A Lagrange method based L-curve for image restoration

G. Landi
Department of Mathematics, University of Bologna, Italy
E-mail: germana.landi@unibo.it

Abstract. The solution of image restoration problems usually requires the use of regularization strategies. The L-curve criterion is a popular heuristic tool for choosing good regularized solutions, when the data noise norm is not a priori known. In this work, we propose replacing the original image restoration problem with a noise-independent equality constrained one and solving it by an iterative Lagrange method. The sequence of the computed iterates defines a discrete L-shaped curve. By numerical results, we show that good regularized solutions correspond with the corner of this curve.

1. Introduction
Image restoration is an inverse problem usually modeled by the linear system

\[ Ax + e = b \]  

where \( b \in \mathbb{R}^{n^2} \) represents the observed image, \( e \in \mathbb{R}^{n^2} \) models additive noise (\( \|b\| > \|e\| \)), and \( x \in \mathbb{R}^{n^2} \) is the original true image. The matrix \( A \) is a large-scale ill-conditioned matrix deriving from the discretization of a Fredholm integral equation. We consider the case of space-invariant blurring operators with periodic boundary conditions, so that \( A \) is block circulant with circulant blocks (BCCB) and matrix-vector products can be performed by using FFTs [1]. We suppose that \( A \) has full rank. Because of the ill-conditioning of \( A \), regularization is necessary in order to reduce the sensitivity of the solution of (1) to the noise in \( b \) [2, 1].

The Conjugate Gradient method applied to the normal equations (CGNR) is a popular regularization technique [3] widely used to solve large-scale problems. The number of performed iterations is the regularization parameter. The L-curve criterion [4, 5] is an heuristic method for the regularization parameter choice which does not require any prior information about the error norm. This criterion is basically based on the plot, in a log-log scale, of the regularized solution norm versus the residual norm, for several regularization parameter values. Very often, this curve is L-shaped and its corner corresponds to a suitable value of the regularization parameter. In [6, 7], the so-called residual L-curve, the graph of the residual norm versus the iteration number, is proposed for computing the CGNR regularization parameter. The recent literature on inverse problems indicates a great interest in the development of heuristic regularization parameter choice rules, as the L-curve-based ones, since they can be used in real-world applications where no information about the error norm is usually available.

The main contribution of this work consists in proposing a new heuristic criterion, referred to as Lagrangian L-curve criterion, for computing regularized solutions of (1). The proposed...
criterium is well-suited for image restoration applications where the coefficient matrix is available only as an operator for matrix-vector products. A wide numerical experimentation has been performed in order to assess the effectiveness of the proposed approach by numerical evidence.

The sequel is organized as follows. The new Lagrangian L-curve criterium is described in section 2 and the numerical results are presented in section 3; conclusions are given in section 4.

2. The Lagrangian L-curve
Consider the following noise-independent minimization problem

\[
\begin{align*}
\text{minimize} & \quad \frac{1}{2} \|x\|^2 \\
\text{subject to} & \quad h(x) = 0
\end{align*}
\]  

(2)

where \( h(x) = \frac{1}{2}\|Ax - b\|^2 - \varepsilon \) and \( \varepsilon \) is a small positive parameter introduced to overcome the non-regularity of the minimizer of \( \|x\|^2 \) subject to \( \|Ax - b\|^2 = 0 \). (Here and in the following, \( \| \cdot \| \) denotes the Euclidean norm.) It should be chosen so that \( \varepsilon \ll \rho^2/2 \), where \( \rho = \|e\| \) is the data noise norm. Even under the hypothesis of no knowledge about the noise norm, this is not a restrictive condition since, in practice, \( \varepsilon \) can be chosen close to the machine precision.

Problem (2) has been introduced in [8] for the regularization of discrete ill-posed problems when the noise norm is supposed to be known. In this work, we suppose the noise norm to be not explicitly known and propose an heuristic regularization strategy.

