Application of the NCP Parameter-Choice Method to the General-Form Tikhonov Regularization of 2-D/TM Inverse Scattering Problems

Puyan Mojabi and Joe LoVetri

Department of Electrical and Computer Engineering
University of Manitoba, Winnipeg, MB, Canada, R3T 5V6
pmojabi@ee.umanitoba.ca lovetri@ee.umanitoba.ca

Abstract: A new method of choosing the regularization parameter, originally developed for a general class of discrete ill-posed problems, is investigated for electromagnetic inverse scattering problems that are formulated using a penalty method. This so-called NCP parameter-choice method uses the entire information available in the residual vector, as opposed to just its norm, and attempts to choose the largest regularization parameter that makes the residual resemble white noise. This is done by calculating the NCP for each choice of the regularization parameter, starting from large values and stopping at the first parameter which puts the NCP inside the Kolmogorov-Smirnoff limits. The main advantage of this method, as compared, for example, to the *L*-curve and GCV techniques, is that it is computationally inexpensive and therefore makes it an appropriate technique for large-scale problems arising in inverse imaging. In this paper, we apply this technique to the general-form Tikhonov-regularized functional arising in the 2-D/TM inverse electromagnetic problem, which is formulated via an integral equation and solved using the Born iterative method.

Keywords: Electromagnetic Inverse Scattering, General-Form Tikhonov Regularization, Normalized Cumulative Periodogram, NCP.

1. Introduction

It is well-known that the inverse scattering problem is ill-posed: the solution to the mathematical problem is not unique and does not depend continuously on the measured data. Therefore, we usually attempt to find a solution to the ill-posed operator by adding some constraints and additional information to the system. Three general classes of handling an ill-posed system of equations are the penalty methods, various projection methods, and hybrid combinations of these (see [1, 2]). The Tikhonov method is the best-known penalty method approach to regularizing an ill-posed system of equations. The main idea behind standard-form Tikhonov regularization is that a regularized solution x_λ with a small norm $\|x_\lambda\|$ and sufficiently small residual norm $\|b - Ax_\lambda\|$ can be considered a good approximation to the desired unknown solution to the ill-posed system $Ax = b$. The second approach, *i.e.*, projection methods, try to project the problem onto a subspace with a good basis for the solution. The most famous projection method is the so-called truncated singular value decomposition (TSVD) [3], but usually projection is achieved using iterative methods such as the conjugate gradient method, GMRES, or other Krylov subspace methods [4]. The last class of approaches are the hybrid methods [5] which are based on regularizing the projected problem. This is done because quite often the projection approach, which casts the problem in a smaller subspace, does not regularize the problem sufficiently.

The regularization in each of these methods usually requires the computationally expensive step of choosing an optimum regularization parameter. This is because the resulting solution can be very sensitive to the choice of regularization parameter. In the Tikhonov method, the regularization parameter controls the weight of the penalty term, while in the projection methods, the dimension of the subspace is considered as the regularization parameter, and therefore in the hybrid methods we need two regularization parameters: one for the dimension of the subspace and the other for regularizing the projected problem. Many regularization parameter-choice methods have been proposed in the literature, for example, the discrepancy principle, generalized cross-validation, and the *L*-curve have been widely used. The discrepancy principle [6] uses the idea that the norm of the residual vector cannot be smaller than the norm of the noise in the measured data, but is difficult to apply to electromagnetic inverse problems. Generalized cross-validation (GCV) [7] is a statistical tool for choosing the regularization parameter by minimizing a specialized functional and does not require any knowledge about the noise variance in the data. The other major parameter-choice method is the *L*-curve method which tries to balance the (semi) norm of the solution and the corresponding residual [8] by choosing the regularization parameter that puts one on the corner of the *L*-curve. All of the aforementioned parameter-choice methods are based on the norm of the residual vector. They are also computationally expensive for inverse methods, such as the Born iterative method, where the optimal regularization parameter must be chosen from an unknown wide range of possible values at each iteration.

