ on the right equality leads to

$$\frac{d^2\eta}{d\rho^2}\rho'(\lambda) = \frac{2\lambda\Phi(\lambda) - \lambda^2\Phi'(\lambda)}{\lambda^4}.$$

Now since $\rho'(\lambda) = \mathbf{x}'(\lambda)/\rho(\lambda) > 0$ by (16), and since by Theorem 2.1, η is a monotonically decreasing convex function of ρ, it follows that $\frac{d^2\eta}{d\rho^2}\rho'(\lambda) > 0$, and the the desired inequality follows. □

3 Fixed-Point Approach for Computing the Tikhonov Regularization Parameter Chosen by GDP

As already commented, in the GDP errors δ_A and δ_g are taken into account and the regularization parameter is chosen as the unique solution of the nonlinear equation $G(\lambda) = 0$. Existence and uniqueness of such solution can be explained by noting that G is a monotonically increasing continuous function of λ, which is immediate to see since $G'(\lambda) > 0$ due to (16). From this observation it is clear that G will have a unique root as long as

$$\lim_{\lambda \to 0^+} G(\lambda) = \delta_0^2 - (\delta_g + \delta_A\|Lf_{LS}\|_2)^2 < 0, \quad \text{and} \quad \lim_{\lambda \to \infty} G(\lambda) = \|\widetilde{g}\|_2^2 - \delta_g^2 > 0,$$

where the limit values are because of (19), (20) in Theorem 2.1. Consequently, the discrepancy equation $G(\lambda) = 0$ will have a unique solution as long as the noise δ_g and δ_A satisfy

$$\delta_0 - \delta_A\|Lf_{LS}\|_2 < \delta_g < \|\widetilde{g}\|_2. \tag{22}$$

If the noisy matrix $\widetilde{A}$ is nonsingular, as seen frequently when the noise is random, then $\delta_0 = 0$ and hence the left inequality is satisfied automatically. When $\widetilde{A}$ is singular, a reasonable condition that ensures the left inequality is $\delta_0 < \delta_g$. As for the right inequality, it holds true when the amount of noise in g satisfies $\delta_g \ll \|g\|_2$.

Having explained the issue of existence and uniqueness of the regularization parameter chosen by GDP, we shall now concentrate on the algorithmic aspects of methods for its computation. The special case $\delta_A = 0$ has been extensively studied by several authors and often treated with Newton's method. An implementation of the Discrepancy principle for

this very especial case is available as the Matlab function discrep in Regularization Tools by Hansen [14]. For the general case the situation is quite different. The only algorithms we are aware of are the Newton-based algorithms by Lu et al. [21] developed in a Hilbert space setting. The convergence and performance of these algorithms depend on user specified parameters.

In the following, we will develop a globally convergent algorithm for solving the discrepancy equation, with convergence properties that do not depend on any extra parameter. We will show, by means of numerical experiments, that this algorithm is capable of capturing the solution of the discrepancy equation very quickly.

We start with the following technical result.

Lemma 3.1 *For $\lambda > 0$ let $\vartheta(\lambda) = \frac{\|\widetilde{A} f_\lambda - \widetilde{g}\|_2}{\delta_g + \delta_A \|L f_\lambda\|_2}$ and $\xi(\lambda) = \frac{\lambda^2}{\vartheta(\lambda)}$. Then both $\vartheta(\lambda)$ and $\xi(\lambda)$ are monotonically increasing functions of λ. Moreover, the following properties hold*

$$\lim_{\lambda \to 0^+} \xi(\lambda) = \begin{cases} 0 & \text{if } \delta_0 \neq 0, \\ C > 0 & \text{if } \delta_0 = 0 \end{cases} \tag{23}$$

for some constant C.

Proof The proof that ϑ increases as λ increases is analogous to the proof what Φ is increasing, see Lemma 2.2. Therefore we will prove the assertion for function ξ only. In fact, differentiation of $\xi(\lambda)\vartheta(\lambda) = \lambda^2$ with respect to λ yields

$$\xi(\lambda)\vartheta'(\lambda) + \xi'(\lambda)\vartheta(\lambda) = 2\lambda.$$

Hence

$$\xi'(\lambda)\vartheta(\lambda) = \frac{2\lambda\vartheta(\lambda) - \lambda^2\vartheta'(\lambda)}{\vartheta(\lambda)}. \tag{24}$$

Further, since $\vartheta(\lambda) = c(\lambda)\Phi(\lambda)$ with $c(\lambda) = \frac{\eta(\lambda)}{\delta_b + \delta_A \eta(\lambda)}$, and $-\frac{\vartheta(\lambda)}{\lambda^2} = c(\lambda)\left(-\frac{\Phi(\lambda)}{\lambda^2}\right)$, taking derivative with respect to λ in this last equality we have

$$\frac{2\lambda\vartheta(\lambda) - \lambda^2\vartheta'(\lambda)}{\lambda^4} = c(\lambda)\left(\frac{2\lambda\Phi(\lambda) - \lambda^2\Phi'(\lambda)}{\lambda^4}\right) - \left(\frac{\Phi(\lambda)}{\lambda^2}\right)c'(\lambda). \tag{25}$$

Since $c'(\lambda) < 0$, which follows from the definition of c, and since $2\lambda\Phi(\lambda) - \lambda^2\Phi'(\lambda) > 0$ by Lemma 2.2, using these inequalities in (25) we obtain $\xi'(\lambda) > 0$, which proves that ξ is increasing.

The limit values in (23) are immediate consequences of (15) together with (19), (20). This concludes the proof.   □

We are in a position to describe the main results of this section.

Theorem 3.2 *Let ϑ and ξ be the functions defined in Lemma 3.1 and let $\zeta(\lambda) = \sqrt{\xi(\lambda)}$. Assume that (22) holds true. Then the following properties hold*

(a) *The parameter chosen by GDP, λ_{GDP}, is the unique solution of the nonlinear equation $\vartheta(\lambda) = 1$.*

(b) *ζ is a monotonically increasing function having a unique nonzero fixed-point at $\lambda = \lambda_{\text{GDP}}$ such that $\zeta(\lambda) \geq \lambda$ if $0 < \lambda \leq \lambda_{\text{GDP}}$, and $\zeta(\lambda) \leq \lambda$ if $\lambda \geq \lambda_{\text{GDP}}$. Additionally,*

$$\lambda_{\text{GDP}} \leq \gamma_1 \sqrt{\frac{\delta_b}{\rho(\gamma_1)} + \frac{\delta_A}{\gamma_1}} \tag{26}$$

as long as $\delta_g \gamma_1 + \delta_A \rho(\gamma_1) \leq \rho(\gamma_1)\gamma_1$.

(c) *For given $\lambda_0 > 0$ consider the sequence $\{\lambda_j\}$ defined by*

$$\lambda_{j+1} = \zeta(\lambda_j), \ j \geq 0. \tag{27}$$

Then $\lim_{j \to \infty} \lambda_j = \lambda_{\mathrm{GDP}}$ irrespective of the initial guess chosen.

Proof The equation $\vartheta(\lambda) = 1$ comes as another way of writing (5) and item (a) holds true because $\vartheta(\lambda)$ is a monotonically increasing function by Lemma 3.1.

We start the proof of item (b) with the observation that ζ increases monotonically because of its definition and Lemma 3.1. Now since λ_{GDP} is a solution of $\vartheta(\lambda) = 1$, then

$$\zeta(\lambda_{\mathrm{GDP}}) = \frac{\lambda_{\mathrm{GDP}}}{\sqrt{\vartheta(\lambda_{\mathrm{GDP}})}} = \lambda_{\mathrm{GDP}},$$

and thus λ_{GDP} is a fixed-point of ζ. Conversely, if λ^* is a fixed-point of ζ, which means $\zeta(\lambda^*) = \lambda^*$, then it follows that $\vartheta(\lambda^*) = 1$ and, due to item (a), we obtain $\lambda^* = \lambda_{\mathrm{GDP}}$. This implies that ζ has a unique nonzero fixed-point at $\lambda = \lambda_{\mathrm{GDP}}$.

