

Article



Improved Generalized Cross-Validation and Unbiased Predictive Risk Estimator Methods Using the RGSVD: Application to Inversion of Potential Field Data

Yuan Fang, Jun Wang *, Xiaohong Meng * and Hanhan Tang

* Correspondence: wangj@cugb.edu.cn (J.W.); mxh@cugb.edu.cn (X.M.); Tel.: +86-135-8199-1194 (J.W.); +86-136-9129-3267 (X.M.)

Abstract: The inversion of potential field data has widely utilized the generalized cross-validation (GCV) and the unbiased predictive risk estimator (UPRE) methods to determine the regularization parameter. However, these two methods are time-consuming and it is difficult for them to determine the optimal linear search range including the optimal regularization. To solve these problems, this article improves the GCV and UPRE methods using the RGSVD (randomized generalized singular value decomposition) algorithm. The improved methods first use the randomized algorithm to compute an approximate generalized singular value decomposition (GSVD) with less computational time. Then, the optimal linear search range is determined based on the generalized singular values. Finally, the GCV and the UPRE functions are efficiently computed on the basis of the results from the RGSVD algorithm. In this way, the GCV and UPRE methods using the RGSVD algorithm are able to determine the optimal regularization parameter fast and effectively. One comparative test shows the effectiveness and efficiency of the GCV and the UPRE methods using the RGSVD algorithm.

Keywords: generalized cross-validation; randomized generalized singular value decomposition; regularization parameter; unbiased predictive risk estimator

1. Introduction

Various geological and geophysical problems can be solved by the inversion of potential field data [1,2]. Generally, the inversion of potential field data is an ill-posed problem, which means that this inversion is usually non-unique and unstable [3]. Tikhonov regularization [4] can solve ill-posed problems. Estimating an optimal regularization parameter is very significant for the Tikhonov regularization, e.g., [5]. In the literature, many methods have been introduced to determine an optimal regularization parameter, such as the Morozov discrepancy principle (MDP) method [6], the generalized cross-validation (GCV) method [7], the L-curve (LC) method [8], and the unbiased predictive risk estimator (UPRE) method [9].

Here, we focus on the GCV and the UPRE methods. The conjugate gradient (CG) method and generalized singular value decomposition (GSVD) can be used to compute the GCV and the UPRE functions in a linear search range to find the optimal regularization parameter. However, when using the CG method, the process is time-consuming, and it is difficult to determine the optimal linear search range including the optimal regularization parameter. When using the GSVD, the optimal linear search range can be determined by analyzing the spectrum (generalized singular values) of the kernel matrix [9], but computational costs and memory requirements limit the application of this method.

In this paper, the RGSVD [10] algorithm was adopted in the GCV and the UPRE methods for determination of the optimal regularization parameter. The RGSVD algorithm uses a randomized algorithm to compute an approximation of the GSVD with less memory requirements and computing time [10], with which the optimal linear search range can be



Citation: Fang, Y.; Wang, J.; Meng, X.; Tang, H. Improved Generalized Cross-Validation and Unbiased Predictive Risk Estimator Methods Using the RGSVD: Application to Inversion of Potential Field Data. *Appl. Sci.* **2021**, *11*, 6326. https:// doi.org/10.3390/app11146326

Academic Editor: Forte Emanuele

Received: 14 May 2021 Accepted: 6 July 2021 Published: 8 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).

School of Geophysics and Information Technology, China University of Geosciences, Beijing 100083, China; 3010180014@cugb.edu.cn (Y.F.); 3010180015@cugb.edu.cn (H.T.)

determined based on the generalized singular values. The result from the RGSVD facilitates efficient computation of the GCV and UPRE functions. Therefore, the regularization parameter can be determined fast and effectively by the improved GCV and UPRE methods using the RGSVD algorithm. One comparative test demonstrated the performances of the GCV and the UPRE methods using the RGSVD algorithm.

