
Direct Current Resistivity Inversion using Various Objective Functions

Rowan Cockett

Department of Earth and Ocean Science
University of British Columbia
rcockett@eos.ubc.ca

Abstract

In geophysical applications direct current (DC) resistivity surveys are collected by injecting current and recording voltage potentials over a field site of interest. These data sets can provide valuable non-invasive images of the sub-surface. To create these images, a minimization problem between model-generated data and observed data is solved. Data is generated through forward modelling and then compared to the observed data in an inversion process. In this paper, both the forward modelling and inversion of the DC resistivity problem are implemented and investigated. A two part objective function, containing data-misfit and model regularization terms, is used in the minimization. The effect of the objective function is investigated through the use of three different functions: (1) the l^2 -norm; (2) the l^1 -norm; and (3) the Student's t -distribution. Synthetic data is produced from a simple conductivity model and various types of random noise are added to the data. Inversions of this synthetic data are presented and indicate that the objective function chosen has a significant effect on the inversion result. Results suggest that the l^1 -norm and the Student's t -distribution are more robust in the face different levels of Gaussian noise, which are considered outliers by the l^2 -norm. It is concluded that consideration should be given to the choice of objective function and that this choice should be guided by *a priori* information about the data noise.

1 Introduction

A set of problems in geophysical applications involves the conversion of observational data into a predictive model that feasibly creates those observations. These problems can be divided into two distinct elements: the forward problem and the inverse problem. The forward problem simulates an observational data set given model parameters; this is described mathematically by a partial differential equation. The inverse problem solves for feasible model parameters given a sparse set of observational data.

One specific context for this problem is in terms of the direct-current (DC) resistivity problem. In this experiment, electrodes are laid out over a field site of interest; source pairs (input at a known current) are sequentially turned on and off and receiver pairs observe the responding voltage potentials. After many voltage measurements have been made, they can be combined in a geophysical inversion to gain information about the conductivity of the subsurface. The conductivity of the ground can be linked to physical parameters of interest. For example, pore-water conductivity is greatly influenced by electrolytic contaminants, these contaminants can thus be imaged through their electrical properties in a DC resistivity survey and appropriate actions can be taken [1]. The context of this paper is solving the optimization problem that arises from the geophysical inverse problem for DC resistivity.

Due to the nature of a relatively sparse data set, the DC resistivity problem is underdetermined and a two part objective function is minimized containing both a data-misfit term and a model regularization term [2]. Although many forms are available for use in these objective functions, the l^2 -norm is commonly used throughout the literature without exploring alternatives (e.g. [3] [4] [5]). In this paper the form of the data-objective function is explored through the use of three different objective functions: (1) the l^2 -norm, (2) the l^1 -norm, and (3) the Student's t -distribution.

Using a forward modelling code for the DC resistivity equation, it is possible to generate synthetic data from simple models and then invert the synthetic data to recreate that model by minimizing data-misfit. Different perturbations can be added to the model, generally in the form of block conductivity changes, to investigate how sensitive the inversion process is to different types of changes. In this paper, the different objective functions are tested with respect to robustness to noise levels and the recovery of the block conductivity perturbations.

2 Forward Problem & Discretization

The DC resistivity problem is governed by Poisson's equation with appropriate boundary conditions applied

$$\nabla \cdot (-\sigma \nabla \phi) = \mathbf{I}(\delta(\mathbf{r} - \mathbf{r}_{s+}) - \delta(\mathbf{r} - \mathbf{r}_{s-})) \quad (1)$$

where σ is the conductivity structure of a medium; ϕ is the electrical potential field induced by a dipole; and \mathbf{I} is the electrical current from a dipole. The dipole is represented by two dirac delta functions centered on the positive and negative source locations (\mathbf{r}_{s+} and \mathbf{r}_{s-} respectively) [3]; where \mathbf{r} is a position vector. Neumann boundary conditions are often applied at infinity or sufficiently far from the area of interest in a discretized model [6]. These boundary conditions are representative of a field experiment where the site is unbounded.

