Fixed-point iterations in determining the Tikhonov regularization parameter[*]

Fermín S Viloche Bazán

Department of Mathematics, Federal University of Santa Catarina, Florianópolis SC,
Santa Catarina, CEP 88040-900, Brazil

E-mail: fermin@mtm.ufsc.br

Abstract
We review a Tikhonov parameter criterion based on the search for local minima of the function $\Psi_\mu(\lambda) = x(\lambda)y^\mu(\lambda), \mu > 0$ where $x(\lambda)$ and $y(\lambda)$ are the squared residual norm and the squared solution norm, respectively, proposed earlier by Regińska (1996, *SIAM J. Sci. Comput.* **3** 740). As a consequence, we demonstrate that extreme points of $\Psi_\mu(\lambda)$ are fixed points of a related function, and then propose a fixed-point algorithm for choosing the Tikhonov parameter. The algorithm constructs a regularization parameter associated with the corner of the *L*-curve in log–log scale, thus yielding solutions with accuracy comparable to that of the *L*-curve method but at a lower computational cost. The performance of the algorithm on representative discrete ill-posed problems is evaluated and compared with results obtained by the *L*-curve method, generalized cross-validation and another fixed-point algorithm from the literature.

1. Introduction

We are concerned with the numerical solution of discrete ill-posed problems

$$\min_{f \in \mathbb{R}^n} \|Af - g\|_2, \qquad g \in \mathbb{R}^m, \qquad A \in \mathbb{R}^{m \times n}, \qquad m \geqslant n \qquad (1.1)$$

where the matrix A has a large condition number with singular values decaying to zero without a particular gap in the singular value spectrum, and g consists of exact data plus noise. In this situation, the naive least squares (LS) solution $f_{ls} = A^\dagger g$ (where $A^\dagger$ denotes the pseudoinverse of A) is contaminated by noise to such an extent that it is of no practical value as an approximation to the exact solution f^{exact}, and the LS problem needs to be regularized. For a survey about regularization methods the reader is referred to Hansen [8] and references

therein. The earliest and most well-known method is due to Tikhonov [17] where (1.1) is replaced by the minimization problem

$$f_\lambda = \arg\min \|Af - g\|_2^2 + \lambda^2 \|Lf\|_2^2, \tag{1.2}$$

where L is chosen so as to incorporate desirable properties on the solution such as smoothness, and λ is a positive parameter called the regularization parameter. The proper choice of the regularization parameter is a nontrivial problem for which several strategies have been proposed, one of them being the *L*-curve method [9]. The *L*-curve method derives its name from a plot of the squared norm of the regularized solution $y(\lambda) = \|f_\lambda\|_2^2$ (or $y(\lambda) = \|Lf_\lambda\|_2^2$) versus the squared norm of the residual vector $x(\lambda) = \|g - Af_\lambda\|_2^2$ in log–log scale, which typically has an *L*-shaped form with distinctive vertical and horizontal parts. The method relies on the observation that the vertical part of the curve for small changes in λ corresponds to rapidly varying regularized solutions f_λ with very little change in the residual norm $x(\lambda)$, while the horizontal part for large values of λ corresponds to slowly varying $y(\lambda)$ with large changes in the residual norm. Thus, a reasonable regularized solution should lie in the vicinity of the 'corner' of the *L*-curve where $y(\lambda)$ is about to start growing and λ remains almost unchanged. For most problems the 'corner' is a *region* and the challenge is to select a point from this region. The *L*-curve method selects the parameter λ which maximizes the curvature of the *L*-curve. Although the idea of the *L*-curve method is intuitively clear, computing the point of the maximum curvature in a robust way is not an easy task. For insightful discussions on properties and drawbacks of the *L*-curve as well as of other methods, the reader is referred to Engl and Grever [2], Hanke [5], Hansen [[8], chapter 6], Regińska [15], Vogel [16] and Hämarik *et al* [3, 4], among others. Another strategy which also looks for the point of the maximum curvature of a certain function is the *U*-curve method; details can be found in [11].

In this work, we propose an iterative procedure for selecting the Tikhonov parameter based on an earlier work of Regińska [15], where the parameter λ that maximizes the curvature of the *L*-curve in log–log scale is related to a local minimum of the function

$$\Psi_\mu(\lambda) = x(\lambda)y^\mu(\lambda), \qquad \mu > 0. \tag{1.3}$$

In this respect, Regińska proved that if the curvature of the *L*-curve is maximized at $\lambda = \lambda^*$, and if the tangent to the *L*-curve at $(\log x(\lambda^*), \log y(\lambda^*))$ has slope $-1/\mu$, then Ψ_μ is minimized at $\lambda = \lambda^*$. However, no further work was done on how to compute such a minimizer. Here, we investigate the behavior of $\Psi_\mu(\lambda)$, and conclude that the minimizers can be calculated through a fixed-point (FP) algorithm that determines a point near the *L*-corner of the maximum curvature. Practically, the FP algorithm only needs computation of the solution norm (or solution seminorm) and the residual norm, while the *L*-curve method requires either the SVD (or GSVD) or the computation of the derivative of the solution norm with respect to the regularization parameter [7]. This makes the FP algorithm simpler and potentially better suited for large-scale problems.

The paper is organized as follows. In section 2, we describe the properties of Ψ_μ on which the proposed parameter choice relies, discussing the existence of minimizers and the role that μ plays in the minimization process. The algorithm and convergence properties are described in section 3. In section 4, the performance of the algorithm on representative discrete ill-posed problems, which include a super-resolution image estimation problem, is evaluated and compared with results obtained by means of the *L*-curve method, generalized cross-validation (GCV) and another fixed-point algorithm by Belge *et al* [1].

We end the section by introducing some preliminary results and notation. As usual, we assume that L is $p \times n$, $\text{rank}(L) = p \leqslant n$, and that the pair (A, L) have a GSVD

$$A = U \begin{bmatrix} \Sigma & 0 \\ 0 & I_{n-p} \end{bmatrix} X, \qquad L = V[M; 0]X. \tag{1.4}$$

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, $X \in \mathbb{R}^{n \times n}$ is nonsingular, and Σ and M are diagonal matrices: $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_p)$, $M = \mathrm{diag}(\mu_1, \ldots, \mu_p)$. Define $\alpha_i = \left| u_i^T g \right|^2$ (the squared Fourier coefficient of g), $\delta_0 = \| (I - UU^T) g \|_2$ (the size of the incompatible component of g that lies outside the column space of A). Let the generalized singular values of the pair (A, L) be denoted by γ_i, i.e., $\gamma_i = \sigma_i / \mu_i$. Then it is easy to see that

$$x(\lambda) = \sum_{i=1}^{p} \frac{\lambda^4 \alpha_i}{\left(\gamma_i^2 + \lambda^2 \right)^2} + \delta_0^2, \qquad y(\lambda) = \sum_{i=1}^{p} \frac{\gamma_i^2 \alpha_i}{\left(\gamma_i^2 + \lambda^2 \right)^2} \tag{1.5}$$

and that for $\lambda > 0$ the derivatives with respect to λ of these functions satisfy

$$x'(\lambda) = 4\lambda^3 \sum_{i=1}^{p} \frac{\gamma_i^2 \alpha_i}{\left(\gamma_i^2 + \lambda^2 \right)^3} > 0, \qquad y'(\lambda) = -4\lambda \sum_{i=1}^{p} \frac{\gamma_i^2 \alpha_i}{\left(\gamma_i^2 + \lambda^2 \right)^3} < 0. \tag{1.6}$$

Moreover, inequalities (1.5) and (1.6) can be used to give

$$\mathrm{d}y / \mathrm{d}x = -1/\lambda^2, \tag{1.7}$$

which shows that y is a monotonically decreasing function of x.