2. Inversion Methodology

Generally, the model domain is discretized into many cells, whose physical properties are the model parameters. According to the linear relationship between model parameters and measured data, the inverse problem of potential field data has the matrix form as

d

$$_{obs} = \mathbf{Gm},\tag{1}$$

where $\mathbf{d_{obs}} \in \mathbb{R}^n$ represents the measured data vector, $\mathbf{m} \in \mathbb{R}^m$ is the unknown parameters, and $\mathbf{G} \in \mathbb{R}^{n \times m}$ represents the kernel matrix. *n* represents the number of measured data, and *m* represents the number of the unknown parameters. In general, *m* is much larger than *n*. Therefore, the inversion of potential field data belongs to an under-determined problem. Through the singular value decomposition (SVD) of the matrix \mathbf{G} or the least squares solution, Equation (1) can be used to calculate the inversion result directly. However, in this way, the inversion result may not conform to reality. These ill-posed problems can be solved by introducing a regularization term [4]. Then, the objective function of this inversion has the form

$$\phi = \|\mathbf{W}_{\mathbf{d}}(\mathbf{Gm} - \mathbf{d}_{\mathbf{obs}})\|_{2}^{2} + \mu^{2} \|\mathbf{W}_{\mathbf{m}}\mathbf{Zm}\|_{2}^{2},$$
(2)

In Equation (2), $\|\mathbf{W}_{\mathbf{d}}(\mathbf{Gm} - \mathbf{d}_{\mathbf{obs}})\|_2^2$ is the data misfit, $\|\mathbf{W}_{\mathbf{m}}\mathbf{Zm}\|_2^2$ is the regularization term, and μ is the regularization parameter, which is used to balance these two terms. In the data misfit, $\mathbf{W}_{\mathbf{d}}$ represents a data weighting matrix. In the regularization term, $\mathbf{W}_{\mathbf{m}} \in \mathbb{R}^{4m \times m}$ represents a smooth constraint matrix and its matrix form is $\mathbf{W}_{\mathbf{m}} = [\mathbf{W}_{\mathbf{s}}; \mathbf{W}_{\mathbf{x}}; \mathbf{W}_{\mathbf{y}}; \mathbf{W}_{\mathbf{z}}]$, where $\mathbf{W}_{\mathbf{s}} \in \mathbb{R}^{m \times m}$, $\mathbf{W}_{\mathbf{x}} \in \mathbb{R}^{m \times m}$, $\mathbf{W}_{\mathbf{y}} \in \mathbb{R}^{m \times m}$, and $\mathbf{W}_{\mathbf{z}} \in \mathbb{R}^{m \times m}$ are different component matrices, respectively [11,12]; and \mathbf{Z} represents a depth-weighting matrix [11,12].

Generally, the physical bound constraint is incorporated into this inversion for obtaining a geologically plausible inversion result. Then, the inverse problem of potential field data becomes the following constrained minimization problem

minimize :
$$\phi = \|\mathbf{W}_{\mathbf{d}}(\mathbf{Gm} - \mathbf{d}_{\mathbf{obs}})\|_{2}^{2} + \mu^{2} \|\mathbf{W}_{\mathbf{m}}\mathbf{Zm}\|_{2}^{2}$$
, (3)
subject to : $\mathbf{m}_{\min} \le \mathbf{m} \le \mathbf{m}_{\max}$

where \mathbf{m}_{min} and \mathbf{m}_{max} are vectors consisting of the lower and upper physical bounds on the unknown model values. Because the matrix \mathbf{Z} is a diagonal matrix, the following transformations can be performed easily:

$$\mathbf{h} = \mathbf{W}_{\mathbf{d}} \mathbf{G} \mathbf{Z}^{-1}, \mathbf{r} = \mathbf{W}_{\mathbf{d}} \mathbf{d}_{\mathbf{obs}}, \text{ and } \mathbf{y} = \mathbf{Z} \mathbf{m}.$$
(4)

Then, Equation (2) is rewritten as

$$\phi = \|\mathbf{h}\mathbf{y} - \mathbf{r}\|_{2}^{2} + \mu^{2} \|\mathbf{W}_{\mathbf{m}}\mathbf{y}\|_{2}^{2}.$$
(5)

