

# Efficient determination of multiple regularization parameters in a generalized L-curve framework

Murat Belge<sup>1</sup>, Misha E Kilmer<sup>2,4</sup> and Eric L Miller<sup>3</sup>

<sup>1</sup> Aware, Inc., 40 Middlesex Turnpike Bedford, MA 01730, USA

<sup>2</sup> Department of Mathematics, Tufts University, 113 Bromfield-Pearson Building, Medford, MA 02155, USA

<sup>3</sup> 315 Sterns Hall, Northeastern University, 360 Huntington Ave., Boston, MA 02115-5000, USA

E-mail: mkilme01@tufts.edu and elmiller@ece.neu.edu

Received 1 March 2002, in final form 31 May 2002

Published 15 July 2002

Online at [stacks.iop.org/IP/18/1161](http://stacks.iop.org/IP/18/1161)

## Abstract

The selection of multiple regularization parameters is considered in a generalized L-curve framework. Multiple-dimensional extensions of the L-curve for selecting multiple regularization parameters are introduced, and a minimum distance function (MDF) is developed for approximating the regularization parameters corresponding to the generalized corner of the L-hypersurface. For the single-parameter (i.e. L-curve) case, it is shown through a model that the regularization parameters minimizing the MDF essentially maximize the curvature of the L-curve. Furthermore, for both the single- and multiple-parameter cases the MDF approach leads to a simple fixed-point iterative algorithm for computing regularization parameters. Examples indicate that the algorithm converges rapidly thereby making the problem of computing parameters according to the generalized corner of the L-hypersurface computationally tractable.

(Some figures in this article are in colour only in the electronic version)

## 1. Introduction

In many inverse problems occurring in the physical sciences the discrete model of the problem takes the form

$$\mathbf{g} = \mathbf{H}(\mathbf{f}) + \mathbf{n}, \quad (1)$$

where  $\mathbf{H}$  is an operator and  $\mathbf{g}$  and  $\mathbf{f}$  are the  $m$ - and  $n$ -length vectors representing the observations and the original object, respectively. The error vector  $\mathbf{n}$ , whose entries consist of zero mean normally distributed random variables, is included in order to account for the inaccuracy introduced in the modelling and measurement phases. The objective of the problem is to obtain an estimate  $\mathbf{f}^*$  of the original object  $\mathbf{f}$  from the noise-corrupted measurements  $\mathbf{g}$ .

<sup>4</sup> Author to whom any correspondence should be addressed.

Attempting to solve (1) in the least-squares sense will most likely fail because the problem is highly ill posed in the sense that the solution is very sensitive to noise in the data. Hence, noise in the data is magnified in the least-squares solution to such an extent that the solution is of little or no practical value. For this reason, many numerical methods treating discrete inverse problems seek to combat the effects of noise by requiring that  $f^*$  be relatively small as measured by the norm of the estimate in some function space. This has the effect of making the problem better conditioned. Such methods are called *regularization methods* and always include one or more parameters, called *regularization parameters*, to control the conditioning of the problem. An appropriate choice of the regularization parameters is of vital importance for the quality of the resulting estimate and has been the subject of extensive research [9, 11, 13–16, 20, 25].

For purposes of this work, we restrict ourselves to linear operators (i.e.  $\mathbf{H}f = \mathbf{H}(f)$ ), which arise, for example, in image and signal processing applications. In this case, the L-curve [11, 13] is one of the simplest and most popular methods for selecting a single regularization parameter when no other noise properties are known. The method is based on a plot, in an appropriate scale, of the size of the solution (measured in an appropriate norm) against the corresponding residual for all valid regularization parameters. Intuitive justifications and numerical experiments indicate that the so-called corner of the L-curve gives a regularization parameter which provides an acceptable compromise between the data misfit and regularization terms in the cost function [11]. The corner is defined to be the point on the L-curve where the curvature reaches a maximum [13]. It is important to note that the L-curve has been shown to have undesirable asymptotic properties [10, 24]; like all parameter choice methods based on no *a priori* knowledge of the noise, the solution estimates based on the parameter chosen via the L-curve fail to converge to the true solution as the noise norm goes to zero. Nevertheless, it can be surmised from the quantity of recent work on the method that, in practical settings when noise is present, the L-curve continues to be a popular and effective parameter selection method when no other information about the noise is available.

In recent years, there has been a growing interest in sophisticated regularization techniques which use multiple constraints as a means of improving the quality of inversion [1, 3, 4, 21]. Examples include the inverse problem of electrocardiography [4] where both temporal and spatial constraints are imposed on the solution, the wavelet domain restoration of blurred images where each subband in the wavelet decomposition of the image is subjected to a different degree of regularization [3], and problems involving depth-varying regularization [19]. We note that, in a sense, what we refer to as multiple constraints can be thought of as a special case of Tikhonov regularization with an appropriate Sobolev norm (see [8, pp 32–4] for a discussion in the Hilbert space setting).

Motivated by the simplicity and the success in practice of the L-curve method, we have developed the L-hypersurface method [2] as a multiparameter generalization of the L-curve. The L-hypersurface, analogous to the L-curve, is the plot of the residual norm against the size of multiple constraints imposed on the solution for all valid regularization parameters. Similar to the one-dimensional case where the corner is defined as the point on the L-curve with maximum curvature, the generalized corner of the L-hypersurface is defined as the point with maximum Gaussian curvature. Note that in the single-parameter case, the L-hypersurface reduces to the L-curve, and as the Gaussian curvature and planar curvature coincide, the L-hypersurface method reduces to the standard L-curve method.

It has been demonstrated in [2] through numerical examples that the Gaussian curvature of the L-hypersurface is closely tied to the error between the original and estimated objects and that regularization parameters maximizing the Gaussian curvature provide acceptable estimates in terms of estimation error. Indeed, the results in [2] showed that by choosing the parameters

as the points that gave maximum Gaussian curvature, the corresponding regularized solutions were close to the optimal solutions (in mean square error).

Determining the point of maximum Gaussian curvature, therefore, is the key computational issue associated with the use of the L-hypersurface. As we show in section 4, the curvature surface itself has many extrema thereby causing difficulties for most gradient ascent-type approaches to maximization. Moreover, computing the Gaussian curvature of the L-hypersurface is computationally very demanding (see section 2) so that any type of grid search either for the maximum point or a starting point for a gradient ascent algorithm is far from attractive.

In this paper, we consider replacing the Gaussian curvature by a surrogate *minimum distance function* (MDF) which is far easier to optimize. A similar approach for the single parameter case with 2-norm constraint is given in [20]. We show that our formulation of the problem leads naturally to an efficient fixed-point iteration to determine the optimal parameters. The performance of this approach is evaluated through numerical experiments involving both standard inverse problems drawn from Hansen's regularization toolbox [12] as well as through practical image restoration examples. These experiments demonstrate that there is little performance loss incurred through the optimization of the MDF as opposed to direct maximization of the Gaussian curvature; however, the computational burden is significantly smaller.

In this single-parameter case, the contribution of this paper is not so much a new means of choosing a parameter but an efficient method for finding the point of max curvature for a wider range of regularization functionals than are normally considered (see section 2). For the multiparameter problem, we offer both a new approach to choosing the parameter as well as an efficient computational method.

In addition to the numerical evaluation of the MDF, we provide theoretical justification for its performance. Here, we concentrate exclusively on the single-parameter (i.e. L-curve) case where analysis is tractable. First, we show that under some mild conditions, the fixed-point iteration for the single-parameter MDF is guaranteed to converge to a critical point of the underlying cost function. Second, we identify general conditions based on the geometry of the L-curve under which the MDF approach can be expected to perform comparably to the curvature method. Clearly, the detailed behaviour of the L-curve for any particular problem is intimately related to the underlying system matrix and regularization functional; however, geometrically the utility of the L-curve (when it works) rests on the fact that there is a discernible corner that separates regions of too little regularization from those where there is too much. Motivated by this observation, we introduce a *model L-curve* parametrized in such a way as to capture the salient geometric characteristics of a true L-curve, especially in the region near the corner. Use of the model leads to an intuitive but still rigorous understanding of the MDF in terms of the basic structure of the corner. Specifically, we relate the 'sharpness' of the corner to the point on the model L-curve which will be selected by MDF optimization. In the appropriate limit where the corner becomes infinitely sharp, we show that MDF will in fact choose the same parameter as the curvature method. While this abstracted model is not derived directly from a true L-curve, we feel that the insight provided by the analysis justifies the consideration of the model.

Our paper is organized as follows. We give a brief overview of Tikhonov-type regularization in section 2 and we formally introduce the L-hypersurface method. In section 3, we introduce the MDF, explore its properties and develop a simple fixed-point iteration to locate its minimum. In section 4, we illustrate the new method by several numerical examples and we conclude the paper with a discussion of results in section 5.

## 2. Regularization and the L-hypersurface

The simplest and the most well known regularization method is Tikhonov's method [22] which consists of solving the following regularized least-squares problem:

$$\mathbf{f}^*(\alpha) = \arg \min_{\mathbf{f}} \{\|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 + \alpha \|\mathbf{f}\|_2^2\}, \quad \mathbf{H} \in \mathcal{R}^{m \times n}, \quad (2)$$

where  $\alpha > 0$  is called the regularization parameter. Tikhonov's method can be extended by using different regularization functionals or multiple constraints. In this paper, we consider the following multiply-constrained regularization approach:

$$\mathbf{f}^*(\alpha) = \arg \min_{\mathbf{f}} \left\{ \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 + \sum_{i=1}^M \alpha_i \Phi_i(\mathbf{R}_i \mathbf{f}) \right\}, \quad \mathbf{R}_i \in \mathcal{R}^{m \times n} \quad (3)$$

where  $M$  denotes the number of constraints,  $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_M]^T$ ,  $\mathbf{R}_i$  are regularization operators and  $\alpha_i$  are the corresponding regularization parameters,  $\Phi_i(\mathbf{R}_i \mathbf{f}) = \sum_{j=1}^m \phi_i([\mathbf{R}_i \mathbf{f}]_j)$  and the notation  $[\mathbf{R}_i \mathbf{f}]_j$  denotes the  $j$ th element of the vector  $\mathbf{R}_i \mathbf{f}$ . We assume that  $\phi_i(t)$  is a continuously differentiable, convex, non-negative ( $\phi_i(t) \geq 0, \forall t$ ), even function which satisfies the following conditions [6]:

- (i)  $\phi'_i(t) \geq 0, \forall t \geq 0,$
- (ii)  $\lim_{t \rightarrow 0^+} \frac{\phi'_i(t)}{t} = C, \quad 0 < C < \infty,$

where prime denotes differentiation. The formulation in (3) includes many popular regularization techniques as its special cases. For example, by taking,  $\phi_i(t) = t^2, i = 1, \dots, M$ , and  $\mathbf{R}_i$  as a discrete approximation to the  $(i-1)$ st-order differentiation, we obtain the conventional Tikhonov method with the discrete Sobolev norm as a constraint. A wavelet domain image restoration algorithm developed by the authors [3] is also a special case of (3). In this case the quantities of interest  $\mathbf{g}, \mathbf{H}, \mathbf{f}, \mathbf{n}$  represent the wavelet decomposition of the related quantities,  $\phi_i(t) = (|t|^2 + \epsilon)^{p/2}, i = 1, \dots, M$  with  $1 \leq p \leq 2$  and  $\epsilon$  a small positive constant, and  $\mathbf{R}_i, i = 1, \dots, M$  are operators extracting the desired portions of the wavelet coefficients.

