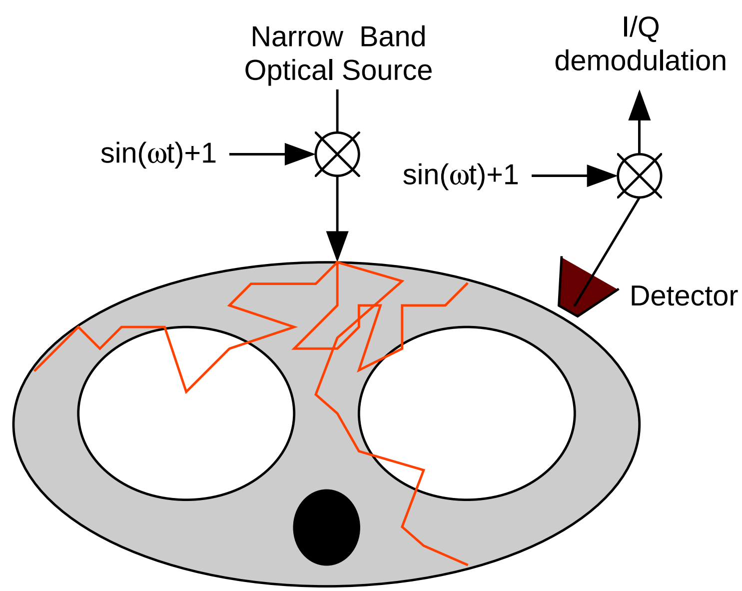
Nonlinear Multigrid Optimization for Bayesian Diffusion Tomography

J.C. Ye[1], *C. A. Bouman*[2], *R. P. Millane*[2], *and K. J. Webb*[2]

[1] University of Illinois at Urbana-Champaign
[2] School of ECE, Purdue University, West Lafayette
http://www.ece.purdue.edu/~bouman

Optical Diffusion Tomography

- Measure light passes through a highly scattering medium
- Light does not travel along a straight line path
- Use measurements to determine unknown absorption cross-section
- Frequency modulate light to reduce measurement noise



Optical Diffusion Model

- The photon flux density, $\psi_k(r,t)$, obeys the **wave** equation

$$\frac{1}{c}\frac{\partial}{\partial t}\psi_k(r,t) - \nabla \cdot D(r)\nabla\psi_k(r,t) + \mu_a(r)\psi_k(r,t) = S(t)\delta(r - s_k)$$

where $D(r) = \dfrac{1}{3(\mu_a(r) + \mu_s'(r))}$

- The frequency modulated light, $\phi_k(r)$, obeys the **PDE**

$$\nabla \cdot D(r)\nabla\phi_k(r) + (-\mu_a(r) + j\omega/c)\phi_k(r) = -\beta\delta(r - s_k).$$

- We need to compute $\mu_a(r)$ from measurements of $\phi_k(r)$

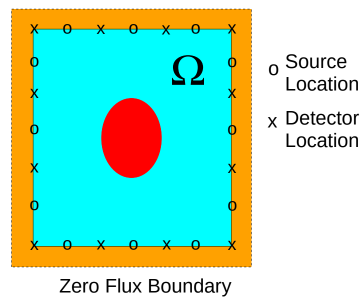
How Does the Forward Model Behave?

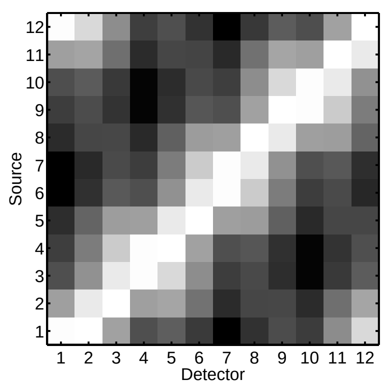
- Nonlinear **forward** model: $\bar{\mathbf{y}} = \mathbf{f}(\mathbf{x})$

  $\bar{\mathbf{y}}$ - noiseless complex optical measurement of $\phi_k(r)$

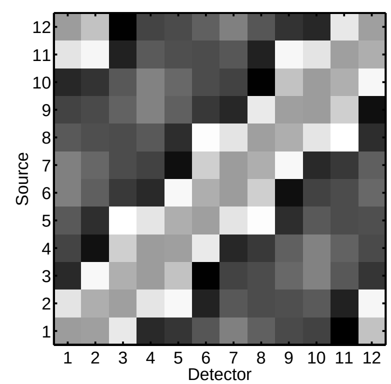
  $\mathbf{x}$ - image of unknown absorbtances, $\mu_a(r)$

- Measurement geometry (8 cm $\times$ 8 cm)





log magnitude of data



phase of data

What Is It Good for?

