Iterative reconstruction algorithm for non-linear operators

Robert A. Eso, Scott Napier, Felix J. Herrmann and Douglas W. Oldenburg, University of British Columbia*

SUMMARY

Iterative soft thresholding of a models wavelet coefficients can be used to obtain models that are sparse with respect to a known basis function. We generate sparse models for non-linear forward operators by applying the soft thresholding operator to the model obtained through a Gauss-Newton iteration and apply the technique in a synthetic 2.5D DC resistivity crosswell tomographic example.

INTRODUCTION

Given a linear (or nonlinear) mapping G between the model space $\mathbf{m}$ and the data space $\mathbf{d}$

$$G\mathbf{m} = \mathbf{d} \tag{1}$$

the goal of geophysical inversion is to recover a model (or suite of models) from some set of observations $\mathbf{d}$. It is often the case that the inverse problem of equation 1 is ill-posed, and solutions must be regularized. Additionally, the data image $\mathbf{d}$ is usually contaminated by noise. To overcome these difficulties, the inverse problem is formulated as an optimization problem in which an estimate $\mathbf{m}_{rec}$ of the true model $\mathbf{m}$ is obtained by minimizing the penalty function

$$\phi(\mathbf{m}) = \phi_d(\mathbf{m}) + \gamma R(\mathbf{m}), \tag{2}$$

where γ is the Tihkonov-regularization parameter, ϕ_d measures the discrepancy between the observed and predicted measurements and $R(\mathbf{m})$ is the regularization function measuring the norm, semi-norm or some combination of the model space $\mathbf{m}$.

The form of the regularization functional $R(\mathbf{m})$ effects the character of the recovered model. Using an ℓ_2 measure on the semi-norm $\iint \frac{\partial}{\partial_x} m \approx W_x m$, for example, can result in reconstructions that exhibit smoothly varying structure. In some instances it may be advantageous to recover models that exhibit a different type of character, namely piecewise blocky. There exist various focusing inversion strategies such as general measures (Farquharson and Oldenburg 1998), minimum support (Zhdanov et al. 2006) and total variation (Vogel and Oman 1998) to achieve this type of structure in an inversion result.

Here we will use the *a priori* knowledge that our desired model is sparse with respect to a known basis function ψ. Daubechies et al. (2004) show that the solution to the ℓ_1 minimization problem

$$\phi(\mathbf{m}) = \|G\mathbf{m} - \mathbf{d}\|_2^2 + \gamma \sum_i |\langle \mathbf{m}, \psi_i \rangle| \tag{3}$$

is obtained through soft thresholding the Landweber iteration

$$\mathbf{m}_{i+1} = S_\mu \left(\mathbf{m}_i + G^\mathsf{T}(\mathbf{d}^{obs} - G\mathbf{m}) \right) \tag{4}$$

which generates models exhibiting very few, but large coefficients of the basis function ψ and thus is sparse in the basis ψ. The nonlinear soft thresholding operator S_μ is defined as

$$S_\mu(x) = \begin{cases} x + \frac{\mu}{2} & \text{if } x \leq -\frac{\mu}{2} \\ 0 & \text{if } |x| < \frac{\mu}{2} \\ x - \frac{\mu}{2} & \text{if } x \geq -\frac{\mu}{2} \end{cases} \tag{5}$$

Also note that the forward and inverse projections onto the ψ basis are encapsulated within the soft thresholding operator S_μ. The step in equation 4 is a Landweber descent iteration.

This methodology has been applied to the interpretation of linear geophysical measurements (Loris et al. 2007). Here we apply the idea of obtaining a sparse representation of the model to nonlinear operators by soft thresholding on the resulting model obtained by a Gauss-Newton step. The basis function ψ in this instance is that of the Haar-wavelet.

NONLINEAR INVERSION

We begin by writing the forward problem as the general linear system

$$A(\mathbf{m})\mathbf{u} = \mathbf{b}, \tag{6}$$

where $A(\mathbf{m})$ is a sparse matrix that depends on the model $\mathbf{m}$. The vector $\mathbf{u}$ is the resulting field due to the source $\mathbf{b}$. The observed data is a projection of the field $\mathbf{d}^{obs} = Q\mathbf{u} + \varepsilon_i$ where the matrix Q is an interpolation operator and ε_i is assumed to be uncorrelated Gaussian noise.