In this paper a cell-centered, finite-volume mesh was written to discretize the DC resistivity equation, in matrix notation Equation 1 can be written:

$$\mathbf{D} \text{diag}(\mathbf{A}_v \mathbf{e}^{\mathbf{m}}) \mathbf{G} \mathbf{u} = \mathbf{q} \quad (2)$$

where \mathbf{D} and \mathbf{G} are matrix representations of the divergence and gradient operators; \mathbf{u} is a vector containing the potential difference field; and \mathbf{q} contains the positive and negative source locations. \mathbf{A}_v averages the conductivity values from cell-centers to cell-faces, and is included on the diagonal. The conductivity model of the medium, \mathbf{m} , is in log-conductivity, and is included via the exponential $\mathbf{e}^{\mathbf{m}}$. Choosing to work in log conductivity (a) enforces a positivity constraint, and (b) allows for interpretations of the results in either conductivity or the inverse: resistivity. The entire forward operator is dependent on the conductivity model and can be written

$$\mathbf{A}(\mathbf{m}) \mathbf{u} = \mathbf{q} \quad (3)$$

where \mathbf{u} is a vector that contains the potential field induced by the source configuration in \mathbf{q} . Neumann boundary conditions were implemented in the forward operator; however, these boundary conditions lead to a constant null space in the forward operator. The null space is removed by modifying cell $\mathbf{A}(1, 1)$ by ± 1 to conform to the sign of that element [3]. It is noted that the data collected in any DC resistivity experiment is a subset of the entire potential field, thus a projection matrix, \mathbf{P} , is used to pick out these measured potentials

$$\mathbf{d}(\mathbf{m}) = \mathbf{P} \mathbf{u} = \mathbf{P} \mathbf{A}(\mathbf{m})^{-1} \mathbf{q} \quad (4)$$

These data, $\mathbf{d}(\mathbf{m})$, can now be directly compared to the observed data, \mathbf{d}_{obs} , generated for that same source configuration. In a DC-resistivity survey there are generally many right-hand sides to Equation 3, and each must be evaluated to compare with the full observational dataset.

3 Objective Function

The DC resistivity inversion results in an unconstrained non-linear optimization problem that is often underdetermined; that is, observational data is much fewer than the number of model parameters. Due to the inverse problem being underdetermined, a weighted regularization parameter is generally added to the optimization problem and the two-part function is minimized

$$\Phi(\mathbf{m}) = \sum_{i=1}^n \rho(\mathbf{d}^{(i)}(\mathbf{m}) - \mathbf{d}_{\text{obs}}^{(i)}) + \frac{\beta}{2} \|\mathbf{G}_w(\mathbf{m} - \mathbf{m}_{\text{ref}})\|_2^2 \quad (5)$$

where the first term minimizes data-misfit between the generated data, $\mathbf{d}(\mathbf{m})$, and the observed data (\mathbf{d}_{obs}); $\rho(\cdot)$ is the data objective function (often taken as $\rho(\cdot) = (\cdot)^2$ to yield the sum of squares). Data is generated for each source configuration, and the individual data-misfit terms are summed over the data corresponding to the n different right-hand sides. The second term is a Tikhonov style term that controls model regularization with regard to a reference model (\mathbf{m}_{ref}) [7]. \mathbf{G}_w is a combination of the gradient operator in Equation 2, which is sensitive to model flatness; and the identity matrix, which is sensitive to model smallness. Both flatness and smallness are with respect to \mathbf{m}_{ref} . The

gradient can be weighted in different directions to create anisotropic regularization; for example, less regularization in the z-axis direction than other axes. β is a regularization parameter that balances the inversion between minimizing data-misfit and model regularization. In place of model smallness, a depth weighting term can be introduced that decreases regularization weight with depth. Depth weighting can be very important if all data is taken on the surface but it is of interest to know what the conductivity structure is at depth. The depth weighting term can replace the identity matrix in \mathbf{G}_w and has the form:

$$\mathbf{W}_d = \text{diag}((\mathbf{Z}(:) + z_0)^{-v}) \quad (6)$$

where $\mathbf{Z}(:)$ is a unwrapped matrix that indicates the depth value of all model cells (z increases with depth); z_0 is the initial depth; and v is a parameter that controls the magnitude of the depth weighting. The calculated vector is placed on the diagonal of \mathbf{W}_d . A v of zero causes \mathbf{W}_d to revert to the identity and act solely as a smallness term. It is clear that regularization can play an important role in the construction of the objective function. The inclusion of *a priori* knowledge into the DC resistivity inversion can be completed by adjusting parameters such as \mathbf{m}_{ref} to include the background conductivity or adjusting flatness regularization in different directions to agree with geologic assumptions.