We close this section with the observation that throughout the paper no assumption is made on rank(A) but to ensure a unique solution to (1.2) we shall assume that rank$([A^T L^T]^T) = n$.

2. Tikhonov parameter-choice via fixed-point criterion

As mentioned in the previous section, we shall analyze the behavior of the function $\Psi_\mu(\lambda)$ defined in (1.3) and interpret its extreme points, if they exist, as fixed points of an iteration function to be defined below. Since $\Psi_\mu(\lambda) \geqslant 0$ for $\lambda \geqslant 0$ and $\Psi_\mu(0) = 0$ when $\delta_0 = 0$, we will concentrate on the existence of *nonzero* extreme points. We start by observing that the derivative of Ψ_μ with respect to λ can be written as

$$\Psi_\mu'(\lambda) = y(\lambda)^\mu y'(\lambda) \left[\mu \frac{x(\lambda)}{y(\lambda)} + \frac{x'(\lambda)}{y'(\lambda)} \right]. \tag{2.1}$$

Since $y(\lambda)^\mu y'(\lambda) \neq 0$ and $x'(\lambda)/y'(\lambda) = -\lambda^2$ by (1.7), the above expression for $\Psi_\mu'(\lambda)$ shows that the necessary condition for Ψ_μ to have a local minimum at $\lambda = \lambda^* \neq 0$, $\Psi_\mu'(\lambda^*) = 0$, requires that

$$\lambda^{*2} = \mu \frac{x(\lambda^*)}{y(\lambda^*)} \Leftrightarrow \lambda^* = \sqrt{\mu} \frac{\eta(\lambda^*)}{\xi(\lambda^*)},$$

where $\eta(\lambda) = \sqrt{x(\lambda)}$ and $\xi(\lambda) = \sqrt{y(\lambda)}$. Thus if Ψ_μ achieves a local minimum at $\lambda = \lambda^* \neq 0$, the parameter λ^* must be a fixed point of $\phi : \mathbb{R}_0^+ \mapsto \mathbb{R}_0^+$, defined by

$$\phi(\lambda) = \sqrt{\mu} \frac{\eta(\lambda)}{\xi(\lambda)}, \qquad 0 \leqslant \lambda < \infty. \tag{2.2}$$

To highlight the dependence of ϕ on the parameter μ, for notational convenience we shall write $\phi(\lambda, \mu)$ instead of $\phi(\lambda)$. It is worthwhile noting that ϕ does not need to have a fixed point for arbitrary μ. So in order to know whether or not Ψ_μ has local extreme, we have to investigate the existence of fixed points of function ϕ. Informative results concerning the behavior of ϕ are given in the following lemma.

Lemma 1. *Assume that $\mu = 1$. Set $\underline{\gamma} = \min_i \gamma_i$, $\overline{\gamma} = \max_i \gamma_i$. If $\delta_0 = 0$ and $0 \leqslant \lambda \leqslant \underline{\gamma}$, then*

$$0 \leqslant \phi(\lambda, 1) \leqslant \lambda. \tag{2.3}$$

Furthermore, if $\lambda \geqslant \overline{\gamma}$, then independently of δ_0 it holds

$$\phi(\lambda, 1) \geqslant \lambda. \tag{2.4}$$

Proof. Since for $\lambda \leqslant \underline{\gamma}$ we have $\lambda^4 \alpha_i \leqslant \underline{\gamma}^4 \alpha_i$, this inequality implies

$$x(\lambda) = \sum_{i=1}^{p} \frac{\lambda^4 \alpha_i}{\left(\gamma_i^2 + \lambda^2\right)^2} \leqslant \underline{\gamma}^4 \sum_{i=1}^{p} \frac{\alpha_i}{\left(\gamma_i^2 + \lambda^2\right)^2}. \tag{2.5}$$

It is also immediate to see that for $\lambda \geqslant 0$ we have

$$y(\lambda) \geqslant \underline{\gamma}^2 \sum_{i=1}^{p} \frac{\alpha_i}{\left(\gamma_i^2 + \lambda^2\right)^2}. \tag{2.6}$$

So for $0 \leqslant \lambda \leqslant \underline{\gamma}$, from (2.5) and (2.6) we get $0 \leqslant \phi(\lambda, 1) \leqslant \underline{\gamma}$. Now for fixed $\check{\lambda}$ and $0 \leqslant \lambda \leqslant \check{\lambda} \leqslant \underline{\gamma}$ we have

$$x(\lambda) \leqslant \check{\lambda}^4 \sum_{i=1}^{p} \frac{\alpha_i}{\left(\gamma_i^2 + \lambda^2\right)^2}.$$

This inequality together with (2.6) implies $\phi(\lambda, 1)^2 \leqslant \check{\lambda}^4 / \underline{\gamma}^2 \leqslant \check{\lambda}^2$, which proves inequality (2.3). Inequality (2.4) for the case where $\delta_0 = 0$ is proved in the same way.

Now let us observe that $\forall \lambda > 0$ the inequality $2\gamma_i \lambda \leqslant \gamma_i^2 + \lambda^2$ leads to

$$\frac{\gamma_i^2}{\left(\gamma_i^2 + \lambda^2\right)^2} \leqslant \frac{1}{4\lambda^2}$$

and that for $\lambda \geqslant \overline{\gamma}$ there holds

$$\sum_{i=1}^{p} \frac{\gamma_i^2 \alpha_i}{\left(\gamma_i^2 + \lambda^2\right)^2} \leqslant \frac{1}{4\overline{\gamma}^2} \sum_{i=1}^{p} \alpha_i. \tag{2.7}$$

Also, since for $\lambda \geqslant \overline{\gamma}$ we have $\lambda^4 \geqslant \overline{\gamma}^2 \gamma_i^2$, this inequality implies

$$\sum_{i=1}^{p} \frac{\lambda^4 \alpha_i}{\left(\gamma_i^2 + \lambda^2\right)^2} \geqslant \overline{\gamma}^2 \sum_{i=1}^{p} \frac{\gamma_i^2 \alpha_i}{\left(\gamma_i^2 + \lambda^2\right)^2}. \tag{2.8}$$

Inequalities (2.7) and (2.8) give for $\lambda \geqslant \overline{\gamma}$

$$\phi(\lambda, 1)^2 \geqslant \overline{\gamma}^2 \left(1 + 4 \frac{\delta_0^2}{\sum_{i=1}^{p} \alpha_i}\right) \geqslant \overline{\gamma}^2,$$

and a reasoning similar to that used to prove (2.3) leads to assertion (2.4). $\qquad\square$

Lemma 1 is useful because it allows us to locate fixed points of ϕ when they exist, and because it helps us to classify local extreme points of Ψ_μ. This is the subject of the following theorem.

