# Regularizing waveform inversion by projection onto intersections of convex sets.

Bas Peters[*]
Brendan Smithyman[#]
Felix J. Herrmann[*]
[*]Seismic Laboratory for Imaging and Modeling (SLIM), University of British Columbia.
[#]University of Western Ontario, Department of Earth Sciences.

## Abstract

A framework is proposed for regularizing the waveform inversion problem by projections onto intersections of convex sets. Multiple pieces of prior information about the geology are represented by multiple convex sets, for example limits on the velocity or minimum smoothness conditions on the model. The data-misfit is then minimized, such that the estimated model is always in the intersection of the convex sets. Therefore, it is clear what properties the estimated model will have at each iteration. This approach does not require any quadratic penalties to be used and thus avoids the known problems and limitations of those types of penalties. It is shown that by formulating waveform inversion as a constrained problem, regularization ideas such as Tikhonov regularization and gradient filtering can be incorporated into one framework. The algorithm is generally applicable, in the sense that it works with any (differentiable) objective function, several gradient and quasi-Newton based solvers and does not require significant additional computation. The method is demonstrated on the inversion of very noisy synthetic data and vertical seismic profiling field data.

## Introduction

Full-waveform inversion (FWI) involves solving a nonlinear optimization problem with the goal of matching observed data with predicted data by iteratively updating the estimate of the subsurface properties. This is an inverse problem and may require some form of regularization to yield solutions, which conform with our knowledge about the Earth at the survey location.

In this expanded abstract we propose a regularization strategy based on solving constrained problems by projection onto intersections of convex sets. We incorporate multiple pieces of prior information about the Earth into the FWI optimization scheme, in an approach similar to Baumstein (2013). This avoids the use of quadratic penalties. It is also shown how the constrained problem formulation incorporates several preexisting regularization ideas. Synthetic and field data examples show the effectiveness of the method.

### Waveform inversion with regularization by projection onto the intersection of convex sets

In search of a flexible and easy to use framework, which can incorporate most existing ideas in the geophysical community to regularize the waveform inversion problem, we propose to base our framework on solving constrained optimization problems of the form

$$\min_{\mathbf{m}} f(\mathbf{m}) \quad \text{s.t.} \quad \mathbf{m} \in \mathscr{C}_1 \bigcap \mathscr{C}_2. \tag{1}$$

The model vector, $\mathbf{m}$ (often velocity), and the objective function can measure the difference between predicted data and observed data, $f(\mathbf{m}) = \frac{1}{2}\|PH(\mathbf{m})^{-1}\mathbf{q} - \mathbf{d}\|_2^2$ as is often used in FWI. $H$ is the discretized Helmholtz equation, $P$ restricts the wavefield to the receiver locations, $\mathbf{q}$ is the vector containing source function and location and $\mathbf{d}$ is the observed data. It is also possible to use other norms and objective functions, for example the unconstrained objective function corresponding to the Wavefield Reconstruction Inversion (WRI) approach (van Leeuwen and Herrmann, 2013): $f(\mathbf{m}) = \frac{1}{2}\|P\bar{\mathbf{u}} - \mathbf{d}\|_2^2 + \frac{\lambda^2}{2}\|H(\mathbf{m})\bar{\mathbf{u}} - \mathbf{q}\|_2^2$ with $\bar{\mathbf{u}} = (\lambda^2 H^*H + P^*P)^{-1}(\lambda^2 H^*\mathbf{q} + P^*\mathbf{d})$.

