Original Research Paper

Volume - 13 | Issue - 07 | July - 2023 | PRINT ISSN No. 2249 - 555X | DOI : 10.36106/ijar

Mathematics

REGULARIZING NONLINEAR INVERSE PROBLEMS THAT WE CALL THE NONLINEAR X2 METHOD

Israa Mohammed Wannas

ABSTRACT Regularizing nonlinear inverse problems uses multiple Theorems to analyze the least squares solution is sometimes called the x2 test. If J (^x) is much larger than the mean of x2 n-m or is outside of some confidence interval, for example say a 95% confidence interval, then this would suggest that the errors in the data are larger than expected. This indicates that either the covariance of the data errors is too small or the forward problem does not accurately model the physical system. Therefore, Theorem supplies some very useful information about the solution to the inverse problem. This result only applies to linear problems, however, it is shown in that it approximately holds for nonlinear problems in a region around ^x. For nonlinear issues, the least squares estimate is not as straightforward. Since the minimum cannot be calculated analytically, an iterative approach must be adopted. Finding the solution falls within the umbrella of the mathematical discipline of optimization. This nonlinear unconstrained optimization problem can be solved using a wide range of techniques, such as genetic algorithms, stochastic algorithms, particle swarm optimization algorithms, quantum optimization algorithms (which require a quantum computer), pattern search algorithms, direct search techniques, steepest descent algorithms, and conjugate gradient algorithms. However, Newton's method-style algorithms are among the fastest and most effective, and they can even provide quadratic convergence if the function behaves properly (that is, if it is continuous and twice differential in the domain)

KEYWORDS : Non-linear inverse problem, regularization, x2 distributions and real distributions

INTRODUCTION

Inverse problems often need to be regularized because they are poorly presented or poorly conditioned. The prevalent approach of Tikhonov regularization necessitates the estimation of an additional parameter known as the regularization parameter. Mead first proposed the x2 technique in, which estimates the regularization parameter for linear inverse problems using the x2 distribution of the Tikhonov functional. The distribution of the Tikhonov functional is unknown for nonlinear inverse situations. In this thesis, we use both Gauss Newton iterations and Levenberg Marquardt iterations to extend the x2 approach to nonlinear problems. For the quadratic functional that appears in Gauss Newton and Levenberg Marquardt iterations, we derive approximate x2 distributions. Two ill-posed nonlinear inverse problems—a nonlinear cross-well tomography problem and a subsurface electrical conductivity estimation challenge—are used to illustrate the methodology. By proving that the theoretical x2 distributions and real distributions closely match each other, we quantitatively verify the viability of the assumptions made in this technique. Two algorithms based on the Gauss Newton and Levenberg Marquardt techniques, which dynamically estimate the regularization parameter via x2 tests, implement the nonlinear x2 approach. On the cross-well tomography problem and the subsurface electrical conductivity estimation challenge, we compare parameter estimates from the nonlinear x2 approach with estimates obtained using Occams inversion and the discrepancy principle. It is demonstrated that the computationally less expensive x2 method provides parameter estimates that are comparable to those discovered using the discrepancy principle. Furthermore, compared to Occams Inversion, the x2 technique offered significantly superior parameter estimates.

Regularization methods for linear problems do not straightforwardly extend to nonlinear least squares problems. Since the nonlinear problems are solved iteratively, the methods for determining the regularization parameter generally breakdown into two approaches. In the first approach, α remains fixed throughout the nonlinear inversion process. In these methods, the inversion is done multiple times for different values of α until the solution meets some criterion. Some criteria used for evaluating the solution are the discrepancy principle [1] and Generalized Cross Validation (GCV) [3]. In the second approach, α is estimated dynamically at each iteration. In this approach, the nonlinear inverse problem is solved only once, but the optimization procedure has to be integrated with the method for estimating α. Some examples of this type of method include Occam's inversion and an implementation of GCV as proposed in [3]. The nonlinear x^2 method used this second approach and is an alternative to these methods and is developed in this article. The first criterion rarely causes problems, while the second and third criteria frequently do. Additionally, these requirements must be met regardless of how effectively an issue is mathematically described, taking into account the accuracy of a computer's computations. For instance, even if (1.3) has a continuous solution, there may be a number of comparable solutions with finite accuracy. Similarly, although though (1.3) depends on the data continuously, it also needs to be computationally stable with respect to modest data perturbations in order to produce a viable solution. A problem is said to be ill-conditioned [12] if it is unstable with respect to tiny perturbations, and the level of instability is measured using a large condition number. A issue is said to be well-conditioned if it is stable in the presence of modest perturbations and has a low condition number. So, one reason an inverse problem may not be well-posed is that it is poorly conditioned.