The objective function to be minimized is written as the constrained optimization problem

$$\phi(\mathbf{m}) = \left\| W_d(Q\mathbf{u} - \mathbf{d}^{obs}) \right\|_2^2 + \gamma R(\mathbf{m}) \tag{7a}$$

$$\text{s.t.} \quad A(\mathbf{m})\mathbf{u} - \mathbf{b} = 0. \tag{7b}$$

To complete the minimization using descent-based algorithms, it is necessary to obtain or estimate the Fréchet derivative $J = \partial_{\mathbf{m}} Q\mathbf{u}$. Here we will absorb the data-weighting matrix W_d into J, $Q\mathbf{u}$ and $\mathbf{d}^{obs}$.

Gauss-Newton using ℓ_2 measure

It is common to regularize using the ℓ_2 norm to measure the models size and complexity, which results in the quadratic objective function

$$\phi(\mathbf{m}) = \left\| Q\mathbf{u} - \mathbf{d}^{obs} \right\|_2^2 + \gamma \|W_m(\mathbf{m} - \mathbf{m}_{\mathbf{ref}})\|_2^2, \tag{8}$$

where $W_{\mathbf{m}}$ is a model weighting matrix. To minimize equation 8, we use a Gauss-Newton strategy with a cooling schedule (Haber et al. 2000) in which the approximate Hessian

$$B = J^\mathsf{T} J + \gamma W_{\mathbf{m}}^\mathsf{T} W_{\mathbf{m}} \approx H \tag{9}$$

is used to obtain model perturbations through solutions of

$$B\delta\mathbf{m} = -\mathbf{g} \tag{10}$$

where $\mathbf{g} = \nabla_{\mathbf{m}}\phi(\mathbf{m})$ is the gradient of the objective function. It is often the case in geophysical inverse problems that the formation of $J^{\mathsf{T}}J$ (or even J for that matter) is prohibitively large. This can be overcome by forming the augmented system

$$\begin{pmatrix} J \\ \sqrt{\gamma}W_m \end{pmatrix} \delta\mathbf{m} = \begin{pmatrix} Q\mathbf{u} - \mathbf{d}^{obs} \\ -\sqrt{\gamma}W_m(\mathbf{m_k} - \mathbf{m}_{ref}) \end{pmatrix} \tag{11}$$

and solving the normal equations using the LSQR technique. The model is updated via $\mathbf{m}_{i+1} = \mathbf{m}_i + \alpha\delta\mathbf{m}$, where the parameter α is chosen using a line search.

General Measures

Following Farquharson (1998), we define the general measurement functional $\phi_g(m)$ to measure the size and complexity of the model in the inversion, where we can then choose the form of the measurement functional to be one that does not penalize large jumps in the model parameters as in the ℓ_2-norm. The general measurement functional is defined as

$$\phi_g(\mathbf{x}) = \sum_{j=1}^{N} \rho(x_j). \tag{12}$$

Given a vector $\mathbf{x} = W_i\mathbf{m}$, where W_i is a matrix, the corresponding gradient and Hessian of the general measure function $\nabla_m\phi_g(x)$ and $\nabla_m\nabla_m\phi_g(x)$, and are given by

$$\mathbf{g_g} = \nabla_m\phi_g(W_i\mathbf{m}) = W_i^{\mathsf{T}}\rho' \tag{13a}$$

$$H_g = \nabla_m\nabla_m\phi_g(W_i\mathbf{m}) = W_i^{\mathsf{T}}\rho'' W_i \tag{13b}$$

where $\rho' = (\partial_{x_1}\rho(x_1), \ldots, \partial_{x_j}\rho'(x_j))^{\mathsf{T}}$ and $\rho'' = diag(\partial_{xx}\rho(x_i))$. Using this general measure function in equation 12, the objective function to be minimized is

$$\phi(\mathbf{m}) = \phi_d(\mathbf{m}) + \gamma\phi_g(\mathbf{m}) \tag{14}$$

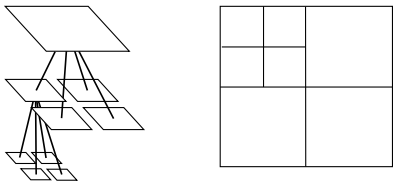
and is solved using a Gauss-Newton strategy, with the approximate Hessian given by

