An Iterative Method for the Solution of Nonlinear Regularization Problems with Regularization Parameter Estimation.

E. Loli Piccolomini* F. Zama†

September 22, 2009

Abstract

Ill posed problems constitute the mathematical model of a large variety of applications. Aim of this paper is to define an iterative algorithm finding the solution of a regularization problem. The method minimizes a function constituted by a least squares term and a generally nonlinear regularization term, weighted by a regularization parameter. The proposed method computes a sequence of iterates approximating the regularization parameter and a sequence of iterates approximating the solution. The numerical experiments performed on 1D test problems show that the algorithm gives good results with different regularization functions both in terms of precision and computational efficiency. Moreover, it could be easily applied to large size regularization problems.

Keywords: Regularization methods, Tikhonov method, Ill-posed problems, Integral equations.
Classification: 65R30, 65R32, 65F22.

1 Introduction

Ill posed problems constitute the mathematical model of a large variety of applications whose physical model is a first kind Fredholm integral equation. Such problems can be expressed by the linear relation

$$Hx = y \tag{1}$$

where H denotes a compact operator between Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$. It is well known that problem (1) is ill posed in the sense of Hadamard since H is not continuously invertible or, equivalently, the set $\{x \in \mathcal{X} : \|Hx - y\| \leq \delta\}$ is unbounded [7]. Regularization methods are introduced to solve a slightly modified well-posed problem whose solution is close to the original one [2].

*Department of Mathematics, University of Bologna, e-mail: piccolom@dm.unibo.it
†Department of Mathematics, University of Bologna, e-mail: zama@dm.unibo.it

One of the most widely used and well established regularization method is the Tikhonov regularization method which requires the solution of the following minimization problem:

$$\min_x \|Hx - y\|^2 + \lambda \mathcal{R}(x) \tag{2}$$

where $\|\cdot\|$ denotes the usual euclidean norm, $\lambda > 0$ is known as *regularization parameter* whose value must be determined. The operator $\mathcal{R}(x)$ is used to impose some constraints about the smoothness of the solution.

The choice of the regularization parameter λ is a crucial issue: a small value gives a good approximation to the original operator but presents instabilities due to the influence of data errors. Conversely, a large value of λ suppresses the data errors but increases the operator approximation error. A wide variety of methods can be found in the literature to compute a suitable value for the regularization parameter. In the absence of any prior knowledge of the entity of data perturbation, a common practice is to use methods based on *a posteriori* criteria. In this case one needs to solve problem (2) more than once for different values of λ. When $\mathcal{R}$ is a euclidean norm operator (i.e. $\mathcal{R}(x) = \|Lx\|^2$) widely used methods are the Generalized Cross Validation method (GCV) [15] and the L-Curve methods [10]. In [11, 16], the values of the regularization parameters are obtained through an iterative step reduction procedure that is suitable for solving large scale problems.

As observed in [1] and [13] the problem of finding a regularizing solution for (1) can be stated as the following constrained optimization problem:

$$\min_x \|Hx - y\|^2 : \ s.t. \quad \mathcal{R}(x) \leq \gamma^* \tag{3}$$

for a given value $\gamma^* > 0$ which represents the level of smoothness required. In order to have a good regularized solution of the original problem (1), the parameter γ^* should be as close as possible to $\mathcal{R}(x^*)$ where x^* is the solution of (1).

In the case $\mathcal{R}(x) = \|x\|^2$ algorithm LSTRS [12] finds the solution of a sequence of modified trust region subproblems, solving each subproblem as a parametric eigenvalue problem [14]. This method is not immediately extensible to the case of a general non linear regularization functional.

On the other hand the approach followed in [1] can be easily extended to the case of a general regularization functional giving raise to linear or nonlinear regularization algorithms. It is not clear however how this algorithm behaves in presence of noisy data when the problem is obtained from the discretization of ill-posed problems. Furthermore none of the previously cited papers gives a procedure to compute a suitable value γ^* for obtaining efficient regularized solutions.

Aim of this paper is to describe an iterative algorithm which computes a succession of approximate solutions x_k and regularization parameters λ_k that converge, in absence of data noise, to the solution of (3) and that can be applied to linear and nonlinear regularization functionals.

When the data are affected by noise, we prove that the method behaves like a regularization method: it finds a solution $(x^\delta, \lambda^\delta)$ that converges to the noisy free solution (x^*, λ^*) as $\delta \to 0$.

Finally we introduce a numerical procedure to compute an estimate $\widehat{\gamma}$ of the parameter γ^* that produces good regularizing solutions of problem (1).

Numerical tests are performed to observe the behavior of the algorithm on small-medium size regularization problems.

In section 2 we introduce our iterative method for a general nonlinear regularization functional and prove the convergence both in the noiseless and noisy data cases. Details about the algorithm are discussed in section 3. The results obtained by several numerical experiments are reported in section 4.

2 The Iterative Method

In order to describe our iterative algorithm we define (x^*, λ^*) as the point that minimizes the Lagrangian functional $\mathcal{L}(x, \lambda)$ of problem (3):

$$\mathcal{L}(x, \lambda) = \|Hx - y\|^2 + \lambda\left(\mathcal{R}(x) - \gamma^*\right). \tag{4}$$

Then (x^*, λ^*) satisfies the following Kurush Khun Tucker (KKT) conditions:

$$H^*Hx + \lambda \nabla_x \mathcal{R}(x) - H^*y = 0 \tag{5}$$

$$\mathcal{R}(x) - \gamma^* = 0 \tag{6}$$

where H^* is the adjioint of H and $\nabla_x \mathcal{R}$ is the gradient of the regularization operator with respect to x.

Proposition 1. *Let $\lambda_0 > 0$ be a given value and $x_0 \equiv x(\lambda_0)$ obtained solving (5) with $\lambda = \lambda_0$. Let $\mathcal{R}(x)$ be a continuous function such that $0 < \mathcal{R}(x_0) < \gamma^*$ with*

$$0 < \mathcal{R}(x(\lambda)) < \gamma^*, \quad \forall \lambda > \lambda^* \tag{7}$$

and

$$\mathcal{R}(x(\lambda)) > \gamma^*, \quad \forall \lambda \in [0, \lambda^*]. \tag{8}$$

Then the succession (x_k, λ_k) computed by the following iterative procedure:

$$\lambda_k = \mathcal{F}_b(\lambda_{k-1}), \quad \mathcal{F}_b(\lambda_{k-1}) = \lambda_{k-1}\left(1 - sign(\gamma^* - \gamma_{k-1})\frac{1}{2}\right) \tag{9}$$

with

$$\begin{array}{l} \gamma_{k-1} = \mathcal{R}(x_{k-1}), \quad \text{and} \quad x_{k-1} \text{ such that} \\ H^*Hx_{k-1} + \lambda_{k-1}\nabla_x \mathcal{R}(x_{k-1}) - H^*y = 0 \end{array}, \quad k = 1, 2, \ldots \tag{10}$$

converges to (x^+, λ^+) that fulfills the KKT conditions (5),(6).