In this paper, we will use a new parameter-choice method for solving the inverse scattering problem which is based on the entire information available in the residual vector, as opposed to just the norm of the residual. This approach is called the normalized cumulative periodogram (NCP) parameter-choice method and was recently introduced by Hansen *et. al.* [9]. The underlying idea of this method can be explained as follows: we can model the measured data, contained in a vector b as the sum of a signal component $\bar{b}$ and a white noise component e but, due to the smoothing effect of the scattering operator, the power spectrum of the signal component will be dominated by low frequencies whereas the power spectrum of the white noise component will have the same expectation at all frequencies. Therefore, the spectral behavior of the signal component is different from the spectral behavior of the white noise and this difference can be used to find a good regularization parameter for our ill-posed problems [9].

2. Formulation of the Problem

The nonlinear integral equation that encapsulates the 2-D time-harmonic, scalar inverse scattering problem for transverse magnetic (TM) fields is written as

$$E_z^s(\boldsymbol{r};\boldsymbol{k}) = k_0^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G(\boldsymbol{r}, \boldsymbol{r}';k_0)E_z(\boldsymbol{r}';\boldsymbol{k})O(\boldsymbol{r}')dx'dy' \tag{1}$$

where $\boldsymbol{r} = x\hat{a}_x + y\hat{a}_y$ represents the observation point in the Cartesian coordinate system, $\boldsymbol{k} = k_x\hat{a}_x + k_y\hat{a}_y$ represents the wavevector, the wavenumber k_0 is related to the wavevector by $k_0 = |\boldsymbol{k}|$. In (1), for a non-magnetic media, $O(\boldsymbol{r}) = \varepsilon_r(\boldsymbol{r}) - 1$ is the contrast profile which must be recovered and it will be assumed that the object profile is lossless in the remainder of this paper. The two-dimensional free-space Green's function, assuming an $e^{j\omega t}$ time dependence, is given as

$$G(\boldsymbol{r}, \boldsymbol{r}';k_0) = \frac{1}{4j}H_0^{(2)}(k_0|\boldsymbol{r} - \boldsymbol{r}'|) \tag{2}$$

where $H_0^{(2)}(x)$ is the zeroeth-order Hankel function of the second kind.

For the results given in this paper, we assume that data collection is done by a set of receivers which are located on a circle around the object and that the object is illuminated by TM plane-waves impinging on the object from different incidence angles. The geometrical configuration is the same as that described in [10]. For obtaining a solution for the contrast in (1), we use the Born iterative method (BIM) (described in [10]). This method proceeds by first using the Born approximation [11] to linearize the problem which is then solved for the unknown contrast using an inverse solver (below we describe our inverse solver that is based on the general-form Tikhonov regularization). The total-field inside the imaging domain, corresponding to this contrast, is then computed using a moment-method forward solver (we use Richmond's method [15]). The newly updated total-field is then used in the subsequent iteration for linearizing the integral equation and the inverse solver is again used for obtaining a new approximation to the contrast. This procedure continues until a termination condition is satisfied: when the change in a norm of the solution is less than a specified value.

3. The General-Form Tikhonov Regularization Inverse Solver

After discretizing the linearized integral equation, we obtain a system of linear equations $Ax = b$, where $A \in \mathrm{C}^{m \times n}$, $b \in \mathrm{C}^{m}$ with $m \geq n$ and x is to be found. The matrix A is a discrete representation of the linearized kernel, while x and b are column-wise stacked representations of the 2-D discrete contrast function, $O(x, y)$, and measured scattered field, $E_z^s(x, y)$, respectively. The pseudo-inverse of A is unbounded due to ill-posedness of the inverse problem. Therefore, for solving this matrix equation, we use the general-form Tikhonov regularization method, which effectively produces a regularized pseudo-inverse operator that is bounded, in conjunction with a parameter-choice method based on NCP that keeps the regularized solution as close as possible to the exact solution. The general-form Tikhonov regularization method can be represented concisely as producing a solution x_λ given as:

$$x_\lambda = A_\lambda^\dagger b = \min_x \{ \|Ax - b\|_2^2 + \lambda^2 \|L(x - x_0)\|_2^2 \}, \tag{3}$$

where λ is the regularization parameter, and $L \in \mathrm{C}^{k \times n}$ is called the regularization matrix which can be any matrix whose nullspace does not intersect with the nullspace of A [12]. The vector x_0 is generally taken as a guess of the solution, and in our case we take it to be the most recent value of the contrast (at the previous iteration). We take L to be the Laplacian operator with zero boundary conditions for the unknown contrast profile. In this case, the nullspace of L is trivial and will not intersect with the numerical nullspace of the ill-posed operator, making the solution to (3) unique.

