[40] G. Whaba, *Practical Approximate Solutions to Linear Operator Equations When The Data are Noisy*, SIAM J. Numer Anal., 14, (1977), pp. 651–667.

[41] G. Whaba, 1990. *Spline Models for Observational Data*, SIAM Philadelphia 1990.

[42] G. Whaba, S. O'Sullivan, 1985. *A cross validated Bayesian retrieval algorithm for nonlinear remote sensing experiments*, J. Comput. Physics, 59, 441-455.

[43] S. H. Ward, G. W. Hohmann, *Electromagnetic theory for geophysical Applications*, in Electromagnetic Methods in Applied Geophysics., 1 131-311. (1988) Soc. Expl. Geophys.

[44] K. P. Whittall, D. W. Oldenburg, *Inversion of Magnetotelluric Data for a One Dimensional Conductivity*, SEG. Monograph, 5 1992.

[45] J. Zhang, R. L. Mackie, T. R. Madden, *3-D Resistivity Forward Modeling and Inversion Using Conjugate Gradients*, Geophysics, 60, 5, (1995), pp. 1313–1325.

[15] G. H. Golub, M. Heath, and G. Whaba, *Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter*, Technometrics, 21, (1979) pp. 215–223.

[16] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Second ed., The Johns Hopkins University Press, Baltimore, MD, 1989.

[17] G. H. Golub, U. Von Matt, *Quadratically Constrained Least Squares and Quadratic Problems*, Numer. Math., 59, (1991) pp. 561–580.

[18] G. H. Golub, U. Von Matt, *Generalized Cross-Validation for Large Scale Problems*, (1997), to appear

[19] E. Haber, *Numerical Strategies for the Solution of Inverse Problems*, Ph.D Thesis, The University of British Columbia, 1997.

[20] M. Hanke, *Regularization with Differential Operators: An Iterative Approach*, Num. Func. Anal. Optim., 13, (1992) pp. 523–540.

[21] M. Hanke, P. C. Hansen, *Regularization methods for large scale problems*, Survey on Mathematics for Industry, 3, (1993) pp. 253–315.

[22] P. C. Hansen *Rank Deficient and Ill-Posed Problems* Ph.D Thesis, Technical University of Denmark, 1995.

[23] C. T. Kelley, *Iterative Methods for Linear and Nonlinear Equations*, SIAM Philadelphia. 1995.

[24] O. Knoth, *A Globalization Scheme for the Generalized Gauss-Newton Method.* Num. Math., 56, 2, (1996) pp. 591–603.

[25] D. G. Luenberger, *Optimization by Vector Space Methods*, New-York: John Wiley and Sons 1969.

[26] W. Menke, *Geophysical Data Analysis: Discrete Inverse Theory*, Academic Press, Inc., Orlando, 1984.

[27] D. W. Oldenburg, *One-Dimensional Inversion of Natural Source Magnetotelluric Measurements*, Geophysics, 44, (1979) pp. 1218–1244.

[28] , D. W. Oldenburg, P. R. McGillvary, R. G Ellis, *Generalized Subspace Methods for Large Scale Inverse Problems*, Geophys. J. Int., 114, (1993) pp. 12–20.

[29] D. W. Oldenburg, Y. Li, *Subspace Linear Inverse Method*, Inverse Problems. 10, (1994) pp. 915–935.

[30] D. P. O'Leary, J. A. Simmons, *A Bidiagonalization-Regularization Procedure for Large Scale Regularization of Ill-Posed Problems*, SIAM J. Sci. Stat. Comut., 2, (1981) pp. 474–489.

[31] R. L. Parker, *Geophysical Inverse Theory*, Princeton University Press, Princeton, New Jersey 1994.

[32] B. N. Parlet, *The Symmetric Eigenvalue Problem*, Prentice-Hall 1980.

[33] Y. Saad, *Iterative Methods for Sparse Linear Systems*, PWS Publishing Company 1996.

[34] J. A. Scales, M. L. Smith, *Introductory Geophysical Inverse Theory*, Samizdat Press 1994.

[35] J. T. Smith, J. R. Booker, *Rapid Inversion of Two and Three Dimensional Magnetotelluric Data*, J. Geophys. Res., 96, (1991) pp. 3905–3922.

[36] A. Tarantola, B. Valette, *Generalized Nonlinear Inverse Problems Solved Using the Least Squares Criterion*, Rev. Geophys. Space Phys., 20, 2, (1982) pp. 219–232.

[37] A. Tarantola, *Inverse Problem Theory*, Elsevier Scientific Publishing Company, Amsterdam 1987.

[38] A. N. Tikhonov, V. Y. Arsenin, *Solutions of Ill-Posed Problems*, John Wiley and Sons, Inc. 1977.

[39] J. M. Varah, *Pitfalls in Numerical Solutions of Linear Ill-Posed Problems*, SIAM J. Sci. Comp., 4 (1983) pp. 164–176.

9. Summary and Conclusions

In this paper we developed two algorithms for solving nonlinear ill-posed problems. Both algorithms are based on understanding the two different processes which we face when solving a nonlinear ill-posed problem. The first process is noise estimation. This process is global in nature, and therefore one has to look for a global technique to estimate the regularization parameter. Here we suggested using two variants of the GCV to estimate the global noise. The second process which has to be addressed is the estimation of the step length. If the step length is too large then the linearization does not hold and the iteration might not reduce the value of the objective function we are trying to minimize. This process is local in its nature and therefore needs a different type of regularization. We suggest using the damped Gauss-Newton for this process. In this way we ensure that the step which we choose is not only regularized against the non-uniqueness of the problem, but we also make sure we descend in each nonlinear step.

References

[1] J. Altman, *Smoothing Data with Correlated Errors,* Tech. Report 280. Dept. of Stat., Stanford, (1987)

[2] L. Armijo, *Minimization of a Function Having Lipschitz-Continuous First Partial Derivatives*, Pacific J. Math., 16, (1966) pp. 1–3.

[3] A. Bakushinsky, and A. Goncharsky, *Ill-Posed Problems: Theory and Applications.* Kluwer Academic Publishing (1994).

[4] P. N. Brown and Y. Saad, *Hybrid Krylov methods for nonlinear systems of equations*, SIAM J. Sci. Statist. Comput., 11 (1990), pp. 450–481.

[5] S. C. Constable, R. L. Parker, and C. G. Constable, *Occam's Inversion: A Practical Algorithm for Generating Smooth Models from Electromagnetic Sounding Data*, Geophysics, 52, (1987), pp. 289–300.

[6] R. S. Dembo, S. C. Eisenstat, and T. Steihaug, *Inexact Newton methods*, SIAM J. Numer. Anal., 19 (1982), pp. 400–408.

[7] R. S. Dembo, T. Steihaug, *Inexact Newton Algorithms for Large Scale Optimization*, Math. Programming, 26, (1983) pp. 190–212.

[8] J. E. Dennis, R. B., Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* SIAM Philadelphia 1996.

[9] L. Elden, *Algorithms for the regularization of ill-conditioned least squares problems*, BIT, 17, (1977), pp. 134–145.

[10] L. Elden, *A Weighted Pseudo-inverse, Generalized Singular Values, and Constrained Least Squares Problems*, BIT, 22, (1982) pp. 487–502.

[11] L. Elden, *Algorithms for the Computation of Functions defined on the Solution of a Discrete ill-posed problem*, BIT, 30, (1990) pp. 466–483.

[12] H. W. Engl, M. Hanke, A. Neubaur, *Regularization of Inverse Problems* Kluwer Academic Publishers. 1996.

