Techniques for Fast Signal Solutions from Massive, Noisy Data

Doug Pewitt
doug@pewitt.org

Why Massive, Noisy Data

- All real data are noisy.
- Massive problems ($\equiv >10^6$ unknowns), e.g.,
  - Image processing
    - Medicine
    - Astronomy
  - Geophysics
    - Mineral exploration
    - Satellite geodesy
  - Seismology
    - Earthquakes
    - Structural analysis
  - Financial Models
  - Bioinformatics
- Much is going on, we'll only touch high points.
- Presentation is a working physicist's approach, **not** mathematically rigorous.

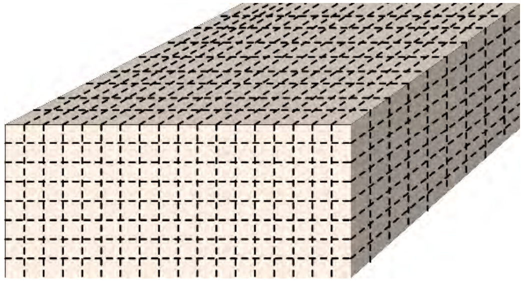
Signal Processing

- We solve a **Forward Problem** such as $\mathbf{F} = \mathbf{G}\,\mathbf{m}_i\,\mathbf{m}_j/\mathbf{r}_{ij}^2$ with known $\mathbf{G}$, $\mathbf{m}_i$, $\mathbf{m}_j$, $\mathbf{r}_{ij}$
- But, if we want to determine a system's internal structure from external **signals** we have an **Inverse Problem**.

E.g., find an unknown density ($\rho = \mathbf{m}/\mathbf{v}$) distribution from gravity **signals** $\mathbf{F}_i$ of M measurements.

$$F_i = G m_i \mathrm{v} \sum_{j=1}^{N} \rho_j / r_{ij}^2$$

Where a volume of interest is divided into N cells of volume $\mathbf{v}$.



Inverse Problems I

- Inverse Problems are of the form
  [**A**] [**x**] = [**b**] where:

  [**A**] is an ***M***×***N*** operator
  matrix from physics,
  e.g., $a_{ij} = \mathrm{Gm}_i\mathrm{v}/r_{ij}^2$,

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ a_{M1} & a_{M2} & \cdots & a_{MN} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ \vdots \\ b_M \end{bmatrix}$$

  [**x**] is a vector of ***N*** unknowns,
  e.g., $x_i = \rho_i$

  [**b**] is the data vector of ***M*** measurements,
  e.g., $b_i = \mathrm{F}_i$

- We write this in matrix form as **Ax = b** and address only **linear problems**.

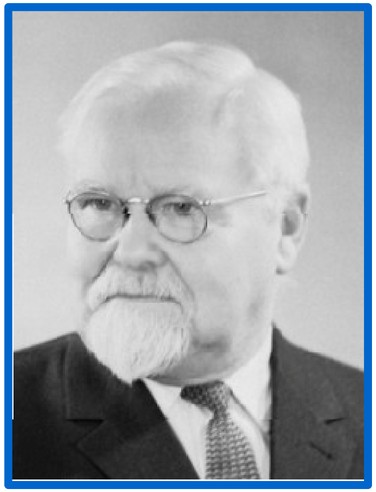
Inverse Problems II

- To solve inverse problems a naïve approach is to invert **A** and get $\mathbf{A}^{-1}$ and have $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$.
  - From: $\mathbf{A}^{-1}\mathbf{A}\,\mathbf{x} = \mathbf{I}\,\mathbf{x} = \mathbf{x}$, where **I** is the identity matrix, ($I_{ij}=1$ if $i=j$, $I_{ij}=\varnothing$ if $i \neq j$)
- **But its a bad idea!**
  - Books might write $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$, but that's not the way to calculate it. **Google** "Don't invert that matrix"
- We mostly avoid or don't face **ill-posed** problems.
  - They require numerical methods largely developed over the last 20-30 years.

Ill-Posed Problems

- **Ill-posed** problems:
  a) do not have unique solutions, or
  b) are *ill-conditioned*, i.e., **any** data noise can cause arbitrarily large perturbations of solutions.
    - Denoising helps but by itself isn't sufficient.
- There is a large Ill-posed problems literature.
  - Mostly it's about **Tikhonov regularization** for linear problems based on work by cold war era Soviet academicians.