- Medical Imaging
  - “See” inside tissues at substantial depths
  - Fluorophors increase contrast
  - Tagging agents can target delivery of fluorophors
- Environmental Imaging
  - Airborne smoke and dust can obscure objects
  - Spectroscopic analysis of materials
  - Doppler shifting of envelope
- Nondestructive evaluation
  - Polymer composites
- Representative of a fundamentally new imaging modality
  - Nonlinear forward problem modeled by PDE
  - Potentially low cost
  - Does not require radioisotopes

What is the Problem?

- This inverse problem is REALLY DIFFICULT
  - Nonlinear forward and inverse problem.
  - Each evaluation of forward problem requires the solution of a PDE.
  - Often highly underdetermined
  - Fundamentally 3-D in nature
- Our approach
  - Use a Bayesian inverse framework
  - Develop general purpose computational tools and models
  - **Nonlinear multigrid optimization framework**

Statistical Measurement Model

$\mathbf{y}$ - complex optical measurements of $\phi_k(r)$

$\mathbf{x}$ - image of unknown absorbtances, $\mu_a(r)$

$\mathbf{f}(\mathbf{x})$ - nonlinear forward model

- Using a shot-noise limited measurement model, then

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{(\pi\alpha)^P|\Lambda|^{-1}} \exp\left[-\frac{||\mathbf{y} - \mathbf{f}(\mathbf{x})||_\Lambda^2}{\alpha}\right],$$

where

α - measurements variance

$\Lambda = \text{diag}1/|\mathbf{y}_k|$ - measurement covariance

Prior Model for Absorbtances

- Generalized Gaussian MRF (GGMRF)

$$\log P(x) = -\frac{1}{p\sigma^p} \sum_{\substack{\text{all neighbors} \\ \{s,r\}}} b_{i-j}|x_s - x_r|^p + \text{constant}$$

- Convex for $p > 1$
- Scalable - $\rho(a\Delta) = a^p \rho(\Delta)$ - eliminates need for a "threshold" parameter.
- Simple parameterization

$$\hat{\sigma}_{ML} = \frac{1}{N} \sum_{\substack{\text{all neighbors} \\ \{s,r\}}} b_{i-j}|x_s - x_r|^p$$

Maximum A Posteriori Estimate

- We perform joint MAP estimation of $\mathbf{x}$ and α.
- Estimation of alpha makes global convergence more robust!

$$
\begin{aligned}
&\hat{\mathbf{x}}_{MAP} \\
&= \arg\max_{\mathbf{x}\geq 0} \{\log p(\mathbf{y}|\mathbf{x}) + \log p(\mathbf{x})\} \\
&= \arg\max_{\mathbf{x}\geq \mathbf{0}} \max_{\alpha} \left\{ -\frac{1}{\alpha}||\mathbf{y} - \mathbf{f}(\mathbf{x})||_{\Lambda}^{2} - P\log\alpha - \frac{1}{p\sigma^{p}} \sum_{\{i,j\}\in\mathcal{N}} b_{i-j}|x_i - x_j|^{p} \right\} \\
&= \arg\max_{\mathbf{x}\geq \mathbf{0}} \left\{ -P\log\left(\frac{1}{P}||\mathbf{y} - \mathbf{f}(\mathbf{x})||_{\Lambda}^{2}\right) - \frac{1}{p\sigma^{p}} \sum_{\{i,j\}\in\mathcal{N}} b_{i-j}|x_i - x_j|^{p} \right\}
\end{aligned}
$$

- Intuition: Logarithm term “reduces size” of local minimum.