We now proceed with the observation that $\zeta'(\lambda_{\mathrm{GDP}}) < 1$. We shall prove this by contradiction. In fact, assume $\zeta'(\lambda_{\mathrm{GDP}}) \geq 1$. Then since $\vartheta(\lambda_{\mathrm{GDP}}) = 1$ by item (a) and since

$$\zeta'(\lambda) = (\vartheta(\lambda))^{-1/2} - \frac{\lambda}{2} \vartheta'(\lambda)(\vartheta(\lambda))^{-3/2}, \tag{28}$$

taking $\lambda = \lambda_{\mathrm{GDP}}$ we obtain

$$1 \leq \zeta'(\lambda_{\mathrm{GDP}}) = 1 - \frac{\lambda_{\mathrm{GDP}}}{2} \vartheta'(\lambda_{\mathrm{GDP}}),$$

or equivalently $\vartheta'(\lambda_{\mathrm{GDP}}) \leq 0$, which is a contradiction as ϑ is increasing. Therefore we must have $\zeta'(\lambda_{\mathrm{GDP}}) < 1$. Using this property, the monotonicity of ζ and the fact ζ has a unique nonzero fixed-point, we conclude that $\zeta(\lambda) \geq \lambda$ if $0 < \lambda \leq \lambda_{\mathrm{GDP}}$ and $\zeta(\lambda) \leq \lambda$ if $\lambda \geq \lambda_{\mathrm{GDP}}$.

To prove the last part of item (b) we use the monotonicity of ϑ and Theorem 2.1 and deduce that for all $\lambda > 0$

$$\zeta(\lambda) \geq \lambda \sqrt{\frac{\delta_b}{\|\widetilde{g}\|_2}}. \tag{29}$$

We now recall from [2, Lemma1] that for $\lambda \geq \gamma_1$ we have $\rho(\lambda)/\eta(\lambda) \geq \lambda$. Thus, for $\lambda \geq \gamma_1$ we have

$$\frac{1}{\vartheta(\lambda)} = \frac{\delta_b + \delta_A \eta(\lambda)}{\rho(\lambda)} \leq \frac{\delta_b}{\rho(\lambda)} + \frac{\delta_A}{\lambda} \leq \frac{\delta_b}{\rho(\gamma_1)} + \frac{\delta_A}{\gamma_1}, \tag{30}$$

where the last inequality is because ρ is increasing, see (17). Combination of (29) and (30) shows that for $\lambda \geq \gamma_1$ we have

$$\sqrt{\frac{\delta_b}{\|\widetilde{g}\|_2}} \lambda \leq \zeta(\lambda) \leq \sqrt{\frac{\delta_b}{\rho(\gamma_1)} + \frac{\delta_A}{\gamma_1}} \lambda. \tag{31}$$

We now use the assumption $\delta_g \gamma_1 + \delta_A \rho(\gamma_1) \leq \rho(\gamma_1)\gamma_1$ and deduce that the slope of the upper bound in (31) is not larger than 1. This proves inequality (26) and ends the proof of item b).

Finally, since ζ is also an increasing function it follows that $\{\lambda_k\}$ will be either an increasing sequence if $\lambda_0 < \lambda_{\mathrm{GDP}}$ or a decreasing sequence if $\lambda_0 > \lambda_{\mathrm{GDP}}$. This concludes the proof. $\square$

From Theorem 3.2 we conclude that the regularization parameter chosen by GDP can be determined by a monotonically and globally convergent fixed-point iteration procedure as defined in (27). We will refer to this algorithm as GDP-FP. Moreover, from the nearly linear behavior of ζ for $\lambda \geq \gamma_1$ and the small slope of the bounds as described in (31), we deduce that the sequence (27) can converge rapidly when the iterative procedure is initialized either with $\lambda_0 \approx \gamma_1$ or with λ_0 chosen as the *a priori* upper bound (26), the latter option being valid for relatively low noise levels. Finally, we note that if $\delta_g = 0$

$$\lambda_{\mathrm{GDP}} \leq \sqrt{\gamma_1}\sqrt{\delta_A}, \tag{32}$$

while if $\delta_A = 0$, the upper bound (26) reduces to

$$\lambda_{\mathrm{GDP}} \leq \gamma_1 \sqrt{\frac{\delta_g}{\rho(\gamma_1)}}, \tag{33}$$

which is closely related to one given by Vinokurov [31].

The statements of Theorem 3.2 as well as the comments above are illustrated in Fig. 1.

We close the section with a description of our fixed-point algorithm for determining the regularization parameter λ_{GDP} and the corresponding regularized solution $f_{\lambda_{\mathrm{GDP}}}$

GDP-FP algorithm :

Input data : $\widetilde{A}, L, \widetilde{g}, \epsilon, \delta_g$ and δ_A.
1.   Compute the GSVD of the matrix pair $(\widetilde{A}, L)$.
2.   Set $j = 0$ and take, e.g., $\lambda_0 = \gamma_1$.
3.   Compute $\lambda_{j+1} = \zeta(\lambda_j)$ with ζ being defined in Thm. 3.2.
4.   **If** $|\lambda_{j+1} - \lambda_j| \geq \epsilon|\lambda_j|$
         do $j \leftarrow j + 1$ and **go to** 3.
         else stop.
5.   Compute f_λ by using (12).

The computation of the GSVD of the matrix pair (A, L) is quite expensive and for this reason, the GSVD is of interest only for small to medium-sized problems. For large-scale problems, we will see in the next section that projection methods are reliable options.
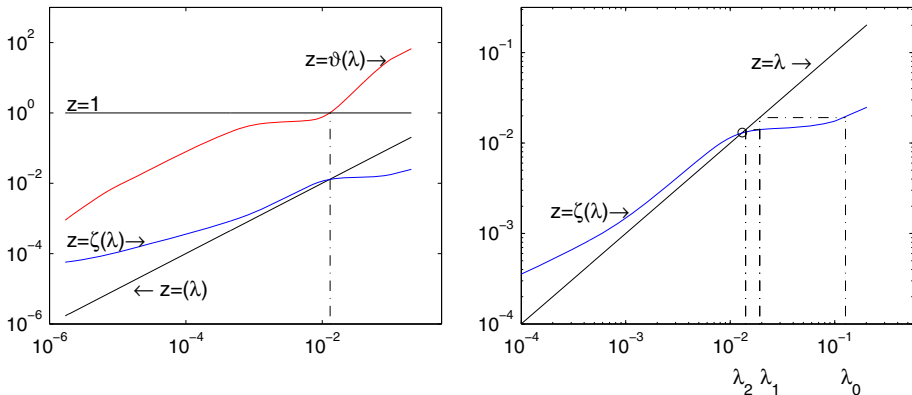

Fig. 1 Functions ϑ and ζ (*left*) and behavior of sequence $\{\lambda_k\}$ for $\lambda_0 > \lambda_{\mathrm{GDP}}$ (*right*) with the choice $\lambda_0 = \gamma_1$. This example corresponds to deriv2 test problem from [14] with $n = 512$. To simulate noise we take $\widetilde{g} = g + e$ and $\widetilde{A} = A + E$, were e is Gaussian random vector such that $\|e\|_2/\|g\|_2 = 0.01$, and E is a Gaussian random matrix such that $\|E\|_2/\|A\|_2 = 0.01$

4 Projection Approach for Large-Scale Problems

In this section we show how to construct regularized solutions f_λ for the case where the dimensions of $\widetilde{A}$ and L are such that the computation of the GSVD of the matrix pair $(\widetilde{A}, L)$ is infeasible or unattractive. The idea here is to construct approximate solutions to $f_{\lambda_{\mathrm{GDP}}}$ by "projecting" the large-scale problem onto a subspace of small dimension. More precisely, let $\{\mathcal{V}_k\}$, $k \geq 1$, be an appropriate nested family of k-dimensional subspaces in $\mathbb{R}^n$. Then the idea is to determine approximations $f_\lambda^{(k)} \in \mathcal{V}_k$ to f_λ by solving the constrained Tikhonov problem

$$f_\lambda^{(k)} = \operatorname*{argmin}_{f \in \mathcal{V}_k} \|\widetilde{A} f - \widetilde{g}\|_2^2 + \lambda^2 \|Lf\|_2^2 \tag{34}$$

where the regularization parameter is chosen in a proper way. Problem (34) is often referred to as *the projected problem* and methods which use $f_\lambda^{(k)}$ as approximations to f_λ are referred to as *projection methods*. For the moment, we assume that the Tikhonov problem is in standard form and we will concentrate on a projection method based on the Golub-Kahan bidiagonalization (GKB) algorithm [6] where the regularization parameter for the projected problem is chosen by GDP. A projection approach for Tikhonov regularization in general form will be addressed at the end of the section.