$\mathscr{C}_1$ and $\mathscr{C}_2$ are convex sets (more than 2 sets can be used). Intuitively, the line connecting any two points in a convex set is also in the set: for all $x, y \in \mathscr{C}$ the following holds: $cx + (1-c)y \in \mathscr{C}$, $0 \leq c \leq 1$. Each set represents a known piece of information about the Earth; a constraint which the estimated model has to satisfy. The model is required to be in the intersection of the two sets, denoted by $\mathscr{C}_1 \bigcap \mathscr{C}_2$. The intersection of convex sets is also convex. To summarize, solving Problem 1 means we try to minimize the data misfit while satisfying constraints, which represent the known information about the geology. Below are two useful convex sets which are also used in the examples. Bound-constraints can be defined element-wise as

$$\mathscr{C}_1 \equiv \{\mathbf{m} \,|\, \mathbf{b}_l \leq \mathbf{m} \leq \mathbf{b}_u\}. \tag{2}$$

If a model is in this set, then every model parameter value is between a lower bound $\mathbf{b}_l$ and upper bound, $\mathbf{b}_u$. These bounds can vary per model parameter and can therefore incorporate spatially-varying information about the bounds. Bound constraints can also incorporate a reference model by limiting the maximum deviation from this reference model as: $\mathbf{b}_l = \mathbf{m}_{\text{ref}} - \delta\mathbf{m}$. The projection onto $\mathscr{C}_1$, denoted as $\mathscr{P}_{\mathscr{C}_1}$, is the elementwise median: $\mathscr{P}_{\mathscr{C}_1}(\mathbf{m}) = \text{median}\{\mathbf{b}_l, \mathbf{m}, \mathbf{b}_u\}$.

The second convex set that is used here is a subset of the wavenumber domain representation of the model (spatial Fourier transformed). This convex set can be defined as

$$\mathscr{C}_2 \equiv \{\mathbf{m} \,|\, E^*F^*(I-S)FE\mathbf{m} = 0\}. \tag{3}$$

This set is defined by a series of matrix multiplications and means that we apply the mirror-extension operator, $E$ to the model $\mathbf{m}$ to prevent wrap-around effects due to applying linear operations in the wavenumber domain. Next, the mirror extended model is spatially Fourier transformed using $F$. In this domain we require the wavenumber content outside a certain range to be equal to zero. This is enforced by the diagonal (windowing) matrix $S$. This represents information that the model should have a certain minimum smoothness. In case we expect an approximately horizontally layered medium without sharp contrasts, we can include this information by requiring the wavenumber content of the

model to vanish outside of an elliptical zone around the zero wavenumber point (see also Brenders and Pratt (2007)). In other words, more smoothness in one direction is required than in the other direction. The adjoint of $F$, $(F^*)$ transforms back from the wavenumber to the physical domain and $E^*$ goes from mirror-extended domain to the original model domain. The projection onto $\mathscr{C}_2$, denoted as $\mathscr{P}_{\mathscr{C}_2}$, is given by $\mathscr{P}_{\mathscr{C}_2}(\mathbf{m}) = E^* F^* S F E \mathbf{m}$.

A basic way to solve Problem 1 is to use the projected-gradient algorithm, which has the main step $\mathbf{m}_{k+1} = \mathscr{P}_{\mathscr{C}}(\mathbf{m}_k - \gamma \nabla_{\mathbf{m}} f(\mathbf{m}))$, with a step length parameter $\gamma$. $\mathscr{P}_{\mathscr{C}} = \mathscr{P}_{\mathscr{C}_1 \cap \mathscr{C}_2}$ is the projection onto the intersection of all convex sets the model is required to be in. "The projection onto" means we need to find the point that is in both sets and closest to the current proposed model estimate generated by $(\mathbf{m}_k - \gamma \nabla_{\mathbf{m}} f(\mathbf{m}))$. To find such a point, defined mathematically as

$$\mathscr{P}_{\mathscr{C}}(\mathbf{m}) = \arg\min_{\mathbf{x}} \|\mathbf{x} - \mathbf{m}\|_2 \quad \text{s.t.} \quad \mathbf{x} \in \mathscr{C}_1 \bigcap \mathscr{C}_2, \tag{4}$$

we can use a splitting method such as Dykstra's projection algorithm (Bauschke and Borwein, 1994). This algorithm finds the projection onto the intersection of the two sets, by projecting onto each set separately. This is a cheap and simple algorithm and enables projections onto complicated intersections of sets. The algorithm is given by Algorithm 1.