The constrained minimization problem (3) becomes the following new problem

minimize:
$$\phi = \|\mathbf{h}\mathbf{y} - \mathbf{r}\|_2^2 + \mu^2 \|\mathbf{W}_{\mathbf{m}}\mathbf{y}\|_2^2$$
,
subjectto: $\mathbf{y}_{\min} \le \mathbf{y} \le \mathbf{y}_{\max}$ (6)

where $\mathbf{y}_{min} = \mathbf{Z}\mathbf{m}_{min}$ and $\mathbf{y}_{max} = \mathbf{Z}\mathbf{m}_{max}$ are vectors of the lower and upper bounds for \mathbf{y} . The logarithmic barrier method is adopted to incorporate the physical bound constraint into the objective function [13–15]. Then, the new objective function has the form

$$\phi = \|\mathbf{h}\mathbf{y} - \mathbf{r}\|_{2}^{2} + \mu^{2} \|\mathbf{W}_{\mathbf{m}}\mathbf{y}\|_{2}^{2} - 2\lambda \sum_{i=1}^{n} [\ln(y_{i} - y_{\min,i}) + \ln(y_{\max,i} - y_{i})],$$
(7)

where $-2\lambda \sum_{i=1}^{n} [\ln(y_i - y_{\min,i}) + \ln(y_{\max,i} - y_i)]$ is the barrier function, and λ is the barrier parameter.

The Newton method [13–15] is used to solve the minimization of (7) iteratively. The regularization parameter μ is kept fixed during the iterations, and the barrier parameter λ decreases with iteration [13]. At the *k*th iteration, one step of the Newton method is applied for (7) to yield

$$\left\{ \mathbf{h}^{T}\mathbf{h} + \mu^{2}\mathbf{W}_{\mathbf{m}}^{T}\mathbf{W}_{\mathbf{m}} + \lambda^{(k)} \left[\left(\mathbf{X}^{(k)} \right)^{-2} + \left(\mathbf{Y}^{(k)} \right)^{-2} \right] \right\} \Delta \mathbf{y}^{(k)} = -\mathbf{h}^{T} \left(\mathbf{h}\mathbf{y}^{(k-1)} - \mathbf{r} \right) - \mu^{2}\mathbf{W}_{\mathbf{m}}^{T}\mathbf{W}_{\mathbf{m}} \left(\mathbf{y}^{(k-1)} - \mathbf{y}^{(0)} \right) + \lambda^{(k)} \left[\left(\mathbf{X}^{(k)} \right)^{-1} + \left(\mathbf{Y}^{(k)} \right)^{-1} \right] \mathbf{e}^{\prime}$$
(8)

where $\mathbf{X}^{(k)} = diag(\mathbf{y}^{(k-1)} - \mathbf{y}_{\min})$, $\mathbf{Y}^{(k)} = diag(\mathbf{y}_{\max} - \mathbf{y}^{(k-1)})$, $\mathbf{e} \in \mathbb{R}^{m \times 1}$ is the vector with all entries one, and $\mathbf{y}^{(0)}$ is an initial model. The solution $\Delta \mathbf{y}^{(k)}$ is the search direction at the *k*th iteration. The strategy of Li and Oldenburg [13] is used to obtain the final inversion result, and Algorithm 1 shows the detailed steps. In step 9, the solution $\Delta \mathbf{y}^{(k)}$ can be obtained by the preconditioned conjugate gradient (PCG) method, and the preconditioner $\mathbf{P}^{(k)}$ has the form

$$\mathbf{P}^{(k)} = diag\left(\sqrt{diag\left(\mathbf{A}^{(k)}\right)}\right),\tag{9}$$

where $\mathbf{A}^{(k)}$ has the form

$$\mathbf{A}^{(k)} = \mathbf{h}^T \mathbf{h} + \mu^2 \mathbf{W}_{\mathbf{m}}^T \mathbf{W}_{\mathbf{m}} + \lambda^{(k)} \left[\left(\mathbf{X}^{(k)} \right)^{-2} + \left(\mathbf{Y}^{(k)} \right)^{-2} \right].$$
(10)

Algorithm 1. Inversion of potential field data

Preparation : d_{obs} , G, W_d , m_{min} , m_{max} and K_{max} .