Theorem 1. *Assume that $\mu = 1$. Let I_1 and I_2 be open intervals such that $\phi(\lambda, 1) < \lambda$, $\forall \lambda \in I_1$ and $\phi(\lambda, 1) > \lambda, \forall \lambda \in I_2$. Then Ψ_μ is increasing in I_1 and decreasing in I_2. In addition to this, the following assertions hold.*

(a) If $\delta_0 = 0$ and $\alpha_i \neq 0, i = 1, \ldots, p$, there exists $\check{\lambda} \in]\underline{\gamma}, \overline{\gamma}[$ such that Ψ_μ achieves a local maximum at $\check{\lambda}$. Additionally, if $\check{\lambda}$ is the fixed point of ϕ nearest to zero and Ψ_μ achieves a local minimum at λ^*, then $\check{\lambda} < \lambda^*$ and Ψ_μ achieves another local maximum inside $]\lambda^*, \overline{\gamma}[$.

(b) *If $\delta_0 \neq 0$ and Ψ_μ achieves a minimum at λ^*, there exists a parameter $\check{\lambda}$ inside $]\lambda^*, \overline{\gamma}[$ at which Ψ_μ achieves a local maximum.*

(c) *Let ϕ have a fixed point λ^* and let P be the point on the L-curve in log–log scale associated with λ^*. Then the L-curve is convex in a neighborhood of P if and only if λ^* locally minimizes Ψ_μ, and it is concave in a neighborhood of P if and only if λ^* locally maximizes Ψ_μ.*

Proof. By property (1.7) the derivative of $\Psi_\mu(\lambda)$ in (2.1) can be rewritten as

$$\Psi_\mu'(\lambda) = y(\lambda)^\mu y'(\lambda)(\phi(\lambda,1)^2 - \lambda^2). \tag{2.9}$$

But since $y(\lambda)^\mu y'(\lambda) < 0$ by (1.6), we conclude that $\Psi_\mu'(\lambda) > 0$ whenever $\lambda \in I_1$ and hence that Ψ_μ increases with λ in I_1. The assertion concerning the decreasing of Ψ_μ follows similarly.

To prove (*a*) we first observe that if $\delta_0 = 0$, neither $\underline{\gamma}$ nor $\overline{\gamma}$ can be fixed points of ϕ, at least not in the context of discrete ill-posed problems. In fact, if λ^* is a fixed point of ϕ then

$$x(\lambda^*) = \lambda^{*2}y(\lambda) \Leftrightarrow \sum_{i=1}^{p} \frac{\lambda^{*4}\alpha_i}{\left(\gamma_i^2 + \lambda^{*2}\right)^2} = \lambda^{*2} \sum_{i=1}^{p} \frac{\gamma_i^2 \alpha_i}{\left(\gamma_i^2 + \lambda^{*2}\right)^2} \Leftrightarrow \sum_{i=1}^{p} \frac{\left(\lambda^{*2} - \gamma_i^2\right)\alpha_i}{\left(\gamma_i^2 + \lambda^{*2}\right)^2} = 0. \tag{2.10}$$

The last equality in (2.10) will remain valid with $\lambda^* = \underline{\gamma}$ only when $\gamma_i = \underline{\gamma}$ for $i = 1, \ldots, p$. But this cannot happen in our context since by assumption the generalized singular values decay to zero without particular gap. Therefore $\underline{\gamma}$ cannot be a fixed point of ϕ. A similar argument explains why $\overline{\gamma}$ cannot be a fixed point of ϕ. From this observation and lemma 1 it then follows that ϕ has at least a fixed point in $]\underline{\gamma}, \overline{\gamma}[$. The statements in (a) follow on using this result together with (2.3), (2.4) and (2.9).

To prove (*b*) we also observe that in this case $\overline{\gamma}$ cannot be a fixed point of ϕ. Observe now that because of lemma 1 and the fact that $y(\lambda) \to 0$ as $\lambda \to \infty$, which is immediate from (1.6), it follows that Ψ_μ monotonically decreases when $\lambda > \overline{\gamma}$ and $\Psi_\mu \to 0$ as $\lambda \to \infty$. This shows that local extreme of Ψ_μ, if any, must be achieved inside the open interval $]0, \overline{\gamma}[$. The statements described in (*b*) are immediate consequences of this observation and the monotonicity properties of Ψ_μ.

The first statement in (*c*) is theorem 1 in [15]. To prove the second statement set $u(\lambda) = \log(x(\lambda))$ and $v(\lambda) = \log(y(\lambda))$. Then

$$\frac{dv}{du} = -\frac{\phi(\lambda,1)^2}{\lambda^2} \qquad \text{and} \qquad \frac{d^2v}{du^2}\frac{du}{d\lambda} = -\frac{2}{\lambda^2}\phi(\lambda,1)\left[\phi'(\lambda,1) - \frac{\phi(\lambda,1)}{\lambda}\right]. \tag{2.11}$$

From calculus and (2.9) it is not hard to prove that $\phi'(\lambda^*, 1) > 1$ in a neighborhood (small enough) of λ^* if and only if λ^* maximizes Ψ_μ. From this result and the fact that $du/d\lambda = x'(\lambda)/x(\lambda) > 0$ by (1.6), the second statement in (*c*) follows on using (2.11). This concludes the proof. $\qquad \square$

We now discuss the existence of local minima for Ψ_μ. A first step toward this was done in ([15], theorem 2), where existence of local minimum inside the interval $]0, \infty[$ is guaranteed to exist under very special conditions on A. A definitive answer for the case where $\delta_0 \neq 0$ and free of any requirement on either A or g is provided in the following corollary.

Corollary 1. *If $\delta_0 \neq 0$ there exists $\mu^* > 0$ such that Ψ_{μ^*} achieves a local minimum in $]0, \overline{\gamma}[$.*

Proof. We first recall that $\phi(0, \mu) > 0$. Assume that $\mu = 1$. If $\phi(\bar{\lambda}, 1) < \bar{\lambda}$ for some $\bar{\lambda} \in]0, \overline{\gamma}[$, by continuity of ϕ it follows that $(\phi(\lambda, 1) - \lambda)$ changes sign in $]0, \bar{\lambda}[$. Thus there
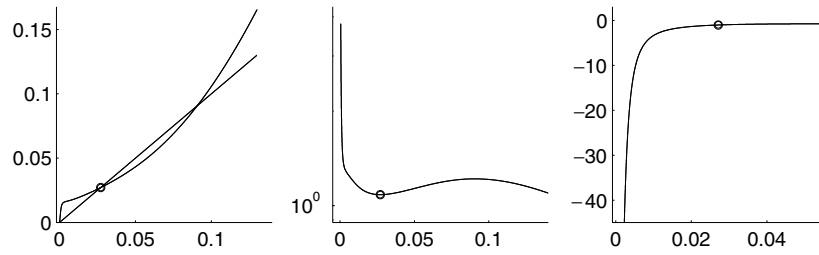
Figure 1. (Left) Function $\phi(\lambda, 1)$ and the straight line $z = \lambda, 0 \leqslant \lambda \leqslant \overline{\gamma}$. (Center) Plot of $z = \log \Psi_\mu(\lambda), 0 \leqslant \lambda \leqslant \overline{\gamma}$ with $\mu = 1$. (Right) $m_L(\lambda)$ for λ near the minimizer λ^* of Ψ_μ. The small circle denotes the point associated with the minimizer λ^*.

exists λ^* such that $\phi(\lambda^*, 1) = \lambda^*$ with the property $\phi(\lambda, 1) > \lambda$ for all λ in some interval to the left of λ^* and $\phi(\lambda, 1) < \lambda$ for all λ in some interval to the right of λ^* (see figure 1). This means that Ψ_μ is decreasing in some interval to the left of λ^* and Ψ_μ is increasing in some interval to the right of λ^* (see theorem 1). Consequently, for $\mu = 1$ the function Ψ_μ achieves a local minimum at $\lambda^* \in \,]0, \overline{\gamma}[$.

