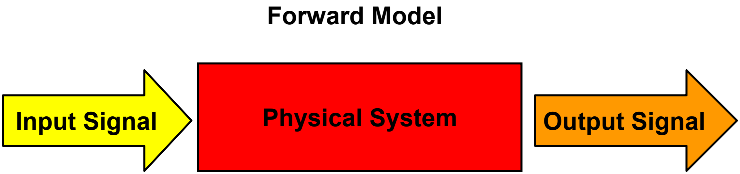
Numerical Methods for Large-Scale Ill-Posed Inverse Problems

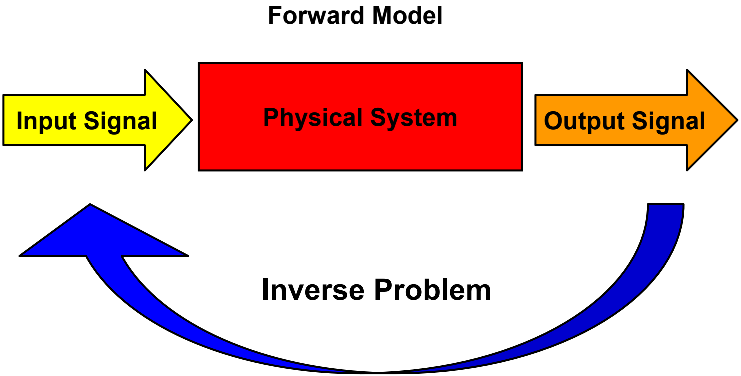
Julianne Chung
University of Maryland

Collaborators: James G. Nagy (Emory)
Eldad Haber (Emory)
Dianne O'Leary (University of Maryland)
Ioannis Sechopoulos (Emory)
Chao Yang (Lawrence Berkeley National Laboratory)

What is an inverse problem?



What is an inverse problem?

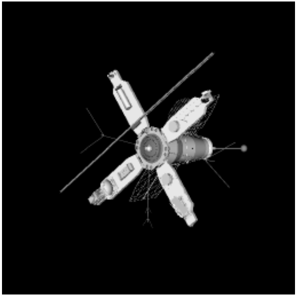


Application: Image Deblurring

- Given: Blurred image and some information about the blurring
- Goal: Compute approximation of true image

Application: Image Deblurring

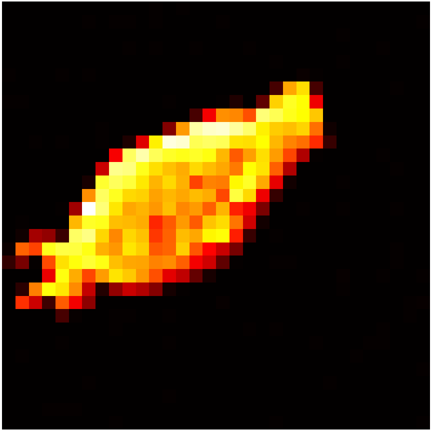
- Given: Blurred image and some information about the blurring
- Goal: Compute approximation of true image



Application: Super-Resolution Imaging

- Given: LR images and some information about the motion parameters

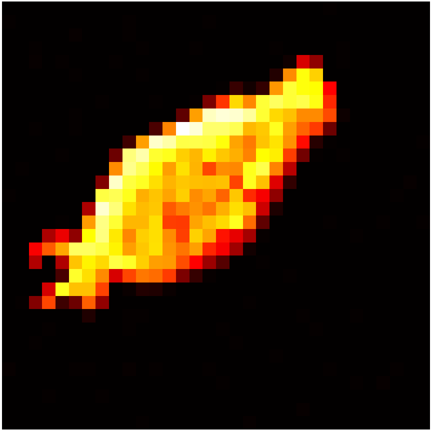
1−th low resolution image



Application: Super-Resolution Imaging

- Given: LR images and some information about the motion parameters

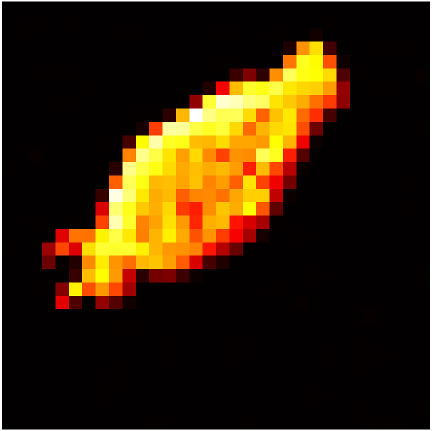
8−th low resolution image



Application: Super-Resolution Imaging

- Given: LR images and some information about the motion parameters

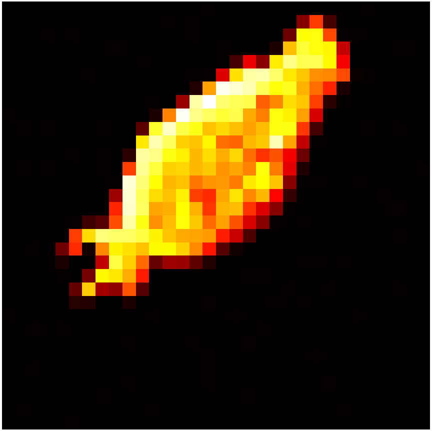
15−th low resolution image



Application: Super-Resolution Imaging

- Given: LR images and some information about the motion parameters

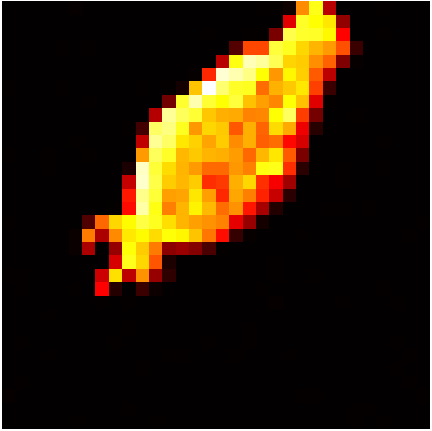
22−th low resolution image



Application: Super-Resolution Imaging

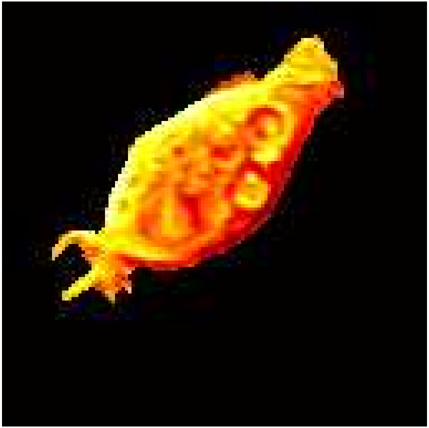
- Given: LR images and some information about the motion parameters