- 1. Calculate W_{m} , and Z.
- 2. Initialize $\mathbf{m}^{(0)} = 0.001$, and k = 0.
- 3. Calculate $\mathbf{h} = \mathbf{W}_{\mathbf{d}}\mathbf{G}\mathbf{Z}^{-1}$, $\mathbf{r} = \mathbf{W}_{\mathbf{d}}\mathbf{d}_{\mathbf{obs}}$, $\mathbf{y}^{(0)} = \mathbf{Z}\mathbf{m}^{(0)}$, $\mathbf{y}_{\mathbf{min}} = \mathbf{Z}\mathbf{m}_{\mathbf{min}}$, and $\mathbf{y}_{\mathbf{max}} = \mathbf{Z}\mathbf{m}_{\mathbf{max}}$.
- 4. Estimate the regularization parameter μ .

5. Calculate
$$\lambda^{(1)} = \frac{\|\mathbf{hy}^{(0)} - \mathbf{r}\|_2 + \mu^2 \|\mathbf{W}_{\mathbf{m}} \mathbf{y}^{(0)}\|_2}{-2\sum_{i=1}^n \left[\ln(y_i^{(0)} - y_{\min,i}) + \ln(y_{\max,i} - y_i^{(0)}) \right]}$$

- 6. while $k < K_{\max}$
- 7. k = k + 1.
- 8. Update $\mathbf{X}^{(k)} = diag(\mathbf{y}^{(k-1)} \mathbf{y}_{\min})$ and $\mathbf{Y}^{(k)} = diag(\mathbf{y}_{\max} \mathbf{y}^{(k-1)})$.
- 9. Solve Equation (8) for $\Delta \mathbf{y}^{(k)}$.
- 10. Update $\mathbf{y}^{(k)} = \mathbf{y}^{(k-1)} + \delta \beta^{(k)} \Delta \mathbf{y}^{(k)}$, where $\delta = 0.925$ and

$$\beta^{(k)} = \begin{cases} 1, \text{ if } \mathbf{y_{min}} < \mathbf{y}^{(k-1)} + \Delta \mathbf{y}^{(k)} < \mathbf{y_{max}} \\ \min \left(\min_{y_i^{(k-1)} + \Delta y_i^{(k)} < y_{\min,i}} \frac{y_i^{(k-1)} - y_{\min,i}}{\left| \Delta y_i^{(k)} \right|}, \min_{y_i^{(k-1)} + \Delta y_i^{(k)} > y_{\max,i}} \frac{y_{\max,i} - y_i^{(k-1)}}{\left| \Delta y_i^{(k)} \right|} \right), \text{ otherwise} \end{cases}$$

- 11. Exit the loop if the termination criterion is satisfied.
- 12. Update $\lambda^{(k+1)} = \left[1 \min\left(\beta^{(k)}, \delta\right)\right] \lambda^{(k)}$.
- 13. End
- 14. Output: Solution $\mathbf{m} = \mathbf{Z}^{-1} \mathbf{y}^{(k)}$.

When the number of iterations k reaches the preset maximum number of iterations K_{max} or the change of the objective function is less than 1%, the iterative process ends. The method of estimating the regularization parameter is discussed in Section 3.

3. Estimation of Regularization Parameter

When estimating the regularization parameter, we only focused on (5) without considering the barrier function. In this way, the regularization parameter is still suitable in the minimization of (7). Here, the GCV [7] and the UPRE [9] methods are used to estimate the regularization parameter.

3.1. Generalized Cross-Validation Method

The randomized trace estimation was introduced into the GCV function to solve the difficulty of a trace calculation in the GCV function [13]. Then, this function for Equation (5) has the form 112 ш

$$GCV(\mu) \approx \frac{\left\| \mathbf{r} - \mathbf{h} \left(\mathbf{h}^{T} \mathbf{h} + \mu^{2} \mathbf{W}_{\mathbf{m}}^{T} \mathbf{W}_{\mathbf{m}} \right)^{-1} \mathbf{h}^{T} \mathbf{r} \right\|_{2}^{2}}{\left\{ n - \mathbf{R}^{T} \mathbf{h} \left(\mathbf{h}^{T} \mathbf{h} + \mu^{2} \mathbf{W}_{\mathbf{m}}^{T} \mathbf{W}_{\mathbf{m}} \right)^{-1} \mathbf{h}^{T} \mathbf{R} \right\}^{2}},$$
(11)

where **R** is a random vector consisting of -1 and 1, each with a probability of 0.5. The parameter μ which minimizes (11) is the optimal regularization parameter μ_{ovt} . This parameter μ_{opt} can be determined by line search within a range for regularization parameter, and the process can be solved by using the CG method [13]. However, when using the CG method, an optimal range for regularization parameter is difficult to determine, and sometimes the optimal regularization parameter μ_{opt} is not within the selected range. Moreover, the GCV method using the CG method is time-consuming.