[13] C. G. Farquharson, D. W. Oldenburg, *Inversion of time domain electromagnetic data for a horizontally layered Earth* Geop. J. Int., (1993), 114, pp. 433-442.

[14] P. E. Gill, W. Murray, M. H. Wright, *Practical Optimization* New York Academic Press. 1981

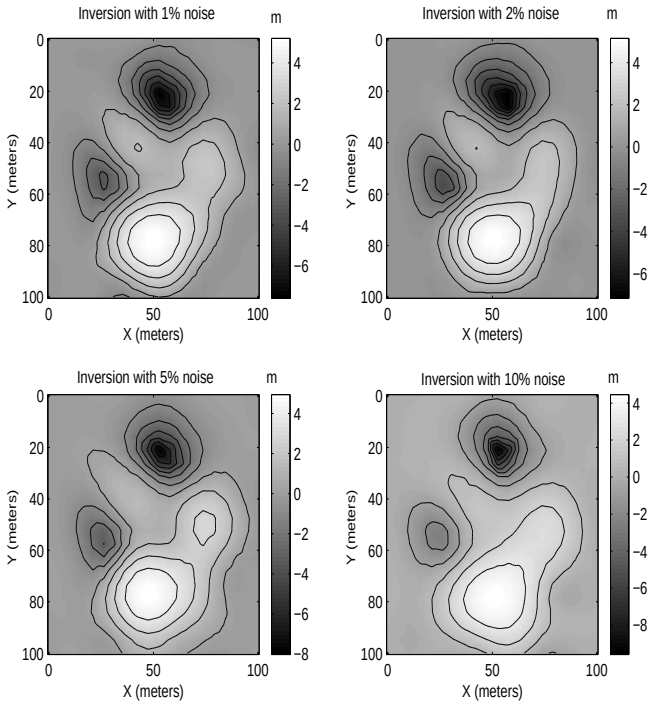

Figure 6. Results of nonlinear gravity inversion for noise levels of $1, 2, 5$ and 10%. The recovered topography can be compared with the true topograhpy of the interface shown in top panel in Figure 4.

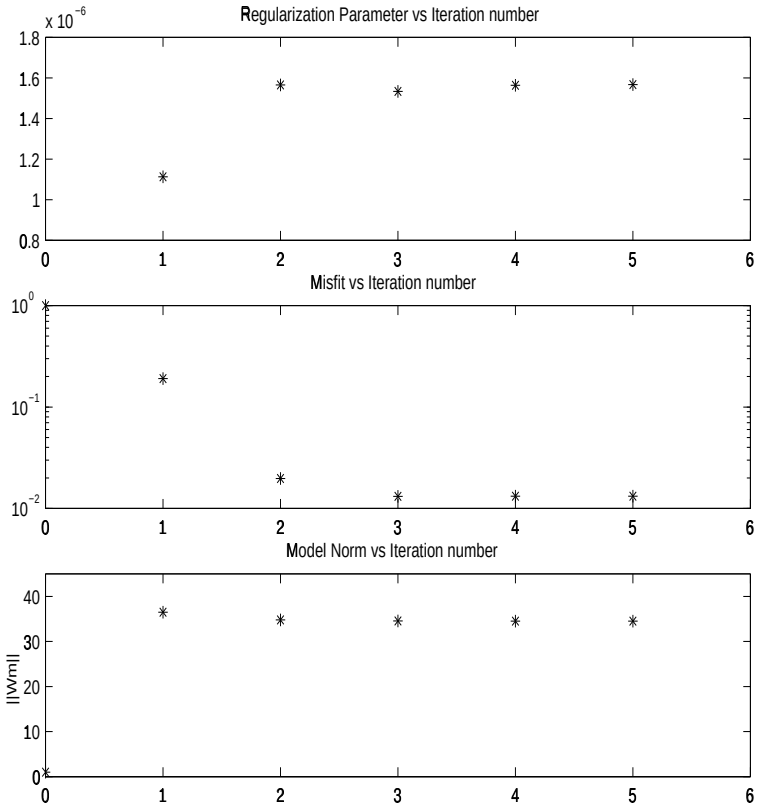

Figure 5. Curves of the regularization parameter, misfit and model norm as a function of iteration in the 1% noise gravity inversion.

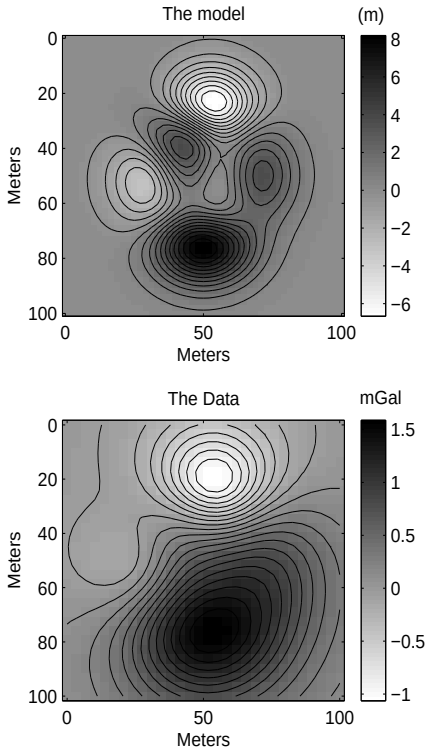

Figure 4. The model and data for the nonlinear gravity example. Positive values in the model indicate a topographic high.

Table 3
Example Two - Numerical experiments with GCV for estimating the noise of the right hand side with hybrid regularization and GCV. Column 1 shows the percentage of noise added to the data. The estimated noise (normalized by $||r_0||$) and shown in column 2, agrees well with the true noise in column 3. E denotes the expected value and is numerically evaluated by computing the average over 50 realizations.

Noise Level	$E(\|\|J(m_0)m_1 - r_0\|\|/\|\|r_0\|\|)$	$E(\|\|\epsilon\|\|/\|\|r_0\|\|)$	$std(\|\|J(m_0)m_1 - r_0\|\|/\|\|r_0\|\|)$
1%	1.0×10^{-2}	1.0×10^{-2}	5.1×10^{-4}
2%	1.9×10^{-2}	1.9×10^{-2}	7.2×10^{-4}
5%	4.9×10^{-2}	4.9×10^{-2}	9.4×10^{-4}
10%	9.8×10^{-2}	9.7×10^{-2}	1.1×10^{-3}
15%	1.4×10^{-1}	1.4×10^{-1}	4.4×10^{-3}
20%	1.9×10^{-1}	1.9×10^{-1}	5.6×10^{-3}

Table 4
Example Two - Numerical experiments with the nonlinear inversion algorithm. The percentage noise added to the data is given in column 1. The magnitude of the estimated noise, shown in column 2, agrees well with the magnitude of the true noise in column 3.

Noise In Data	$\|\|F[m] - b\|\|/\|\|b\|\|$	$\|\|\epsilon\|\|/\|\|b\|\|$
1%	9.4×10^{-3}	9.5×10^{-3}
2%	2.1×10^{-2}	2.0×10^{-2}
5%	5.0×10^{-2}	4.8×10^{-2}
10%	8.4×10^{-2}	8.4×10^{-2}
15%	1.4×10^{-1}	1.4×10^{-1}
20%	1.9×10^{-1}	1.9×10^{-1}

we would also need the Frechet derivatives of the operator. By differentiating equation 8.4 with respect to m we get:

$$\frac{\partial b_j}{\partial m} = \gamma \int\int_D \frac{h + m(x,y)}{[(x-x_j)^2 + (y-y_j)^2 + (h+m(x,y))^2]^{\frac{3}{2}}}\, dx\, dy \qquad (8.5)$$

In order to experiment with this type of problem we assume that 30×30 gravity data are measured on the surface. The data are equally gridded in the interval $[0, 100] \times [0, 100]$ meters. We assume that the reference height is $h(x, y) =$ 20 meters in equation 8.4. The model, $m(x, y)$, has zero mean about this surface. To evaluate the data, the integral 8.4 is calculated using the midpoint rule. The 100×100 square domain is divided into 49×49 grid points. The model and the data are plotted in Figure 4.