We note that most of the literature on single-parameter choice methods ( $M = 1$ ) has been devoted to the 2-norm case  $\phi_i(t) = t^2$  (see for example [14, 15] and references therein). Although our primary goal was to construct a parameter selection method for multiple parameters, one benefit of our MDF approach is its applicability to single-parameter cases with  $\phi_i(t) \neq t^2$  (see example 1). If  $\phi_i(t) = t^2$  and  $M = 1$ , there may be more (computationally) efficient approaches, some of which were cited in the introduction, and it is not within the scope of this paper to perform a comparison between those methods and ours in this instance. For a performance comparison of some of the methods applicable in the  $\phi_i(t) = t^2, M = 1$  case, the reader is referred to [14, ch 7] and the references therein.

By taking the gradient of (3) with respect to  $\mathbf{f}$  and setting the result equal to zero we obtain the following first-order condition that must be satisfied by  $\mathbf{f}^*(\alpha)$ :

$$\mathbf{K}_{\mathbf{f}^*} \mathbf{f}^* = \mathbf{H}^T \mathbf{g}, \quad (4)$$

where

$$\mathbf{K}_{\mathbf{f}^*} = \mathbf{H}^T \mathbf{H} + \frac{1}{2} \sum_{i=1}^M \alpha_i \mathbf{R}_i^T \text{diag}_{k=1, \dots, m} \left[ \frac{\phi'_i([\mathbf{R}_i \mathbf{f}^*]_k)}{[\mathbf{R}_i \mathbf{f}^*]_k} \right] \mathbf{R}_i. \quad (5)$$

A key issue with the use of this regularization scheme is the selection of the  $\alpha_i$  parameters. Here we propose the L-hypersurface as a multi-dimensional extension of the classical L-curve

method [11]. To construct the L-hypersurface, we first introduce the following quantities:

$$z(\alpha) = \|g - Hf^*(\alpha)\|_2^2 \quad (6)$$

$$x_j(\alpha) = \Phi_j[R_j f^*(\alpha)], \quad j = 1, \dots, M \quad (7)$$

where  $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_M]^T$  and  $f^*$  is defined in (3). With the above definitions, the L-hypersurface is defined as a subset of  $\mathcal{R}^{M+1}$  associated with the map  $\beta(\alpha) : \mathcal{R}_+^M \rightarrow \mathcal{R}^{M+1}$ , such that

$$\beta(\alpha) = (\psi[x_1(\alpha)], \dots, \psi[x_M(\alpha)], \psi[z(\alpha)]), \quad (8)$$

where  $\psi$  is an appropriate scale such as  $\psi(t) = \log(t)$  or  $\psi(t) = \sqrt{t}$ . For a single constraint, the L-hypersurface reduces to the conventional L-curve which is simply a plot of the residual norm versus the norm of the regularized solution drawn in an appropriate scale for a set of admissible regularization parameters. In this way, the L-curve displays the compromise between the minimization of these two quantities. It has been argued and numerically shown that the so-called ‘corner’ of the L-curve corresponds to a point where regularization errors and perturbation errors are approximately balanced [11].

Analogous to the one-dimensional case, the L-hypersurface is a plot of the residual norm  $z(\alpha)$  against the constraint norms  $x_j(\alpha)$ ,  $1 \leq j \leq M$  drawn in an appropriate scale. Intuitively, the ‘generalized corner’ of the L-hypersurface should correspond to a point where regularization errors and perturbation errors are approximately balanced. By a generalized corner, we mean a point on the surface around which the surface is maximally warped. We can quantitatively measure how much a surface is warped around a point by computing its Gaussian curvature [5, 7]. In this work, we shall assume that the surface  $\beta(\alpha)$  is *regular* [5] and therefore its Gaussian curvature is well defined and exists  $\forall \alpha \in \mathcal{R}_+^M$ .

The Gaussian curvature of  $\beta(\alpha)$  can be computed given the first- and second-order partial derivatives of  $\psi[z(\alpha)]$  with respect to  $\psi[x_i(\alpha)]$ ,  $1 \leq i \leq M$ , and is given by the following expression [7]:

$$\kappa(\alpha) = \frac{(-1)^M}{w^{M+2}} |P|, \quad (9)$$

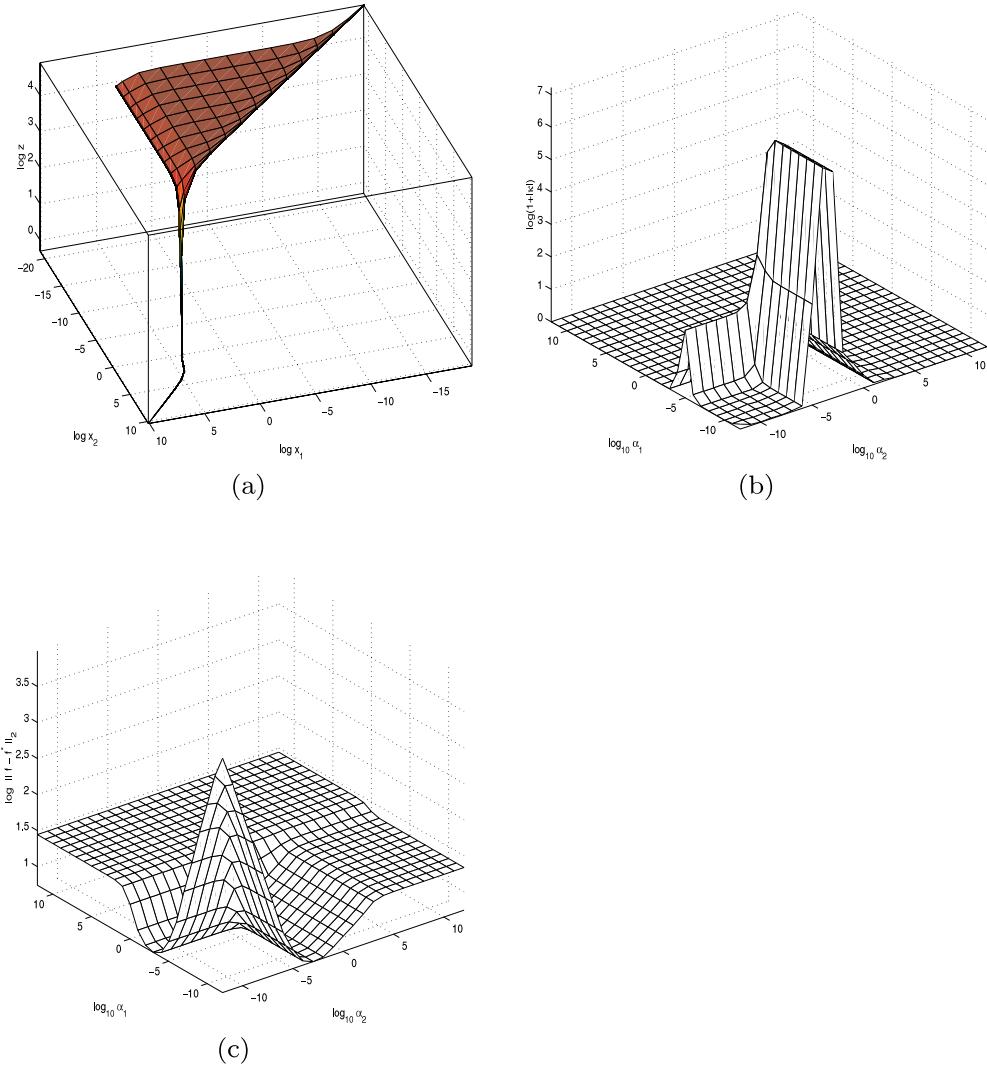
where  $w^2 = 1 + \sum_{i=1}^M (\frac{\partial \psi(z)}{\partial \psi(x_i)})^2$ ,  $P_{i,j} = \frac{\partial^2 \psi(z)}{\partial \psi(x_i) \partial \psi(x_j)}$  and derivatives are evaluated at  $q = (x_1(\alpha), \dots, x_M(\alpha), z(\alpha))$ . These derivatives are obtained by a simple transformation of partial derivatives of related quantities with respect to regularization parameters [2]. One difficulty with this calculation is that each of the required partial derivatives can only be found by solving a linear system of equations of the same size as that of the original inverse problem.

A sample L-hypersurface is shown in figure 1(a) for a least-squares problem with first- and second-order derivatives of the object as constraints. The middle and the rightmost plots are the curvature of the L-hypersurface and the norm of the error between the original and estimated objects. We observe that the points on the curvature plot where the curvature achieves a local maximum seems to track the local minimum of the estimation error surface. This behaviour is not limited to this example but is also observed in a variety of different problems [3].

### 3. Parameter selection via the minimum distance function

#### 3.1. The minimum distance function defined

A major shortcoming of the L-hypersurface method is that direct maximization by evaluating the Gaussian curvature for a large number of regularization parameters is expensive. Furthermore, use of a conventional optimization technique to locate the maximum Gaussian



**Figure 1.** (a) A typical L-surface; (b) the Gaussian curvature of the L-surface in (a); (c) corresponding estimation error surface.

curvature point is hampered by the fact that the Gaussian curvature function possesses multiple extrema. Considering these difficulties, we propose replacing the Gaussian curvature function by a surrogate function which is far easier to optimize. Our ultimate goal is to choose the surrogate function so that the regularization parameters obtained from the optimization of this function are close to those chosen by the L-hypersurface method. To give a flavour of the simple geometrical ideas behind our approach, we consider a typical L-curve as displayed in figure 4. We denote the points where extreme solution norm and extreme residual norm regions start by  $a$  and  $b$  respectively. Formally, the points  $a$  and  $b$  are defined as

$$a = \psi(\|g - Hf^*(\alpha_a)\|_2^2), \quad b = \psi(\Phi[Rf^*(\alpha_b)]), \quad (10)$$

where  $\alpha_a$  is the regularization parameter to the left of the corner where the L-curve becomes approximately horizontal and  $\alpha_b$  is the regularization parameter to the right of the corner where

the L-curve becomes approximately vertical (see figure 2). We define an origin,  $O = (a, b)$ , and compute the distance from our origin  $O$  to the L-curve. Suppose that there is a slowly expanding bubble located exactly at the origin  $O$ . From the geometry, it is easy to see that the first point on the L-curve that the bubble touches will be close to the corner of the L-curve. Furthermore, as the bubble continues to expand, the circle describing the boundaries of the bubble intersects the L-curve at exactly two points at the left and right of the corner until the circle reaches extreme residual norm and extreme solution norm regions. The radius of the circle is in fact the value of our distance function. The statements concerning the behaviour of the bubble describes our distance function. That is, the distance function has a minimum at a point close to the corner and the function increases as we go away from the origin until we reach extreme residual or signal norm regions. In this way, we define a function whose minimum occurs at or near the corner and possesses a single minimum for a wide range of regularization parameters. Hopefully, the newly defined distance function will make our optimization task much easier.

We begin by defining our surrogate function as follows.