$$B = J^{\mathsf{T}}J + W_i^{\mathsf{T}}\rho'' W_i \approx H. \tag{15}$$

This approximate expression is used to obtain model perturbations through $B\delta\mathbf{m} = -\mathbf{g}$. The gradient $\mathbf{g}$ of the objective function is

$$\mathbf{g} = J^{\mathsf{T}}(Q\mathbf{u} - \mathbf{d}^{obs}) + \gamma W_i^{\mathsf{T}}\rho' \tag{16}$$

and the model is updated using a line-search, $\mathbf{m}_{i+1} = \mathbf{m}_i + \alpha\delta\mathbf{m}$. The form of the measuring function $\rho(x)$ in equation 12



a.) b.)

Figure 1: Data structure for a simple quadtree a.) hierarchal relationship b.) discretization

is chosen to be the Huber-norm:

$$\rho(x) = \begin{cases} x^2 & |x| \le c, \\ 2c|x| - c^2 & |x| > c \end{cases} \tag{17}$$

whereby the parameter c is chosen to vary the behavior of the norm between an ℓ_1-like behavior or an ℓ_2-like behavior.

Iterative sparse recovery

The Landweber iteration in equation 4 can be extended to nonlinear operators operators (Ramlau 1999). Soft thresholding the nonlinear Landweber iteration

$$\mathbf{m}_{i+1} = S_\mu\left(\mathbf{m}_i + J^{\mathsf{T}}(\mathbf{d}^{obs} - Q\mathbf{u})\right). \tag{18}$$

However this type of iteration can exhibit slow convergence properties. This nonlinear Landweber iteration can be accelerated by incorporating the curvature information of the Newton-step (Egger 2005). Using this approach, we chose to perform the soft thresholding on the wavelet coefficients

$$\mathbf{m}_{i+1} = S_\mu\left(\mathbf{m}_i + \alpha\delta\mathbf{m}\right), \tag{19}$$

where the step $\delta\mathbf{m}$ is obtained through the Gauss-Newton equation $B\delta\mathbf{m} = -\mathbf{g}$ corresponds to either the ℓ_2-norm solution, or the Huber-norm solution. The basic algorithm for this methodology is as follows:

Algorithm 1 nonlinear soft thresholding

1: μ=μ_0
2: $\mathbf{m}_0 = \mathbf{m}_{start}$
3: $\gamma = \gamma_0$
4: **while** not done **do**
5: $\quad\delta\mathbf{m} = -B^{-1}\mathbf{g}$
6: $\quad$linesearch $\alpha \rightarrow \tilde{\mathbf{m}}_i = \mathbf{m}_i + \alpha\delta\mathbf{m}$
7: $\quad\mathbf{m}_{i+1} = S_\mu(\tilde{\mathbf{m}}_i)$
8: $\quad\mu = \eta_1\mu$
9: $\quad\gamma = \eta_2\gamma$
10: **end while**

Here the parameters $\eta_1, \eta_2 < 1$ are parameters that control the cooling schedules of the tradeoff parameter γ and the soft thresholding parameter μ. We note that this algorithm is similar to that proposed by Ramlau and Teschke (2006). The cooling of the tradeoff parameter can be done continuously each iteration, or in stages.

2.5D DC RESISTIVITY

DC resistivity is a widely used geophysical imaging technique to investigate the Earth's electrical conductivity structure. DC resistivity experiments involve injecting a controlled DC current source at two locations and making measurements of the resulting potential field at locations away from the injection location.

The current distribution due to a point source is inherently 3D, but it is common to restrict the conductivity model to being 2D. The resulting 2.5D Helmholtz equation describing the electric potential due to a point source at the location $\delta(r - r_0)$ is described in Dey & Morrison (1977) and is

$$\nabla\cdot\{\sigma(x,z)\nabla\tilde{u}(x,k_y,z)\} - k_y^2\sigma(x,z)\tilde{u}(x,k_y,z) = \frac{I}{2}\delta(r - r_0), \tag{20}$$

where $\tilde{\mathbf{u}}$ is the transformed-potential in the Fourier-cosine domain, k_y is the wave-number and $\sigma(x,z)$ is the 2D conductivity model. The potential $\mathbf{u}$ in the spatial domain is obtained by an application of the inverse-Fourier transformation

$$\mathbf{u}(x,y,z) = F^{-1}[\tilde{\mathbf{u}}(x,k_y,z)] = \frac{2}{\pi}\int_0^\infty \tilde{\mathbf{u}}(x,k_y,z)cos(k_y y)\mathrm{dk_y} \tag{21}$$