Methods For Regularizing Non-linear Inverse Problem

Linear X^2 Method; For the linear x^2 method, one has to consider the following If $\varepsilon \sim N(0, \sigma_\varepsilon^2 \mathrm{I})$ and $f \sim N(0, \sigma_f^2 \mathrm{I})$,

then xM is identical to the estimate found by minimizing the Tikhonov functional with L as the identity, z as the initial estimate, and with. $\alpha = \sigma_\varepsilon / \sigma_f$. Of course, many times in inverse problems, the prior covariance for f is not available. However, all is not lost. Mead in [8] suggested capitalizing on Theorem 2 to estimate α.

Theorem 2; If $F : \mathbb{R}^m \to \mathbb{R}^n$ is a linear function and the following holds;

$$d = F(x) + \varepsilon,$$

$$x = x_p + f,$$

then JM(x) from (3.7) at its minimum value, i.e. $J_M(x_M)$, follows a x^2 distribution with n degrees of freedom.

Proof; Since, $F : \mathbb{R}^m \to \mathbb{R}^n$ at its minimum value, i.e. $J_M(x_M)$, follow a x^2 distribution with n degree of freedom. Also, $x \in \mathbb{R}^m$, $d \in \mathbb{R}^n$, C_ε has dimension n x n and C_f has dimension m x m. Hence,

$$\mathcal{J}_M(x) = \left\| \begin{matrix} C_\varepsilon^{-1/2}(Ax - d) \\ C_f^{-1/2}(x - x_p) \end{matrix} \right\|_2^2$$

$$x_M = \arg\min_x \mathcal{J}_M(x).$$

This can be written as an ordinary least squares problem

$$\mathcal{J}(x) = \left\| \begin{bmatrix} C_\varepsilon^{-1/2} A \\ C_f^{-1/2} \end{bmatrix} x - \begin{bmatrix} C_\varepsilon^{-1/2} d \\ C_f^{-1/2} x_p \end{bmatrix} \right\|_2^2.$$

For the sake of simplicity, let

$$A^* = \begin{bmatrix} C_\varepsilon^{-1/2} A \\ C_f^{-1/2} \end{bmatrix}, \; d^* = \begin{bmatrix} C_\varepsilon^{-1/2} d \\ C_f^{-1/2} x_p \end{bmatrix}, \; \varepsilon^* = \begin{bmatrix} C_\varepsilon^{-1/2} \varepsilon \\ C_f^{-1/2} f \end{bmatrix}.$$

Then

$$\mathcal{J}(x) = \| d^* - A^* x \|_2^2$$

$$\hat{x} = \arg\min_x \mathcal{J}(x),$$

where A^* has dimension (n + m) x m, d^* has dimension (n + m) x 1, and "£~N(0; I n+m). By Theorem 1, in the previous chapter, $\mathcal{J}(\hat{x}) \sim \chi_n^2$.

The method proposed in [8] is called the x^2 method and it says choose α such that the minimum of the functional (3.7) has a value that is consistent with its x^2 distribution. This is implemented in [8] as finding the α that makes the minimum of the functional equal to the mean of the x^2 distribution. Also, Mead showed in [8] that Theorem 2 holds asymptotically when " and if are not normally distributed, which allows this method to be applied in a more general sense.

X^2 Tests for Gauss-Newton Method

Theorem 2 has only been shown for linear problems. For nonlinear problems, the

Distribution of $\mathcal{J}_M(x_M)$ is not usually known. However, if the nonlinear inverse problem is solved with a sequence of linearizations, then it is possible to find appropriate x^2 tests at each iteration.

