Publicaciones en Ciencias y Tecnología. Vol 3, 2009 N^0 1, pp.15–22, ISSN:1856-8890,Dep.Legal pp200702LA2730

Inexact searches on the $L-$curve.

*Hugo Lara † Adalys Alvarez ‡ Freddy Torrealba

Recibido: 5 de mayo 2009, Aceptado: 28 de septiembre 2009

Abstract

In a Tikhonov regularization scheme to solve discrete linear ill-posed problems, selecting the parameter value is a key task. We use Wolfe inexact search on the $L-$curve to choose a λ regularization parameter value far from critical areas of the $L-$curve. Numerical results are shown comparing the inexact scheme with other exact searches.
Keywords: Ill-posed problems, Thikonov regularization, L-curve, inexact searches.

Búsquedas sobre curvas inexactas
Resumen

En un esquema de regularización para resolver problemas inversos lineales dicretos, la selección del valor del parámetro es una tarea clave. Usamos la busqueda inexacta de Wolfe sobre la curva L, para elegir un valor del parámetro de regularización λ lejos de áreas críticas de la curva L. Se muestran resultados numéricos que comparan el esquema inexacto con otras búsquedas inexactas.
Palabras clave: Curva L, regularización de Thikonov, búsquedas inexactas.

Introduction

Many problems encountered in science and engineering are ill posed inverse problems. Given an $n \times N$ matrix M and a data vector $y \in \mathbb{R}^n$, a discrete linear inverse problem involves solving approximately the following system of linear equations:

$$Mx = y. \tag{1}$$

According to Hadamard, an inverse problem is well posed if it satisfy the requirements of existence, uniqueness and stability of it solutions. If one of these requirements is not satisfied, the problem is said to be ill-posed. A linear least-squares solution for system (1) is a solution for

$$\text{minimize}_{x\in\mathbb{R}^N} \quad \tfrac{1}{2}\|y - Mx\|^2. \tag{2}$$

When a linear inverse problem is ill-posed, the least squares solution for problem (2) is not satisfactory, frequently giving poor reconstructions. To overcome these difficulties Regularization schemes are introduced (for regularization schemes see for example [2]). Regularization methods for computing stable solutions to inverse problems involve a trade-off between the "size" of the regularization solution (or it difference with a known default solution) and the quality of the fit that it provides to the given data. The well known Tikhonov regularization scheme (see [2]) consists in replacing the least squares problem (2) by

$$\text{minimize}_{x\in\mathbb{R}^N} \quad \tfrac{1}{2}\|y - Mx\|^2 + \tfrac{\lambda}{2}\|L(x - x_c)\|^2 \tag{3}$$

*Departamento de Investigación de Operaciones, Universidad Centroccidental "Lisandro Alvarado", Barquisimeto 3001 Apdo 400, Venezuela, hugol@ucla.edu.ve

†Universidad Nacional Abierta, Barquisimeto 3001 Apdo 400, Venezuela, adalys.alvarez@gmail.com

‡Departamento de Física, Universidad Centroccidental "Lisandro Alvarado", Barquisimeto 3001 Apdo 400, Venezuela, ftorre@ucla.edu.ve

where λ is called the regularized parameter. Here, the "size" of the regularized solution is, for a given matrix L, measured by $\|L(x - x_c)\|$; while the fit is measured by $\|y - Mx\|$. x_c is a priory estimate of x which represents our previous knowledge about the solutions. If no a priori information is available, x_c is set to zero.