Proof. Let's define

$$G(\lambda) \equiv \mathcal{R}(x(\lambda)) - \gamma^* \tag{11}$$

for the hypotheses we have that $G(\lambda_0) < 0$, and $G(0) > 0$. The bisection method applied in the interval $[0, \lambda_0]$ converges linearly to the root λ^+ such that $G(\lambda^+) = 0$

It is not difficult to observe that (9) computes the sequence of iterates λ_k applying the bisection method in the interval $[0, \lambda_0]$. Concluding we have that λ_k converges linearly to λ^+ fulfilling condition (6) and, using the continuity of (5), $x(\lambda_k)$ converges to $x^+ \equiv x(\lambda^+)$ that fulfills the KKT conditions (5). □

In order to speed up the convergence modified Newton methods can also be applied, in particular both Newton and Hebden methods are used in [1]. In the following proposition we report the convergence conditions for the secant method that can be coupled with the bisection method to obtain a globally convergent method.

Proposition 2. *Let the succession* (x_k, λ_k) *be computed by the iterative procedure* (10) *with*

$$\lambda_k = \mathcal{F}_s(\lambda_{k-1}), \quad \mathcal{F}_s(\lambda_{k-1}) = \lambda_{k-1} - \frac{\gamma_{k-1} - \gamma *}{\gamma_{k-1} - \gamma_{k-2}} (\lambda_{k-1} - \lambda_{k-2}) \tag{12}$$

If $D \subseteq \mathbb{R}$ *is an open interval such that* $\lambda^* \in D$ *and*

$$\left| \frac{d}{d\lambda} \mathcal{R}(x(\lambda)) \right| \geq \rho, \;\; \rho > 0, \quad \forall \lambda \in D \tag{13}$$

$$|\mathcal{R}(x(\lambda)) - \mathcal{R}(x(\mu))| \leq \alpha \, |\lambda - \mu| \, , \quad \alpha > 0 \tag{14}$$

Then $\exists I \subseteq D$ *open interval of size* 2η*:* $I = \{\lambda : |\lambda - \lambda^*| \leq \eta\}$ *such that the succession* $\{\lambda_k\}$ *computed by* (12) *converges to* λ^**,* $\forall \lambda_0, \lambda_1 \in I$ *and the pair* (x_k, λ_k) *obtained by* (10), (12) *converges to* (x^*, λ^*) *that fulfills the KKT conditions* (5),(6)*.*

Proof. We observe that (12) is the secant method applied to the nonlinear equation (11), which is convergent to λ^* for the hypotheses (13) [6]. We have that λ_k converges to λ^* fulfilling condition (6) and, using the continuity of (5), $x(\lambda_k)$ converges to $x^* \equiv x(\lambda^*)$. In order to start up properly the sequence (12) the values λ_0, λ_1 can be obtained by performing several iterations of the bisection method (9). □

If (4) has a unique global minimizer (x^*, λ^*), then we have $x^+ = x^*$ and $\lambda^+ = \lambda^*$, otherwise the values computed by the method in proposition 1 define a local minimizer (x^+, λ^+).

In applicative problems the data are usually affected by noise i.e. the observed data y^δ are such that: $\|y - y^\delta\| < \delta$. We can prove that the iterative method (10) converges to a regularized solution $(x^\delta, \lambda^\delta)$ such that $x^\delta \to x^+$ and $\lambda^\delta \to \lambda^+$ as $\delta \to 0$.

Proposition 3. *Let* $(x_k^\delta, \lambda_k^\delta)$*,* $k = 0, 1, \ldots$ *be the succession obtained by the method* (10) *applied to the noisy problem:*

$$\min_x \|Hx - y^\delta\|^2 : \;\; s.t. \quad \mathcal{R}(x) \leq \gamma^* \tag{15}$$

then the succession $(x_k^\delta, \lambda_k^\delta)$ *converges to* $(x^\delta, \lambda^\delta)$ *such that*

$$\|x^\delta - x^+\|^2 + |\lambda^\delta - \lambda^+|^2 < \tau\delta^2, \quad \tau > 0. \tag{16}$$

Proof. Let $y^\delta = y + \delta\eta$ with $\|\eta\| = 1$ and $\delta > 0$.

Following the arguments in proposition 1, $(x^\delta, \lambda^\delta)$ satisfies the KKT conditions:

$$H^*Hx^\delta + \lambda^\delta \nabla_x (\mathcal{R})(x^\delta) - H^*y^\delta = 0 \tag{17}$$

$$\mathcal{R}(x^\delta) - \gamma^* = 0 \tag{18}$$

Subtracting (5) from (17) we obtain

$$H^*H(x^\delta - x^+) + \lambda^\delta \nabla_x (\mathcal{R})(x^\delta) - \lambda^+ \nabla_x (\mathcal{R})(x^+) = \delta H^*\eta$$

substituting the first order approximation of $\nabla_x (\mathcal{R})(x^+)$, given by:

$$\nabla_x (\mathcal{R})(x^+) \simeq \nabla_x (\mathcal{R})(x^\delta) + (x^+ - x^\delta)\nabla_{xx} (\mathcal{R})(x^\delta)$$

we obtain

$$H^*H(x^\delta - x^+) + (\lambda^\delta - \lambda^+)\nabla_x (\mathcal{R})(x^\delta) \simeq \delta H^*\eta$$

Setting $\Gamma \equiv \nabla_x (\mathcal{R})(x^\delta)$ and defining $\Delta\Gamma$ as an error vector we obtain:

$$H^*H(x^\delta - x^+) + (\lambda^\delta - \lambda^+)\Gamma = \delta (H^*\eta + \Delta\Gamma) \tag{19}$$

Let H be an $m \times n$ matrix with $m > n > 0$, introducing $B \in \mathbb{R}^{n,n+1}$ and $z \in \mathbb{R}^{n+1}$ such that:

$$B = \begin{bmatrix} H^*H & \Gamma \end{bmatrix}, \quad z = [z_x, z_\lambda]^t, \text{ with } z_x = (x^\delta - x^+) \text{ and } z_\lambda = (\lambda^\delta - \lambda^+)$$

then we can write relation (19) as

$$Bz = \delta (H^*\eta + \Delta\Gamma) \tag{20}$$

where:

$$\|Bz\|^2 = z_x^* (H^*H)^2 z_x + 2z_\lambda \Gamma^* H^*H z_x + z_\lambda^* \Gamma^*\Gamma z_\lambda \tag{21}$$