To deal with the non-linear aspect of this problem, the objective function can be linearized about the current model [3]

$$\Phi(\mathbf{m}) = \sum_{i=1}^n \rho \left(\left(\mathbf{d}^{(i)} + \frac{\partial \mathbf{d}^{(i)}}{\partial \mathbf{m}} \cdot \partial \mathbf{m} \right) - \mathbf{d}_{\text{obs}}^{(i)} \right) + \frac{\beta}{2} \|\mathbf{G}_w(\mathbf{m} - \mathbf{m}_{\text{ref}})\|_2^2 \quad (7)$$

The gradient and Hessian of the objective function using this linearization, for one source configuration, are:

$$\nabla \Phi(\mathbf{m}) = \mathbf{J}^T \nabla \rho + \beta \mathbf{G}_w^T \mathbf{G}_w (\mathbf{m} - \mathbf{m}_{\text{ref}}) = \mathbf{g} \quad (8)$$

$$\nabla^2 \Phi(\mathbf{m}) = \mathbf{J}^T \nabla^2 \rho \mathbf{J} + \beta \mathbf{G}_w^T \mathbf{G}_w = \mathbf{H} \quad (9)$$

Here \mathbf{J} is the Jacobian of the objective function ($\partial \mathbf{d} / \partial \mathbf{m}$) and describes the sensitivity of data to changes in model parameters. Note that this gradient must be computed once for each right-hand side. The Jacobian is a large and dense matrix, and it is preferable to avoid explicitly forming the matrix to reduce memory costs. Haber [5] shows the Jacobian in this problem has the form:

$$\mathbf{J} = \mathbf{P} \mathbf{A}^{-1} (\mathbf{D} \text{diag}(\mathbf{G} \mathbf{u}) \mathbf{A}_v \text{diag}(\mathbf{e}^m)) \quad (10)$$

Using this expression for the Jacobian, it is seen that the full matrix \mathbf{J} need not be formed explicitly, and only the effect of this matrix on a vector is needed. Two functions are needed to calculate a matrix-vector product, one that calculates $\mathbf{J}(\mathbf{v})$ and one that calculates $\mathbf{J}^T(\mathbf{w})$ for some vectors \mathbf{v} and \mathbf{w} . Both of these functions require three parts, $\mathbf{J}(\mathbf{v})$ can be calculated through

$$\mathbf{z} = (\mathbf{D} \text{diag}(\mathbf{G} \mathbf{u}) \mathbf{A}_v \text{diag}(\mathbf{e}^m)) \mathbf{v} \quad (11a)$$

$$\mathbf{A} \mathbf{x} = \mathbf{z} \quad (11b)$$

$$\mathbf{J}(\mathbf{v}) = \mathbf{P} \mathbf{x} \quad (11c)$$

where the vector is first multiplied sequentially by all operators in Equation 11a to obtain a new vector \mathbf{z} . A linear system with \mathbf{A} is then solved for \mathbf{x} , here the explicit inverse of \mathbf{A} is not necessary and may be solved with an iterative solver rather than a direct solver if necessary. The vector \mathbf{x} is then multiplied by the projection matrix \mathbf{P} to yield $\mathbf{J}(\mathbf{v})$. A similar process is used for calculating $\mathbf{J}^T(\mathbf{w})$. The accuracy of this function can be tested by computing $\mathbf{w}^T \mathbf{J}(\mathbf{v})$ and comparing it to $\mathbf{v}^T \mathbf{J}^T(\mathbf{w})$; the results of this adjoint test should be identical. It is noted that computing the Jacobian in this form requires solving a linear system with \mathbf{A} (Equation 11b) and computing this many times to high accuracy could be computationally prohibitive.

3.1 Data Objective Function

Three data objective functions $\rho(\mathbf{r}_i)$ are explored in this paper, where \mathbf{r}_i is the i^{th} data residual ($\mathbf{d}_m^{(i)} - \mathbf{d}_{\text{obs}}^{(i)}$): (a) the l^2 -norm, which results in the sum of squares in Equation 5; (b) the l^1 -norm; and (c) the Student's t -distribution.

$$\rho_2(\mathbf{r}_i) = \frac{1}{2} \mathbf{r}_i^T \mathbf{r}_i \quad (12a)$$

$$\rho_1(\mathbf{r}_i) = \sqrt{\mathbf{r}_i^T \mathbf{r}_i + \epsilon} \quad (12b)$$

$$\rho_t(\mathbf{r}_i) = \log(1 + \mathbf{r}_i^T \mathbf{r}_i) \quad (12c)$$

In the l^1 -norm ϵ is some small positive constant to allow for differentiability at zero. This approximation of the true absolute value function does not significantly change the results of the inversion as the errors introduced are small in regards to other modelling and regularization errors. The statistical interpretation of these three objective functions can be considered through “tail-heaviness” of the error distribution [8]. As seen in Figure 1 the l^2 -norm is the most sensitive to outliers, as they are not expected in Gaussian noise modelling [8]. The l^1 -norm is usually chosen as a robust approach to outlying data, such that they do not have a disproportionate effect on the inverse solution [9]. Another robust approach is to use the Student’s t -distribution to model error noise [8]. Care must be taken with the Student’s t -distribution as this function is non-convex; however, in this non-linear inversion, convexity is unlikely.