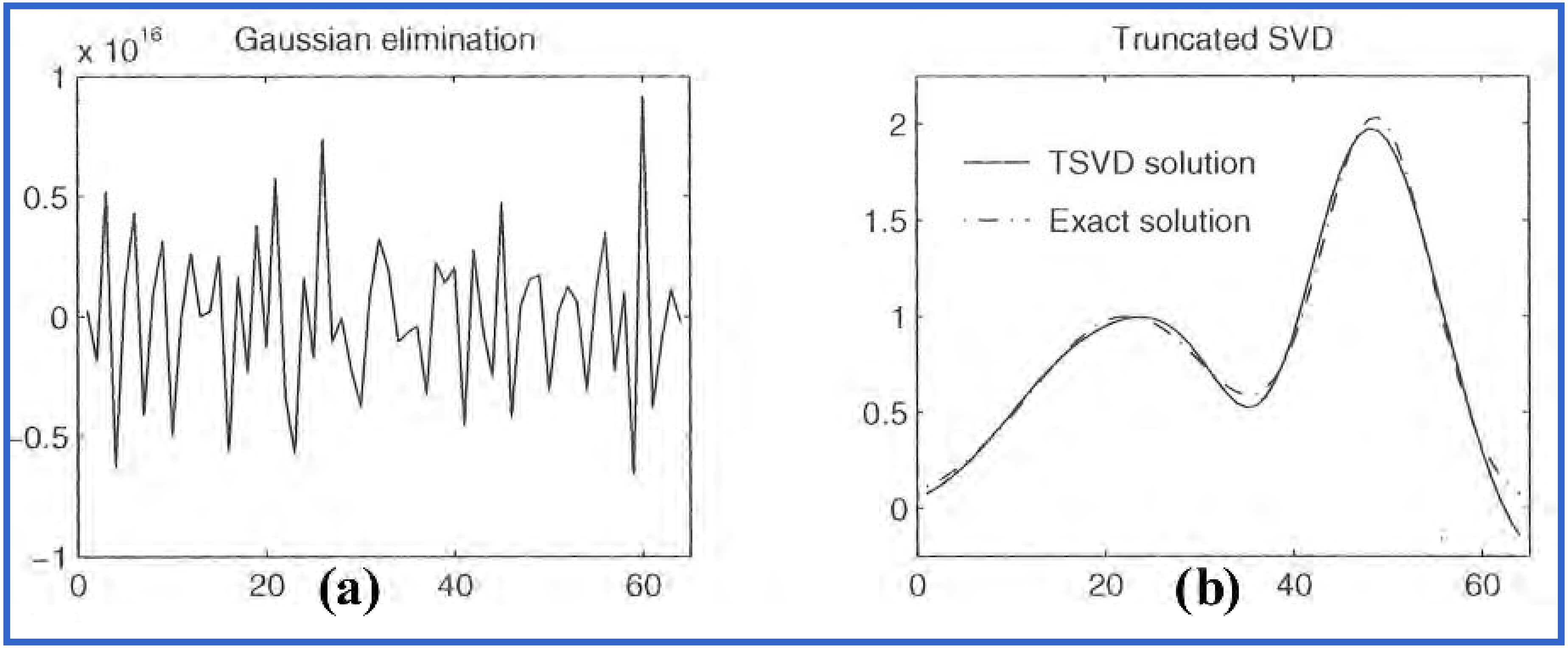
    "*For a long time mathematicians felt that ill-posed problems cannot describe real phenomena and objects. However [...] such problems have important applications.* Tikhonov"



Regularization

- Allows introduction of **additional information** to solve linear ill-posed problems.
- Tikhonov regularization finds an **approximate** solution $\mathbf{x}_\lambda$ that is a **unique** minimizer of
$$\|\mathbf{Ax}-\mathbf{b}\|^2 + \lambda\|\mathbf{Qx}\|^2$$
  - The first term is the usual least squares, the second is the regularizer; **λ** is a problem dependent regularization parameter and **I** often works for **Q**.
- Other methods are useful with ill-posed problems, e.g., singular value decomposition (SVD) can be helpful in analyzing ill-posed problems.