Using the the singular value decomposition (SVD) of the matrix H:

$$H = U\Sigma V^*, \quad \text{where } U = [u_1, \dots, u_m], V = [v_1, \dots, v_n]$$

are unitary matrices and $\Sigma = diag(\sigma_1, \dots \sigma_n)$ is the $m \times n$ diagonal matrix with $\sigma_1 \geq \sigma_2, \geq \cdot \geq \sigma_p > 0)$, $p \leq n$, and substituting in (21):

$$\|Bz\|^2 = z_x^* V S^4 V^* z_x + |z_\lambda|^2 \|\Gamma\|^2 + 2z_\lambda \Gamma^* V S^2 V^* z_x$$

Where S^4 defines the diagonal matrix of size n with non zero elements given by σ_i^4, (the same definition applies to S^2). Setting $w = V^* z_x$ and $g = V^*\Gamma$ we obtain:

$$\|Bz\|^2 = \sum_{i=1}^n \sigma_i^4 w_i^2 + |z_\lambda|^2 \sum_{i=1}^n g_i^2 + C \tag{22}$$

where $C = 2z_\lambda \sum_{i=1}^n \sigma_i^2 g_i w_i$. Let's define a positive parameter μ such that

$$|C| = \mu \left(\sum_{i=1}^n \sigma_i^4 w_i^2 + |z_\lambda|^2 \sum_{i=1}^n g_i^2 \right) \tag{23}$$

and the sign function:

$$sgn(C) = \begin{cases} 1 & C \geq 0 \\ -1 & C < 0 \end{cases}$$

then from (22) and (23) it follows:

$$\|Bz\|^2 = (1 + sgn(C)\mu) \left(\sum_{i=1}^n \sigma_i^4 w_i^2 + |z_\lambda|^2 \sum_{i=1}^n g_i^2 \right).$$

When $C > 0$ we have

$$\|Bz\|^2 > \left(\sum_{i=1}^n \sigma_i^4 w_i^2 + |z_\lambda|^2 \sum_{i=1}^n g_i^2 \right)$$

otherwise:

$$\|Bz\|^2 = (1 - \mu) \left(\sum_{i=1}^n \sigma_i^4 w_i^2 + |z_\lambda|^2 \sum_{i=1}^n g_i^2 \right) \tag{24}$$

with $\mu < 1$ since $\|Bz\|^2 > 0$. By setting $\|h\|^2 = \sum_{i=1}^p \sigma_i^4 w_i^2$ and $\|g\|^2 = \sum_{i=1}^n g_i^2$ and defining a value $\alpha \in [0, 1)$ such that

$$\alpha = \begin{cases} \mu & C < 0 \\ 0 & C \geq 0 \end{cases}$$

supposing $\|h\|^2 > 0$, we can rewrite the relation (24) as follows:

$$(1 - \alpha) \left(\|h\|^2 + |z_\lambda|^2 \|g\|^2 \right) \leq \|Bz\|^2. \tag{25}$$

Let's define $s = (\sigma_1^2, \ldots, \sigma_p^2, 0, \ldots, 0)$, such that $s \in \mathbb{R}^n$. We have:

$$(s^t w)^2 = \|h\|^2 + 2 \sum_{i=1}^p \sigma_i^2 w_i \left(\sum_{j \neq i} \sigma_j^2 w_j \right)$$

and obtain that

$$\|h\|^2 = \left[1 - \frac{2}{(s^t w)^2} \sum_{i=1}^p \sigma_i^2 w_i \left(\sum_{j \neq i} \sigma_j^2 w_j \right) \right] (s^t w)^2 \tag{26}$$

Using the Cauchy-Schwartz inequality $(s^t w)^2 \leq \|s\|^2 \|w\|^2$, we can define a positive parameter $0 < \theta \leq 1$ such that

$$(s^t w)^2 = \theta \|s\|^2 \|w\|^2$$

Substituting in (26) we can define a value $M > 0$ such that:

$$\|h\|^2 = M\|s\|^2\|w\|^2, \quad \text{where} \quad M = \theta\left[1 - \frac{2}{(s^t w)^2}\sum_{i=1}^{p}\sigma_i^2 w_i\left(\sum_{j\neq i}\sigma_j^2 w_j\right)\right]$$

Substituting in equation (25) we obtain:

$$(1-\alpha)\left(M\|s\|^2\|w\|^2 + |z_\lambda|^2\|g\|^2\right) \leq \|Bz\|^2$$

Defining $s_{min} = \min(M\|s\|^2, \|g\|^2)$ we have $s_{min} > 0$ and obtain from (25):

$$s_{min}(1-\alpha)(\|w\|^2 + |z_\lambda|^2) \leq \|Bz\|^2$$

Using (20) and recalling that $\|w\| = \|z_x\|$ we have:

$$\|z_x\|^2 + |z_\lambda|^2 \leq \frac{\delta^2}{(1-\alpha)s_{min}}\left(\|H^*\eta\|^2 + \|\Delta\Gamma\|^2\right)$$

Since η is a unitary vector we have $\|H^*\eta\|^2 \leq \sigma_1^2$:

$$\|z_x\|^2 + |z_\lambda|^2 < \frac{\delta^2\sigma_1^2}{(1-\alpha)s_{min}}\left(1 + \frac{\|\Delta\Gamma\|^2}{\sigma_1^2}\right)$$

then we can define $\tau > 0$:

$$\tau \equiv \frac{\sigma_1^2}{(1-\alpha)s_{min}}\left(1 + \frac{\|\Delta\Gamma\|^2}{\sigma_1^2}\right)$$

and conclude:

$$(x^\delta - x^+)^2 + (\lambda^\delta - \lambda^+)^2 \leq \delta^2\tau, \quad \text{with} \quad \tau > 0$$

□

We observe that when the data are affected by noise, the method behaves like a regularization method: it finds a solution $(x^\delta, \lambda^\delta)$ that converges to the noisy free solution (x^*, λ^*) as $\delta \to 0$.

3 Algorithmic features

Using the results stated in propositions 1, 2 and 3 we can define the linear iterative regularization algorithm named `TKit` as follows:

Definition 4. Iterative Regularization Algorithm `TKit`.