The GSVD of the matrix pair $[\mathbf{h}, \mathbf{W}_{\mathbf{m}}]$ can be used to compute the GCV function [5]. Based on the GSVD, the spectrum of h can be obtained. The optimal linear search range can be determined by analyzing the spectrum of h [9]. However, the calculation of GSVD is time-consuming and has a large memory requirement. When the scale of data is large, the calculation of GSVD is even impractical. Instead, the RGSVD algorithm is adopted. The RGSVD algorithm uses the randomized algorithm to provide a low-rank approximation of the GSVD [10]. Compared with the GSVD, the RGSVD can reduce computational costs and memory demands, and results with good accuracy can be obtained. Algorithm 2 shows the RGSVD algorithm [10], which is used in this study. The parameter q determines the accuracy and efficiency of the RGSVD algorithm. As the value of the parameter *q* increases, the result becomes more accurate and the computation time is longer.

Algorithm 2. Randomized generalized singular value decomposition (RGSVD) algorithm. Given $\mathbf{h} \in \mathbb{R}^{n \times m}$ ($n \le m$) and $\mathbf{W}_{\mathbf{m}} \in \mathbb{R}^{4m \times m}$, a target matrix rank q ($q \le n$), calculate an approximate GSVD of the matrix pair $[\mathbf{h}, \mathbf{W}_{\mathbf{m}}]$: $\mathbf{h} \approx \mathbf{UCX}, \mathbf{W}_{\mathbf{m}} \approx \mathbf{VSX}$ with $\mathbf{U} \in \mathbb{R}^{n \times q}$, $\mathbf{V} \in R^{4m \times q}$, $\mathbf{C} \in R^{q \times q}$, $\mathbf{S} \in R^{q \times q}$, and $\mathbf{X} \in R^{q \times m}$.

- 1. Generate a $q \times n$ Gaussian random matrix **A**.
- 2. Calculate the $q \times m$ matrix $\mathbf{Y} = \mathbf{A}\mathbf{h}$.

4.

- Calculate the $m \times q$ orthonormal matrix **Q** via QR factorization $\mathbf{Y}^T = \mathbf{Q}\mathbf{R}$. 3.
 - Form the $n \times q$ matrix $\mathbf{B}_1 = \mathbf{h}\mathbf{Q}$ and the $4m \times q$ matrix $\mathbf{B}_2 = \mathbf{W}_{\mathbf{m}}\mathbf{Q}$.
- 5. Use $[\mathbf{U}, \mathbf{V}, \mathbf{W}, \mathbf{C}, \mathbf{S}] = \text{gsvd}(\mathbf{B}_1, \mathbf{B}_2, 0)$ to calculate the economy-sized GSVD of the matrix $\begin{bmatrix} \mathbf{C} \\ \mathbf{S} \end{bmatrix} \mathbf{W}^T.$
 - $= \left[\begin{array}{cc} \mathbf{U} & \\ & \mathbf{V} \end{array} \right] \left| \right.$ pair $[\mathbf{B}_1, \mathbf{B}_2]$: $\begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{bmatrix}$
- Form the $q \times m$ matrix $\mathbf{X} = \mathbf{W}^T \mathbf{Q}^T$. 6.
- Note : $h \approx UCX$ and $W_m \approx VSX$. 7.

