
Rosemary A Renaut
**Efficiently estimating the Regularization Parameter for
Least Squares problems: a Newton algorithm using the χ^2
distribution of the cost functional**

Department of Mathematics and Statistics
871804 Arizona State University
Tempe
AZ 85287-1804
USA
renaut@asu.edu
Jodi Mead

We consider the ill-posed system of equations $A\mathbf{b} = \mathbf{b}$, $A \in \mathcal{R}^{m \times n}$, $\mathbf{b} \in \mathcal{R}^m$, $\mathbf{b} \in \mathcal{R}^n$ for which an approximate solution is to be obtained by solving the regularized weighted least squares problem

$$\hat{\mathbf{x}} = \operatorname{argmin} J(\mathbf{x}) = \operatorname{argmin}\{\|A\mathbf{x} - \mathbf{b}\|_{W_{\mathbf{b}}}^2 + \|\mathbf{x} - \mathbf{x}_0\|_{W_{\mathbf{x}}}^2\}. \quad (1)$$

A new approach for determining the weighting matrix $W_{\mathbf{x}}$ was discussed in [2] and is based on the following observation, presented originally in [4]:

Theorem Let $J(\mathbf{x}) = (\mathbf{b} - A\mathbf{x})^T W_{\mathbf{b}} (\mathbf{b} - A\mathbf{x}) + (\mathbf{x} - \mathbf{x}_0)^T W_{\mathbf{x}} (\mathbf{x} - \mathbf{x}_0)$, where the elements in \mathbf{x} and the elements in \mathbf{b} are each independent and identically distributed. If $W_{\mathbf{b}}$ and $W_{\mathbf{x}}$ are symmetric positive definite (SPD) weighting matrices on the data and model errors, resp., then for large m , the minimum value of J is a random variable which follows a χ^2 distribution with m degrees of freedom.

Specifically, the result suggests that in the limit, for statistical measurements, the minimum value of J should be a χ^2 random variable with approximately m degrees of freedom. In [3] we extended this result for the more general Tikhonov regularization term $\|D(\mathbf{x} - \mathbf{x}_0)\|_2^2$, obtaining that the functional has fewer degrees of freedom, $\tilde{m} = m - n + p$, where $D \in \mathcal{R}^{p \times n}$. Mead used the first result to design an algorithm in which one finds the regularization weighting matrix such that the obtained minimum of J has m degrees of freedom within some given confidence interval. But in the special case where the inverse covariance matrix $W_{\mathbf{x}}$ on the model parameters \mathbf{x} is limited to be a diagonal scaling of the identity $\sigma_{\mathbf{x}}^{-2}I$, the problem simplifies to the determination of the scalar regularization parameter $\lambda = 1/\sqrt{\sigma_{\mathbf{x}}}$. For this case λ can be found using a scalar Newton method for solving $F(\lambda) = J - \tilde{m} = 0$. Moreover, using the generalized singular value decomposition, Mead and Renaut [3] showed that F is monotonic in λ , and thus when a solution exists, it is unique. Their

results validated the viability of the use of the Newton method to obtain good estimates for λ , completely consistent with values obtained by generalized cross-validation, and unbiased predictive risk algorithms, see for example [5]. Indeed the method is far cheaper than these methods, only requiring a few steps of the Newton iteration, as compared to these other methods and to the L-curve, all of which usually involve evaluating the relevant functionals for many more values of λ . All of these methods, however, are limited in applicability for large scale problems, if they are based on the use of the generalized singular value decomposition (GSVD) (of singular value decomposition when appropriate) for obtaining updates of the solutions for varying λ . Additionally, the χ^2 approach appears to not be relevant when there is no information on covariance structure on the data \mathbf{b} .

In this work we address these issues. The difficulty with lack of statistical information on \mathbf{b} , is easily addressed by the realization that the lack of the covariance matrix prevents the whitening of the noise in the data, so that the number of degrees of freedom is reduced. Fortunately, the implementation of the algorithm explicitly provides the number of degrees of freedom so that \tilde{m} can be reestimated. For the development of efficient implementations of the χ^2 estimate for the regularization parameter λ , so as to extend the approach for large scale problems, we consider two directions which avoid the GSVD. For image deblurring applications, operations with the matrix A are equivalent to convolutions which are more efficiently handled in the Fourier domain than in space. For a given λ an *exact* minimization of J is found using the benefits of fast Fourier transforms, see for example [5]. The Newton algorithm also requires then the evaluation of the derivative $(d/d\lambda)(J)$ which for exact solutions $\mathbf{x}(\lambda)$ can be seen to easily simplify, and the GSVD is avoided. For more general problems, where the FFT approach is not relevant, it is appropriate to consider the use of inexact solves for minimizing $J(\lambda)$, and to replace the Newton method by the secant algorithm for rootfinding. Theoretical considerations and numerical validations describing the new algorithms will be discussed. We will use not only examples from the test set of Hansen [1], but also restoration of real seismic deep earthquake signals.

Bibliography

- [1] Hansen, P. C., 1994, Regularization Tools: A Matlab Package for Analysis and Solution of Discrete Ill-posed Problems, *Numerical Algorithms* **6** 1-35.
- [2] Mead J., 2007, A priori weighting for parameter estimation, *J. Inv. Ill-posed Problems*, to appear.
- [3] J Mead, R. A. Renaut, The χ^2 -curve method of parameter estimation for generalized Tikhonov regularization submitted, SIAM SISC, August 2007.
- [4] Rao, C. R., 1973, *Linear Statistical Inference and its applications*, Wiley, New York.
- [5] Vogel, C. R., 2002. *Computational Methods for Inverse Problems*, (SIAM), Frontiers in Applied Mathematics.