Nonlinear Multigrid Inversion Algorithms with Applications to Statistical Image Reconstruction

Seungseok Oh
School of Electrical and Computer Engineering
Purdue University

Ph.D. Final Examination
December 1, 2004

Inverse Problems

- Forward model

$$x \xrightarrow{f(\cdot)} y = f(x) + noise$$

  x : image

  y : measurement vector

- Inverse Problem

$$y \xrightarrow{\text{How?}} x$$

  - Need to search for x which solves equation $y = f(x)$
  - Can be formulated as optimization problem

- Applications include: tomography, image restoration, remote sensing, machine vision

Regularized Inverse

- First formulate regularized cost function

$$c(x) = \frac{P}{2} \log ||y - f(x)||^2_\Lambda + S(x)$$

where

x : unknown image

y : measurement

$f(x)$: forward model

Λ : weight matrix

P : no. of real-valued unknowns in y

$S(x)$: stabilizing functional

- Then solution is given by minimum

$$\hat{x} \quad = \quad \arg \min_x c(x)$$

Fixed-grid Optimization

- Optimization problem is to minimize

$$c(x;y) = \frac{P}{2}\log||y - f(x)||^2_\Lambda + S(x)$$

- Operates on a fixed-grid resolution to reduce $c(x)$

- Shortcomings
  - Finest scale solutions of forward model are very costly
  - Convergence speed depends on spectral characteristics of error
  - Very sensitive to initial condition
  - Tends to become trapped in local minima

Our Approach: Multigrid Inversion

- Formulate cost function at each scale
- Use different forward model resolutions at each scale

$$y^{(q)} = f^{(q)}(x^{(q)}) + noise$$

  q indexes the forward model scale

- Move up and down between scales to solve problem
  - Coarse-to-fine: Fine is more accurate
  - Fine-to-coarse: Coarse is less accurate!

    $\Rightarrow$ Dynamically adjust cost functions for consistency
- Advantages
  - Dramatically reduces computation of forward solution and inversion
  - Incorporates the true *nonlinear* forward model at *all* scales
  - Rapid and robust convergence

Multigrid Cost Functions

- Original finest scale cost function

$$c(x) = \frac{P}{2} \log ||y - f(x)||^2_\Lambda + S(x)$$

- We will use cost function at scale q

$$c^{(q)}(x^{(q)}; y^{(q)}, r^{(q)}) = \frac{P}{2} \log ||y^{(q)} - f^{(q)}(x^{(q)})||^2_\Lambda + S^{(q)}(x^{(q)}) - r^{(q)} x^{(q)}$$

