An inversion algorithm for inverse Sturm-Liouville problems

Fernando Guevara Vasquez

Rice University CAAM Dept.

March 31 2004
Outline

Optimal finite difference grids

BV grids

Future research
The Model Forward Problem

Layered medium in an infinite strip in \( \mathbb{R}^2 \) with \( \sigma > 0 \), bounded. Fourier transform in \( x \) to obtain:

\[
\frac{d}{dz} \left( \sigma(z) \frac{du(z)}{dz} \right) - \lambda \sigma(z) u(z) = 0, \quad z \in (0, 1),
\]

\[
-\sigma(0) \frac{du(0)}{dz} = 1, \quad u(1) = 0.
\]

Optimal grids are designed to achieve a very accurate finite difference approximation of the NtD map \( F^\sigma(\lambda) = u(0) \) (Druskin, Knizherman, Igerman, Moskow).

The impedance is an Stieltjes function (Kac and Krein):

\[
F^\sigma(\lambda) = \sum_{j=1}^{\infty} \frac{\xi_j}{\lambda + \theta_j^2}.
\]
Finite difference discretization

\[
\begin{aligned}
\hat{z}_0 &= z_1, \\
\hat{z}_1 &= z_2, \\
\hat{z}_2 &= z_3, \\
\hat{z}_3 &= z_4
\end{aligned}
\]

\[
\begin{aligned}
\hat{h}_1 &= h_1, \\
\hat{h}_2 &= h_2, \\
\hat{h}_3 &= h_3
\end{aligned}
\]

where \( \gamma_j = \frac{h_j}{\hat{\sigma}_j} = \int_{z_j}^{z_{j+1}} \frac{dz}{\sigma(z)}, \quad \hat{\gamma}_j = \hat{h}_j \sigma_j = \int_{\hat{z}_j}^{\hat{z}_{j-1}} \sigma(z) dz. \)

The discrete impedance is also a Stieltjes function:

\[
F_k^\gamma(\lambda) = U_1 = \sum_{j=1}^{k} \frac{\xi_{k,j}}{\lambda + \theta_{k,j}^2}.
\]
The discrete inverse problem

- In principle, the grid and hence $\gamma_j$ and $\hat{\gamma}_j$ can be arbitrary.
- We choose $\gamma_j$ and $\hat{\gamma}_j$, for $j = 1, \ldots, k$, s.t. $2k$ measurements of $F^\sigma(\lambda)$ are satisfied exactly by its approximation $F_k^{\gamma}(\lambda)$. Examples of measurement sets are:
  1. The Truncated Measure (TM) set:
     \[
     \theta_{k,j} = \theta_j \quad \text{and} \quad \xi_{k,j} = \xi_j, \quad \text{for} \quad j = 1, \ldots, k.
     \]
  2. Padé approximation:
     Multipoint: $F^\sigma(\lambda_j) = F_k^{\gamma}(\lambda_j)$, for $2k$ distinct $\lambda_j > 0$.
     Simple: $\frac{d^j}{d\lambda^j} F_k^{\gamma}(\lambda_0) = \frac{d^j}{d\lambda^j} F^\sigma(\lambda_0), \quad j = 0, \ldots, 2k - 1$.
- We can find $\gamma_j$ and $\hat{\gamma}_j$, $j = 1, \ldots, k$, by basically solving a Jacobi inverse eigenvalue problem (Lanczos, Stieltjes).
- Given $\sigma(z)$, we can find the optimal grid.
Imaging on optimal grids

Algorithm 1

Step 1: Calculate the grid $G_k^0$ for $\sigma^0 = 1$.

Step 2: Find $\gamma_j$ and $\hat{\gamma}_j$, for $1 \leq j \leq k$, by solving the Jacobi inverse eigenvalue problem.