Let us denote by $x(\lambda)$ the unique optimal solution for problem (3). Regularization is necessary when solving inverse problems because the "naive" least squares solution ($\lambda = 0$), denoted by x_{LS}, is completely dominated by contributions from data errors and rounding errors. Regularization is introduced to damp these contributions and keep the norm $\|L(x - x_c)\|$ with reasonable size. If too much regularization, or damping, is imposed on the solution, then it will not fit the given data y properly and the residual $\|y - Mx\|$ will be too large. On the other hand, if too little regularization is imposed then the fit will be good but the solution will be dominated by the contributions from the data errors, so $\|L(x - x_c)\|$ will be too large. The regularization parameter λ plays an important roll in balancing these norms. An important device used to choose a proper value for parameter λ is the so called $L-$curve criterium (see [2]). The $L-$curve is the plot $(\|y - Mx(\lambda)\|, \|L(x(\lambda) - x_c)\|)$ for $\lambda > 0$. It is a curve parametrized by λ. It help us to control the tradeoff between these two quantities. The $L-$curve is of our interest because it shows how the regularized solution changes as the regularization parameter λ changes. A distinct $L-$shaped corner of the $L-$curve is located exactly where the solution $x(\lambda)$ changes, form being dominated by the regularization errors of the data. That is why the corner of the $L-$curve corresponds to a good balance between the minimization of the sizes, and the corresponding regularization parameter λ is a good one.

When calculating points $x(\lambda)$ on the $L-$curve, we should solve the optimization problem (3) for each value of λ. A search on the curve to choose λ on the corner involves possibly an iterative procedure where is defined a sequence λ_k which converges to λ^* on the corner. When the inverse problem is large, it is desirable to choose λ^* in a few number of these expensive steps. Inexact searches exist in the optimization literature (see for example [5]). The objective of an inexact search is to look for an approximated optimal solution in a small number of steps, when accuracy is not necessary. A possible large search interval is chosen where the minimizer lies and ensuring sufficient decreasing of the objective function. Searches like Armijo search, or Wolfe search use first derivatives. In this paper we shall implement the Wolfe inexact search on the rotated $L-$curve to choose a fast approximated λ^*. We compare the number of iterations in other searches, like bisection, and demonstrate finite convergence of the procedure.

A standard tool to analyze the regularized solutions is the singular value decomposition (SVD) of the matrix M, which is a decomposition of the form

$$M = \sum_{i=1}^{N} u_i \sigma_i v_i^T \tag{4}$$

where the left and right singular vectors u_i and v_i are orthonormal, and the singular values σ_i are nonnegative and appears in a nondecreasing order. It is straightforward to show that, if $L = I$ and $x_c = 0$, the regularized solution is given by

$$x(\lambda) = \sum_{i=1}^{N} f_i \frac{u_i^T y}{\sigma_i} v_i$$

where $f_1, \ldots, f_N$ are the Tikhonov filter factors $f_i = \frac{\sigma_i^2}{\sigma_i^2 + \lambda}$.

The norm of the solution and the norm of the residual vector which characterizes the misfit are given in terms of the SVD by

$$\|x(\lambda)\|^2 = \sum_{i=1}^{N} (f_i \frac{u_i^T y}{\sigma_i})^2 \tag{5}$$

and $\|y - Mx(\lambda)\|^2 = \sum_{i=1}^{N} ((1 - f_i) u_i^T y)^2$. These expressions form the base to analyze the $L-$curve in [2, **?**, 6] and [**?**]. Nevertheless, in large problems the singular value decomposition could be expensive, and

iterative solutions for problem (3) becomes more reasonable. We shall give our analysis in the context of these iterative solutions.

Let us denote $\tau = \frac{1}{2}\|y - Mx\|^2$ and $\eta = \frac{1}{2}\|L(x - x_c)\|^2$. The log-log $L-$curve is the curve given by $(\log\tau, \log\eta)$. It is known (see [6]) that it has concave areas at the ends near the axes and convex $L-$shaped area where the curvature is maxima. We are interested in establishing inexact procedures to search for balanced λ in the log-log $L-$curve. These inexact procedures use the derivatives to identify regions where local minima are encountered. The main tool are derivatives for τ and η. In [2] Hansen provides derivatives for η and τ in terms of their SVD expressions. In our approach we use iterative resolution of the optimization problem (3) instead of SVD, like Gullikson and Wedin [1]. The expressions are equivalent.Let us denote $\hat{\tau} = \log\tau$ and $\hat{\eta} = \log\eta$. So $d\hat{\tau} = \frac{d\tau}{\tau}$ and $d\hat{\eta} = \frac{d\eta}{\eta}$. We also have