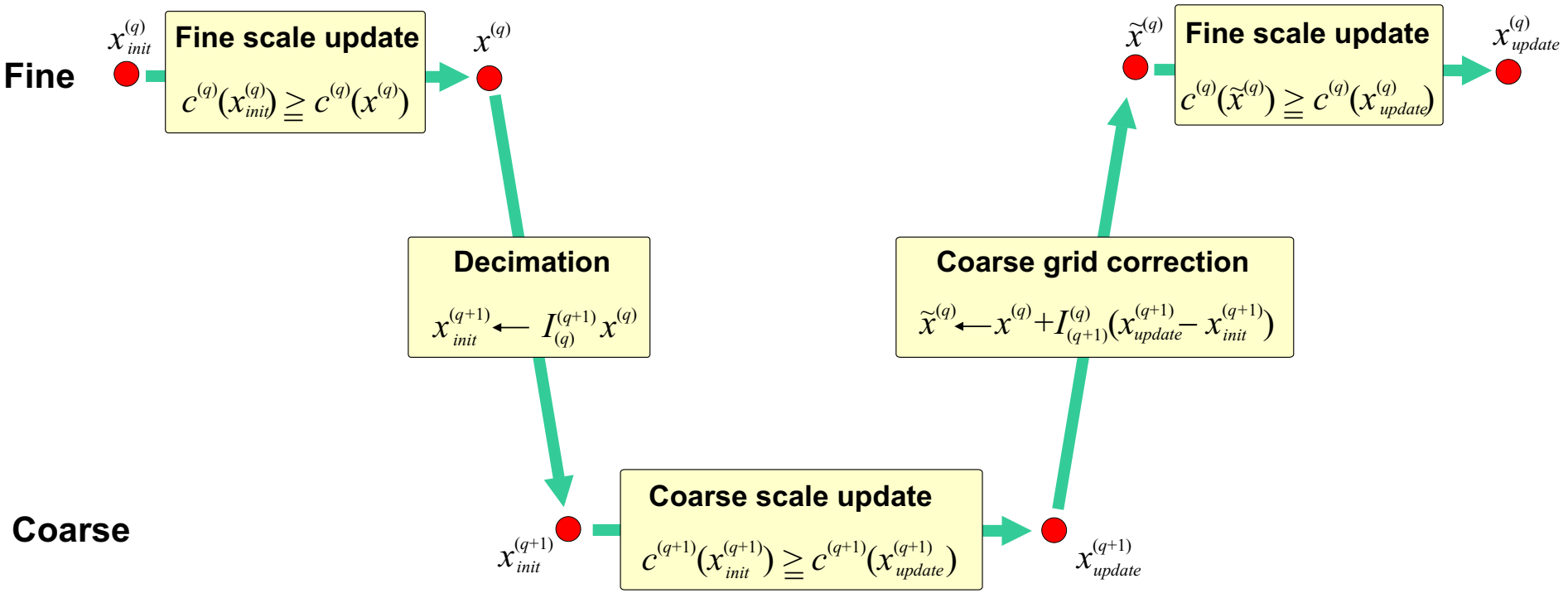
$f^{(q)}(\cdot)$ - coarse scale forward model