**Definition 1 (Minimum distance function).** Let  $O = (a, b_1, \dots, b_M)$  denotes the coordinates of our origin. The MDF,  $v(\alpha)$ , is the distance from the origin  $O$  to the point  $\beta(\alpha)$  on the L-curve:

$$v(\alpha) = |\psi[z(\alpha)] - a|^2 + \sum_{i=1}^M |\psi[x_i(\alpha)] - b_i|^2. \quad (11)$$

**Definition 2 (Minimum distance point).** The minimum distance point (MDP) is the point where the Gaussian curvature of the L-curve is positive and  $v(\alpha)$  reaches a local minimum:

$$\alpha^* = \arg \min_{\alpha \in \mathcal{R}^{M+1}} v(\alpha).$$

### 3.2. A fixed-point approach for finding the MDP

For the moment let us assume that a minimum to the MDF does in fact exist. If this is the case, then in general we may use any appropriate optimization technique for finding the  $\alpha^*$ ; however, many optimization algorithms require higher-order partial derivatives of  $z(\alpha)$  and  $x_i(\alpha)$  with respect to  $\alpha_i$ ,  $i = 1, \dots, M$ . It can be easily shown that each of these partials can be computed from  $\frac{\partial f^*(\alpha)}{\partial \alpha_i}$ , which in turn, are obtained by solving a linear system whose size is the same as that of the original problem. Clearly, the computational effort associated with computing the required partials can be prohibitively large if the size of the problem is big. However, using elementary properties of the MDF we can easily derive a fixed-point algorithm for  $\alpha^*$ . In the following, for ease of notation we have dropped the explicit dependence of  $f^*$  on  $\alpha$ . Differentiating (11) with respect to  $\alpha_j$ , and equating the result to zero we obtain the following equation:

$$\sum_{i=1}^M (\psi[x_i] - b_i) \psi'[x_i] \frac{\partial x_i}{\partial \alpha_j} + (\psi[z] - a) \psi'[z] \frac{\partial z}{\partial \alpha_j} = 0. \quad (12)$$

Using (6), (7), it is easy to show the following:

$$\frac{\partial z}{\partial \alpha_j} = 2(\mathbf{H}f^* - \mathbf{g})^T \mathbf{H} \frac{\partial}{\partial \alpha_j} f^* \quad (13)$$

$$\frac{\partial x_i}{\partial \alpha_j} = (f^*)^T \mathbf{R}_i^T \underset{k=1, \dots, m}{\text{diag}} \left[ \frac{\phi'_i([\mathbf{R}_i f^*]_k)}{[\mathbf{R}_i f^*]_k} \right] \mathbf{R}_i \frac{\partial}{\partial \alpha_j} f^*. \quad (14)$$

Next, we consider (13):

$$\begin{aligned}
\frac{\partial z}{\partial \alpha_j} &= 2(\mathbf{H}\mathbf{f}^* - \mathbf{g})^T \mathbf{H} \frac{\partial}{\partial \alpha_j} \mathbf{f}^* = 2(\mathbf{H}\mathbf{K}_{\mathbf{f}^*}^{-1} \mathbf{H}^T \mathbf{g} - \mathbf{g})^T \mathbf{H} \frac{\partial}{\partial \alpha_j} \mathbf{f}^* \\
&= 2(\mathbf{H}^T \mathbf{H} \mathbf{K}_{\mathbf{f}^*}^{-1} \mathbf{H}^T \mathbf{g} - \mathbf{H}^T \mathbf{g})^T \frac{\partial}{\partial \alpha_j} \mathbf{f}^* = 2\mathbf{g}^T \mathbf{H} \mathbf{K}_{\mathbf{f}^*}^{-T} (\mathbf{H}^T \mathbf{H} - \mathbf{K}_{\mathbf{f}^*})^T \frac{\partial}{\partial \alpha_j} \mathbf{f}^* \\
&= -2(\mathbf{f}^*)^T \left( \frac{1}{2} \sum_{i=1}^M \alpha_i \mathbf{R}_i^T \text{diag} \left[ \frac{\phi_i'([\mathbf{R}_i \mathbf{f}^*]_k)}{[\mathbf{R}_i \mathbf{f}^*]_k} \right] \mathbf{R}_i \right) \frac{\partial}{\partial \alpha_j} \mathbf{f}^* \\
&= -\sum_{i=1}^M \alpha_i \frac{\partial x_i}{\partial \alpha_j},
\end{aligned} \tag{15}$$

where the last step follows from (14). Substituting (15) into (12) we obtain the following equation for  $j = 1, \dots, M$ :

$$\sum_{i=1}^M ((\psi[x_i] - b_i)\psi'[x_i] - \alpha_i(\psi[z] - a)\psi'[z]) \frac{\partial x_i}{\partial \alpha_j} = 0. \tag{16}$$

Note that (16) is actually a collection of  $M$  different equations. We can arrange those  $M$  equations into a matrix–vector equation:

$$\mathbf{J}\mathbf{r} = 0, \tag{17}$$

where  $[\mathbf{J}]_{j,i} = \frac{\partial x_i}{\partial \alpha_j}$  and

$$[r]_i = (\psi[x_i] - b_i)\psi'[x_i] - \alpha_i(\psi[z] - a)\psi'[z], \quad i = 1, \dots, M. \tag{18}$$

If  $\mathbf{J}$  is nonsingular, (17) has only the trivial solution  $\mathbf{r} = \mathbf{0}$ . However, the non-singularity of  $\mathbf{J}$  follows from our assumption that the surface is regular and (15) [5].

Thus (17) implies  $\mathbf{r} = \mathbf{0}$ . Therefore, the solution of (16) is given by

$$\alpha_i^* = \frac{\psi'[x_i] \psi(x_i[\alpha^*]) - b_i}{\psi'[z] \psi(z[\alpha^*]) - a}, \quad i = 1, \dots, M. \tag{19}$$

If  $\psi(t) = \log t$ , (19) reduces to

$$\alpha_i^* = \frac{z(\alpha^*)}{x_i(\alpha^*)} \left( \frac{\log x_i(\alpha^*) - b_i}{\log z(\alpha^*) - a} \right), \quad i = 1, \dots, M. \tag{20}$$

Because (19)  $x_i = x_i(\alpha^*)$  and  $z = z(\alpha^*)$  are also functions of  $\alpha^*$ , (20) defines  $\alpha^*$  implicitly. Based on the formula in (20), we propose the following iterative algorithm to approximate  $\alpha^*$  in log scale:

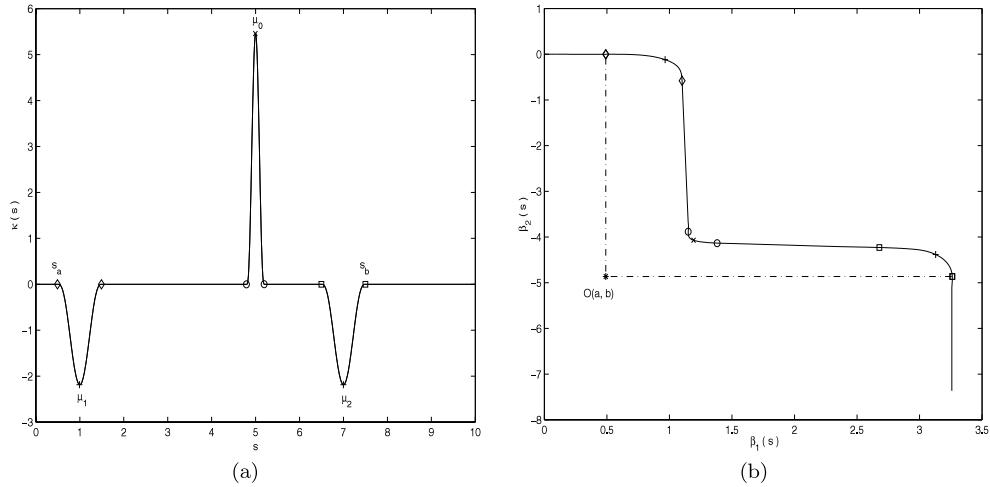
$$\alpha_i^{(k+1)} = \frac{z(\alpha^{(k)})}{x_i(\alpha^{(k)})} \left( \frac{\log[x_i(\alpha^{(k)})] - b_i}{\log[z(\alpha^{(k)})] - a} \right), \quad i = 1, \dots, M \tag{21}$$

where  $\alpha^{(k)}$  is the vector of regularization parameters at step  $k$ . The algorithm is started with an appropriate initial regularization parameter vector  $\alpha^{(0)}$  and iterated until the relative change in the iterates is determined to be sufficiently small.

Under some assumptions, we are able to prove that if  $M = 1$  (with  $\phi(t) = \log_{10}(t)$ ), the fixed-point iteration converges to a minimum of  $v$ .

**Theorem 1.** Assume  $v(t)$  has only one critical point, say  $t^* > 0$ , in some interval  $(\epsilon_1, \epsilon_2)$  with  $\alpha_a < \epsilon_1 < \epsilon_2 < \alpha_b$ , and that  $v$  is a minimum at that critical point. Further, assume that  $z(t), -x(t)$  are strictly increasing functions of  $t$  over the interval<sup>5</sup>. Let the origin  $(a, b)$  be

<sup>5</sup> This can be shown to be the case for a particular class of  $\phi(t)$ .



**Figure 2.** (a) Assumed curvature function. (b) The corresponding L-curve. The coordinates of the origin are given by  $O = (a, b) = (\beta_1(s_a), \beta_2(s_b))$ .

defined as in (10). Then if the starting guess  $t^{(0)}$  satisfies either  $\epsilon_1 < t^{(0)} < t^*$  or  $t^* < t^{(0)} < \epsilon_2$  with

$$\frac{z(t)}{z(\alpha_a)} \geq 10,$$

and

$$\frac{x(t)}{x(\alpha_b)} > 10,$$

for all  $t$  between  $t^{(0)}$  and  $t^*$  then the fixed-point iteration (21) converges to  $t^*$ .

**Proof.** See appendix. □

As a consequence, we know that if we pick a starting point for the single-dimensional fixed-point iteration which satisfies the hypotheses, the iterates will all be positive. For the multi-dimensional fixed-point algorithm, it is difficult to determine if and when the algorithm is guaranteed to converge. We therefore leave this study of convergence for the multi-dimensional case for future research, but note that in practice, this has not been a difficulty given judicious choice of origin.

### 3.3. Properties of the MDF

In this section, we describe the relationship of the minimum MDF point to the point of maximum curvature. For this discussion, we restrict our attention to the case of a single regularization parameter (i.e. an L-curve) where the necessary analysis is quite tractable. We emphasize that numerical experiments demonstrating the utility of this method for the multi-parameter case for a variety of applications are provided in section 4. We leave to future work the substantially more difficult analysis of the L-hypersurface.

For the L-curve we wish to show that under appropriate circumstances, the point chosen by the MDF is ‘close’ if not the same as that selected by maximizing the curvature. Essentially then we are interested in an analysis based on the geometric structure of the L-curve specifically in the neighbourhood of the corner. To facilitate this work we introduce a parametrized model

for the L-curve which emphasizes this geometry. All analysis is done in the context of this model. This model-based approach allows us to state that for true L-curves (i.e. ones generated by a linear inverse problem) possessing an appropriate corner geometry one can expect the MDF method to work comparably to the more intensive curvature maximization approach.