$$d\eta = [L^T L(x - x_c)]^T dx, \tag{6}$$

$$d\tau = [M^T(Mx - y)]^T dx. \tag{7}$$

On the other hand, we know from [2] that

$$d\eta = -\frac{d\tau}{\lambda}, \tag{8}$$

from which we obtain

$$\frac{d\hat{\eta}}{d\hat{\tau}} = \frac{\tau}{\eta}\frac{d\eta}{d\tau} = -\frac{\tau}{\lambda\eta}. \tag{9}$$

The second derivatives are also necessary:

$$\frac{d^2\hat{\eta}}{d\hat{\tau}^2} = \frac{d}{d\hat{\tau}}\left(\frac{d\hat{\eta}}{d\hat{\tau}}\right) = \frac{d}{d\tau}\left(-\frac{\tau}{\lambda\eta}\right)\frac{d\tau}{d\hat{\tau}}. \tag{10}$$

Since $x(\lambda)$ is calculated by minimizing (3) for fixed λ, with respect to x, then the derivatives for x should be calculated in a point satisfying

$$M^T(Mx - y) + \lambda L^T L(x - x_c) = 0. \tag{11}$$

Implicitely differenciating this expression and regrouping we obtain

$$(M^T M + \lambda L^T L)dx = -L^T L(x - x_c)d\lambda.$$

Since the Hessian matrix is positive definite we have

$$dx = -(M^T M + \lambda L^T L)^{-1} L^T L(x - x_c)d\lambda$$

and from (7), (11) and the last expression we get

$$\frac{d\tau}{d\lambda} = \lambda\beta \tag{12}$$

where $\beta := \beta(\lambda) = (x(\lambda) - x_c)^T L^T L(M^T M + \lambda L^T L)^{-1} L^T L(x(\lambda) - x_c)$. Now, since

$$\frac{d}{d\tau}\left(-\frac{\tau}{\lambda\eta}\right) = \frac{-\frac{d\tau}{d\tau}\lambda\eta + \tau\left(\eta\frac{d\lambda}{d\tau} + \lambda\frac{d\eta}{d\tau}\right)}{(\lambda\eta)^2},$$

and merging (8) and (12) in (10) we obtain

$$\frac{d^2\hat{\eta}}{d\hat{\tau}^2} = \left(\frac{-\lambda^2\beta\eta + \tau\eta - \lambda\beta\tau}{\lambda^3\beta\eta^2}\right)\tau. \tag{13}$$

We need efficient procedures to calculate (9) and (13) which requires calculating $x(\lambda)$ by an iterative optimization procedure; then we evaluate $\tau(x(\lambda))$, $\eta(x(\lambda))$ and $\beta(x(\lambda))$. To evaluate β we need to solve a linear system of equations (instead of calculating the inverse of the Hessian). Since this procedure involves solving a possible large optimization problem, it is desirable to perform as less evaluations for these expressions as possible.

Inexact search on the rotated $L-$curve

We rotate the $L-$curve like Reginska [6], obtaining the curve $(\theta, G(\theta))$, where $\theta = \hat{\tau} - \hat{\eta}$ and $G(\theta) = \hat{\tau} + \hat{\eta}$. We use the minimization of G as a selection criterium for the regularization parameter λ, avoiding regularization parameter values in regions of the $L-$curve dominated either for the quality of the fit or by the size of the regularized solution. First and second derivatives for G are needed to develop the fast minimization procedure.