4. The NCP Parameter-Choice Method

Consider the measured data, *i.e.,* the scattered electric field, as a matrix $B \in \mathrm{C}^{p \times q}$ where p denotes the number of different angles at which the TM plane wave illuminates the object and q denotes the angle at which the q^{th} receiver is located (on a circle around the object). As mentioned previously, the measured values in the matrix B consist of signal and noise components and therefore this matrix can be represented as a signal component matrix $\bar{B}$ and a white noise component matrix E: $B = \bar{B} + E$.

For simplicity of discussion assume that in (3), $L = I$ where I is the identity matrix and $x_0 = 0$, then the residual vector of Tikhonov solution can be written as:

$$r_\lambda = b - Ax_\lambda = U \Lambda U^H \bar{b} + U \Lambda U^H e, \; \Lambda = \mathrm{diag}\{\lambda^2 / (\lambda^2 + \sigma_i^2)\} \tag{4}$$

where U is the matrix of left singular vectors, u_i, of the matrix A, with each u_i corresponding to σ_i, a singular value. The vectors $\bar{b}$ and e are obtained by stacking the columns of $\bar{B}$ and E into a vector of length

$p \times q$. For the case where $L \neq I$, the singular values will be substituted by generalized singular values of the pair (A, L) and U will be the orthonormal matrix in the decomposition of A using the Generalized Singular Value Decomposition of (A, L) [13].

The diagonal components $\lambda^2/(\lambda^2 + \sigma_i^2)$ look like a high-pass filter when plotted against the index i, because the singular values decrease rapidly for ill-posed problems. The regularization parameter λ determines the "cut-off" index *k* of this high-pass characteristic: the smaller the value of λ, the larger the cut-off index. This means that as we decrease the regularization parameter λ, the first term in the residual, $U\Lambda U^H \bar{b}$ will have little contribution from these initial vectors since it can be written as:

$$\sum_{i=1}^{n} u_i \left[\frac{\lambda^2}{\lambda^2 + \sigma_i^2} u_i^H \bar{b} \right] = \sum_{i=1}^{k} u_i \left[\frac{\lambda^2}{\lambda^2 + \sigma_i^2} u_i^H \bar{b} \right] + \sum_{i=k+1}^{n} u_i \left[\frac{\lambda^2}{\lambda^2 + \sigma_i^2} u_i^H \bar{b} \right] \approx \sum_{i=k+1}^{n} u_i \left[\frac{\lambda^2}{\lambda^2 + \sigma_i^2} u_i^H \bar{b} \right]. \quad (5)$$

It has been argued by Hansen *et al.* [9] that $\bar{b}$ has few significant (*i.e.* non-zero) components in the SVD basis corresponding to the first few left singular vectors of A and the remaining components are almost zero. That's why the smaller the λ (*i.e.*, the larger the cut-off index), the less contribution from the first term in the residual vector (in (4)). Thus, using a cut-off index that suppresses all of the significant components of $\bar{b}$ in the residual means that we've used as much information as possible in the solution, and choosing the smallest such index ensures a stable solution (giving an acceptable trade-off between the regularization and perturbation errors). The regularization parameter corresponding to this cut-off index can be considered as the optimum regularization parameter because it singles out the most stable solution whose residual does not have any important component of $\bar{b}$. The residual vector for this optimum regularization parameter will be dominated by $U\Lambda U^H e$, but $U\Lambda U^H e$ behaves statistically like white noise for ill-posed problems because:

$$\mathrm{cov}\{U\Lambda U^H e\} = U\Lambda U^H [\mathrm{cov}\{e\}] U\Lambda U^H = U\Lambda U^H [\eta^2 I](U\Lambda U^H) = \eta^2 (U\Lambda^2 U^H) \approx \eta^2 I_k \quad (6)$$

where η is the standard deviation of the additive white noise and I_k is the identity matrix with the first *k* diagonal elements set to zero. For ill-posed problems *k* is very small and this covariance will be very similar to the covariance of white noise. Therefore, the optimum regularization parameter can be considered as the largest λ which makes the residual vector behaves like white noise.