Algorithm 1 (TKit(H, y^δ, L, γ^*, λ_0, k_s, $maxit$)).

set $k = 0$
compute x_0 *s.t.* $H^*Hx_0 + \lambda_0 \nabla_x(\mathcal{R}(x_0)) = H^*y^\delta$
repeat
  $k = k + 1$
  if $k < k_s$, $\lambda_k = \mathcal{F}_b(\lambda_{k-1})$ *as in* (9)
  else $\lambda_k = \mathcal{F}_s(\lambda_{k-1})$*as in* (12)
  compute x_k *s.t.* $H^*Hx_k + \lambda_k \nabla_x(\mathcal{R}(x_k)) - H^*y^\delta = 0$
  $\gamma_k = \mathcal{R}(x_k)$
until convergence

where $maxit$ is the maximum number of allowed iterations and the starting value λ_0 is supposed to fulfil condition (7). The parameter k_s defines the iteration for starting the secant method (12). Setting $k_s = maxit$ the bisection method (9) is used until the maximum number of allowed iterations $maxit$ is reached.

In the solution of inverse ill-posed problems different kind of regularization functionals $\mathcal{R}$ are used. From the algorithmic point of view, we distinguish between the regularization operators that produce a linear system in the KKT condition (17) and those that give raise to a nonlinear system. In the first case a class of regularization functional that are widely applied is $\mathcal{R}(x) \equiv \|Lx\|_2^2$ where L is either the identity matrix or a discrete approximation of a derivative operator. In this case $\mathcal{R}$ is a convex function having its minimum at $x = 0$. Substituting

$$\nabla_x(\mathcal{R}(x_k)) = L^*Lx_k$$

in the KKT conditions (5) we obtain a system of linear equations:

$$(H^*H + \lambda_k L^*L)x_k = H^*y^\delta \tag{27}$$

whose matrix is symmetric and positive definite provided that $\mathcal{N}(H) \bigcap \mathcal{N}(L) = \emptyset$, ($\mathcal{N}(\cdot)$ is the kernel operator).

The main computational load in algorithm TKit is given by the solution of linear system (27). Iterative methods such as Conjugate Gradient (CG) can be used to exploit the sparsity of the matrix L.

As an example of regularization functional that gives raise to a nonlinear system in the KKT conditions (5) we consider here:

$$\mathcal{R}(x) \equiv \left(\sum_{i=1}^{n} |x_i|^p \right), \qquad 1 < p < \infty, \quad p \neq 2. \tag{28}$$

In this case the KKT conditions (5) are a system of nonlinear equations where:

$$(\nabla_x \mathcal{R}(x))_i \equiv p \cdot sgn(x_i)|x_i|^{p-1}, \quad i = 1, \ldots, n$$

In this case $\mathcal{R}$ may be a non convex function depending on p and the values of the vector x. In particular the Hessian matrix Z is a diagonal matrix with components:

$$Z_i = p(p-1)|x_i|^{p-2}, \quad i = 1, \ldots, n$$

It is not well defined if $x_i = 0$ and $p < 2$ and it is semi positive definite if $x_i = 0$ and $p > 2$. In these cases we expect difficulties in the numerical solution of the nonlinear system (5) and some suitable stopping criteria must be considered.

The solution of the nonlinear system (5) can be efficiently performed by the Newton method [6] since the Jacobian matrix requires only a diagonal update:

$$J(x) = H^*H + \lambda_k Z.$$

The Newton directions can be efficiently computed by truncated Conjugate Gradient iterations provided $J(x)$ is nonsingular.

In the solution of ill-posed problems the parameter $\gamma^* \equiv \mathcal{R}(x^*)$ is not known and an approximate value $\widehat{\gamma}$ is usually heuristically given by the user. We propose here to compute $\widehat{\gamma}$ as follows:

$$\widehat{\gamma} = \mathcal{R}(\widehat{x}) \tag{29}$$

where $\widehat{x}$ is a rough approximation of solution of the least squares problem $\min \|Hx - y^\delta\|$ obtained by performing several iterations of `cgls` algorithm [5].

We exploit the decreasing behavior of the norm of the residual vector $r_k \equiv H^*y^\delta - H^*Hx_k$, which is fast in the first few iterations and then becomes very slow when the iterations increase. We observe that the ratio between the norm of two successive residual vectors $\|r_k\|/\|r_{k-1}\|$ is small in the first few iterations an then it increases becoming very close to 1. An approximate solution $\widehat{x}$ is computed by stopping the iterations of the `cgls` method as soon as $\|r_k\|/\|r_{k-1}\| \geq 0.9$. For the regularization properties of the CG iterations ([5], [4]), $\widehat{x}$ can be considered as a low pass filtered version of the true solution x^*.

Since $\widehat{\gamma} \neq \gamma^*$, the algorithm `TKit` converges to a an approximate solution $(\widetilde{x}, \widetilde{\lambda})$ which is different from the optimal (x^*, λ^*). Therefore the truncation of the iterative process is usually a better solution in terms of errors and time efficiency. Hence, the stopping criterium used for the algorithm convergence is the following:

$$\texttt{if} \;\; |\lambda_{k+1} - \lambda_k| < Tol|\lambda_k| + \varepsilon \;\; \texttt{or} \;\; \lambda_{k+1} > \lambda_k \quad \texttt{then} \quad \texttt{stop} \tag{30}$$

where $\epsilon \simeq 1.e-16$ and Tol is a fixed tolerance.

The criterium $\lambda_{k+1} > \lambda_k$ is based on the observation that, instead of executing all the iterations until convergence, we stop them as soon as $x(\lambda_k)$ is not feasible, i.e. $\mathcal{R}(x(\lambda_k)) > \widehat{\gamma}$. For the hypotheses in proposition 3, the algorithm starts from a value of λ_0 such that $\mathcal{R}(x(\lambda_0)) \leq \widehat{\gamma}$ then it decreases its value as long as $x(\lambda_k)$ is an admissible point for the constraints (i.e. $\mathcal{R}(x(\lambda_k)) \leq \widehat{\gamma}$) and increases it otherwise. Since the succession λ_k is decreasing in the first steps and may oscillate near the limit $\widetilde{\lambda}$, we choose to stop the iterations as soon as $\lambda_{k+1} > \lambda_k$.

4 Numerical results

In this section we present the numerical results obtained with the proposed method. The numerical tests have been performed on a Pentium PC 2GHz equipped with 2Gb of RAM and using Matlab 7.4.0.

In paragraph 4.1 we analyze the convergence of algorithm `TKit` both in absence and presence of noise. To this purpose we define test problems with known solution (x^*, λ^*); in this case $\gamma^* \equiv \mathcal{R}(x^*)$ is used in (3).

Successively (par. 4.2) we report the results obtained in the solution of ill-posed test problems. Only the true solution x^* is known and the parameter γ^* is substituted by the estimate $\widehat{\gamma}$ computed in (29). The algorithm is tested both in the linear and nonlinear case.

Finally (par. 4.3) we compare the results obtained by `TKit` with the functions available in Hansen's regtool based on the Generalized Singular Values Decomposition.