4.1 Review of GKB and LSQR

Application of $k < n$ GKB steps to $\widetilde{A}$ with initial vector $\widetilde{g}/\|\widetilde{g}\|_2$ yields three matrices: a lower bidiagonal matrix $B_k \in \mathbb{R}^{(k+1) \times k}$ and two matrices $U_{k+1} \in \mathbb{R}^{m \times (k+1)}$ and $V_k \in \mathbb{R}^{n \times k}$ with orthonormal columns, such that

$$\beta_1 U_{k+1} e_1 = \widetilde{g} = \beta_1 u_1, \tag{35}$$

$$\widetilde{A} V_k = U_{k+1} B_k, \tag{36}$$

$$\widetilde{A}^T U_{k+1} = V_k B_k^T + \alpha_{k+1} v_{k+1} e_{k+1}^T, \tag{37}$$

where e_k denotes the k-th unit vector in $\mathbb{R}^{k+1}$. The columns of V_k provide an orthonormal basis for the generated Krylov subspace $\mathcal{K}_k(\widetilde{A}^T \widetilde{A}, \widetilde{A}^T \widetilde{g})$, an excellent choice for use when solving ill-posed problems [6,7,14]. GKB iterations constitute the basis for the LSQR method [24, 25]. LSQR is designed to construct approximate solutions to the least squares problem (4) defined by $f^{(k)} = V_k y^{(k)}$, where $y^{(k)}$ solves the *projected* least squares problem

$$\min_{y \in \mathbb{R}^k} \|B_k y - \beta_1 e_1\|_2. \tag{38}$$

In practical computations $f^{(k)}$ is computed via a QR factorization of B_k which allows for an efficient updating of the LSQR iterates; the reader is referred to [24] for details. LSQR is also well suited for solving the "damped least squares problem" [24]

$$f_\lambda = \arg \min_{f \in \mathbb{R}^n} \left\| \begin{pmatrix} \widetilde{g} \\ 0 \end{pmatrix} - \begin{pmatrix} \widetilde{A} \\ \lambda I_n \end{pmatrix} f \right\|_2, \tag{39}$$

where λ is a fixed regularization parameter. In this case, the kth approximate solution is taken to be

$$f_\lambda^{(k)} = V_k y_\lambda^{(k)}, \tag{40}$$

where $y_\lambda^{(k)}$ solves the regularized projected problem

$$y_\lambda^{(k)} = \arg\min_{y \in \mathbb{R}^k} \left\| \begin{pmatrix} \beta_1 e_1 \\ 0 \end{pmatrix} - \begin{pmatrix} B_k \\ \lambda I_k \end{pmatrix} y \right\|_2 \tag{41}$$

which can be handled efficiently using the QR factorization of $\begin{pmatrix} B_k \\ \lambda I_k \end{pmatrix}$. As for $f_\lambda^{(k)}$, it can be computed using (40). Alternatively, as shown by Paige and Saunders [24], for fixed λ the regularized solution $f_\lambda^{(k)}$ can be computed through an updating formula which does not require any storage of V_k. Note that due to (35) and (36), the residual vector $r_\lambda^{(k)} = \tilde{g} - \tilde{A} f_\lambda^{(k)}$ and the regularized solution $f_\lambda^{(k)}$ satisfy

$$\|r_\lambda^{(k)}\| = \|U_{k+1}^T (\beta_1 e_1 - B_k y_\lambda^{(k)})\| = \|\beta_1 e_1 - B_k y_\lambda^{(k)}\|, \tag{42}$$

$$\|f_\lambda^{(k)}\| = \|y_\lambda^{(k)}\|. \tag{43}$$

In addition, if we let the singular value decomposition of B_k be

$$B_k = P_k \begin{pmatrix} \Sigma_k \\ 0 \end{pmatrix} Q_k^T = \sum_{i=1}^k \sigma_i^{(k)} p_i q_i^T, \tag{44}$$

where $P_k \in \mathbb{R}^{(k+1) \times (k+1)}$ and $Q_k \in \mathbb{R}^{k \times k}$ are orthogonal, and $\Sigma_k = \mathrm{diag}(\sigma_1^{(k)}, \ldots, \sigma_k^{(k)})$ with $\sigma_1^{(k)} \geq \sigma_2^{(k)} \geq \cdots \geq \sigma_k^{(k)} > 0$. Then it is immediate to check that

$$\|y_\lambda^{(k)}\|_2^2 = \beta_1^2 \sum_{i=1}^k \frac{[\sigma_i^{(k)}]^2 \xi_{1i}^2}{([\sigma_i^{(k)}]^2 + \lambda^2)^2}, \qquad \|r_\lambda^{(k)}\|_2^2 = \beta_1^2 \sum_{i=1}^k \frac{\lambda^4 \xi_{1i}^2}{([\sigma_i^{(k)}]^2 + \lambda^2)^2} + \delta_0^{(k)2} \tag{45}$$

where ξ_{1i} denotes the i-th component of the first row of matrix P_k, and $\delta_0^{(k)}$ is the 2-norm of the incompatible part of $\beta_1 e_1$ that lies outside $\mathcal{R}(B_k)$. Moreover, analogous to the squared residual norm $\mathbf{x}(\lambda)$ and the squared solution norm $\mathbf{y}(\lambda)$, $\|r_\lambda^{(k)}\|_2^2$ is a monotonically increasing function and $\|y_\lambda^{(k)}\|_2^2$ is a monotonically decreasing function.

The following technical result will be used in the sequel.

Theorem 4.1 *For fixed $\lambda > 0$ the norm of the solution $f_\lambda^{(k)}$ and corresponding norm of the residual vector $r_\lambda^{(k)}$ satisfy*

$$\|f_\lambda^{(k+1)}\| \geq \|f_\lambda^{(k)}\|, \qquad \|r_\lambda^{(k+1)}\| \leq \|r_\lambda^{(k)}\|, \quad k = 0, \ldots, n-1. \tag{46}$$

Proof See [4, Theorem 2.1].

4.2 Proposed Method

In order to introduce our projection method we shall assume that the GKB procedure is used to compute the approximate solutions $f_\lambda^{(k)}$ and the corresponding residuals $r_\lambda^{(k)}$. Based on these functions let us consider the finite sequence of functions $\vartheta^{(k)} : \mathbb{R}^+ \to \mathbb{R}, k = 1, \ldots, n$, defined by

$$\vartheta^{(k)}(\lambda) = \frac{\rho^{(k)}(\lambda)}{\delta_g + \delta_A \eta^{(k)}(\lambda)}, \tag{47}$$

where $\rho^{(k)}(\lambda) = \|r_\lambda^{(k)}\|_2$, and $\eta^{(k)}(\lambda) = \|y_\lambda^{(k)}\|_2$, and the finite sequence of functions $\zeta^{(k)} : \mathbb{R}^+ \to \mathbb{R}$, $k = 1, \ldots, n$, defined by

$$\zeta^{(k)}(\lambda) = \frac{\lambda}{\sqrt{\vartheta^{(k)}(\lambda)}}. \tag{48}$$

Since $\rho^{(k)}$ and $\eta^{(k)}$ are monotonic, some calculations show that $\vartheta^{(k)'}(\lambda) > 0$ and $\zeta^{(k)'}(\lambda) > 0$. Therefore, both $\vartheta^{(k)}$ and $\zeta^{(k)}$ are monotonically increasing functions. Our projected method relies on the fact that we can construct a sequence of fixed-points $\lambda^{(k)*}$ of $\zeta^{(k)}$ which approximates λ_{GDP} as k increases. For this, due to the monotonicity of $\vartheta^{(k)}$ and $\zeta^{(k)}$ it is clear that $\lambda^{(k)*}$ is a fixed-point of $\zeta^{(k)}$ if and only if $\lambda^{(k)*}$ is the unique solution of the "projected discrepancy equation" $\vartheta^{(k)}(\lambda) = 1$.

The following theorem describes how the sequence $\zeta^{(k)}$ relates to the function ζ of the original and large-scale problem and shows that $\zeta^{(k)}$ always has a unique nonzero fixed-point.

Theorem 4.2 *Assume that exact arithmetic is used and that the GKB procedure runs n steps without interruption. Then for all $\lambda > 0$ it holds*

$$\zeta^{(k+1)}(\lambda) \geq \zeta^{(k)}(\lambda), \quad k = 1, \ldots, n-1. \tag{49}$$

Consequently, $\zeta(\lambda) \geq \zeta^{(k)}(\lambda)$ for $k = 1, \ldots, n$. Additionally, if the condition (22) is satisfied, then $\zeta^{(k)}$ always has a unique nonzero fixed-point that is the unique solution of the nonlinear equation (6).