The metric that is used to see if the residual "looks" like white noise is the NCP of the residual [14]. So the regularization procedure that we follow is to start with a large λ, which is usually less than one, in which case the NCP of the residual vector will look like that of the data—meaning that we have a lot of the data information left in the residual. We then decrease λ until the NCP of the residual first becomes like that of white noise (*i.e.*, a straight line between the Kolmogorov-Smirnov limits). Once this happens, we can be sure that all the important information available in $\bar{b}$ has been used in calculating x_λ. Notice that if we decrease λ further, the residual is still white noise (or slightly highpass filtered white noise) but the solution is more likely to be unstable due to perturbation error.

One note regarding our use of the NCP parameter-choice method is that NCP is usually defined for real vectors—because it is generally used as a statistical time-series analysis tool [14]—but here we use the same definition for the NCP of a complex vector. For finding the NCP of the vector r_λ, we first find the power spectrum of this vector as

$$P = [\mathrm{DFT}\{r_\lambda\}] \odot [\mathrm{DFT}\{r_\lambda\}]^* \quad (7)$$

where $\odot$ denotes the Hadamard product (*i.e.*, element-wise multiplication) and * denotes complex conjugate. The components of the NCP vector, $C \in \mathbb{R}^{n-1}$, can be calculated as

$$C_i = [\|P\|_1 - P_1]^{-1} \sum_{j=2}^{i+1} P_i,\ i = 1 \ldots n-1 . \tag{8}$$

In our case, the Kolmogorov-Smirnoff (KS) limit lines as a function of index i are given as $i/(n-1) \pm \delta$ where, for a significance level of 5 percent, we set $\delta = 1.36/\sqrt{n}$.

As was mentioned above, $\bar{b}$, and consequently b, have only a few components that are significant in the SVD basis of the ill-posed operator. Due to the similarity of the SVD basis and the Fourier basis [9], b will also have only a few significant components in the Fourier basis. At the index location where these significant components occur there will also occur step changes in the NCP of b. This means that the NCP of b will look like a staircase plot where the step-locations correspond to the location of the significant components of b in its Fourier basis. In [9], a permutation matrix, Π, has been introduced to reorder the elements of the power spectrum of b such that all the significant information inside the NCP of b, *i.e.* the steps in the original NCP plot, are moved to the first elements of the NCP vector. We've observed that using this permutation matrix has no effect on choosing the optimum regularization parameter but we include results using this permutation matrix in addition to the NCP vector obtained directly, because this helps us visualize the effect of the filtering term in (4), *i.e.* $\Lambda = \text{diag}\{\lambda^2/(\lambda^2 + \sigma_i^2)\}$, more easily.

5. Numerical Results

We present the results for two cases: a sinusoidal contrast with an amplitude of 2.0 and also two spatially separated sinusoidal contrasts of amplitudes 2.0 and 1.0. Figure 1 shows the resulting reconstruction using both the identity and the Laplacian operator as the regularization matrices. The synthetic data was produced by an MoM solver with triangular meshes and then white noise was added such that the signal to noise ratio was $\|\bar{b}\|_2/\|e\|_2 = 10$. In the different iterations of the BIM, the forward solution was obtained by Richmond's method [15] using a pulse basis over the imaging domain. As can be seen, the general trend of the NCP method is to over-regularize the problem, and sometimes the solution converges before reaching the maximum value of the unknown profile. This is expected because there are many regularization parameters producing a residual NCP that lies within the Kolmogorov-Smirnoff limits but this method choose the largest one to make sure that the solution is stable. However, we've observed that using the Laplacian operator as the regularization matrix helps to not only produce ripple-free solutions but also to better reconstruct the peak value of the profile as compared to the identity operator. In Fig. 2, the NCP of b as well as the NCP of some residual vectors corresponding to different regularization parameters are shown for the Born approximation of the first test case, both with and without using the permutation matrix. As seen in Fig. 2, for large values of λ the NCP of the residual looks like the NCP of b, showing that we have not used all of the available information in reconstructing the profile. As we decrease the λ, we include less information in the residual and more for the solution. The first NCP which fits the Kolmogorov-Smirnoff limits is the NCP corresponding to $\lambda = 0.02$. As a comparison, we've solved these two problems using the L-curve parameter-choice method and it seems that L-curve chooses a smaller regularization parameter in comparison with the NCP method. For example, in the Born approximation of the first test case, the L-curve chooses $\lambda = 0.013$ as the optimum regularization parameter (as compared to the NCP's $\lambda = 0.02$). This fact can be seen in Fig. 3: in this figure, we've plotted the L-curve for the Born approximation of the first test case using 100 different λ's. λ_{NCP} is the regularization parameter chosen by the NCP method but the L-curve chooses λ_c which is smaller than λ_{NCP}. This simply reflects the fact that there is no unique solution to the inverse problem.