---

**Algorithm 1** Dykstra.

$x_0 = \mathbf{m}$, $p_0 = 0$, $q_0 = 0$
For $k = 0, 1, \ldots$
$\quad y_k = \mathscr{P}_{\mathscr{C}_1}(x_k + p_k)$
$\quad p_{k+1} = x_k + p_k - y_k$
$\quad x_{k+1} = \mathscr{P}_{\mathscr{C}_2}(y_k + q_k)$
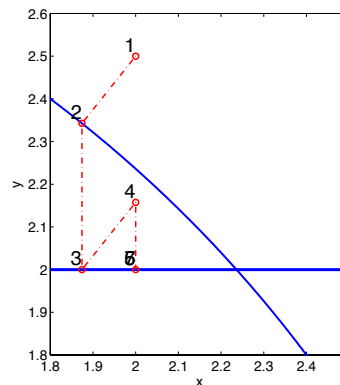$\quad q_{k+1} = y_k + q_k - x_{k+1}$
End

---



Figure 1: The trajectory of Dykstra's algorithm for a toy example with constraints $y \leq 2$ and $x^2 + y^2 \leq 3$. Iterates 5, 6 and 7 coincide; the algorithm converged to the point closest to point number 1 and satisfying both constraints. Note that the projection onto convex sets (POCS) algorithm would converge at point number 3 and is clearly unsuitable for this type of projection problem.

While the gradient-projection algorithm is a solid approach, it can also be relatively slowly converging. A potentially much faster algorithm is the class of quasi-Newton methods, which iteratively try to approximate the Hessian by using just gradient and function value information. However, in general it is not possible to straightforwardly project quasi-Newton steps onto a convex set. The use of second-order information can cause the algorithm to converge to a solution which does not correspond to Problem 1. There are slightly more complicated algorithms, which properly implement a second-order approach. We select the projected-quasi-Newton (PQN) algorithm by Schmidt et al. (2009). This algorithm finds a search direction which is in the intersection of the convex sets using the spectral projected gradient

algorithm (SPG) as a sub-problem of an l-BFGS-like algorithm. It does not simply project an update. This algorithm needs the objective function value, $f(\mathbf{m})$, its gradient and an algorithm to solve the projection Problem 4. PQN has a very similar computational cost as the standard l-BFGS algorithm, because the projections are cheap to compute. See Schmidt et al. (2009) for some examples of strong empirical performance of PQN versus projected-gradient for some non-geophysical examples. The final algorithm is given by Algorithm 2.

---

**Algorithm 2** Projected-Quasi-Newton based waveform inversion with projections onto convex sets.

1. define convex sets $\mathscr{C}_1, \mathscr{C}_2, \ldots$
2. Set up algorithm (for example Dykstra's) to find the projection onto the intersection of $\mathscr{C}_1, \mathscr{C}_2, \ldots$
3. $\min_{\mathbf{m}} f(\mathbf{m})$   s.t.   $\mathbf{m} \in \mathscr{C}_1 \bigcap \mathscr{C}_2$ using the PQN algorithm