Multigrid Optimization Approach

- Advantages of multigrid:
  - Fast convergence
  - Robustness to local minima
  - Suitable for non-quadratic optimization (nonlinear problems)
  - Allows simple enforcement of positivity constraints
  - Not just a "multiresolution" algorithm
- Approach:
  - Reformulate nonlinear multigrid in optimization framework
  - Derive general expressions for multigrid recursions
  - Use iterative re-linearization (Born approximation)
  - Iterative estimation of α

Multigrid Cost Functions

- Fine grid cost function is defined by problem

$$c^{(0)}(\mathbf{x}^{(0)}) = \frac{1}{\alpha}||\mathbf{z} - \mathbf{A}\mathbf{x}^{(0)}||_\Lambda^2 + \frac{1}{p\sigma^p} \sum_{\{i,j\}\in\mathcal{N}} b_{i-j} \left| x_i^{(0)} - x_j^{(0)} \right|^p$$

- Choose coarse grid cost functions which are a good approximation to fine grid

$$c^{(k)}(\mathbf{x}^{(k)}) = \frac{1}{\alpha}||\mathbf{z}^{(k)} - \mathbf{A}^{(k)}\mathbf{x}^{(k)}||_\Lambda^2 + \frac{4^k}{p\sigma^p} \sum_{\{i,j\}\in\mathcal{N}} b_{i-j} \left| \frac{x_i^{(1)} - x_j^{(1)}}{2^k} \right|^p$$

- We will approximately correct any errors later

General 2-Grid Optimization Approach

- Let $\mathbb{I}_{(1)}^{(0)}$ and $\mathbb{I}_{(0)}^{(1)}$ be interpolation and decimation operators
- 2-Grid Algorithm:
  1. Approximately optimize fine grid cost function $c^{(0)}(\mathbf{x}^{(0)})$
  2. Initalized coarse grid to $\mathbf{x}^{(1)} \leftarrow \mathbb{I}_{(0)}^{(1)}\hat{\mathbf{x}}^{(0)}$, then approximately optimize coarse grid cost function $c^{(1)}(\mathbf{x}^{(1)})$
  3. Update fine grid result

$$\mathbf{x}^{(0)} \leftarrow \hat{\mathbf{x}}^{(0)} + \mathbb{I}_{(1)}^{(0)}(\mathbf{x}^{(1)} - \mathbb{I}_{(0)}^{(1)}\hat{\mathbf{x}}^{(0)})$$

- Problem
  - True solution is not a fixed point of algorithm!
  - Coarse grid cost function needs to be corrected

Fine Grid Residual Term

- Use fine grid solution to compute correction term

$$\min_{\mathbf{x}^{(1)} \geq \mathbf{0}} \left\{ c^{(1)}\left(\mathbf{x}^{(1)}\right) - \mathbf{r}^{(1)} \, \mathbf{x}^{(1)} \right\}$$

- Choose the row vector $\mathbf{r}^{(1)}$ so that:
  - Gradients of coarse and fine grid cost functions are equal
  - Exact solution is fixed point of algorithm

- General formula for $\mathbf{r}^{(1)}$

$$\mathbf{r}^{(1)} = \nabla c^{(1)}\left(\mathbb{I}_{(0)}^{(1)} \hat{\mathbf{x}}^{(0)}\right) - \nabla c^{(0)}\left(\hat{\mathbf{x}}^{(0)}\right) \mathbb{I}_{(1)}^{(0)}$$