6. Conclusions

We've shown that the NCP parameter-choice method can be very useful in large-scale inverse problems because it only needs to check the NCP of the residuals for a relatively small number of regularization parameters. The L-curve method requires that one calculate the norm of the residual and the regularized solution for a large number of regularization parameters. When using the GCV method, we need to find both the residual norm and the trace of $(I - A(A^H A + \lambda^2 I)^{-1} A^H)$ for a relatively large number of regularization parameters. These calculations can be done fast when the SVD of the ill-posed operator is available but the SVD is difficult to compute for large-scale problems. In addition, the NCP method is easier to implement than well-known methods like L-curve.

As has been mentioned, the drawback of the NCP method is that it tends to over-regularize the solution. The NCP method, as presented, is only applicable when the residual tends to behave like white noise as we decrease the regularization parameter. This requirement can be checked by finding the Tikhonov solution for a very small regularization parameter, say $\lambda = 10^{-6}$, and checking the NCP of the corresponding residual. If it satisfies the NCP criteria, then this method should work well. On the other hand, if it does not satisfy the NCP criteria for such a small λ, then the method must be somewhat modified. We will present details of the modified version of the NCP method in future work.

7. Acknowledgements

The authors would like to acknowledge the financial support of the Natural Sciences and Engineering Research Council. We would also like to thank Prof. P. C. Hansen for making some of his recent work on the NCP method available to us and Messrs. Colin Gilmore and Ian Jeffrey for providing the MoM code that was used for obtaining the synthetic data.

8. References

[1] A. N. Tikhonov and V. Y. Arsenin, *Solution of Ill-Posed Problems*, Scripta Series in Mathematics, John Wiley & Sons, New York, 1977.

[2] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems*, SIAM, Philadelphia, 1998.

[3] P. C. Hansen, "Truncated Singular Value Decomposition Solutions to Discrete Ill-Posed Problems With Ill-determined Numerical Rank," *SIAM, J. Sci. Stat. Comput.*, Vol. 11, No. 3, pp. 503-518, 1990.

[4] M. E. Kilmer and Dianne P. O'Leary "Choosing Regularization Parameters In Iterative Methods For Ill-Posed Problems," *SIAM J. Matrix. Anal. Appl.*, Vol. 22, No. 4, pp. 1204-1221, 2001.

[5] D. P. O'Leary and J. A. Simmons, "A Bidiagonalization-Regularization Procedure For Large Scale Discretization of Ill-Posed Problems," *SIAM J. Sci. Statist. Comput.*, 2, pp. 474-489, 1981.

[6] V. A. Morozov, *Methods For Solving Incorrectly Posed Problems*, Springer-Verlag, 1984.

[7] G. Golub, M. Heath and G. Wahba, "Generalized Cross-Validation As A Method For Choosing A Good Ridge Parameter," *Technometrics*, 21, pp. 215-223, 1979.

[8] P. C. Hansen, "Analysis of Discrete Ill-Posed Problems By Means of The L-Curve," *SIAM, Review*, 34, pp. 561-580, 1992.

[9] P. C. Hansen, M. E. Kilmer and R. H. Kjeldsen, "Exploiting Residual Information In The Parameter Choice For Discrete Ill-Posed Problems," *BIT*, *Numerical Mathematics*, 2006.

[10] Y. M. Wang, and W. C. Chew, "An Iterative Solution of Two-Dimensional Electromagnetic Inverse Scattering Problem," *Int. J. Imaging Syst. Technol.*, Vol. 1, No. 1, pp. 100-108, 1989.