The Gauss-Newton method to find $x_M = \arg\min_x \mathcal{J}_M(x)$ from is as follows. First find the first and second Frechet derivative of $\mathcal{J}_M$,

$$\nabla \mathcal{J}_M(x) = J^T C_\varepsilon^{-1} (d - F(x)) - C_f^{-1}(x - x_p),$$

where J is the Jacobian of F at x. Now

$$\nabla^2 \mathcal{J}_M(x) = J^T C_\varepsilon^{-1} J + Q(x) + C_f^{-1},$$

Where, Q (T_k) is the second-order information of $\mathcal{J}_M$. Gauss-Newton method ignores this Q, so we get the following iteration:

$$x_{k+1} = x_k + \Delta x_k$$

$$\Delta x_k = -\left(J_k^T C_\varepsilon^{-1} J_k + C_f^{-1}\right)^{-1} \left(J_k^T C_\varepsilon^{-1} (d - F(x_k)) - C_f^{-1}(x_k - x_p)\right).$$

The Gauss-Newton method can be converted to a sequence of linear OLS problems with the following manipulations

$$x_{k+1} = x_k + \left(J_k^T C_\varepsilon^{-1} J_k + C_f^{-1}\right)^{-1} \left(J_k^T C_\varepsilon^{-1} r_k - C_f^{-1}(x - x_p)\right),$$

where rk = d - F(xk). Now multiplying both sides with $\left(J_k^T C_\varepsilon^{-1} J_k + C_f^{-1}\right)$ gives

$$\left(J_k^T C_\varepsilon^{-1} J_k + C_f^{-1}\right) x_{k+1} = \left(J_k^T C_\varepsilon^{-1} J_k + C_f^{-1}\right) x_k + \left(J_k^T C_\varepsilon^{-1} r_k - C_f^{-1}(x_k - x_p)\right).$$

$C_f^{-1} x_k$ subtracts out and gives

$$\left(J_k^T C_\varepsilon^{-1} J_k + C_f^{-1}\right) x_{k+1} = \left(J_k^T C_\varepsilon^{-1} J_k\right) x_k + \left(J_k^T C_\varepsilon^{-1} r_k + C_f^{-1} x_p\right).$$

Further via simplification

$$\tilde{d}_k = J_k x_{k+1} + \varepsilon_k,$$

$$x_{k+1} = x_p + f$$

Where, $\varepsilon_k = \varepsilon + \nu_k$ with Cov (E_k) = C_{EK} represents error introduced by the linearization i.e. $\nu_k = F(x) - F(x_k - J_k(x - x_k))$. The following theorem gives x^2 distribution for the Gauss-Newton functional Jk(x) under several assumptions.

Theorem 3; If nonlinear error is zero, and the following are true:

$$\tilde{\mathcal{J}}_k(x) = \|\tilde{d}_k - J_k x\|_{C_\varepsilon^{-1}}^2 + \|x - x_p\|_{C_f^{-1}}^2, \; \hat{x}_{k+1} = \arg\min_x \tilde{\mathcal{J}}_k(x),$$

$$d_k = J_k x_{k+1} + \varepsilon_k \qquad \varepsilon_k \sim N(0, C_{\varepsilon_k})$$

$$x_{k+1} = x_p + f \qquad f \sim N(0, C_f)|$$

Then

$$\tilde{\mathcal{J}}_k(\hat{x}_{k+1}) \sim \chi_n^2.$$

Proof; This follows trivially from Theorem 2.

If the nonlinear error is zero, then the problem is likely linear. However, this theorem can still be used to develop the x^2 test for nonlinear problems by making the assumption that $C_{\varepsilon_k} \approx C_\varepsilon$. This approximation will get better as the iterations gets closer to the solution and the nonlinear error is reduced. Under this assumption, the x^2method can be applied at each iterations to achieve increasingly better estimates for C_f^{-1}. In the next chapter, we show that this assumption works well for two inverse problems given. Now we consider a more general case where L is used as in (3.1). It is not difficult to see that in a similar way we can minimize

$$\mathcal{J}_M(x) = \| d - F(x) \|_{C_\varepsilon^{-1}}^2 + \| Lx - z \|_{C_f^{-1}}^2$$

with the sequence of linear OLS problems:

$$\tilde{\mathcal{J}}_k(x) = \|\tilde{d}_k - J_k x\|_{C_\varepsilon^{-1}}^2 + \|Lx - z\|_{C_f^{-1}}^2.$$

Often when L is chosen to represent a derivative operator, it is not a square matrix. In this case, the x^2 distribution of $\tilde{\mathcal{J}}_k(x)$ has different degrees of freedom given in Theorem 4.