29−th low resolution image



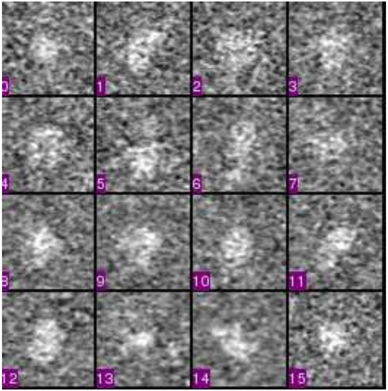
Application: Super-Resolution Imaging

- Given: LR images and some information about the motion parameters
- Goal: Improve parameters and approximate HR image



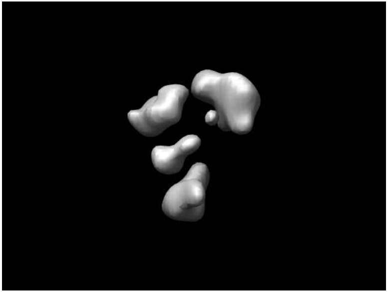
Application: Tomographic Imaging

- Given: 2D projection images
- Goal: Approximate 3D volume



Application: Tomographic Imaging

- Given: 2D projection images
- Goal: Approximate 3D volume



What is an Ill-Posed Inverse Problem?

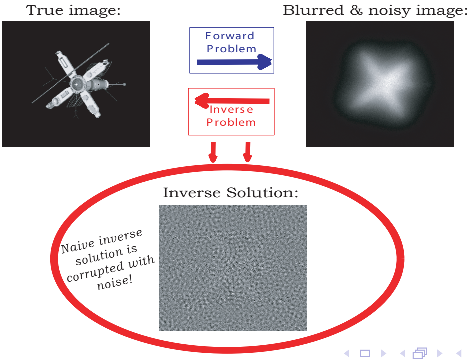
Hadamard (1923): A problem is ill-posed if the solution

- does not exist,
- is not unique, or
- does not depend continuously on the data.

What is an Ill-Posed Inverse Problem?

Hadamard (1923): A problem is ill-posed if the solution

- does not exist,
- is not unique, or
- does not depend continuously on the data.



Outline

1. Regularization for Least Squares Systems
2. High Performance Implementation
3. Polyenergetic Tomosynthesis
4. Concluding Remarks

Outline

1. Regularization for Least Squares Systems
2. High Performance Implementation
3. Polyenergetic Tomosynthesis
4. Concluding Remarks

The Linear Problem

$$\mathbf{b} = \mathbf{A}\mathbf{x} + \varepsilon$$

where

- $\mathbf{x} \in \mathcal{R}^n$ - true data
- $\mathbf{A} \in \mathcal{R}^{m \times n}$ - large, ill-conditioned matrix
- $\varepsilon \in \mathcal{R}^m$ - noise, statistical properties may be known
- $\mathbf{b} \in \mathcal{R}^m$ - known, observed data

Goal: Given **b** and **A**, compute approximation of **x**

Regularization

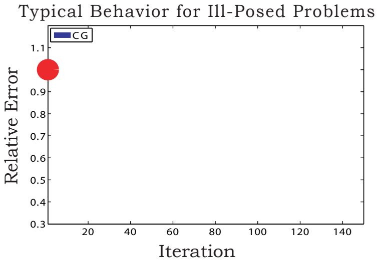
Tikhonov Regularization

$$\min_{\mathbf{x}} \left\{ ||\mathbf{b} - \mathbf{A}\mathbf{x}||_2^2 + \lambda^2 ||\mathbf{L}\mathbf{x}||_2^2 \right\} \quad \Leftrightarrow \quad \min_{\mathbf{x}} \left\| \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix} - \begin{bmatrix} \mathbf{A} \\ \lambda \mathbf{L} \end{bmatrix} \mathbf{x} \right\|_2$$

- Selecting a good regularization parameter, λ, is difficult
  - Discrepancy Principle
  - Generalized Cross-Validation
  - L-curve
- Difficult for large-scale problems

Illustration of Semi-convergence Behavior

$$\min_{\mathbf{x}} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2$$



Iteration 0

Illustration of Semi-convergence Behavior

$$\min_{\mathbf{x}} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2$$

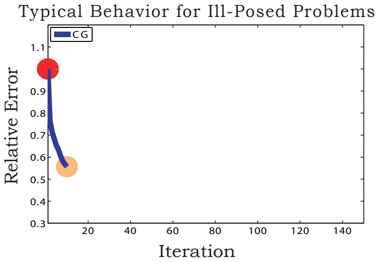
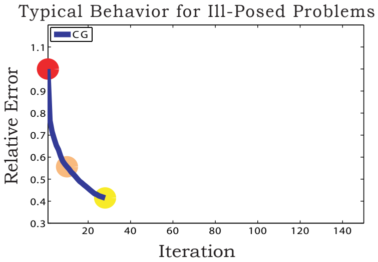