---

**Example 1: Inversion of very noisy data**

This is a synthetic example where the goal is to invert 10 Hz data with $\|\text{noise}\|_2 / \|\text{signal}\|_2 = 0.3$. Inverting the noisy data will yield a noisy image, even if we use bound-constraints and a good start model. When the bound-constraints are used together with the projection onto set $\mathscr{C}_2$, the inverted result is quite satisfying. The projection method is very effective for this type of situation, because the noise in the data induces highly oscillatory components in the model, which are outside the spatial Fourier spectrum of the expected model. Therefore, those components are projected out and can never enter the model. When used in this way, the projection onto $\mathscr{C}_2$ can be interpreted as an image-domain noise filter. Besides the model-domain noise filter interpretation, the projection also keeps the model within the range of physically acceptable models. This restriction of the space, from $\mathbf{m} \in \mathbb{R}^N$ to $\mathbf{m} \in \mathscr{C}_1 \bigcap \mathscr{C}_2 \subset \mathbb{R}^N$ may prevent the optimization of $f(\mathbf{m})$ to get stuck in local minimizers. Note that $f(\mathbf{m})$ is unaltered by the projection approach, just the space where the model can be. The true model, initial model and results with and without projection onto $\mathscr{C}_2$ are shown in Figure 2
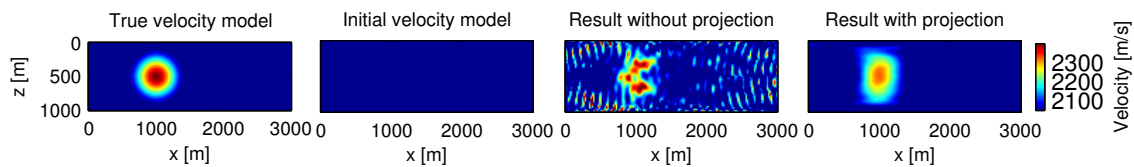


Figure 2: Models corresponding to Example 1, inversion of very noisy data, 10 Hertz only.

**Example 2: Land data in a vertical seismic profiling experiment**

Vertical Seismic Profiling (VSP) data was acquired with receivers on the surface and in part of the well. The sources are on the surface only. Available data has a frequency content of 8 to 15.75 Hertz and all frequencies were used simultaneously in one batch. No density, elastic and attenuation effects are accounted for in the modeling. No significant anisotropy is expected in this type of geology from the Permian Basin.

The results are obtained by solving Problem 1. The set $\mathscr{C}_2$ is specified as follows. First the spatial Fourier spectrum of the initial model is analyzed. Then the filter $S$ in Equation 3 is designed to allow the spatial Fourier spectrum of the initial model plus a bit higher spatial frequency content, based on what wavenumber spectrum can be expected from the startmodel in combination with the frequency content of the data. The results are shown in Figure 3.

**Relation to existing work, discussion and summary**

Conceptually, the approach presented is related to the work of Baumstein (2013). There too, projections onto convex sets are used to regularize the waveform inversion problem. The optimization, computation of projections, as well as the sets are different.

Tikhonov regularization is a well known regularization technique and adds quadratic penalties to an objective as $\min_{\mathbf{m}} f(\mathbf{m}) + \frac{\alpha}{2}\|R_1\mathbf{m}\|_2^2 + \frac{\beta}{2}\|R_2\mathbf{m}\|_2^2$, where two regularizers are used. $\alpha$ and $\beta$ are the
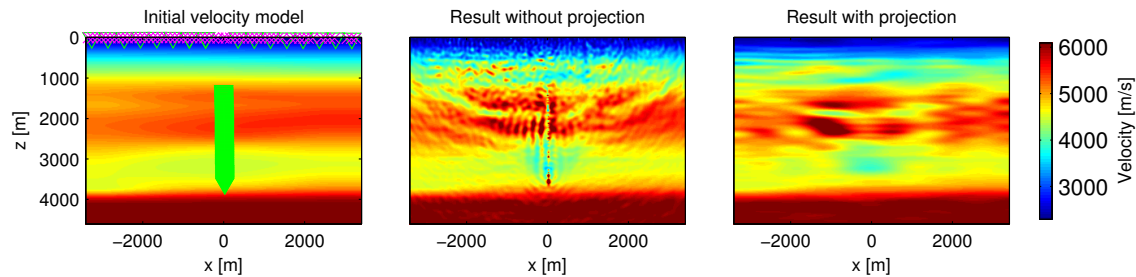
Figure 3: Models corresponding to example 2. Inversion of field data.