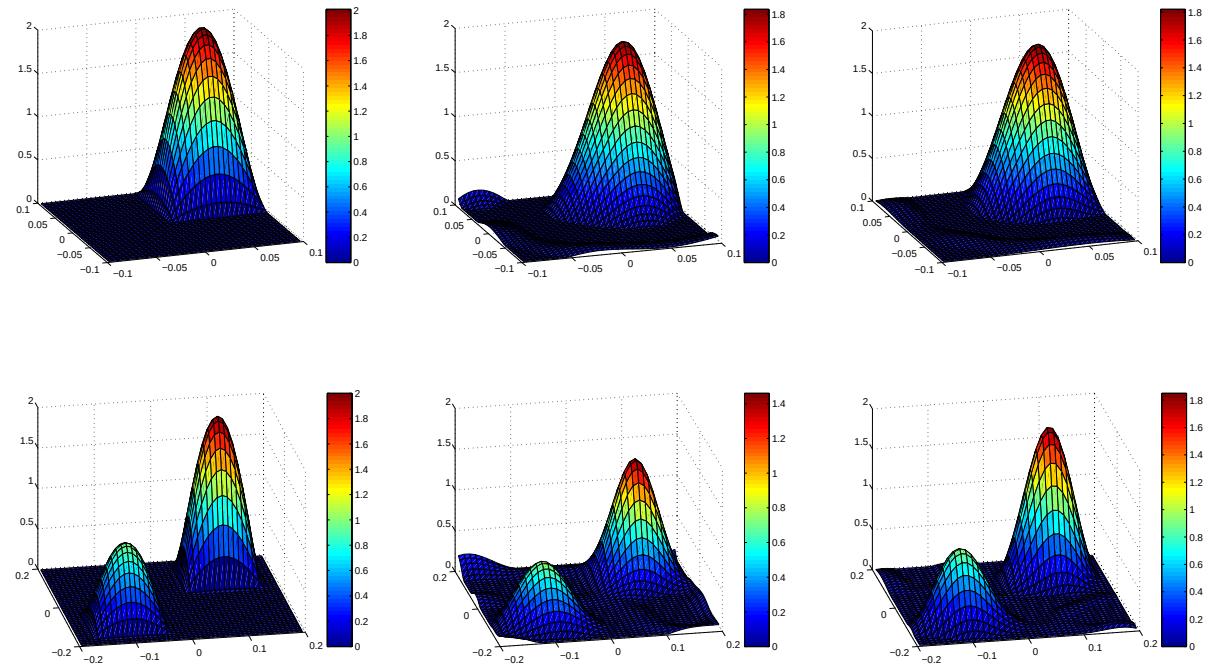
[11] M. Born, and E. Wolf, *Principles of Optics,* 7th ed., Cambridge University Press, 1999.

[12] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Kluwer Academic Publishers, Dordrecht, 2000.

[13] P. C. Hansen, "Regularization, GSVD and Truncated GSVD," *BIT* 29, pp. 491-594, 1989.

[14] W. A. Fuller, *Introduction to Statistical Time Series*, 2nd ed., Wiley, New York, 1976.

[15] J. H. Richmond, "Scattering by a Dielectric Cylinder of Arbitrary Cross Section Shape," *IEEE Trans. Antennas. Propag.*, AP-13, pp. 334-341, 1965.



(a) (b) (c)

Figure 1. Top: first test case (a) true profile, sinusoidal profile with peak-permittivity is 3.0 (contrast is 2.0) (b) reconstruction after 3 iterations with $L = I$, (c) after 3 iterations with L the Laplacian. Bottom: second test case (a) true profile, two sinusoidal profiles with peak-contrast equal 2.0 and 1.0 (b) reconstruction after 5 iterations with $L = I$, (c) after 5 iterations with L the Laplacian.