Illustration of Semi-convergence Behavior

$$\min_{\mathbf{x}} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2$$



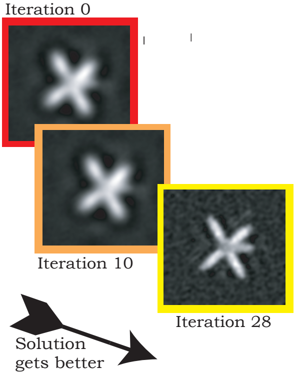


Illustration of Semi-convergence Behavior

$$\min_{\mathbf{x}} \|\mathbf{b} - \mathbf{Ax}\|_2$$

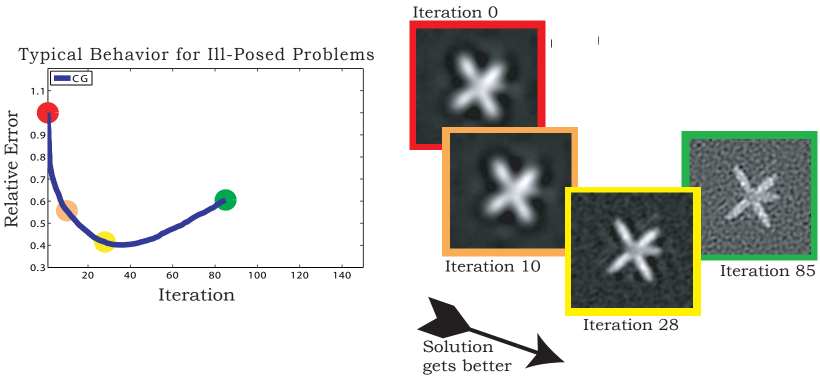
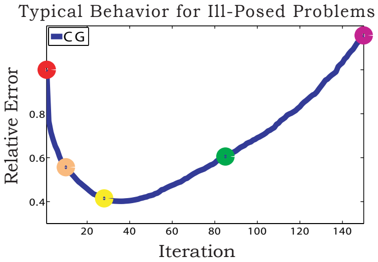


Illustration of Semi-convergence Behavior

$$\min_{\mathbf{x}} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2$$



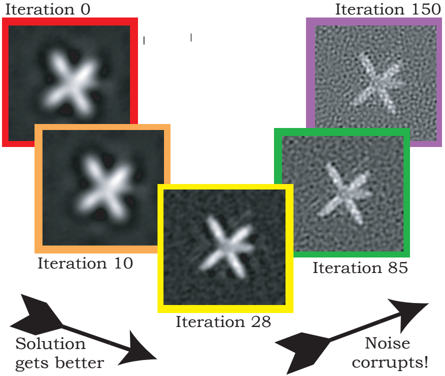
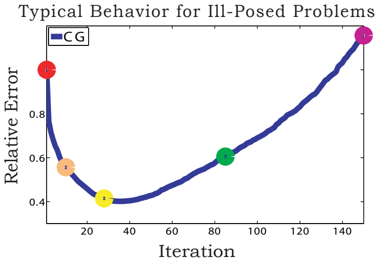
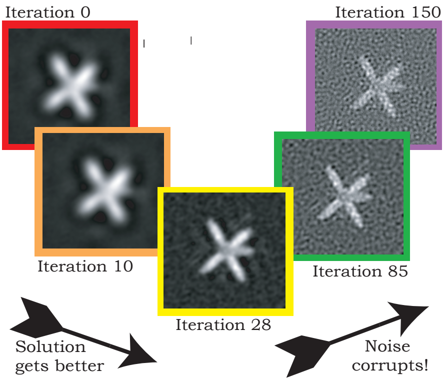


Illustration of Semi-convergence Behavior

$$\min_{\mathbf{x}} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2$$

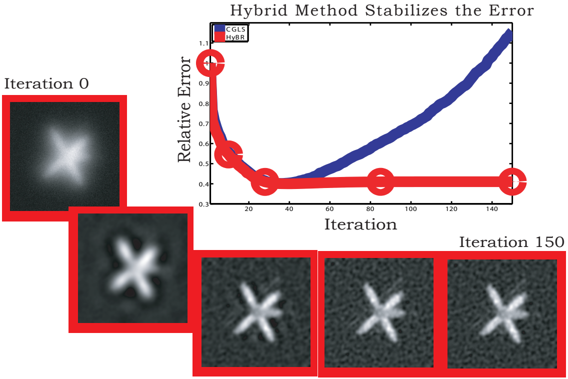




Either find a good stopping criteria or ...

Motivation to use a Hybrid Method

... avoid semi-convergence behavior altogether!



Previous Work on Hybrid Methods

Regularization embedded in iterative method:

- O'Leary and Simmons, SISSC, 1981.
- Björck, BIT 1988.
- Björck, Grimme, and Van Dooren, BIT, 1994.
- Larsen, PhD Thesis, 1998.
- Hanke, BIT 2001.
- Kilmer and O'Leary, SIMAX, 2001.
- Kilmer, Hansen, Espanol, 2006.

Use iterative method to solve regularized problem:

- Golub, Von Matt, Numer. Math.,1991
- Calvetti, Golub, Reichel, BIT, 1999
- Frommer, Maass , SISC, 1999

Lanczos Bidiagonalization(LBD)

Given **A** and **b**, for $k = 1, 2, ...,$ compute

- $\mathbf{W} = \begin{bmatrix} \mathbf{w}_1 & \mathbf{w}_2 & \cdots & \mathbf{w}_k & \mathbf{w}_{k+1} \end{bmatrix}, \quad \mathbf{w}_1 = \mathbf{b}/||\mathbf{b}||$
- $\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_k \end{bmatrix}$
- $\mathbf{B} = \begin{bmatrix} \alpha_1 & & & \\ \beta_2 & \alpha_2 & & \\ & \ddots & \ddots & \\ & & \beta_k & \alpha_k \\ & & & \beta_{k+1} \end{bmatrix}$

where **W** and **Y** have orthonormal columns, and

$$\mathbf{AY} = \mathbf{WB}$$

The Projected Problem

After k steps of LBD, we solve the *projected* LS problem:

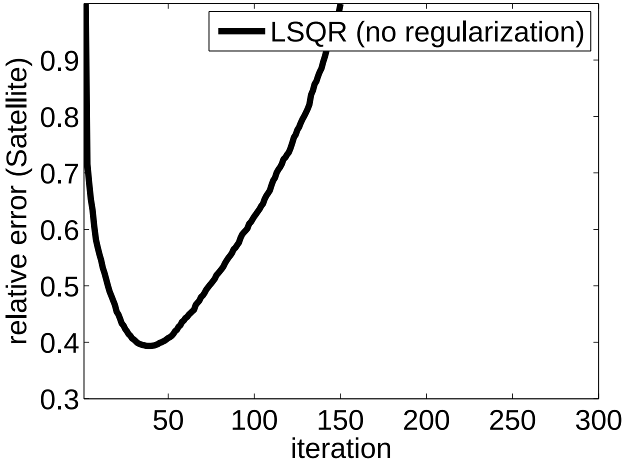
$$\min_{\mathbf{x} \in R(\mathbf{Y})} ||\mathbf{b} - \mathbf{A}\mathbf{x}||_2 = \min_{\mathbf{f}} ||\mathbf{W}^T\mathbf{b} - \mathbf{B}\mathbf{f}||_2$$

where $\mathbf{x} = \mathbf{Y}\mathbf{f}$.

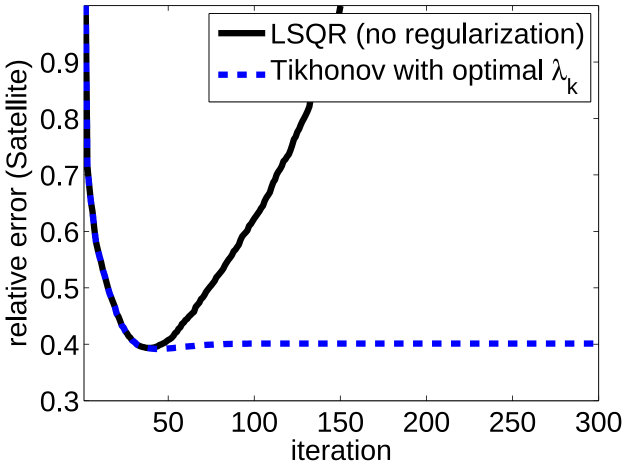
Remarks:

- Ill-posed problem $\Rightarrow$ **B** may be very ill-conditioned.
- **B** is much smaller than **A**
- Standard techniques (e.g. GCV) to find λ and stopping point

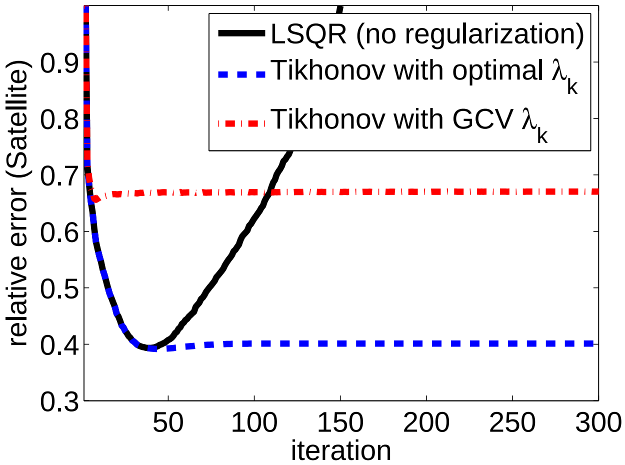
Lanczos Hybrid Method in Action: Satellite



Lanczos Hybrid Method in Action: Satellite



Lanczos Hybrid Method in Action: Satellite



A Novel Approach: Weighted GCV

$$\min_{\mathbf{f}} ||\mathbf{W}^T\mathbf{b} - \mathbf{B}\mathbf{f}||_2$$

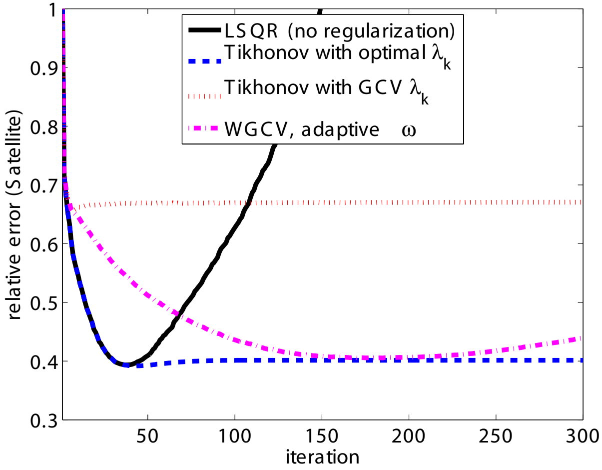
GCV tends to over smooth, use weighted GCV function with $\omega < 1$:

$$G(\omega, \lambda) = \frac{n||(I - \mathbf{B}\mathbf{B}_\lambda^\dagger)\mathbf{W}^T\mathbf{b}||^2}{\left[\text{trace}(I - \omega\mathbf{B}\mathbf{B}_\lambda^\dagger)\right]^2}$$

New adaptive approach to select ω
MATLAB implementation:

$$>> \mathbf{x} = HyBR(\mathbf{A}, \mathbf{b});$$

Results for Satellite



The Nonlinear Problem

$$\mathbf{b} = \mathbf{A}(\mathbf{y})\mathbf{x} + \varepsilon$$

where

$\mathbf{x}$ - true data

$\mathbf{A}(\mathbf{y})$ - large, ill-conditioned matrix defined by parameters $\mathbf{y}$ (registration, blur, etc.)

ε - additive noise

$\mathbf{b}$ - known, observed data

Goal: Approximate $\mathbf{x}$ and improve parameters $\mathbf{y}$

Mathematical Representation

- We want to find **x** and **y** so that

$$\mathbf{b} = \mathbf{A}(\mathbf{y})\mathbf{x} + \mathbf{e}$$

- With Tikhonov regularization, solve

$$\min_{\mathbf{x},\mathbf{y}} \left\| \begin{bmatrix} \mathbf{A}(\mathbf{y}) \\ \lambda \mathbf{I} \end{bmatrix} \mathbf{x} - \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix} \right\|_2^2$$

Some Considerations:

- Problem is linear in **x**, nonlinear in **y**.
- $\mathbf{y} \in \mathcal{R}^p$, $\mathbf{x} \in \mathcal{R}^n$, with $p \ll n$.