$S^{(q)}(\cdot)$ - coarse scale stabilizing function

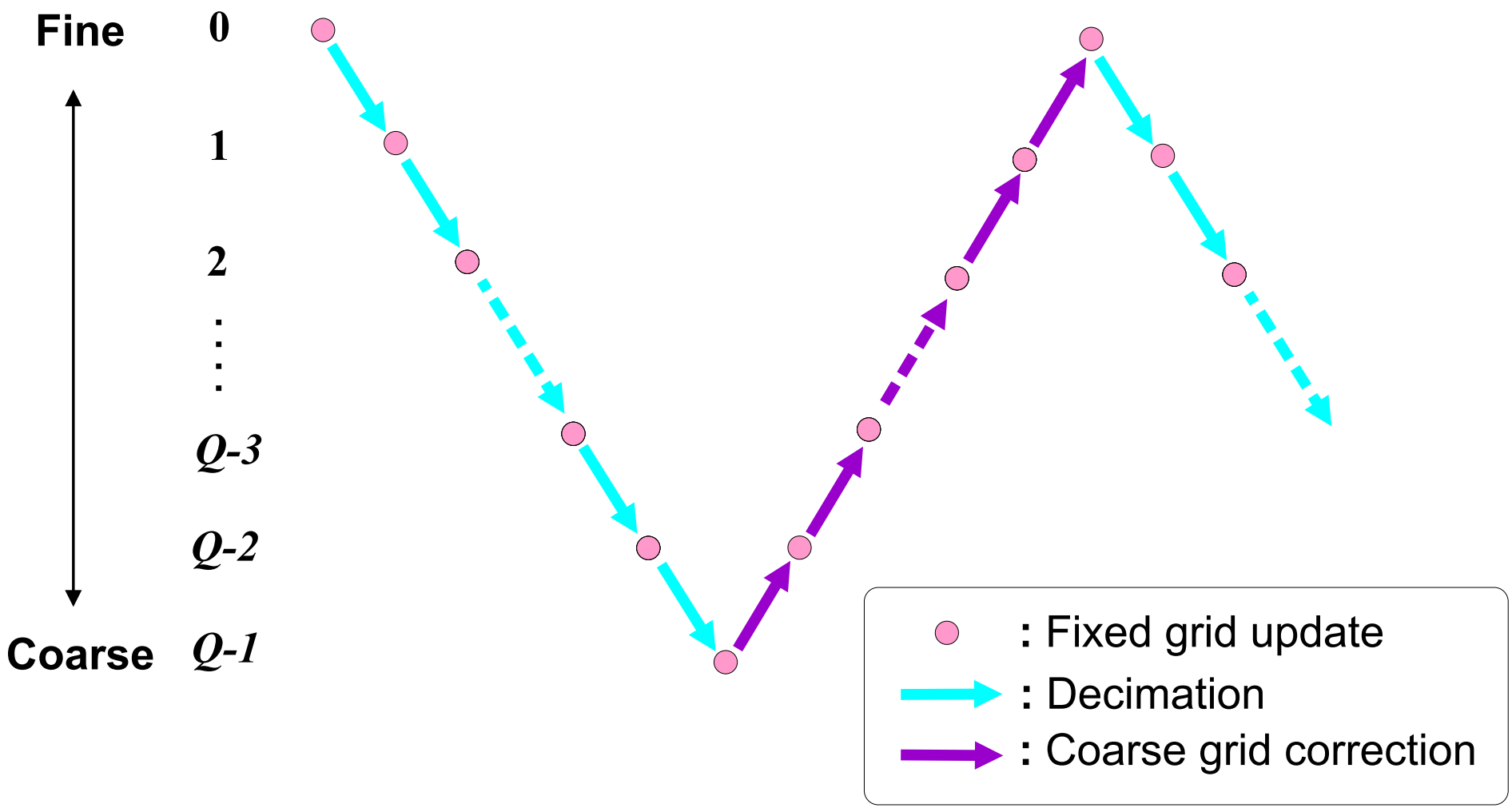
$y^{(q)}$ - coarse scale measurement

$r^{(q)}$ - gradient adjustment factor at scale q

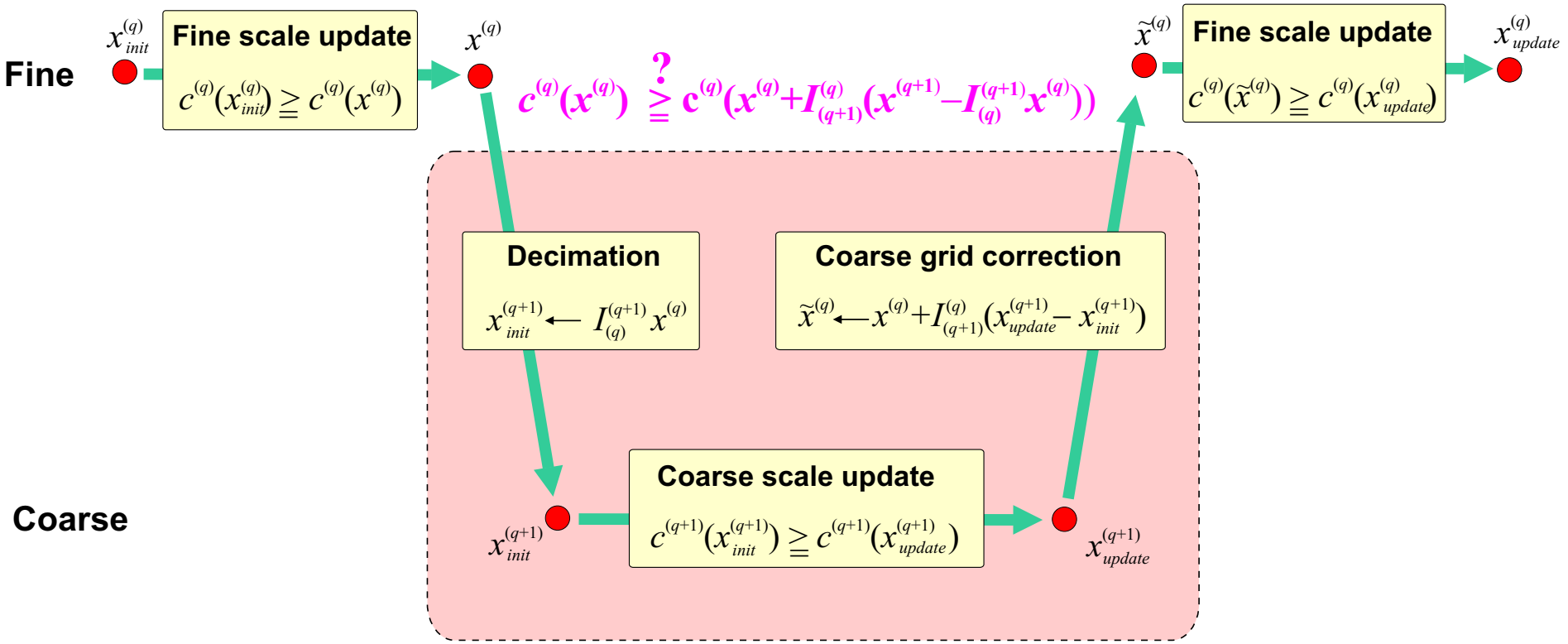
Two-Grid Algorithm



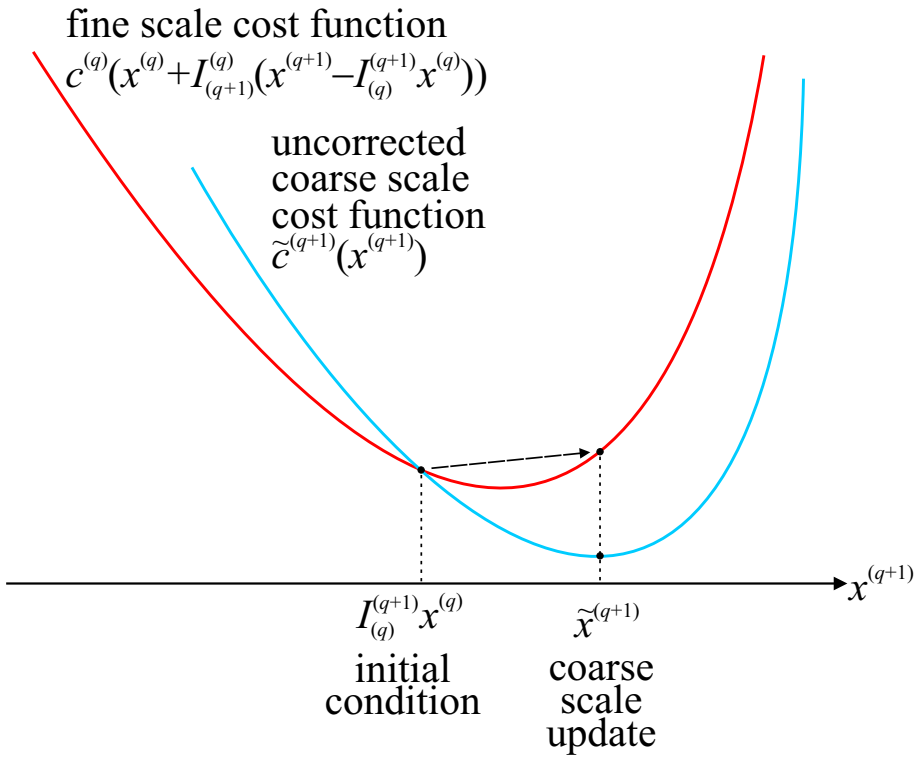
Multigrid-V Algorithm



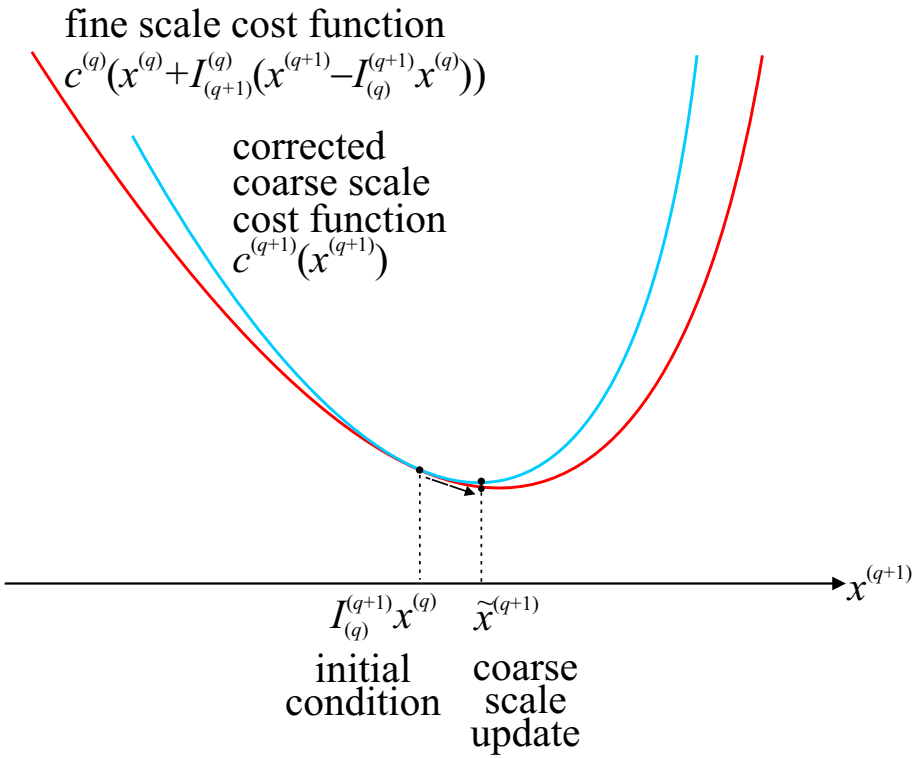
Convergence Issue



Choosing Consistent Cost Functions



Inconsistent Update



Consistent Update

- Coarse scale cost should:
  - Upper bound fine scale cost function
  - Be tangent to fine scale cost function at initial solution

Design Concept for Coarse Scale Cost Functions

- Adjust dynamically $c^{(q+1)}(\cdot)$ so that