4.1 Tests with assigned γ^*

In these test problems we used the data vector y and the matrix H obtained from the Matlab functions `Shaw` and `Phillips` of Hansen Regtool package [9] with problem size $n = 100$. Then we fixed a value for λ^* (i.e. $\lambda^* = 0.1$) and computed x^* solving (17). Finally the value γ^* si computed from (18). We eventually added white noise to y in order to obtain the noisy data y^δ.

These test problems constitute examples of different kind of ill-posedness. In the first case (`Shaw`) the singular values decay to values smaller than machine epsilon, showing that the regularization parameters are affected mainly by the amount of noise present in y^δ (figure 1(a)). In the second case (`Phillips`) the singular values decay to $\approx 10^{-5}$ (figure 1(b)). In this case not only the noise present in y^δ but also the data errors affect the regularization parameter [8].

The regularization functions $\mathcal{R}(x)$ are computed by the Regtool function `get_l(n,d)`:

$$\mathcal{R}(x) = \|Lx\|_2 \quad \text{where} \quad L = \begin{cases} I & (\mathtt{d} = \mathtt{0}) \\ D_1 & (\mathtt{d} = \mathtt{1}) \\ D_2 & (\mathtt{d} = \mathtt{2}) \end{cases} \tag{31}$$

where the matrix $D_1 \in \mathbb{R}^{(n-1)\times n}$ is the forward difference approximation of the first derivative and the matrix $D_2 \in \mathbb{R}^{(n-2)\times n}$ is the difference approximation of the second derivative computed in the interior points:

$$D_1(i,j) = \begin{cases} -1 & j = i \\ 1 & j = i+1 \\ 0 & \text{otherwise} \end{cases} \qquad D_2(i,j) = \begin{cases} -1 & j = i, i+2 \\ 2 & j = i+1 \\ 0 & \text{otherwise.} \end{cases}$$

The iterations of the algorithm `TKit` are stopped by means of the absolute error criterium (16):

$$\|x_k - x^*\|^2 + |\lambda_k - \lambda^*|^2 < Tol, \quad TOL = 10^{-10}.$$

In the first experiment we ran `TKit` without noise added using both the bisection method (i.e. $k_s = maxit$) and the secant method with $k_s = 3$. In tables 1 and 2 we report the number of iterations performed (it) and the relative errors for the computed solution (x_{it}, λ_{it}):

$$Erel_x = \|x_{it} - x^*\|/\|x^*\|, \quad Erel_\lambda = |\lambda_{it} - \lambda^*|/\lambda^*, \tag{32}$$

with the initial value $\lambda_0 = 1$. We observe a better performance of the secant

method	`d`	it	$Erel_x$	$Erel_\lambda$
Bisect	0	37	4.384e-012	8.731e-011
	1	37	7.254e-012	8.731e-011
	2	35	1.982e-012	3.492e-010
Secant	0	10	2.934e-014	3.192e-015
	1	12	1.195e-013	6.273e-014
	2	10	5.015e-013	8.841e-012

Table 1: `TKit` algorithm with bisection and secant methods applied to noiseless data: `Shaw` test problem. `d`: regularization functional (31); it: number of iterations; $Erel_x$, $Erel_\lambda$: relative errors(32).

method	`d`	it	$Erel_x$	$Erel_\lambda$
Bisect	0	34	1.836e-011	2.328e-010
	1	34	7.206e-012	2.328e-010
	2	34	2.726e-012	2.328e-010
Secant	0	9	8.504e-014	5.024e-014
	1	10	5.915e-012	1.654e-012
	2	10	8.470e-013	4.395e-011

Table 2: `TKit` algorithm with bisection and secant methods applied to noiseless data: `Phillips` test problem. `d`: regularization functional (31); it: number of iterations; $Erel_x$, $Erel_\lambda$: relative errors(32).

method in terms of iterations and relative errors. An example of the converging behavior of the sequence λ_k is shown in figure 2 both for the `shaw` and the `phillips` test problems with $\lambda_0 \neq 1$.

The convergence of `TKit` algorithm is also tested in the case of noisy data y^δ computed as follows:

$$y^\delta = y + \delta\eta, \quad \|\eta\| = 1,$$

with η random unitary vector. The noisy problem (15) is solved with different values of δ. In this case we consider a hybrid bisection-secant algorithm setting $k_s = 5$. In order to fulfil the hypotheses of proposition 1, we estimated the starting value λ_0 by setting:

$$\lambda_0 = \min_\ell \frac{\ell + 1}{2}, \quad \ell = 0, 1, \ldots \ \texttt{so that} \ \mathcal{R}(x(\lambda_0)) < \gamma^*$$

Usually $\ell \le 2$.

In table 3 we report the results obtained with three noise levels: $\delta = 10^{-6}$ (low noise), $\delta = 10^{-4}$ (medium noise) and $\delta = 10^{-2}$ (high noise). From the

Test	d	δ	it	$Erel_x$	$Erel_\lambda$
Shaw	0	1.e-006	12	1.696e-006	7.428e-006
		1.e-004	12	1.697e-004	7.424e-004
		1.e-002	12	1.736e-002	6.979e-002
	1	1.e-006	13	3.029e-005	2.721e-004
		1.e-004	13	3.080e-003	2.702e-002
		1.e-002	13	4.505e-001	3.660e-001
	2	1.e-006	12	1.540e-004	3.447e-003
		1.e-004	13	1.554e-002	1.530e-001
		1.e-002	14	1.017e+000	**3.395e+001**
Phillips	0	1.e-006	11	1.647e-005	1.124e-004
		1.e-004	11	1.656e-003	1.100e-002
		1.e-002	12	1.342e-001	3.483e-001
	1	1.e-006	12	4.574e-005	3.522e-003
		1.e-004	12	5.513e-003	2.093e-001
		1.e-002	15	6.464e-002	**7.610e+001**
	2	1.e-006	12	1.543e-004	3.241e-002
		1.e-004	13	1.273e-002	4.547e-001
		1.e-002	23	6.538e-002	**2.712e+003**

Table 3: TKit regularization algorithm with hybrid method ($k_s = 5$), shaw and Phillips test problems. d: regularization functional (31); δ noise level; it: number of iterations; $Erel_x$, $Erel_\lambda$: relative errors(32).