Solution of Ill-Posed Problems



Solutions of a 64 × 64 operator matrix computed by (a) Gaussian elimination and (b) with the seven largest truncated SVD components.

Noisy Data: Summary

- Available numerical methods can produce solutions for problems that do not have unique solutions (e.g., potential fields) or where traditional methods (e.g., least squares) are unstable with noisy data.

Noisy Problems Doable

How about Massive Problems?

Massive Problems: $N > 10^6$

- In signal processing we learned:
  - Shannon showed that sampling rate should be twice the maximum frequency present in the signal (Nyquist rate).
  - Classical least squares "solves" overdetermined systems (**M** $\geq$ **N**) of linear equations.
- This leads to $\geq 10^{12}$ matrix elements.
  - If the problem is "sparse" (mostly Øs) this may be easy but if it is "dense" **real time solutions** with classical techniques are impractical even with supercomputers.
- **Recent developments provide methods to manage such "dense" problems.**

Reducing Problem Size: I
Compressed Sampling (CS)

- Introduced by David Donoho in 2004.*
  - http://www-stat.stanford.edu/~donoho/reports.html
- Shows the way to acquire data in compressed form so far fewer samples are needed (**$$$**).
- The central idea is that the samples needed depends primarily on content not bandwidth.
  - Exploits **compressibility** of many natural signals.
    - An example is **.jpg** image compression.
  - For **N** unknowns the number of measurements, **M** can be on the order of $\mathbf{N}^{1/4} \log^{5/2}(\mathbf{N})$, e.g., for $\mathbf{N}=10^6$, $\mathbf{M} \sim 2{,}800$; $\mathbf{N}=10^9$, $\mathbf{M} \sim 43{,}200$

* **Patent # 7,646,942** see Google Patent search: US007646924B2

Reducing Problem Size: II
Randomized Matrix Approximation

- Decades of history with significant recent advances.
  - “Statistical theory of energy levels of complex systems”, Freeman Dyson, 1962.
  - Made accessible with Sept. 2009 survey by Halko, Martinsson, and Tropp.
    - Available at http://arxiv.org/abs/0909.4061.
  - An approach toward petascale (10^{16}) data analysis.
- Still working on understanding this.
  - May be “better” than or complementary to **CS**.
  - Donoho’s **CS** patent references randomized matrices.

Reducing Computation

- **Interior Point ℓ_1 minimization.**
  - ℓ_1, minimizes the sum of absolute values $\Sigma|r_i|$, where $r_i = \Sigma a_{ij} x_j - b_i$, least squares $[\Sigma r_i^2]^{1/2}$ is ℓ_2.
    - Proposed by Boscovich (1760) , furthered by Laplace (1789) but faded after Gauss described ℓ_2 (1794).
  - Karmarkar (1984) showed interior point minimization to be faster than simplex on large problems.
  - Reported to typically yield a sparser solution in **X**.
  - Can solve problems with 10^6 variables in minutes on a PC.
  - Impressive when used together with compressed sampling see: http://www.acm.caltech.edu/l1magic/

