



## Analysis of Quasi-Newton Inversion of MT data

Alex Fabrício de Andrade Silva, and Cícero Régis,  
CPGf–UFPA, INCT–GP, Brazil.

Copyright 2019, SBGf - Sociedade Brasileira de Geofísica

This paper was prepared for presentation during the 16<sup>th</sup> International Congress of the Brazilian Geophysical Society, held in Rio de Janeiro, Brazil, 19–22 August 2019.

Contents of this paper were reviewed by the Technical Committee of the 16<sup>th</sup> International Congress of The Brazilian Geophysical Society and do not necessarily represent any position of the SBGf, its officers or members. Electronic reproduction or storage of any part of this paper for commercial purposes without the written consent of The Brazilian Geophysical Society is prohibited.

### Abstract

This paper presents an analysis of the Quasi-Newton BFGS Limited Memory method to invert geophysical electromagnetic data, in comparison with the traditional Gauss-Newton approach. The application is to inversion of magnetotelluric data, as an illustration of the methods. The comparison is in terms of the number of operations performed in each iteration of the inversion process and the total number of iterations needed in each case to achieve stable solutions. The results indicate that the BFGS Limited Memory method can achieve the same solutions with significantly less memory requirements than the Gauss-Newton method, although with a bigger number of iterations.

### Introduction

The regularized Gauss-Newton method is widely used in applications of geophysical inversion Constable et al. (1987). In this method, it is necessary to assemble two usually large matrices: 1) the Jacobian, or Sensitivity matrix, which is composed of the derivatives of the synthetic data in relation to the interpretive model parameters; and 2) the Hessian matrix, which approximates the second derivatives with products of first derivatives. Besides the large memory requirements to store these matrices during the inversion, calculating sensitivities is usually the most demanding task in Gauss-Newton inversion in terms of computer time.

Memory use and processing time are limiting factors in large problems, particularly in 3D applications. The Limited-Memory BFGS Quasi-Newton method was devised to reduce the memory requirements in the inversion process. This method uses an approximation to the inverse of the hessian matrix calculated directly from the gradient of the objective function in each iteration, and operates to generate a new estimate of the parameter vector without storing the complete hessian, nor its inverse. The gradient is calculated in each iteration by using a column of sensitivities at a time, which precludes the storage of the complete jacobian in memory.

The LM-BFGS method has been used for many large scale problems. In geophysical electromagnetic inverse problems, the quasi-Newton methods are gaining wide acceptance since the beginning of the 21<sup>st</sup> century, as

exemplified by the works of Newman and Boggs (2004), Haber (2005), Avdeeva and Avdeev (2006), and Avdeev and Avdeeva (2009).

To assess the gain in memory usage achieved with the Quasi-Newton BFGS Limited Memory method in the inversion of geophysical electromagnetic problems, this paper presents an analysis of an application to the inversion of magnetotelluric data from layered earth models. This simple case allows an easy estimation of the operation counts for the iterations in both the traditional Gauss-Newton method and the Limited Memory approach.

The example presented here shows the early results of an ongoing work which will culminate in the application to 3D problems.

The true importance of the Limited Memory method is in its application for very large problems, which easily reach the memory limits of ordinary workstations. However, this first application to 1D problems generates insights that will be useful in the later work with more realistic and demanding situations.

### Method

The inverse problem is defined as the fitting of a set of synthetic data ( $\mathbf{y}$ ) to a set of observations ( $\mathbf{y}^o$ ), by the determination of a set of parameters that define an interpretive model. The mathematical model of the synthetic data is expressed by the functional relation between the parameter vector  $\mathbf{p}$  and the synthetic observation vector  $\mathbf{y}$ :

$$\mathbf{y} = \mathbf{f}(\mathbf{x}, \omega, \mathbf{p}), \quad (1)$$

with the vector  $\mathbf{x}$  representing coordinates of the measurement stations and  $\omega$ , the angular frequencies used in the survey.