values reported we can see that the iteration number does not strongly depend on the noise levels since $11 \le it \le 13$ in all cases where $\max(Erel_x, Erel_\lambda) < 1$ (i.e. when a good approximation is obtained). The relative errors ($Erel_x$, $Erel_\lambda$) are very strictly connected to the noise level δ. In some cases (shown in boldface type in table 3) the relative errors are high. In order to better understand this phenomenon, we try to estimate the value of the parameter τ of (16). We compute the values

$$M(\delta) = \frac{1}{\delta}\sqrt{\frac{\|x_{it} - x^*\|^2 + |\lambda_{it} - \lambda^*|^2}{\|x^*\|^2 + |\lambda^*|^2}} \tag{33}$$

where x_{it} and λ_{it} are the computed solution and regularization parameter, respectively, using 100 values of δ from a logarithmical distribution in the interval $[10^{-6}, 10^{-2}]$ and obtaining the mean value $\overline{M}$ for each test problem and each value of d. The value $\overline{M}^2$ can be considered as an estimate of the parameter τ of proposition 3, i.e.

$$\tau \approx \overline{M}^2.$$

Then, substituting it in (16) we have:

$$\sqrt{\|x^\delta - x^+\|^2 + |\lambda^\delta - \lambda^+|^2} < \overline{M}\delta.$$

The data reported in table 4 show the different values $\overline{M}$ obtained. When the product $\overline{M} \cdot \delta$ is high, the errors on the computed values of x or λ are high, too. This is, for example, the case of the `shaw` test problem with $\delta = 10^{-2}$ and `d` = 2 ($\overline{M} = 2$ in this case).

	$\overline{M}$	
d	Shaw	Phillips
0	1.7	1.7e+1
1	3.5e+1	1.4e+2
2	1.5e+2	5.3e+2

Table 4: Mean value $\overline{M}$ of $M(\delta)$ computed from 100 values of δ with logarithmical distribution in the interval $[10^{-6}, 10^{-2}]$. `d`: regularization functional (31).

4.2 Regularization tests

In this section we report the results obtained from the previous `Shaw` and `Phillips` test problems.

In this case the data vector y and the matrix H and the true solution x^* are obtained from the Matlab functions `Shaw` and `Phillips` with problem size $n = 100$. The noisy vector y^δ is obtained by adding white noise as described before. Numerical experiments are performed using $\gamma^* \equiv \mathcal{R}(x^*)$ and computing the estimate $\widehat{\gamma}$ as in (29). We are interested in the regularized solution x_{reg} that filters out the noise present in the data.

In these experiments he algorithm is tested both in the linear and nonlinear cases. The algorithm has been stopped with the criterium (30), with $Tol = 10^{-3}$.

As an example we report in table 5 the results obtained in the case of $R(x) = \|Lx\|_2^2$ with noise ranging from 10^{-6} (low noise) to 10^{-2} (high noise). The table is split into two parts, in order to compare the results obtained with exact γ^* (columns T1) and with approximated $\widehat{\gamma}$ (columns T2). Columns $k(it_{cg})$ report the number k of external iterations to compute λ_k and the total number it_{cg} of CG iterations required by the solution of the linear system (5). In all the experiments the CG iterations are stopped with a relative tolerance of 10^{-12}.

Analyzing the results reported in table 5 we point out the following issues.

- The estimated parameter $\widehat{\gamma}$ is generally a good approximate value of γ^* and it depends by the data noise: it decreases when the noise increases. This produces large values in the regularization parameters λ that can be observed in the $T2$ column with respect to the same column in $T1$.

- In the 61, 1% of the cases, the values in column $Erel_x$ reported in $T2$ are smaller or equal to those reported in $T1$. Thus proving that the parameter $\widehat{\gamma}$ together with the stopping criterium (30) eventually enhances the quality of the regularized solution. The remaining cases (38.9% in table 5) are mostly related to low noise tests.

- An increase in the computational efficiency can be observed by comparing the smaller number of total CG iterations in columns $T2$ with respect to those in column $T1$.

We can conclude that algorithm `TKit` can be used to efficiently compute regularized solutions in presence of medium/high data noise.

As an example of algorithm `TKit` in the non linear case we report in table 6 the results obtained with regularization function as in (28) with $p = 1.4$ for the `Shaw` test problem and $p = 1.8$ `Phillips` test problem.

The Newton steps in the solution of nonlinear equations (5) are stopped when the step size is less or equal to 10^{-3}, while the CG inner iterations are stopped with a relative tolerance of 10^{-3}.

The good quality of regularized solutions for different noise levels can also be observed in this case.

4.3 Comparison with other methods

In this section we compare the results obtained by solving the regularization problem with $R(x) = \|Lx\|_2^2$ with the `TKit` algorithm with those obtained by the `tikhonov` and the `lsqi` functions available in Hansen's Regtool. These functions are based on the Generalized Singular Value Decomposition `gsvd` [3] and because of the high computational complexity and memory requirements are not suitable for large size problems. In the case of `tikhonov` function, the regularization parameter λ is computed by means of the Generalized Cross Validation function `gcv` [15], while the function `lsqi` computes the regularization parameter by solving the nonlinear equation (6) using the Hebden method [1].

In table 7 we report the relative error $Erel_x$ and the regularization parameter λ in the cases of `tikhonov`, `lsqi` called with γ^* and `lsqi` called with $\widehat{\gamma}$.

Comparing the columns of the relative errors $Erel_x$ obtained with approximated values $\widehat{\gamma}$ in tables 5 and 7, we observe that the errors $Erel_x$ obtained with the proposed algorithm `TKit` are better than those obtained with the `lsqi` function in the case of γ^* (i.e. compare columns $T1$ in 5 and $LSQI(\gamma^*)$). When the estimate $\widehat{\gamma}$ is used, algorithms `TKit` and $LSQI(\widehat{\gamma})$ give very similar errors.

For what concerns the Tikhonov algorithm, the main difficulty is the computation of the regularization parameter: if the `gcv` function computes a good regularization parameter, then the errors are small and comparable to `TKit` (as in the `shaw` test problem); otherwise the errors are not comparable with those obtained with the `TKit` algorithm (as in the `phillips` test problem). As an example we report in figure 3 the solutions obtained with all the methods in the case of the `shaw` test problem with (`d=1`) and $\delta = 5 \cdot 10^{-4}$.

We can conclude that algorithm `TKit` with $\widehat{\gamma}$ computed by (29) computes good regularized solutions at a low computational cost.

5 Conclusions

We defined an iterative algorithm `TKit` which computes a succession of solution approximations x_k and regularization parameters λ_k that converge (in absence of data noise) to the solution of (3). In the case of noisy data we proved that the method finds a solution $(x^\delta, \lambda^\delta)$ that converges to the noisy free solution (x^*, λ^*) as $\delta \to 0$. Finally we introduced an estimation procedure of the parameter γ^* and suitable stopping criteria to compute efficient regularized solutions.

The numerical experiments carried out on linear and nonlinear regularization problems prove that `TKit` is an efficient method that can be easily applied to large size regularization problems. Future developments concern the extension to different non linear regularization functionals such as the total variation functional applied to large size inverse problems in imaging.