Suppose now that $\phi(\lambda, 1) > \lambda$ for all $\lambda > 0$. Let the slope of the L-curve in log–log scale at the point $(\log x(\lambda), \log y(\lambda))$ be denoted by $m_L(\lambda)$. Then, since $m_L(\lambda) = -\phi(\lambda, 1)^2/\lambda^2$ by (2.11), we have that $\phi(\lambda, \mu)^2 = \lambda^2$ with $\mu = -1/m_L(\lambda)$. This shows that *any* $\lambda > 0$ can be regarded as a fixed point of $\phi(\lambda, \mu)$ as long as μ is the negative reciprocal of $m_L(\lambda)$. To conclude the proof, it suffices to select any λ^* at which the L-curve is convex, which is always guaranteed to exist near zero (see, e.g., lemma 3 in [15]), and then define $\mu = -1/m_L(\lambda^*)$. This implies that $\phi(\lambda^*, \mu^*) = \lambda^*$. The proof concludes on using item (c) of theorem 1.   □

The fixed-point criterion for choosing the regularization parameter relies on the heuristic that the fixed point λ^* that minimizes $\Psi_\mu(\lambda)$ leads to a point near the corner of the L-curve in log–log scale. The key ingredient that supports the heuristic is the behavior of the slope, $m_L(\lambda) = -\phi(\lambda, 1)^2/\lambda^2$. To explain this more precisely, assume that $\phi(\lambda, 1)$ behaves as in figure 1 so that Ψ_μ is minimized at the first fixed point of ϕ, say $\lambda = \lambda^*$. We then see that $m_L(\lambda^*) = -1$, and that for $\lambda < \lambda^*$ the slope $m_L(\lambda)$ rapidly reaches *large* negative values, whereas for $\lambda > \lambda^*$ the slope moderately changes so that $-1 < m_L(\lambda) < 0$ or $a < m_L(\lambda) < b < 0$ where a and b do not differ very much from -1 (see figure 1). Thus the interval $]0, \lambda^*[$ corresponds to the vertical part of the L-curve, the interval $]\lambda^*, \overline{\gamma}[$ corresponds to the flat part of the L-curve and the minimizer λ^* corresponds to a point in the vicinity of the corner of the L-curve.

We observe that for L-curves with a distinct corner separating the vertical and horizontal parts, the L-curve corner might as well be defined as the point for which $m_L(\lambda) = -1$. However, although intuitively acceptable, this cannot be extended for general problems since, as we will see, there are cases where $m_L(\lambda) < -1$ for all $\lambda > 0$. We postpone the analysis of these cases to the following section.

We emphasize that an *a priori* parameter choice due to Miller ([13], method 2) is related to the proposed criterion. If we assume that the (semi)norm of the exact solution is α and the norm of the error $g - g^{\mathrm{exact}}$ is δ, Miller's method suggests the choice $\lambda = \delta/\alpha$. Thus if the residual norm associated with the regularized solution and $\|Lf_\lambda\|_2$ are good estimates of δ and α, respectively, which is reasonable when λ is close to optimal, then Miller's method leads to the fixed-point criterion $\lambda^* = \phi(\lambda^*, 1)$, thereby providing an alternative way to support the proposed criterion. Further details about Miller's method in connection with discrete ill-posed

Fixed-point algorithm
Input: $\overline{\gamma}$, *tol*
1. Set $\mu = 1$, $k = 0$, and choose λ_0 small enough such that $\phi(\lambda_0, \mu) > \lambda_0$.
2. Compute $s_0 = \phi(\lambda_0, \mu), \lambda_0$.
3. **while** $(|s_k - 1| > tol \ \& \ \lambda_k < \overline{\gamma})$ **do**
        $\lambda_{k+1} = \phi(\lambda_k, \mu), \quad s_k = \lambda_{k+1}/\lambda_k$
        k = k+1
    end while
4. **if** $(\lambda_k < \overline{\gamma} \ \& \ \phi'(\lambda_k, 1) < 1)$ **do**
        $\lambda^* = \lambda_k,$
    elseif $(\lambda_k < \overline{\gamma} \ \& \ \phi'(\lambda_k, 1) = 1)$ **do**
        $\sqrt{\mu} = 2\overline{\gamma}/(\phi(\overline{\gamma}, 1) + \overline{\gamma})$
        Set $\lambda_0 = \lambda_k$, $k = 0$, and go to step 2
    else do
        Find k^* such that $s_{k^*} = \min s_k$.
        Find $\theta \gtrapprox s_{k^*}$ (e.g., the smallest integer larger than s_{k^*}),
        Choose $\sqrt{\mu} = 2/(s_{k^*} + \theta)$,
        Set $\lambda_0 = \lambda_{k^*}$, $k = 0$, and go to step 2
    end if

Figure 2. Fixed-point algorithm for determining Tikhonov regularization parameter.

problems can be found in [8, section 7.5.3]. Another related criterion is the one that looks for the roots of the function $B(\lambda) = -x(\lambda) + \lambda^2 y(\lambda)$. This method is known as the zero-crossing method [10]. Note that searching for fixed points of $\phi(\lambda, 1)$ amounts to searching for roots of $B(\lambda)$.

3. Proposed algorithm and convergence analysis

Given an initial guess, λ_0, our algorithm starts by setting $\mu = 1$ as a default value and proceeds by computing the sequence $\lambda_{k+1} = \phi(\lambda_k, 1), k = 0, 1, \ldots,$ until convergence is reached or until divergence is detected. When convergence is reached, a fixed point λ^* is computed and a test for convexity of the *L*-curve at $(\log x(\lambda^*), \log y(\lambda^*))$ based on the condition $\phi'(\lambda^*, 1) < 1$ is made (see (2.11)). When divergence is detected or when convergence is reached and the convexity test is not satisfied, the parameter μ is adjusted in a manner to be specified in section 3.2 and the iterations restart. A brief description of our algorithm is depicted in figure 2.

In order to make the algorithm clear, we shall give a detailed description of quantities such as input parameters and others.

3.1. Initial guess and input parameters $\overline{\gamma}$ and tol

A critical issue when using fixed-point-based algorithms is the proper choice of λ_0. In our context the choice of the initial guess poses no difficulty when $\delta_0 \neq 0$ since independently of μ we have $\phi(0, \mu) > 0$; hence, it suffices to choose $\lambda_0 \approx 0$. The choice of λ_0 for the case when $\delta_0 = 0$ is not as simple as before but an upper bound for γ can be appropriate.

As for the parameter $\overline{\gamma}$, it must be estimated efficiently when the G(SVD) is not available. A way to do this is by means of the Lanczos method applied to $[A^T L^T]^T$, as done, e.g., in *pnu* in the Toolbox Regularization Tools by Hansen [9].