Step 3: Obtain the solution as

$$\sigma_j = \frac{\hat{\gamma}_j}{\hat{h}_j^0} \quad \text{and} \quad \hat{\sigma}_j = \frac{h_j^0}{\gamma_j}, \quad 1 \leq j \leq k.$$

By construction $F^\gamma_k(\lambda) = F^{\sigma^k}(\lambda) \rightarrow F^\sigma(\lambda)$. Moreover Borcea, Druskin and Knizherman, proved the convergence of Algorithm 1 to the continuum solution assuming:

- TM measurements are used,
- $\sigma(z)$ is smooth enough (requirement: $\theta_n - \theta_0^0$ and $\xi_n - \xi_0^0$ decay fast enough as $n \rightarrow \infty$).
Optimal grids in action

\[ TV(\ln(\sigma)) = 4.57 \]

\[ TV(\ln(\sigma)) = 12.46 \]

\( (k = 20) \)
Total Variation

Definition
Let $f : [0, 1] \to \mathbb{R}$, the Total Variation (TV) is the supremum over all the partitions $0 = x_0 < x_1 < \ldots < x_m = 1$ of the quantity
\[
\sum_{i=1}^{m} |f(x_i) - f(x_{i-1})| .
\]

Fact
If $f$ is piecewise constant on a grid $0 = x_0 < x_1 < \ldots < x_m = 1$, then
\[
TV(f) = \sum_{i=1}^{m} |f(x_i) - f(x_{i-1})| .
\]

Helly’s selection theorem
A sequence of functions with uniformly bounded total variation, and uniformly bounded at one point has a subsequence that converges pointwise and in $L^1[0, 1]$. 
Convergence of Algorithm 1

Sketch of proof

i. $\text{TV} (\ln (\sigma^k)) \leq C$, where $C$ is a constant independent of $k$: very technical proof, involves perturbation analysis of the Jacobi inverse eigenvalue problem (gives uniformly bounded variation of the $\sigma^k$).

ii. $\sigma^k(0) = \sigma_1 \to 1$ as $k \to \infty$ (gives uniformly boundedness of $\sigma^k(0)$).

iii. Assume for contradiction $\sigma^k \not\to \sigma$ in $L^1[0, 1]$, then $\exists \epsilon > 0$, and a subsequence $\sigma^{k_l}$ s.t. $\|\sigma^{k_l} - \sigma\|_{L^1} > \epsilon$. By Helly's selection theorem, there is another subsequence, call it also $\sigma^{k_l}$, that converges pointwise and in $L^1[0, 1]$, to some limit $\tilde{\sigma}$. Hence we have $F_{\sigma^{k_l}} (\lambda) \to F_{\tilde{\sigma}} (\lambda)$.

iv. But by construction $F_{\sigma^{k_l}} (\lambda) \to F_{\sigma} (\lambda)$, so by uniqueness of the solution to the inverse problem, we must have $\sigma = \tilde{\sigma}$, a contradiction.
Imaging on grids enforcing BV of the reconstructed $\sigma$

Idea
Substitute i. and ii. by an appropriate stretching of the grid, in a way that ensures $\sigma^k$ has uniformly bounded variation.

Algorithm 2

Step 1: Find $\gamma_j$ and $\hat{\gamma}_j$, for $1 \leq j \leq k$, by solving the Jacobi inverse eigenvalue problem.

Step 2: Obtain a grid (and hence a $\sigma^k$) as the solution of

$$
\min_{\text{s.t. } \{h_j, \hat{h}_j\} \text{ grid}} \ TV(\ln(\sigma(h)))
$$

where $\{h_j, \hat{h}_j\}$ is considered to be a valid grid when $\sum_{j=1}^{k} h_j = 1$, and the points $z_j, \hat{z}_j$ alternate adequately.

By construction we also have $F_k^{\gamma}(\lambda) = F^{\sigma^k}(\lambda) \rightarrow F^{\sigma}(\lambda)$. 
Imaging on grids enforcing BV of the reconstructed $\sigma$

Idea
Substitute i. and ii. by an appropriate stretching of the grid, in a way that ensures $\sigma^k$ has uniformly bounded variation.

Algorithm 2

Step 1: Find $\gamma_j$ and $\hat{\gamma}_j$, for $1 \leq j \leq k$, by solving the Jacobi inverse eigenvalue problem.

Step 2: Obtain a grid (and hence a $\sigma^k$) as the solution of

$$
\min_{\text{s.t. } \{h_j, \hat{h}_j\} \text{ grid}} \text{TV}(\ln(\sigma(h))) + \text{penalty}(h),
$$

where $\{h_j, \hat{h}_j\}$ is considered to be a valid grid when $\sum_{j=1}^k h_j = 1$, and the points $z_j, \hat{z}_j$ alternate adequately.