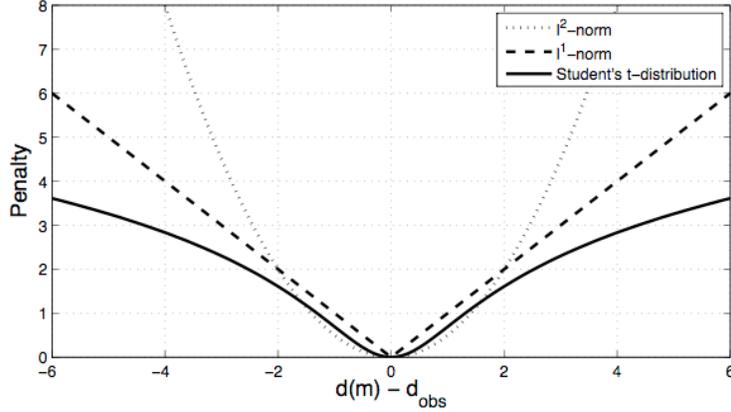


Figure 1: Penalties of three objective functions (Equations 12a - 12c).

3.2 Data Weighting

The magnitude of the data varies by an order of magnitude or more, thus when fitting with the objective function it is necessary to change the weight of the inversion so that it is equally sensitive to large and small data if they are valid. This data weighting is done by adding a weighting matrix to the data residual ($\mathbf{d}_m^{(i)} - \mathbf{d}_{\text{obs}}^{(i)}$), for example in the least-squares objective function:

$$\Phi(\mathbf{m}) = \sum_{i=1}^n \left\| \mathbf{W}^{(i)} (\mathbf{d}^{(i)}(\mathbf{m}) - \mathbf{d}_{\text{obs}}^{(i)}) \right\|_2^2 + \frac{\beta}{2} \left\| \mathbf{G}_w (\mathbf{m} - \mathbf{m}_{\text{ref}}) \right\|_2^2 \quad (13)$$

Where $\mathbf{W}^{(i)}$ is a diagonal weighting matrix that is multiplied by the i^{th} residual vector:

$$\mathbf{W}^{(i)} = \text{diag} \left(\frac{1}{\left| \mathbf{d}_{\text{obs}}^{(i)} \right| \cdot SD(\mathbf{d}_{\text{obs}}^{(i)}) + \epsilon} \right), \text{ for } i = 1, 2, \dots, n \quad (14)$$

where $SD(\mathbf{d}_{\text{obs}}^{(i)})$ is the element wise standard deviation of each data; ϵ is a small positive constant that ensures a cut-off value so an extremely large weight is not given to very low amplitude data. This operation is done element wise on the vector $\mathbf{d}_{\text{obs}}^{(i)}$ corresponding to one right-hand-side vector \mathbf{q}_i ; the result is placed on the diagonal of the weighting matrix $\mathbf{W}^{(i)}$. A unique weighting matrix is necessary for each right-hand side, as seen in Equation 13. The standard deviation of data is often unknown in a field situation and is assumed to be 2-3% of the data [6]. The weighting matrix multiplies the residual in the data objective function, and must also be included in the gradient and Hessian.

4 Minimization

A Newton update of the form of $\mathbf{H}\partial\mathbf{m} = -\mathbf{g}$ can be obtained by combining Equations 8 and 9. For the least squares data objective function the Newton update has the form

$$(\mathbf{J}^T \mathbf{J} + \beta \mathbf{G}_w^T \mathbf{G}_w) \partial \mathbf{m} = -(\mathbf{J}^T (\mathbf{P} \mathbf{A}^{-1} \mathbf{q} - \mathbf{d}_{\text{obs}}) + \beta \mathbf{G}_w^T \mathbf{G}_w (\mathbf{m} - \mathbf{m}_{\text{ref}})) \quad (15)$$