Separable Nonlinear Least Squares

Variable Projection Method:

- Implicitly eliminate linear term.
- Optimize over nonlinear term.

Some general references:

Golub and Pereyra, SINUM 1973 (also IP 2003)
Kaufman, BIT 1975
Osborne, SINUM 1975 (also ETNA 2007)
Ruhe and Wedin, SIREV, 1980

Variable Projection Method

Instead of optimizing over both **x** and **y**:

$$\min_{\mathbf{x},\mathbf{y}} \phi(\mathbf{x},\mathbf{y}) = \min_{\mathbf{x},\mathbf{y}} \left\| \begin{bmatrix} \mathbf{A}(\mathbf{y}) \\ \lambda \mathbf{I} \end{bmatrix} \mathbf{x} - \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix} \right\|_2^2$$

Minimize the reduced cost functional:

$$\min_{\mathbf{y}} \psi(\mathbf{y}), \quad \psi(\mathbf{y}) = \phi(\mathbf{x}(\mathbf{y}),\mathbf{y})$$

where $\mathbf{x}(\mathbf{y})$ is the solution of

$$\min_{\mathbf{x}} \phi(\mathbf{x},\mathbf{y}) = \min_{\mathbf{x}} \left\| \begin{bmatrix} \mathbf{A}(\mathbf{y}) \\ \lambda \mathbf{I} \end{bmatrix} \mathbf{x} - \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix} \right\|_2^2$$

Variable Projection Method

Instead of optimizing over both **x** and **y**:

$$\min_{\mathbf{x},\mathbf{y}} \phi(\mathbf{x},\mathbf{y}) = \min_{\mathbf{x},\mathbf{y}} \left\| \begin{bmatrix} \mathbf{A}(\mathbf{y}) \\ \lambda \mathbf{I} \end{bmatrix} \mathbf{x} - \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix} \right\|_2^2$$

Minimize the reduced cost functional:

$$\boxed{\min_{\mathbf{y}} \psi(\mathbf{y}), \quad \psi(\mathbf{y}) = \phi(\mathbf{x}(\mathbf{y}),\mathbf{y})}$$

where $\mathbf{x}(\mathbf{y})$ is the solution of

$$\min_{\mathbf{x}} \phi(\mathbf{x},\mathbf{y}) = \min_{\mathbf{x}} \left\| \begin{bmatrix} \mathbf{A}(\mathbf{y}) \\ \lambda \mathbf{I} \end{bmatrix} \mathbf{x} - \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix} \right\|_2^2$$

Gauss-Newton Algorithm

choose initial $\mathbf{y}_0$

for $k = 0, 1, 2, \ldots$

$$\mathbf{x}_k = \arg\min_{\mathbf{x}} \left\| \begin{bmatrix} \mathbf{A}(\mathbf{y}_k) \\ \lambda_k \mathbf{I} \end{bmatrix} \mathbf{x} - \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix} \right\|_2$$

$$\mathbf{r}_k = \mathbf{b} - \mathbf{A}(\mathbf{y}_k)\,\mathbf{x}_k$$

$$\mathbf{d}_k = \arg\min_{\mathbf{d}} \|\mathbf{J}_\psi \mathbf{d} - \mathbf{r}_k\|_2$$

$$\mathbf{y}_{k+1} = \mathbf{y}_k + \mathbf{d}_k$$

end

Gauss-Newton Algorithm with HyBR

choose initial $\mathbf{y}_0$

for $k = 0, 1, 2, \ldots$

$$\mathbf{x}_k = \arg\min_{\mathbf{x}} \left\| \begin{bmatrix} \mathbf{A}(\mathbf{y}_k) \\ \lambda_k \mathbf{I} \end{bmatrix} \mathbf{x} - \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix} \right\|_2 \quad \Rightarrow \mathbf{x}_k = \text{HyBR}(\mathbf{A}(\mathbf{y}_k), \mathbf{b})$$

$$\mathbf{r}_k = \mathbf{b} - \mathbf{A}(\mathbf{y}_k)\,\mathbf{x}_k$$

$$\mathbf{d}_k = \arg\min_{\mathbf{d}} \|\mathbf{J}_\psi \mathbf{d} - \mathbf{r}_k\|_2$$

$$\mathbf{y}_{k+1} = \mathbf{y}_k + \mathbf{d}_k$$

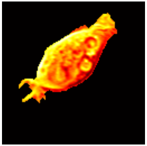
end

Numerical Results: Super-resolution

Inverse Problem

Given:

Goal:



Gauss-Newton Iterations

	Error of $\mathbf{y}_k$	λ_k
0	0.5810	0.2519
1	0.3887	0.2063
2	0.2495	0.1765
3	0.1546	0.1476
4	0.1077	0.1254
5	0.0862	0.1139
6	0.0763	0.1102
7	0.0706	0.1077
8	0.0667	0.1067

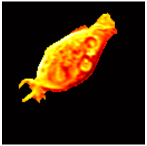
Reconstructed Image

Numerical Results: Super-resolution

Inverse Problem

Given:

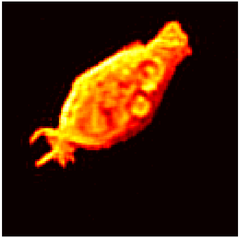
Goal:



Gauss-Newton Iterations

	Error of $\mathbf{y}_k$	λ_k
0	0.5810	0.2519
1	0.3887	0.2063
2	0.2495	0.1765
3	0.1546	0.1476
4	0.1077	0.1254
5	0.0862	0.1139
6	0.0763	0.1102
7	0.0706	0.1077
8	0.0667	0.1067

Reconstructed Image



Outline

1. Regularization for Least Squares Systems
2. High Performance Implementation
3. Polyenergetic Tomosynthesis
4. Concluding Remarks