The  $L_2$  norm fitting function  $\phi(\mathbf{p})$  is defined as

$$\phi(\mathbf{p}) = \|\mathbf{y}^o - \mathbf{f}\|^2 = (\mathbf{y}^o - \mathbf{f})^T (\mathbf{y}^o - \mathbf{f}). \quad (2)$$

The problem is to find a parameter vector  $\mathbf{p}^*$  that minimizes the function  $\phi$ , i.e., a  $\mathbf{p}$  for which the gradient of  $\phi$  is zero.

### Gauss-Newton

For a problem with  $n$  observations and  $m$  parameters, the sensitivity matrix  $\mathbf{A}$  is defined as

$$\mathbf{A}_{n \times m} = \begin{bmatrix} \frac{\partial f_1}{\partial p_1} & \frac{\partial f_1}{\partial p_2} & \cdots & \frac{\partial f_1}{\partial p_{m-1}} & \frac{\partial f_1}{\partial p_m} \\ \frac{\partial f_2}{\partial p_1} & \frac{\partial f_2}{\partial p_2} & \cdots & \frac{\partial f_2}{\partial p_{m-1}} & \frac{\partial f_2}{\partial p_m} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{\partial f_{n-1}}{\partial p_1} & \frac{\partial f_{n-1}}{\partial p_2} & \cdots & \frac{\partial f_{n-1}}{\partial p_{m-1}} & \frac{\partial f_{n-1}}{\partial p_m} \\ \frac{\partial f_n}{\partial p_1} & \frac{\partial f_n}{\partial p_2} & \cdots & \frac{\partial f_n}{\partial p_{m-1}} & \frac{\partial f_n}{\partial p_m} \end{bmatrix} \quad (3)$$

The Gauss-Newton method approximates the hessian matrix  $\mathbf{H}$  of the second derivatives of the functional as the product of the first derivatives:

$$\mathbf{H} = \mathbf{A}^T \mathbf{A}. \quad (4)$$

To achieve a stable solution, it is necessary to impose a priori constraints to the solution. The most common practice is to use smoothing constraints that impose proximity between adjacent parameters by enforcing the following equality in the least squares sense:

$$\begin{bmatrix} p_2 - p_1 \\ p_3 - p_2 \\ p_4 - p_3 \\ \vdots \\ p_n - p_{n-1} \end{bmatrix} = \mathbf{0}, \quad (5)$$

$$\begin{bmatrix} -1 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 & 0 \\ 0 & 0 & -1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -1 & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ \vdots \\ p_{n-1} \\ p_n \end{bmatrix} = \mathbf{M}\mathbf{p} = \mathbf{0}. \quad (6)$$

The most widely used approach to control the convergence of the inversion is the Marquardt-Levenberg method, which adds to the diagonal of the hessian matrix a variable ( $\lambda$ ) that changes at each iteration to make the parameter estimation close to either a small step in the direction opposed to the gradient or to the Gauss-Newton step.

Including constraints with a relative weight given by the variable  $\mu$  and using the Marquardt-Levenberg method, at the  $k^{\text{th}}$  iteration, the estimation of the parameter vector is

$$\mathbf{p}_{k+1} = \mathbf{p}_k + \left[ \mathbf{A}^T \mathbf{A} + \mu \mathbf{M}^T \mathbf{M} + \lambda \mathbf{I} \right]^{-1} \left[ \mathbf{A}^T (\mathbf{y}^0 - \mathbf{f}) - \mu \mathbf{M}^T \mathbf{M} \mathbf{p} \right], \quad (7)$$