We start the development of the model L-curve by specifying its curvature function which we represent with three little bumps as seen in figure 2:

$$\kappa(s) = \theta_0 \frac{1}{\sigma_0} \mathcal{G}\left(\frac{s - \mu_0}{\sigma_0}\right) - \theta_1 \frac{1}{\sigma_1} \mathcal{G}\left(\frac{s - \mu_1}{\sigma_1}\right) - \theta_2 \frac{1}{\sigma_2} \mathcal{G}\left(\frac{s - \mu_2}{\sigma_2}\right) \quad (22)$$

where subscripts 1 and 2 represent false corners where the L-curve is concave, subscript 0 represents the desired corner, and  $\mu$  and  $\sigma$  are parameters adjusting the location and spread of the bumps. The function  $\mathcal{G}(s)$  is given by

$$\mathcal{G}(s) = \begin{cases} \frac{35}{32}(1 - s^2)^3, & \text{for } |s| \leq 1, \\ 0 & \text{for } |s| > 1. \end{cases} \quad (23)$$

For convenience, we choose to parametrize the curvature function  $\kappa(s)$  in terms of arclength,  $s = \int_{\alpha_0}^{\alpha} \sqrt{|\frac{d\psi(x)}{d\alpha}|^2 + |\frac{d\psi(z)}{d\alpha}|^2} d\alpha$ . Since  $x(\alpha)$  and  $z(\alpha)$  are continuous functions of  $\alpha$  and  $\frac{ds}{d\alpha}$  never vanishes, there is a one-to-one correspondence between  $s$  and  $\alpha$ . That is, given any  $s$  we can uniquely determine the corresponding  $\alpha$  value [17].

There is a unique plane curve (up to a rigid motion) realizing  $\kappa(s)$  in (22), as its curvature [5]:

$$\beta(s) = \left( \int_0^s \cos \rho(u) du, \int_0^s \sin \rho(u) du \right) \quad (24)$$

$$\rho(u) = \int_0^u \kappa(t) dt. \quad (25)$$

In our case, an explicit analytic formula for  $\beta(s)$  cannot be found. However, (25) can be numerically solved in differential form. Figure 2 shows the L-curve obtained from  $\kappa(s)$  in (22). It is easy to see from this figure that  $\theta_0 > \frac{\pi}{2}$  determines the angle between the approximately horizontal and vertical parts of the L-curve to the right and left of the corner. On the other hand,  $\sigma_0$  is a measure of the width of the crossover region (i.e. the regions separating horizontal and vertical parts). As  $\sigma_0$  approaches zero, the corner of the L-curve becomes sharper.

Using this L-curve model we now show that the  $\alpha$  minimizing the MDF essentially maximizes the curvature of the L-curve. To begin with, we define the tangent vector to the L-curve,  $t(s)$ :

$$t(s) = \frac{d\beta(s)}{ds} = \left( \frac{d\beta_1(s)}{ds}, \frac{d\beta_2(s)}{ds} \right). \quad (26)$$

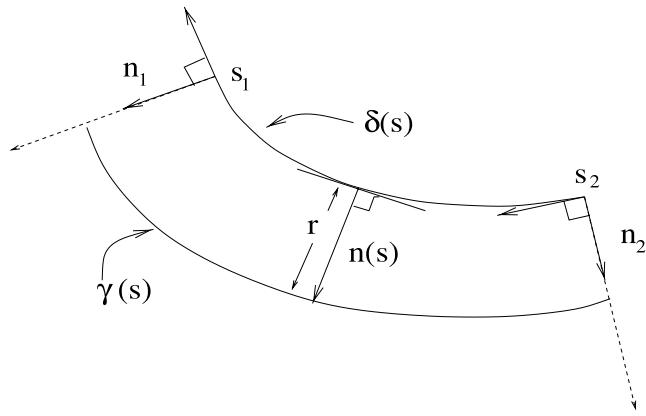
Based on (26), the unit normal,  $n(s)$ , is defined as the unit length vector perpendicular to  $t(s)$ .

Considering the actual behaviour of an L-curve, it is easy to see that  $t(0)$  is parallel to the  $\beta_1$  axis (L-curve becomes horizontal as  $\alpha \rightarrow 0$ ) and that  $t(s)$  becomes parallel to the  $\beta_2$  axis as  $s \rightarrow \infty$  (L-curve becomes vertical as  $\alpha \rightarrow \infty$ ). This, in turn imposes the following constraints on  $\theta_i$ ,  $i = 0, 1, 2$ :

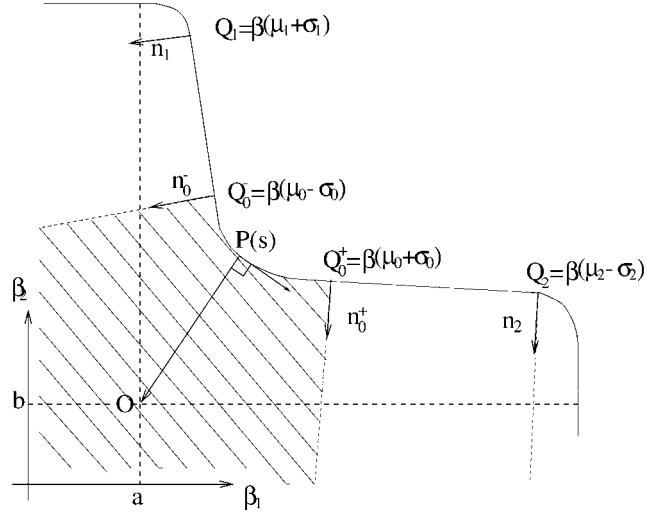
$$\theta_1 - \theta_0 + \theta_2 = \frac{\pi}{2} \quad (27)$$

$$\theta_0, \theta_1, \theta_2 < \frac{\pi}{2}. \quad (28)$$

Equation (27) ensures that there are exactly  $\frac{\pi}{2}$  degrees between the tangent vectors  $t(0)$  and  $t(\infty)$  and (28) is a natural consequence of the fact that  $\beta_2(s)$  is a monotonically decreasing function of  $\beta_1(s)$  [20].



**Figure 3.** Parallel curve of a convex plane curve.



**Figure 4.** A typical L-curve.

We denote the origin chosen for the computation of  $v(s)$  by  $O = (a, b)$ . Point  $a$  is associated with a value  $\beta_1(s_a)$  where  $s_a < \mu_0 - \sigma_0$  and point  $b$  is associated with a value  $\beta_2(s_b)$  where  $s_b > \mu_0 + \sigma_0$ . Points on the L-curve where  $s = \mu_i \pm \sigma_i, i = 0, 1, 2$  carry a special importance for us, namely they represent the points where  $\kappa(s)$  is zero and the L-curve switches between linear and non-linear.

Before proceeding any further, we introduce the concept of *parallel transport* of a convex plane curve (i.e.  $\kappa(s) \geq 0$ ).

**Definition 3.** Let  $\delta(s)$  be a convex plane curve ( $\kappa(s) \geq 0$ ) positively oriented. The curve

$$\gamma(s) = \delta(s) + rn(s) \quad (29)$$

where  $r > 0$  is a constant and  $n$  is the unit normal, is called a parallel curve to  $\delta(s)$  (figure 3).

It is easy to see from the definition that the parallel curve of a plane curve is obtained by simply expanding the curve by a constant amount along the direction of the normal. The

*parallel transport* of a convex plane curve  $\delta(s)$ , denoted by  $\mathcal{P}(\delta)$ , is defined as the region covered by the parallel curves  $\gamma(s)$  for all  $r \neq 0$  and all  $s$  such that  $\gamma(s)$  is defined. We use the notation  $\delta_{[s_1, s_2]}(s)$  to denote the part of the curve  $\delta$  restricted to  $s \in [s_1, s_2]$ . Thus for the L-curve  $\beta$  in our example,  $\mathcal{P}(\beta_{[s_1, s_2]})$  covers all those points on the plane from which we can draw a line perpendicularly intersecting the curve  $\beta_{[s_1, s_2]}$ . This property of a parallel transport will play a crucial role in the proofs of theorems 2–5.

Our first two results, theorems 2 and 3, tell us that by placing the origin  $O = (a, b)$  inside the region bounded by the perpendicular lines at zero-curvature points  $s = \mu_0 \pm \sigma_0$  on the sides and the L-curve above, we can actually create a function  $v(s)$  such that the minimum of  $v(s)$  is close to the corner of the L-curve (point on the L-curve for which  $s = \mu_0$ ) and that  $v(s)$  possesses a unique minimum for a wide range of  $s$  values. These are, of course, desirable properties for a surrogate function replacing the curvature since our initial goal was to create a function approximating the corner of the L-curve and having nice characteristics for the purpose of optimization.

**Theorem 2.**  *$v(s)$  has a unique local minimum at  $s^* \in [s_0^-, s_0^+] \equiv [\mu_0 - \sigma_0, \mu_0 + \sigma_0]$  if  $O = (a, b)$  is in the region bounded above by the part of the L-curve lying between  $Q_0^- = \beta(\mu_0 - \sigma_0)$  and  $Q_0^+ = \beta(\mu_0 + \sigma_0)$  and the semi-infinite rays  $n_0^-, n_0^+$  emanating from the points  $Q_0^-$  and  $Q_0^+$  and perpendicular to the L-curve at the cited points (shaded region in figure 4).*

**Proof.** See appendix. □

**Theorem 3.** *Let  $O$  be an origin satisfying the hypothesis of theorem 2. Let  $s^*$  be the corresponding point minimizing  $v(s)$  in  $[s_0^-, s_0^+] \equiv [\mu_0 - \sigma_0, \mu_0 + \sigma_0]$ . Then  $s^*$  is the unique minimum of  $v(s)$  for all  $s$  in  $(\mu_1 + \sigma_1, \mu_2 - \sigma_2)$ .*

**Proof.** See appendix. □

Although theorems 2 and 3 tell us a great deal about the behaviour of  $v(s)$ , they do not tell us how to choose an appropriate origin satisfying the condition in theorem 2. However, as we will see in the next two theorems the choice of the origin  $O = (a, b)$  is not crucial for a well-behaved L-curve.

**Theorem 4.** *As  $\theta_0$  approaches  $\frac{\pi}{2}$ , any origin  $O = (a, b)$ , such that  $a = \beta_1(s_a)$  where  $s_a < \mu_0$  and  $b = \beta_2(s_b)$  where  $s_b > \mu_0$ , lies in  $\mathcal{P}(\beta_{[\mu_0 - \sigma_0, \mu_0 + \sigma_0]})$  for  $s \in (\mu_0 - \sigma_0, \mu_0 + \sigma_0)$  and  $r > 0$ .*

**Proof.** See appendix. □

Now, we are ready to prove our final result.

