Research on Numerically Solving the Inverse Problem Based on L1 Norm

Guo Yanan¹,², Cao Xiaoqun²*, Liu Bainian², Peng Kecheng¹,², Wang Guangjie³, Gao Mei¹,²

¹College of Computer, National University of Defense Technology, Changsha Hunan 410073, China
²College of Meteorology and Oceanography, National University of Defense Technology, Changsha Hunan 410073, China
³Chinese People's Liberation Army 61741, Beijing 100000, China
*Corresponding author’s e-mail: caoxiaoqun@nudt.edu.cn

Abstract. In this paper, the numerical solution method based on L1 norm for inverse problem is studied. First, according to the theory of the L1 norm and the characteristics of the problem to be solved, a cost function is constructed. Further, for complex parameter estimation problems and derivative discontinuities, the regularization method is used to construct the cost function and the Huber function is used in the derivative discontinuity. Finally, the problem is solved numerically by the semi-smooth Newton method. The experimental results show that the method based on L1 norm is an effective method under the interference of non-Gaussian noises. For the parameter estimation of complex models, the results based on the L1 norm method are very closer to the real parameters when there is impulse noise. For parameter estimation under non-Gaussian noise, the L1 norm estimation method has significant advantages over the L2 norm method.

1. Introduction
There are a large number of physical model parameter estimation problems in the fields of geophysical sciences such as atmosphere, ocean, and space [¹,²]. Although the parameter estimation method based on the L2 norm has been widely used in solving the above problems, there are still major limitations. For example, in the data assimilation of atmospheric observation data, researchers have found that when severe weather changes, observation data often cannot be effectively used, and problems such as poor prediction results often occur. The main reason is that in the data assimilation, it is usually assumed that the observation error obeys the Gaussian distribution, that is, the cost function in the form of L2 norm; and the actual observations (such as satellite remote sensing data) usually contain significant non-Gaussian errors, which leads to one of the important reasons for low data utilization and poor assimilation effect. In recent years, a large number of studies by scientists have shown that the L1 norm has many advantages that the L2 norm does not have. For example, when the observation data is affected by gross errors, the method of minimizing the L1 norm is more stable and reliable, and has better anti-interference performance [³,⁴]. Because the cost function in the form of L1 norm has non-differentiable computational problems during minimization, its application range is far less extensive than that of L2 norm. However, with the continuous development of mathematical calculation methods and calculation conditions, the parameter estimation method based on the L1
norm has attracted more and more attention from researchers. At present, many solving algorithms have been proposed, such as the alternating direction multiplier method (ADMM), the Homotopy method \([5-7]\), which greatly promoted the application of L1 norm optimization method, and achieved good results in many fields.

2. Parameter estimation method based on L1 norm

2.1. The concept and properties of L1 norm

The L1 norm is also called the Manhattan distance and is defined as the sum of the absolute values of the elements in the vector. At present, the L1 norm has attracted more and more attention from researchers due to its sparseness and other properties. The first is in the fields of machine learning and pattern recognition \([8]\). When constructing a model, if only a few features contribute to the model and most of the features are not contributing or the contribution is small, researchers often use L1 regularization. L1 norm can achieve sparseness to extract main features, so L1 norm is also called the sparse rule operator. In addition, the L1 norm sparsification feature is also widely used in the fields of compressed sensing, image denoising, and signal processing. Taking oil and gas exploration as an example: geophysicists use geological echo signals to invert geological structures. During the inversion process, the L1 norm sparsity can be used to extract as much information as possible from the limited echo signals. In addition, a large number of studies have found that cost functions based on the L1 norm form can effectively resist the interference of abnormal observation data. For example, in the multiple regression problem, when there is abnormal observation interference, the experimental result of the least square method based on the L1 norm which reflects the good resistance of noisy is obviously better than that the least square method based on the L2 norm.

2.2. Problem definition

In many scientific and engineering problems, it is often necessary to establish a physical model to describe complex physical movements or phenomena, which usually includes two important parts: model structure and model parameters. In most cases, model parameters cannot be directly observed or detected, but need to be estimated based on the observation information of the model state quantities. Such parameter solving problems are often proposed in the form of inverse problems\([9,10]\):

\[ M(v) = y^\delta \]

Where \(v\) is the target parameter and \(y^\delta\) is the observation data with noise. \(M\) is an integral operator or partial differential operator. \(M\) maps the parameter \(v\) to the function value \(y\). Calculating the value of the parameter \(v\) based on the observation data \(y^\delta\) has important applications in the data assimilation of atmosphere, geophysical sciences, and image processing.