#### Quasi-Newton BFGS

In the Quasi-Newton BFGS method, the inverse of the hessian matrix for the  $(k+1)^{\text{th}}$  iteration is approximated by

$$\mathbf{H}_{k+1}^{-1} = (\mathbf{I} - \rho_k \mathbf{s}_k \mathbf{y}_k^T) \mathbf{H}_k^{-1} (\mathbf{I} - \rho_k \mathbf{y}_k \mathbf{s}_k^T) + \rho_k \mathbf{s}_k \mathbf{s}_k^T, \quad (8)$$

with

$$\rho_k = \frac{1}{\mathbf{y}_k^T \mathbf{s}_k}, \quad (9)$$

$$\mathbf{s}_k = \mathbf{p}_k - \mathbf{p}_{k-1}, \quad (10)$$

$$\mathbf{y}_k = \nabla \phi_k - \nabla \phi_{k-1}. \quad (11)$$

The parameter estimation is

$$\mathbf{p}_{k+1} = \mathbf{p}_k - \alpha_{k+1} \mathbf{H}_{k+1}^{-1} \nabla_p \phi(\mathbf{p}_j), \quad (12)$$

where the value for the step size  $\alpha$  is chosen to satisfy the Wolfe conditions Nocedal and Wright (1999):

$$\phi(\mathbf{p}_{k+1}) \leq \phi(\mathbf{p}_k) + c_1 \alpha_k \nabla \phi_k^T \delta \mathbf{p}_k, \quad (13)$$

$$|\nabla \phi_{k+1}^T \delta \mathbf{p}_k| \leq c_2 |\nabla \phi_k^T \delta \mathbf{p}_k|, \quad (14)$$

with  $0 < c_1 < c_2 < 1$ .

Given an estimation of  $\mathbf{H}_k^{-1}$  at the  $k^{\text{th}}$  iteration, the method yields the next estimation without the need to actually calculate the inverse of a matrix, and the parameter vector step is calculated with a single matrix-vector product.

The whole process depends on the determination of an initial estimation for the hessian. Nocedal and Wright (1999) present the approximation

$$\mathbf{H}_0^{-1} = \beta \|\nabla_p \phi(\mathbf{p}_0)\|^{-1} \mathbf{I}, \quad (15)$$

where  $\beta$  is an arbitrary scalar.

#### Limited Memory Quasi-Newton BFGS

A step beyond the estimation of the inverse of the hessian given by equation (8) was developed by Broyden, Fletcher, Goldfarb, and Shanno to produce an estimate of the product of the inverse hessian by the gradient by a recursive process that avoids the storing of the full hessian inverse in memory. This product is then used in the parameter estimation expressed in equation (12). The method known as BFGS stores the vectors  $\mathbf{s}_k = \mathbf{p}_k - \mathbf{p}_{k-1}$  and  $\mathbf{y}_k = \nabla \phi_k - \nabla \phi_{k-1}$  generated in several iterations and uses them to generate a new  $(k+1)$  estimation for  $\alpha_{k+1} \mathbf{H}^{-1} \nabla_p \phi(\mathbf{p}_j)$  to update the parameter vector in equation (12). There is, then, a significant reduction in the memory requirement to perform the inversion.

#### Application to the inversion of magnetotelluric data

The first results in this study are from the inversion of MT data from layered earth models. This problem presents little demand for memory and takes very little time, even in the traditional Gauss-Newton approach. However, it provides a preliminary a simple case that allows the fast development of the inversion code to be adapted to higher dimension problems in later phases of the study. It also allows a comparison in terms of the operation counts in both methods.

The model to generate the synthetic data is composed of four layers with the following resistivities and thicknesses:

$$\begin{array}{ll} \rho_1 = 500\Omega\text{m} & h_1 = 3000 \text{ m} \\ \rho_2 = 60\Omega\text{m} & h_2 = 1000 \text{ m} \\ \rho_3 = 700\Omega\text{m} & h_3 = 2000 \text{ m} \\ \rho_4 = 50\Omega\text{m}. & \end{array}$$

The data are apparent resistivities from a sounding using 61 frequencies in the range from  $10^{-4}$  to  $10^3$  Hz.