Proof The first inequality is straightforward from Theorem 3.2. The second inequality follows upon taking $k = n - 1$ in (49). To prove the last assertion we shall analyze the sign of the function $h^{(k)}(\lambda) = \zeta^{(k)}(\lambda) - \lambda$ for $\lambda > 0$. In fact, if (22) is satisfied, then λ_{GDP} is a nonzero fixed-point of ζ due to Theorem (3.2), and thus we must have $\zeta^{(k)}(\lambda) < \lambda$ when $\lambda > \lambda_{\text{GDP}}$ as we have seen that $\zeta(\lambda) \geq \zeta^{(k)}(\lambda)$ for $\lambda > 0$. Therefore, for $\lambda > \lambda_{\text{GDP}}$ we have

$$h^{(k)}(\lambda) < 0. \tag{50}$$

Now notice that the behavior of h near the origin depends on the constant $\delta_0^{(k)} \neq 0$, see (45) and (48). Notice also that by Lemma 3.1 we have $\lim_{\lambda \to 0^+} h^{(k)}(\lambda) = 0$ if $\delta_0^{(k)} \neq 0$ and $\lim_{\lambda \to 0^+} h^{(k)}(\lambda) > 0$ if $\delta_0^{(k)} = 0$. This tells us that only the case $\delta_0^{(k)} \neq 0$ must be considered when analyzing the function $h^{(k)}$ for λ near zero and that this analysis requires information about the derivative of h. We then use (28) and obtain

$$h^{(k)'}(0) = \zeta^{(k)'}(0) - 1 = \sqrt{\frac{\delta_g + \delta_A \|f_\lambda^{(k)}\|_2}{\delta_0^{(k)}}} - 1 > 0, \tag{51}$$

the inequality being valid as $\delta_0^{(k)} \leq \delta_g$. But this implies that $\zeta^{(k)'}(0) > 1$. This inequality implies that there must be an open interval, say $I = (0, \bar{\lambda}) \subset (0, \lambda_{\text{GDP}})$, such that $\zeta^{(k)}(\lambda) > \lambda$ for all $\lambda \in I$, and thus we must have

$$h^{(k)}(\lambda) > 0 \quad \forall \lambda \in I. \tag{52}$$

We now use (50)–(52) and conclude that $h^{(k)}$ has a root $\lambda^{(k)*}$, or equivalently that $\lambda^{(k)*}$ is a nonzero fixed-point of $\zeta^{(k)}$. Uniqueness of $\lambda^{(k)*}$ follows from the monotonicity properties of $\zeta^{(k)}$ and $\vartheta^{(k)}$, and this concludes the proof. $\square$

Fig. 2 Functions ζ and $\zeta^{(k)}$ for some values of k. This example corresponds to deriv2 test problem from [14] with the same data as in Fig. 1
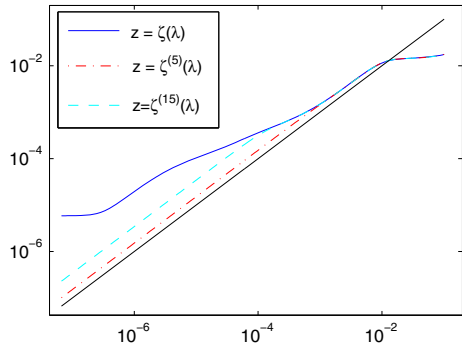
To better explain the relevance of Theorem 4.2, it is instructive to note that as the GKB algorithm proceeds, more and more information associated with the largest singular values of $\widetilde{A}$ is captured [6,15]. As a consequence, the sequence $\zeta^{(k)}(\lambda)$ will quickly capture the information of $\zeta(\lambda)$ for a range of values of λ inside the part of the singular values of A that is captured by the GKB procedure in k steps. This is illustrated in Fig. 2 where are depicted some functions $\zeta^{(k)}$ as well as its limit value $\zeta(\lambda)$ for the deriv2 test problem. Notice that in a wide range of λ values which include the parameter λ_{GDP}, the functions $\zeta^{(k)}$ remain remarkably close to ζ even for small k.

Now for given $k \geq 1$ and arbitrarily chosen initial guess $\lambda_0^{(k)} > 0$, consider the sequence $\{\lambda_j^{(k)}\}$ defined by

$$\lambda_{j+1}^{(k)} = \zeta^{(k)}(\lambda_j^{(k)}), \tag{53}$$

and assume it converges to a fixed-point $\lambda^{(k)*}$ of $\zeta^{(k)}$; when this is true we prove that the sequence $\lambda^{(k)*}$ converges to λ_{GDP} from below in at most n GKB steps. This is the subject of the following theorem.

Theorem 4.3 *Under the hypotheses of Theorem 4.2, for $k \geq 1$ let $\lambda^{(k)*}$ be the unique nonzero fixed-point of $\zeta^{(k)}$. Then the sequence of fixed-points $\{\lambda^{(k)*}\}_k$ is nondecreasing and $\lambda^{(k)*}$ converges to λ_{GDP} in at most n GKB steps.*

Proof Let $\lambda^{(k)*}$ be a fixed-point of $\zeta^{(k)}(\lambda)$. Due to Theorem 4.2 we have

$$\zeta^{(k+1)}(\lambda^{(k)*}) \geq \zeta^{(k)}(\lambda^{(k)*}) = \lambda^{(k)*}. \tag{54}$$

If $\lambda^{(k)*}$ is also a fixed-point of $\zeta^{(k+1)}$ there is nothing to prove. Assume then that $\zeta^{(k+1)}(\lambda^{(k)*}) > \lambda^{(k)*}$ and consider the sequence $\lambda_{j+1} = \zeta^{(k+1)}(\lambda_j)$, $j \geq 0$, with starting value $\lambda_0 = \lambda^{(k)*}$. Based on the fact that $\zeta^{(k+1)}$ increases as λ increases, it follows that λ_j forms a nondecreasing sequence, and therefore $\{\lambda_j\}$ converges to a fixed-point of $\zeta^{(k+1)}$, i.e.,

$$\lim_{j \to \infty} \lambda_j = \lambda^{(k+1)*} = \zeta(\lambda^{(k+1)*})$$

with $\lambda^{(k+1)*} \geq \lambda^{(k)*}$. Now since after n GKB steps the Krylov space equals $\mathbb{R}^n$, it follows that $\zeta^{(n)}$ equals ζ and so $\{\lambda^{(k)*}\}$ must converge to λ_{GDP} in at most n GKB steps, and the proof concludes.   □

The main consequence of Theorem 4.3 is that we can construct approximations to λ_{GDP} using a nondecreasing finite sequence of fixed-points $\lambda^{(k)*}$ of $\zeta^{(k)}$, which as illustrated above can converge very quickly due to the approximation properties of the Krylov subspace generated by the GKB procedure. From the practical point of view, this gives rise to an algorithm for computing λ_{GDP} and the corresponding regularized solution which we will refer to as PGDP-FP. The main steps of PGDP-FP can be summarized as follows:

PGDP-FP :

Input: $\widetilde{A}, \widetilde{g}, p > 1, \epsilon$.

Output: Regularized solution $f_{\lambda^*}^{(k)}$

1. Apply p GKB steps to $\widetilde{A}$ with starting vector $\widetilde{g}$ and form the matrix B_p.

2. Set $k = p$. Compute the fixed-point $\lambda^{(k)*}$ of $\zeta^{(k)}$ and set
$\lambda_0 = \lambda^{(k)*}, \lambda_{\text{old}} = \lambda_0 , k \leftarrow k + 1$.

3. Perform one more GKB step and compute the fixed-point
$\lambda^{(k)*}$ of $\zeta^{(k)}$ taking λ_0 as starting value.
Set $\lambda_{\text{old}} = \lambda_0, \lambda_0 = \lambda^{(k)*}$.

4. **If** stopping criterion is satisfied **do**
$\lambda^* = \lambda_{\text{old}}$
else do
$k \leftarrow k + 1$
Go to step **3.**
end if

5. Compute the regularized solution $f_{\lambda^*}^{(k)}$

We will refer to each GKB iteration as an *outer iteration* and to iterations during fixed computations for each k as *inner iterations*. For practical purposes, we note that computing fixed-points for each k requires solving the projected problem (41) for several values of λ and for increasing values of k. This can be done following the ideas of the LSQR algorithm at a cost of approximately $\mathcal{O}(k)$ arithmetic operations [10,25]. Moreover, since very often a fairly small number of steps is required for convergence, the overall cost of PGDP-FP is dominated by k matrix-vector products with $\widetilde{A}$ and $\widetilde{A}^T$ for some moderate number k. In addition, to introduce savings for the algorithm to be efficient and competitive, the fixed-point $\lambda^{(k)*}$ at stage k can be used as initial guess for computing the fixed-point at stage $(k+1)$ since, as demonstrated before, the sequence of fixed-points $\lambda^{(k)*}$ is nondecreasing. Taking into account the monotonicity of $\lambda^{(k)*}$, we choose to stop the outer iterations when the relative change of consecutive fixed-points,

$$\left| \lambda^{(k+1)*} - \lambda^{(k)*} \right| < \epsilon \lambda^{(k)*}, \tag{55}$$

is sufficiently small, where ϵ is a tolerance parameter.

4.3 Extension to General Form Tikhonov Regularization

For some problems, standard-form Tikhonov regularization does not produce suitable regularized solutions and it is convenient to consider instead general-form Tikhonov regularized solutions. The purpose of this subsection is to briefly describe two approaches to extend GDP-FP to general-form Tikhonov regularization: one approach which implicitly transforms the general-form Tikhonov problem (3) into a related problem in standard form, and another approach where such transformation is not necessary.

