Estimating Primaries by Sparse Inversion in a Curvelet-like Representation Domain

Tim T.Y. Lin and Felix J. Herrmann
EPSI Problem

Estimation of Primaries by Sparse Inversion (van Groenestijn and Verschuur, 2009)

recorded data predicted data from primary IR

\[ P = G(Q + RP) \]

- **P**: total up-going wavefield
- **Q**: down-going source signature
- **R**: reflectivity of free surface (assume -1)
- **G**: primary impulse response

(all monochromatic data matrix, implicit)
Estimation of Primaries by Sparse Inversion (van Groenestijn and Verschuur, 2009)

recorded data predicted data from primary IR

\[ P = G(Q + RP) \]

Inversion objective:

\[ f(G, Q) = \frac{1}{2} \| P - G(Q + RP) \|_2^2 \]
In time domain (lower-case: whole dataset in time domain)

recorded data  predicted data from primary IR

\[ p = \mathcal{M}(g, q) \]

\[ \mathcal{M}(g, q) := \mathcal{F}_t^\dagger \text{BlockDiag}_{\omega_1, \ldots, \omega_n} [(q(\omega)I - P)^\dagger \otimes I] \mathcal{F}_t g \]

Inversion objective:

\[ f(g, q) = \frac{1}{2} \| p - \mathcal{M}(g, q) \|_2^2 \]
EPSI Problem

Linearizations

\[ p = \mathcal{M}(g, q) \]

\[ M_{\tilde{q}} = \left( \frac{\partial \mathcal{M}}{\partial g} \right)_{\tilde{q}} \]

\[ M_{\tilde{g}} = \left( \frac{\partial \mathcal{M}}{\partial q} \right)_{\tilde{g}} \]

In fact it is bilinear:

\[ M_{\tilde{q}}g = \mathcal{M}(g, \tilde{q}) \quad M_{\tilde{g}}q = \mathcal{M}(q, \tilde{g}) \]
EPSI Problem

Linearizations

\[ p = M(g, q) \]

\[ M_{\tilde{q}} = \left( \frac{\partial M}{\partial g} \right)_{\tilde{q}} \]

\[ M_{\tilde{g}} = \left( \frac{\partial M}{\partial q} \right)_{\tilde{g}} \]

Associated objectives:

\[ f_{\tilde{q}}(g) = \frac{1}{2} \| p - M_{\tilde{q}} g \|_2^2 \]
\[ f_{\tilde{g}}(q) = \frac{1}{2} \| p - M_{\tilde{g}} q \|_2^2 \]

Thursday, June 16, 2011
EPSI Procedure

Do:

\[ g_{k+1} = g_k + \alpha \nabla f_{q_k}(g_k) \]
\[ q_{k+1} = q_k + \beta \nabla f_{g_{k+1}}(q_k) \]

Alternating updates (Gauss-Sidel) to the linearized problem
**EPSI Procedure**

Do:

\[ g_{k+1} = g_k + \alpha \mathcal{S}(\nabla f_{q_k}(g_k)) \]
\[ q_{k+1} = q_k + \beta \nabla f_{g_{k+1}}(q_k) \]

**Gradient sparsity**

\( \mathcal{S} : \text{pick largest } \rho \text{ elements per trace} \)
Related to two underlying sub-problems:

\[ \min_{g} \| p - M_{\tilde{q}}g \|_2 \quad \text{s.t.} \quad \text{nnz}(g) \leq \rho \]

\[ \min_{q} \| p - M_{\tilde{g}}q \|_2 \]

Which approximates:

\[ \min_{g} \text{nnz}(g) \quad \text{s.t.} \quad \| p - M_{\tilde{q}}g \|_2 \leq \sigma \]

(notion of sparsest solution)
EPSI Procedure

Can be made non-combinatorial (convex) by:

\[
\begin{align*}
\min_{\mathbf{g}} & \quad \|\mathbf{g}\|_1 \quad \text{s.t.} \quad \|\mathbf{p} - \mathbf{M}\tilde{q}\mathbf{g}\|_2 \leq \sigma \\
\min_{\mathbf{q}} & \quad \|\mathbf{p} - \mathbf{M}\tilde{q}\mathbf{q}\|_2
\end{align*}
\]

(minimum L1 solution usually the sparsest solution)
Convex EPSI

Do:

\[ g_{k+1} = g_k + \alpha \text{SoftTh}_\phi(\nabla f_{q_k}(g_k)) \]
\[ q_{k+1} = q_k + \beta \nabla f_{g_{k+1}}(q_k) \]

Soft-thresholding solves an L1 minimization problem, but how is \( \phi \) determined?
Projected gradient. Our application of the SPG algorithm to solve $(\mathbb{L}_\Sigma \tau)$ follows Birgin, Martínez, and Raydan [5] closely for the minimization of general nonlinear functions over arbitrary convex sets. The method they propose combines projected-gradient search directions with the spectral step length that was introduced by Barzilai and Borwein [1]. A nonmonotone line search is used to accept or reject steps. The key ingredient of Birgin, Martínez, and Raydan's algorithm is the projection of the gradient direction onto a convex set, which in our case is defined by the constraint in $(\mathbb{L}_\Sigma \tau)$. In their recent report, Figueiredo, Nowak, and Wright [27] describe the remarkable efficiency of an SPG method specialized to $(\mathbb{Q}_\lambda \Sigma)$. Their approach builds on the earlier report by Dai and Fletcher [18] on the efficiency of a specialized SPG method for general bound-constrained quadratic programs $(\mathbb{Q}_\lambda s)$.