The interpretive model is composed of 50 layers with fixed thicknesses, starting at 50 m and increasing at a 5% rate, limited at a depth of 10 km.

The same data was inverted using Gauss-Newton and the LM-BFGS method. The same stable solution was achieved using the same stopping criterium by both methods. The Gauss-Newton inversion run for 23 iterations, whereas the LM-BFGS program took 88 iterations to reach the same solution, as illustrated in figure 1. The resulting smooth models from both methods are virtually identical, as shown in figure 2.

In the comparison between the two methods, LM-BFGS has the advantage in two key aspects for this example:

- The LM-BFGS method stored vectors  $s$  and  $y$  from 5 iterations. This means that a total of 500 real numbers were stored in memory simultaneously. In the Gauss-Newton inversion, on the other hand, the hessian matrix alone is  $50 \times 50$ , and the sensitivity matrix is  $61 \times 50$ , which means a total of 5550 elements in memory at the same time, more than 10 times the demand of the LM-BFGS method.
- The number of operations performed in an iteration of the LM-BFGS method is much smaller than that of the Gauss-Newton method because the former does not need to solve a large system of linear equations, requiring only vector multiplications for which the operation count is relatively small. The difference is enough to overcome the greater number of iterations of the LM-BFGS program.

### Final Remarks

The example shown here illustrates the gain in memory usage for performing inversion with the Limited-Memory BFGS Quasi-Newton method in comparison with the Gauss-Newton method. The conclusion drawn for the 1D MT case is an indication that the LM-BFGS method has the potential to significantly improve the inversion of electromagnetic data in 2D and 3D cases.

### Acknowledgments

The authors thank Petrobras for the grant and the scholarship that support this work through the research project number 2017/00424-6.

### References

- Avdeev, D., and A. Avdeeva, 2009, 3d magnetotelluric inversion using a limited-memory quasi-newton optimization: *Geophysics*, **74**, F45–F57.
- Avdeeva, A., and D. Avdeev, 2006, A limited-memory quasi-newton inversion for 1d magnetotellurics.: *Geophysics*, **71**, G191–G196.
- Constable, S. C., R. L. Parker, and C. G. Constable, 1987, Occam's inversion: a practical algorithm for generating smooth models from em sounding data: *Geophysics*, **52**, 289–300.
- Haber, E., 2005, Quasi-newton methods for large-scale electromagnetic inverse problems: *Inverse Problems*, **21**, 305–323.
- Newman, G. A., and P. T. Boggs, 2004, Solution accelerators for large-scale three-dimensional electromagnetic inverse problems: *Inverse Problems*, **20**, S151–S170.
- Nocedal, J., and S. J. Wright, 1999, *Numerical optimization*: Springer.