2.3. Parameter estimation algorithm based on L1 norm

The above inverse problem is different from the forward problem. It calculates the target parameters in the model through observation. Because of the errors in the observation information and the nonlinearity of the system itself, the inverse problem is often ill-posed. Therefore, methods such as Tikhonov regularization are widely used to solve the problem. The main idea of regularization is to use the loss term and regularization term in the form of L2 norm to construct a cost function, and then obtain the optimal model parameter value through minimization. Considering that the L1 norm has many advantages such as robustness, this article attempts to use the L1 norm to estimate parameters in complex physical models, and constructs a new cost function as follows\([11,12]\):

\[ J(v, \alpha) = \|M(v) - y^\delta\|_t + \frac{\alpha}{2}\|v\|_t^2 \]

If \(M\) is strictly differentiable, the generalized gradient sum rule and chain rule are used to convert the problem into a problem of solving the equation:
\[ \nabla_v J = \alpha \| v \|_{L_2} + M'(v) \eta = 0 \]  
(3)

Where \( \eta \) is determined by Huber function:

\[
\eta = \begin{cases} 
1 & \text{if } (M(v) - y^\delta) > \beta \\
-1 & \text{if } (M(v) - y^\delta) < -\beta \\
(M(v) - y^\delta) \beta^{-1} & \text{if } |M(v) - y^\delta| \leq \beta 
\end{cases}
\]
(4)

In the above formula, \( \beta \) is the adjustment factor of the Huber function. The value of \( \beta \) has an important influence on the effect of the L1 norm estimation. If the value of \( \beta \) is too large, the effect of the L1 norm estimation is not obvious. If the value of \( \beta \) is too small, the convergence radius of the semi-smooth Newton method will shrink to 0 as \( \beta \) decreases. At the same time, the \( \beta \) value is limited in regularization conditions. When the semi-smooth Newton method is used to solve formula (2), the value of \( \alpha \) depends on the magnitude of the noise. But in general, the magnitude of the noise is unknown. Therefore, in the parameter estimation process, the regularization term coefficient \( \alpha \) and the adjustment factor \( \beta \) of the Huber function need to be iteratively updated at the same time to obtain the optimal \( \alpha \) and \( \beta \) values.

The specific steps of the parameter estimation algorithm based on the L1 norm are explained as follows:

Step 1: Transform the cost function and introduce Huber function for non-differentiable points. A new equation to be solved is obtained by gradient operation, and the initial guess value \( v_0 \) and initial regularization coefficient \( \alpha_0 \) of the equation are given.

Step 2: According to the semi-smooth Newton method, give the initial value \( \beta_0 \) of the adjustment factor in the Huber function, and then start to solve iteratively until the number of iterations is greater than the maximum number or the gradient norm is less than the set threshold. Then use the automatic parameter selection method to reduce the value of the adjustment factor \( \beta \) in the Huber function, update the objective equation and repeat the above-mentioned solution process. If the number of iterations exceeds the maximum number of iterations, or \( \beta \) is less than the preset minimum, the optimization result is taken from the previous result, and the loop exits to the next step.

Step 3: When the number of updates of \( \alpha \) is less than the maximum number of updates, update the regularization coefficient \( \alpha \). Specifically, update \( \alpha \) by \( \alpha_{k+1} = (\mu - 1) \| M(v_{\alpha_k}) - y^\delta \|_{L_2} / J'(\alpha) \), where \( J'(\alpha) = \frac{1}{2} \| v_{\alpha_k} \|_{L_2}^2 \). Set \( \mu > 1 \) and \( \mu \) fixed in the experiment. The value of \( \mu \) depends on the smoothness of the data. If the difference between the \( \alpha \) update result and the previous result is small (the change value is less than the threshold value), then the loop goes out to the next step, otherwise it moves to the second step.

Step 4: Output the solution result and the final regularization coefficient \( \alpha \) and the value of the adjustment factor \( \beta \) in the Huber function.