Mathematical Model

$$\min_{\mathbf{x}} \frac{1}{2}||\mathbf{A}\mathbf{x} - \mathbf{b}||^2$$

where

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_m \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_m \end{bmatrix}$$

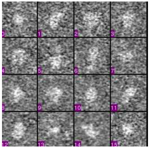
Some Applications:

- Super-resolution
- Tomography - Cryo-Electron Microscopy

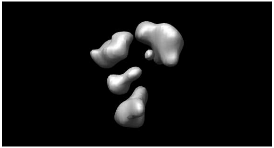
An Application: Cryo-EM

Inverse Problem

Given:



Goal:



$$\min_{\mathbf{x}} \rho(\mathbf{x}) = \frac{1}{2} \sum_{i=1}^{m} ||\mathbf{A}_i \mathbf{x} - \mathbf{b}_i||^2$$

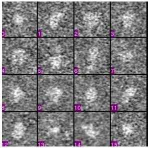
where

- $\mathbf{x} \in \mathcal{R}^{n^3}$ represents the 3-D electron density map
- $\mathbf{b}_i \in \mathcal{R}^{n^2} (i = 1, 2, ..., m)$ represents 2-D projection images
- $\mathbf{A}_i = \mathbf{A}(\mathbf{y}_i) \in \mathcal{R}^{n^2 \times n^3}$ represents projection
  - $\mathbf{y}_i$ - translation parameters and Euler angles

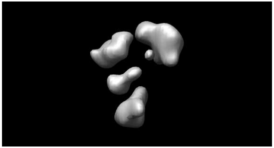
An Application: Cryo-EM

Inverse Problem

Given:



Goal:



$$\min_{\mathbf{x}} \rho(\mathbf{x}) \equiv \frac{1}{2}\sum_{i=1}^{m} ||\mathbf{A}_i\mathbf{x} - \mathbf{b}_i||^2$$

where

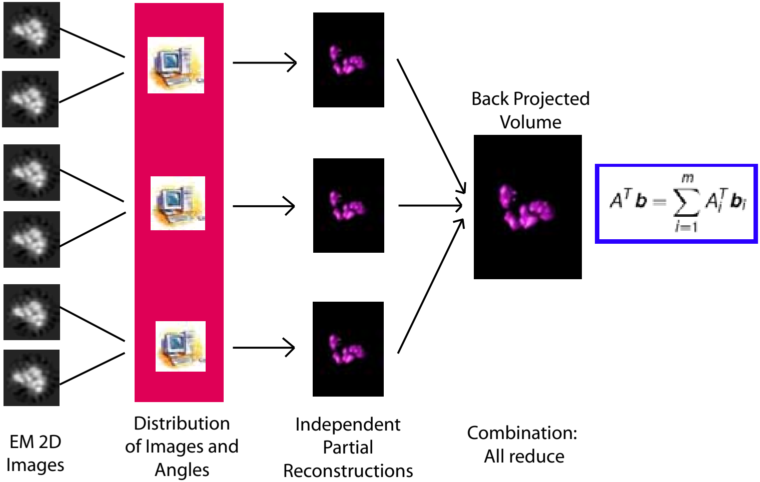
$\mathbf{x} \in \mathcal{R}^{n^3}$ represents the 3-D electron density map

$\mathbf{b}_i \in \mathcal{R}^{n^2} (i = 1, 2, ..., m)$ represents 2-D projection images

$\mathbf{A}_i = \mathbf{A}(\mathbf{y}_i) \in \mathcal{R}^{n^2 \times n^3}$ represents projection

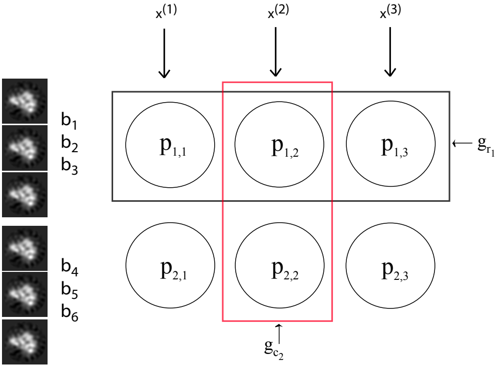
$\mathbf{y}_i$ - translation parameters and Euler angles

Parallelization using 1D data distribution



New Parallelization using 2D data distribution

- Distribute images along rows.
- Distribute volume along columns.



Forward and Back Projection on 2D Topology

$$\mathbf{A}_i = \begin{bmatrix} \mathbf{A}_i^{(1)} & \mathbf{A}_i^{(2)} & \cdots & \mathbf{A}_i^{(n_c)} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} \mathbf{x}^{(1)} \\ \mathbf{x}^{(2)} \\ \vdots \\ \mathbf{x}^{(n_c)} \end{bmatrix}, \nabla\rho = \begin{bmatrix} \nabla\rho_{\mathbf{x}^{(1)}} \\ \nabla\rho_{\mathbf{x}^{(2)}} \\ \vdots \\ \nabla\rho_{\mathbf{x}^{(n_c)}} \end{bmatrix}$$

$$\mathbf{A}_i\mathbf{x} = \sum_{j=1}^{n_c} \mathbf{A}_i^{(j)}\mathbf{x}^{(j)} \quad \Rightarrow \quad \text{All Reduce along Rows}$$

$$\nabla\rho_{\mathbf{x}^{(j)}} = \sum_{i=1}^{m} (\mathbf{A}_i^{(j)})^T \mathbf{r}_{(i)} \quad \Rightarrow \quad \text{All Reduce along Columns}$$

New MPI Parallel Performance

- Good for very large problems
- Adenovirus Data Set: 500 × 500 pixels, 959 (×60) images

n_r	n_c	Wall clock seconds	speedup
137	7	9635	1
959	2	4841	2
959	4	2406	4
959	8	1335	7.2
959	16	609	15.8

- SPARX software package

Outline

1. Regularization for Least Squares Systems
2. High Performance Implementation
3. Polyenergetic Tomosynthesis
4. Concluding Remarks