We solve equation 20 using a finite-volume technique using a quadtree discretization, similar to the finite volume octree formulation described in Haber & Heldmann (2007). The quadtree is a tree-based data structure and allows for very efficient spatial discretization by allowing cells to have multiple-neighbors. Figure 1 shows a simple quadtree displayed as a hierarchial data structure in figure 1a and the corresponding discretization in figure 1b. Using this discretization we derive finite-volume

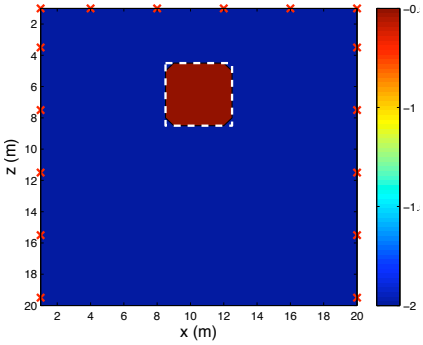


Figure 2: Synthetic geometry for DC resistivity crosswell tomography. Red crosses indicate the location of electrodes used in the survey acquisition.

approximations to the differential operators $\nabla\cdot$, and ∇. Our formulation places the discrete electrical potential u at cell centers, and current densities J at the center of cell edges. Additionally, the electrical conductivity is harmonically averaged and located at cell-edges. The resulting sparse-linear system is written as

$$A(k_y,\mathbf{m})\tilde{\mathbf{u}} = \mathbf{b} \tag{22}$$

where $\tilde{\mathbf{u}}$ is the discrete potential in the Fourier-cosine domain.

Electrical conductivity is known to range over several orders of magnitude in Earth materials thus we chose to invert for the model $m = log(\sigma)$ in the inversion. The Fréchet derivative $\partial_{\sigma_i} u_j$ is calculated using the adjoint-Green's function method described in McGillivray and Oldenburg (1990) and Zhou and Greenlaugh (1999). We efficiently calculate the Fréchet derivative by evaluating

$$\frac{\partial \mathbf{u}}{\partial \mathbf{m}} = -\frac{I\sigma}{2}F^{-1}\left[\int_{\Omega_i}\left(\nabla\tilde{\mathbf{G}}\cdot\nabla\tilde{\mathbf{G}}' + k_y^2\tilde{\mathbf{G}}\tilde{\mathbf{G}}'\right)\mathrm{dxdz}\right], \tag{23}$$

where $\tilde{\mathbf{G}}$ and $\tilde{\mathbf{G}}'$ are the Green's functions of the partial differential equation 20. Note that in the 2.5D case the potentials are not the Green's function of the equation 20, but are related to the transformed potential $\tilde{u}$ $\tilde{\mathbf{G}} = 2\tilde{\mathbf{u}}$. Computing the Fréchet derivative using the adjoint Green's function method typically requires on the order of N solutions to the forward problem, where N is the number of electrodes deployed in DC resistivity survey. This is a substantial computational savings from the $M+1$ solutions to the forward problem required to compute the Fréchet derivative through finite-difference techniques where M is the number of model parameters.

Numerical Implementation

DC resistivity measurements are typically collected using quadrapole configurations in which 4 electrode locations are used for each measurement, with two electrode locations being the transmitter (i.e. current sources) and two electrodes being used for the receivers (i.e. a voltmeter). The linear nature of equation 20 with respect to the potential allows any quadrapole measurement to be made through the super position of the pole-pole measurements. Typical DC resistivity surveys will deploy a small number of fixed electrode locations and then will make many quadrapole measurements. Often the number of quadrapole measurements is much greater than the number of electrodes making numerical solutions via pole-pole substitution an efficient numerical strategy.

The sparse linear system in 22 is solved directly using an LU decomposition. The LU decomposition is stored for each wave-number k_y and used to compute the transformed potential $\tilde{\mathbf{u}}$.