Theorem 4; If

$$\tilde{\mathcal{J}}_k(x) = \|\tilde{d}_k - J_k x\|_{C_\varepsilon^{-1}}^2 + \|Lx - z\|_{C_f^{-1}}^2, \; where \; L : \mathbb{R}^m \to \mathbb{R}^q$$

is a linear operator and $\hat{x}_{k+1} = \arg\min_x \tilde{\mathcal{J}}_k(x)$, the invertibility condition holds: henece, $\mathcal{N}(J_k) \cap \mathcal{N}(L) = 0 \; where \; \mathcal{N}(A)$ is the null space of A, the nonlinear error is zero, and the following are true:

$$d_k = J_k x + \varepsilon_k, \qquad \varepsilon_k \sim N(0, C_{\varepsilon_k})$$

$$Lx = z + f \qquad f \sim N(0, C_f)$$

$$Then, \tilde{\mathcal{J}}_k(\hat{x}) \sim \chi_{n-m+q}^2.$$

Proof;

$L : \mathbb{R}^m \to \mathbb{R}^q$ is a linear operator and $x \in \mathbb{R}^m, z \in \mathbb{R}^q \; d \in \mathbb{R}^n, C_\varepsilon$ has dimension n x n, and Cf has dimension q x q. Rewrite as an ordinary least squares problem:

$$\mathcal{J}(x) = \left\| \begin{bmatrix} C_\varepsilon^{-1/2} J_k \\ C_f^{-1/2} L \end{bmatrix} x - \begin{bmatrix} C_\varepsilon^{-1/2} \tilde{d}_k \\ C_f^{-1/2} z \end{bmatrix} \right\|_2^2.$$

For sake of simplicity, let

$$A^* = \begin{bmatrix} C_\varepsilon^{-1/2} J_k \\ C_f^{-1/2} L \end{bmatrix}, \; d^* = \begin{bmatrix} C_\varepsilon^{-1/2} \tilde{d}_k \\ C_f^{-1/2} z \end{bmatrix}, \; \varepsilon^* = \begin{bmatrix} C_\varepsilon^{-1/2} \varepsilon \\ C_f^{-1/2} f \end{bmatrix}.$$

The least squares problem can be written as:

$$\tilde{\mathcal{J}}_k(x) = \| d^* - A^* x \|_2^2,$$

$$\hat{x} = \arg\min_x \mathcal{J}(x),$$

Where A* has dimension (n+q)xm, d* has dimension

By Theorem

$$\tilde{\mathcal{J}}_k(\hat{x}) \sim \chi_{n-m+q}^2.$$

Nonlinear X^2 Method

Theorems 3 and 4, we derived approximate x^2 distributions for the regularized Gauss-Newton functional J_k at $\hat{x}_{k+1}$, i.e. $\tilde{\mathcal{J}}_k(\hat{x}_{k+1})$. we found an approximate x^2 distribution for $\tilde{\mathcal{J}}_k(x_{k+1}^{LM})$. In keeping with the approach proposed by Mead in [8] for the linear x^2 method, we suggest using these x^2 distributions to estimate the regularization parameter. However, when solving real problems, only one sample of Jek is available because there is only one realization of the error "in the data. Therefore, the best we can do is to use a single characteristic of the distribution to find the regularization parameter. However, for an x^2 distribution the median is approximately equal to the mean. This implies that if a perfectly weighted $\mathcal{J}_k(x_{k+1})$ is sampled multiple times, about half of these samples will be greater than the mean. If we estimate the regularization parameter such that $\mathcal{J}_k(x_{k+1})$ is always equal to the mean, then about half of the time the regularization parameter will have to be made smaller to compensate for the realization of data error that makes a perfectly weighted $\mathcal{J}_k(x_{k+1})$ larger than the mean. This approach is implemented in Algorithm 1, which uses the Gauss-Newton method to solve the nonlinear inverse problem and dynamically estimates the regularization parameter at each Gauss-Newton iteration.

Inversion Results

In some ways, this inverse problem is more difficult than the cross-well tomography problem. The Gauss-Newton step doesn't always lead to a reduction in the nonlinear cost function and it is not always possible to and a regularizing parameter for which the solution satisfies the discrepancy principle. In the authors used an implementation of the Levenberg-Marquardt algorithm to minimize the un-regularized least squares problem to demonstrate the ill posedness of this problem. This solution, plotted, is wildly oscillating, has extreme values and is not even close to being a physically possible solution. However, this isn't evident from just looking at the data misfit as this solution actually fits the data quite well. The inverse problem was also solved in using Occam's Inversion method. Occam's Inversion is given as the following algorithm. Occam's inversion was able to and good solutions for some realization of ", such as the solution plotted, the results indicate that sometimes it found poor estimates. Conversely, the large value of σ for Occam's inversion indicates that these solutions were not consistent with each other. Since both methods estimate the regularization parameter dynamically, the computational cost should be about the same and both methods took about the same speed in terms of wall-clock time.