In the RGSVD algorithm, $\mathbf{C} = diag(c_1, c_2, \dots, c_q) \in \mathbb{R}^{q \times q}$ with

$$0 < c_1 \le c_2 \le \dots \le c_q < 1, \tag{12}$$

and $\mathbf{S} = diag(s_1, s_2, \dots, s_2) \in \mathbb{R}^{q \times q}$ with

$$s_1 > s_1 \ge s_2 \ge \dots \ge s_q > 0,$$
 (13)

 $c_i^2 + s_i^2 = 1$, $\forall i = 1 : q$, that is $\mathbf{C}^T \mathbf{C} + \mathbf{S}^T \mathbf{S} = \mathbf{I}_q$. According to the result of the RGSVD, Equation (11) has the form

$$GCV(\mu) \approx \frac{\left\| \mathbf{r} - \mathbf{h} \left[\sum_{i=1}^{q} \frac{\gamma_i^2}{\gamma_i^2 + \mu^2} \frac{\mathbf{u}_i^T \mathbf{r}}{c_i} (\mathbf{X}^{\dagger})_i \right] \right\|_2^2}{\left\{ n - \mathbf{R}^T \mathbf{h} \left[\sum_{i=1}^{q} \frac{\gamma_i^2}{\gamma_i^2 + \mu^2} \frac{\mathbf{u}_i^T \mathbf{R}}{c_i} (\mathbf{X}^{\dagger})_i \right] \right\}^2},$$
(14)

where $\gamma_i = c_i/s_i$ denotes the *i*th generalized singular value, \mathbf{u}_i represents the *i*th column of matrix \mathbf{U} , and $(\mathbf{X}^{\dagger})_i$ denotes the *i*th column of the Moore-Penrose inverse of matrix \mathbf{X} . The parameter μ_{opt} can be found between the minimum and maximum of the generalized singular value γ_i . Here, the parameter *q* is set as *n*. Therefore, the full generalized singular values can be obtained. The value of *n* is slightly large, but the computation time of the RGSVD algorithm with q = n is still much shorter than that of GSVD.

3.2. Unbiased Predictive Risk Estimator Method

The UPRE function for Equation (5) has the form [16]

$$UPRE(\mu) = \left\| \mathbf{r} - \mathbf{h} \left(\mathbf{h}^T \mathbf{h} + \mu^2 \mathbf{W}_{\mathbf{m}}^T \mathbf{W}_{\mathbf{m}} \right)^{-1} \mathbf{h}^T \mathbf{r} \right\|_2^2 ,$$

+2 trace $\left[\mathbf{h} \left(\mathbf{h}^T \mathbf{h} + \mu^2 \mathbf{W}_{\mathbf{m}}^T \mathbf{W}_{\mathbf{m}} \right)^{-1} \mathbf{h}^T \right] - n$ (15)

where *trace*[] is the trace of the term in the brackets. The randomized trace estimation [9] was also introduced in Equation (15) to solve the difficulty from the calculation of the trace. Then, the UPRE function is approximated by

$$UPRE(\mu) \approx \left\| \mathbf{r} - \mathbf{h} \left(\mathbf{h}^{T} \mathbf{h} + \mu^{2} \mathbf{W}_{\mathbf{m}}^{T} \mathbf{W}_{\mathbf{m}} \right)^{-1} \mathbf{h}^{T} \mathbf{r} \right\|_{2}^{2} .$$

$$+ 2 \mathbf{R}^{T} \mathbf{h} \left(\mathbf{h}^{T} \mathbf{h} + \mu^{2} \mathbf{W}_{\mathbf{m}}^{T} \mathbf{W}_{\mathbf{m}} \right)^{-1} \mathbf{h}^{T} \mathbf{R} - n$$
(16)

The result of the RGSVD is introduced into Equation (16), and Equation (16) has the form

$$UPRE(\mu) \approx \left\| \mathbf{r} - \mathbf{h} \left[\sum_{i=1}^{q} \frac{\gamma_i^2}{\gamma_i^2 + \mu^2} \frac{\mathbf{u}_i^T \mathbf{r}}{c_i} (\mathbf{X}^{\dagger})_i \right] \right\|_2^2 + 2 \mathbf{R}^T \mathbf{h} \left[\sum_{i=1}^{q} \frac{\gamma_i^2}{\gamma_i^2 + \mu^2} \frac{\mathbf{u}_i^T \mathbf{R}}{c_i} (\mathbf{X}^{\dagger})_i \right] - n$$
(17)

Similarly, when the value of Equation (17) is the minimum, its corresponding parameter μ is the optimal regularization parameter μ_{opt} , which can be found between the minimum and maximum of the generalized singular value γ_i as well. Here, the parameter q is also set as n.