2. The Pareto curve. The function $\phi$ defined by (1.1) yields the optimal value of the constrained problem $(\mathbb{L}_\Sigma \tau)$ for each value of the regularization parameter $\tau$. Its graph traces the optimal trade-off between the one-norm of the solution $x$ and the two-norm of the residual $r$, which defines the Pareto curve. Figure 2.1 shows the graph of $\phi$ for a typical problem. The Newton-based root-finding procedure that we propose for locating specific points on the Pareto curve—e.g., finding roots of (1.2)—relies on several important properties of the function $\phi$. As we show in this section, $\phi$ is a convex and differentiable function of $\tau$. The differentiability of $\phi$ is perhaps unintuitive, given that the one-norm constraint in $(\mathbb{L}_\Sigma \tau)$ is not differentiable. To deal with the nonsmoothness of the one-norm constraint, we appeal to Lagrange duality theory. This approach yields significant insight into the properties of the trade-off curve. We discuss the most important properties below.

2.1. The dual subproblem. The dual of the Lasso problem $(\mathbb{L}_\Sigma \tau)$ plays a prominent role in understanding the Pareto curve. In order to derive the dual of $(\mathbb{L}_\Sigma \tau)$, we first recast $(\mathbb{L}_\Sigma \tau)$ as the equivalent problem

$$\begin{align*}
\text{minimize} & \quad \|r\|_2 \\
\text{subject to} & \quad Ax + r = b, \quad \|x\|_1 \leq \tau.
\end{align*}$$

\text{(van den Berg, Friedlander, 2008)}
Projected gradient. Our application of the SPG algorithm to solve (LS\_\tau) follows Birgin, Martínez, and Raydan [5] closely for the minimization of general nonlinear functions over arbitrary convex sets. The method they propose combines projected-gradient search directions with the spectral step length that was introduced by Barzilai and Borwein [1]. A nonmonotone line search is used to accept or reject steps. The key ingredient of Birgin, Martínez, and Raydan’s algorithm is the projection of the gradient direction onto a convex set, which in our case is defined by the constraint in (LS\_\tau). In their recent report, Figueiredo, Nowak, and Wright [27] describe the remarkable efficiency of an SPG method specialized to (QP\_\lambda). Their approach builds on the earlier report by Dai and Fletcher [18] on the efficiency of a specialized SPG method for general bound-constrained quadratic programs (QP\_\lambda).

2. The Pareto curve. The function \( \phi \) defined by (1.1) yields the optimal value of the constrained problem (LS\_\tau) for each value of the regularization parameter \( \tau \). Its graph traces the optimal trade-off between the one-norm of the solution \( x \) and the two-norm of the residual \( r \), which defines the Pareto curve. Figure 2.1 shows the graph of \( \phi \) for a typical problem. The Newton-based root-finding procedure that we propose for locating specific points on the Pareto curve—e.g., finding roots of (1.2)—relies on several important properties of the function \( \phi \). As we show in this section, \( \phi \) is a convex and differentiable function of \( \tau \). The differentiability of \( \phi \) is perhaps unintuitive, given that the one-norm constraint in (LS\_\tau) is not differentiable. To deal with the nonsmoothness of the one-norm constraint, we appeal to Lagrange duality theory. This approach yields significant insight into the properties of the trade-off curve. We discuss the most important properties below.

2.1. The dual subproblem. The dual of the Lasso problem (LS\_\tau) plays a prominent role in understanding the Pareto curve. In order to derive the dual of (LS\_\tau), we first recast (LS\_\tau) as the equivalent problem (2.1) minimize \( \|r\|_2 \) subject to \( Ax + r = b, \|x\|_1 \leq \tau \).
Pareto curve

minimize $\|x\|_1$
subject to $\|Ax - b\|_2 \leq \sigma$

Look at the solution space and the line of optimal solutions (Pareto curve)

Derivative given by $\|A^T r\|_\infty$
Pareto curve

\[
\begin{align*}
\text{minimize} & \quad \|x\|_1 \\
\text{subject to} & \quad \|Ax - b\|_2 \leq \sigma
\end{align*}
\]

Look at the solution space and the line of optimal solutions (Pareto curve)

---

**Fig. 6.1.** Corrupted and interpolated images for problem seismic. Graph (c) shows the Pareto curve and the solution path taken by SPGL1. However, as might be expected of an interior-point method based on a conjugate-gradient linear solver, it can require many matrix-vector products.