scalar weights, determining the importance of each regularizer. The matrices $R$ represent properties of the model, which we would like to penalize. This method should in principle be able to achieve similar results as the projection approach, but it has a few disadvantages. The first one is that it may be difficult or costly to determine the weights $\alpha$ and $\beta$. Multiple techniques for finding a 'good' set of weights are available, including the L-curve, cross-validation and the discrepancy principle, see Sen and Roy (2003) for an overview in a geophysical setting. A second issue with the penalty approach is the effect of the penalty terms on the condition number of the Hessian of the optimization problem, see Nocedal and Wright (2000). This means that certain choices of $\alpha$, $\beta$, $R_1$ and $R_2$ may lead to an optimization problem which is infeasibly expensive to solve. A third issue is related to the two-norm squared, $\| \cdot \|_2^2$ not being a so called 'exact' penalty function Nocedal and Wright (2000). This means that constraints can be approximated with the penalty functions, but generally not satisfied exactly. Note that the projection and quadratic penalty strategies for regularization can be easily combined as $\min_{\mathbf{m}} f(\mathbf{m}) + \frac{\alpha}{2}\|R_1\mathbf{m}\|_2^2 + \frac{\beta}{2}\|R_2\mathbf{m}\|_2^2$   s.t.   $\mathbf{m} \in \mathscr{C}_1 \bigcap \mathscr{C}_2$.

The concept of filtering the gradient is used by multiple researchers, for example Brenders and Pratt (2007). A gradient filter can be represented as applying a linear operation to the gradient step as $\mathbf{m}_{k+1} = \mathbf{m}_k - \gamma F \nabla_{\mathbf{m}} f(\mathbf{m})$. Brenders and Pratt (2007) apply a low-pass filter to the gradient to prevent high-wavenumber updates to the model while inverting low-frequency seismic data. A very similar effect is achieved by the framework proposed in this abstract and using the set defined in 3. The projection based approach has the advantage that it straightforwardly generalizes to multiple constraints on the model in multiple domains, whereas gradient filtering does not. Restricting the model to be in a certain space is quite similar to reparametrization of the model in, for example the (partial) spatial Fourier domain.

In summary, in this abstract we combine Dyksta's algorithm with the projected quasi-Newton algorithm in order to merge several regularization strategies for waveform inversion, such as Tikhonov regularization, gradient filtering and reparameterization into one framework, which is simple to use and extend. Although the projection onto the intersection of multiple convex sets strategy does not necessarily yield better results than other regularization strategies, it may be easier to obtain improved results, because the projection approach makes is clear what constraints are satisfied exactly at every iteration.

## References

Baumstein, A. [2013] Pocs-based geophysical constraints in multi-parameter full wavefield inversion. EAGE.

Bauschke, H. and Borwein, J. [1994] Dykstras alternating projection algorithm for two sets. *Journal of Approximation Theory*, **79**(3), 418 – 443, ISSN 0021-9045, doi:http://dx.doi.org/10.1006/jath.1994.1136.

Brenders, A.J. and Pratt, R.G. [2007] Full waveform tomography for lithospheric imaging: results from a blind test in a realistic crustal model. *Geophysical Journal International*, **168**(1), 133–151, doi:10.1111/j.1365-246X.2006.03156.x.

Nocedal, J. and Wright, S.J. [2000] *Numerical optimization*. Springer.

Schmidt, M., van den Berg, E., Friedlander, M. and Murphy, K. [2009] Optimizing costly functions with simple constraints: A limited-memory projected quasi-newton algorithm. JMLR, vol. 5, 456–463.

Sen, M. and Roy, I. [2003] Computation of differential seismograms and iteration adaptive regularization in prestack waveform inversion. *GEOPHYSICS*, **68**(6), 2026–2039, doi:10.1190/1.1635056.

van Leeuwen, T. and Herrmann, F.J. [2013] Mitigating local minima in full-waveform inversion by expanding the search space. *Geophysical Journal International*, **195**, 661–667, doi:10.1093/gji/ggt258.