The parameter *tol* plays the role of convergence test parameter; it points out convergence when the relative change in λ_k: $|\lambda_{k+1} - \lambda_k|/|\lambda_k| = |s_k - 1|$ is small enough. It is worth
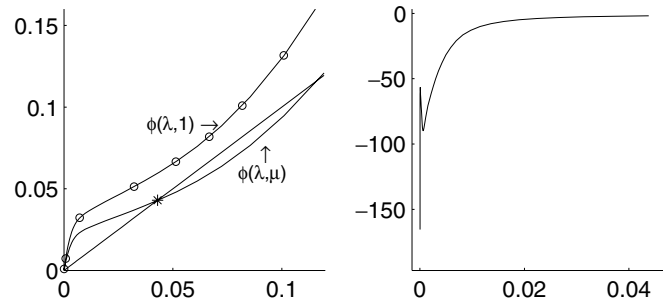
Figure 3. (Left) Function $\phi(\lambda, 1)$ with no fixed point and the adjusted one $\phi(\lambda, \mu)$ with $\mu < 1$. The small circles denote a few points of the diverging sequence of points associated with λ_k (for $\mu = 1$). (Right) Slope $m_L(\lambda)$.

emphasizing that when the else loop at step 4 is not visited, which means we have found the fixed point λ^* nearest to λ_0, the FP algorithm finds out a point on the *L*-curve for which the slope is $m_L(\lambda^*) = -1/\mu = -1$ and therefore located in the vicinity of the 'corner'.

3.2. Choice of parameter μ

Clearly, only the case $\delta_0 \neq 0$ needs to be discussed. When $\phi(\lambda, 1)$ has no fixed point the sequence of values λ_k must diverge and, as described above, the algorithm interrupts computations after λ_k exceeds $\overline{\gamma}$, and then restarts with a parameter μ chosen such that the function $\phi(\lambda, \mu)$ is guaranteed to have a fixed point in $]0, \overline{\gamma}[$. In order to choose the parameter μ, we first recall that if $\phi(\lambda, 1)$ has a fixed point λ^* that locally minimizes Ψ_μ, then m_L is *maximized* at a point that is to the right of λ^* (more precisely, located between two fixed points of $\phi(\lambda, 1)$), with the maximum value of $m_L(\lambda)$ being between $m_L(\lambda^*) = -1$ and 0, see figure 1 and the discussion after corollary 1. This property is not valid when $\phi(\lambda, 1)$ has no fixed point because $m_L(\lambda) < -1$ for all $\lambda > 0$, but the slope $m_L(\lambda)$ continues to behave like the 'ideal situation' where $\phi(\lambda, 1)$ has a fixed point that minimizes Ψ_μ, see figure 3 and compare with figure 1. Our choice of μ relies on this observation.

Before describing our strategy of selecting the parameter μ when $\phi(\lambda, 1)$ fails to have a fixed point, observe that because $\phi(\lambda, \mu) = \sqrt{\mu}\phi(\lambda, 1)$ (see definition (2.2)), for $\phi(\lambda, \mu)$ to have a fixed point in the interval $]0, \overline{\gamma}[$, the parameter μ must be smaller than 1; this is graphically illustrated in figure 3.

Now note that if λ^* is a fixed point of $\phi(\lambda, \mu)$, then $\sqrt{\mu} = \lambda^*/\phi(\lambda^*, 1)$ and $m_L(\lambda^*) = -1/\mu$. As a consequence, whenever $\phi(\lambda, 1) > \lambda$ for all $\lambda > 0$, *any* $\lambda > 0$ can be regarded as a fixed point of $\phi(\lambda, \mu)$ with μ being defined by $\sqrt{\mu} = \lambda/\phi(\lambda, 1)$. Naturally, $\phi(\lambda, \mu)$ does not necessarily need to have a fixed point for *any* $\mu < 1$. This can be explained as follows. Define $\varphi(\lambda) = \lambda/\phi(\lambda, 1)$ and assume that the function φ achieves a maximum inside $]0, \overline{\gamma}[$, say ω. It is straightforward to see that if $\omega < \sqrt{\mu} < 1$, then $\phi(\lambda, \mu)$ has no fixed point since, $\phi(\lambda, \mu) > \lambda$ for all $\lambda > 0$. We thus conclude for $\phi(\lambda, \mu)$ to have a fixed point in $]0, \overline{\gamma}[$, the parameter $\sqrt{\mu}$ cannot exceed the maximum value of $\varphi(\lambda)$.

In order to describe our choice of μ, recall that maximizing $\varphi(\lambda)$ amounts to maximizing $m_L(\lambda)$, which in turns amounts to minimizing $\phi(\lambda, 1)/\lambda$. Our algorithm minimizes approximately the function $\phi(\lambda, 1)/\lambda$ and proposes as parameter μ the one calculated by $\sqrt{\mu} = 2/(s_k^* + \theta)$, where θ is the smallest integer larger than s_k^*. The main virtue of this choice is that the parameter λ^* so computed yields a slope that does not differ very much

from -1 (recall that $m_L(\lambda^*) = -1/\mu$), e.g., if $1 < s_k^* < 2$, $-4 < m_L(\lambda^*) < -9/4$, thereby ensuring that the computed λ^* does not correspond to the vertical part of the L-curve where the perturbation errors dominate and therefore avoiding local corners with large slopes, as illustrated in figure 5. Another characteristic of our choice, revealed by numerical experiments with discrete ill-posed problems, is that the parameter μ is relatively insensitive to small changes in θ, this being valid for the associated parameter λ^* as well.

It remains to explain the choice of μ for the case where $\phi(\lambda, 1) \geqslant \lambda$ for all $\lambda > 0$, that is, the case where the straight line $z = \lambda$ is tangent to the curve $z = \phi(\lambda, 1)$ at some point $\lambda^* \in \,]0, \overline{\gamma}[$. Although this case rarely happens, we need to consider it. Obviously, the choice of μ according to the above proposal might be used, in which case the selected parameter would be $\mu = 4/9$. We note, however, that this choice might not be suitable for all problems since the parameter μ is maintained fixed. In order to allow certain flexibility when choosing μ, while incorporating information of the problem, we propose the choice $\sqrt{\mu} = 2\overline{\gamma}/(\phi(\overline{\gamma}, 1) + \overline{\gamma})$. Geometrically, $\sqrt{\mu}/\overline{\gamma}$ is the reciprocal of the distance of the point $(\overline{\gamma}, 0)$ to the point $Q = (P_1 + P_2)/2$, where $P_1 = (\overline{\gamma}, \overline{\gamma})$ and $P_2 = (\overline{\gamma}, \phi(\overline{\gamma}, 1))$.

We now give a theorem that describes the convergence properties of our algorithm.