3. Numerical experiment results and analysis
In order to verify the applicability of the parameter estimation algorithm based on the L1 norm in complex systems, this section estimates the unknown coefficients in the variable coefficient Helmholtz equation. Helmholtz equation is a kind of elliptic partial differential equation, which mainly describes a kind of physical phenomenon of harmonic propagation. It has a wide range of applications in many fields, such as electromagnetic wave propagation, acoustic wave scattering, seismic wave propagation,
and ocean wave propagation \cite{13,14}. Therefore, studying the coefficient estimation problem of the Helmholtz equation with variable coefficients is very important to deal with many physical problems in the fields of electromagnetics, acoustics, and so on. First give the equation as follows:

$$\frac{\partial^2 \omega}{\partial x^2} + \varphi(x) \omega = C$$  \hspace{1cm} (5)

Set the real coefficients \( \varphi(x) = 2 - \cos(x^3) + |\sin(x)| \) and \( C = 1 \) in the experiment. First, the interval \([-3,3]\) is equally divided into \( N = 1000 \) parts, and the partial differential equation is discretized using the finite element method. The solution interval includes \( N + 1 \) nodes and \( N \) units. After discretization, the problem of estimating unknown parameters of partial differential equations is transformed into solving equations \( M(\varphi(x_i)) = y_i, i = 1, L, N + 1 \). Use the following methods to construct observations with impulsive noise:

$$y^\varphi = \begin{cases} y + \tau \psi & \text{with probability } \gamma \\ y & \text{otherwise} \end{cases}$$  \hspace{1cm} (6)

Where \( y \) is the real data and \( \psi \) is a random variable that obeys the standard normal distribution. The value of \( \tau \) depends on how much the observation data is affected by noise, where \( \tau \) is set to \( \|y\|_L \). \( \gamma \in (0, 1) \) is the proportion of abnormal data. The value of \( \gamma \) in the experiment is 0.3. In the following, the cost function is constructed based on the L1 norm and L2 norm respectively, and regularization method is introduced to calculate the coefficient \( \varphi(x) \):

$$J(\varphi) = \|M(\varphi) - y^\varphi\|_L + \frac{\alpha}{2} \|\varphi\|_L^2$$  \hspace{1cm} (7)

$$J(\varphi) = \|M(\varphi) - y^\varphi\|^2_L + \frac{\alpha}{2} \|\varphi\|^2_L$$  \hspace{1cm} (8)

When solving (7), the solution algorithm in Section 2.3 is adopted. The problem is first transformed, and then the initial iteration value \( v_0 \) and the initial regularization coefficient \( \alpha_0 \) are given. The fixed point iteration method is used to update \( \alpha \), and the semi-smooth Newton method is used for iterative calculation. When solving (8), perform gradient calculation to obtain the optimization target firstly: \( 2M'(\varphi)\|M(\varphi) - y^\varphi\|_L + \alpha \|\varphi\|_L = 0 \), where \( \alpha \) is 0.001, use Newton’s method to solve the equation. In each iteration, calculations are performed using the biconjugate gradient stabilized method. The following experimental results were obtained:
The solid black line in Figure 1 is the true coefficient value, the dotted line is the estimated value based on the L2 norm method, and the plus sign is the estimated value based on the L1 norm method. By comparison, it can be seen that the results obtained by the method based on the L1 norm are closer to the true value, while the results obtained by the method based on the L2 norm estimation are greatly affected by the impulse noise, and there are obvious fluctuations. The advantages and disadvantages of the two methods are analyzed by comparing the estimated error sizes. Taking the difference between the estimated result and the true value on the interval $[-3, 3]$ as the standard. By calculation, the estimation error $err_2 = 15.1113$ based on the L2 norm method, and the estimation error $err_1 = 0.1796$ based on the L1 norm method. Obviously, the results obtained based on the L1 norm method have smaller errors and better results. This result indicates that in the parameter estimation problem, the method based on the L1 norm is superior to the method based on the L2 norm when the error of the observation data follows a non-Gaussian distribution. It also shows that for data with different error distributions, it is important to choose the appropriate data analysis method.

In the experiment, an automatic update method of regularization coefficients is used. The following specifically analyzes the update method of $\alpha$ and its convergence process:

Figure 2 shows the variation of the estimation error $\varepsilon$ with different iteration times for different regularization coefficients, where $\varepsilon = \|v - v_\epsilon\|_{L^2}$, $v$ represent the estimation result, and $v_\epsilon$ represents
the true coefficient. When iterating with fixed regularization coefficient $\alpha$, the initial value $\beta_0$ of the adjustment factor in the Huber function is taken as 1.0. After each iteration, $\beta$ is reduced to 0.5 times the previous result, so Figure 2 essentially reflects that when the regularization coefficient is determined, the estimation error varies with $\beta$ changes. It can be seen from Figure 2 that when the regularization coefficient is determined, the estimation error gradually converges to a stable value as $\beta$ decreases. In Figure 2 (a), $\alpha = 1$, the estimation error has stably converged to around 52.2 at the third iteration. In Figure 2 (b), $\alpha = 0.0644$, the estimation error has stably converged to around 25.0 when the iteration is performed for the sixth time. In Figure 2 (c), $\alpha = 0.0030$, the estimation error converges to around 0.2 at the 20th iteration. In Figure 2 (d), $\alpha = 0.0015$, the estimation error also converged to around 0.2 at the 20th iteration. In the experiment, because the difference between the fourth update result and the third result of $\alpha$ is less than the given threshold, the loop breaks out. It can be concluded from the above results that the estimation error decreases rapidly with the update, indicating that the automatic update method is very effective for reducing the estimation error.