**Theorem 5.** *Denote the point where  $v(s)$  achieves a local minimum in  $[s_0^-, s_0^+] \equiv [\mu_0 - \sigma_0, \mu_0 + \sigma_0]$  by  $s^*$ . Let  $O = (a, b) \in \mathcal{P}(\beta_{[s_0^-, s_0^+]})$  for  $s \in [\mu_0 - \sigma_0, \mu_0 + \sigma_0]$  and  $r > 0$ . Then,  $\lim_{\sigma_0 \rightarrow 0} s^* = \mu_0$ .*

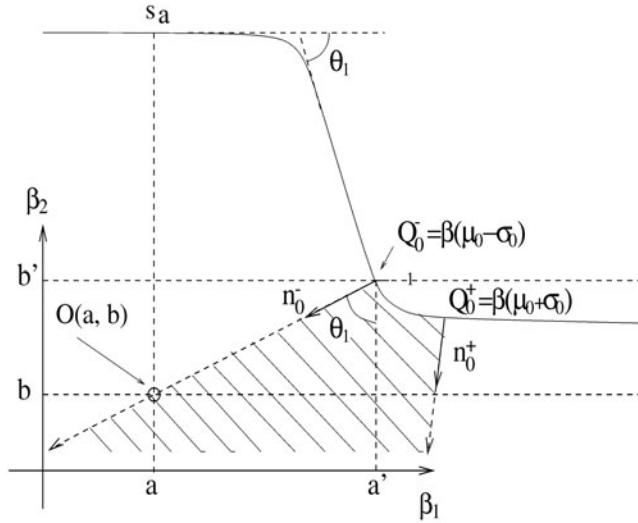
**Proof.** By theorem 2,

$$\mu_0 - \sigma_0 \leq s^* \leq \mu_0 + \sigma_0.$$

The desired result is obtained by letting  $\sigma_0 \rightarrow 0$ . □

By combining the results of theorems 4 and 5 we obtain the following result.

**Corollary 1.** *Denote the point where  $v(s)$  achieves a local minimum in  $s \in (\mu_0 - \sigma_0, \mu_0 + \sigma_0)$  by  $s^*$ . Let  $O = (a, b)$  be such that  $a = \beta_1(s_a)$  where  $s_a < \mu_0$  and  $b = \beta_2(s_b)$  where  $s_b > \mu_0$ . Then,  $\lim_{\sigma_0 \rightarrow 0, \theta_0 \rightarrow \frac{\pi}{2}} s^* = \mu_0$ .*



**Figure 5.** Illustration for the proof of theorem 4. Shaded region is  $\mathcal{P}(\beta_{[\mu_0 - \sigma_0, \mu_0 + \sigma_0]})$ .

In other words, corollary 1 says that as  $\theta_0 \rightarrow \frac{\pi}{2}$  and  $\sigma_0 \rightarrow 0$  (e.g. the more the curve looks like the letter L), the point  $s^*$  for which  $v(s^*)$  is a minimum coincides with the corner of the L-curve  $\mu_0$  no matter where we choose the origin (provided that  $a$  falls to the right of the corner and  $b$  falls below the corner).

In the case of multiple regularization parameters, we can make a qualitative argument as to how one might begin extending the above discussion to the L-hypersurface. To begin, examining figure 1(b) reveals that the L-hypersurface is convex in the vicinity of the maximum Gaussian curvature point,  $\alpha^*$  (i.e.  $\kappa(\alpha) > 0$ ). Therefore, the L-hypersurface has a bowl-shaped appearance around  $\alpha^*$  and any point on the L-hypersurface lies above the tangent plane at  $\alpha^*$ . Hence, the unit normal to the L-hypersurface at  $\alpha^*$ ,  $N(\alpha^*)$ , defines a line whose points, when used as an origin for the computation of  $v(\alpha)$ , yields a MDF which has a local minimum at  $\alpha^*$ . Thus, if we choose our origin  $O$  in the close vicinity of the line defined by  $N(\alpha^*)$ , the minimum of  $v(\alpha)$  hits a close point to the generalized corner of the L-hypersurface. A rigorous analysis of these ideas is a far from trivial task and one which we leave to future efforts.

### 3.4. Origin selection in practice

The previous discussion indicated that for well behaved L-curves we expect that the ability of the MDF to find the corner of the L-curve should be fairly insensitive to the choice of MDF origin. *A priori* of course, we have no knowledge as to how well behaved the L-curve is and therefore a method for selecting the origin must be provided to practically implement the MDF technique. We give a heuristic and its justification below. A more robust means of selecting the origin is a subject for future research.

We have found the following heuristic approach to be reliable (see section 4). For the single-parameter case, we take  $O = (\log z(\sigma_{min}^2), \log x(\sigma_{max}^2))$ , where  $\sigma_{min}, \sigma_{max}$  denote the smallest and largest singular values (or approximations thereof) of  $H$ . The justification is as follows. We would like the origin to fall inside the shaded region (see figure 5) to guarantee a good estimate of the corner. The more well behaved the curve, the less sensitive the method is to the origin placement (i.e. the ‘wider’ the shaded region and the easier it is to put the origin in the shaded region).

Pick a small parameter that is either near the smallest singular value or, in the case of numerically singular matrices, just large enough so that the regularized problem is well defined. As long as the curve is monotonically decreasing, the horizontal component of the origin chosen in this way will be to the left of corner as desired. Then it remains to pick a vertical component so that the origin falls in the shaded region. Unless the ‘large’ parameter we use to pick the vertical component is very large (i.e. very far down the near vertical part past the lower right ‘knee’) and the curve is not well behaved (i.e. the normal  $n_0^+$  in figure 5 is far from being vertical), an origin selected this way will be inside the shaded region. In analogy with the two-norm analyses of Reginska [20] and Hansen [14], we suggest a value about as large as the largest singular value should give a point near the bend in the lower right knee and therefore we should be able to avoid this potential problem with using too large a parameter.

This approach can be extended to cases where there are more than one regularization functional. Noting that the two points in  $R^3(z(\sigma_{min}^2, 0), x_1(\sigma_{max}, 0), 0)$  and  $(z(0, \sigma_{min}^2), 0, x_2(0, \sigma_{max}))$  would be the 1D origin guesses if either  $\alpha_2$  or  $\alpha_1$  were 0, respectively, for  $M = 2$  we take<sup>6</sup>  $O = (\log z(\sigma_{min}^2, \sigma_{min}^2), \log x_1(\sigma_{max}^2, 0), \log x_2(0, \sigma_{max}^2))$ , and analogously for  $M > 2$ . (Similar to the discussion in the preceding paragraph, if the matrix is numerically singular than an appropriate multiple of the machine precision should replace  $\sigma_{min}$ .)

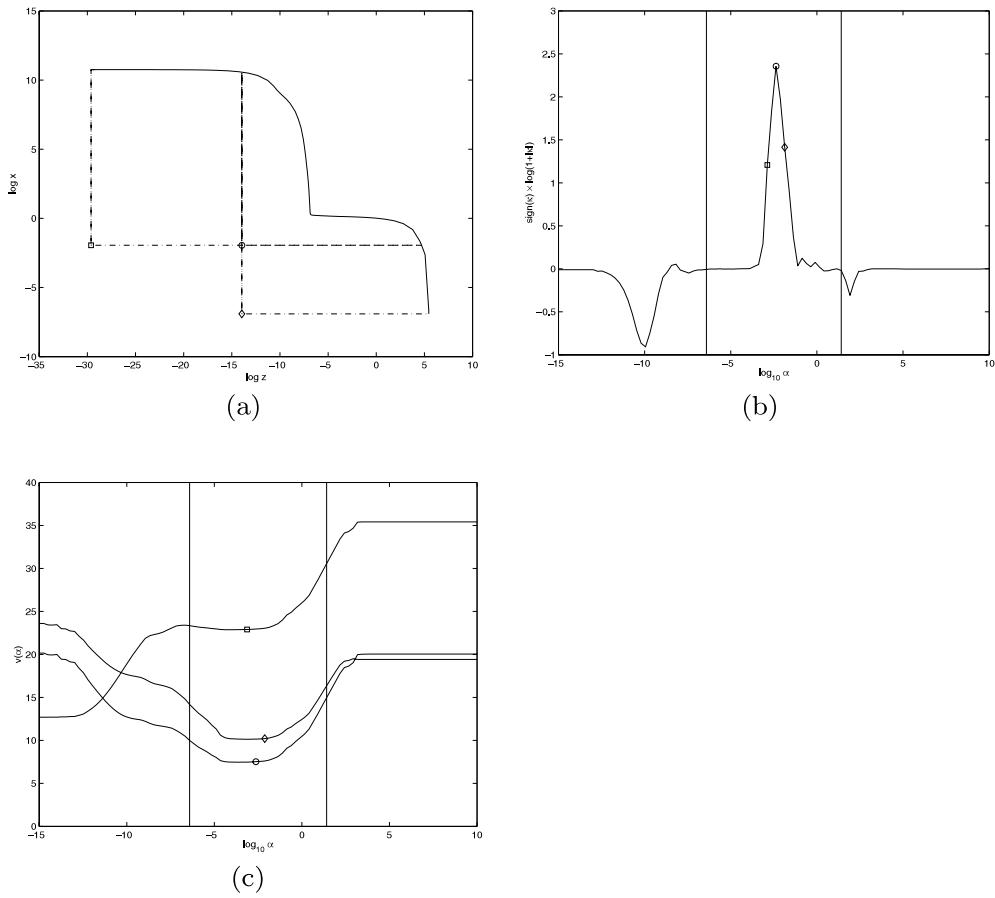
#### 4. Numerical examples

In this section, we verify the statements made concerning the behaviour of the MDF and demonstrate the effectiveness of the iterative algorithm derived in section 3.2 for both one- and multi-dimensional parameter selection problems.

##### 4.1. Example 1

We generated a test problem of the form  $Hf = \bar{g}$  by using the function `phillips` (100) in Hansen’s regularization toolbox [12] in MATLAB. We modified the exact right-hand side  $\bar{g}$  by adding normally distributed noise,  $n$ , scaled so that  $\frac{\text{variance}(n)}{\text{variance}(\bar{g})} = 10^{-5}$ . We employed total variation (TV) regularization [18, 23] to obtain the regularized solution  $f$ . TV is a special case of (3) where  $M = 1$ ,  $\phi_1(t) = |t|$  and  $R_1$  is a discrete approximation to first-order differentiation. In order to facilitate an iterative solution as in [23] we slightly modified the cost functional by substituting  $\phi_1(t) = \sqrt{|t|^2 + \beta}$  where  $\beta > 0$  is a stabilization parameter. We used  $\beta = 10^{-8}$  in our simulations. The L-curve for this problem was computed by sampling  $\alpha$  in 100 logarithmically equi-spaced points between  $10^{-15}$  and  $10^{10}$ . The resulting L-curve is displayed in figure 6(a). We chose three different origins and computed corresponding  $v(\alpha)$  functions. Each one of the three origins chosen were indicated by the following symbols (figure 6(a)): a circle, a diamond and a square. The origin indicated by a circle is especially important since it is the one we advocated using. The other two origins, indicated by the diamond and the square symbols, are arbitrary and included here to see how sensitive the resulting regularization parameter is to the selection of the origin. Coordinates of the origin indicated by the circle were calculated as follows: first we estimated the smallest and largest singular values of the matrix  $H$  which were found to be  $\sigma_{min} \approx 2.2 \times 10^{-6}$  and  $\sigma_{max} \approx 5.8$ . According to the preceding discussion, we took the coordinates of the origin as  $O = (a, b) = (\log z(\sigma_{min}^2), \log x(\sigma_{max}^2))$  (where  $\sigma_i$  denote the generalized singular values rather than singular values). In figure 6(c) we display  $v(\alpha)$  functions for each of the three origins selected. The minimum of  $v(\alpha)$  for each

<sup>6</sup> If easy to compute, the appropriate  $\sigma$  values should be replaced by estimates of the generalized singular values for the corresponding 1D problems.



