A Nonlinear Sparsity Promoting Formulation and Algorithm for Full Waveform Inversion

Aleksandr Aravkin[1], Tristan van Leeuwen[1], James V. Burke[2] and Felix Herrmann[1]
[1] Dept. of Earth and Ocean sciences University of British Columbia Vancouver, BC, Canada
[2] Dept. of Mathematics University of Washington Seattle, WA

January 14, 2011

Abstract

Full Waveform Inversion (FWI) is a computational procedure to extract medium parameters from seismic data. FWI is typically formulated as a nonlinear least squares optimization problem, and various regularization techniques are used to guide the optimization because the problem is ill-posed. In this paper, we propose a novel sparse regularization which exploits the ability of curvelets to efficiently represent geophysical images. We then formulate a corresponding sparsity promoting constrained optimization problem, which we call Nonlinear Basis Pursuit Denoise (NBPDN) and present an algorithm to solve this problem to recover medium parameters. The utility of the NBPDN formulation and efficacy of the algorithm are demonstrated on a stylized cross-well experiment, where a sparse velocity perturbation is recovered with higher quality than the standard FWI formulation (solved with LBFGS). The NBPDN formulation and algorithm can recover the sparse perturbation even when the data volume is compressed to 5% of the original size using random superposition.

Introduction

Full Waveform Inversion (FWI) is a data-fitting procedure based on full wavefield modeling designed to extract medium parameters (velocity and density) from seismograms. Computational methods for waveform inversion go back more than 20 years (see, e.g, Tarantola (1984)) and the problem has been consistently formulated as a nonlinear least squares or similar type of optimization problem (Virieux and Operto (2009)). It is useful at this point to provide an explicit framework for a typical FWI approach:

$$\min_m \quad \phi(m) := \|D - PH[m]^{-1}Q\|_F^2 \tag{1}$$

where $\|\cdot\|_F^2$ is the Frobenius norm, m is a vector of velocity parameters in a 2D or 3D grid, $D \in \mathbf{R}^{k\times l}$ contains results of l source experiments (as k-dimensional columns), $H[m]$ is a discretization of the Helmholtz operator with boundary conditions, $Q \in \mathbf{R}^{p\times l}$ specifies l source experiments, $H^{-1}[m]Q$ describes the solution of the Helmholtz equation for the sources Q, and P is a restriction of this solution to the surface where the data was observed.

FWI is widely known to be an ill-posed problem, and so regularization strategies are applied in practice (see Virieux and Operto (2009) and sources within). A common strategy is least squares regularization, where given reasonable guess of prior parameters m^*, one solves the problem

$$\min_m \quad \phi(m) + (m - m^*)^T W(m - m^*) \tag{2}$$

where W is some weighting matrix that encodes the confidence in the prior guess m^* as well as relationship (correlations) between the parameters. Alternative models using total variation (TV) regularization have also been proposed (see e.g. Vogel and Oman (1996)).

The aim of the present paper is to formulate an alternative regularization approach based on sparsity

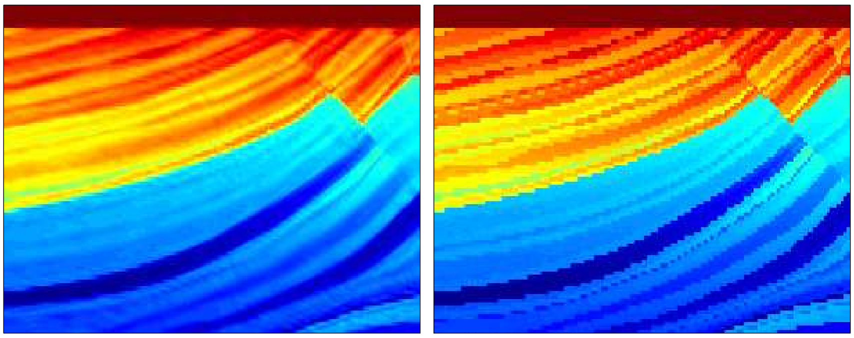

Figure 1 *Partial Marmoussi model in curvelets: 5% of the largest curvelet coefficients capture most of the features of the full representation.*

promotion, to develop an algorithm for the solution of the resulting optimization problem, and to show the results of the new algorithm on a toy example.

Sparsity promotion for seismic data using curvelets

The curvelet frame was presented as an effective nonadaptive representation for objects with edges in the seminal paper Candes and Donoho (2000). The key result in that paper is that the curvelet frame provides a stable, efficient, and near-optimal representation of otherwise smooth objects having discontinuities along smooth curves. While there may be limitations to this characterization of geophysical images, it is important to note that such images are layered due to geological sedimentation, and this feature allows for efficient representations using curvelets. Motivated by this observation, researchers have used curvelet representations in migration, dimensionality reduction, simulation, and sparse sampling applications (see Hennenfent et al. (2010); Herrmann et al. (2009, 2008, 2007)). See also Figure 1 for a simple demonstration.