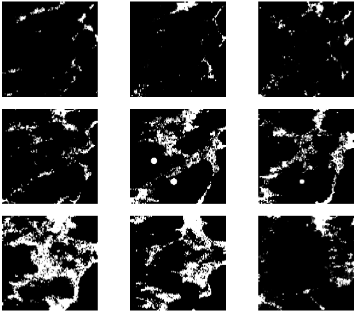
Digital Tomosynthesis

- X-ray Mammography
- Digital Tomosynthesis
- Computed Tomography

An Inverse Problem

- Given: 2D projection images
- Goal: Reconstruct a 3D volume

True Images

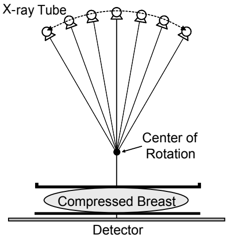


Simulated Problem

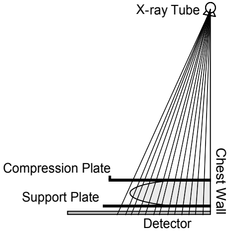
Original object:
$300 \times 300 \times 200$ voxels
($7.5 \times 7.5 \times 5$ cm)

21 projection images:
200×300 pixels
(10×15 cm)
-30° to 30°, every 3°

Reconstruction:
$150 \times 150 \times 50$ voxels
($7.5 \times 7.5 \times 5$ cm)



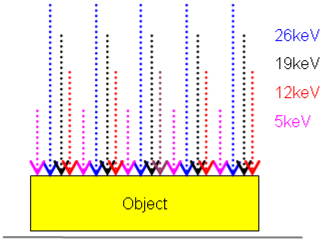
Front view



Side view with X-ray tube at 0°

Polyenergetic Model

- Incident X-ray has a distribution of energies
- 43 energy levels: 5keV - 26keV

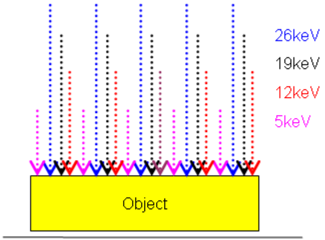


Consequences:

- Beam Hardening: Low energy photons preferentially absorbed
- Unnecessary radiation
- Linear attenuation coefficient depends on energy

Polyenergetic Model

- Incident X-ray has a distribution of energies
- 43 energy levels: 5keV - 26keV



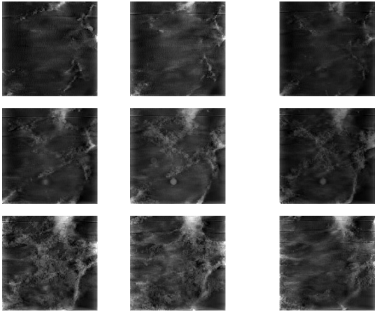
Consequences:

- Beam Hardening: Low energy photons preferentially absorbed
- Unnecessary radiation
- Linear attenuation coefficient depends on energy

Monoenergetic Algorithm

- Lange and Fessler's Convex MLEM Algorithm
- Beam hardening artifacts

Monoenergetic Reconstruction



Previous Methods

Methods for eliminating beam hardening artifacts:

- Dual Energy Methods

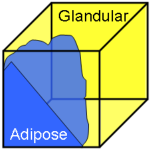
  Alvarez and Macovski (1976), Fessler et al (2002)
- FBP + Segmentation

  Joseph and Spital (1978)
- Filter function based on density

  De Man et al (2001), Elbakri and Fessler (2003)

A Polyenergetic Mathematical Representation

Voxel j



Energy-dependent Attenuation Coefficient:

$$\mu(e)^{(j)} = s(e)x^{(j)} + z(e)$$

where

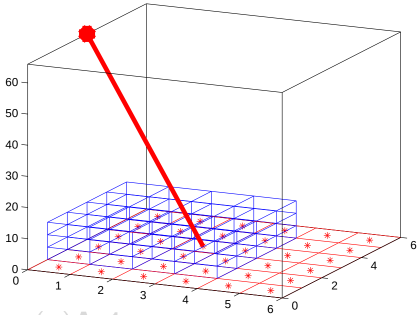
$x^{(j)}$ represents unknown glandular fraction of j^{th} voxel

$s(e)$ and $z(e)$ are known linear fit coefficients

Computing Image Projections

- Ray Trace:

$$\int_{L_i} \mu(e) dl \approx \sum_{j=1}^{N} \mu(e)^{(j)} a^{(ij)}$$



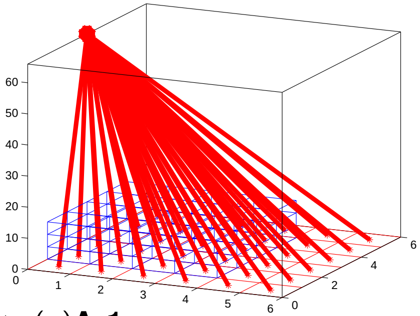
- Vector Notation

$$\mu(e) = s(e)\mathbf{x} + z(e) \quad \Rightarrow \quad s(e)\mathbf{A}_\theta \mathbf{x} + z(e)\mathbf{A}_\theta \mathbf{1}$$

Computing Image Projections

- Ray Trace:

$$\int_{L_i} \mu(e)dl \approx \sum_{j=1}^{N} \mu(e)^{(j)} a^{(ij)}$$



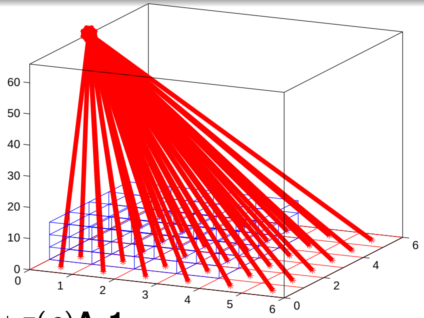
- Vector Notation

$$\mu(e) = s(e)\mathbf{x} + z(e) \quad \Rightarrow \quad s(e)\mathbf{A}_\theta \mathbf{x} + z(e)\mathbf{A}_\theta \mathbf{1}$$