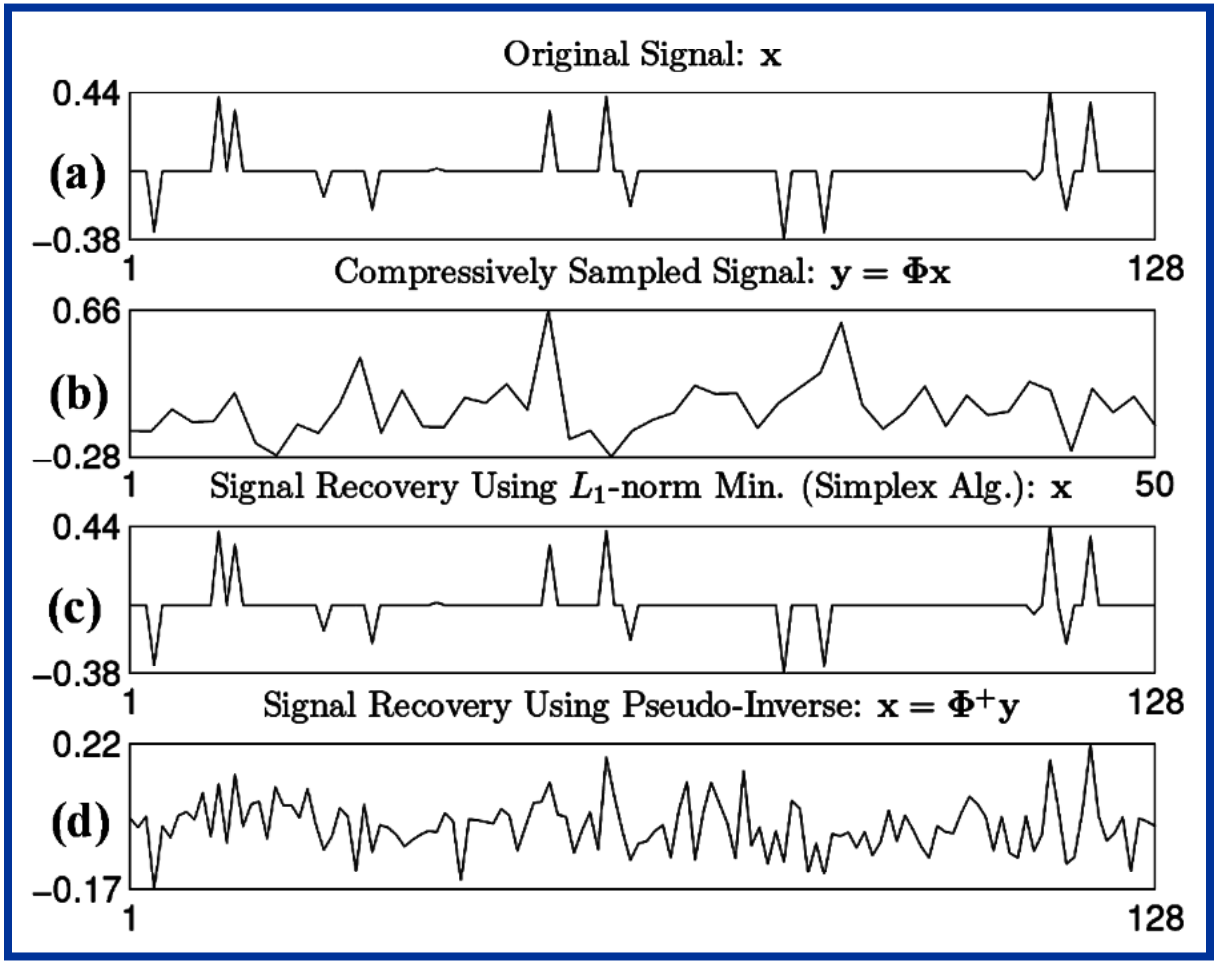
ℓ_1 Compressed Sampling

(a) The original signal **x** of length **N** = 128.

(b) The compressively sampled signal **y** of length **M** = 50.

(c) The **perfectly recovered signal** using **CS** and ℓ_1-norm minimization.

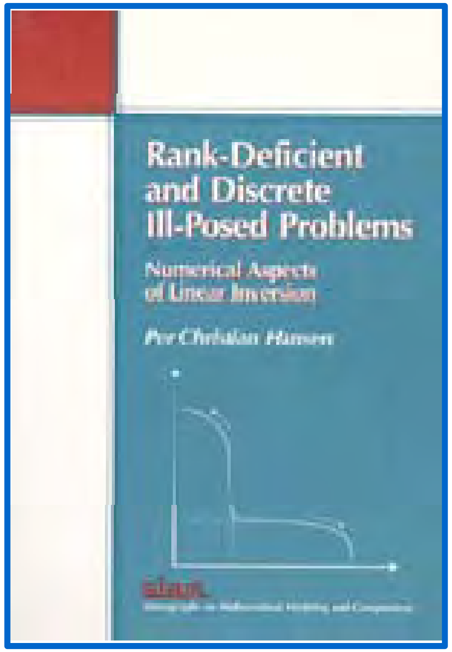
(d) The ℓ_2-norm solution with **M = N**.