Lemma 0.1 *Given the rotated $L-$curve $(\theta, G(\theta))$. The first and second derivatives are*

1. $\frac{dG}{d\theta} = \frac{-\tau+\lambda\eta}{\tau+\lambda\eta}$ *and*
2. $\frac{d^2G}{d\theta^2} = \frac{2\eta\tau(\eta\tau-\lambda\beta(\tau+\lambda\eta))}{\beta(\tau+\lambda\eta)^3}$.

Proof 0.1 *We know $d\theta = d\hat{\tau} - d\hat{\eta}$ and $dG(\theta) = d\hat{\tau} + d\hat{\eta}$. By using $d\hat{\tau} = \frac{d\tau}{\tau}$, $d\hat{\eta} = \frac{d\eta}{\eta}$ and $d\eta = -\frac{d\tau}{\lambda}$ we obtain*

$$\frac{dG(\theta)}{d\theta} = \frac{-\tau + \lambda\eta}{\tau + \lambda\eta} \tag{14}$$

$$\text{and } \frac{d\theta}{d\tau} = \frac{1}{\tau} + \frac{1}{\lambda\eta} \tag{15}$$

Now, to show the second statement note that

$$\frac{d^2G}{d\theta^2} = \frac{d}{d\theta}\left(\frac{dG}{d\theta}\right) = \frac{d}{d\tau}\left(\frac{dG}{d\theta}\right)\frac{d\tau}{d\theta} \tag{16}$$

so

$$\begin{aligned}
\tfrac{d}{d\tau}\left(\tfrac{dG}{d\theta}\right) &= \tfrac{d}{d\tau}\left(\tfrac{-\tau+\lambda\eta}{\tau+\lambda\eta}\right) \\
&= \tfrac{d}{d\tau}[(-\tau+\lambda\eta)](\tau+\lambda\eta)^{-1} \\
&+(-\tau+\lambda\eta)\tfrac{d}{d\tau}[(\tau+\lambda\eta)^{-1}] \\
&= \left(-1+\tfrac{d\lambda}{d\tau}\eta+\lambda\tfrac{d\eta}{d\tau}\right)(\tau+\lambda\eta)^{-1} \\
&+(-\tau+\lambda\eta)\left(-(\tau+\lambda\eta)^{-2}\left(1+\tfrac{d\lambda}{d\tau}\eta+\lambda\tfrac{d\eta}{d\tau}\right)\right)
\end{aligned}$$

then using $\frac{d\tau}{d\lambda} = \lambda\beta$ and $\frac{d\eta}{d\tau} = -\frac{1}{\lambda}$ we get

$$\frac{d}{d\tau}\left(\frac{dG}{d\tau}\right) = \frac{2(\tau\eta - \lambda\beta(\tau+\lambda\eta))}{\lambda\beta(t+\lambda\eta)^2}.$$

Merging this expression and (15) in (16) we obtain the second derivative.

As a direct consequence of this lema we have that $\lim_{\lambda\to 0}\frac{d}{d\theta}G(\lambda) = -1$ and $\lim_{\lambda\to\infty}\frac{d}{d\theta}G(\lambda) = 1$.

To calculate these derivatives we need to obtain $x(\lambda)$ as the result of an optimization procedure, and then evaluate τ, η and β.

In the sequel we establish an inexact procedure on the $L-$curve to choose the regularization parameter, avoiding portions of the curve dominated by the size of the regularization solution or the data misfit. In the literature there are some inexact line searches to choose approximated minimizers of a function on an interval, when accuracy is not necessary (see for example [5]). We here shall implement the so called Wolfe search to choose the regularization parameter, by minimizing the rotated $L-$curve.