The first-order optimality conditions for (2) are

\[
\begin{align*}
\nabla_x \mathcal{L}(x, \lambda) &= 0 \\
h(x) &= 0
\end{align*}
\]  

(3)

where \( \mathcal{L}(x, \lambda) = \frac{1}{2}\|x\|^2 + \lambda h(x) \) is the Lagrangian function and \( \lambda \in \mathbb{R} \) is the Lagrange multiplier.

Problem (2) is equivalent to the strictly convex problem of minimizing \( 1/2\|x\|^2 \) over the convex set \( h(x) \leq 0 \), then its solution is unique with positive Lagrange multiplier [9, 10].

Given an initial iterate \((x_0, \lambda_0)\), an iteration of the Lagrange method to solve (2) is stated as

\[
\begin{align*}
x_{k+1} &= x_k + \alpha_k \Delta x_k \\
\lambda_{k+1} &= \lambda_k + \alpha_k \Delta \lambda_k
\end{align*}
\]  

(4)

where \((\Delta x_k, \Delta \lambda_k)\) is the Newton step solving the linear system

\[
\begin{pmatrix}
\nabla^2_{xx} \mathcal{L}(x_k, \lambda_k) & \nabla_x h(x_k) \\
\nabla_x h(x_k)^T & 0
\end{pmatrix}
\begin{pmatrix}
\Delta x \\
\Delta \lambda
\end{pmatrix}
= - \begin{pmatrix}
\nabla_x \mathcal{L}(x_k, \lambda_k) \\
h(x_k)
\end{pmatrix}
\]  

(5)

and the step-length \( \alpha_k \) is chosen in order to guarantee a sufficient decrease of the merit function

\[
m(x, \lambda) = \frac{1}{2} \left( \|\nabla_x \mathcal{L}(x, \lambda)\|^2 + |h(x)|^2 \right).
\]  

(6)

In this work, we make use of the Armijo rule for the selection of the step-length \( \alpha_k \) [11]. Under standard assumptions for the Lagrange method convergence, the sequence \( \{(x_k, \lambda_k)\} \) converges to a point satisfying the first-order necessary conditions for a solution to (2).

Let \( r_k = b - Ax_k \) be the residual at the k-th iterate. The following lemma show that, even if the sequences \( \{\|x_k\|\} \) and \( \{|r_k|\} \) do not have a monotonic behavior, an eventually finite non empty index subset \( K \) exists such that \( \{\|x_k\|\}_K \) and \( \{|r_k|\}_K \) are respectively increasing and decreasing with \( k \in K \).
Lemma 2.1. Let $x^*$ be a minimum point of (2) and let $\lambda^*$ be the corresponding Lagrange multiplier. If the sequence $\{x_k, \lambda_k\}_N$ generated by the iterative procedure (4) converges to $(x^*, \lambda^*)$, then, given $x_0 = 0$ and $\varepsilon \ll \rho^2 / 2$, the solution norm sequence $\{||x_k||\}_N$ and the residual norm sequence $\{||r_k||\}_N$ respectively converge to $||x^*||$ and $||r^*||$ such that

$$
||x^*|| > ||x_0|| = 0,
$$

$$
||r^*|| < ||r_0||.
$$

Moreover, a non empty, eventually finite, index subset $K$ exists such that $\{||x_k||\}_K$ and $\{||r_k||\}_K$ are respectively increasing and decreasing with $k \in K$.

Proof. The initial guess $x_0$ is the solution of the unconstrained problem of minimizing $||x||^2$; however, $x_0$ is not the solution of (2) since $||r_0||^2 = ||b||^2 \geq \rho^2 > 2\varepsilon$. Thus, $x^* \neq x_0$ and $||r_0|| > \sqrt{2\varepsilon} = ||r^*||$, $||x^*|| > ||x_0||$. Let us now consider the sequences $a_k := ||x_0|| - ||x_k||$ and $b_k := ||r_0|| - ||r_k||$. They respectively converge to $a^* = ||x_0|| - ||x^*|| < 0$ and $b^* = ||r_0|| - ||r^*|| > 0$. Then, two positive integers $K_1 \in N$ and $K_2 \in N$ exist such that $a_k < 0$ for all $k \geq K_1$ and $b_k > 0$ for all $k \geq K_2$. By setting $k = \max\{K_1, K_2\}$ we get $a_k < 0$ and $b_k > 0$ for all $k \geq k$. Moreover, we have $||x_0|| < ||x_k||$ and $||r_0|| > ||r_k||$ for all $k \geq k$. This shows that a non empty index subset $K$ always exists and, in the worst case, it reduces to $\{0, k\}$.