4. Conclusion

Aiming at the problem that the parameter estimation method based on the L2 norm is not effective in the case of non-Gaussian errors, a parameter estimation method based on the L1 norm is proposed. The minimization algorithm research and numerical experiments are carried out. In this paper, experiments are performed to estimate unknown coefficients of partial differential equations. Aiming at the problem of impulse noise interference, a loss function based on the L1 norm is introduced. Huber function is introduced at the discontinuity of the derivative. Then, the new cost function is solved by using the semi-smooth Newton method. By using the method based on the L1 norm, better experimental results are obtained. The experimental results show that the method based on L1 norm is more robust than the method based on L2 norm under non-Gaussian noise interference.

In addition, the regularization method has an important effect on the parameter estimation effect. In this paper, an L2 regular term is added to the L1 norm loss term to prevent overfitting problems. Taking into account the sparseness of the L1 norm, L1 regular terms will be added to the cost function in the future to study the impact of sparseness on the parameter estimation problem.

Acknowledgement

National key research and development program of China (Grant of No. 2018YFC1506704)

References

[1] Schillings, C., & Stuart, A. M. (2017). Analysis of the ensemble kalman filter for inverse problems. SIAM Journal on Numerical Analysis, 55(3), 1264-1290.
[2] Yang, Y. X., & Zeng, A. M. (2009). Adaptive filtering for deformation parameter estimation in consideration of geometrical measurements and geophysical models. Science in China Series D: Earth Sciences, 52(8), 1216-1222.
[3] Yanyong Xiang, N.R. Thomson, & J.F. Sykes. (1992). Fitting a groundwater contaminant transport model by l1 and l2 parameter estimators. Advances in Water Resources, 15(5), 303–310.
[4] Seyoung Oh, Sunjoo Kwon, & Jae heon Yun.(2006) . A parameter estimation method for model analysis. Journal of Applied Mathematics & Computing, 22(1-2), 373-385.
[5] Zhu, L., Fu, C. W., Jin, Y., Wei, M., Qin, J., & Heng, P. A. (2016). Non-local sparse and low-rank regularization for structure-preserving image smoothing. Computer Graphics Forum, 35(7), 217-226.
[6] Shi, B., Pang, Z. F., & Yang, Y. F. (2012). Image restoration based on the hybrid total-variation-type model. Abstract and Applied Analysis, 2012(13-16), 97-112.
[7] Wang, T., Yang, Z., & Wu, Y. (2010). A maximum entropy function method for signal reconstruction in compressed sensing. Mathematica Applicata, 23(2), 345-352.
[8] Wang, Y., Li, D., Du, Y., & Pan, Z. (2015). Anomaly detection in traffic using L1-norm minimization extreme learning machine. (Vol.149, pp.415-425).

[9] Aster, R. C., Thurber, C. H., & Borchers, B. (2012). Parameter estimation and inverse problems. International Geophysics, 67, 411-421.

[10] Fu, H., Han, B., & Liu, H. (2012). A wavelet multiscale iterative regularization method for the parameter estimation problems of partial differential equations. Neurocomputing, 104, 138-145.

[11] Christian Clason, Bangti Jin. (2012). A Semismooth Newton Method for Nonlinear Parameter Identification Problems with Impulsive Noise. SIAM Journal on Imaging Sciences, 5:2, 505-536

[12] Christian Clason, Tuomo Valkonen. (2017) Primal-Dual Extragradient Methods for Nonlinear Nonsmooth PDE-Constrained Optimization. SIAM Journal on Optimization 27:3, 1314-1339.

[13] Chen, Y., & Rokhlin, V. (1992). On the inverse scattering problem for the helmholtz equation in one dimension. Inverse Problems, 8(3), 365-391.

[14] Barceló, Juan Antonio, Fanelli, L., Ruiz, A., & Vilela, M. (2013). A priori estimates for the helmholtz equation with electromagnetic potentials in exterior domains. Proceedings of the Royal Society of Edinburgh: Section A Mathematics, 143(01), 1-19.