SYNTHETIC CROSSWELL TOMOGRAPHY

A synthetic crosswell DC resistivity tomography is used to demonstrate the DC resistivity experiment. The experimental setup is shown in Figure 2. Electrodes are deployed in two boreholes separated by $20\,m$ with each borehole extending to a depth of $20\,\mathrm{m}$. There are 5 electrode locations distributed equally in each borehole and an additional 6 electrodes distributed equally along the surface, resulting in a total of 16 electrode locations. These 16 electrode locations have the capability of generating 5460 unique non-reciprocal quadrapole measurements (Noel and Xu, 1991). The survey configuration used in this example contains transmitter dipoles made up of a combination of common-hole, cross-hole and borehole-to-surface transmitter configurations. The resulting dataset contains 4186 measurements. The synthetic model consists of a $10\,\mathrm{S/m}$ conductive body with dimensions of $5\,\mathrm{m}\times 4\,\mathrm{m}$ in a background halfspace of $0.01\,\mathrm{S/m}$.

Results

The inversion result using an ℓ_2 model norm and the Gauss-Newton methodology is shown in Figure 3. Although the block is clearly detected, the resulting model shows a smoothly varying structure typical of ℓ_2 inversion results.

The result of an inversion using the Huber-norm measure of model structure and the Gauss-Newton inversion is shown in Figure 4. This model exhibits more structure and steeper-gradients between the conductive body and the background.

Finally, the model obtained using iterative soft thresholding on the Gauss-Newton model with the ℓ_2-norm is shown in figure 5, while the model obtained by iteratively soft thresholding the Huber-norm Gauss-Newton model is shown in Figure 6. The Huber-norm in this case is applied only to the model norm

in the objective function. The data-misfit term is calculated using an ℓ_2-norm. Both of the models obtained through iterative soft thresholding are very similar.

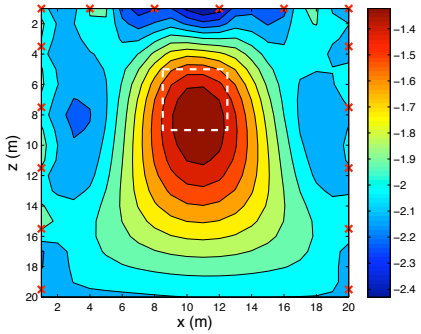


Figure 3: ℓ_2 Gauss-Newton inversion result

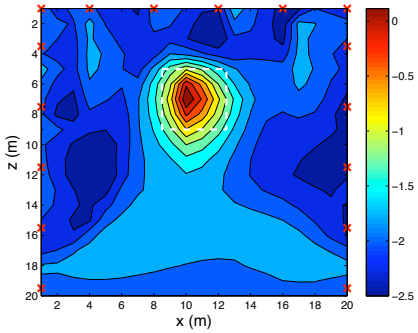


Figure 4: Huber-norm Gauss-Newton inversion result

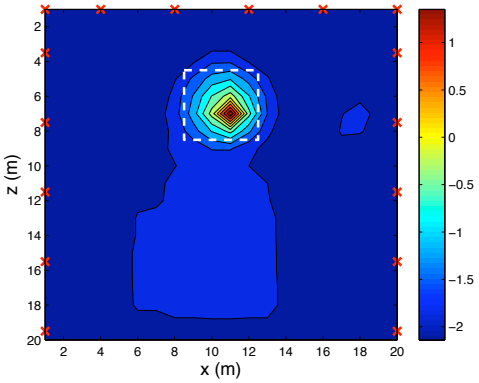


Figure 5: Iterative soft thresholding using the ℓ_2 Gauss-Newton step

The models are all plotted on different color scales corresponding to the log of the resistivity (i.e. $\log(\sigma^{-1})$). All the inversion models recover approximately the correct background halfspace of $0.01\,\mathrm{S/m}$. The ℓ_2 inversion result generates a model exhibiting a lower conductivity of the anomalous body than the true result. The behavior of all the recovered models show the trend that the more compact the body recovered in the inversion, a higher conductivity is required to reproduce the data.