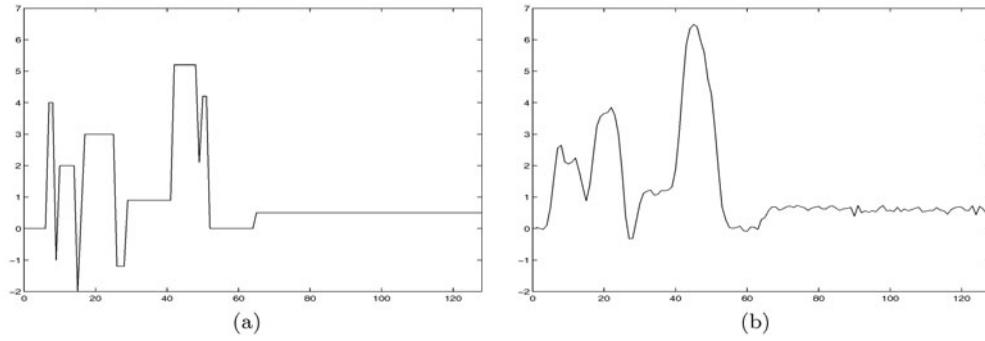
**Figure 6.** Example 1. (a) L-curve for problem phillips. (b) Curvature of the L-curve in (a). (c) Three different MDF for different choices of the origin.

case is marked with the appropriate symbol from figure 6(a). Also shown in figure 6(c) is the plot of  $-\text{sign}(\kappa(\alpha)) \times \log(1 + |\kappa(\alpha)|)$  and the location of the minimum of the  $v(\alpha)$  functions for each origin. The region in between the vertical lines in figures 6(b) and (c) represents the part of the L-curve for which  $\kappa(\alpha) \geq 0$ . It is nicely seen from figure 6(c) that for all three cases the minimum of  $v(\alpha)$  is inside the crossover region of the L-curve and that the minimum of  $v(\alpha)$  for origin indicated by the circle comes very close to the maximum curvature point.

#### 4.2. Example 2

In our second example, we demonstrate the utility of the MDF in a multiple regularization parameter setting. The test problem of interest was generated by convolving the 128-point signal displayed in figure 7(a) with a Gaussian kernel,  $h(i) = \frac{1}{\sqrt{2\sigma^2\pi}} \exp\{i^2/2\sigma^2\}$ , with  $\sigma = 2.0$  and  $i = -30, \dots, 30$ . The exact right-hand side is again modified by adding normally distributed random noise scaled so that  $\frac{\text{variance}(n)}{\text{variance}(\hat{g})} = \times 10^{-2}$ . The blurred and noisy signal obtained in this way is displayed in figure 7(b). We obtained the regularized solution by using Tikhonov's regularization in the following way:

$$f^*(\alpha_1, \alpha_2) = (\mathbf{H}^T \mathbf{H} + \alpha_1 \mathbf{R}_1 + \alpha_2 \mathbf{R}_2)^{-1} \mathbf{H}^T \mathbf{g} \quad (30)$$



**Figure 7.** Example 2. (a) Original 128-point signal. (b) Blurred and noisy version of the signal in (a).

where  $\mathbf{R}_1$  is an operator extracting the first half of the signal and  $\mathbf{R}_2$  is an operator extracting the second half of the signal so that  $\mathbf{R}_1 + \mathbf{R}_2 = \mathbf{I}_{128 \times 128}$  is the  $128 \times 128$  identity matrix. Such a selection for the regularization operators is motivated by the fact that the second half of the blurred, noisy signal (from point 65 to point 128) appears much smoother than the first half. Thus, it is possible to decrease noise artifacts in the second half of the restored signal by applying more regularization.

For this problem, the minimum and the maximum singular values of  $\mathbf{H}$  were  $\sigma_{\min} \approx 1.3 \times 10^{-4}$  and  $\sigma_{\max} \approx 1.3$ . The origin chosen for the computation of the MDF was  $O = (\log z(\sigma_{\min}^2, \sigma_{\min}^2), \log x_1(\sigma_{\max}^2, 0), \log x_2(0, \sigma_{\max}^2))$ .

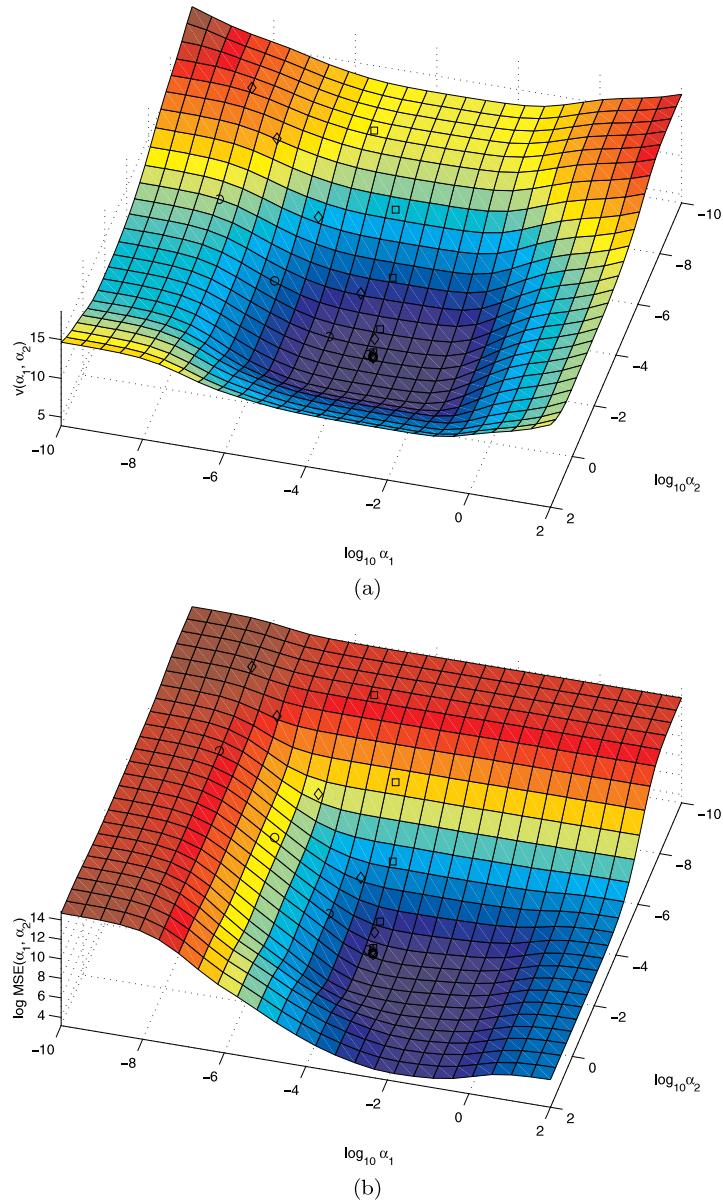
We computed the MDF ( $v$ ) and the mean square error (MSE),  $\frac{1}{N} \|f - f^*(\alpha_1, \alpha_2)\|_2^2$ , by sampling regularization parameters at 26 logarithmically equi-spaced points between  $10^{-10}$  and  $10^2$ . The resulting MDF and MSE surfaces are displayed in figures 8(a) and (b). We compared the performance of BFGS with line search, which is guaranteed to converge to a minimum of the MDF, with our fixed-point iteration. Ideally, we are only interested in  $\alpha \geq 0$ . We note that neither of these methods is guaranteed to have non-negative iterates, but since the non-negativity constraint is not violated at the minimum of  $v$  and both converge to the minimum for reasonable starting points, we chose to ignore the non-negativity constraint.

We started our iterative algorithm with three different initial values of  $\alpha = [\alpha_1^0, \alpha_2^0]^T$ . The stopping criteria we used for BFGS was that the norm of the gradient be less than  $10^{-6}$ . The stopping criteria we used for our fixed point algorithm was

$$\max_{\forall i} |\alpha_i^k - \alpha_i^{k-1}| / |\alpha_i^{k-1}| \leq 10^{-4}. \quad (31)$$

In fact,  $10^{-4}$  may be a smaller tolerance than is necessary to get good  $\alpha_1, \alpha_2$ ; however, since BFGS tended to converge to solutions where this measure was of order of  $10^{-4}$ , this tolerance was useful for comparison purposes.

For each run, the points computed by our algorithm at each iteration are indicated on both the MDF and MSE surfaces. Although it may be difficult to see from the figure, the optimal parameters were an order of magnitude apart, thereby justifying the use of two regularization terms. In figures 9(a) and (b), circles indicate the trajectory of the algorithm for  $\alpha^{(0)} = [10^{-8}, 10^{-5}]$ , diamonds indicate the trajectory of the algorithm for  $\alpha^{(0)} = [10^{-8}, 10^{-8}]$  and squares indicate the trajectory of the algorithm for  $\alpha^{(0)} = [10^{-5}, 10^{-8}]$ . In all three cases, our fixed-point algorithm converged to the same point ( $\alpha = [3.5 \times 10^{-3}, 2.2 \times 10^{-2}]^T$ ) in fewer than nine iterations. Recall that for every iteration, one system of the form (30) needs to be solved.



**Figure 8.** Example 2. (a) Plot of  $v$  function. (b) Corresponding MSE surface. Circles, diamonds and squares indicate the trajectory of our iterative algorithm for different starting points.

In contrast, BFGS took 26 or 27 iterations, depending on the starting point, to converge using the stopping criteria based on the norm of the gradient. Further, it took about 47 function evaluations plus the same number of gradient evaluations to reach convergence for each of the three cases. Each function evaluation is equivalent to solving one linear system of the form (30). Additionally, each gradient evaluation requires solving two additional linear systems of the form (30). Thus, about  $47 \times 3$  linear systems need to be solved before convergence is reached. Therefore, the number of linear systems required for BFGS with line search to reach convergence was roughly 15 times more than for our fixed-point algorithm.