Theorem 2. *Assume that $\delta_0 \neq 0$, that λ_0 is selected according to step 1 of the FP algorithm, and that no fixed point of $\phi(\lambda, 1)$ is to the left of λ_0. Assume further that if $\phi(\lambda, 1)$ has no fixed point in $\,]0, \overline{\gamma}[$, the finite sequence of values s_k generated by the FP algorithm achieves a minimum that differs from s_0. Then, the sequence λ_k converges either to the closest fixed point of $\phi(\lambda, 1)$ which locally minimizes Ψ_μ for $\mu = 1$, or to a fixed point of $\phi(\lambda, \mu)$ with $\mu < 1$ which locally minimizes Ψ_μ.*

Proof. Assume that $\phi(\lambda, 1)$ has a fixed point λ^* near λ_0. Then, since $\phi'(\lambda, 1) > 0$ for all $\lambda > 0$ by (1.6), it follows that λ_k is an increasing sequence and that $|\lambda_{k+1} - \lambda^*| < |\lambda_k - \lambda^*|$. This is sufficient to ensure that $\lambda_k \to \lambda^*$ as $k \to \infty$.

Assume now that $\phi(\lambda, 1)$ has no fixed point. Then a finite sequence of values s_k is generated from which an integer k^* and a parameter $\mu^* < 1$ are computed. Using the definition of μ^* according to the algorithm we have

$$\phi(\lambda_{k^*}, \mu^*) = \sqrt{\mu^*}\phi(\lambda_{k^*}, 1) < \lambda_{k^*},$$

and a continuity argument shows then that $\phi(\lambda, \mu^*)$ has a fixed point that is to the left of λ_{k^*}. Restarting the algorithm with λ_{k^*} as initial guess, the sequence λ_k so generated converges to the nearest fixed point because of the fact that $\phi(\lambda, \mu)$ is an increasing function of λ. This concludes the proof. $\qquad\square$

4. Numerical results

To evaluate the performance of our FP algorithm we have solved three discrete ill-posed problems: two test problems from Hansen's regularization toolbox, and one from image super-resolution. For each problem we ran 500 instances with different vectors $g = g^{\text{exact}} + e$ where e contains zero-mean Gaussian random numbers (with the seed value of the random generator set to zero), and report average relative errors and standard deviations. Parameter θ at step 4 is chosen as the smallest integer larger than s_{k^*}. For comparison, we also report results associated with L-curve, GCV, and another fixed-point-based algorithm due to Belge *et al* [1], which we call BFP. The BFP method selects the regularization parameter α (Belge uses symbol α instead of λ^2 in (1.2)) by minimizing the so-called distance function $v(\alpha) = |\log x(\alpha) - a|^2 + |\log y(\alpha) - b|^2$, where $(a, b) = (\log x(\underline{\gamma}^2), \log y(\overline{\gamma}^2))$ is an origin, which also leads to a fixed point procedure.
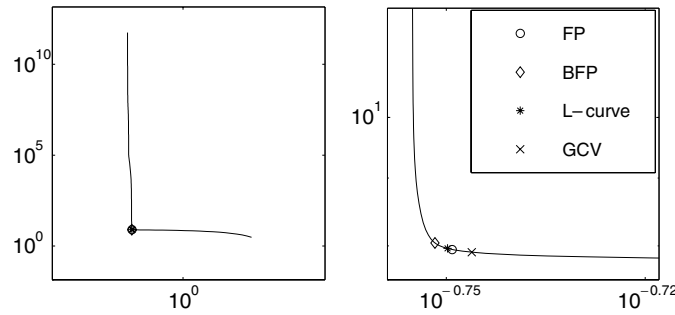
Figure 4. Tikhonov *L*-curve for Shaw problem. The right figure shows a close-up view of the points on the *L*-curve associated with the regularization parameters found by FP, BFP, *L*-curve and GCV.

Our implementation was made in MATLAB using the SVD(GSVD) and functions *l_curve.m* and *gcv.m* from the regularization tools by Hansen [9]. FP and BFP were initialized with $\lambda_0 = 10^{-4}$ and terminated when $|\lambda_{k+1} - \lambda_k| \leqslant \lambda_k 10^{-4}$. In our description we use the following symbols:

- SR: successful runs (%);
- $\bar{E}_f$ (resp. $\bar{\lambda}$): average values of relative error in f_λ (resp. computed Tikhonov parameter);
- $E_{\max}$ (resp. $E_{\min}$): maximum error (resp. minimum) occurring in all realizations;
- $I_{\max}$ (resp. $I_{\min}$): maximum (resp. minimum) number of iterations;
- STD: standard deviation of computed errors,
- NL: noise level defined by $\|g - g^{\text{exact}}\|_2 / \|g^{\text{exact}}\|_2$.

Example 1 Shaw test problem**.** We apply Tikhonov regularization in standard form (i.e., $L = I$) to a problem with matrix A and vector g^{exact} generated by using the function *Shaw.m* from [9], with $n = 64$. Matrix A is numerically rank-deficient with rank$(A) = 20$ and effective condition number $\kappa(A) = \sigma_1/\sigma_{20} = 5.2759e + 12$. This test problem has been analyzed in several places and its Tikhonov *L*-curve is known to be well behaved (see, e.g., figure 5.3 in Hansen's book [9], and figure 4). Numerical results displayed in table 1 show that with the exception of the GCV method, all remaining methods perform very well, with a slight advantage in favor of FP. For this test problem, a GCV solution is considered successful when the corresponding error does not exceed the maximum error associated with the other methods. We see that among the tested methods, the one that estimates more consistently (smallest variance) the regularization parameters is FP, while the one in the opposite direction (largest variance) is BFP. We also see that the smallest average error corresponds to FP and that the largest one corresponds to BFP. For this test problem the choice $\mu = 1$ works well in all the instances and the else loop at step 4 of our algorithm is *not* visited.

Example 2 (Helio test problem)**.** This example considers Tikhonov method in general form applied to a problem from inverse helioseismology. Matrix A in this problem is 212×100 and matrix L is a discrete second-order differential operator of size 98×100. Matrices A, L and vectors f^{exact}, g^{exact} are all generated by function *Helio.m*; for further information the reader is referred to [6]. A discussion about the performance of several methods for this problem is in [8, chapter 7, section 7.7.1]; there are reported results for noisy data involving e with standard deviation $\sigma_0 = 10^{-2}$. Here, we report results for $\sigma_0 = 2 \times 10^{-2}$ (which implies NL $\approx 5\%$);
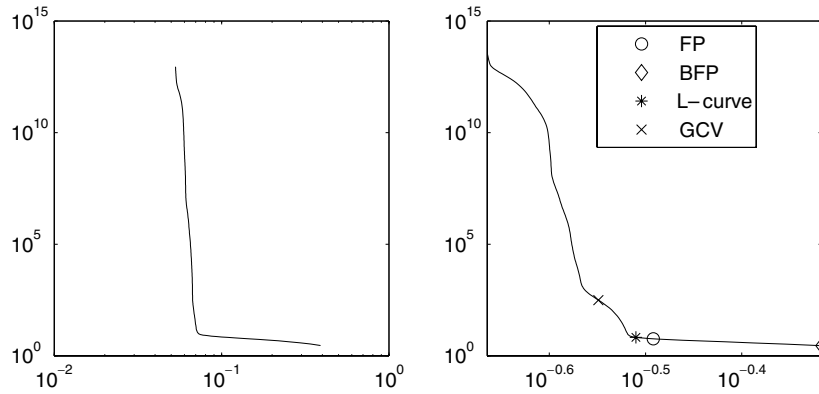
Figure 5. (Left) L-curve for Helio problem for data at low-noise level. (Right) L-curve for Helio problem and points on the L-curve generated by FP, BFP, L-curve and GCV, for NL = 5%.

Table 1. Numerical results for Shaw test problem.