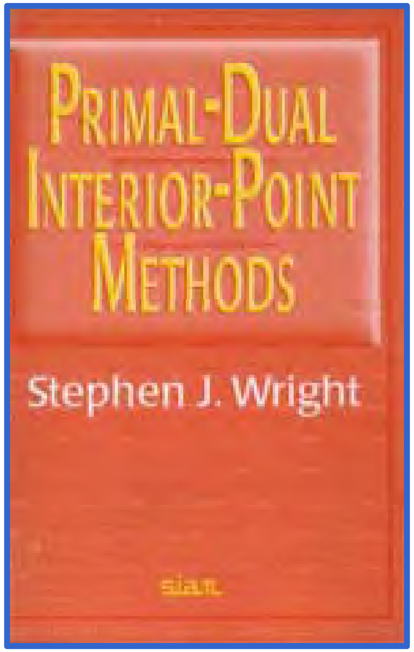


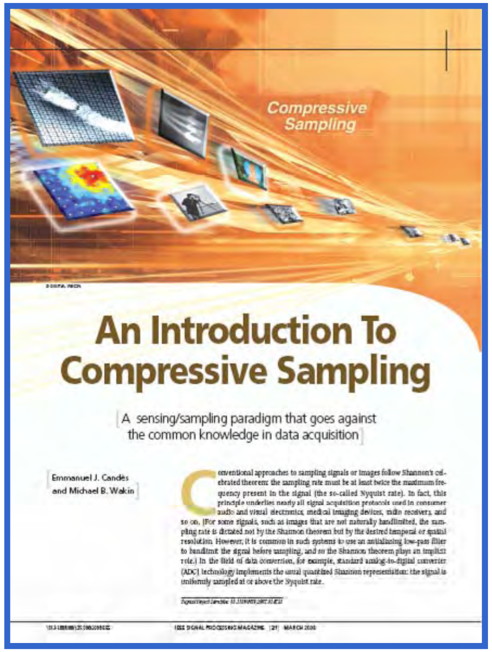
Computing Resources

- General-purpose computing on graphics processing units (**GPGPU**) is a **game changer!**
- Personal supercomputers with up to 2,048 GPUs.
  - Multi teraflop computing for under $20,000.
  - NVIDIAs GPU CUDA architecture using C, C++, and Fortran is optimized for scientific applications.

Numerical Maths Starting Points







Compressive Sensing Resources at: dsp.rice.edu/cs and www.acm.caltech.edu/l1magic/

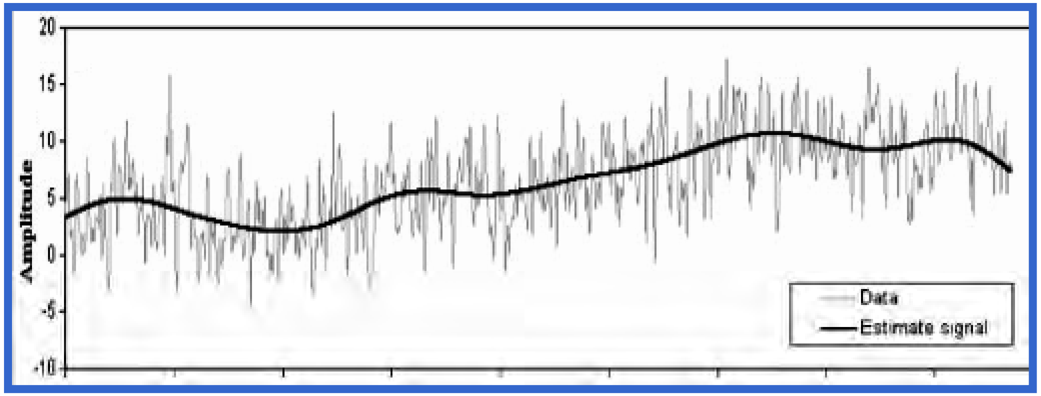
For Randomized Matrix Approximation see Joel Tropp's homepage at: www.acm.caltech.edu/~jtropp/

Backup

Numerical Methods: Other

- **Google**
  - PageRank—The World's Largest Eigenvalue Problem
    - Google assigns a PageRank (importance weight) to each page which is computed via the eigenvalue problem $\boldsymbol{Pw} = \lambda \boldsymbol{w}$ where $\boldsymbol{P}$ is based on the link structure of the Internet.

- Denoising
  - Gaussian white noise can be reduced by wavelet denoising.



  - Donoho and others showed many ill-conditioned problems can be transformed to be well-conditioned by using wavelet thresholding.