4. Synthetic Example Tests

In this study, one example was used to demonstrate the performances of the GCV and the UPRE methods using the RGSVD algorithm. For comparison, the CG method and the

GSVD method were also used in the GCV and UPRE methods. The following tests were run on a computer with a 3.00 GHz processor and 32 GB RAM.

The model consists of two cuboids, and Figure 1a shows its 3D perspective view. The dimensions of the two cuboids are both 300 m × 300 m × 200 m, and the depths of their tops are both 50 m. Both of them have a density contrast of 1 g/cm³. The gravity data were generated on a grid with $31 \times 21 = 651$ points and 50 m × 50 m spacings. Meanwhile, random Gaussian noise, whose standard deviation is 0.02 mGal plus 2% of each datum, were incorporated in the gravity data (Figure 1b). The subsurface was divided into $30 \times 20 \times 10 = 6000$ cells, and dimension of each cell is 50 m × 50 m × 50 m. In the following inversions, the density range was 0–1 g/cm³ and K_{max} was set as 50.



Figure 1. (a) 3D perspective view of the model consisting of two cuboids; (b) gravity data produced by this model.

First, we used the GCV method to determine the regularization parameter in inversion. In the GCV method, the CG method, the GSVD, and the RGSVD algorithm were used, respectively. Figure 2a–c show the inversion results from the GCV methods using the CG method, the GSVD, and the RGSVD algorithm, respectively. These three inversion results are very similar: two source bodies with cuboid shape are in the actual position of the true model. Their corresponding GCV function curves are shown in Figure 3a–c, and these three curves have similar trends. Meanwhile, Table 1 records the corresponding optimal regularization parameters μ_{opt} , and the corresponding elapsed times for the process of choosing the regularization parameter. The GCV methods using the GSVD and the RGSVD algorithm have very similar μ_{opt} , and the μ_{opt} of the GCV method using the CG method is slightly large. The GCV method using the RGSVD algorithm has the shortest elapsed time, at only 1.3 s.

Then, the UPRE method was implemented in inversion. Similarly, the UPRE method used the CG method, GSVD, and the RGSVD algorithm, respectively. Their inversion results are shown in Figure 2d–f, respectively. These inversion results are similar to those obtained using the GCV method. The curves for the UPRE function are demonstrated in Figure 3d–f, and they also have similar trends. Table 1 also lists the value of optimal regularization parameters and the elapsed times for the UPRE method. Their optimal regularization parameters are the same as them from the GCV method, and the elapsed times are close to them from the GCV method. The UPRE method using the RGSVD algorithm also has the shortest elapsed time.

Through the above comparative test, it was concluded that the GCV and the UPRE methods using the RGSVD algorithm can quickly provide an optimal regularization parameter that can generate a satisfying inversion result.



Figure 2. Inversion results from different regularization methods. The generalized cross-validation (GCV) methods using (a) the conjugate gradient (CG) method, (b) the generalized singular value decomposition (GSVD), and (c) the randomized generalized singular value decomposition (RGSVD algorithm; the unbiased predictive risk estimator (UPRE) methods using (d) the CG method, (e) the GSVD, and (f) the RGSVD algorithm. In all these panels, the black lines indicate the real position of the model.



Figure 3. The GCV functions using (**a**) the CG method, (**b**) the GSVD, and (**c**) the RGSVD algorithm; the UPRE functions using (**d**) the CG method, (**e**) the GSVD, and (**f**) the RGSVD algorithm. The blue curve is GCV function or UPRE function, and the red \bigcirc is the optimal regularization parameter.

	GCV Method			UPRE Method		
	CG	GSVD	RGSVD	CG	GSVD	RGSVD
µ _{opt}	100.0	45.3	44.4	100.0	45.3	44.4
Time	117.5 s	113.0 s	1.3 s	111.2 s	114.3 s	1.4 s

Table 1. The optimal regularization parameters and the elapsed times for different regularization methods.

Note: μ_{opt} is the optimal regularization parameter, and Time denotes the elapsed time for different regularization method.