Let now $K$ be an index subset such that $\{||x_k||\}_K$ and $\{||r_k||\}_K$ are respectively increasing and decreasing with $k \in K$. Finally, consider the two-dimensional graph of the points

$$
(\log_{10} ||Ax_k - b||, \log_{10} ||x_k||), \quad k \in K.
$$

(9)

We will refer to it as the Lagrangian L-curve since it is based on iterations of the Lagrange method for (2) and since, very often, this graph looks like the letter L. We propose to choose the iterate $x_k$ corresponding to the vertex point of the Lagrangian L-curve as an approximated solution to (1).

Basically, the proposed Lagrangian L-curve criterium is used to single out one of the Lagrange iterates that may be a good regularized solution, even if the limit of the entire Lagrange sequence is not a useful one. A justification for why one of the Lagrange iterate can be used as a regularized solution can be given as follows. Given $x_0 = 0$, the sequence $x_k$ converges to the solution of (2) and a finite index $\ell$ exists such that $||Ax_\ell - b|| < \rho$. Such an iterate $x_\ell$ satisfies the discrepancy principle [12] and, as shown in [8], is a useful regularized solution. An extensive numerical experimentation has shown that $x_\ell$ is close to the Lagrangian L-curve corner.

In this paper, we present numerical evidence that the Lagrangian L-curve criterium can be used to determine good regularized solution to (1).

3. Implementation details

Initial choice $\lambda_0$. The Lagrange method requires an initial guess for $\lambda_0$. As suggested in §14.4 of [11], it can be chosen such that

$$
\lambda_0 = \left(\nabla_x h(x_0)^T \nabla_x h(x_0)\right)^{-1} \left[h(x_0) - \nabla_x h(x_0)^T x_0\right].
$$

(10)

Computation of the search direction $(\Delta x_k, \Delta \lambda_k)$. When $A$ is BCCB, $\nabla^2_{xx} L(x_k, \lambda_k)$ is BCCB again and can be easily inverted by means of FFTs. Hence, an explicit solution of (5) can be computed as (§14.1 of [11]):

$$
\Delta \lambda_k = - \left(D(x_k, \lambda_k)^{-1} \nabla_x h(x_k)^T (\nabla^2_{xx} L(x_k, \lambda_k))^{-1} \nabla_x L(x_k, \lambda_k) - h(x_k)\right)
$$

$$
\Delta x_k = - \left(\nabla_{xx}^2 L(x_k, \lambda_k)^{-1} (\nabla_x h(x_k) \Delta \lambda_k + \nabla_x L(x_k, \lambda_k)\right)
$$

(11)
where $D(x_k, \lambda_k) = \nabla_x h(x_k)^T (\nabla^2_{xx} L(x_k, \lambda_k))^{-1} \nabla_x h(x_k)$ is a scalar.

**Computation of the index subset $K$.** In order to create the index subset $K$, we propose to check the monotonicity of the solution and residual norm sequences and discard those points where the monotonicity condition is not satisfied. In our implementation, the algorithm for the index subset $K$ computation starts setting $0 \in K$. Given $k \in K$, the term in $K$ which follows $k$ is the least index $\phi(k) \in N$ such that $||x_{\phi(k)}|| > ||x_k||$ and $||r_{\phi(k)}|| < ||r_k||$.

**Stopping criteria.** In our implementation, the iterative procedure (4) has been terminated when the relative distance between two successive iterates $(x_k, \lambda_k)$ has became smaller than a given positive tolerance $\tau$ or when a maximum number of allowed iterations has been performed.

