Random Convolution and $\ell_1$ Filtering

Justin Romberg, Georgia Institute of Technology

Duke CS Workshop
February 26, 2009

Collaborators: Ramesh Neelamani (ExxonMobil), Joel Tropp (Caltech), Salman Asif (Georgia Tech)
Deconvolution

- Transmitter sends out a “pulse” (or probe), receiver listens to the echo(s)

- Transmit $p$, receive $y = p \ast x$

- If we observe all of $y$, it is deconvolution
  More generally, it is a linear inverse problem: $y = \Phi x$

- Common sensing framework:
  radar, sonar, seismic, channel estimation, coded imaging, etc.

- This talk: we will use a random probe,
  and study what is fundamentally possible
Random Convolution

- Motivations for randomness:
  - wide band, high resolution, but low peak power
  - diverse
  - easy to generate
  - different probes do not interfere with one another

- Recent novel applications:

  “random radar/sonar” [Axellson 07]

  “fluttered shutter” [Raskar et al 06]
The Inverse Problem

- Gaussian probe of length $m$, convolved with the channel response of length $n$

- Key: conditioning of an $m \times n$ Gaussian Toeplitz matrix

- We know a lot about the conditioning of $m \times n$ iid Gaussian matrix (play a key role in compressed sensing)
  - for $m = n$, the condition number is $\sim n$ (bad)
  - very well conditioned when $m \geq \text{Const} \cdot n$
The recovery problem boils down to solving

\[ y = \Phi x, \]

where

\[ \Phi = F^*GF \] (Gaussian circulant), and \( x \) supported on first \( n \) terms

- \( \Phi \) is badly conditioned in general
Submatrices of Random Toeplitz matrices

- Let $\Phi = F^* G F$ be an $m \times m$ Gaussian circulant matrix. Then with high probability

\[ (1 - \delta) \|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \delta) \|x\|_2^2 \]

for every $s$-sparse $x$ with

\[ s \leq \text{Const} \cdot \frac{m}{\log^5 m} \]

- Even though $\Phi$ is ill-conditioned as a whole, it is well-conditioned when operating a sparse vectors

- Draws on Rudelson and Vershynin’s estimates for random sums of rank-1 matrices
Stable Recovery

• Consequence 1:
  a channel response of length $n$ can be recovered (stably) using a probe of length $m \gtrsim n \log^5 n$

\[
\hat{x} = \arg\min_x \|\Phi \Gamma x - y\|_2^2, \quad \hat{x} = (\Phi^*_\Gamma \Phi_\Gamma)^{-1} \Phi^*_\Gamma y
\]

\[
\Gamma = \{0, \ldots, n - 1\} = \text{support of } x
\]
• Consequence 2: probe of length $m$ can recover (stably) a channel response with $s \lesssim m \log^{-5} m$ active components

\[
y = \text{sparse } x
\]

using $\ell_1$ minimization (CRT ’06)

\[
\hat{x} = \arg\min_x \tau \|x\|_1 + \frac{1}{2} \|\Phi x - y\|_2^2,
\]

or CoSAMP (Needell and Tropp ’08), or iterative thresholding (Blumensath and Davies ’08)

(related results by Bajwa, Haupt, Nowak, Raz ’08, Saligrama ’08, R. ’08)
Multiple Channels

Efficient Seismic Modeling and Acquisition using Random Sources and Sparsity

Ramesh Neelamani (Neelsh), Chris Krohn, Jerry Krebs, Max Deffenbaugh, John E Anderson

EMURC
Thursday, February, 2008

Seismic Data Acquisition

• Acquisition time
  - Time to shake and record at different locations (+ time to move the setup to different locations)
• Goal: High efficiency acquisition
  - Less $ without compromising data quality (significantly)
• Approach: Simultaneous sourcing
Multiple Channels

In Pictures

0.0

2.5

Time (s)

E1

E2

E3

E4

From Data to Subsurface Image

All receiver measurements (~1TB)

Subsurface properties

(velocity, reflection coefficients, porosity, permeability, etc)

Total data = Earth's response at all locations on the surface to sources sequentially placed at on the surface

i.e, total data = 5D cube (2D receivers locations x 2D source locations x time)

Modeling Summary

• "Goal:"
  • Efficiently estimate 5D Green's function for an acoustic/elastic model
  • "5D = 2D receiver * 2D source * 1D time"

• "Why:"
  • "Brute force" approach is very expensive (impractical?)
  • "Sequentially energizes each source and computes response via forward modeling (say, using a finite difference (FD) scheme"
  • "Brute force efficiency ! 1/(number of sources * T)"

• "Contribution:"
  • "Setup:"
    • Model by simultaneously setting off random noise sources at all source locations and compute responses at all receivers (by running FD once)
  • "Key problem:"
    • Separate interfering source-to-receiver Green's functions from each receiver measurement
  • "Realization:"
    • Formulate as a ill-posed linear inverse problem and solve by exploiting sparsity of the Green's function (say, in the curvelet domain)
    • "Random efficiency ! 1/(Tr+ T)."
Multiple Channels