The notion that velocity parameters should be sparse (or at least compressible) in the curvelet representation leads to a 'sparse' regularization of FWI (compare with eq. (2)):

$$\min_{x} \quad \phi(C^{*}x) + \lambda \|x\|_1, \tag{3}$$

where C denotes the curvelet basis and x is the vector of curvelet coefficients corresponding to the velocity parameters m, i.e. $m = C^{*}x$, the term $\|x\|_1$ serves to promote sparsity in this representation, and λ is a parameter that balances sparsity in curvelets vs. model fit. While this is a reasonable formulation, λ must be known ahead of time, and it is not clear how to choose it. Rather than working with eq. (3), we go to a closely related constrained reformulation

$$\begin{aligned} \min_{x} \quad & \|x\|_1 \\ \text{s.t.} \quad & g(C^{*}x) \leq \sigma, \end{aligned} \tag{4}$$

where as before the objective $\|x\|_1$ serves to promote sparsity in this representation, and the parameter σ is a regularization parameter that determines the acceptable value of the residual $\|D - RH[C^{*}x]^{-1}Q\|_F$ (i.e. noise level in the data). Unlike λ in eq. (3), the parameter σ in eq. (4) is likely to be known to scientists working with inverse problems in geophysics. Note that the formulation (4) is a natural nonlinear extension to the Basis Pursuit Denoise (BPDN) formulation used in compressive sensing literature to for sparse signal recovery from under-sampled noisy data (see van den Berg and Friedlander (2008)). The optimization formulation (4) is harder to solve then (1), and requires a custom algorithm. The main contribution of this paper is to describe such an algorithm and demonstrate its performance on a simplified problem of the form

$$\begin{aligned} \min_{m} \quad & \|m\|_1 \\ \text{s.t.} \quad & \|D - PH[m_0 + m]^{-1}Q\|_F \leq \sigma, \end{aligned} \tag{5}$$

for a situation where we are trying to recover a velocity perturbation m relative to a constant background velocity m_0, and so the perturbation is sparse in the physical domain. The algorithm for (5) can also be used to solve (4), but requires special care to maintain the feasibility of transformed velocities C^*x.

Nonlinear Basis Pursuit Denoise (NBPDN) algorithm

To solve (5), we implement an iterated algorithm of the form

$$m^{\nu+1} = m^{\nu} + \tau_{\nu} s^{\nu}, \tag{6}$$

where s^{ν} is the solution to a particular subproblem at step ν, and τ_{ν} is a step size chosen by a line search strategy. In developing the algorithm, especially in the line search for τ_{ν}, we follow ideas presented in Burke (1989) and Burke (1992). To obtain the subproblem, at each step ν, we linearize the functions m and $D - RH[m_0 + m]^{-1}Q$, and solve the resulting optimization problem, using a ν-dependent parameter σ_{ν}:

$$\begin{array}{ll} \min\limits_{s} & \|m^{\nu} + s\|_1 \\ \text{s.t.} & \|D - \mathscr{F}(m^{\nu}) - \nabla\mathscr{F}(m^{\nu})(s)\|_F \leq \sigma_{\nu}, \end{array} \tag{7}$$

where $\mathscr{F}(m) = PH[m_0 + m]^{-1}Q$ and $\nabla\mathscr{F}(m^{\nu})$ denotes the linearized Born scattering operator. The solution to this problem is the direction s^{ν} that appears in eq. (6). To solve this problem, we use the substitution $y = m^{\nu} + s$ to obtain

$$\begin{array}{ll} \min\limits_{y} & \|y\|_1 \\ \text{s.t.} & \left\| \left(D - \mathscr{F}(m^{\nu}) + \nabla\mathscr{F}(m^{\nu}) m^{\nu} \right) - \nabla\mathscr{F}(m^{\nu})(y) \right\|_F \leq \sigma_{\nu}. \end{array} \tag{8}$$