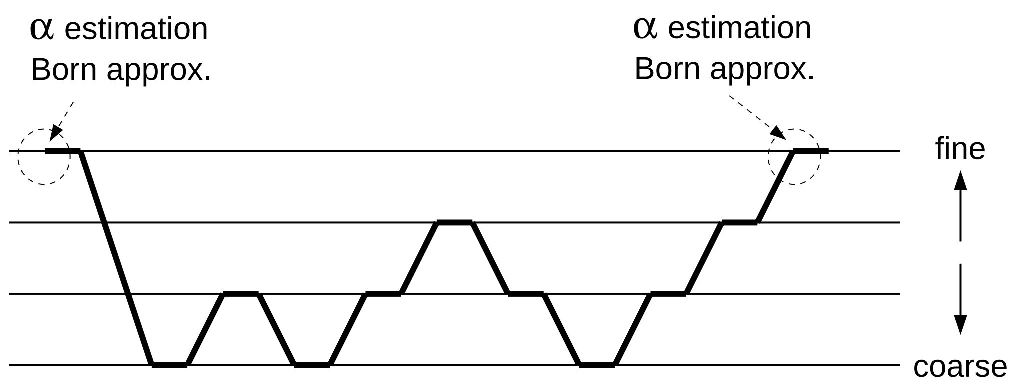
Formula for Residual Term

- Explicit expression for residual term

$$\left[\mathbf{r}^{(1)}\right]_k = \frac{4}{\sigma^p}\sum_{j\in\mathcal{N}_k} b_{k-j}\frac{1}{2}\left|\frac{x_k^{(1)}-x_j^{(1)}}{2}\right|^{p-1} \operatorname{sgn}(x_k^{(1)}-x_j^{(1)})$$
$$-\frac{4}{\sigma^p}\sum_l \left[\mathbb{I}_{(0)}^{(1)}\right]_{k,l}\left(\sum_{m\in\mathcal{N}_l} b_{l-m}\left|x_l^{(0)}-x_m^{(0)}\right|^{p-1}\operatorname{sgn}(x_l^{(0)}-x_m^{(0)})\right)$$

- Optimize $c^{(1)}(x^{(1)})$ to minimize $\mathbf{r}^{(1)}$?

Basic Iteration for Algorithm



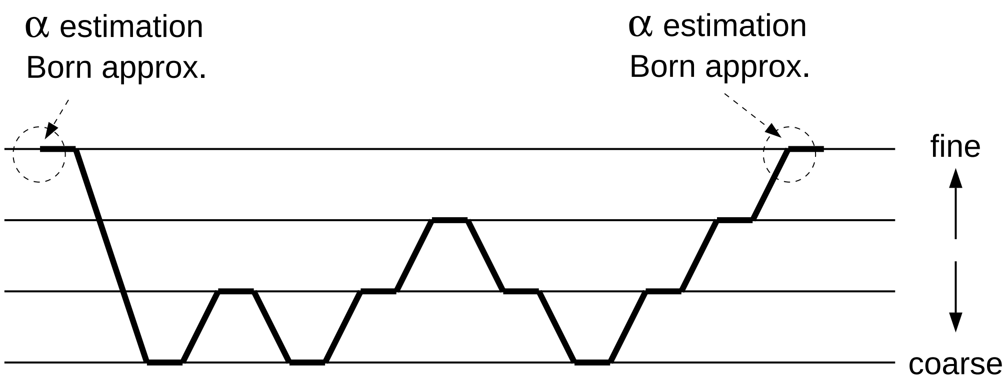
- For each iteration:
  - Update α
  - Linearize about current point (Born approximation)
  - Apply nonlinear multigrid optimization

$$\hat{\alpha} \leftarrow \frac{1}{P}||\mathbf{y} - \mathbf{f}(\hat{\mathbf{x}})||_\Lambda^2$$

$$\mathbf{A} \leftarrow \nabla \mathbf{f}(\hat{\mathbf{x}}) \qquad \mathbf{z} \leftarrow \mathbf{y} - \mathbf{f}(\hat{\mathbf{x}}) + \nabla \mathbf{f}(\hat{\mathbf{x}})\hat{\mathbf{x}}$$

$$\hat{\mathbf{x}} \leftarrow \text{Multigrid} \min_{\mathbf{x} \geq \mathbf{0}} \left\{ \frac{1}{\hat{\alpha}}||\mathbf{y} - \mathbf{A}\mathbf{x}||_\Lambda^2 + \frac{1}{p\sigma^p} \sum_{\{i,j\} \in \mathcal{N}} b_{i-j}|x_i - x_j|^p \right\}$$