- We now have an underdetermined random system

\[
\begin{bmatrix}
    y \\
\end{bmatrix}
= 
\begin{bmatrix}
    G_1 & G_2 & \cdots & G_p \\
\end{bmatrix}
\begin{bmatrix}
    x_1 \\
    x_2 \\
    \vdots \\
    x_p \\
\end{bmatrix}
\]
Multiple Channels

• Let $y = \Phi x$ be a $m \times mp$ multiple convolution system

\[
\begin{pmatrix}
y \\
x_1 \\
x_2 \\
\vdots \\
x_p
\end{pmatrix}
= \begin{pmatrix}
\begin{bmatrix}
\end{bmatrix} \\
\begin{bmatrix}
\end{bmatrix} \\
\begin{bmatrix}
\end{bmatrix} \\
\begin{bmatrix}
\end{bmatrix}
\end{pmatrix}
\begin{pmatrix}
x_1 \\
x_2 \\
\vdots \\
x_p
\end{pmatrix}
\]

• With high probability

\[
(1 - \delta)\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \delta)\|x\|_2^2
\]

for all $s$-sparse $x$ with

\[
s \leq \text{Const} \cdot \frac{m}{(\log m + \log p)^5}
\]

• Consequence: we can recover such $x$ stably using

\[
\hat{x} = \arg\min_x \tau\|x\|_\ell_1 + \frac{1}{2}\|\Phi x - y\|_2^2
\]
Seismic Imaging Simulation

(a) (b) (c)

Figure 1: (a) Shot and receiver geometry. (b) Sample simultaneous sweeps. (c) Receiver measurements.

(c) Difference ((a) minus (b)) (d) Cross-correlation estimate.

Figure 2: Simulation results.

64 shots x 128 receivers
Seismic Imaging Simulation

Figure 1: (a) Shot and receiver geometry. (b) Sample simultaneous sweeps. (c) Receiver measurements.
(c) Difference ((a) minus (b)) (d) Cross-correlation estimate.

Figure 2: Simulation results.

Representation $\Psi = \text{curvelets}, \sim 8 \times \text{decrease in time}$
(courtesy of R. Neelamani, ExxonMobil)
Decoding with $\ell_1$

- Applying a tall random matrix can protect a message against *sparse errors*.

- Recover message by solving

$$\min_x \|y - Ax\|_{\ell_1}$$

- You can correct $\rho \cdot m$ errors

(Candes and Tao ’06)
• Observe a combination of (unknown) shifts of a coded (unknown) message
• "Blind deconvolution"
• Recovering $x$ and channel response $h$ can be recast as a sparse multichannel deconvolution problem
• For $m = Cn$, can protect against a channel with $\sim C / \log^5 m$ taps
Streaming sparse recovery: $\ell_1$ filtering
Streaming sparse recovery: $\ell_1$ filtering

- Solving an optimization program like

$$\min_x \tau \|x\|_{\ell_1} + \frac{1}{2} \|Ax - y\|_2^2$$

...can be costly

- We want to update the solution when
  1. the underlying signal changes slightly, or
  2. we add measurements
Time Varying Sparse Signal

• System model  
\[ y_0 = Ax_0 + e_0 \]

• Estimate using Lasso  
minimize  
\[ \tau \| x \|_1 + \frac{1}{2} \| Ax - y_0 \|_2^2 \]

• New measurements after some time  
\[ y_1 = Ax_1 + e_1 \]

• Estimate again using Lasso  
minimize  
\[ \tau \| x \|_1 + \frac{1}{2} \| Ax - y_1 \|_2^2 \]

• Homotopy:  
minimize  
\[ \tau \| x \|_1 + (1 - \epsilon) \frac{1}{2} \| Ax - y_0 \|_2^2 + \epsilon \frac{1}{2} \| Ax - y_1 \|_2^2 \]

Homotopy parameter: change from 0 to 1.
Update Direction

\[
\min_{x} \tau \|x\|_1 + \frac{1 - \epsilon}{2} \|Ax - y_{old}\|_2^2 + \frac{\epsilon}{2} \|Ax - y_{new}\|_2^2
\]

- Optimality conditions:

\[
A_{\Gamma}^T (Ax - (1 - \epsilon)y_{old} - \epsilon y_{new}) = -\tau \text{sign } x_{\Gamma}
\]

\[
\|A_{\Gamma c}^T (Ax - (1 - \epsilon)y_{old} - \epsilon y_{new})\|_{\infty} < \tau
\]

- Update direction:

\[
\partial x = \begin{cases} 
-(A_{\Gamma}^T A_{\Gamma})^{-1} (y_{old} - y_{new}) & \text{on } \Gamma \\
0 & \text{otherwise}
\end{cases}
\]
Experiment 1: Evolving Piecewise Poly