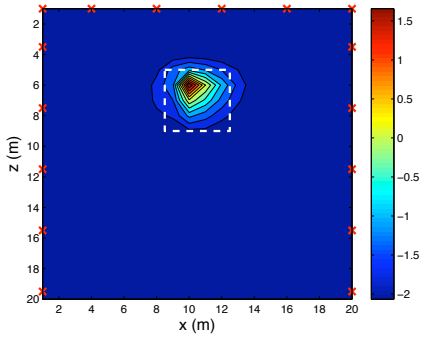


Figure 6: Iterative soft thresholding using the Huber-norm Gauss-Newton step

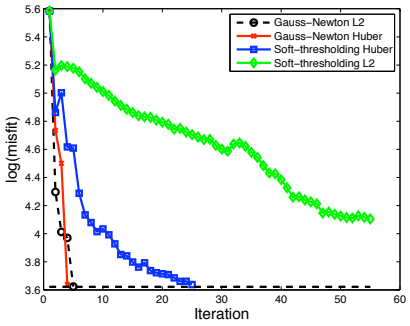


Figure 7: Convergence

The convergence of the 4 different imaging techniques is shown in Figure 7. The Gauss-Newton inversion using either the ℓ_2 or Huber-norms converges to the desired misfit rapidly, while the iterative soft thresholding techniques converge more slowly. The iterative soft thresholding performed on the ℓ_2 Gauss-Newton step did not converge within the maximum number of iterations at 55, but likely would have converged if allowed to go continue running.

CONCLUSION

We have applied iterative nonlinear soft thresholding to invert nonlinear geophysical operators. As expected, the resulting images exhibit blocky structure and are considerably more focused on the anomalous block than the ℓ_2 or Huber-norm inversion techniques. Soft thresholding on the Gauss-Newton model obtained through the Huber-norm converges quicker than soft thresholding on the Gauss-Newton model obtained using the ℓ_2 norm. We will continue to test and expand the iterative soft thresholding algorithms applied to nonlinear operators to improve the inversion images, and investigate the use of different numerical solvers such as SPgℓ_1. The eventual goal of interpreting DC resistivity measurements collected in the field with these imaging techniques.

REFERENCES

Daubechies, I., M. Defrise, and C. De Mol, 2004, An iterative thresholding algorithm for linear inverse problems with Pure Appl. Math, **57**, 1413–1541.

Dey, A. and H. Morrison, 1977, Resistivity modelling for arbitrarily shaped two-dimensional structures: Geophysical Prospecting, **27**, 106–136.

Egger, H., 2005, Accelerated newton-landweber iterations for regularizing nonlinear inverse problems: SFB-Report, **3**.

Farquharson, C. G. and D. W. Oldenburg, 1998, Non-linear inversion using general measures of data misfit and model structure: Geophys. J. Int, **134**, 213–227.

Haber, E., U. Ashcer, and D. Oldenburg, 2000, On optimization techniques for solving nonlinear inverse problems: Inverse Problems, **16**, 1263–1280.

Haber, E. and S. Heldmann, 2007, An octree multigrid method for quasi-static maxwell's equations with highly discontinuous coefficients: Journal of Computational Physics, **223**, 783–796.

Loris, I., G. Nolet, I. Daubechies, and F. Dahlen, 2007, Tomographic inversion using l1-norm regularization of wavelet coefficients: Geophys. J. Int., **170**, 359–379.

McGillivray, P. and D. Oldenburg, 1990, Methods for calculating frechet derivatives and sensitivities for the non-linear inverse problem: a comparative study: Geophysical Prospecting, **38**, 499–524.

Noel, M. and B. Xu, 1991, Archaeological investigation by electrical resistivity tomography: a preliminary study: Geophysical Journal International, **107**, 95–102.

Ramlau, R., 1999, A modified landweber-method for inverse problems: Journal for Numerical Functional Analysis and Optimization, **20**, 79–98.

Ramlau, R. and G. Teschke, 2006, A tikhonov-based projection iteration for nonlinear ill-posed problems with sparsity constraints: Numerische Mathematik, **104**, 177–203.

Vogel, C. and M. Oman, 1998, Fast total variation based reconstruction of noisy, blurred images: IEEE Trans. Image Processing, **7**, 813–824.

Zhdanov, M., G. Vignoli, and T. Ueda, 2006, Sharp boundary inversion in crosswell travel-time tomography: J. Geophys. Eng, **3**, 122–134.

Zhou, B. and Greenlaugh, 1999, Explicit expressions and numerical calculations for the frechet and second derivatives in 2.5d helmholtz equation inversion: Geophysical Prospecting, **47**, 443–468.