8.2. Numerical Experiments

Just as in Section 6, the numerical experiments are built from two parts. In the first part we demonstrate the ability of the GCV, in a restricted subspace, to differentiate between random noise and the nonlinear terms. In the second part we show that our method is able to obtain a model which minimizes the objective function and fits the data to the right extent.

Once again, we calculate the sensitivities with respect to the initial model. Using these sensitivities we carry out one iteration and solve 6.5. The experiment is repeated and each time the data, b, is contaminated with a different amount of noise. For each noise level we carried out 50 realizations. In each realization new random errors are assigned to the data and the process of estimating that noise is repeated. The results of some of these experiment are in Table 3.

The experiments demonstrate the ability of hybrid regularization with GCV to differentiate between the correlated nonlinear terms and the random noise. In this case we have found that the hybrid regularization with GCV is extremely close to the true noise level.

In the second experiment we evaluate the whole process by inverting the data after adding different levels of noise. The results of the experiment are summarized in Table 4 and we again conclude that GCV has worked well in estimating the true noise level.

In Figure 5 we plot the regularization parameter β, the misfit $||F[m] - b||$ and the model norm $||m||$ as a function of the iteration number. We see that the regularization parameter quickly stabilizes to its final value which leads to a regular Gauss-Newton convergence rate. Finally in Figure 6 we plot the results of the inversion for noise levels of $1, 2, 5$ and 10%.

code for these experiments can be obtained through anonymous ftp at *ftp* : *geop.ubc.ca/pub/haber/MLIB.tar*.

8.1. Description of the Problem

A surface gravity survey is carried out to measure the anomalous gravitational acceleration in the vertical direction. From Newton's law we know that the gravitational acceleration at location $(x_i, y_i, 0)$ on the surface of the earth due to a mass anomaly density, $\Delta\rho$, at position (x, y, z) inside the earth is:

$$\Delta g_i = \Delta g(x_i, y_i, 0) = \gamma \frac{\Delta\rho(x, y, z)}{(x - x_i)^2 + (y - y_i)^2 + z^2}\, dx\ dy\ dz \tag{8.1}$$

where γ is the gravitational constant and $dx\ dy\ dz$ is the volume of that mass. In three-dimensions, the anomalous mass can be anywhere in the region D. Integrating over that region and taking the vertical component, we obtain the expression for the measured gravity datum:

$$b_i = b(x_i, y_i, 0) = \gamma \int_D \frac{\Delta\rho(x, y, z)\ z}{[(x - x_i)^2 + (y - y_i)^2 + z^2]^{\frac{3}{2}}}\, dx\ dy\ dz \tag{8.2}$$

It is possible to solve for $\Delta\rho$ as a linear inverse problem however, for some situations a different formulation is more suitable. Assume now that the earth has two layers with known densities and the density contrast between them is $\Delta\rho$. The first layer has a mean depth of $h(x, y)$ which is assumed to be known. Changes in the gravitational field are due to changes in the depth of the first layer relative to h, which is given by the function $m(x, y)$. Assuming $\Delta\rho = 1$, we can write:

$$b_j = \gamma \int\int_D \int_h^{h+m(x,y)} \frac{z}{[(x - x_j)^2 + (y - y_j)^2 + z^2]^{\frac{3}{2}}}\, dx\ dy\ dz \tag{8.3}$$

Integrating the expression with respect to z gives:

$$b_j = \gamma \int\int_D \left(\frac{1}{r_{h_j}} - \frac{1}{r_{m_j}}\right) dx\ dy \tag{8.4}$$

where:

$$r_{h_j}^2 = (x - x_j)^2 + (y - y_j)^2 + h^2$$

and

$$r_{m_j}^2 = (x - x_j)^2 + (y - y_j)^2 + (h + m)^2$$

The goal of the inverse problem presented in this example is to recover the surface $m(x, y)$ from the given gravity anomaly data b. In order to do this

7.2. Using Hybrid Methods in the Nonlinear Inversion

We now have a method to regularize a large scale problem and to estimate the noise. We therefore use this method to solve the nonlinear problem iteratively. The problem can be formulated as follows:

$$\begin{aligned} &minimize \quad ||F[m] - b||^2 + \beta||m||^2 \\ &subject\ to \quad m \in \mathcal{K}\ (J(m), r(m), n) \end{aligned} \tag{7.6}$$

Similar algorithms have been proposed to solve well-posed problems with no regularization [4, 6, 7, 23]. In every nonlinear step we use the combination of Krylov space and Tikhonov regularization in order to regularize the problem and obtain a suggested model m_{k+1}. The new model solves the linearized problem, but it does not necessarily reduce the nonlinear function which we try to minimize. A simple solution to this problem is to use again a step length strategy. This is done by calculating the direction $\delta m = m_{k+1} - m_k$ and using the same method as in the damped Gauss-Newton method to calculate a step length which ensures the reduction of the nonlinear function. The algorithm can be summarized as follows:

Krylov Hybrid Subspace - Damped Gauss-Newton Method
Choose an initial model m_0. Calculate the misfit $\phi_d = ||F[m_0] - b||^2$.
For $k = 1, 2...$
1. Calculate $J(m_k)$ and $r(m_k)$.
2. Solve: $J(m_k)m_{k+1}^p = b - F[m_k] + J(m_k)m_k$
   using the hybrid method and the GCV criterion for β_k.
3. Calculate the perturbation $\delta m = m_{k+1}^p - m_k$, the new misfit
   ${\phi_d}^{new} = ||b - F[m_{k+1}^p]||^2$, and the new model norm ${\phi_m}^{new} = ||m_{k+1}^p||^2$
4. If: ${\phi_d}^{new} + \beta_k {\phi_m}^{new} \leq \phi_d + \beta_k \phi_m$, set $m_{k+1} = m_{k+1}^p$ go to 1.
5. Elseif: ${\phi_d}^{new} + \beta_k {\phi_m}^{new} > \phi_d + \beta_k \phi_m$, set $m_{k+1}^p = m_k + \omega \delta m$ go to 3.
6. If criterion 3.6 is fulfilled, terminate the process.

In the implementation of this process we chose $\omega = 0.5$ and the stopping criterion in 3.6 was $\delta = 10^{-3}$.

8. Example Two - The Interface Problem

In this section we test the algorithm discussed in Section 7 on a large inverse problem. The description of the problem is given in Section 8.1 and the numerical experiments are described in Section 8.2. The MATLAB

Restricting m to the Krylov subspace we can write:

$$Jm = U_{n+1} B_n V_n^T m = r$$

Let $z = V_n m$, and multiplying both sides of the above equation by U_{n+1}^T we get

$$B_n z = \beta_1 e_1 = \tilde{e}_1 \tag{7.3}$$

where $e_1 = [1, 0...0]^T$. The matrix B_n is usually ill-conditioned because some of its singular values are numerically close to zero. In this case one cannot simply invert the matrix, and some regularization is needed. O'Leary and Simmons [30] suggested using the truncated singular value decomposition, [39], to solve 7.3, however Tikhonov regularization could be used as well. The solution is given by:

$$m(n, \beta) = V_n (B_n^T B_n + \beta I)^{-1} B_n^T \tilde{e}_1 \tag{7.4}$$

Note that the hybrid solution depends on the space size as well as the regularization parameter. If we want to make this regularization close to the Tikhonov regularization then we have to ensure that all of the vectors associated with significant singular values have converged. This requires that the decomposition is continued long enough so only small singular values are obtained with succesive iterations. This can be done by noting that the bidiagonal matrix B_n contains an approximation to the singular values of the system [32]. To ensure that we capture some of the small singular values we carry out the SVD of the small sparse matrix B_n. We continue the decomposition until the matrix B_n has n_1 singular values which are smaller than a small number δ. The number δ is chosen such that it is practically zero according to working precision. In this work we chose $\delta = 10^{-6}$. After the decomposition has been obtained it is used in equation 7.4 to calculate the solution. In most cases we found that the results were satisfactory if the number of almost zero singular values is roughly $10 - 20\%$ of the total number of singular values of B_n .