Although \mathbf{J} does not need to be explicitly formed, every product involves the solution of the linear system involving \mathbf{A} (Equation 11b). Solving this system could become prohibitively expensive if a high tolerance solution is required [3]. Additionally, because the Hessian is a very large and dense matrix, solving for the model update, $\partial\mathbf{m}$, is expensive. To address these numerical difficulties a quasi-Newton approach was taken using a limited memory BFGS update to incorporate changes in gradient information into an approximation of the Hessian [10]. In limited memory BFGS only a certain number of previous gradients are held in memory and incorporated into the search direction for any given iteration. The limited memory BFGS algorithm requires an initial Hessian \mathbf{H}_0 ; commonly this is set to the identity. In this application the initial Hessian was set to:

$$\mathbf{H}_0 = \mathbf{G}_w^T \mathbf{G}_w \quad (16)$$

Setting the initial Hessian to the regularization term of the true Hessian helps to combat some issues of scaling as the discretization gets finer [6].

4.1 Line Search

After a model update, $\partial\mathbf{m}$, is found, it is incorporated into the current model by

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha_k \partial\mathbf{m} \quad (17)$$

where α_k is the current line search parameter and \mathbf{m}_{k+1} is the new model. The line search parameter is used to control the magnitude of the update. The update must be scaled appropriately so as not to violate the linearization about the current model. Additionally, α_k is systematically decreased such that a sufficient decrease in the objective function is obtained. The sufficient decrease is enforced by a backtracking Armijo line search [11] where α_k is decreased until the following inequality is satisfied

$$\Phi(\mathbf{m}_{k+1} + \alpha_k \partial\mathbf{m}) \geq \Phi(\mathbf{m}_k) + c_1 \alpha_k \mathbf{g}^T \partial\mathbf{m} \quad (18)$$

where c_1 is a constant usually chosen to be quite small (e.g. 10^{-4}) [6] [3]; \mathbf{g} is the gradient of the function. This inequality is evaluated iteratively decreasing α_k by a constant c_2 each time the inequality is not satisfied; c_2 is generally chosen to be between 0.4 and 0.8, with higher values indicating a finer line search and requiring more evaluations of the objective function [12]. Evaluation of the objective function is expensive, and an extensive line search should be avoided. To combat excessive line search iterations, adaptive scaling of the model update is implemented. A maximum model update based on $\|\partial\mathbf{m}\|_\infty$ is chosen, if the model update is larger than this tolerance, it is scaled back. The model update tolerance is increased by a factor of two, to a maximum of one, if no line search iterations are required. The update tolerance is decrease by a factor of 3/4 if more than one line search iteration is needed. The adaptive scaling of the model update leads to fewer iterations in the line search routine. Typically, between 0 and 2 line search iterations were needed with a maximum of around 4. Additionally, without this adaptive scaling technique, the model update can be too large and violate the linearity assumption in Equation 7 leading to problems in evaluating the objective function.

It is noted that although a BFGS update is used, Wolfe conditions were not implemented, although they are theoretically required [12]. Enforcing the Wolfe conditions requires more evaluations of the objective function, which is expensive. Additionally, in this application it was found that these additional conditions were unnecessary when the inverse problem converged to a reasonable solution; unreasonable solutions were characterized by over-fitting of the data, and were fixed by adding more regularization in the objective function.

4.2 Stopping Criteria

The minimization procedure is repeated with a new linearization until either the norm of the gradient falls below a certain tolerance, or the maximum number of iterations have been exhausted. Values for these stopping criteria are often problem dependent [3]. Running synthetic models that have elements similar to observed data can often give insight into appropriate stopping criteria [2]. A stopping criteria that can be used effectively is the absolute size of the model step [6]. Monitoring the step size can be used to cut-off an inversion when negligible progress is being made. If the updates being added to the model are orders of magnitude smaller than the magnitude of the features of interest, this additional model structure is not of use. The inversions converged in 30-50 iterations using these stopping criteria. Additionally, as discussed briefly in Section 4.1, scaling can be an issue in the gradient of the function, so using it as a valid stopping criteria can be difficult. Testing stopping criteria on synthetic models where the true solution is known, as in this paper, is the best way to understand what stopping criteria to use.

5 Experimental Setup

A forward modelling code was written that implemented Equations 2-4 over a unit cube with Neumann boundary conditions. The forward operators were tested for analytical potential fields with the appropriate boundary conditions. A series of electrode arrays (surveys) were written to produce and collect data from the forward model; the survey used in this paper considered all receiver permutations in a grid on the top surface of the model. It is noted that it is not possible experimentally to collect data at the same location as the source electrodes; although these configurations were programmed, the standard deviation of the data was set to infinity in Equation 14 and $\mathbf{W}_{jj}^{(i)} = 0$ so this residual was not included in the objective function.