CONCLUSION

We presented a method regularizing nonlinear inverse problems that we call the nonlinear x^2 method. This approach uses statistical information about the data to determine the proper level of regularization and is an extension of the linear x^2 method proposed by Mead in [8]. The x^2 tests used in the linear x^2 method were extended to nonlinear problems in Section 3.2. The x^2 method was extended to nonlinear problems using the Gauss-Newton method and the Levenberg-Marquardt method in Algorithms 1 and 2, respectively. We gave numerical results in Sections 4.1.1 and 4.2.1 illustrating the statistical theory developed in Chapter 3 and demonstrated that it was valid for two complex nonlinear problems. Two new algorithms were implemented on two nonlinear problems from [1] and compared against several existing methods for nonlinear regularization. It was shown that Algorithm 1 provided parameter estimates that were of similar accuracy as the discrepancy principle in a nonlinear cross-well tomography problem from [1]. In a subsurface electrical conductivity problem from [1], Algorithm 2 proved to be more robust than Occam's inversion, providing parameter estimates without the use of a smoothing operator.

Algorithm 2 also provided much better estimates than Occam's inversion on average when the smoothing operator was used. The high computational cost of the first forward problems should be considered and this is where the x^2 method prevails. The discrepancy principle solves the nonlinear inverse problem several times for different regularization parameters and thus it requires more forward model evaluations, making it computationally expensive. The nonlinear x^2 method is cheaper because it only solves the inverse problem once and dynamically updates the regularization parameter. We conclude that the nonlinear x^2 method is an attractive alternative to the discrepancy principle and Occam's inversion. However, it does share a disadvantage with these methods in that they all require the covariance of the data to be known. If an estimate of the data covariance is not known, then the nonlinear x^2 method will not be appropriate for solving such a problem. Future work includes estimating more complex covariance matrices for the parameter estimates. In [9], Mead shows that it is possible to use multiple x^2 tests to estimate such a covariance and it seems likely that this could also be extended to solving nonlinear problems.

REFERENCES

1. Richard C. Aster, Brian Borchers, and Cli_ord H. Thurber. Parameter Estimation and Inverse Problems. Elsevier Academic Press, 2005.
2. Philip E. Gill, Walter Murray, and Margaret H. Wright. Practial Optimization. Academic Press, 1981.
3. Eldad Haber and Douglas Oldenburg. A GCV based method for nonlinear illposed problems. Computational Geosciences, 4(1):41{63, 2000.
4. Per Christian Hansen. Discrete Inverse Problems, Insight and Algorithms. SIAM,2010.
5. Corwin D L; Lesch S M; et al Hendrickx, J M H; Borchers. Inversion of soil conductivity profiles from electromagnetic induction measurements: Theory and experimental verification. Soil Science Society of America Journal, 66(3):673; 685, 2002.
6. Enting Ian G. Inverse Problems in Atmospheric Constituent Transport. Cambridge University Press, 2002.
7. John M. Lewis, S. Lakshmivarahan, and Sudarshan Dhall. Dynamic Data Assimilation. Cambridge University Press, 2006.
8. Jodi L Mead. Parameter estimation: A new approach to weighting a priori information. Journal of Inverse and Ill-Posed Problems, 16(2):175{194, 2008.
9. Jodi L Mead;. Discontinuous parameter estimates with least squares estimators. Applied Mathematics and Computation, In Revision, 2011
10. Jodi L Mead and Rosemary A Renaut. A Newton root-_ending algorithm for estimating the regularization parameter for solving ill-conditioned least squares problems. Inverse Methods, 2009.
11. Cleve Moler. Numerical Computing with Matlab. SIAM, 2006.
12. Sybil P Parker. McGraw-Hill dictionary of scientific and technical terms. New York : McGraw-Hill, 1994.
13. G. A. F. Seber and C. J. Wild. Nonlinear Regression. Wiley, 2003.
14. George A.F. Seber and Alan J. Lee. Linear Regression Analysis. John Wiley and Sons, Inc, 2003.
15. Albert Tarantola. Inverse Problem Theory and Methods for Model Parameter Estimation. SIAM, 2005.