Computing Image Projections

- Ray Trace:

$$\int_{L_i} \mu(e)dl \approx \sum_{j=1}^{N} \mu(e)^{(j)} a^{(ij)}$$



- Vector Notation

$$\mu(e) = s(e)\mathbf{x} + z(e) \quad \Rightarrow \quad s(e)\mathbf{A}_\theta\mathbf{x} + z(e)\mathbf{A}_\theta\mathbf{1}$$

Polyenergetic Projection:

$$\sum_{e=1}^{n_e} \varrho(e) \exp\left(-[s(e)\mathbf{A}_\theta\mathbf{x}_{true} + z(e)\mathbf{A}_\theta\mathbf{1}]\right)$$

Statistical Model

Given **x**, define for pixel *i* the expected value:

$$\bar{b}_\theta^{(i)} = \sum_{e=1}^{n_e} \varrho(e) \exp\left(-[s(e)\mathbf{A}_\theta \mathbf{x} + z(e)\mathbf{A}_\theta \mathbf{1}]\right).$$

Let $\bar{\eta}^{(i)}$ be the statistical mean of the noise.

Then $\bar{b}_\theta^{(i)} + \bar{\eta}^{(i)} \in \mathcal{R}$ is the expected or average observation.

Observed Data: $b_\theta^{(i)} \sim \text{Poisson}(\bar{b}_\theta^{(i)} + \bar{\eta}^{(i)})$

Statistical Model

Likelihood Function:

$$p(\mathbf{b}_\theta, \mathbf{x}) = \prod_{i=1}^{M} \frac{e^{-(\bar{b}_\theta^{(i)} + \bar{\eta}^{(i)})} (\bar{b}_\theta^{(i)} + \bar{\eta}^{(i)})^{b_\theta^{(i)}}}{b_\theta^{(i)}!}$$

Negative Log Likelihood Function:

$$\begin{aligned} -L_\theta(\mathbf{x}) &= -\log p(\mathbf{b}_\theta, \mathbf{x}) \\ &= \sum_{i=1}^{M} (\bar{b}_\theta^{(i)} + \bar{\eta}^{(i)}) - b_\theta^{(i)} \log(\bar{b}_\theta^{(i)} + \bar{\eta}^{(i)}) \end{aligned}$$

Volume Reconstruction

Maximum Likelihood Estimate:

$$\mathbf{x}_{MLE} = \underset{\mathbf{x}}{\text{argmin}} \left\{ \sum_{\theta=1}^{n_\theta} -L_\theta(\mathbf{x}) \right\}$$

Numerical Optimization:

- Gradient Descent:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k \nabla L(\mathbf{x}_k), \quad \text{where} \quad \nabla L(\mathbf{x}_k) = \mathbf{A}^T \mathbf{v}_k$$

- Newton Approach:

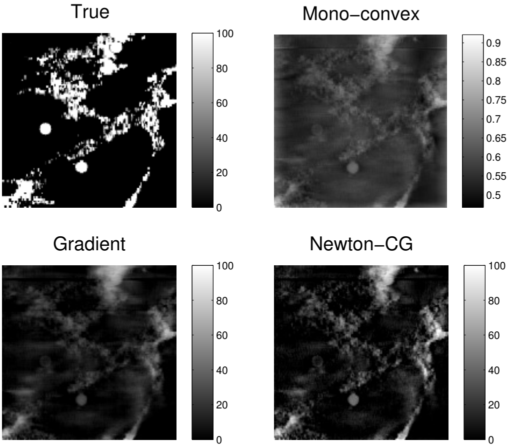
$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k \mathbf{H}_k^{-1} \nabla L(\mathbf{x}_k), \quad \text{where} \quad \mathbf{H}_k = \mathbf{A}^T \mathbf{W}_k \mathbf{A}$$

Numerical Results

- Initial guess: 50% glandular tissue
- Newton-CG inner stopping criteria:
  - Max 50 inner iterations
  - residual tolerance < 0.1

Gradient Descent		Newton Iteration		
Iteration	Relative Error	Iteration	Relative Error	CG Iterations
0	1.7691	0	1.7691	-
1	1.0958	1	1.1045	3
5	0.8752	2	0.8630	2
10	0.8320	3	0.8403	2
25	0.8024	4	0.7925	16

Compare Images



Some Considerations

- Convexity

  Severe nonlinearities $\Rightarrow$ Cost function is not convex

- Regularization

$$\mathbf{x}_{MAP} = \underset{\mathbf{x}}{\operatorname{argmin}}\{-L(\mathbf{x}) + \lambda \mathbf{R}(\mathbf{x})\}$$

  - Need good regularizer, $\mathbf{R}(\mathbf{x})$:

    Huber penalty, Markov Random Fields, Total Variation

  - Need good methods for choosing λ

Outline

1. Regularization for Least Squares Systems
2. High Performance Implementation
3. Polyenergetic Tomosynthesis
4. Concluding Remarks

Concluding Remarks

- Inverse problems arise in many imaging applications.
- Hybrid methods:
  - efficient solvers for large scale LS problems
  - effective linear solvers for nonlinear problems
- Separable nonlinear LS models exploit high level structure
- High performance implementation allows reconstruction of large volumes with high resolution
- Polyenergetic tomosynthesis:
  - Novel mathematical framework
  - Standard optimization made feasible
  - Better reconstructed images

References

- Linear LS (HyBR):
  - Chung, Nagy, O'Leary. ETNA (2008)
  - `http://www.cs.umd.edu/~jmchung/Home/HyBR.html`
- Nonlinear LS:
  - Chung, Haber, Nagy. Inverse Problems (2006)
  - Chung, Nagy. Journal of Physics Conference Series (2008)
  - Chung, Nagy. SISC (Accepted 2009)
- High Performance Computing:
  - Chung, Sternberg, Yang. Int. J. High Perf. Computing (Accepted 2009)
  - Project featured in DOE publication, DEIXIS 2009
- Digital Tomosynthesis:
  - Chung, Nagy, Sechopoulos. (Submitted 2009)

Thank you!