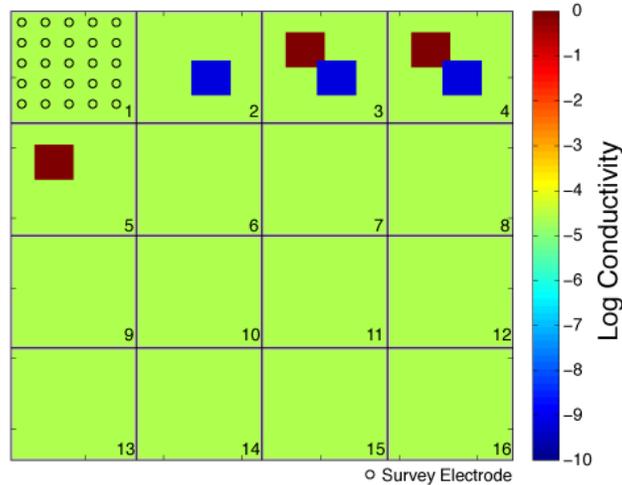


Figure 2: True model used for creation of synthetic data; showing locations of survey electrodes and conductivity values of block perturbations in log-conductivity.

5.1 Model & Data Setup

The models used for this experiment were $16 \times 16 \times 16$ with the survey grid spacing of 3 cells centered on the upper surface of the model (Figure 2). There were a total of 25 electrodes, 300 source configurations, and 253 active measurements per source dipole. This gave rise to 75 900 total measurements half of which are symmetric and likely would not have been collected in a field experiment, but were collected in this numerical experiment. The test model had a constant conductivity of 10^{-2}Sm^{-1} and contained two block perturbations seen in Figure 2. The entire model is presented in 16 sub-blocks that show horizontal depth slices. The top left sub-block is the top most horizontal slice, and all sources and receivers are located in this layer; depth increases going left to right and then continues on the next row. The true model contains two blocks of differing conductivity, the resistive block has a conductivity of 10^{-4}Sm^{-1} (log conductivity of -9.2103) and is shown in blue in Figure 2. The conductive block has a conductivity of 10^0Sm^{-1} (log conductivity of 0) and is shown in red in Figure 2. Both the conductive and resistive block perturbations have a size of $5 \times 5 \times 3$, but the resistive block is one cell closer to the surface as well as overlaps the conductive block by one cell (Figure 2).

Data was produced from the forward operator using the true model, and random Gaussian noise was subsequently added. The standard deviation of this noise varied from 0% (true data) to 2%. To test the different objective functions, additional spurious noise was added to some data sets. This spurious noise type adds additional Gaussian noise with a standard deviation of 25% to a random subset of the data. Three data-sets with differing noise levels were generated using the model in Figure 2: data with no noise, data with 2% random noise added, and data with 2% noise as well as 20% of the data with spurious noise. These data sets will be referred to as data with no noise, 2% noise, and spurious noise. In the following experiments data weighting in Equation 14 was assumed to be 2% regardless of the noise levels in the data.

Table 1: Recovered average conductivity for the resistive and the conductive block for three different noise levels, all values are in log conductivity. The true conductive block has a value of 0, and the resistive block has a value of -9.2103.

Objective Function	No Noise		2% Noise		Spurious Noise	
	Conductive	Resistive	Conductive	Resistive	Conductive	Resistive
l^2 -norm	-2.3631	-6.4601	-2.4002	-6.3991	-2.9646	-6.0689
l^1 -norm	-2.6804	-6.0715	-2.4448	-6.1217	-2.7754	-5.9339
Student's t -dist.	-2.4815	-6.0001	-2.5179	-5.8983	-2.6081	-5.9072