References

[1] T. F. Chan, J. A. Olkin, and D. W. Cooley, *Solving quadratically constrained least squares using black box solvers*, BIT **32** (1992), 481–495.

[2] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of inverse problems*, Kluwer Academic Publishers, 1996.

[3] G. H. Golub and F. Van Loan, *Matrix computations*, The Johns Hopkins University Press, 1996.

[4] M. Hanke, *Conjugate gradient type methods for ill posed problems*, Pitman Reserch Notes in Mathematics Series, Longman, 1995.

[5] P. C. Hansen, *Rank–deficient and discrete ill–posed problems*, SIAM, 1997.

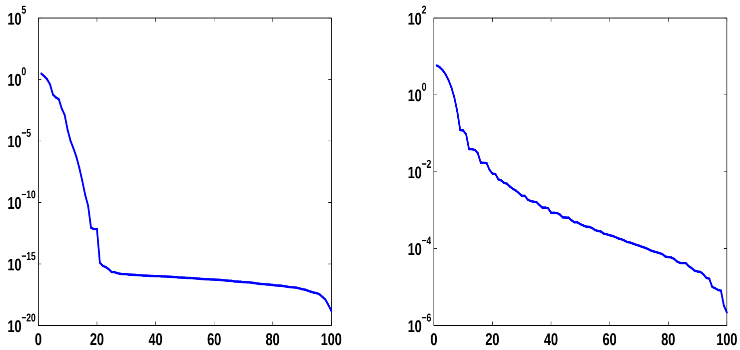
[6] J.E.Dennis and R.B. Schnabel, *Numerical methods for unconstrained optimization and nonlinear equations*, SIAM, 1996.

[7] Rainer Kress, *Linear integral equations*, Applied Mathematica Sciences, vol. 62, Springer,, 1999.

[8] P.C.Hansen, *The discrete Picard condition for discrete ill–posed problems*, BIT **30** (1990), 658–672580.

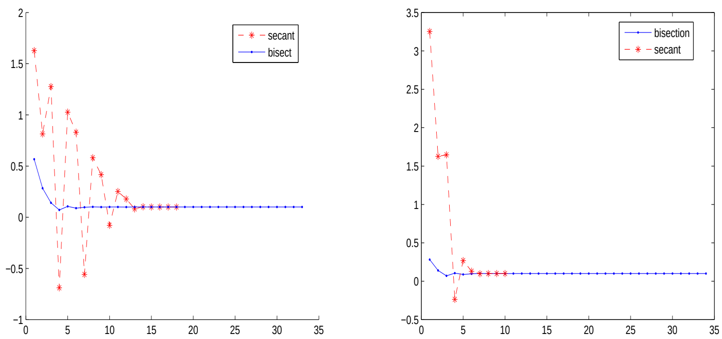
[9] ———, *Regularization Tools: A Matlab package for analysis and solution of discrete ill–posed problems*, Numeric. Algor. **6** (1994), 1–35.

[10] D. P. O'Leary P.C.Hansen, *The use of the L–curve in the regularization of discrete ill–posed problems*, SIAM J. Sci. Comp. **14** (1993), 1487–1503.



(a) (b)

Figure 1: Singular Values: (a) `Shaw` test problem (b)`Phillips` test problem



(a) (b)

Figure 2: Converging behavior of the λ_k succession with bisection and secant methods and (`d=1`). (a) `shaw` test problem (b) `phillips` test problem

[11] A. Frommer P.Maass, *Fast cg–based methods for tikhonov–phillips regularization*, SIAM J. Sci. Comput. **20** (1999), no. 6, 1831–1850.

[12] M. Rojas, S. Santos, and D. Sorensen, *Lstrs a. matlab software for large scale trust-region subproblems and regularization*, ACM Trans. on Math. Software **32** (2008), no. 2.

[13] M. Rojas and D. Sorensen, *A trust regions approach to the regularization of large-scale discrete forms of ill-posed problems*, SIAM J. on Sci. Comput. **23** (2002), no. 6, 1842–1860.

[14] T. Steihaug, *The conjugate gradient method and trust regions in large scale optimization*, SIAM J. Numer. Anal. **20** (1983), no. 3, 626–637.

[15] G. Wahba, *Practical approximate solutions to linear operator equations when the data are noisy*, SIAM J. Num. Anal. **14** (1977), 651–667.

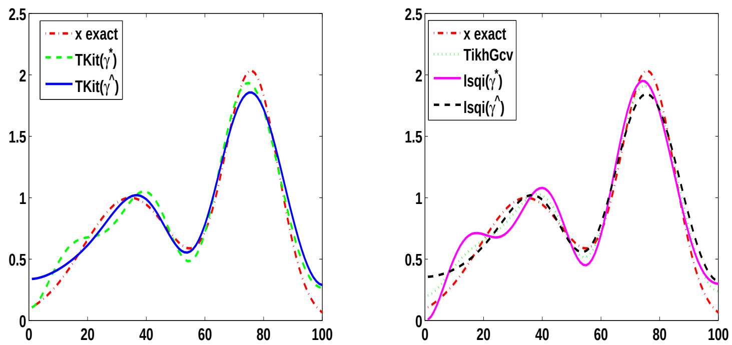
[16] F. Zama and E. Loli Piccolomini", *A Descent Method for Regularization of Ill–Posed Problems*, Journal on Optimization Methods and Software **20** (2005), no. 14-5, 615–628.