It may be progressively more difficult to solve \((\text{QP}_\lambda)\) as \(\lambda \to 0\) because the regularizing effect from the one-norm term tends to become negligible, and there is less control over the norm of the solution. In contrast, the \((\text{LS}_\tau)\) formulation is guaranteed to maintain a bounded solution norm for all values of \(\tau\).

6.4. Sampling the Pareto curve. In situations where little is known about the noise level \(\sigma\), it may be useful to visualize the Pareto curve in order to understand the trade-offs between the norms of the residual and the solution. In this section we aim to obtain good approximations to the Pareto curve for cases in which it is prohibitively expensive to compute it in its entirety.

We test two approaches for interpolation through a small set of samples \(i = 1, \ldots, k\). In the first, we generate a uniform distribution of parameters \(\lambda_i = \left(\frac{i}{k}\right)\|A^Tb\|_\infty\) and solve the corresponding problems \((\text{QP}_{\lambda_i})\). In the second, we generate a uniform distribution of parameters \(\sigma_i = \left(\frac{i}{k}\right)\|b\|_2\) and solve the corresponding problems \((\text{BP}_{\sigma_i})\). We leverage the convexity and differentiability of the Pareto curve to approximate it with piecewise cubic polynomials that match function and derivative values at each end. When a nonconvex fit is detected, we switch to a quadratic interpolation.

---

\[
\begin{align*}
\text{minimize} & \quad \|Ax - b\|_2 \\
\text{subject to} & \quad \|x\|_1 \leq \tau
\end{align*}
\]

solve with SPG (spectral projected gradients)
SPG start

Thursday, June 16, 2011
SPG at Pareto curve
Pareto curve

\[
\begin{align*}
\text{minimize} & \quad \|x\|_1 \\
\text{subject to} & \quad \|Ax - b\|_2 \leq \sigma
\end{align*}
\]

Only solve least-squares matching for \( q \) when solution reaches Pareto curve

\[
\begin{align*}
\text{minimize} & \quad \|Ax - b\|_2 \\
\text{subject to} & \quad \|x\|_1 \leq \tau
\end{align*}
\]
Robust EPSI procedure

While $\|p - M(g_k, q_k)\|_2 > \sigma$

- determine new $\tau_k$ from the Pareto curve

$$g_{k+1} = \operatorname{arg\,min}_{g} \|p - Mq_k g\|_2 \text{ s.t. } \|g\|_1 \leq \tau_k$$
  (Solve with SPGL1 until Pareto curve reached)

$$q_{k+1} = \operatorname{arg\,min}_{q} \|p - Mg_k g_{k+1} q\|_2$$
  (Solve with LSQR)
REPSI in transform domain

Modify just the problem for $g$:

$$\min_g \|g\|_1 \quad \text{s.t.} \quad \|p - M\tilde{q}g\|_2 \leq \sigma$$

$$\min_q \|p - M\tilde{g}q\|_2$$
REPSI in transform domain

Modify just the problem for \( g \):

\[
\begin{align*}
\min_x & \quad \| x \|_1 \quad \text{s.t.} \quad \| p - M_{\tilde{q}} S^\dagger x \|_2 \leq \sigma, \quad g = S^\dagger x \\
\min_q & \quad \| p - M_{\tilde{q}} q \|_2 
\end{align*}
\]

\( S \): sparsifying representation for seismic signals
- Should have spatially localized support
- ex: nd-Wavelets, Curvelets, etc...

\( S^\dagger \): synthesis operator for \( S \)
REPSI in transform domain

While \( \| p - \mathcal{M}(g_k, q_k) \|_2 > \sigma \)

determine new \( \tau_k \) from the Pareto curve

\[
x_{k+1} = \arg \min_x \| p - M_{q_k} S^\dagger x \|_2 \quad \text{s.t.} \quad \| x \|_1 \leq \tau_k
\]

(Solve with SPGL1 until Pareto curve reached)

\[
g_{k+1} = S^\dagger x_{k+1}
\]

\[
q_{k+1} = \arg \min_q \| p - M g_{k+1} q \|_2
\]

(Solve with LSQR)
North sea
EPSI + Sp
North sea SRME
North sea data
North sea pred. mul
(show Gulf of Suez results here)
summary

- L1-convexification behaves nicely and has few free parameters
- Follows the Pareto curve into a series of projected gradient problems
- Easily incorporates seeking the solution in a transform domain that promotes continuity
Acknowledgements

Special thanks to G.J. van Groenestijn, Eric Verschuur, and the rest of the members of DELPHI

M. Friedlander and E. van den Berg

This work was in part financially supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant (22R81254) and the Collaborative Research and Development Grant DNOISE II (375142-08). This research was carried out as part of the SINBAD II project with support from the following organizations: BG Group, BP, Chevron, ConocoPhillips, Petrobras, Total SA, and WesternGeco.
(van Groenestijn and Verschuur 08)