For a fixed m^{ν}, this problem is now equivalent to the basis pursuit denoise (BPDN) problem detailed in van den Berg and Friedlander (2008). The algorithm in that paper, called SPGℓ_1, allows us to solve (8) quickly, and moreover allows a functional representation of $\nabla\mathscr{F}$ to be provided (specifying its action on vectors y) rather than requiring an explicit matrix representation. The parameters σ_{ν} are chosen to start large and decrease untill it reaches the σ parameter specified by the user. To obtain the step parameter τ_{ν}, we first define an auxiliary penalty function

$$P_{\alpha}(m) = \|m\|_1 + \alpha \left(\|D - \mathscr{F}(m)\|_2 - \sigma \right)_+,$$

which includes both the sparsity promoting objective $\|m\|_1$ and a measure of the distance from optimality. The parameter α_{ν} is then selected to ensure that s^{ν}, the solution to (5), is a descent direction for $P_{\alpha_{\nu}}(m)$. In other words, the choice of α_{ν} ensures $P_{\alpha_{\nu}}(m^{\nu} + s^{\nu}) - P_{\alpha_{\nu}}(m^{\nu}) < 0$. We then use the backtracking Armijo line search (see e.g. Nocedal and Wright (1999)) using the merit function $P_{\alpha_{\nu}}(m)$. The resulting step τ_{ν} is used to update the model as described in (6).

Results

To illustrate the new algorithm and the power of sparsity regularization, we considered a stylized cross-well problem. The true velocity consists of three small features embedded in a constant background of 2 km/s and is depicted in figure 2. The features are sparse in the pixel-basis so we can directly enforce sparsity on the recovered perturbation. We use a 9-point discretization of the Helmholtz operator with absorbing boundaries on a grid with 10 m spacing. The data are generated for 101 equi-spaced sources and receivers located in vertical wells 800 m apart for (randomly chosen) frequencies $[5.0, 6.0, 11.5, 14.0, 15.5, 17.5, 23.5]$ Hz. We consider two different scenarios: inversion with *all* the sources and inversion using only 5 randomly synthesized 'supershots'. These are generated by weighting all the sources with random Gaussian weights and stacking. Such techniques have recently been proposed to dramatically reduce the costs of FWI (Krebs et al., 2009; Moghaddam and Herrmann, 2010; Haber et al., 2010) (see also other contributions of the authors to these proceedings). We compare the use of unregularized L-BFGS on (1) and the newly proposed algorithm on (8). The results are depicted in figure 2. We see that when using all the sources the unregularized approach produces a reasonable image. The resolution is not very high, as expected, and the vertical sides of the circle are not well recovered. The NBPDN algorithm, however, produces a nearly exact recovery. The circle is now recovered completely but the horizontal bar is somewhat distorted. When using only 5 supershots (a reduction of a factor 20 in the data volume) the L-BFGS approach produces an unusable image. The artifacts introduced by the crosstalk between the shots completely obscures the recovered velocity perturbations. The NBPDN formulation, remarkably, gives us almost the same result as before.

Conclusion

We formulated FWI as a non-linear, sparsity promoting optimization problem. The underlying assumption is that the medium parameters that we are trying to recover have a sparse representation in some basis. In particular, we envision that typical velocity structures are sparse in curvelets. Instead of adding a penalty term to the misfit term with a regularization parameter, as is commonly done in for example TV regularization, we propose to minimize the penalty subject to the misfit being smaller than some preset error level. The advantage of this formulation compared to other sparsity promoting strategies (e.g. LASSO) is that this error level may be easier to determine than the regularization parameter.

We demonstrate the algorithm on a toy cross-well example, where the unknown velocity perturbation is sparse in the pixel-basis. Compared to an unregularized least-squares inversion, our approach gives a superior result with much higher resolution. We also consider using randomly synthesized data to reduce the computational cost of the inversion. Such a reduction comes at the cost of introducing crosstalk between the shots. In the unregularized inversion, this crosstalk overshadows the reconstructed velocity perturbations. With the regularized inversion, however, we obtain a result nearly identical to the earlier case at roughly 5% of the computational cost. The latter result may be tentatively explained by invoking results from compressive sensing; a sparse signal may be reconstructed from severely undersampled data by solving a linear sparsity promoting program as long as the sampling satisfies some additional criteria. Most notably, the sampling must be random. The current formulation is a direct generalization of the sparsity promoting linear formulation used in compressive sensing. Future research will be aimed at further exploiting the connection to compressive sensing.