For the first approach, note that the Tikhonov problem (3) can always be transformed into a Tikhonov problem in standard form [10]

$$\overline{f}_\lambda = \underset{\overline{f} \in \mathbb{R}^n}{\operatorname{argmin}} \|\overline{A}\,\overline{f} - \overline{g}\|_2^2 + \lambda^2 \|\overline{f}\|_2^2. \tag{56}$$

For the case where L is invertible the standard-form transformation becomes $\overline{A} = \widetilde{A}L^{-1}$, $\overline{g} = \widetilde{g}$, and the back-transformation is $f_\lambda = L^{-1}\overline{f}_\lambda$.

When L is not invertible the transformation takes the form

$$f_\lambda = L_{\widetilde{A}}^\dagger \overline{f}_\lambda + f_N, \quad \overline{A} = \widetilde{A}L_{\widetilde{A}}^\dagger, \quad \overline{g} = \widetilde{g} - \widetilde{A}f_N, \tag{57}$$

where f_N lies in the null space of L and

$$L_{\widetilde{A}}^\dagger = \left(I_n - \left(\widetilde{A}\left(I_n - L^\dagger L \right) \right)^\dagger \widetilde{A} \right) L^\dagger,$$

is the $\widetilde{A}$-weighted generalized inverse of L. If we know a full rank matrix W whose columns span the null space $\mathcal{N}(L)$, then $f_N = W\left(\widetilde{A}W\right)^\dagger \widetilde{g}$, and the $\widetilde{A}$-weighted generalized inverse of L reduces to $L_{\widetilde{A}}^\dagger = \left(I_n - W\left(\widetilde{A}W\right)^\dagger \widetilde{A} \right) L^\dagger$.

Therefore our first extension of GDP-FP to large-scale general-form Tikhonov regularization is to apply PGDP-FP to the transformed problem (56) assuming that the transformed matrix $\overline{A}$ is never explicitly formed. Once $\overline{f}_\lambda$ is determined, the regularized solution f_λ is computed by using the back-transformation (57). This approach is well suited when the matrix-vector products with $\overline{A}$ and $\overline{A}^T$ are performed efficiently. That is, this approach works well when the computation of matrix-vector products with $L_{\widetilde{A}}^\dagger$ is feasible and performed efficiently, as often occurs in applications involving sparse regularization matrices e.g., banded regularization matrices with small bandwidth and regularization matrices defined via Kronecker products, among others. For examples and situations where the transformation to standard form is feasible, the reader is referred to [5,26].

Our second approach considers the case where L is a regularization matrix such that the computation of matrix-vector products with $L_{\widetilde{A}}^\dagger$ is infeasible or unattractive. We follow an approach suggested by Hochstenbach and Reichel [16] recently exploited in connection with an extension of GKB-FP, an algorithm for large-scale general-form Tiknonov problems discussed in [5]. The main idea here is to determine regularized solutions as defined in (34) where the subspace $\mathcal{V}_k$ is chosen as the Krylov subspace $\mathcal{K}_k(\widetilde{A}^T\widetilde{A}, \widetilde{A}^T\widetilde{g})$ generated by GKB and the regularization parameter is chosen by GDP. In other words, we apply the GKB procedure to $\widetilde{A}$, which produces the partial decomposition (35)–(37), and compute regularized solutions as

$$f_\lambda^{(k)} = V_k y_\lambda^{(k)}, \quad y_\lambda^{(k)} = \underset{y \in \mathbb{R}^k}{\operatorname{argmin}}\{\|\widetilde{A}V_k y - \widetilde{g}\|_2^2 + \lambda^2 \|LV_k y\|_2^2\}, \tag{58}$$

where the regularization λ is determined as the unique solution of the nonlinear equation (6). Note that if we use the QR factorization of the product LV_k, $LV_k = Q_k R_k$, based on (35), (36) the above least squares problem reduces to one of the form

$$y_\lambda^{(k)} = \underset{y \in \mathbb{R}^k}{\operatorname{argmin}}\{\|B_k y - \beta_1 e_1\|_2^2 + \lambda^2 \|R_k y\|_2^2\} \tag{59}$$

which can be computed efficiently in several ways, e.g., by a direct method or by first transforming the stacked matrix $[B_k^T \ \lambda R_k^T]^T$ to upper triangular form, as done when implementing GKB-FP [4]. We can now compute the functions $\vartheta^{(k)}$ and $\zeta^{(k)}$ as defined in (47), (48) and

then follow the same ideas as PGKB-FP. That is, for chosen $p > 1$ and $k \geq p$, we determine the solution $\lambda^{(k)*}$ of the nonlinear function $\vartheta^{(k)}(\lambda) = 1$, repeating the process until the stopping criterion (55) is satisfied.

In practice the QR factorization $L V_k = Q_k R_k$ is calculated only once at step $k = p$ and updated in subsequent steps. Algorithms for updating the QR factorization can be found in [13, Chapter 12].

5 Numerical Examples

We illustrate the effectiveness of GDP-FP and PGDP-FP considering two test problems involving first kind integral equations

$$\int_0^1 K(s, t)\mathbf{f}(t)dt = \mathbf{g}(s), \quad 0 \leq s \leq 1, \tag{60}$$

which have been discussed in [21], and a third test problem from inverse scattering. Our computations were carried out in Matlab with about 16 significant decimal digits. Since the special case $\delta_A = 0$ has been discussed and illustrated in several papers, we will only concentrate here on the case $\delta_A \neq 0$.

Example 1:

In this test problem the kernel $K(s, t)$ and the functions $\mathbf{f}$ and $\mathbf{g}$ are defined by

$$K(s, t) = \begin{cases} s(1 - t) & \text{for} \quad s \leq t, \\ t(1 - s) & \text{for} \quad s \geq t, \end{cases}, \quad \mathbf{f}(t) = 4t(t - 1), \quad \mathbf{g}(s) = \frac{s}{3}(s^3 - 2s^2 + 1).$$

Discretization of (60) leads to a discrete ill-posed problem where A, f, and g depend on how the approximation is made. In this example the matrix $A \in \mathbb{R}^{n \times n}$ is generated by using the Matlab function deriv2 from [14], the vector f has coordinates $f_j = \sqrt{\Delta t}(t_j - \Delta t/2)$, where $t_j = j\Delta t$, $\Delta t = 1/n$, and $g = Af$. To simulate noisy data we take $\widetilde{A} = A + E$ and $\widetilde{g} = g + e$, where e and E are a scaled Gaussian random vector and a scaled Gaussian random matrix, respectively, such that for chosen ϵ_A and ϵ_g we have $\|E\|_2/\|A\|_2 = \epsilon_A$ and $\|e\|_2/\|g\|_2 = \epsilon_g$. In this example we take $\epsilon_A = \epsilon_g = 0.03$. GDP-FP is implemented by using the SVD of A, which means we choose $L = I$, and PGDP-FP is implemented with complete reorthogonalization, taking $\epsilon = 10^{-5}$ and $p = 3$. To simulate a medium sized problem we choose $n = 1,200$.

Figure 3 displays the results obtained with GDP-FP and PGDP-FP, as well as the results obtained with the L-curve method (LC) [15] which are included here for comparison purposes. From the results we see that LC may not be a good option in cases where both the matrix A and the vector b are noise-corrupted, and that the solutions produced by PGDP-FP and by GDP-FP, respectively, are indistinguishable (albeit slightly oscillating), as seen in Fig. 3 (second row). We also report important numbers such as the computed regularization parameters, relative errors associated with each regularized solution, etc., see Table 1. This table shows that GDP-FP converged in $j = 6$ iterations and that PGDP-FP spent $k = 4$ outer iterations, i.e., the regularized solution determined with PGDP-FP lives in a Krylov subspace of dimension $k = 4$.