Test	d	δ	T1				T2			
			$k(it_{cg})$	$Erel_x$	λ	γ^*	$k(it_{cg})$	$Erel_x$	λ	$\widehat{\gamma}$
`Shaw`	0	1e-006	29(522)	2.535e-002	1.863e-009	9.964e+001	15(163)	4.789e-002	9.155e-005	9.908e+001
	0	1e-004	19(272)	8.823e-002	1.907e-006	9.964e+001	15(172)	5.454e-002	9.155e-005	9.896e+001
	0	1e-002	8(88)	2.064e-001	3.906e-003	9.964e+001	6(48)	1.748e-001	4.688e-002	9.180e+001
	1	1e-006	25(2968)	1.930e-002	2.980e-008	3.632e-001	11(1095)	4.169e-002	1.465e-003	3.361e-001
	1	1e-004	14(1559)	6.613e-002	6.104e-005	3.632e-001	11(1094)	4.318e-002	1.465e-003	3.270e-001
	1	1e-002	5(555)	8.282e-001	3.125e-002	3.632e-001	5(437)	8.282e-001	9.375e-002	2.391e-001
	2	1e-006	21(6757)	8.189e-003	4.768e-007	6.981e-003	13(2981)	3.400e-002	3.662e-004	5.345e-003
	2	1e-004	12(3002)	1.302e-001	2.441e-004	6.981e-003	9(1831)	9.649e-002	5.859e-003	5.233e-003
	2	1e-002	3(682)	1.108e+000	8.750e-001	6.981e-003	3(463)	1.108e+000	2.625e+000	4.019e-003
`Phillips`	0	1e-006	8(152)	1.126e-002	3.906e-003	8.996e+000	10(183)	6.211e-003	2.930e-003	8.998e+000
	0	1e-004	9(209)	1.579e-002	1.953e-003	8.996e+000	10(209)	1.906e-002	2.930e-003	8.998e+000
	0	1e-002	4(74)	1.470e-001	6.250e-002	8.996e+000	3(34)	9.869e-002	3.750e-001	8.674e+000
	1	1e-006	9(671)	4.146e-003	1.953e-003	4.725e-002	4(222)	1.281e-002	1.875e-001	4.686e-002
	1	1e-004	7(592)	1.506e-002	7.813e-003	4.725e-002	5(332)	1.433e-002	9.375e-002	4.690e-002
	1	1e-002	8(710)	5.284e-002	6.852e+000	4.725e-002	8(636)	5.284e-002	2.056e+001	4.099e-002
	2	1e-006	18(2947)	1.904e-003	3.433e-005	7.286e-004	2(156)	1.914e-002	6.750e+000	6.524e-004
	2	1e-004	8(1413)	1.808e-002	3.516e-002	7.286e-004	3(313)	1.834e-002	3.375e+000	6.541e-004
	2	1e-002	12(3560)	6.726e-002	1.326e+002	7.286e-004	11(3132)	4.941e-002	7.954e+002	4.799e-004

Table 5: Results obtained from `TKlin` with γ^* and $\widehat{\gamma}$. Columns T1 are relative to `TKlin` with γ^* and iterations stopped with rule (**??**). Columns T2 are relative to `TKlin` with $\widehat{\gamma}$ computed by (29) and the iterations stopped with rule (30). k denotes the external iterations while it_{cg} is the total number of Conjugate Gradient iterations.

Test	p	δ	$k(it_{cg})$	$Erel_x$	λ	$\widehat{\gamma}$
`Shaw`	1.4	1.e-6	22(1444)	6.202e-003	9.537e-007	8.864e+001
	1.4	1.e-4	21(1337)	7.068e-002	1.907e-006	8.879e+001
	1.4	1.e-2	10(420)	1.524e-001	3.906e-003	8.617e+001
`Phillips`	1.8	1.e-6	22(1403)	4.471e-003	9.537e-007	1.015e+001
	1.8	1.e-4	14(644)	2.266e-002	2.441e-004	1.014e+001
	1.8	1.e-2	6(147)	1.055e-001	6.250e-002	9.902e+000

Table 6: Results obtained from `TKit` with regularization function as in (28): p is the norm parameter in (28); k denotes the external iterations while it_{cg} is the total number of inner Conjugate Gradient iterations performed by the Newton steps. The values of γ^* are 88.095 for `Shaw` and 10.137 for `Phillips`



(a) (b)

Figure 3: `shaw` test problem with (`d=1`) and $\delta = 5 \cdot 10^{-4}$ (a) True solution, `TKit`(γ^*), `TKit`$(\widehat{\gamma})$. (b) `tikhonov` function with λ computed by `gcv`, `lsqi` with γ^* and $\widehat{\gamma}$

Test	d	δ	Tikh Gcv		LSQI (γ^*)			LSQI ($\widehat{\gamma}$)		
			$Erel_x$	λ	$Erel_x$	λ	γ^*	$Erel_x$	λ	$\widehat{\gamma}$
`Shaw`	0	1e-006	2.932e-001	7.480e-006	2.837e-002	5.286e-005	9.964e+001	4.829e-002	8.337e-003	9.908e+001
	0	1e-004	5.993e-002	4.646e-003	9.474e-002	1.692e-003	9.964e+001	5.453e-002	8.165e-003	9.896e+001
	0	1e-002	1.811e-001	1.200e-001	2.656e-001	6.487e-002	9.964e+001	1.760e-001	1.956e-001	9.180e+001
	1	1e-006	3.094e-001	2.044e-005	2.372e-002	1.826e-004	3.632e-001	4.344e-002	3.841e-002	3.361e-001
	1	1e-004	5.119e-002	1.665e-002	7.268e-002	9.523e-003	3.632e-001	4.488e-002	3.791e-002	3.270e-001
	1	1e-002	5.626e-001	3.227e+000	8.934e-001	2.096e-001	3.632e-001	8.186e-001	2.570e-001	2.391e-001
	2	1e-006	3.362e-001	6.517e-005	1.041e-002	7.722e-004	6.981e-003	4.328e-002	2.165e-002	5.344e-003
	2	1e-004	9.683e-002	6.167e-002	1.376e-001	1.712e-002	6.981e-003	9.461e-002	6.766e-002	5.233e-003
	2	1e-002	6.549e-001	8.770e+001	1.122e+000	1.259e+000	6.981e-003	1.025e+000	1.751e+000	4.019e-003
`Phillips`	0	1e-006	1.259e+000	2.264e-006	8.775e-003	6.613e-002	8.996e+000	7.331e-003	5.367e-002	8.998e+000
	0	1e-004	1.259e+002	2.264e-006	1.611e-002	5.845e-002	8.996e+000	1.705e-002	5.172e-002	8.998e+000
	0	1e-002	1.259e+004	2.264e-006	1.675e-001	3.161e-001	8.996e+000	9.857e-002	5.005e-001	8.674e+000
	1	1e-006	1.259e+000	1.142e-006	4.022e-003	5.959e-002	4.725e-002	1.333e-002	3.755e-001	4.686e-002
	1	1e-004	1.259e+002	1.142e-006	1.548e-002	1.193e-001	4.725e-002	1.529e-002	3.064e-001	4.690e-002
	1	1e-002	1.259e+004	1.142e-006	7.083e-002	2.698e+000	4.725e-002	5.471e-002	4.700e+000	4.099e-002
	2	1e-006	1.259e+000	5.746e-007	2.014e-003	7.093e-003	7.286e-004	1.931e-002	2.162e+000	6.524e-004
	2	1e-004	1.259e+002	5.746e-007	1.861e-002	2.555e-001	7.286e-004	2.010e-002	1.887e+000	6.541e-004
	2	1e-002	1.259e+004	5.746e-007	7.620e-002	1.451e+001	7.286e-004	6.381e-002	3.191e+001	4.799e-004

Table 7: Results obtained from `tikhonov` function and λ computed by `gcv` and `lsqi` with γ^* and $\widehat{\gamma}$.