#### 4.3. Example 3

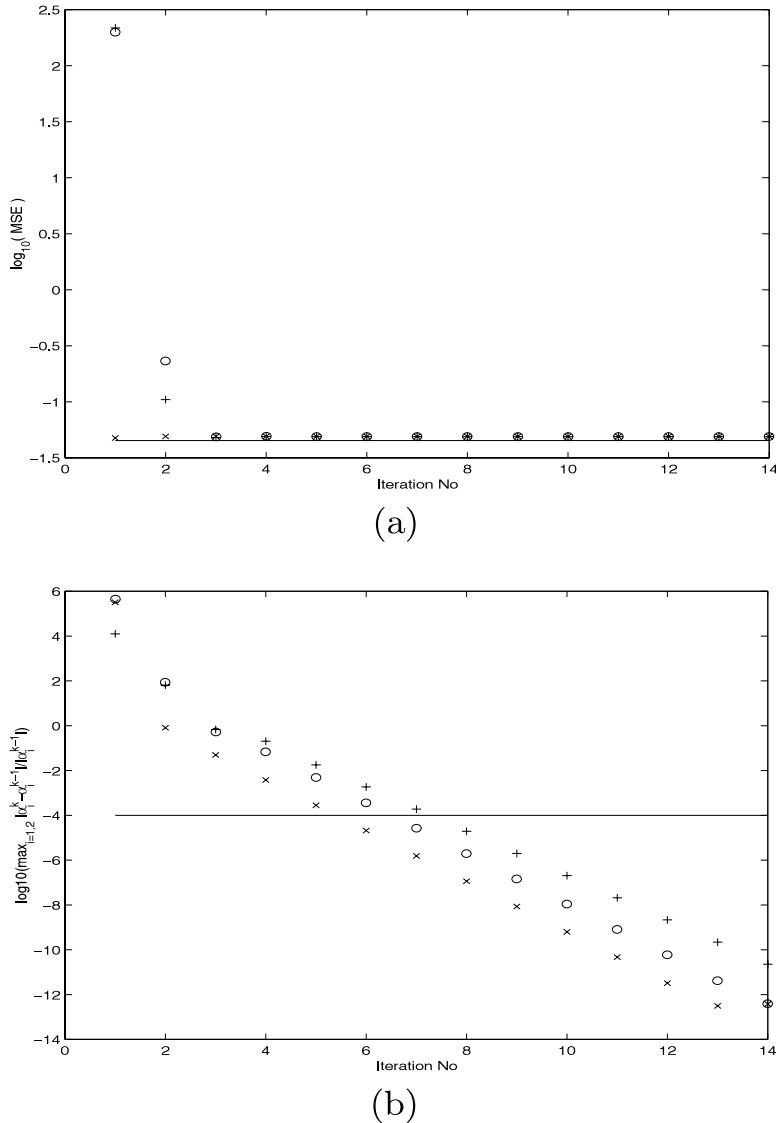
Our next example is a wavelet domain signal restoration experiment where we use  $L_1$  norm as the regularization functional. The original 256 point *Blocks* sequence, extracted from Donoho's WaveLab package<sup>7</sup>, is convolved with a Gaussian kernel,  $h(i) = \frac{1}{\sqrt{2\sigma^2\pi}} \exp\{(i-m)^2/2\sigma^2\}$ , with parameters  $\sigma = 4.0$  and  $m = 128$ . Zero mean normally distributed random noise was added to the degraded signal so that  $\frac{\text{variance}(n)}{\text{variance}(\hat{g})} = 4.62 \times 10^{-5}$ . We obtained the regularized solution by a wavelet domain signal restoration algorithm developed by the authors [3]. This regularization approach is a special case of (3) where  $\mathbf{g}$ ,  $\mathbf{H}$ ,  $\mathbf{f}$  are the wavelet domain representations of the measurements, the convolution operator and the original signal,  $M = 2$ ,  $\phi_i(t) = \sqrt{|t|^2 + 0.01}$ ,  $\mathbf{R}_1$  is the operator extracting the scaling coefficients and  $\mathbf{R}_2$  is the operator extracting the wavelet coefficients of  $\mathbf{f}$ . Hence, in this experiment,  $\alpha_1$  is the regularization parameter for the scaling coefficients and  $\alpha_2$  is the regularization parameter for the wavelet coefficients of the original signal. For this problem, we determined that the minimum and the maximum singular values of  $\mathbf{H}$  were  $\sigma_{\min} \approx 2.2 \times 10^{-19}$  and  $\sigma_{\max} \approx 1.0$ . The origin chosen for the computation of the MDF was  $O = (\log z(\sigma_{\min}^2, \sigma_{\min}^2), \log x_1(\sigma_{\max}^2, 0), \log x_2(0, \sigma_{\max}^2))$ .

As in the previous example, we started our iterative algorithm with three different initial values for  $\alpha = [\alpha_1^0, \alpha_2^0]^T$ , and ran our algorithm until (31) was satisfied. In figures 9(a) and (b), '+' indicates the trajectory of the algorithm for  $\alpha^{(0)} = [10^{-3}, 10^{-10}]$ , 'o' is the trajectory of the algorithm for  $\alpha^{(0)} = [10^{-10}, 10^{-10}]$  and 'x' indicates the trajectory of the algorithm for  $\alpha^{(0)} = [10^{-10}, 10^{-3}]$ . In figure 9, a solid line represents the minimum attainable MSE,  $\frac{1}{N} \|\mathbf{f} - \mathbf{f}^*(\alpha_1, \alpha_2)\|_2^2$  (as determined by a search over hundreds of  $\alpha = [\alpha_1, \alpha_2]$ ) for this problem. The final MSE value obtained by our iterative algorithm was 0.0490 while the minimum MSE was 0.0457. Figure 9 shows the convergence of our iterative algorithm. As observed from this figure the convergence criterion in (31) is satisfied by the eighth iteration for all starting points.

#### 4.4. Example 4

Our final example is concerned with the restoration of an image from its blurred and noisy version. The original cameraman image selected for this example is displayed in figure 10. This image was convolved with a 2D Gaussian kernel,  $h(i, j) = \frac{1}{2\sigma^2\pi} \exp\{((i-m)^2 + (j-n)^2)/2\sigma^2\}$ , with parameters  $\sigma = 2.0$  and  $m = n = 128$ . Zero-mean normally distributed random noise was added to the degraded image so that  $\frac{\text{variance}(n)}{\text{variance}(\hat{g})} = 10^{-4}$ . To obtain the restored image, we used a three-parameter regularization scheme where  $L_2$  norms of the horizontal, vertical and diagonal first-order differences of the image were used as regularization functionals (i.e.  $\phi_i(t) = |t|^2$ ,  $\mathbf{R}_i$ ,  $i = 1, 2, 3$  are the horizontal, vertical and diagonal differencing operators in the image domain in (3)). For this problem, we determined that the minimum and the maximum singular values of  $\mathbf{H}$  were  $\sigma_{\min} \approx 10^{-18}$  and  $\sigma_{\max} \approx 10^{-2}$ . The origin chosen for the computation of the MDF was  $O = (\log z(\sigma_{\min}^2, \sigma_{\min}^2, \sigma_{\min}^2), \log x_1(\sigma_{\max}^2, 0, 0), \log x_2(0, \sigma_{\max}^2, 0), \log x_3(0, 0, \sigma_{\max}^2))$ . To obtain the regularized solution which uses the regularization parameters obtained by our algorithm, we started our iterative algorithm with the initial point  $\alpha = [10^{-10}, 10^{-10}, 10^{-10}]^T$ , and ran our algorithm until (31) was satisfied. It was found that the algorithm converged within 17 iterations. The restored image obtained in this way is displayed in figure 10(d). For comparison, we also found the regularization parameters that minimize the MSE

<sup>7</sup> The WaveLab documentation and package is available from <http://www-stat.stanford.edu/wavelab>.



**Figure 9.** (a) MSE versus iteration number for three different choices of initial regularization parameters for the *Blocks* problem. Solid line shows the minimum attainable MSE. (b) Convergence of the iterative algorithm for different choices of initial regularization parameters.

$\frac{1}{N} \|f - f^*(\alpha_1, \alpha_2, \alpha_3)\|_2^2$ , by performing an extensive search<sup>8</sup> over  $\alpha_1, \alpha_2, \alpha_3$ . Figure 10(c) displays the restored image obtained by using the best regularization parameters (in terms of minimizing the MSE). The ratio of the MSE value of our iterative algorithm to that of minimum MSE value was found to be 1.2686.

<sup>8</sup> Minimum MSE was determined by sampling  $\alpha_i$ -space using the MATLAB command `logspace(-9, 0, 20)`, for a total of 8000  $\alpha$  vectors tested.



**Figure 10.** (a) Original cameraman image. (b) Blurred, noisy cameraman image. (c) Restored by using the regularization parameters minimizing the MSE. (d) Restored by using the regularization parameters obtained by our iterative algorithm.

## 5. Conclusions and discussion

In this paper, we considered the problem of estimating multiple regularization parameters in a generalized L-curve framework. We defined a surrogate function, called the MDF, to replace the curvature function. The analysis we carried out on a hypothetical L-curve model indicated that, in the single-parameter case, the regularization parameters minimizing the MDF approximately maximize the curvature as the corner of the L-curve becomes sharper. This latter point was confirmed by numerical examples performed on actual L-curves. Further, numerical results indicated that the parameters chosen using our approach gave regularized solutions that were comparable to the optimal regularized solutions.

We also developed an iterative fixed-point algorithm to approximate the regularization parameters minimizing the MDF. In the case of a single regularization parameter, we were able to prove that the fixed point converges to a minimum of the MDF under certain assumptions. It was shown through numerical experiments that the iterative algorithm quickly converges. Thus, the computational effort associated with computing approximations to the regularization parameters that correspond to the generalized corner of the L-hypersurface has been greatly reduced. The potential tradeoff is a slight degradation in the MSE of the reconstruction if the origin chosen is not optimal.

### Acknowledgments

This work was supported by an ODDR&E MURI under Air Force Office of Scientific Research contract F49620-96-1-0028, a CAREER Award from the National Science Foundation MIP-9623721, and the Army Research Office Demining MURI under Grant DAAG55-97-1-0013.

### Appendix. Proofs of theorems

**Proof of theorem 1.** First, define the iteration function as

$$\Theta(t) \equiv \frac{z(t)}{x(t)} \left( \frac{\log[x(t)] - b}{\log[z(t)] - a} \right).$$

Then the fixed point iteration is written  $t^{(k+1)} = \Theta(t^{(k)})$ .

The case  $t^{(0)} = t^*$  is trivial, so in the remainder, we assume  $t^{(0)} < t^*$ .

Let  $\mathcal{I}_1$  denote the interval  $[\epsilon_1, t^*]$  and let  $\mathcal{I}_2$  denote the interval  $[t^*, \epsilon_2]$ . For all  $t$  in  $\mathcal{I}_1$ , our assumptions imply  $v'(t) < 0$  whereas for  $t$  in  $\mathcal{I}_2$ , we must have  $v'(t) > 0$ . Differentiating (11) and using (15) we get

$$v'(t) = \frac{2}{\ln 10} x'(t) \left( \frac{1}{x(t)} (\log[z(t)] - a) - t \frac{1}{z(t)} (\log[x(t)] - b) \right) < 0.$$

But since  $x'(t) < 0$ , it follows that for all  $t$  in  $\mathcal{I}_1$ ,

$$\Theta(t) > t$$

whereas for all  $t$  in  $\mathcal{I}_2$ ,

$$\Theta(t) < t.$$

Now it is straightforward to show that the derivative of the iteration function is given as

$$\Theta'(t) = \zeta(t) \frac{z'(t)}{x} \left( 1 - \frac{1}{\log[z(t)] - a} \right) + \frac{z(t)(-x'(t))}{x^2(t)} \left( \zeta(t) - \frac{1}{\log[z(t)] - a} \right),$$

where

$$\zeta(t) = \left( \frac{\log(x(t)) - b}{\log(z(t)) - a} \right).$$

Using the fact that  $\frac{x(t)}{\alpha_b} \geq \frac{x(t^*)}{\alpha_b} > 10$  and  $\frac{z(t)}{\alpha_a} \geq 10$  together with the positivity of  $x, z, -x', z'$ , it is easy to show that  $\Theta'(t) > 0$  for  $t$  in  $\mathcal{I}_1$  and  $\mathcal{I}_2$ .