4. Numerical results

The numerical experiments have been executed on a Sun Fire V40z server consisting of four 2.4GHz AMD Opteron 850 processors with 16GB RAM using Matlab 7.10 (Release R2010a). For all the tests, a maximum number of 30 iterations and the value $\tau = 10^{-8}$ have been set for terminating the Lagrange iteration. The value of the parameter $\varepsilon$ has been set equal to the machine precision. The proposed Lagrangian L-curve criterium has been compared with the traditional L-curve criterium [4, 5] and the residual L-curve criterium [7] for the CGLS method. In all cases, the L-corner method [13] has been used for the detection of the curves corner. A number of 150 CGLS iterations have been performed in order to determine the corresponding L-curve and residual L-curve.

The test problems are based on 9 famous 256 × 256 test images: the Lena, Cameraman, Mandril, Peppers, Jetplane, Walkbridge, Woman, Boat and Lake images which can be obtained from http://decsai.ugr.es/cvg/dbimagenes/g256.php. These images have been corrupted by Gaussian blur with variance $\sigma = 2$ and 4, and defocus blur with radius $R = 7.5$ and 15, respectively obtained with the code psfGauss and psfDefocus from [1]. Then Gaussian noise has been added with several noise level values. (The noise level NL is defined as $NL := ||e||/||Ax_{\text{exact}}||$ where $x_{\text{exact}}$ is the exact image.) More specifically, the following values for the noise level have been used: $NL=5 \cdot 10^{-4}, 7.5 \cdot 10^{-4}, 1 \cdot 10^{-3}, 2.5 \cdot 10^{-3}, 5 \cdot 10^{-3}, 7.5 \cdot 10^{-3}, 1 \cdot 10^{-2}, 2.5 \cdot 10^{-2}, 5 \cdot 10^{-2}$. For each noise level, 10 Gaussian noise vectors have been obtained with different noise realizations. This amounts to 360 experiments for each image and to 3240 experiments in total.

For all the considered criterium (Lagrangian L-curve, L-curve and residual L-curve), let $k_{\text{corner}}$ and $k_{\text{best}}$ denote the iteration respectively to the L-curve corner and to the smallest relative error. The performance of the three criteria has been evaluated in terms of percentage of experiments producing a reconstructed image such that

$$\|x_{k_{\text{corner}}} - x_{\text{exact}}\| > C\|x_{k_{\text{best}}} - x_{\text{exact}}\|, \quad \text{with} \quad C = 1.05, 1.25, 2.5.$$

For each image, the corresponding percentage values are reported in table 1; for the total 3240 experiments, the percentage values are also shown. The numerical results in the table indicate that, in all the experiments, the proposed criterium outperforms the other ones. Figure 1 depicts the data images, corresponding to Gaussian blur with $\sigma = 2$ and Gaussian noise with $NL=2.5 \cdot 10^{-3}$; figure 2 shows the corresponding reconstructions obtained by the Lagrangian L-curve criterium. For the Lena image degraded by Gaussian blur and Gaussian noise ($\sigma = 2$, $NL=10^{-2}$), figure 3 represents the reconstructions obtained by the Lagrangian L-curve, the L-curve and the residual L-curve criteria. The L-curve image tends to be under-regularized while the Residual L-curve image tends to be over-regularized. The same behavior of the L-curve and the Residual L-curve criteria have been observed in almost all the experiments. Figure 4 shows the relative error behavior for the Lagrange and GCLS iterations together with the graph of the Lagrangian L-curve, L-curve and residual L-curve. Observe that, for the Lagrange iterations, the relative error grows rather slowly after the iteration with the smallest relative error even if...
ε is set to a value equal to the machine precision. This behavior of the relative error has been observed in all the performed numerical tests and is indeed a desirable property for a numerical technique aimed to solve ill-posed inverse problems. In fact, when the relative error grows slowly, then performing too many iterations does not dramatically deteriorate the reconstructed image quality.

5. Conclusions
In this work, a new heuristic criterium, referred to as Lagrangian L-curve criterium, has been presented for the solution of image restoration problems in the absence of prior information about the data noise norm. The effectiveness of the proposed criterium is shown numerically by performing numerous test problems on several images with various blur operators and noise level values. The numerical results are promising and indicate that the proposed criterium can be more effective than other state-of-the-art criteria for the CGLS method.