$$c^{(q+1)}\left(x^{(q+1)}\right) \cong c^{(q)}\left(x^{(q)} + I^{(q)}_{(q+1)}\left(x^{(q+1)} - I^{(q+1)}_{(q)} x^{(q)}\right)\right) + const.$$

- Done by adjusting $y^{(q+1)}$ and $r^{(q+1)}$ in

$$\begin{aligned} c^{(q+1)}(x^{(q+1)}; y^{(q+1)}, r^{(q+1)}) \quad = \quad & \frac{P}{2} \log ||y^{(q+1)} - f^{(q+1)}(x^{(q+1)})||^2_{\Lambda} \\ & + S^{(q+1)}(x^{(q+1)}) - r^{(q+1)} x^{(q+1)} \end{aligned}$$

Choosing Coarse Scale Measurement $y^{(q+1)}$

- Match error in data term at coarse and fine scales

$$y^{(q+1)} - f^{(q+1)}(I_{(q)}^{(q+1)} x^{(q)}) = y^{(q)} - f^{(q)}(x^{(q)})$$

- Simply use

$$y^{(q+1)} \quad \leftarrow \quad y^{(q)} - \left[f^{(q)}(x^{(q)}) - f^{(q+1)}(I_{(q)}^{(q+1)} x^{(q)}) \right]$$

- This compensates for the forward model mismatch between resolutions

Choosing Gradient Adjustment Factor $r^{(q+1)}$

- Match derivatives in cost function at coarse and fine scales*

$$r^{(q+1)} \leftarrow \nabla \tilde{c}^{(q+1)}(x^{(q+1)})\Big|_{x^{(q+1)}=I_{(q)}^{(q+1)}x^{(q)}} - \left(\nabla \tilde{c}^{(q)}(x^{(q)}) - r^{(q)}\right) I_{(q+1)}^{(q)}$$

where

$$\tilde{c}^{(q)}(x^{(q)}) = \frac{P}{2}\log||y^{(q)} - f^{(q)}(x^{(q)})||_{\Lambda}^{2} + S^{(q)}(x^{(q)})$$

- True optimum solution is a fixed point in multigrid inversion
- Theorem: If the difference between cost functions is convex, then multigrid iterations generate monotone decreasing cost.