Algorithm 0.1 *Given $\gamma \in (0,1)$, λ_0 satisfying $\frac{dG}{d\theta}(\lambda_0) < 0$*

For $i = 1, 2, \ldots$

$\lambda_i := 2^i\lambda_0$

if $\frac{dG}{d\theta}(\lambda_i) > 0$ *stop*
end
$\underline{\lambda}_0 = \lambda_{i-1}$, $\overline{\lambda}_0 = \lambda_i$
for $k = 1, 2, \ldots$
$\lambda_{k+1} := \frac{\underline{\lambda}_k + \overline{\lambda}_k}{2}$
if $\frac{d}{d\lambda}G(\lambda_{k+1}) > 0$,
then $\underline{\lambda}_{k+1} = \underline{\lambda}_k$, $\overline{\lambda}_{k+1} = \lambda_{k+1}$
else $\underline{\lambda}_{k+1} = \lambda_{k+1}$, $\overline{\lambda}_{k+1} = \overline{\lambda}_k$
end
$k := k + 1$
Take $\hat{\lambda}_0$ *the first* $\underline{\lambda}_k$ *such that* $\frac{d^2}{d\theta^2}G(\underline{\lambda}_k) > 0$.
if $|\frac{d}{d\theta}G(\lambda_k)| < \gamma|\frac{d}{d\theta}G(\hat{\lambda}_0)|$
and $\frac{d^2}{d\theta^2}G(\lambda_k) > 0$ *stop.*

The objective of the first loop is to choose an interval where a minimum is guaranteed to be in. The second loop reduce the size of the interval by a half in each iteration, and keep an interval $[\underline{\lambda}_k, \overline{\lambda}_k]$ containing a local minimum. The stopping rule ask for any λ_k in these possible large interval, such that sufficient decrease of the function is guaranteed. It also ask for the second derivative to be sure λ_k is in the convex part of the $L-$curve.

The stoping rule $|\frac{d}{d\theta}G(\lambda_k)| < \gamma|\frac{d}{d\theta}G(\hat{\lambda}_0)|$ in the above procedure is called Wolfe condition [5]. We also call Wolfe interval to $[\underline{\lambda}, \overline{\lambda}]$ defined by the above condition, in which sufficient decreasing of the objective function is guaranteed.

Theorem 0.1 *The algorithm 0.1 terminates in a finite number of steps with a near optimal parameter value.*

Proof 0.2 *Since* $\lim_{\lambda\to 0}\frac{d}{d\theta}G(\lambda) = -1$ *then there exist* λ_0 *such that* $\frac{d}{d\theta}G(\lambda_0) < 0$. *Also, since* $\lim_{\lambda\to\infty}\frac{d}{d\theta}G(\lambda) = 1$, *we have that there exist* $\hat{i}$ *such that* $\frac{d}{d\theta}G(\lambda_{\hat{i}}) < 0$ *and* $\frac{d}{d\theta}G(\lambda_{\hat{i}+1}) > 0$ *so the first loop is well defined.*

Since $\frac{d}{d\theta}G(\underline{\lambda}_0) < 0$, $\frac{d}{d\theta}G(\overline{\lambda}_0) > 0$ *and the continuity of this derivative we have that there exist* $\lambda^* \in [\underline{\lambda}_0, \overline{\lambda}_0]$ *satisfying* $\frac{d}{d\theta}G(\lambda^*) = 0$ *that is* λ^* *is a stationary point for* G. *Furthermore,* λ^* *is a local minima because the local convexity of* G. *Additionally, since* $\frac{d^2}{d\theta^2}G(\lambda) \geq 0$ *for* λ *near* λ^* *there exists an interval* $[\underline{\lambda}, \overline{\lambda}]$ *satisfying* $\lambda^* \in [\underline{\lambda}, \overline{\lambda}] \subset [\underline{\lambda}_0, \overline{\lambda}_0]$, $\frac{d}{d\theta}G(\underline{\lambda}) < 0$, $\frac{d}{d\theta}G(\overline{\lambda}) > 0$ *and* $\frac{d^2}{d\theta^2}G(\lambda) \geq 0\ \forall\lambda \in [\underline{\lambda}, \overline{\lambda}]$. *By the construction of sequences* $\{\underline{\lambda}_k\}$, $\{\overline{\lambda}_k\}$ *and* $\{\lambda_k\}$ *we keep an interval* $[\underline{\lambda}_k, \overline{\lambda}_k]$ *with* λ^* *in it, and such that* $0 \leq \lambda^* - \underline{\lambda}_k \to 0$ *and* $0 \leq \overline{\lambda}_k - \lambda^* \to 0$. *So, eventually we get* $\hat{k}$ *such that* $[\underline{\lambda}_{\hat{k}}, \overline{\lambda}_{\hat{k}}] \subset [\underline{\lambda}, \overline{\lambda}]$. *In this interval , the second derivative condition holds. It remains to prove that there exist* k^* *such that* $\lambda_{k^*} \in [\underline{\lambda}_{k^*}, \overline{\lambda}_{k^*}] \subset [\underline{\lambda}_{\hat{k}}, \overline{\lambda}_{\hat{k}}]$ *and Wolfe condition holds, which is immediate because* $\lambda_k \to \lambda^*$ *and* $\frac{d}{d\theta}G(\lambda^*) = 0$.