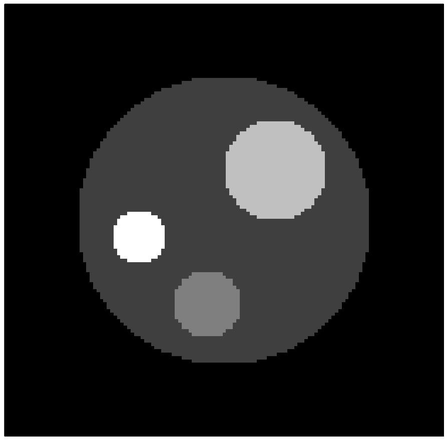
Multigrid Recursion



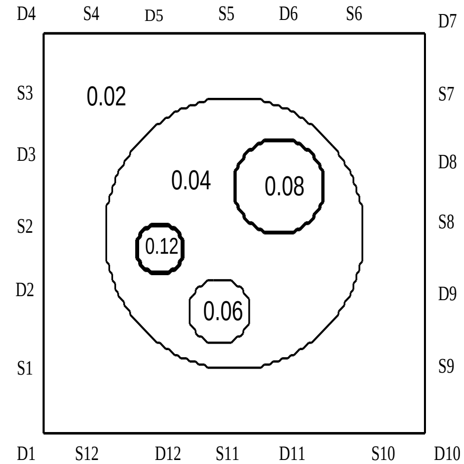
- Multigrid(k)
  - Apply v optimization iterations to $c^{(k)}(\mathbf{x}^{(k)})$
  - Apply Multigrid(k+1) to $c^{(k+1)}(\mathbf{x}^{(k+1)}) + (\mathbf{r}^{(k+1)})^T\mathbf{x}^{(k+1)}$
  - Apply v optimization iterations to $c^{(k)}(\mathbf{x}^{(k)})$

- Fixed grid optimizer must have good **high frequence** convergence
- We use ICD/Born optimizer

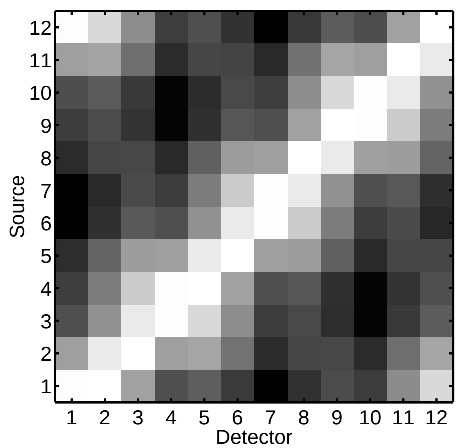
Data (Simulated)