Using the mean value theorem and an induction argument, it follows from the fact that  $\Theta(t) > t$ ,  $t \in \mathcal{I}_1$  or  $\Theta(t) < t$ ,  $t \in \mathcal{I}_2$  together with the fact that  $\Theta'(t) > 0$  that if  $t^{(0)} < t^*$ , the iterates remain to the left of  $t^*$  while if the initial guess is to the right, the iterates remain to the right. Hence, the iteration is either producing an increasing sequence  $\{t^{(k)}\}_{k=0}^\infty$  bounded above by  $t^*$  or a decreasing sequence bounded below by  $t^*$ , depending on where the starting guess is. Therefore, the sequence started to the left of  $t^*$  converges to a point  $\hat{t}_1 \in \mathcal{I}_1$ , and the sequence started on the right converges to a point  $\hat{t}_2 \in \mathcal{I}_2$ . It is easy to show these limit points are fixed points. However, if  $\hat{t}_i \neq t^*$ , and yet  $\hat{t}_i$  is a fixed point,  $v'(\hat{t}_i) = 0$ , which is a contradiction of our assumption on  $v$  over  $\mathcal{I}_1$  and  $\mathcal{I}_2$ . Therefore, either sequence must converge to  $t^*$ .

To prove the results in section 3.3, we require the following theorem.

**Theorem 6.** Let  $Q$  be a point on the plane below the convex curve  $\delta(s)$ , which does not lie on  $\delta(s)$ . A line from  $Q$  to  $\delta(s)$  intersecting  $\delta(s)$  perpendicularly can be drawn iff  $Q \in \mathcal{P}(\delta)$ .

**Proof.** Follows from the definition of  $\mathcal{P}(\delta)$ .  $\square$

**Proof of theorem 2.** Given the point  $O = (a, b)$ , by the definition of parallel transport and theorem 6 there exists a scalar  $r^* > 0$  and  $s^* \in [s_0^-, s_0^+]$  such that  $O = \beta_{[s_0^-, s_0^+]}(s^*) + r^*n(s^*)$ . Define the point on  $\beta_{[s_0^-, s_0^+]}$  at  $s^*$  as  $P = \beta_{[s_0^-, s_0^+]}(s^*)$ . Then we can write the vector from  $P$  to  $O$ ,  $\overrightarrow{PO}$ , as  $r^*n(s^*)$ . If  $t(s^*)$  denotes the tangent to  $\beta_{[s_0^-, s_0^+]}$  at  $P$ , it follows that  $\overrightarrow{PO} \cdot t(s^*) = 0$ . But by our earlier definitions, we have

$$t(s^*) = \left[ \frac{d\psi[z(s^*)]}{ds}, \frac{d\psi[x(s^*)]}{ds} \right], \quad \overrightarrow{PO} = [(\psi[z(s^*)] - a), (\psi[x(s^*)] - b)].$$

Thus, the condition  $\overrightarrow{PO} \cdot t(s^*) = 0$  implies

$$(\psi[z(s^*)] - a) \frac{d\psi[z(s^*)]}{ds} + (\psi[x(s^*)] - b) \frac{d\psi[x(s^*)]}{ds} = 0,$$

which is precisely the condition  $v'(s^*) = 0$ . Therefore  $s^*$  is a critical point of  $v(s)$  for  $s \in [s_0^-, s_0^+]$ .

Now suppose the  $s^*$  is not a minimum. Then there exists  $\epsilon$  such that  $v(s^* + \epsilon) < v(s^*)$ . This would imply that  $\beta_{[s_0^-, s_0^+]}(s^* + \epsilon)$  must lie below  $t(s^*)$  since it is closer to  $O$  than  $P$  is. But this would contradict the fact that  $\beta_{[s_0^-, s_0^+]}$  is convex. Therefore  $s^*$  must be a minimum of  $v(s)$  in  $[s_0^-, s_0^+]$ . By a similar argument  $s^*$  is also seen to be unique.

**Proof of theorem 3.** By theorem 2 we know that  $v(s)$  has a single minimum,  $v(s^*)$ , at  $s^* \in [\mu_0 - \sigma_0, \mu_0 + \sigma_0]$ , so it suffices to prove that  $v(s^*)$  is the unique minimum in  $(\mu_1 + \sigma_1, \mu_0 - \sigma_0) \cup (\mu_0 + \sigma_0, \mu_2 - \sigma_2)$ . Since the argument is the same for either subinterval, we assume without loss of generality that there is another minimum in the left subinterval,  $s^{**} \in (s_0^+, s_2) \equiv (\mu_0 + \sigma_0, \mu_2 - \sigma_2)$ . Let  $P = \beta(s^{**})$ . By theorem 6 applied to the curve  $\beta_{[s_0^+, s_2]}$ , the only way to draw a perpendicular line to  $P$  from  $O$  is if  $O \in \mathcal{P}(\beta_{[s_0^+, s_2]})$ . Since  $O$  is not in  $\mathcal{P}(\beta_{[s_0^+, s_2]})$  by our assumption, we cannot draw a perpendicular line to  $P$ . Therefore  $\overrightarrow{PO} \cdot t(s^{**}) = v'(s^{**}) \neq 0$ , a contradiction. It follows that  $s^*$  is the unique minimum of  $v(s)$  for  $s \in (\mu_1 + \sigma_1, \mu_2 - \sigma_2)$ .

**Proof of theorem 4.**  $O = (a, b)$  falls to the outside of  $\mathcal{P}(\beta_{[\mu_0 - \sigma_0, \mu_0 + \sigma_0]})$  only if either  $O$  is in the region to the left of  $n_0^- = n(\mu_0 - \sigma_0)$  or  $O$  is in the region to the right of  $n_0^+ = n(\mu_0 + \sigma_0)$  (figure 5). We investigate, without loss of generality, the first case.

Let us consider an origin  $O(a, b) = (\beta_1[s_a], \beta_2[s_b])$  with  $s_b < \mu_0$  fixed and whose  $\beta_1$  coordinate  $a$  is determined by the intersection of the horizontal line  $\beta_2 = b$  with  $n_0^-$ . Since  $\mathcal{P}(\beta_{[\mu_0 - \sigma_0, \mu_0 + \sigma_0]})$  (shaded region in figure 4) for  $r > 0$  lies below  $n_0^-$ ,  $O = (a, b)$  as well as any origin whose  $\beta_1$  coordinate is smaller than  $a$  results in an MDF whose minimum is outside the region covered by  $s \in (\mu_0 - \sigma_0, \mu_0 + \sigma_0)$ . In other words, the intersection of the line  $\beta_2 = b$  with  $n_0^-$  determines the boundary beyond which an origin falling outside the region  $\mathcal{P}(\beta_{[\mu_0 - \sigma_0, \mu_0 + \sigma_0]})$  can be found.

Since  $\theta_1 - \theta_0 + \theta_2 = \frac{\pi}{2}$  and  $\theta_0, \theta_1, \theta_2 < \frac{\pi}{2}$  by our assumptions,  $\theta_0 \rightarrow \frac{\pi}{2}$  implicates that  $\theta_1, \theta_2 \rightarrow \frac{\pi}{2}$  and as  $\theta_0 \rightarrow \frac{\pi}{2}$  the L-curve takes the limiting shape shown in figure 4 by the dashed lines. In the limit, as  $\theta_0 \rightarrow \frac{\pi}{2}$ ,  $n_0^-$  becomes horizontal and the line  $\beta_2 = b$  cannot intersect  $n_0^-$  meaning that an  $O(a, b)$  that is outside  $\mathcal{P}(\beta_{[\mu_0 - \sigma_0, \mu_0 + \sigma_0]})$  cannot be found.

Hence, as  $\theta_0$  approaches  $\frac{\pi}{2}$ ,  $\mathcal{P}(\beta_{[\mu_0 - \sigma_0, \mu_0 + \sigma_0]})$  for  $r > 0$  extends in such a way that the  $\beta_1$  coordinate,  $a$ , of any origin  $O(a, b)$ , satisfying the conditions in theorem 4, falls in  $\mathcal{P}(\beta_{[\mu_0 - \sigma_0, \mu_0 + \sigma_0]})$ .

## References

- [1] Banham M R and Katsaggelos A K 1996 Spatially adaptive wavelet-based multiscale image restoration *IEEE Trans. Image Process.* **5** 619–33
- [2] Belge M, Kilmer M E and Miller E L 1998 Simultaneous multiple regularization parameter selection by means of the L-hypersurface with applications to linear inverse problems posed in the wavelet domain *Proc. SPIE'98–Bayesian Inference for Inverse Problems* **3459**
- [3] Belge M and Miller E L Wavelet domain Bayesian image restoration using edge-preserving prior models *Proc. ICIP'98 (Chicago, IL, Oct. 1998)*
- [4] Brooks D H, Ahmad G F, MacLeod R S and Maratos G M 1998 Inverse electrocardiography by simultaneous imposition of multiple constraints *IEEE Trans. Biomed. Eng.*
- [5] Do Carmo M P 1976 *Differential Geometry of Curves and Surfaces* (Englewood Cliffs, NJ: Prentice-Hall)
- [6] Charbonnier P, Aubert G, Blanc-Feraud L and Barlaud M 1997 Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images *IEEE Trans. Image Process.* **6** 298–311
- [7] Flanders H 1989 *Differential Forms with Applications to the Physical Sciences* (New York: Dover)
- [8] Groetsch C W 1984 *The Theory of Tikhonov Regularization for Fredholm Integral Equations of the First Kind* (Boston, MA: Pitman)
- [9] Hanke M and Raus T 1996 A general heuristic for choosing the regularization parameter in ill-posed problems *SIAM J. Sci. Comput.* **17** 956–72
- [10] Hanke M 1996 Limitations of the L-curve method on ill-posed problems *BIT* **36** 287–301
- [11] Hansen P C 1992 Analysis of discrete ill-posed problems by means of the L-curve *SIAM Rev.* **34** 561–80
- [12] Hansen P C 1993 Regularization tools: A MATLAB package for analysis and solution of discrete ill-posed problems *SIAM J. Sci. Comput.* **14** 1487–503
- [13] Hansen P C 1993 The use of the L-curve in the regularization of ill-posed problems *SIAM J. Sci. Comput.* **14** 1487–503
- [14] Hansen P C 1997 *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion* (Philadelphia, PA: SIAM)
- [15] 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
- [16] Morozov V A 1984 *Methods for Solving Incorrectly Posed Problems* (New York: Springer)
- [17] Oprea J 1997 *Differential Geometry and its Applications* (Upper Saddle River, NJ: Prentice-Hall)
- [18] Osher S, Rudin L I and Fatemi E 1992 Nonlinear total variation based noise removal algorithms *Physica D* **60** 259–68
- [19] Pogue B W, McBride T O, Prewitt J, Osterberg U L and Paulsen K D 1999 Spatially variant regularization improves diffuse optical tomography *Appl. Opt.* **38** 2950–61
- [20] Regińska T 1996 A regularization parameter in discrete ill-posed problems *SIAM J. Sci. Comput.* **17** 740–9
- [21] Stephanakis I M 1997 Regularized image restoration in multiresolution spaces *Opt. Eng.* **36** 1738–44
- [22] Tikhonov A N and Arsenin V Y 1977 *Solutions of Ill-posed Problems* (New York: Wiley)
- [23] Vogel C R and Oman M E 1998 Fast, robust total variation-based reconstruction of noisy, blurred images *IEEE Trans. Image Process.* **7** 813–24
- [24] Vogel C R 1996 Non-convergence of the L-curve regularization parameter selection method *Inverse Problems* **12** 535–47
- [25] Wahba G 1977 Practical approximate solutions to linear operator equations when data are noisy *SIAM J. Numer. Anal.* **14** 651–67