Numerical experiments

We present numerical experiments run in some test problems. The first problem Shaw, by Hansen in [2] is a model of an image reconstruction problem. In figure 1a we show curves of this reconstruction obtained with extreme parameter values (in the horizontal and vertical parts of the $L-$curve), and the original curve. The doted curve represents the original data; the low continuous curve corresponds to a large parameter value, and the scrambled curve is associated to a very small parameter value. Figure 1b exibit curves calculated with parameters in Wolfe interval $[10^{-3}, 10^{-2}]$, obtained in 1 iteration. Note that the main characteristics of such curves are kept. In figure 1c the doted curve is associated to exact (bisection, 9 iterations) and the continuous one corresponds to Wolfe search.

The second test problem, an exponential sum parameter estimation problem of the form

$$y(t_j) = \sum_{i=1}^{N} a_i \exp(-b_i t_j) + \epsilon_j, \ , \ j = 1 \ldots, n \tag{17}$$

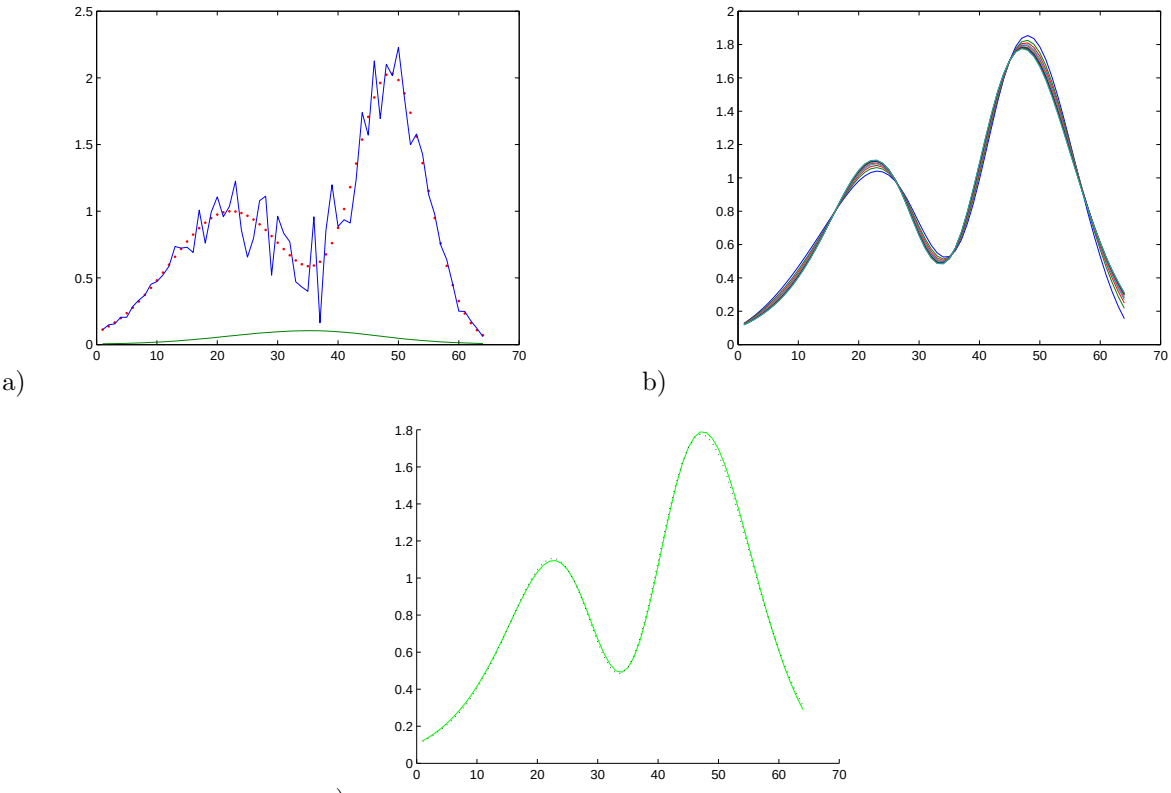


Figure 1: a) Extreme parameter values in Shaw. b) Curves in Wolfe parameter interval for Shaw. c) Curves with parameters from bisection and Wolfe searches.

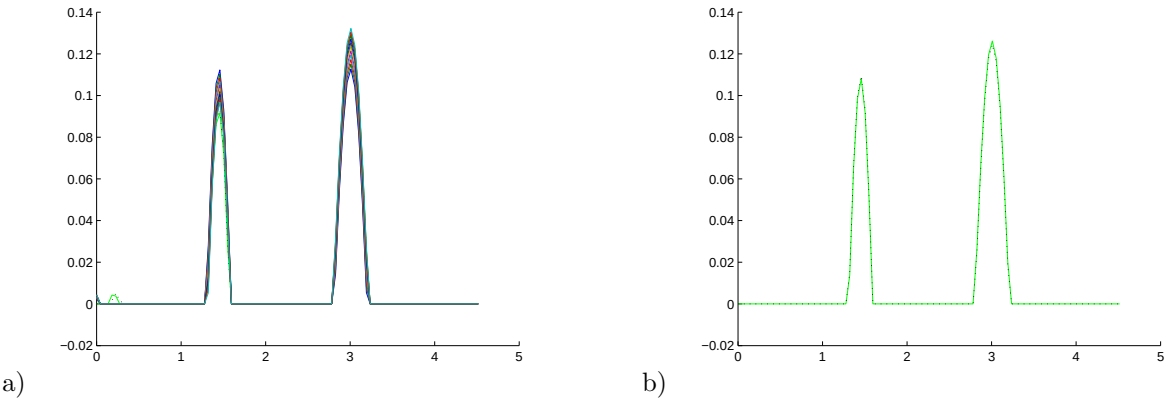
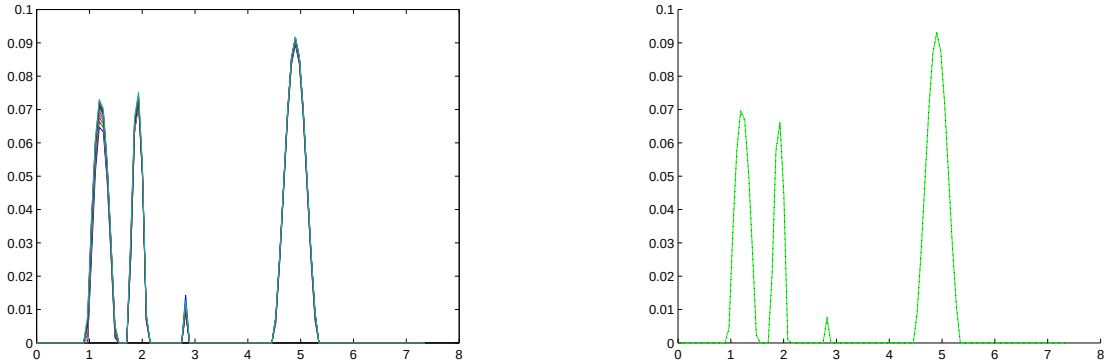