By construction we also have $F_k^{\gamma}(\lambda) = F^{\sigma^k}(\lambda) \rightarrow F^\sigma(\lambda)$.

If a positive penalty term is added, the $\sigma_k$ still have bounded variation!
Formulation of the optimization problem

\[
\hat{\sigma}_0 = \sigma_1 \quad \hat{\sigma}_1 \quad \sigma_2 \quad \hat{\sigma}_2 \quad \sigma_3 \quad \hat{\sigma}_3
\]

\[
\hat{z}_0 = z_1 \quad \hat{z}_1 \quad z_2 \quad \hat{z}_2 \quad z_3 \quad \hat{z}_3 \quad z_4
\]

\[
\min \sum_{j=1}^{k} |\ln(\hat{\sigma}_j) - \ln(\sigma_j)| + \sum_{j=1}^{k-1} |\ln(\sigma_{j+1}) - \ln(\hat{\sigma}_j)|
\]

subject to

\[
H_j \geq 0, \quad j = 1 \ldots 2k,
\]

\[
H_1 + H_2 + \cdots + H_{2k} = 1,
\]

\[
H_0 = 0, \quad H_1 = \hat{\gamma}_1.
\]

Nonlinear, Nonconvex, Nondifferentiable problem.
Formulation of the optimization problem

\[
\begin{align*}
\hat{\sigma}_0 &= \sigma_1 \\
\hat{\sigma}_1 &= \sigma_2 \\
\hat{\sigma}_2 &= \sigma_3 \\
\hat{\sigma}_3 &= \sigma_4
\end{align*}
\]

\[
\begin{align*}
\hat{z}_0 &= z_1 \\
\hat{z}_1 &= z_2 \\
\hat{z}_2 &= z_3 \\
\hat{z}_3 &= z_4
\end{align*}
\]

\[
\begin{align*}
\hat{h}_1 &= h_1 \\
\hat{h}_2 &= h_2 \\
\hat{h}_3 &= h_3
\end{align*}
\]

\[
\begin{align*}
H_1 &= H_1 \\
H_2 &= H_2 \\
H_3 &= H_3 \\
H_4 &= H_4 \\
H_5 &= H_5 \\
H_6 &= H_6
\end{align*}
\]

\[
\begin{align*}
\min \quad & \sum_{j=1}^{k} \left| \ln(h_j/\gamma_j) - \ln(\hat{\gamma}_j/\hat{h}_j) \right| + \sum_{j=1}^{k-1} \left| \ln(\hat{\gamma}_{j+1}/\hat{h}_{j+1}) - \ln(h_j/\gamma_j) \right| , \\
\text{subject to} \quad & H_j \geq 0, \quad j = 1 \ldots 2k, \\
& H_1 + H_2 + \cdots + H_{2k} = 1, \\
& H_0 = 0, \quad H_1 = \hat{\gamma}_1.
\end{align*}
\]

Here \( h_j = H_{2j} + H_{2j-1} \), \( \hat{h}_j = H_{2j-1} + H_{2j-2} \) for \( j = 1, \ldots, k \).

Nonlinear, Nonconvex, Nondifferentiable problem.
Formulation as a *differentiable* optimization problem

\[
\begin{align*}
\min & \quad s_1 + s_2 + \cdots + s_{2k-1} \\
\text{subject to} & \\
H \text{ defines a valid grid} & \\
0 & \leq s_j, \quad j = 1 \ldots 2k - 1, \\
-s_j & \leq c_j(H) \leq s_j, \quad j = 1 \ldots 2k - 1,
\end{align*}
\]

where

\[
\begin{align*}
c_j(H) &= \ln(h_j/\gamma_j) - \ln(\hat{\gamma}_j/\hat{h}_j), \quad \text{for } j = 1 \ldots k, \\
c_{k+j}(H) &= \ln(\hat{\gamma}_{j+1}/\hat{h}_{j+1}) - \ln(h_j/\gamma_j), \quad \text{for } j = 1 \ldots k - 1.
\end{align*}
\]

Nonlinear, Nonconvex problem.
Current implementation uses Matlab’s `fmincon` (SQP, quasi Newton Hessian updates, dense matrices).