In order to solve 7.4 we need to choose a regularization parameter β. Since the problem has been reduced to an n-dimensional space where usually $n \ll N$, it is possible to use the GCV procedure in this space. The use of the GCV on the rotated problem is justified since the GCV is not sensitive to orthogonal transformation [15]. Therefore we choose a regularization parameter which minimizes:

$$GCV(n, \beta) = \frac{||(I - B_n (B_n^T B_n + \beta I)^{-1} B_n^T) \tilde{e}_1||^2}{[trace(I - B_n (B_n^T B_n + \beta I)^{-1} B_n^T)]^2} \tag{7.5}$$

The evaluation of the GCV function in this case is very cheap and the minimization can be carried out easily.

7. Application for Large Scale Problems

So far we have discussed Tikhonov-style regularization for nonlinear problems which is based on the ability to estimate a regularization parameter and invert the regularized Frechet derivatives. However the inversion of the regularized Frechet derivatives is computationally burdensome for large scale problems and should be avoided. Noticing that at each iteration we have to regularize the linearized system 2.9, we turn to other possible regularization methods. One of the most stable methods for large scale regularization is hybrid type regularization, [16, 21, 22, 29, 30]. We now review this method.

7.1. Hybrid Regularization Methods

Hybrid regularization is a combination of two regularization methods, subspace regularization and Tikhonov regularization. In the first stage the problem is reduced to a small and tractable subspace, and then Tikhonov regularization is used inside that subspace. In this subsection we assume that the weighting matrix $W = I$. This assumption is generic since the case of arbitrary W can be reduced to the case $W = I$ (see [10, 11, 12, 13, 19, 22]). A general hybrid approach solves the problem:

$$\begin{aligned} &minimize \quad \phi = ||Jm - r||^2 + \beta||m||^2 \\ &subject\ to \quad m \in S_n \end{aligned} \tag{7.1}$$

There are many possible ways to choose the subspace S_n [19, 22, 29, 31]. In this work we consider the case that $S_n = \mathcal{K}(J, r, n)$ which is the Krylov space, that is,

$$S_n = [J^T r, (J^T J)J^T r, ..., (J^T J)^{n-1} J^T r]$$

In order to obtain the subspace decomposition we use Lanczos bidiagonalization procedure [16, 33]. After n iterations of the bidiagonalization process, we have a partial decomposition:

$$J \approx U_{n+1} B_n V_n^T \tag{7.2}$$

where $U_{n+1}^T U_{n+1} = I_{n+1}$ with $U(:,1) = r/\beta_1$, $V_n^T V_n = I_n$ and B_n is an $(n+1)\times n$ matrix

$$B_n = \begin{bmatrix} \alpha_1 & 0 & ... & 0 \\ \beta_1 & \alpha_2 & ... & 0 \\ .. & .. & ... & .. \\ .. & .. & ... & 0 \\ 0 & ... & \beta_{n-1} & \alpha_n \\ 0 & ... & 0 & \beta_n \end{bmatrix}$$

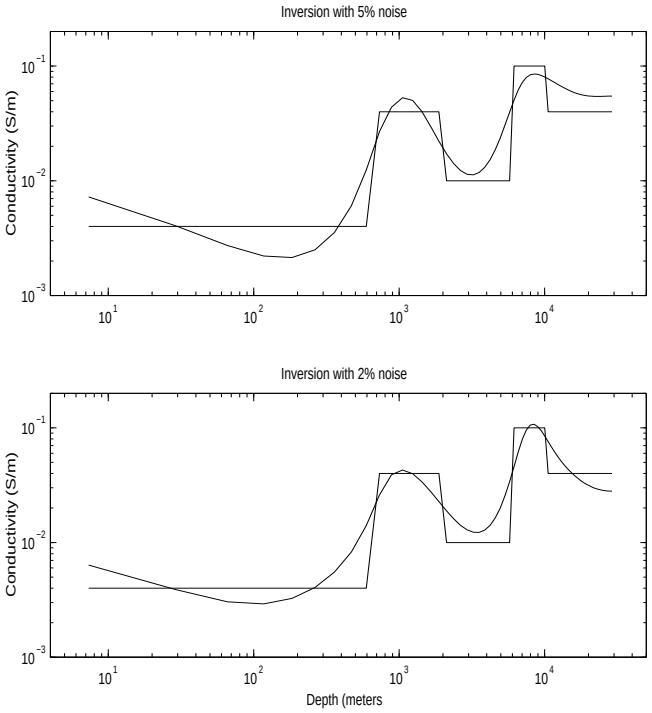

Figure 3. Results of nonlinear MT inversion. Data are contaminated with 5% (top) and 2% (bottom) noise.

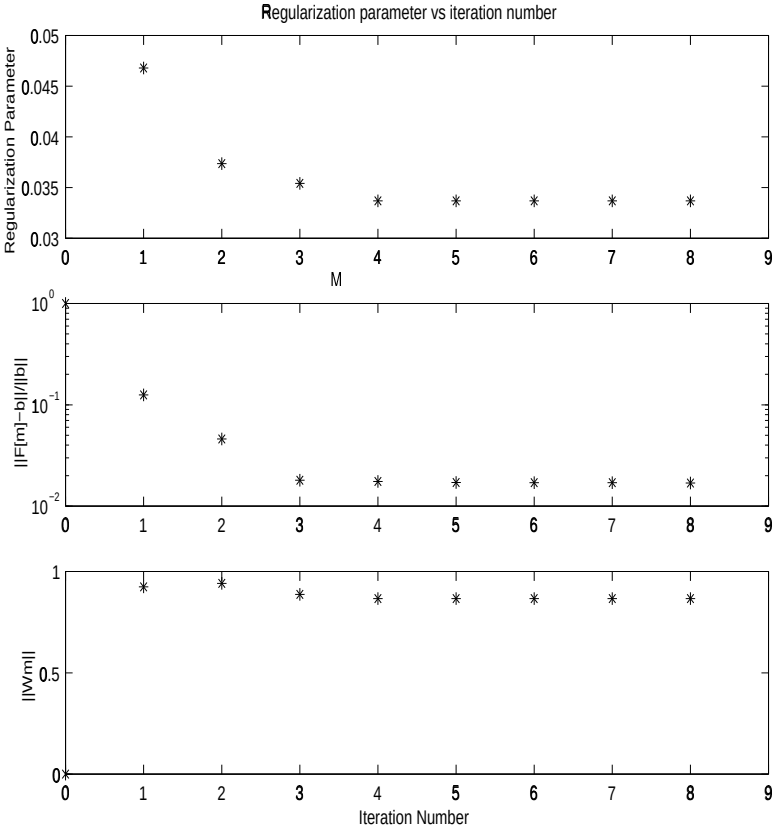

Figure 2. Curves of the regularization parameter, misfit and model norm as a function of iteration in an inversion with data contaminated with 5% noise.