We now show that the small oscillations in the regularized solution can be suppressed provided that a suitable regularization matrix is chosen. To this end we follow [21] and choose an invertible regularization matrix defined by $L = \frac{(n+1)}{\pi} T^{1/2}$ with
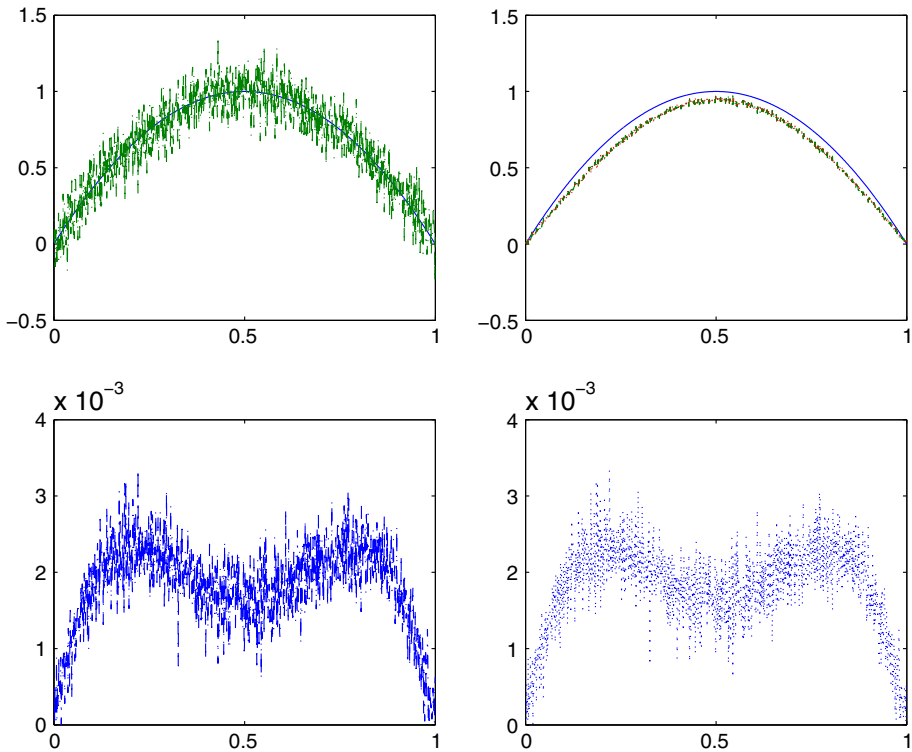
Fig. 3 First row (*left*): solution produced by L-curve. First row (*right*): Exact solution f (*solid line*), solution f_λ determined by GDP-FP (*dash-dotted line*) and solution $f_\lambda^{(4)}$ determined by PGDP-FP (*dotted line*). Second row: relative error in f_λ (*left*) and relative error in $f_\lambda^{(4)}$ (*right*)

Table 1 Index j refers to the number of iterates spent by GDP-FP, and index k refers to the dimension of the Krylov subspace used to approximate λ_{GDP}

	LC	GDP-FP	PGDP-FP
λ	0.005457603	0.0303640471	0.0303640471
$\|f - f_\lambda\|/\|f\|$	0.1529722780	0.08972004460	0.0897200529
j	–	6	–
k	–	–	4

$$
T = \begin{pmatrix} 2 & -1 & & \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 2 \end{pmatrix}. \tag{61}
$$

With this choice of L we transform the general-form Tikhonov problem (3) into a problem in standard form, as described in Sect. 4.3, and then determine the solution of the transformed problem via PGDP-FP. The solution $f_\lambda^{(k)}$ obtained with PGDP-FP is then recovered using

Fig. 4 f (*solid line*) and $f_\lambda^{(5)}$ (*dotted line*) with the choice $L = \frac{n+1}{\pi} T^{1/2}$. In this case we obtain $\lambda = 0.0096$, and $\|f - f_\lambda^{(5)}\| / \|f\|_2 = 0.0339$
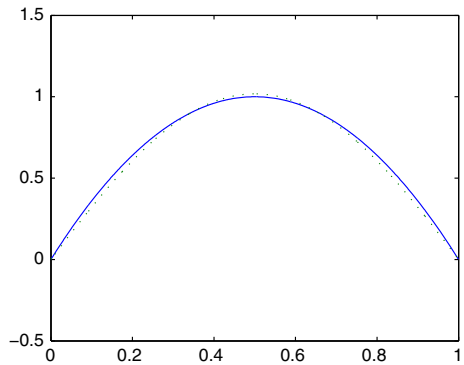
the corresponding back-transformation. Figure 4 displays results for the case $n = 1,200$ and $\epsilon_A = \epsilon_g = 0.003$. In this case PGDP-FP performed $k = 5$ outer iterations. Note that the choice of L yields a smooth regularized solution.

Example 2:

We now consider the first kind integral equation (60) with the same kernel $K(s, t)$ as in Example 1 but with the following two choices for **f** and **g**

(a) $\mathbf{f}(t) = t$, $\quad \mathbf{g}(s) = \frac{s}{6}(s^2 - 1)$.
(b) $\mathbf{f}(t) = e^t$, $\quad \mathbf{g}(s) = e^s + (1 - e)s - 1$

Although in many problems the matrix L is chosen as a scaled discrete differential operator, the purpose of the above test problems is to illustrate that such a choice may not be appropriate if the operator is not properly scaled. To this end we choose again $n = 1,200$ and use the same noise level as in the previous example.

For the test problem with **f** and **g** defined in **(a)** we consider one case where $L = I$ and two additional cases where $L \neq I$. Figure 5 (first row) displays the regularized solutions obtained with GDP-FP and PGDP-FP for the first case, as well as the relative errors associated with each solution. In this case both GDP-FP and PGDP-FP are implemented as in Example 1. The solution produced by PGDP-FP lives in a Krylov subspace of dimension $k = 5$ and remains very close to that produced by GDP-FP, as it is clearly visible in Fig. 5, first row (left). However, the large relative errors in both solutions indicate that the choice $L = I$ is unsatisfactory and motivates the use of other regularization matrices.

The second part of the example considers two choices for the regularization matrix: $L = \frac{n+1}{\pi} T^{1/2}$ as defined in the previous example, and $L = L_2$ where

$$L_2 = \begin{pmatrix} -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \end{pmatrix} \tag{62}$$

is a rectangular matrix of size $(n - 2) \times n$.

For the case where L is invertible we use the extension of GDP-FP to general-form Tikhonov regularization considered in Example 1, while for the case where L is rectangular, we use the extension of GDP-FP which performs computations with the $\widetilde{A}$-weighted pseudoinverse of L, as described in Sect. 4.3, taking advantage of the fact that L has a known null subspace.

The computed regularized solutions f_λ (obtained with the GSVD of the matrix pair $(\widetilde{A}, L)$) and corresponding relative errors are displayed in Fig. 5, second row. The solutions obtained
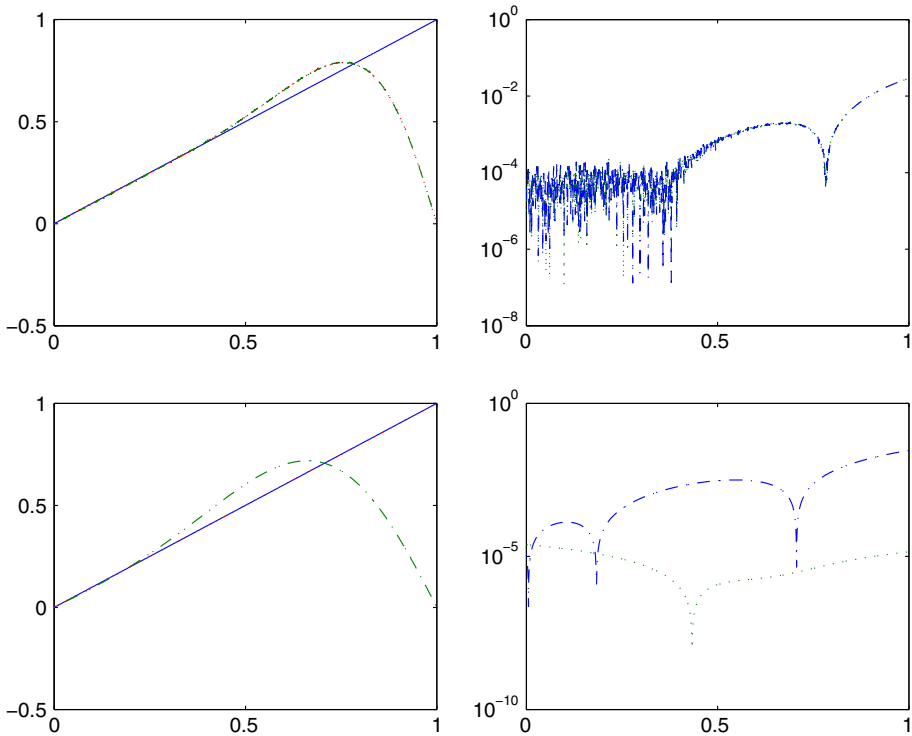
Fig. 5 First row (*left*): Exact solution f (*solid line*), f_λ obtained with GDP-FP (*dash-dotted line*) and $f_\lambda^{(5)}$ obtained with PGDP-FP (*dotted line*). First row (*right*): relative error in f_λ and relative error in $f_\lambda^{(5)}$. Second row (*left*): f (*solid line*), f_λ for $L = \frac{n+1}{\pi} T^{1/2}$ (*dash-dotted line*) and f_λ for $L = L_2$ (*dotted line*). Second row (*right*): relative errors associated with the solutions obtained with $L = \frac{n+1}{\pi} T^{1/2}$ (*dash-dotted line*) and $L = L_2$ (dotted line), respectively

with the projection approach are very similar to those obtained with the GSVD and therefore are not shown here. Note that the results obtained with L invertible are very similar to those obtained with $L = I$ (same figure, first row) and therefore of poor quality. This does not occur with the regularized solution obtained with $L = L_2$ whose quality is apparent.