5.2 Regularization Setup

The regularization parameters used influence the inversion results substantially and must be selected carefully. The operator \mathbf{G}_w contains the gradient in the x , y , and z directions as well as a depth weighting term. The relative magnitude of these sub-operators was chosen to be the same in the x , y , and z directions with the depth weighting term having twice the relative magnitude of the other flatness terms. The regularization parameters in the depth weighting term were chosen as $z_0 = 1$ and $v = \pi$; this gave a weighting difference of one order of magnitude between the top and bottom of the model. These regularization parameters were experimented with thoroughly, and the values reported represent regularizations that produced the best results. The magnitude of the entire regularization objective function is controlled by β in Equation 5. The regularization parameter was selected by varying β between 10^{-1} and 10^4 , running the inversion for 50 iterations on a 9x9x9 model and plotting the data objective function against the regularization objective function; this analysis produces a L-shaped trade-off curve between data misfit and model regularization. The region of maximum positive curvature was visually chosen for each objective function yielding β values of 10^1 , $10^{0.5}$, and $10^{0.5}$, for the l^2 -norm, l^1 -norm, and Student's t -distribution respectively. These values of β were used as an initial guess for subsequent inversion, and were varied up to an order of magnitude on either side. The reference model used in the regularization was a homogeneous model containing the background conductivity of the model (10^{-2}Sm^{-1}).

6 Results

Data is produced from the true model (Figure 2) and three noise levels are added to the data: no noise, 2% noise, and spurious noise. The inversion was run for each of these data sets using the three objective functions, the l^2 -norm, the l^1 -norm, and the Student's t -distribution. The inversions for the spurious noise data set are seen in Figure 3 and are visually similar to the inversions of data sets with other noise levels. In Figure 3a the inversion for the l^2 -norm is shown in full, and can be compared to the true model in Figure 2; only the top 8 horizontal layers of the l^1 -norm and the Student's t -distribution inversions are shown for comparison to the l^2 -norm inversion. As seen in the inversion results, both the conductive and resistive blocks are recovered by the various inversions. It is noted that the blocks are smoothed as a result of the regularization applied. The smoothing results in a conservative estimate of the magnitudes of both the resistive and conductive blocks (i.e. values closer to the reference model). The location of the block in the horizontal plane is well constrained by the inversion; however, in the depth axis there is considerably more smoothing and regularization. This result is consistent with the layout of the survey electrodes that are in the horizontal plane; no data is collected in the depth axis. All inversions placed the most conductive cell in the fourth horizontal slice, which is in the middle of the conductive block of the true model; a similar result was found with the resistive block, with the inversions finding the most resistive cell in the third horizontal slice. All of the models also found that the top horizontal layer of the model had no significant perturbations from the background conductivity. The inversion with the Student's t -distribution as the objective function consistently had the best estimates of this first layer.

To compare the inversions quantitatively the average log conductivity value was found where the true model contained the block perturbation; both the conductive and resistive blocks were averaged for all inversions and are recorded in Table 1. The values of log conductivity can be compared to the true values of 0 for the conductive block and -9.2103 for the resistive block. It is seen that the l^2 -norm produces averages closer to the true log-conductivity values for the

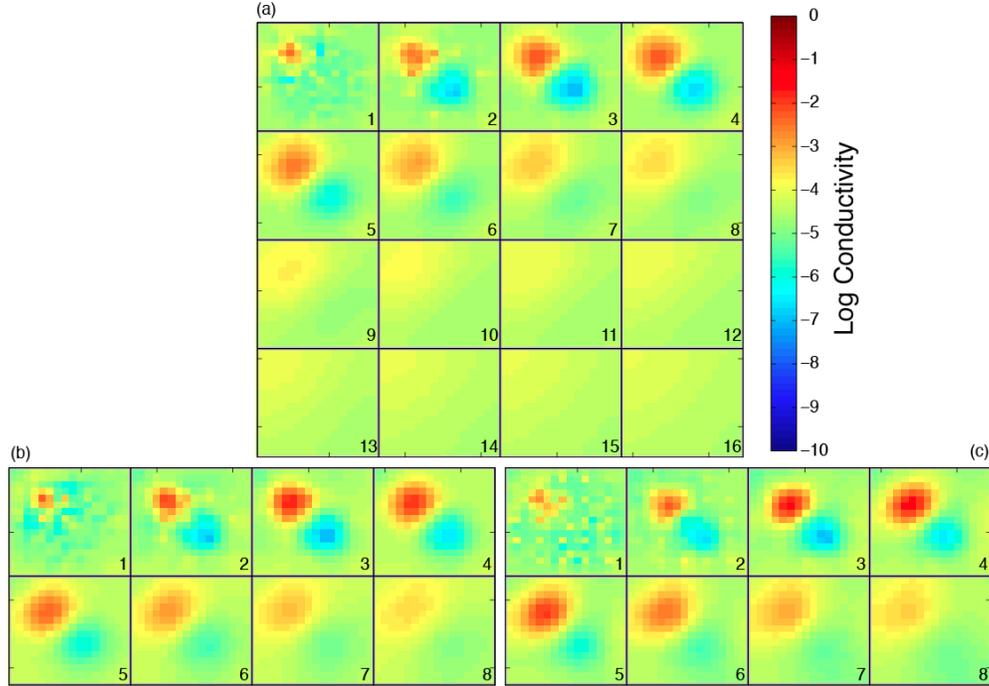


Figure 3: Inversion results in log conductivity for the data set with spurious noise using three objective functions: (a) the l^2 -norm, (b) the l^1 -norm, and (c) the Student's t -distribution.