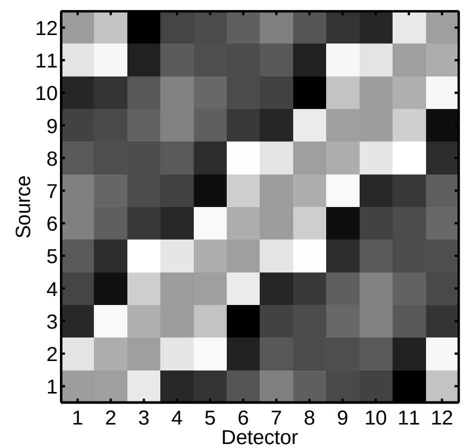
129×129 phantom



absorbtance in cm^{-1}

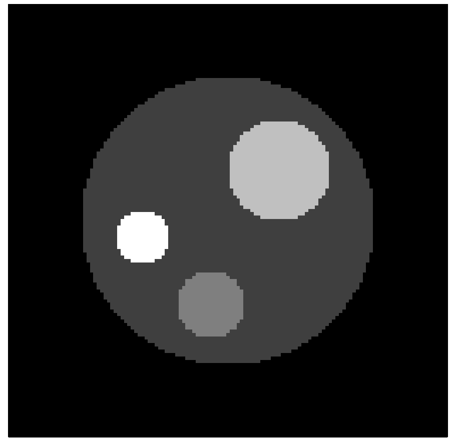


log magnitude of data

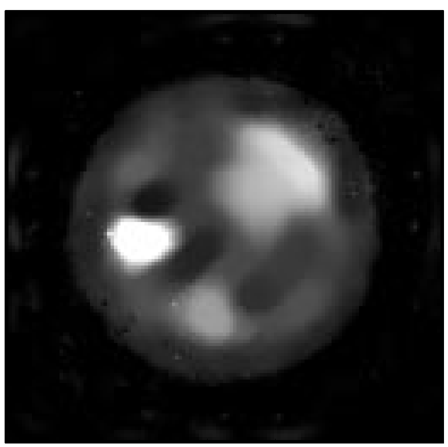


phase of data

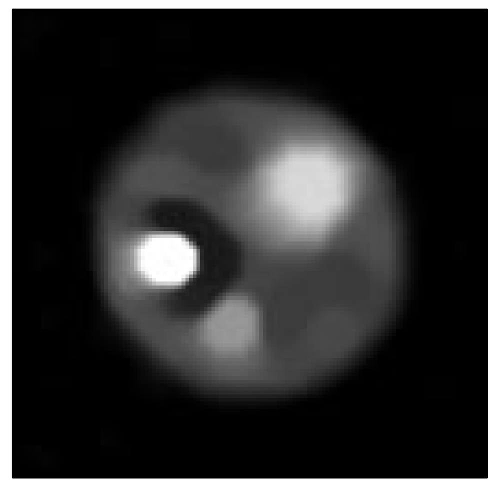
Reconstructions



129×129 phantom

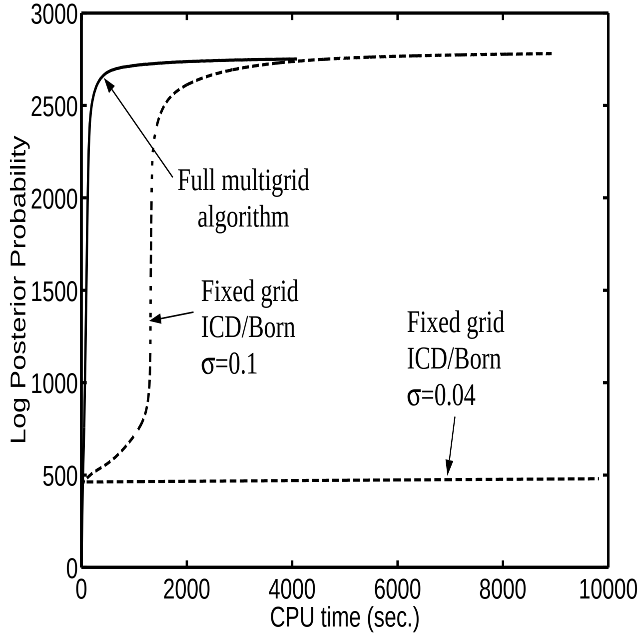


Fixed grid solution

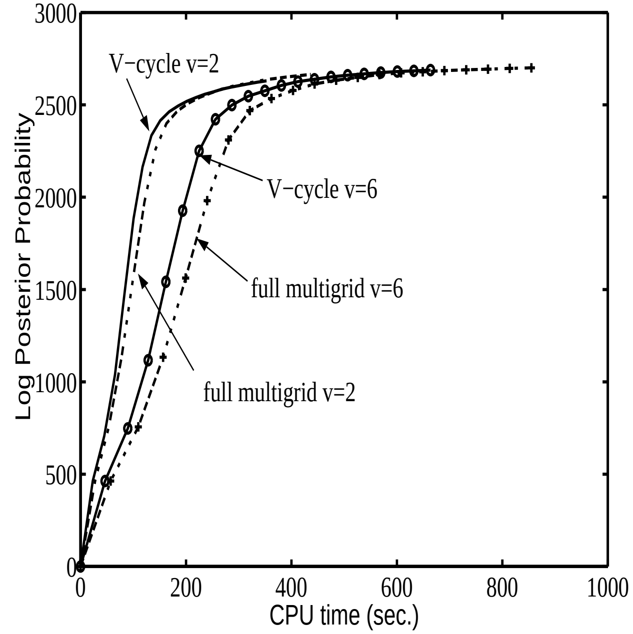


Multigrid solution

Convergence Speed



Fixed grid solution



Multigrid solution

Conclusions

- Optical tomography represents a fundmentally new imaging modality with potentially important applications.
- Optical tomography problem is representative of a very important class of nonlinear inverse problems.
- Multigrid algorithms offer great potential in reducing computation and providing robustness to local minima.
- Multigrid algorithms are well suited to nonlinear optimization problems and the enforcement of positivity constraints.
- Direct formulation of multigrid algorithms in an **optimization framework** has many analytical advantages.