We now describe results for the test problem with **f** and **g** defined in **(b)**. For this problem, in addition to the choices for L used in the test problem with data in **(a)**, we consider the choice $L = D$, where D is a rank-deficient square matrix defined by

$$D = \beta (L_2^T L_2)^{1/4}, \quad \beta > 0, \tag{63}$$

with scaling factor $\beta = (n+1)/\pi$. This choice is motivated by a theoretical result by Lu et al. [21] that, for invertible $\mathcal{L}$, suggests choosing a scaling factor β such that $L = \beta \mathcal{L}$ satisfies $\|L^{-1}\|_2 \leq 1$, see Theorem 2.1 in [21], and justifies the choice of the square regularization matrix L in Example 1 defined via the matrix $T^{1/2}$. In order to mimic this result inverse is replaced by Moore-Penrose pseudo inverse and the scaling factor β is chosen so that $\|D^\dagger\|_2 \leq 1$. The computed regularized solutions f_λ obtained with the GSVD of the matrix pair $(\tilde{A}, L)$ are displayed in Fig. 6. Notice that, contrary to the excellent results obtained with the regularization matrix $L = L_2$ in Example 1, the results in this case are of poor
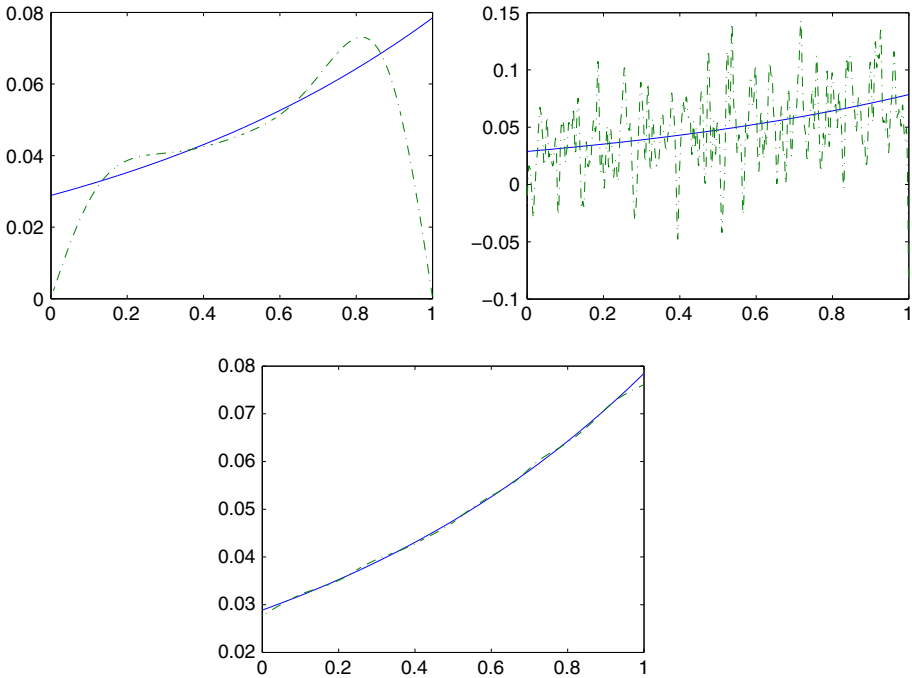
Fig. 6 Exact solution (*solid line*) and regularized solution f_λ obtained with GDP-FP (*dash-dotted line*). First row (*left*): results for $L = \frac{n+1}{\pi} T^{1/2}$. First row (*right*): results for $L = L_2$. Second row: results for $L = \mathsf{D}$. For this this example, the computed regularization parameters are $\lambda = 0.00093$, 0.00097 and 0.00019, while the corresponding relative errors are $\|f - f_\lambda\|_2 / \|f\|_2 = 0.3111$, 0.6030 and 0.0091, respectively

quality; a similar observation applies to the results obtained with the square and invertible regularization matrix. On the other hand, notice that the quality of the regularized solution obtained with $L = \mathsf{D}$ is excellent.

Example 3: Inverse scattering

The propagation of time-harmonic acoustic fields in a homogeneous medium, in the presence of a sound soft obstacle D, is modeled by the exterior boundary value problem (direct obstacle scattering problem) [8]

$$\begin{cases} \triangle u(x) + k^2 u(x) = 0, & x \in \mathbb{R}^2 \setminus \overline{D} \\ u(x) \doteq u^s + u^i(x) = 0, & x \in \partial D \\ (\frac{\partial u^s}{\partial r} - iku^s) = O(\frac{1}{r}), & r = |x| \to \infty \end{cases}$$

where k is a real positive wavenumber, u^i is an incident wave, and u^s is the scattered wave outside D. In addition, u^s has the following asymptotic behavior

$$u^s(x) = u_\infty(\hat{x}) \frac{e^{ikr}}{\sqrt{r}} + O(r^{-3/2}), \tag{64}$$

where u_∞ is defined on the unit circle Ω and called the far field-pattern of the scattered wave. Our interest here is to identify the shape of D from measurements of u_∞ by using a variant

of the linear sampling method (LSM) [9]. The basis of LSM is the far-field equation

$$(F\,w_z)(\hat{x}) = \frac{e^{i\pi/4}}{\sqrt{8\pi k}}e^{-ik\hat{x}\cdot z}, \quad \hat{x} = x/|x|, \ \ z \in \mathbb{R}^2, \tag{65}$$

where $F : L^2(\Omega) \to L^2(\Omega)$ is a linear compact operator, the Far-field operator, given by

$$(F\,w)(\hat{x}) = \int_\Omega u_\infty(\hat{x}; \hat{d})\,w(\hat{d})\,ds(\hat{d}), \quad \hat{d} \in \Omega. \tag{66}$$

LSM is based on the the observation that the solution g_z of (65) has a large norm for z outside and close to ∂D. Hence, characterizations of D are obtained by plotting the norm $\|w_z\|$. However, a difficulty with LSM is that (65) is not always solvable. Kirsch [17] was able to overcome this difficulty with the introduction of a variant of LSM based on the following linear equation

$$(F^\star F)^{1/4}w_z = \frac{e^{i\pi/4}}{\sqrt{8\pi k}}e^{-ik\hat{x}\cdot z}. \tag{67}$$

In this variant, the shape of D is characterized analogously as in LSM, i.e., by inspecting for which z the norm $\|g_z\|$ is large. In practice, Kirsch's method deals with a finite dimensional counterpart of (67) obtained by replacing F for a perturbed finite-dimensional far-field operator of the form $\widetilde{F}_d = F_d + E \in \mathbb{C}^{n\times n}$, where n denotes the number of observed incident waves and the number of outgoing directions, and F_d has singular values decaying quickly to zero, which means F_d is severely ill-conditioned. Hence, regularization is needed in order to compute stable approximations to the solution w_z of (67). Following [17], we approximate w_z by using a regularized solution $w_{\lambda,z}$ constructed by the method of Tikhonov:

$$w_{\lambda,z} = \underset{w\in\mathbb{C}^n}{\mathrm{argmin}}\{\|r_z - \widetilde{A}_d\,w\|_2^2 + \lambda^2\|w\|_2^2\} \tag{68}$$

where the regularization parameter is chosen by the GDP. Here, r_z denotes the discrete right hand side of (67) and $\widetilde{A}_d = (\widetilde{F}_d^*\widetilde{F}_d^*)^{1/4}$. Regarding GDP, note that the discrepancy function in this case becomes

$$G_z(\lambda) = \|\widetilde{A}_d\,w_{\lambda,z} - r_z\|_2^2 - \delta_A^2\|w_{\lambda,z}\|_2^2,$$

which, in terms of the SVD of $\widetilde{F}_d$ reads

$$G_z(\lambda) = \sum_{i=1}^n \frac{\lambda^4 - \delta_A^2\sigma_i}{(\sigma_i + \lambda^2)^2}\alpha_i \tag{69}$$

where σ_i denote the singular values de $\widetilde{F}_d$ and $\alpha_i = |v_i^*r_z|^2$, the squared Fourier coefficients of r_z with respect to the singular vectors of $\widetilde{F}_d$.

Summarizing, the characterization of the object can be done as follows:

- For each z in a grid containing the object D, solve the Tikhonov problem (68) with the regularization parameter being chosen by GDP.
- Characterize the object by using a plot of the map $z \to W(z) = 1/\|w_{\lambda,z}\|_2^2$.