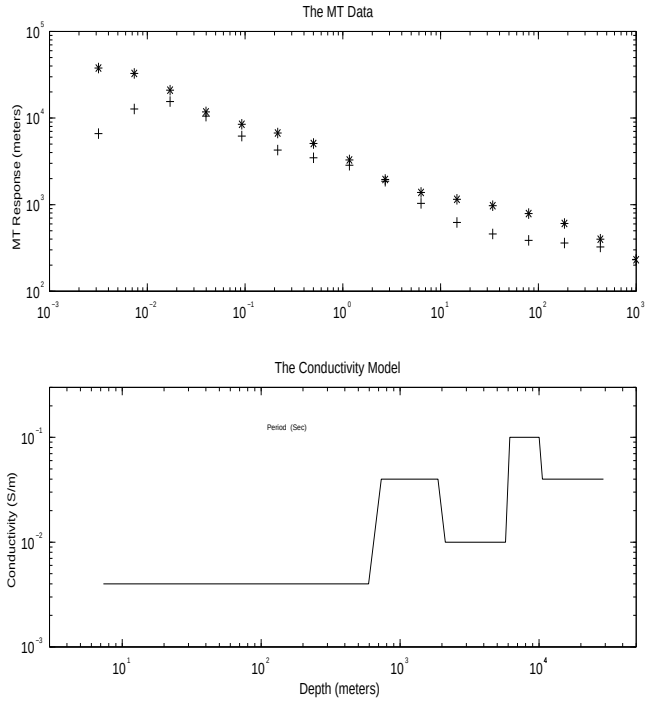

Figure 1. The conductivity model used for the MT experiment (bottom) and the data from that model (top). The imaginary part of the data is denoted by $*$ and the real part is denoted by $+$.

Table 1
Example one - Numerical experiments with GCV for estimating the noisy parts of the right hand side of the linearized equation. Column 1 shows the percentage of noise added to the data. The estimated noise (normalized by $||r_0||$) and shown in column 2, agrees well with the true noise in column 3. E denotes the expected value and is numerically evaluated by computing the average over 50 realizations.

Noise Level	$E(\|\|J(m_0)m_1 - r_0\|\|/\|\|r_0\|\|)$	$E(\|\|\epsilon\|\|/\|\|r_0\|\|)$	$std(\|\|J(m_0)m_1 - r_0\|\|/\|\|r_0\|\|)$
1%	8.7×10^{-3}	9.0×10^{-3}	5.2×10^{-4}
2%	1.7×10^{-2}	1.8×10^{-2}	6.4×10^{-3}
5%	4.2×10^{-2}	4.4×10^{-2}	7.2×10^{-3}
10%	8.4×10^{-2}	8.8×10^{-2}	9.1×10^{-3}
15%	1.2×10^{-1}	1.3×10^{-1}	5.7×10^{-2}
20%	1.6×10^{-1}	1.7×10^{-1}	6.2×10^{-2}

Table 2
Example One - Numerical experiments with the nonlinear inversion algorithm. The percentage noise added to the data is given in column 1. The magnitude of the estimated noise, shown in column 2, agrees well with the magnitude of the true noise in column 3.

Noise In Data	$\|\|F[m] - b\|\|/\|\|b\|\|$	$\|\|\epsilon\|\|/\|\|b\|\|$
1%	7.0×10^{-3}	7.3×10^{-3}
2%	1.9×10^{-2}	1.7×10^{-2}
5%	4.6×10^{-2}	4.2×10^{-2}
10%	7.3×10^{-2}	7.1×10^{-2}
15%	1.2×10^{-1}	1.2×10^{-1}
20%	1.6×10^{-1}	1.5×10^{-1}

leads to a regular Gauss-Newton convergence rate. Finally, in Figure 3, we plot the result of the inversion for the 2 and the 5% noise levels. Both inversions are smooth representations of the blocky model and no manifestation of artificial structure caused by fitting the noise is observed.

where W_d is a diagonal scaling matrix with the inverse of the estimated magnitude of standard deviation of each datum. In this way each datum has the same order of standard deviation. Below, for sake of brevity in notation, we use F and b instead of $\hat{F}$ and $\hat{b}$.

6.2. Numerical Experiments

The numerical experiments are built from two parts. In the first part we demonstrate the ability of the GCV to differentiate between random noise and the nonlinear terms. In the second part we show that our method is able to obtain a model that minimizes the objective function and fits the data to the right extent.

To investigate the ability for the GCV criterion to estimate the random noise in the nonlinear problem we carry out the following analysis that emulates what is done at the first iteration in solving the nonlinear inverse problem. Let the starting model m_0 be equal to the reference model ($m_{ref} = 0.04S/m$). This is substantially different from the true model in Figure 1. We next calculate the sensitivities at the model m_0 and solve

$$\begin{bmatrix} J(m_0) \\ \sqrt{\beta}W \end{bmatrix} m_1 = \begin{bmatrix} b - F[m_0] + J(m_0)m_0 \\ \sqrt{\beta}W m_0 \end{bmatrix} \tag{6.5}$$

for a range of β values to generate the GCV curve. The optimum β is obtained by finding the minimum in the GCV curve. Equation (6.5) is then solved for m_1 and the estimated error for the noise, $||J(m_0)m_1 - r_0||$, is easily obtained. This estimated noise can be compared with the true added Gaussian noise $||\epsilon||$ that was added to the data b. The numerical results are likely dependent upon the magnitude and particular realization of the additive Gaussian noise. To account for this we consider a number of different experiments in which the data b is contaminated with a different percentage of Gaussian noise. For each noise level we carried out 50 realizations. In each realization, new random errors are added to the data and the process of estimating that noise is repeated. In Table 1 we give the results of these experiments in terms of the mean of the estimated noise in the 50 realizations, the mean of the real noise, and the standard deviation of the 50 realizations. We note that the estimated noise (Column 2) is in good agreement with the true noise (Column 3).

In the second set of tests we evaluate the complete inversion process by inverting the data set for different noise levels. Only one realization of the noise is used at each noise level. The results of the experiment are summarized in Table 2. We see that the algorithm presented in Section 5 estimated a noise level that was very close to the real noise level.

In Figure 2, we plot the regularization parameter β, the relative residual $||F[m]-b||/||b||$ and the model norm $||Wm||$ as a function of the iteration number. We see that the regularization parameter quickly stabilizes to its final value which

$$E(0) = 1$$

The MT data, measured at the surface of the earth are

$$c_0(\omega, \sigma(z)) = -\frac{E(z=0,\omega)}{\partial_z E(z=0,\omega)} \tag{6.3}$$

Our goal is to recover the conductivity profile $\sigma(z)$ from the complex measurements c_0 at different frequencies ω.

To carry out forward modelling and inversion we discretize the earth into 64 layers. The thickness of the layers increases quadratically with depth such that $z_i = \zeta_i^2$ and ζ is linearly spaced. The total thickness of the model is chosen such that $E \simeq 0$ at the bottom boundary. The differential equation 6.2 is solved in each layer, along with the conditions that E and its derivative E' are continuous at each interface. This leads to the recursive relation for $C(z,\omega) = -E(z,\omega)/E'(z,\omega)$ which constitutes the forward modelling [43]:

$$C(z_{j-1},\omega) = \frac{1}{k_j}\frac{tanh(k_j h_j) + k_j C(z_j,\omega)}{1 + k_j C(z_j,\omega) tanh(k_j h_j)} \qquad j = M...1 \tag{6.4}$$

where

$$k_j = (1+i)\sqrt{\omega \mu_0 \sigma_j},$$

h_j is the thickness of the j^{th} layer and σ_j is the conductivity of the j^{th} layer. The data is given simply by $c_0(\omega) = C(0,\omega)$.