			NL = 1%			NL= 5%		
	FP	BFP	L-curve	GCV	FP	BFP	L-curve	GCV
SR	100%	100%	100%	85.80%	100%	100%	100%	82%
$\overline{E}_f$	0.1213	0.1585	0.1281	0.1315	0.1728	0.2193	0.1738	0.2087
E_{max}	0.2985	0.4011	0.3382	0.3176	0.2867	0.5653	0.2771	0.2837
E_{min}	0.0475	0.0511	0.0466	0.0493	0.0939	0.0697	0.0929	0.0866
STD	0.0410	0.0555	0.0453	0.0424	0.0335	0.0821	0.0388	0.0908
$\overline{\lambda}$	0.0221	0.0122	0.0189	0.0329	0.1155	0.0622	0.0997	0.0860
I_{max}	11	8	–	–	12	11	–	–
I_{min}	5	4	–	–	7	5	–	–

in this case the associated L-curve is not well behaved and with convex and concave parts inside the interval $[\underline{\gamma}, \overline{\gamma}]$ (see figure 5), which makes the identification of the corner difficult. This problem provides an excellent way to illustrate that the Tikhonov parameter choice via BFP can be very sensitive to the origin (a,b). We illustrate this point by calculating Tikhonov parameters for three distinct choices of the origin; results displayed in columns BFP(1) and BFP(2) in table 2 are apparent. Also, we see that the best average relative error corresponds to FP and that the worst one corresponds to BFP. The computed standard deviations, on the other hand, show that FP computes solutions in a much more consistent way than the other methods. The line labeled with symbol $\overline{\mathcal{E}}_f$ contains average relative errors of solutions whose first 12 elements are discarded (as done in [8]). The results do not change so much the trend of the methods. For this example the choice $\mu = 1$ fails to work in all the instances and the else loop at step 4 of our algorithm is *always* visited.

Example 3 (super-resolution image reconstruction problem). We consider the problem of estimating a high-resolution (HR) image from observed multiple low-resolution (LR) images. Let **f** be the HR image of size $M = M_1 \times M_2$ written in lexicographic notation as $\mathbf{f} = [f_1, \ldots, f_M]^T$, where M is the number of pixels of the HR image, and let the kth LR image of size $N = N_1 \times N_2$ be denoted in lexicographic notation as $\mathbf{g}_k = [g_{k,1}, g_{k,2}, \ldots, g_{k,N}]^T$, $k = 1, 2, \ldots, p$, with $M_1 = N_1 \times L_1$, $M_2 = N_2 \times L_2$, where
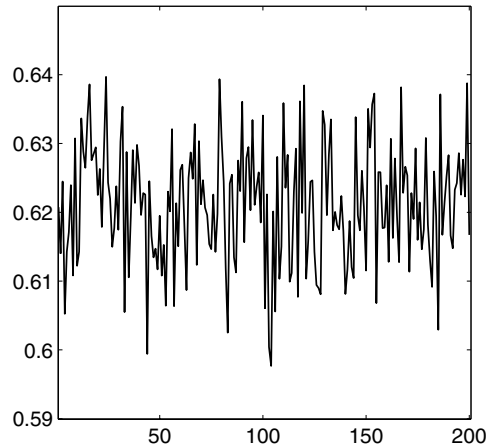
Figure 6. Behavior of $\sqrt{\mu}$ computed by the FP algorithm of first 200 runs for Helio problem.

Table 2. Numerical results for Helio problem. Columns BFP, BFP(1) and BFP(2), contain values corresponding to origins $(\log x(\underline{\gamma}^2), \log y(\overline{\gamma})^2)$, $(\log x(\underline{\gamma}^2), \log y(1.2\overline{\gamma}^2))$, and $(\log x(\underline{\gamma}^2), \log y(1.4\overline{\gamma}^2))$, respectively.

	FP	BFP	BFP(1)	BFP(2)	L-curve	GCV
SR	100%	100%	100%	100%	99.20%	99%
$\overline{E}_f$	0.1540	0.2015	0.2470	0.3005	0.1883	0.2190
E_{max}	0.1946	0.2536	0.3024	0.3585	0.2949	0.2989
E_{min}	0.1454	0.1623	0.1963	0.2489	0.1470	0.0709
$\overline{\mathcal{E}}_f$	0.0794	0.1418	0.1705	0.2035	0.0603	0.0508
STD	0.0074	0.0152	0.0178	0.0174	0.0255	0.059
$\overline{\lambda}$	0.0319	0.073	0.0973	0.1163	0.0190	0.0059
I_{max}	25	21	24	12	–	–
I_{min}	18	4	4	6	–	–

L_1 and L_2 represent down-sampling factors in the observation model for the horizontal and vertical directions, respectively. Assuming that the acquisition process of the LR sequence involves blurring, motion, subsampling and additive noise, an observation model that relates **f** to $\mathbf{g}_k$ is written as [14]

$$\mathbf{g}_k = \mathbf{C}_k \mathbf{f} + \epsilon_k \tag{4.1}$$

where $\mathbf{C}_k$ is $N \times M$, and ϵ_k stands for noise. The HR image problem is to estimate **f** from all LR images $\mathbf{g}_k$. In this context, the Tikhonov method consists of solving the optimization problem

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{f}} \|\mathbf{g} - \mathbf{C}\mathbf{f}\|_2^2 + \lambda^2 \|\mathbf{R}\mathbf{f}\|_2^2 \tag{4.2}$$

where $\mathbf{g} = \left[\mathbf{g}_1^T \ldots \mathbf{g}_p^T\right]^T$, $\mathbf{C} = \left[\mathbf{C}_1^T \ldots \mathbf{C}_p^T\right]^T$, **R** is a discrete first-order differential operator in the horizontal and vertical directions, and λ is the regularization parameter.

We address the problem of estimating the 18×24 HR image depicted in figure 7 from a sequence of five noisy LR images with downsampling factors $L_1 = L_2 = 2$. Thus **C** is of size 540×432 and **R** is of size 822×432. The noisy LR images were obtained adding white noise
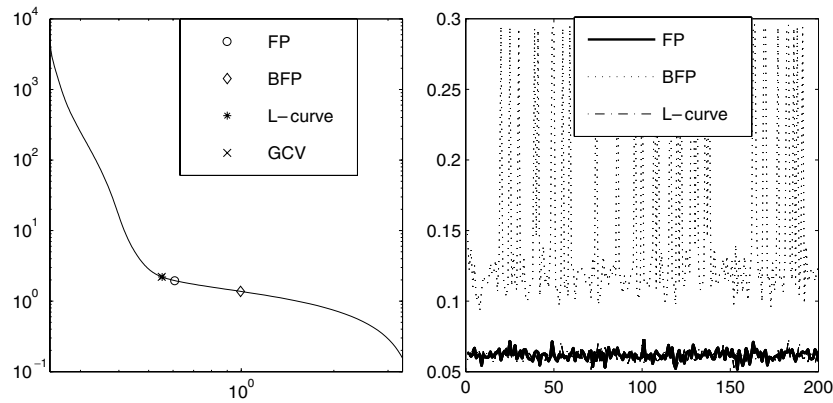
Figure 7. (Left) *L*-curve for HR problem. (Right) Relative errors corresponding to first 200 runs.

Table 3. Numerical results for HR estimation problem.