5. Conclusions

We introduced the GCV and the UPRE methods using the RGSVD algorithm, with which the optimal regularization parameter can be determined fast in the inversion of potential field data. We demonstrated the effect of these two methods through a comparative test.

Author Contributions: Conceptualization, J.W. and X.M.; Investigation, H.T.; Methodology, Y.F.; Software, Y.F.; Visualization, H.T.; Writing—original draft, Y.F.; Writing—review & editing, J.W. and X.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by (1) The National Natural Science Foundation of China (Grant numbers: 41804099 and 41974161), and (2) The Fundamental Research Funds for the Central Universities (Grant number: 2-9-2019-040).

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Yan, J.; Lü, Q.; Chen, X.; Qi, G.; Liu, Y.; Guo, D.; Chen, Y. 3D lithologic mapping test based on 3D inversion of gravity and magnetic data: A case study in Lu-Zong ore concentration district, Anhui Province. *Acta Petrol. Sin.* **2014**, *30*, 1041–1053.
- Yan, J.; Chen, X.; Meng, G.; Lü, Q.; Deng, Z.; Qi, G.; Tang, H. Concealed faults and intrusions identification based on multi-scale edge detection and 3D inversion of gravity and magnetic data: A case study in Qiongheba area, Xinjiang, Northwest China. *Interpretation* 2019, 7, T331–T345. [CrossRef]
- 3. Wang, J.; Meng, X.; Li, F. Fast Nonlinear Generalized Inversion of Gravity Data with Application to the Three-Dimensional Crustal Density Structure of Sichuan Basin, Southwest China. *Pure Appl. Geophys.* **2017**, *174*, 4101–4117. [CrossRef]
- 4. Tikhonov, A.N.; Arsenin, V.Y. Solutions of Ill-Posed Problems. Math. Comput. 1977, 32, 491.
- Vatankhah, S.; Ardestani, V.E.; Renaut, R.A. Automatic estimation of the regularization parameter in 2D focusing gravity inversion: Application of the method to the Safo manganese mine in the northwast of Iran. *J. Geophys. Eng.* 2014, *11*, 045001. [CrossRef]
- 6. Morozov, V.A. On the solution of functional equations by the method of regularization. Sov. Math. Dokl. 1966, 7, 414–417.
- Golub, G.H.; Heath, M.; Wahba, G. Generalized cross-validation as a method for choosing a good ridge regression parameter. *Technometrics* 1979, 21, 215–223. [CrossRef]
- 8. Hansen, P.C. Analysis of discrete ill-posed problems by means of the L-curve. SIAM Rev. 1992, 34, 561–580. [CrossRef]
- 9. Vogel, C.R. Computational Methods for Inverse Problems. In *SIAM Frontiers in Applied Mathematics;* SIAM: Philadelphia, PA, USA, 2002.
- 10. Wei, Y.; Xie, P.; Zhang, L. Tikhonov Regularization and Randomized GSVD. *SIAM J. Matrix Anal. Appl.* **2016**, *37*, 649–675. [CrossRef]
- 11. Li, Y.; Oldenburg, D.W. 3-D inversion of magnetic data. *Geophysics* 1996, 61, 394–408. [CrossRef]
- 12. Li, Y.; Oldenburg, D.W. 3-D inversion of gravity data. Geophysics 1998, 63, 109–119. [CrossRef]
- 13. Li, Y.; Oldenburg, D.W. Fast inversion of large-scale magnetic data using wavelet transforms and a logarithmic barrier method. *Geophys. J. Int.* 2003, *152*, 251–265. [CrossRef]
- 14. Li, Z.; Yao, C.; Zheng, Y.; Wang, J.; Zhang, Y. 3D magnetic sparse inversion using an interior-point method. *Geophysics* **2018**, *83*, J15–J32. [CrossRef]
- 15. Li, Z.; Yao, C. 3D sparse inversion of magnetic amplitude data when strong remanence exists. *Acta Geophys.* **2020**, *68*, 365–375. [CrossRef]
- 16. Vatankhah, S.; Renaut, R.A.; Ardestani, V.E. Total variation regularization of the 3-D gravity inverse problem using a randomized generalized singular value decomposition. *Geophys. J. Int.* **2018**, *213*, 695–705. [CrossRef]