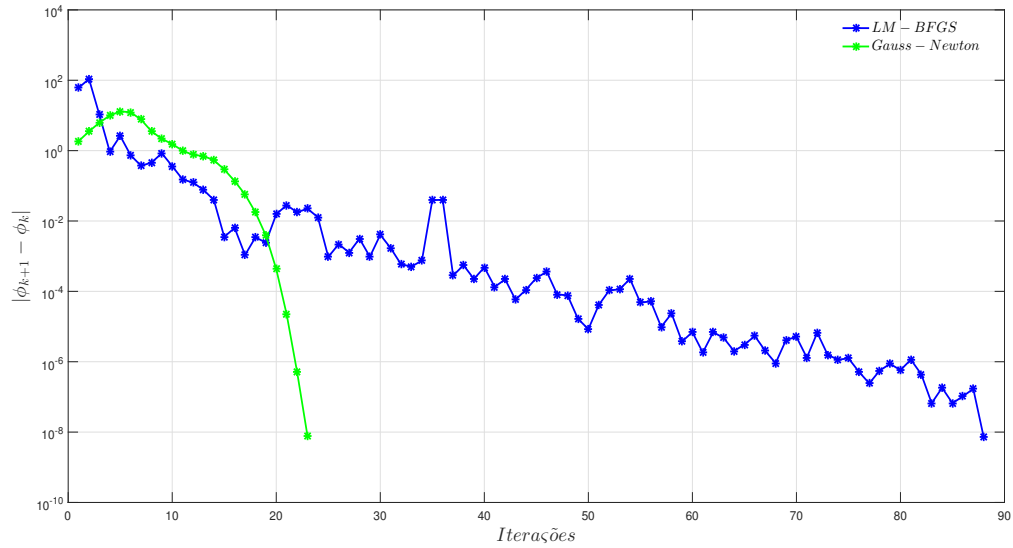


Figure 1: Convergence of the difference  $|\phi_k(\mathbf{p}_j) - \phi_{k-1}(\mathbf{p}_j)|$  in both inversion methods

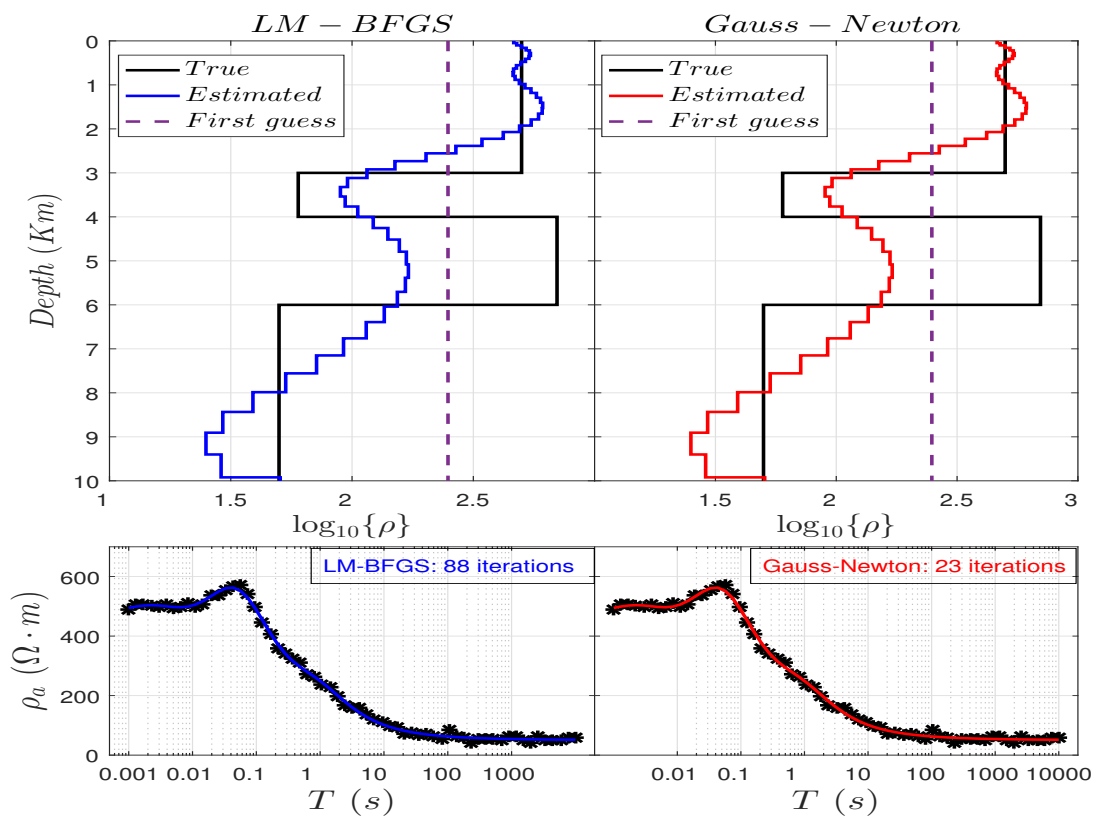


Figure 2: Inversion results and data fitting achieved by both methods.