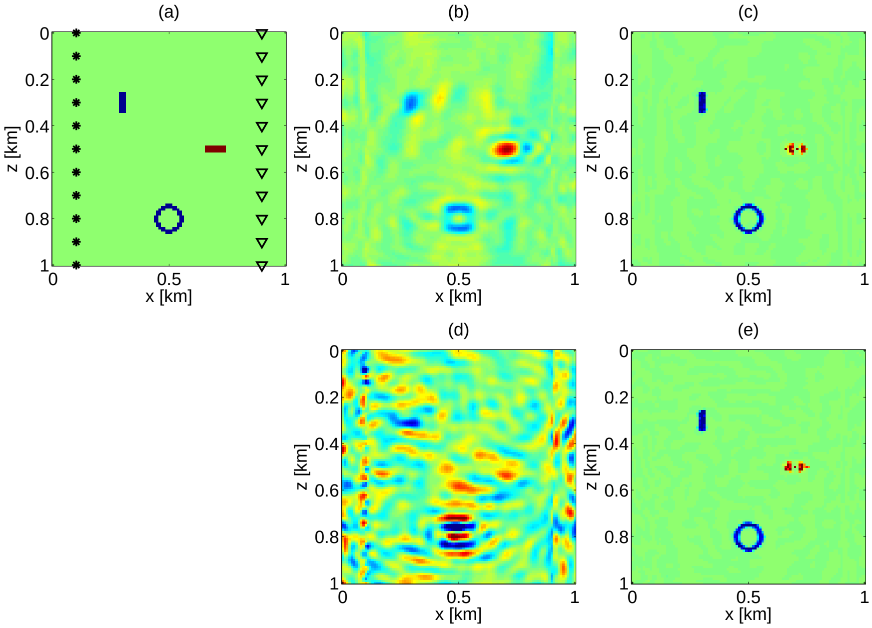

Figure 2 (a) True model for cross-well experiment; asterisks are sources and triangles are receivers. (b) LBFGS recovery using full data (101 shots). (c) NBPDN recovery using full data (101 shots). (d) LBFGS recovery using five supershots (20 x speedup). (e) NBPDN recovery using five supershots (20 x speedup).

References

Burke, J.V. [1989] A sequential quadratic programming method for potentially infeasible mathematical programs. *Journal of Mathematical Analysis and Applications*, **139**(2), 319–351.

Burke, J.V. [1992] A robust trust region method for constrained nonlinear programming problems. *SIAM Journal on Optimization*, **2**(2), 325–347, doi:10.1137/0802016.

Candes, E.J. and Donoho, D.L. [2000] *Curves and Surfaces*, Vanderbilt University Press, Nashville, TN., chap. Curvelets - a surprisingly effective nonadaptive representation for objects with edges.

Haber, E., Chung, M. and Herrmann, F.J. [2010] An effective method for parameter estimation with pde constraints with multiple right hand sides. Tech. Rep. TR-2010-4, UBC-Earth and Ocean Sciences Department.

Hennenfent, G., Fenelon, L. and Herrmann, F.J. [2010] Nonequispaced curvelet transform for seismic data reconstruction: a sparsity-promoting approach. Tech. Rep. TR-2010-2, UBC-Earth and Ocean Sciences Department.

Herrmann, F.J., Brown, C., Erlangga, Y. and Moghaddam, P. [2008] Curvelet-based migration preconditioning and scaling. Tech. Rep. TR-2008-10, The University of British Columbia, submitted for publication.

Herrmann, F.J., Erlangga, Y.A. and Lin, T. [2009] Compressive simultaneous full-waveform simulation. *Geophysics*, **74**, A35.

Herrmann, F.J., Moghaddam, P.P. and Stolk, C.C. [2007] Sparsity- and continuity-promoting seismic image recovery with curvelet frames. Accepted for publication in the Journal of Applied and Computational Harmonic Analysis.

Krebs, J.R. et al. [2009] Fast full-wavefield seismic inversion using encoded sources. *Geophysics*, **74**(6), WCC177–WCC188, doi:10.1190/1.3230502.

Moghaddam, P.P. and Herrmann, F.J. [2010] Randomized full-waveform inversion: a dimenstionality-reduction approach. *SEG Technical Program Expanded Abstracts*, **29**(1), 977–982, doi:10.1190/1.3513940.

Nocedal, J. and Wright, S.J. [1999] *Numerical Optimization*. Springer Series in Operations Research, Springer.

Tarantola, A. [1984] Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, **49**(8), 1259–1266, doi:10.1190/1.1441754.

van den Berg, E. and Friedlander, M.P. [2008] Probing the pareto frontier for basis pursuit solutions. *SIAM Journal on Scientific Computing*, **31**(2), 890–912, doi:10.1137/080714488.

Virieux, J. and Operto, S. [2009] An overview of full-waveform inversion in exploration geophysics. *Geophysics*, **74**, 127–+, doi:10.1190/1.3238367.

Vogel, C.R. and Oman, M.E. [1996] Iterative methods for total variation denoising. *SIAM J. Sci. Comput.*, **17**, 227–238, ISSN 1064-8275, doi:10.1137/0917016.