Figure 2: a) Curves in Wolfe parameter interval for Expsum1. b) Curves with parameter from bisection and Wolfe for Expsum1.



a) b)

Figure 3: a) Curves in Wolfe parameter interval for Expsum2. b) Curves with parameter from bisection and Wolfe for Expsum2.

is known to be ill-posed (see [3]). In figures 2a and 3a we show Wolfe interval curves for two different data sets, in which 4 and 3 inexact search iterations were needed; while figures 2b and 3b exibit curves associated to bisection (20 and 21 iterations respectively) and Wolfe parameter values.

Concluding remarks

Iterative search procedures to solve minimization in one dimension keep a search interval in each step, and either reduce it to a fixed tolerance in "exact" procedures like bisection, or take the approximate minimizer in a larger search interval, like Wolfe interval.

We propose using a fast inexact search on the $L-$curve to choose a balanced λ parameter. In our experiments we observe that the difference between an exact search and the inexact counterpart is too small, and so it does not worth to perform a more expensive algorithm.

Figure 1a shows images of the Shaw problem reconstructed from parameter values in the critical areas of the $L-$curve. Note that it is necessary to choose a parameter value far from the "horizontal" and "vertical" portions of the $L-$curve. In figures 1b, 2a and 3a we show reconstructions obtained form parameters in Wolfe interval. We note that the "band" described by curves in this interval is thin, and not too much differences in the main properties of the resulting reconstructed images is observed. In figures 1c, 2b y3b we show images obtained by bisection and Wolfe procedures. Note that the difference is too small compared with the numerical effort required to perform each of the searches.

In our framework do not use the singular value decomposition as a tool to calculate the reconstruction, instead we use optimization procedures. This is appropriate to deal with large problems, or problems that should be solved several times as part of another procedure. Therefore, some procedure which save computational effort is adequate in this framework. In our opinion inexact searches are promising choices when dealing with large ill-posed linear inverse problems in which accuracy is not too important.

Other choices of the minimization problem can be done. In fact, the more popular maximization of the curvature of the $L-$curve, or the left corner of the $U-$curve (see [**?**]) are natural extensions. The main difficulty in applying the maxima curvature criterion is the efficient representation of the derivatives.

References

[1] M. E Gulliksson and P... Wedin. Using the Nonlinear L-curve and its Dual, *Progress in Industrial Math.*, ECMI, pp. 162-170, Eds., L. Arkeryd, J. Bergh, P. Brenner, and R. Pettersson, B.G. Teubner Stuttgart, Leibniz, 1999

[2] P. C. Hansen, The $L-$curve and its use in the numerical treatment of inverse problems, *Computational Inverse Problems in Electrocardiology*, Ed. P. Johnston. Advances in Computational Bioengineering, 2001, Vol 5.

[3] K. Holmstrom and J. Petersson, A review of the parameter estimation problem of fitting positive exponential sums to empirical data, *Applied Mathematics and Computation*, 2002, Vol 126, No 1, 31–61.

[4] D. Krawczyk-Stando and M. Rudnicki, Regularization parameter selection in discrete ill-posed problems – the use of the U–curve, *Int. J. Appl. Math. Comput. Sci.* 2007, Vol 17, No. 2, 157–164.

[5] S. Nash and A. Sofer, *Linear and Nonlinear Programming*. McGraw Hill. 1996.

[6] T. Reginska, A regularization parameter in discrete ill-posed problems, *SIAM J. Sci. Comput* 1996, Vol 17, No 3, 740–749.