To illustrate the efficiency of GDP-FP in this class of problems, we choose a test problem which has been considered in [3, 19]. In this case the object to be characterized is a kite located in a grid of 50×50 points, the far field matrix $\widetilde{F}_d$ is 32×32 (i.e., we use 32 incident observed directions), and the relative noise level in $\widetilde{F}_d$ is 5 % (which implied a relative noise level in $\widetilde{A}_d$
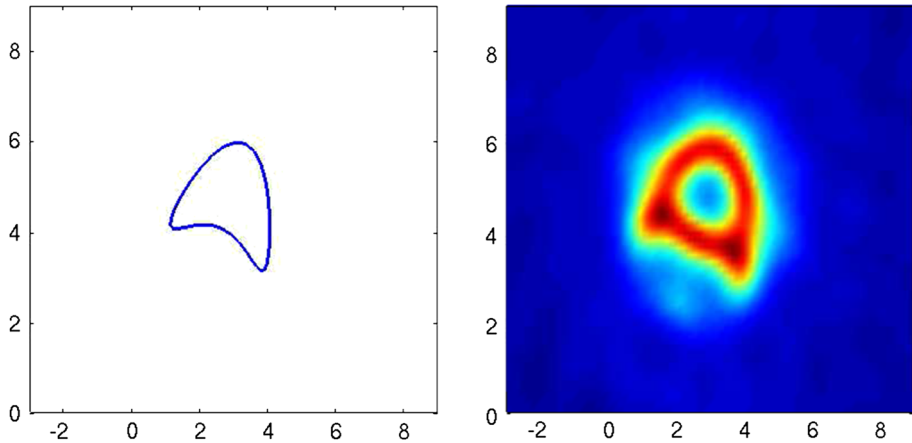
Fig. 7 Kite and its reconstruction via Kirsch's method with GDP as parameter choice rule

of approximately 15 %). To simulate perturbed data, we generate Gaussian random matrices E_1, E_2 and use a far-field matrix defined by

$$\widetilde{F}_d = F_d + \delta(E_1 + i\, E_2)\|F_d\|_2, \quad \delta > 0,$$

for given δ, where F_d is constructed by using the Nyström method [8].

To assess the efficiency of GDP-FP when compared to other method used to solve the scattering inverse problem, we also compute the regularization parameter by using *regula falsi* method as done in [17]. The tolerance parameter for stopping both GDP-FP and regula falsi is set to $\epsilon = 10^{-5}$. For this example, the reconstruction obtained with GDP-FP is about 12 times faster than the one obtained with regula falsi. The profile of the kite and its reconstructed version are displayed in Fig. 7.

6 Conclusions

We derived a fixed-point algorithm for determining the Tikhonov regularization parameter chosen by the GDP for discrete ill-posed problems. The algorithm is monotonically and globally convergent and easy to implement using the GSVD of the coefficient matrix. A special version of the algorithm based on the Golub-Kahan bidiagonalization procedure, that is well suited for large-scale problem, is also developed and discussed in detail. The algorithms are numerically illustrated using three test problems from the Regularization Tools [14] and an additional test problem from inverse scattering [9]. In addition, the issue of selecting a proper scaling in connection with discrete differential operators was also illustrated. The numerical results indicate that the algorithms are quite fast and therefore promising for discrete ill-posed problems with noisy operator and noisy data.

References

1. Ang, D.D., Gorenflo, R., Le, V.K., Trong, D.D.: Moment Theory and Some Inverse Problems in Potential Theory and Heat Conduction, Lect. Notes Math. 1792. Springer, Berlin (2002)

2. Bazán, F.S.V.: Fixed-point iterations in determining the Tikhonov regularization parameter. Inverse Probl. **24**, 1–15 (2008)
3. Bazán, F.S.V., Francisco, J.B., Leem, K.H., Pelekanos, G.: A maximum product criterion as a Tikhonov parameter choice rule for Kirsch's factorization method. J. Comput. Appl. Math. **236**, 4264–4275 (2012)
4. Bazán, F.S.V., Borges, L.S.: GKB-FP:an algorithm for large-scale discrete ill-posed problems. BIT **50**(3), 481–507 (2010)
5. Bazán, F.S.V., Cunha, M.C., Borges, L.S.: Extension of GKB-FP to large-scale general-form Tikhonov regularization. Numer. Linear Algebra Appl. **21**, 316–339 (2014)
6. Björck, Å.: A bidiagonalization algorithm for solving ill-posed systems of linear equations. BIT **28**, 659–670 (1988)
7. Chung, J., Nagy, J.G., O'Leary, D.P.: A weighted-GCV method for Lanczos-hybrid regularization. Electron. Trans. Numer. Anal. **28**, 149–167 (2008)
8. Colton, D., Kress, R.: Inverse Acoustic and Electromagnetic Scattering Theory. Springer, New York (1992)
9. Colton, D., Piana, M., Potthast, R.: A simple method using Morozov's discrepancy principle for solving inverse scattering problems. Inverse Probl. **13**, 1477–1493 (1999)
10. Eldén, L.: A weighted pseudoinverse, generalized singular values, and constrained least square problems. BIT **22**, 487–502 (1982)
11. Engl, H.W., Hanke, M., Neubauer, A.: Regularization of Inverse Problems (Mathematics and its Applications), vol. 375. Kluwer Academic Publishers Group, Dordrecht (1996)
12. Golub, G.H., Hansen, P.C., O'Leary, D.P.: Tikhonov regularization and total least squares. SIAM J. Matrix Anal. Appl. **21**, 185–194 (1999)
13. Golub, G.H., Van Loan, C.F.: Matrix Computations, 3rd edn. The Johns Hopkins University Press, Baltimore (1996)
14. Hansen, P.C.: Regularization tools: a matlab package for analysis and solution of discreet Ill-posed problems. Numer. Algorithms **6**, 1–35 (1994)
15. Hansen, P.C.: Rank-Deficient and Discrete Ill-Posed Problems. SIAM, Philadelphia (1998)
16. Hochstenbach, M.E., Reichel, L.: An iterative method for Tikhonov regularization with a general linear regularization operator. J. Integral Equ. Appl. **22**, 463–480 (2010)
17. Kirsch, A.: Characterization of the shape of a scattering obstacle using the spectral data of the far field operator. Inverse Probl. **14**, 1489–1512 (1998)
18. Lampe, J., Voss, H.: Large-scale Tikhonov regularization of total least squares. J. Comput. Appl. Math. **238**, 95–108 (2013)
19. Leem, K.H., Pelekanos, G., Bazán, F.S.V.: Fixed-point iterations in determining a Tikhonov regularization parameter in Kirsch's factorization method. Appl. Math. Comput. **216**, 3747–3753 (2010)
20. Lu, S., Pereverzev, S.V., Tautenhahn, U.: Regularized total least squares: computational aspects and error bounds. SIAM J. Matrix Anal. Appl. **31**, 918–941 (2009)
21. Lu, S., Pereverzev, S.V., Shao, Y., Tautenhahn, U.: On the generalized discrepancy principle for Tikhonov regularization in Hilbert Scales. J. Integral Equ. Appl. **22**(3), 483–517 (2010)
22. Markovsky, I., Van Huffel, S.: Overview of total least squares methods. Signal Process. **87**, 2283–2302 (2007)
23. Morozov, V.A.: Regularization Methods for Solving Incorrectly Posed Problems. Springer, New York (1984)
24. Paige, C.C., Saunders, M.A.: LSQR: an algorithm for sparse linear equations and sparse least squares. ACM Trans. Math. Softw. **8**(1), 43–71 (1982)
25. Paige, C.C., Saunders, M.A.: Algorithm 583. LSQR: sparse linear equations and least squares problems. ACM Trans. Math. Softw. **8**(2), 195–209 (1982)
26. Reichel, L., Ye, Q.: Simple square smoothing regularization operators. Electron. Trans. Numer. Anal. **33**, 63–83 (2009)
27. Renaut, R.A., Guo, G.: Efficient algorithms for solution of regularized total least squares. SIAM J. Matrix Anal. Appl. **21**, 457–476 (2005)
28. Sima, D., Van Huffel, S., Golub, G.H.: Regularized total least squares based on quadratic eigenvalue problems solvers. BIT **44**, 793–812 (2004)
29. Tautenhahn, U.: Regularization of linear ill-posed problems with noisy right hand side and noisy operator. J. Inverse Ill Posed Probl. **16**, 507–523 (2008)
30. Tikhonov, A.N., Arsenin, V.Y.: Solution of Ill-Posed Problems. Wiley, New York (1977)
31. Vinokurov, V.A.: Two notes on the choice of regularization parameter. USSR Comput. Math. Math. Phys. **12**(2), 249–253 (1972)