Fernando Guevara Vasquez

Inverse Sturm-Liouville problems
Numerical results: smooth $\sigma$

TV grids, Gaussian, $k=20$, Truncated Measure

$(k = 20, \ TV(\ln(\sigma)) = 3.80)$
Numerical results: discontinuous $\sigma$

\begin{align*}
&TV\text{ grids, Step1, } k=20, \text{ Truncated Measure} \\
&(k = 20, \ TV(\ln(\sigma)) = 6.22)
\end{align*}
Improvements for the optimization

- Use a better optimization routine:
  - Denis’s own Interior Point SQP solver
  - commercial packages: LOQO, KNITRO (limited in number of variables).

- Can the problem be formulated as a convex problem, a SDP?

- Can we use an optimization method that handles the non-differentiabilities of the objective function? (subgradient . . .)

- Can we incorporate ideas from the image/signal processing community?
Regularization of the grid

The optimization problem that we solve is:

\[
\min \quad TV(\sigma(h)) + \alpha \text{penalty}(h)
\]
\[
\text{s.t. } \{h_j, \hat{h}_j\} \text{ grid.}
\]

To improve the imaging on these grids for smooth conductivities, we penalize the distance to the homogeneous grid \(\{h_0^j, \hat{h}_0^j\}:

\[
\text{penalty}(h) = \|h - h_0^j\|_p + \|\hat{h} - \hat{h}_0^j\|_p, \text{ for } p = 1, 2.
\]

The parameter \(\alpha \geq 0\) allows us to transition from one method to the other.
Numerical results: penalty

TV(\ln(\sigma)) = 3.80

TV(\ln(\sigma)) = 6.22

\alpha = 0.1, k = 20
Numerical results: penalty

\[
TV(\ln(\sigma)) = 3.80
\]

\[
TV(\ln(\sigma)) = 6.22
\]

\(\alpha = 1, k = 20\)
Numerical results: penalty

\[TV(\ln(\sigma)) = 3.80\]

\[TV(\ln(\sigma)) = 6.22\]

\[\alpha = 3, k = 20\]
Numerical results: penalty

TV(log(σ)) = 3.80

TV(log(σ)) = 6.22

α = 10, k = 20
Numerical results: penalty

TV(ln(σ)) = 3.81
TV(log(σ)) = 3.813890e+00

TV grids + regularization α = 20, Gaussian, k=20, Truncated Measure

TV grids + regularization α = 20, Step1, k=20, Truncated Measure

α = 20, k = 20

TV(ln(σ)) = 6.23
TV(log(σ)) = 6.229929e+00
Numerical results: penalty

TV(ln(\(\sigma\))) = 3.83

\(\alpha = 30\), \(k = 20\)

TV(ln(\(\sigma\))) = 6.24

\(\alpha = 30\), Step1, \(k = 20\)
Numerical results: penalty

\[ TV(\ln(\sigma)) = 3.85 \]

\[ TV(\ln(\sigma)) = 6.22 \]

\( \alpha = 40, k = 20 \)
Numerical results: penalty

TV(\ln(\sigma)) = 3.87

\alpha = 50, k = 20

TV(\ln(\sigma)) = 6.24
Numerical results: penalty

\[
TV(\ln(\sigma)) = 3.97 \\
\alpha = 100, k = 20
\]

\[
TV(\ln(\sigma)) = 6.30 \\
\alpha = 100, \text{Step1, } k=20
\]
Numerical results: penalty

TV(ln(σ)) = 4.04

TV(ln(σ)) = 6.34

α = 200, k = 20
Numerical results: penalty

TV(\ln(\sigma)) = 4.26

\alpha = 1000, k = 20

TV(\ln(\sigma)) = 6.86
Padding with artificial data

What if we wanted more points than measurements?

Idea
As \( k \to \infty \) the asymptotic behavior of \( \xi_k \) and \( \xi_0^k \) is the same, and similarly for \( \theta_k \) and \( \theta_0^k \) (Coleman, McLaughlin).

Method
Given \( 2k \) spectral measurements,

\[
\theta_1, \ldots, \theta_k \text{ and } \xi_1, \ldots, \xi_k,
\]

we use as data for the inversion the measurements

\[
\theta_1, \ldots, \theta_k, \theta_0^{k+1} \cdots \theta_0^{k+N} \text{ and } \xi_1, \ldots, \xi_k, \xi_0^{k+1}, \ldots, \xi_0^{k+N}
\]

to obtain a \( 2(k + N) \) grid points.