To generate simulated field data, we need to specify both the conductivity structure and also the frequencies at which the measurements are made. We choose the conductivity model used by Whittall and Oldenburg [44]. Data are calculated at 16 frequencies over a range from $10^{-3} - 10^3$ Hz. The conductivity model and the real and imaginary parts of the data are plotted in Figure 1.

In order to carry out the inversion we need to pick a reference model and a weighting matrix W. We set the reference model m_{ref} so that it corresponds to a background conductivity of $0.04 \quad S/m$. To specify W we use the middle difference discretization of the operator $0.001I - \nabla_\zeta^2$, which means that the finite difference operator is discretized with respect to ζ.

The MT problem is characterized by data that span a few orders of magnitude. In such cases the errors in the data are generally proportional to the data magnitude, and thus the standard deviation of the data varies significantly. However, the GCV is theoretically designed for problems with uniform standard deviation. To accomodate this, we change the data into

$$\hat{b} = W_d b$$

and the forward modelling of the MT problem to

$$\hat{F}[m] = W_d F[m]$$

attraction point assuming that the Lipschitz constant of this fixed point iteration is smaller than one. Our experience so far is that this has happened and that the regularization parameter tends to converge quickly to its final value. The complete algorithm is as follows:

Nonlinear Inversion Using GCV and Damped Gauss-Newton

1. Calculate the sensitivities $J(m)$ and the right hand side $r(m)$.
2. Evaluate a suitable regularization parameter, β, using the GCV.
3. Use the regularization parameter to calculate m_{k+1} and $\delta m = m_{k+1} - m_k$
4. Update using step length strategy and criterion 3.3.
5. Check for convergence and go to 1.

6. Example I: The One-dimensional Magnetotelluric Problem

The one dimensional magnetotelluric (MT) problem is a well-studied nonlinear inverse problem in electromagnetics, (see for example [27, 31, 43, 44]) and therefore it is very suitable as a test example. We first provide the mathematical description of the problem and then proceed to the numerical experiments. The matlab code for these experiments can be obtained through anonymous ftp at $ftp : geop.ubc.ca/pub/haber/MLIB.tar$.

6.1. Description of the Problem

We start with the quasi-static Maxwell's equations in the frequency domain:

$$\nabla \times \mathcal{E} = -i\omega\mu_0 \mathcal{H} \tag{6.1}$$

$$\nabla \times \mathcal{H} = \sigma \mathcal{E}$$

where $\mathcal{E}$ and $\mathcal{H}$ are the three-component electric and magnetic field intensities, μ_0 is the magnetic permeability, which is assumed to be constant, ω is the angular frequency and σ is the electrical conductivity.

Assuming that $\mathcal{E}$ and $\mathcal{H}$ are plane waves, i.e.

$$\mathcal{E} = (E, 0, 0) \qquad \mathcal{H} = (0, H, 0)$$

and that $\sigma = \sigma(z)$, we can take the curl of the first equation in 6.1 and get the governing equation

$$\frac{d^2 E}{dz^2} = i\omega\mu_0 \sigma(z) E \tag{6.2}$$

The boundary conditions are

$$E(\infty) = 0$$

The minimum of the CV function with respect to β represents the β for which the data predicted would change the least if we omit one data point. The CV as it is defined above is not very practical to compute. Let

$$\phi = ||Jm - r||^2 + \beta||Wm||^2 \tag{4.3}$$

and let $r(\beta) = Jm(\beta)$. Whaba [41] proved that

$$CV(\beta) = \sum_{n=1}^{N} \frac{(r(\beta)_n - r_n)^2}{(1 - c_{nn})^2} \tag{4.4}$$

where c_{nn} is the nn term of $C(\beta) = J(J^TJ + \beta W^TW)^{-1}J^T$ and $r(\beta)_n$ is the n^{th} component of $r(\beta)$. The Cross Validation function was replaced by the Generalized Cross Validation in order to keep it's minimum under orthogonal transformation (Golub *et al.* [15]):

$$GCV(\beta) = \frac{||r(\beta) - r||^2}{trace(I - C(\beta))^2} \tag{4.5}$$

.

While the GCV has a good theoretical background, it is often hard to calculate in practice since the calculation of C is computationally expensive. A new approach by Golub and Von Matt [18] makes the computation less intensive and feasible for large scale overdetermined problems. We will use a similar approach in Section 7.

5. Summary of Our Inversion Algorithm

We now summarize our methodology. It is based on three main components.
1. A method to pick a regularization parameter (GCV).
2. A criterion to accept/reject a step
3. A method to alter a step size (in the case of rejection).

In the above list, item one is based on the ability of GCV to differentiate between noise and signals, while items two and three are based on the Damped Gauss-Newton method. In general our methodology can be simply viewed as a variation of the Damped Gauss-Newton method, which uses a new regularization parameter β that is changing in each iteration. If the regularization parameter approaches a specific value β^* then our algorithm turns into a standard Damped Gauss-Newton algorithm. Note also that our algorithm is a fixed point iteration of the form

$$m_{k+1} = f(m_k, \beta(m_k))$$

where $f(m_k, \beta(m_k))$ is defined in equation 2.8 and β_k is defined by the GCV methods. Therefore our algorithm should converge if we start close enough to its

Instead, we replace our convergence criteria with one that is commonly used for least squares problems [8]:

$$max(\frac{|g_k^{(i)} m_k^{(i)}|}{\phi(m_k)};\quad i = 1...M) \;<\; tol \tag{3.5}$$

An additional convergence criterion is to ensure that the model is stationary [14], that is

$$\frac{||m_{k+1} - m_k||}{max(||m_{k+1}||, ||m_k||)} < \delta \tag{3.6}$$

We now return to the explanation of stage two of our algorithm in which we need to choose the regularization parameter β for each iteration. We require a method that can differentiate between the nonlinear terms and the noisy terms. The nonlinear terms are well correlated and therefore, we can use methods that were developed for linear problems with uncorrelated noise. The method which we use in this paper is the Generalized Cross Validation. We review this method next.

4. Estimating Gaussian Noise In a Linear Inverse Problem

The problem of estimating noise in a linear inverse problem has been treated by [15, 22, 40, 41]. One popular method is the Generalized Cross Validation (GCV). The GCV has a firm theoretical foundation and numerical experiments [1, 41], justify our choice for GCV as a tool to differentiate between correlated and uncorrelated noise. We therefore review the GCV here. For ease of notation in this section we let $m = m_{k+1} - m_{ref}$, $r_k = r(m_k) - Jm_{ref}$ and $J = J(m_k)$.

The major idea of the GCV is that we do not want our computed model to be sensitive to the elimination of one data point. This means that our solution should predict a datum fairly well even if that datum is not used when calculating the model. In order to do that we introduce the following notation. Let $m_n(\beta)$ be the minimizer of

$$\phi_n = ||Jm - r||^2 - (a_n^T m - r_n)^2 + \beta||Wm||^2 \tag{4.1}$$

where a_n is the n^{th} row of J and r_n is the n^{th} element of r. Note that ϕ_n is the same linearized objective function as ϕ^{lin} in 2.10 but with the n^{th} element in r, and the corresponding row in J, missing.