Table 1: Numerical results.

| Image  | Lagr. L-curve | L-curve | Res. L-curve |
|--------|---------------|---------|--------------|
|        | > 1.05 | > 1.25 | > 2.5 | > 1.05 | > 1.25 | Z > 2.5 |
| Lena   | 25.56 | 8.33  | 0.00  | 60.83 | 42.50 | 6.94  | 88.89 | 61.11 | 0.00 |
| Cameraman | 9.17  | 5.56  | 0.00  | 58.06 | 50.00 | 10.56 | 84.72 | 36.11 | 0.00 |
| Mandril | 8.61  | 5.56  | 0.00  | 67.22 | 45.56 | 5.83  | 63.89 | 17.22 | 5.56 |
| Peppers | 15.83 | 5.56  | 0.00  | 60.00 | 48.89 | 5.56  | 86.94 | 31.39 | 0.83 |
| Jetplane | 17.50 | 9.44  | 2.50  | 81.11 | 58.33 | 22.22 | 83.89 | 38.33 | 6.11 |
| Walkbridge | 11.94 | 10.28 | 0.00  | 52.78 | 33.61 | 5.28  | 80.56 | 22.22 | 0.00 |
| Woman  | 31.11 | 17.78 | 6.11  | 97.22 | 85.83 | 47.22 | 76.94 | 34.44 | 10.56 |
| Boat   | 16.94 | 8.33  | 0.00  | 64.44 | 50.56 | 12.22 | 84.44 | 37.22 | 1.11 |
| Lake   | 16.67 | 10.00 | 0.00  | 57.22 | 45.00 | 7.78  | 86.39 | 44.72 | 0.28 |
| Total  | 17.04 | 8.98  | 0.96  | 66.54 | 51.14 | 13.73 | 81.85 | 35.86 | 2.72 |

Figure 1: Data images (Gaussian blur σ = 2, Gaussian noise NL=2.5·10⁻³). Top to bottom, left to right: Lena, Cameraman, Mandril, Peppers, Jetplane, Walkbridge, Woman, Boat, Lake.
Figure 2: Images restored by the Lagrangian L-curve criterium.

Figure 3: Lena data image ($\sigma = 2$, NL=10$^{-2}$). Top to bottom, left to right: data, Lagr. L-curve, L-curve, Res. L-curve images.

Figure 4: Lena image ($\sigma = 2$, NL=10$^{-2}$). Top to bottom, left to right: relative error (the circles and asterisks indicate the reconstruction and best relative error values), Lagr. L-curve, L-curve, Res. L-curve (residual norm versus solution norm).

[1] Hansen P C, Nagy J and O’Leary D P 2006 Deblurring images. Matrices, Spectra and Filtering (SIAM)
[2] Engl H W, Hanke M and Neubauer A 1996 Regularization of inverse problems (Kluwer Academic Publishers, Dordrecht)
[3] Hanke M 1995 Conjugate Gradient Type Methods for Ill-Posed Problems (Pitman Research Notes in Mathematics, Longman House, Harlow, Essex)
[4] Hansen P and O’Leary D 1993 SIAM J. Sci. Comput. 14 1487–1503
[5] Hansen P C 1992 SIAM Rev. 34 561–580
[6] Reichel L and Sadok H 2008 J. Comput. Appl. Math. 219 493-508
[7] Reichel L and Rodriguez G 2012 Num. Alg. DOI 10.1007/s11075-012-9612-8
[8] Landi G and Piccolomini E L 2010 Comput. Math. App 60 1723–1738
[9] Bertsekas D 2003 Convex Analysis and Optimization (Athena Scientific)
[10] Landi G 2008 Comput. Optim. Appl. 39 347–368
[11] Luenberger D G 2003 Linear and nonlinear programming (Kluwer Academic Publishers, (2nd Ed.))
[12] Morozov V A 1984 Methods for solving incorrectly posed problems (Springer-Verlag, New York)
[13] Hansen P, Jensen T and Rodriguez G 2006 J. Comput. Appl. Math. 198 483492