- Fix singularities, smooth parts “wiggle”
- Average # steps = 12
Experiment 2: Slices of an Image

- Fix singularities, smooth parts “wiggle”
- Average # steps = 70
Recursive Least Squares

• Classical least-squares:
  
  solve a system of linear eqns  \( y = Ax + e \)
  
  min energy solution  \( \min_x \|Ax - y\|_2^2 \)
  
  analytical solution  \( \hat{x} = (A^* A)^{-1} A^* y \)

• Suppose we add new measurements  \( w = B^* x \)

\[
\begin{bmatrix}
A \\
B
\end{bmatrix} x = \begin{bmatrix}
y \\
w
\end{bmatrix}
\]

\( \hat{x}_0 = (A^* A + B^* B)^{-1} (A^* y + B^* w) \)

• Recursive Least-Squares (RLS): easy low-rank update

\[
\hat{x}_1 = \hat{x}_0 + (I + B(A^* A)^{-1} B^*)^{-1} (A^* A)^{-1} B^* (w - B\hat{x}_0)
\]
Dynamic Lasso

- We want the analog of RLS for the LASSO. Adding one measurement

\[
\begin{bmatrix}
y \\
w
\end{bmatrix} = \begin{bmatrix} A \\ b \end{bmatrix} x + \begin{bmatrix} e \\ d \end{bmatrix} \quad \rightarrow \quad \min_x \tau \|x\|_{\ell_1} + \frac{1}{2} \|Ax - y\|_2^2 + \frac{1}{2} \|bx - w\|_2^2
\]

- Challenges:
  - not as smooth as least-squares update
  - solution can change drastically with just one new measurement
  - need to move slowly, use a homotopy method

(see also work by Garrigues et al. 08)
### LASSO Update

**Dynamic Lasso:**

\[
\begin{bmatrix}
  y \\
  w
\end{bmatrix} = \begin{bmatrix}
  A \\
  b
\end{bmatrix} x + \begin{bmatrix}
  e \\
  d
\end{bmatrix}
\]

\[
\text{minimize} \quad \tau \| \tilde{x} \|_1 + \frac{1}{2} \| A \tilde{x} - y \|_2^2 + \frac{1}{2} \| b \tilde{x} - w \|_2^2
\]

\[
\text{minimize} \quad \tau \| \tilde{x} \|_1 + \frac{1}{2} (\| A \tilde{x} - y \|_2^2 + \epsilon \| b \tilde{x} - w \|_2^2),
\]

Updated system with 1 new measurement

Homotopy parameter: change it from 0 to 1.

**How to change \( \epsilon \)? (direction, step size)**

\[
\| A^T (Ax^{(\epsilon)} - y) + \epsilon b^T (bx^{(\epsilon)} - w) \|_\infty \leq \tau,
\]

\((L_{\text{opt}})\)

(M1) \[ A^T (Ax^{(\epsilon)} - y) + \epsilon b^T (bx^{(\epsilon)} - w) = -\tau z_{\epsilon} \]

(M2.) \[ \| A^T_{\text{re}} (Ax^{(\epsilon)} - y) + \epsilon b^T_{\text{re}} (bx^{(\epsilon)} - w) \|_\infty < \tau \]
Update Direction

**RLS**

One homotopy step

\[
x_1 = x_0 + \frac{(A^T A)^{-1}b^T(w - bx_0)}{1 + b(A^T A)^{-1}b^T}
\]

**L1 update**

Multiple homotopy steps

\[
\partial x = \begin{cases} 
U^{-1}b^T_\Gamma(w - bx^{(\epsilon_0)}) & \text{on } \Gamma \\
0 & \text{elsewhere.}
\end{cases}
\]

\[
U = (A^T_\Gamma A_\Gamma + \epsilon_0 b^T_\Gamma b_\Gamma)
\]

\[
u = b_\Gamma U^{-1}b^T_\Gamma
\]

\[
x^{(\epsilon)} = x^{(\epsilon_0)} + \frac{(\epsilon - \epsilon_0)}{1 + (\epsilon - \epsilon_0)u}U^{-1}b^T_\Gamma(w - bx^{(\epsilon_0)}) \text{ on } \Gamma.
\]
Number of steps per update

Measurements $m = 150$
Signal length $n = 256$
Summary

• Effectiveness of channel estimation using a random probe boils down to the spectral properties of Gaussian circulant matrices

\[ \Phi = F^* G F \]

• A probe of length \( m \) can recover an \( m/ \log^5 m \) sparse channel
  – support known: least-squares
  – support unknown: \( \ell_1 \)-regularization (LASSO)

• Uses many of the mathematical tools from compressed sensing

• Results extend to the multichannel case

\[ \Phi = \begin{bmatrix} F^* G_1 F & F^* G_2 F & \cdots & F^* G_p F \end{bmatrix} \]

• We have dynamic recovery algorithms that use low-rank updates