For each regularization parameter, β, the function ϕ_n can be minimized to yield a solution $m_n(\beta)$. The Cross Validation function is defined as the sum of square differences between the predicted right hand side without the n^{th} element and the actual n^{th} element:

$$CV(\beta) = \sum_{n=1}^{N} (a_n^T m_n(\beta) - r_n)^2 = \sum_{n=1}^{N} (r_n(\beta) - r_n)^2 \tag{4.2}$$

4. Update using step length strategy.
5. Check for convergence and go to 1.

There are three important points in this algorithm that need explanation. One of the most important is how to pick a regularization parameter (stage 2) but we delay this until the next section. We also need to explain how to make the update (stage 4) and how to check for convergence (stage 5). We first consider the updating.

At the k^{th} iteration we have β_k, m_k and equation (2.6) can be solved for m_{k+1}. Acceptance of m_{k+1} (or the perturbation $\delta m = m_{k+1} - m_k$) when a fixed β was used in the last section, required that $\phi(\beta, m_{k+1}) < \phi(\beta, m_k)$. The main problem in applying this criterion here is that the objective function changes from iteration to iteration because the regularization parameter can vary. At the k^{th} iteration

$$\phi(\beta_k, m_k) = \beta_k ||W(m_k - m_{ref})||^2 + ||F[m_k] - b||^2$$

while at the $(k+1)^{th}$ iteration

$$\phi(\beta_{k+1}, m_{k+1}) = \beta_{k+1} ||W(m_{k+1} - m_{ref})||^2 + ||F[m_{k+1}] - b||^2$$

Since the global objective function ϕ is changing at each iteration, the demand of decreasing the value of the objective function is not reasonable. We therefore replace it with the consistent demand:

$$\phi(\beta_{k+1}, m_{k+1}) < \phi(\beta_{k+1}, m_k) \tag{3.2}$$

or more specifically:

$$\beta_{k+1} ||W(m_{k+1} - m_{ref})||^2 + ||F[m_{k+1}] - b||^2 < \beta_{k+1} ||W(m_k - m_{ref})||^2 + ||F[m_k] - b||^2 \tag{3.3}$$

Thus every step in the algorithm is also equivalent to a one-step descent from the model m_k to m_{k+1} with regularization parameter β_{k+1}. This is important if we want our algorithm to be consistent with the objectives of the minimization.

The second important point is the convergence (stage 5). For every iteration k we need to know whether to stop the process or to continue. Since the objective function is changing, the question is what criterion should we take? The answer is again given by consistency. If each iteration is a damped Gauss-Newton iteration with different parameter β_k then our convergence criterion is the same as for a damped Gauss-Newton algorithm. Let:

$$g_k = \beta_k W^T W(m_k - m_{ref}) + J(m_k)^T (F[m_k] - b) \tag{3.4}$$

be the gradient of the k^{th} functional. A simple way to test convergence is to demand that $||g_k|| \approx 0$. This is not easily implemented because scaling is a factor.

observation, Armijo [2] suggested trying a step size of $\omega\delta m$ where $0.1 < \omega < 0.5$ and to repeat the process. Finally, for a small enough step, the nonlinear function decreases and we get closer to the minimum of the nonlinear functional. Since every step is a descent step it is reasonable to demand that at the k^{th} iteration

$$\phi(\beta, m_{k+1}) < \phi(\beta, m_k) \tag{2.11}$$

This iterative procedure is repeated until convergence. The result is a model $m^{\dagger}(\beta)$ which minimizes 1.2.

As stated before, in most applications the regularization parameter cannot be chosen *a priori* and therefore an adaptive method to find regularization parameter is needed. This is done in the next section.

3. Nonlinear Inversion Combined with the Damped Gauss-Newton Method

In this section we develop the main idea of our paper. Our goal is to simultaneously find an m and regularization parameter β such that ϕ in 1.2 is minimized and ϕ_d is a good estimate of the true noise. To accomplish this we notice that at the k^{th} iteration we must solve the linear system in 2.9. This system can be written as

$$J(m_k)m_{k+1} = b^c - F[m_k] + J(m_k)m_k + \ nonlinear\ terms\ +\ noise \tag{3.1}$$

where b^c is the exact data i.e., the data with no noise. We therefore have to deal with two problems. The first is the measurement noise and the second is the nonlinear terms. While the noise has to be treated through a global regularization (i.e. choosing a regularization parameter for the whole problem), the nonlinear terms have to be treated by reducing the step size which is part of a local regularization process (i.e. the step size regularizes one iteration). As we get closer to the minimum, full steps are taken and no local regularization is needed [8]. These processes are distinct and therefore we want to treat them separately and differently. First we deal with the measurement noise. This requires a method which can differentiate between the Gaussian noise and the correlated nonlinear terms. This method yields the value of β for the current iteration. Now with a β in hand we can carry out a damped Gauss-Newton iteration. The damping in a Gauss-Newton strategy takes the nonlinearity into consideration and makes sure that the perturbation is small enough so that nonlinear terms are actually small. In general, our algorithm can be summarized as follows:

Steps for Solving Nonlinear Inverse Problem

1. Calculate the sensitivities $J(m)$ and the right hand side $r(m)$.
2. Calculate a suitable regularization parameter β.
3. Use the regularization parameter to calculate m_{k+1}.

where $R(m, \delta m) = \mathcal{O}(\delta m^2)$. With the assumption that $R(m, \delta m)$ is small, this linearization yields the Gauss-Newton equations

$$\beta W^T W(m + \delta m - m_{ref}) + J(m)^T(F[m] + J(m)\delta m - b) = 0 \qquad (2.4)$$

Rearranging terms in this equation gives a linear system of equations for δm.

The minimization problem is solved iteratively. At the k^{th} iteration we solve

$$(J(m_k)^T J(m_k) + \beta W^T W)\delta m = J(m_k)^T(b - F[m_k]) - \beta W^T W(m_k - m_{ref}) \quad (2.5)$$

to find the perturbation δm. Solving 2.5 is identical to solving the least squares problem

$$\begin{bmatrix} J(m_k) \\ \sqrt{\beta}W \end{bmatrix} \delta m = \begin{bmatrix} b - F[m_k] \\ -\sqrt{\beta}W(m_k - m_{ref}) \end{bmatrix} \qquad (2.6)$$

At each iteration we have an option of solving directly for a perturbation or solving for an updated model. To formulate the latter option we write $m_{k+1} = m_k + \delta m$ and substitute into equation 2.5 to obtain

$$(J(m_k)^T J(m_k) + \beta W^T W)m_{k+1} = J(m_k)^T(b - F[m_k] + J(m_k)m_k) - \beta W^T W m_{ref} \qquad (2.7)$$

The corresponding linear system is

$$\begin{bmatrix} J(m_k) \\ \sqrt{\beta}W \end{bmatrix} m_{k+1} = \begin{bmatrix} b - F[m_k] + J(m_k)m_k \\ \sqrt{\beta}W m_{ref} \end{bmatrix} \qquad (2.8)$$

Note that equation 2.8 is equivalent to solving the linear system

$$J(m_k)m_{k+1} = b - F[m_k] + J(m_k)m_k = r_k(m_k) \qquad (2.9)$$

using Tikhonov regularization with W as a weighting matrix and m_{ref} as a reference model. Therefore, at each iteration the new update m_{k+1} minimizes the linearized Tikhonov function

$$\phi^{lin} = ||F[m_k] + J(m_k)(m_{k+1} - m_k) - b||^2 + \beta||W(m_{k+1} - m_{ref})||^2 = \phi_d^{lin} + \beta\phi_m \qquad (2.10)$$

Using the above formulation, we can solve for either the perturbation δm or for the model m_{k+1} and proceed to the next iteration. If we start from a point which is close enough to the solution, the method converges. In most geophysical problems however, we do not start very close to the solution and therefore the linearization process which involves neglecting the residual $R(m, \delta m)$ from equation 2.3 might not be justified. The Taylor residual $R(m, \delta m)$ is of $\mathcal{O}(\delta m^2)$ and therefore when the step size is large the minimization of the linearized equation 2.10 might cause an increase in the objective function 1.2. Such a step is obviously a bad choice because it does not take us closer to the minimum of the nonlinear problem. This defect can easily be corrected. The Gauss-Newton step is a descent direction [8], and therefore if the step size is small enough, the nonlinear function will have a similar behaviour to the linearized problem. Based on this

one, the misfit ϕ_d is reduced to the target misfit, T, and in stage two this target misfit is kept constant while the model norm ϕ_m is reduced. Since the process is made from two stages we call this method the Two-Stage Method. Although the Two-Stage Method is very popular in practice, it does not have a proof of convergence and it can be shown to diverge in some cases [19]. In order to avoid divergence, "safety steps" have been suggested in [28], however these steps are *ad hoc* in nature and employ parameters which are taken from practical experience.