inversions of data with no noise and with 2% noise. However, when spurious noise was added to the data, the l^2 -norm no longer outperformed the l^1 -norm and Student's t -distribution. With spurious noise in the inversion the robust objective functions produced better averages than the l^2 -norm of the conductive block and very similar averages of the resistive block. Inversions of other models with the same noise levels introduced into the data produced similar results; with the l^2 -norm providing the closest average estimates to the true conductivity for low levels of gaussian noise. When spurious noise is introduced into the data sets the robust data objective functions outperform the l^2 -norm.

7 Conclusions & Future Work

In this paper, I have implemented a forward modelling DC resistivity code to create synthetic data. In the implementation of the geophysical inversion significant care is placed on memory-constraints by not explicitly forming the sensitivity matrix J . Additionally, the line-search implemented allows for an adaptive technique to deal with a scaled gradient as well as minimizing the number of iterations required by the line search. The synthetic data was inverted to produce models comparable to the true model. Block perturbations in the true model were well constrained by the inversion in the horizontal plane. However, there was significant regularization introduced in the depth axis by the inversions; this is due to the nature of the survey design that resides only on the top surface of the model. More optimal survey designs that collect information at depth as well as on the surface of the model could be used to better constrain the z -dimension of anomalies.

The inversions were completed with three different objective functions. Results suggest that the robust object functions (l^1 -norm and Student's t -distribution) produced better results than the standard l^2 -norm in the synthetic experiments that contain significantly different levels of Gaussian noise. The objective function as well as the regularization parameters have a large effect on the inversion results and can be guided through the use of *a priori* information about data noise. It is recommended that various objective functions be considered in these geophysical inversions, as their implementation is simple but significant improvements in inversion results could be observed.

Acknowledgments

Eldad Haber provided invaluable guidance on the discretization of the partial differential equation, as well as provided many tips for increasing inversion resolution through regularization techniques. The three anonymous reviewers also provided many helpful ideas and edits.

References

- [1] William Daily, Abelardo Ramirez, and Andrew Binley. Electrical resistance tomography. *The Leading Edge*, 2:438–442, 2004.
- [2] Richard C Aster, Brian Borchers, and Clifford Thurber. *Parameter Estimation and Inverse Problems*. Elsevier Inc., 2004.
- [3] Adam Pidlisecky, Eldad Haber, and Rosemary Knight. RESINVM3D : A 3D resistivity inversion package. *Geophysics*, 72(2), 2007.
- [4] Esben Auken and Anders Vest Christiansen. Layered and laterally constrained 2D inversion of resistivity data. *Geophysics*, 69(3):752–761, 2004.
- [5] Eldad Haber, U.M. Ascher, and D.W. Oldenburg. On optimization techniques for solving nonlinear inverse problems. *Inverse Problems*, 16:1263–1280, 2000.
- [6] Eldad Haber. *Computational Methods for Simulation Based Optimization*. UBC Math, 2011.
- [7] A.N. Tikhonov and V.Y. Arsenin. *Solutions of Ill-Posed Problems*. W.H. Winston and Sons., 1977.
- [8] Aleksandr Aravkin, Michael P Friedlander, and Tristan van Leeuwen. Robust inversion via semistochastic dimensionality reduction. *October*, October 2011.
- [9] CG Farquharson. Nonlinear inversion using general measures of data misfit and model structure. *Geophysical Journal*, 134:213–227, 1998.
- [10] Richard H Byrd, Jorge Nocedal, and Robert B Schnabel. Representations of quasi-Newton matrices and their use in limited memory methods. *Mathematical Programming*, 63:129–156, 1994.
- [11] Larry Armijo. Minimization Of Functions Having Lipschitz Continuous First Partial Derivatives. *Pacific Journal of Mathematics*, 16(1):1–3, 1966.
- [12] Jorge Nocedal and Stephen J Wright. *Numerical Optimization*. Springer New York, New York, NY, 1 edition, 1999.