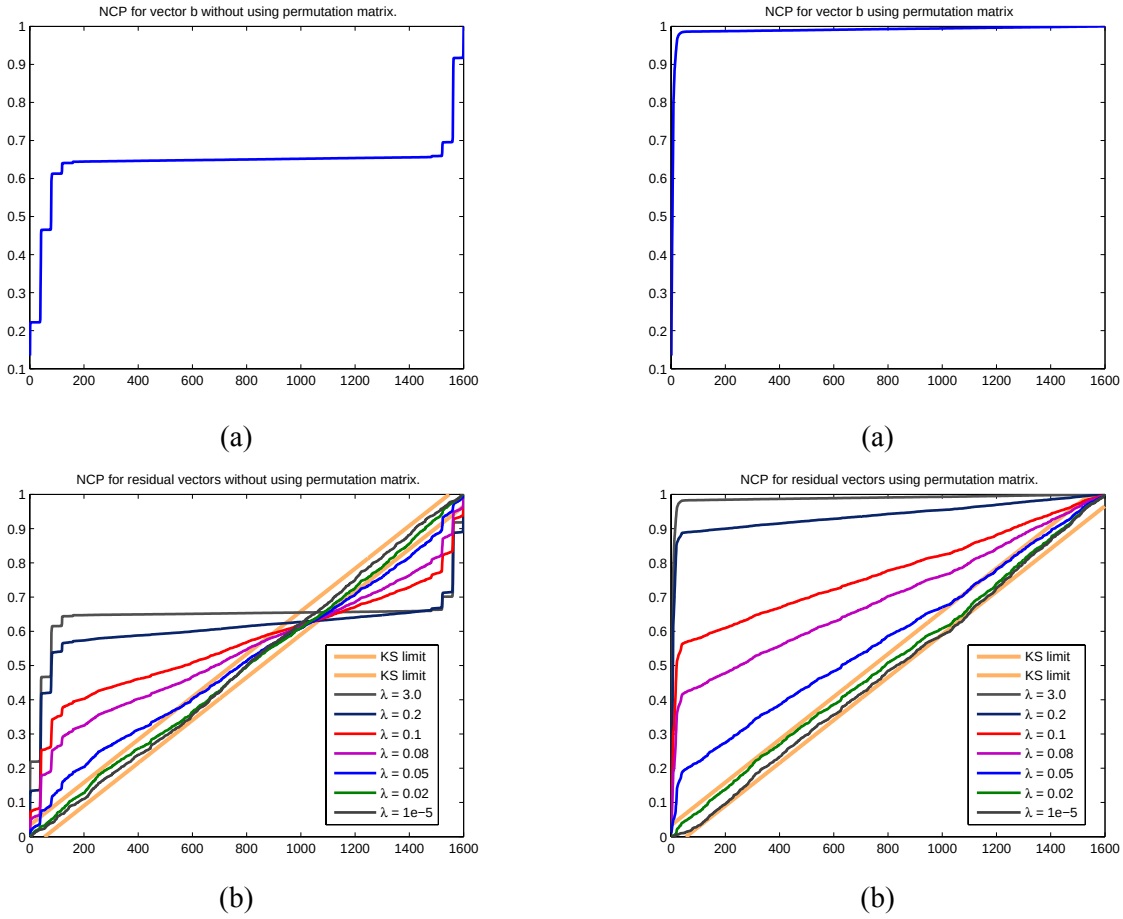


Figure 2. Left: Without using permutation matrix, Right: Using permutation matrix
(a) NCP of vector *b* (b) NCP of residual vector corresponding to seven different regularization parameters.

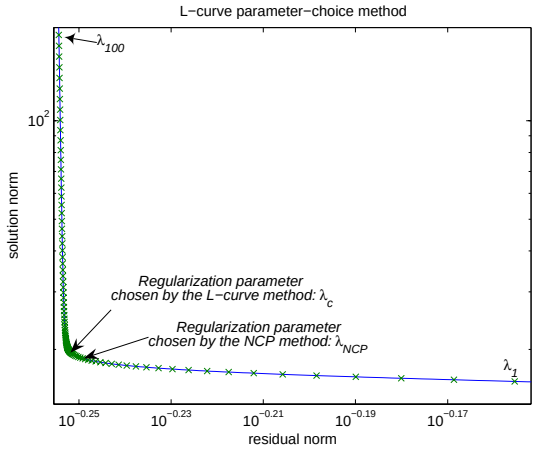


Figure 3. Comparison between the regularization parameters chosen by the L-curve and the NCP methods.