	FP	BFP	*L*-curve	GCV
$\overline{E}_f$	0.0617	0.1471	0.0615	0.0622
$E_{\max}$	0.0748	0.2975	0.0751	0.0911
STD	0.0034	0.0670	0.0035	0.0045
$\overline{\lambda}$	0.3318	1.7753	0.1968	0.1944
$I_{\max}$	26	386	–	–
$I_{\min}$	19	21	–	–

ϵ to the degraded images so that SNR = 15dB; two LR images are depicted in figure 7. For this problem we verified that rank($\mathbf{C}$) = 390 (to machine precision) and that rank($\mathbf{R}$) = 341. Since $\mathbf{R}$ is overdetermined, we compute the QR decomposition of $\mathbf{R}$, $\mathbf{R} = QR$, and replace $\|\mathbf{Rf}\|_2$ by $\|L\mathbf{f}\|_2$ in (4.2) where L is the top block of $\mathbf{R}$ of size 341×342. We also verified that the associated *L*-curve is not as well behaved as in example 1, and that the *L*-shape deteriorates significantly as the noise level increases. The *L*-curve for noise level of 15 dB of the last run is depicted on the left of figure 7; on the right of the same figure are displayed relative errors corresponding to the first 200 runs. The estimated images obtained by means of FP, BFP and *L*-curve of the last run are depicted on the second line in figure 8. The estimated image obtained by GCV is very similar to those obtained by FP and *L*-curve and it is not displayed there. Visual inspection shows that the estimated images obtained by FP and *L*-curve are much more accurate than that obtained by BFP, which is also verified from the average relative errors shown in table 3. Table 3 also shows for this problem that the number of iterations needed for BFP to reach convergence is significantly larger than that for FP and that BFP calculates solutions that can vary very much from one run to another (as explained by the large value of the associated standard deviation and by the relative errors depicted in figure 7). For this problem the choice $\mu = 1$ fails to work in approximately 5% of the cases.

5. Conclusions

We proposed a fixed point algorithm that selects a Tikhonov parameter associated with a point near the 'corner' of the *L*-curve in log–log scale, thus producing solutions with accuracy
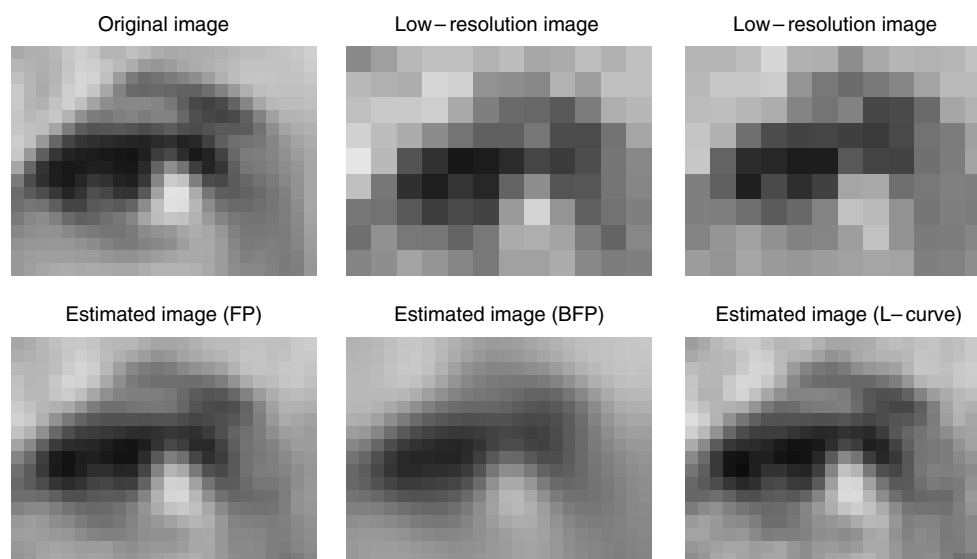
Figure 8. HR image, two LR images and estimated images by FP, BFP and *L*-curve method.

comparable to that of the *L*-curve method but at a lower computational cost. Numerical results suggest that the FP algorithm can be much more robust than BFP, and that BFP can be very sensitive to the choice of the origin specially when the associated *L*-curve is not well behaved. Although all computations were performed satisfactorily using the SVD(GSVD), we are aware that this may not be the case when the problem is too large. However, we expect the ideas behind the algorithm can be implemented efficiently in conjunction with iterative regularization methods such as those based on LSQR and GC [12]. Less heuristic choices of μ should be investigated. The extension of FP to the selection of multiple regularization parameters is also of interest and is the subject of future research.

Acknowledgments

References

[1] Belge M, Kilmer M E and Miller E L 2002 Efficient determination of multiple regularization parameters in a generalized *L*-curve framework *Inverse Problems* **18** 1161–83
[2] Engl H W and Grever W 1994 Using the *L*-curve for determining optimal regularization parameters *Num. Math.* **69** 25–31
[3] Hämarik U and Raus T 2006 On the choice of the regularization parameter in ill-posed problems with approximately given noise level of data *J. Inverse Ill-Posed Problems* **14** 251–66
[4] Hämarik U, Palm R and Raus T 2007 Use of extrapolation in regularization methods *J. Inverse Ill-Posed Problems* **15** 277–94

 [5] Hanke M 1996 Limitations of the *L*-curve method in ill-posed problems *BIT* **36** 287–301
 [6] Hanke M and Hansen P C 1993 Regularization methods for large-scale problems *Surv. Math. Ind.* **3** 253–315
 [7] Hansen P C 2001 The *L*-curve and its use in he numerical treatment of inverse problems *Computational Inverse Problems in Electrocardiology* ed P Johnston (Southampton: WIT Press) pp 119–42 (invited chapter)
 [8] Hansen P C 1998 *Rank-Deficient and Discrete Ill-Posed Problems* (Philadelphia, PA: SIAM)
 [9] Hansen P C 1994 Regularization tools: a MATLAB package for analysis and solution of discrete ill-posed problems *Numer. Algorithms* **6** 1–35
[10] Johnston P R and Gulrajani R M 2002 An analysis of the zero-crossing method for choosing regularization parameter *SIAM J. Sci. Comput.* **24** 428–42
[11] Krawczy-Stando D and Rudnicki M 2007 Regularization parameter selection in discrete-ill posed problems—the use of the *U*-curve *Int. J. Appl. Math. Comput. Sci.* **17** 157–64
[12] Kilmer M E and O'Leary D P 2001 Choosing regularization parameters in iterative methods for ill-posed problems *SIAM J. Matrix Anal. Appl.* **22** 1204–21
[13] Miller K 1970 Least squares methods for ill-posed problems with a prescribed bound *SIAM J. Math. Anal.* **1** 52–74
[14] Park S C, Park M K and Kang M G 2003 Super-resolution image reconstruction: a technical overview *IEEE Signal Process. Mag.* **20** 21–36
[15] Regińska T 1996 A regularization parameter in discrete ill-posed problems *SIAM J. Sci. Comput.* **3** 740–9
[16] Vogel C R 1996 Non-convergence of the *L*-curve regularization parameter selection method *Inverse Problems* **12** 535–47
[17] Tikhonov A N 1963 Solution of incorrectly formulated problems and the regularization method *Sov. Math.—Dokl.* **4** 1035–8