- Another way of regularizing
- To try for TM, and Padé spectral measurements
The 2D problem

A problem that arises in geologic prospecting applications is

\[
\nabla \cdot [\sigma(r, \theta) \nabla u] = 0 \quad \text{in } \Omega,
\]

\[
-s(1, \theta) \frac{\partial u}{\partial n}(1, \theta) = I(\theta),
\]

\[
u(L, \theta) = 0.
\]
Homogeneous grid in 2D

- Consider an homogeneous medium, e.g. $\sigma = 1$, then the problem simplifies to:

$$
\Delta u = 0 \quad \text{in } \Omega \\
- \frac{\partial u}{\partial n}(1, \theta) = I(\theta), \quad u(L, \theta) = 0.
$$

- Use polar coordinates and take the Fourier transform w.r.t. $\theta$:

$$
r \frac{\partial^2 \hat{u}}{\partial r^2}(r, \omega) + \frac{\partial \hat{u}}{\partial r}(r, \omega) - \frac{1}{r} \omega^2 \hat{u}(r, \omega) = 0 \quad \text{for } r \in [1, L],
$$

$$
- \frac{\partial \hat{u}}{\partial r}(1, \omega) = \hat{I}(\omega), \quad \hat{u}(L, \omega) = 0.
$$

- Using similar ideas, we have an optimal placement $\{r_i, \hat{r}_i\}$.
- We will take as the homogeneous grid the tensor product of this optimal placement $\{r_i, \hat{r}_i\}$ and of an angular grid (uniform).
Imaging in 2D

- Discretizing the equation with Finite Volumes, the discrete inverse problem becomes a resistor network problem, where the resistors to find are averages of $\sigma$ over grid cells, analogously to the $\{\gamma_j, \hat{\gamma}_j\}$ in 1D.

- How to find the resistors that reproduce the data? One possibility is circular planar graphs (Ingerman, . . .).

- Then use the homogeneous grid, and the obtained “resistors” to reconstruct $\sigma$ via the averaging relations.

- BV grids idea could also be used: given the resistors find the grid such that the reconstructed $\sigma$ has minimal total variation.
Summary

Work done

▸ an inversion algorithm for Sturm-Liouville type inverse problems that can be proven to converge to the continuum solution, and that works well with discontinuous parameters $\sigma$

▸ a working implementation

Future work

▸ improve the implementation

▸ explore the idea of regularization of the grid

▸ extension to 2D problems

▸ application to geologic prospecting
What the objective looks like $1/2$

For the case $k = 2$, the minimization can be done in $\mathbb{R}^2$. Assume $\{\gamma_j, \hat{\gamma}_j\}_{j=1,2}$ are known, then the problem becomes:

$$
\min |\ln(\hat{\sigma}_1) - \ln(\sigma_1)| + |\ln(\sigma_2) - \ln(\hat{\sigma}_1)| + |\ln(\hat{\sigma}_2) - \ln(\sigma_2)|
$$

subject to $H_2, H_3 \geq 0$, $H_2 + H_3 \leq 1 - \hat{\gamma}_1$,

where

$$
\sigma_1 = \frac{\gamma_1}{\hat{h}_1} = 1,
\hat{\sigma}_1 = \frac{h_1}{\gamma_1} = \frac{\hat{\gamma}_1 + H_2}{\gamma_1},
\sigma_2 = \frac{\gamma_2}{\hat{h}_2} = \frac{\hat{\gamma}_2}{H_2 + H_3},
\hat{\sigma}_2 = \frac{h_2}{\gamma_2} = \frac{1 - \hat{\gamma}_1 - H_2}{\gamma_2}.
$$
What the objective looks like 2/2
Numerical results: L-curve

\( (k=20) \)
Other general measurements

► What if we had noisy measurements?
► Given \( \{ f(\lambda_i) \}_{i=1}^k \), find the \( \{ \theta_j, \xi_j \}_{i=1}^k \) that best fits the data:

\[
f(\lambda) = \sum_{j=1}^{k} \frac{\xi_j}{\lambda + \theta_j^2}.
\]

Then use this \( \{ \theta_j, \xi_j \}_{i=1}^k \), as spectral data.
► Is this a stable imaging algorithm?