The goal of this paper is to present another method to choose an adaptive regularization parameter. The method is stable, does not contain *ad hoc* parameters, and it does not need a pre-defined target misfit.

The paper is built as follows. We start with a review of the Damped Gauss-Newton method as applied to a minimization problem with a constant regularization parameter. We then present the major ideas of our algorithm. Our algorithm uses the Generalized Cross validation and we review it in Section 4. Section 5 is used to summarized the algorithm. In Section 6, we give a 1-D example taken from magnetotelluric (MT) experiments. Section 7 deals with the application of this technique to large scale problems. Finally, Section 8 gives an example of inverting data from a large scale gravity problem.

2. Damped Gauss-Newton Method

If β is specified then the minimization of the global objective function 1.2 is a well-posed nonlinear optimization problem. In this section we assume β is known and review the Damped Gauss-Newton method for the minimization of 1.2 as suggested in [8, 14, 24, 25].

Since we look for an m that minimizes ϕ, we differentiate 1.2 with respect to m and set the result to zero. We obtain

$$\frac{\partial \phi}{\partial m} = g(m) = \beta W^T W(m - m_{ref}) + J(m)^T(F[m] - b) = 0 \qquad (2.1)$$

where $g(m)$ is the gradient of 1.2 and $J(m)$ is the sensitivity:

$$J(m) = \frac{\partial F}{\partial m} \qquad (2.2)$$

If we find $m^\dagger$ that solves 2.1, then we have found the desired solution (with respect to that fixed β). A Newton method for solving 2.1 requires the differentiation of 2.1 with respect to m. The main difficulty with the Newton method is that 2.1 is a nonlinear equation, which already involves the Frechet derivative of F with respect to m. In order to avoid the calculation of the second Frechet derivative of F to m, the forward modeling, F is linearized by:

$$F[m + \delta m] = F[m] + J(m)\delta m + R(m, \delta m) \qquad (2.3)$$

The inverse problem is usually ill-posed which means that there is more than one model, m, which fits the noisy data, b. In order to solve such problems, regularization and minimization are commonly used [12, 26, 31, 38]. Our goal is to find the model m^* which minimizes the Tikhonov functional:

$$\phi = ||F[m] - b||^2 + \beta||W(m - m_{ref})||^2 = \phi_d + \beta\phi_m \tag{1.2}$$

where $||.||$ represents the two-norm, W is a positive weighting function, m_{ref} is a reference model and β is the regularization parameter. The quantities ϕ_m and ϕ_d are the model norm and the misfit respectively.

There are two main computational difficulties in the solution of 1.2. First, the regularization parameter is unknown, and second, a nonlinear functional has to be minimized. These impose special difficulties if the problem is large.

The choice of regularization parameter is addressed in [3, 5, 12, 26, 34, 37, 38]. The methods for the selection of regularization parameter can be roughly divided into two groups. In [3, 12, 26, 36, 37, 42], the regularization parameter is estimated *a priori* and problem 1.2 is solved using a fixed regularization parameter, β. A second approach suggested in [3, 31, 34], is to estimate the regularization parameter at each iteration. This approach is used in many practical applications [13, 28, 35, 45].

In the first method which uses a fixed regularization parameter, the minimization of 1.2 is done a few times. Each minimization is solved with a regularization parameter β_{est} and a solution m_{est} is obtained. If the solution is satisfactory as judged by some criterion, then the inverse problem is considered to have been solved. Let β^* be the regularization parameter which accomplishes this goal. The criterion suggested in [3, 5, 34, 37, 38], is the discrepancy principle which means that:

$$||F[m] - b||^2 \leq T \tag{1.3}$$

where the number T is the target misfit. This requires knowledge of the noise, ϵ, but unfortunately for most field data the noise is not known. The GCV has been successful in estimating Gaussian and uncorrelated additive noise in linear problems [41]. Whaba and O'Sullivan [42] also used a GCV criterion in a nonlinear problem but their implementation fall into the category of fixing the regularization parameter and therefore it requires that the minimization of 1.2 be carried out a number of β values.

The main difficulty with the above approach is that we need to solve a new nonlinear problem for every new regularization parameter and therefore the algorithm is computationally expensive. The algorithm is substantially cheaper if we estimate β to be close to the optimal β^*. In [3, 12], considerable effort is made to choose such a β, however, the estimates are asymptotic in nature and therefore 1.2 still has to be solved many times.

The second method for solving the inverse problem is to use an adaptive regularization parameter [5, 31]. The process is divided into two stages. In stage

A GCV Based Method for Nonlinear Ill-Posed Problems

Eldad Haber and Douglas Oldenburg

Dept. of Computer Science UBC
E-mail: haber@cs.ubc.ca
Dept. of Earth and Ocean Science UBC
E-mail: doug@geop.ubc.ca

This paper discusses the inversion of nonlinear ill-posed problems. Such problems are solved through regularization and iteration and a major computational problem arises because the regularization parameter is not known *a priori*. In this paper we show that the regularization should be made of two parts. A global regularization parameter is required to deal with the measurement noise, and a local regularization is needed to deal with the nonlinearity. We suggest the Generalized Cross Validation as a method to estimate the global regularization parameter and the Damped Gauss-Newton to impose local regularization. Our algorithm is tested on the magnetotelluric problem.

In the second part of this paper we develop a methodology to implement our algorithm on large-scale problems. We show that hybrid regularization methods can successfully estimate the global regularization parameter. Our algorithm is tested on a large gravimetric problem.

Keywords: Damped Gauss-Newton, Generalized Cross Validation, Magnetotelluric, Noise, Nonlinear Inverse Problems, Regularization

AMS Subject classification: 35R30, 65R30, 65Y99

1. Introduction

In this paper we deal with the solution of nonlinear ill-posed inverse problems. Let:

$$\mathcal{F}[m] + \epsilon = b \tag{1.1}$$

represent a nonlinear inverse problem. Our goal is to reconstruct the model, $m \in \mathcal{H}$ from the noisy data $b \in R^N$, where $\mathcal{H}$ denotes the Hilbert space. The transformation $\mathcal{F} : \mathcal{H} \rightarrow R^N$ is the forward modelling and ϵ, the measurement noise, is assumed to be Gaussian and non-correlated. The operator $\mathcal{F}$ is assumed to be twice Frechet differentiable. Since $m \in \mathcal{H}$ it is common to discretize the model with M unknowns such that $M > N$ (see for example [22, 26, 31, 34, 37, 41]). The transformation $\mathcal{F}$ becomes a discrete transformation: $F : R^M \rightarrow R^N$.