* Ye, Bouman, Millane, Webb 1999, 2001.

Choosing Stabilizing Function $S^{(q)}(\cdot)$

- We would like

$$S^{(q)}(x^{(q)}) \cong S(x)$$

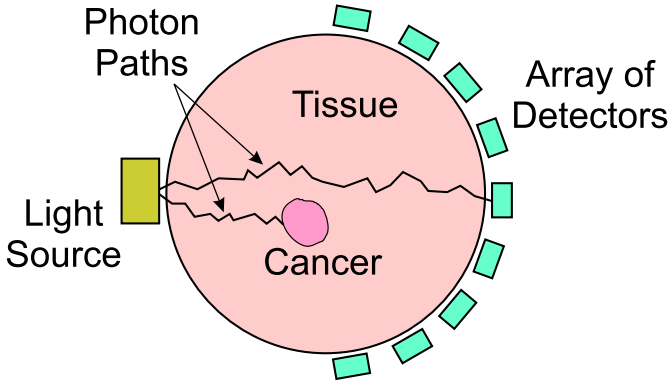
- For the generalized Gaussian Markov random field model (GGMRF)

$$S^{(q)}(x^{(q)}) = \frac{1}{p(\sigma^{(q)})^p} \sum_{\{i,j\} \in \mathcal{N}} b_{i-j} \left| x_i^{(q)} - x_j^{(q)} \right|^p$$

where $\sigma^{(q)} = 2^{q(1-\frac{dim}{p})} \cdot \sigma^{(0)}$

Application: Optical Diffusion Tomography (ODT)

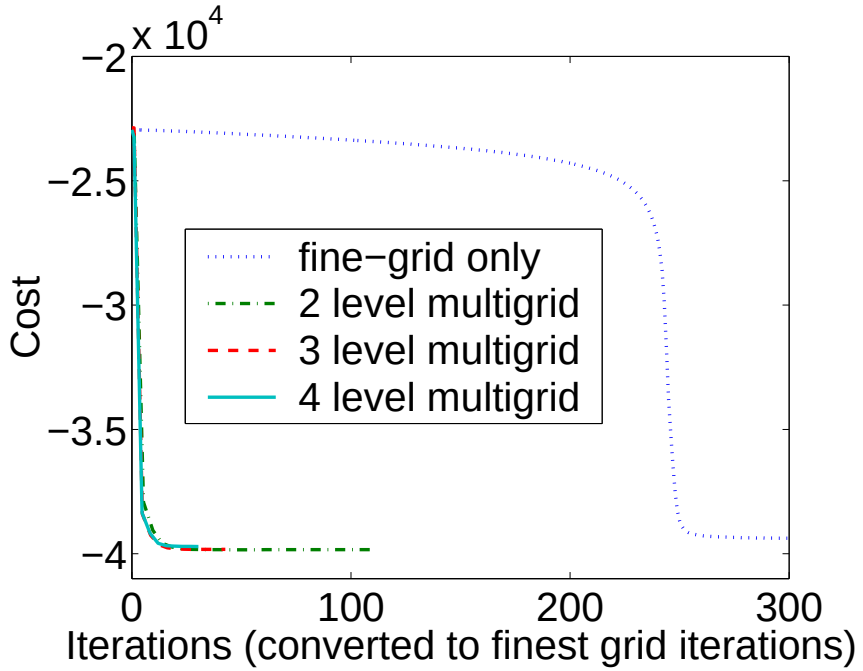
- Measure light that passes through a highly scattering medium
  - y : set of measurements for complex amplitude ϕ_k of modulation envelope
- Determine unknown optical properities of material
  - x : image of unknown absorption μ_a and diffusion D



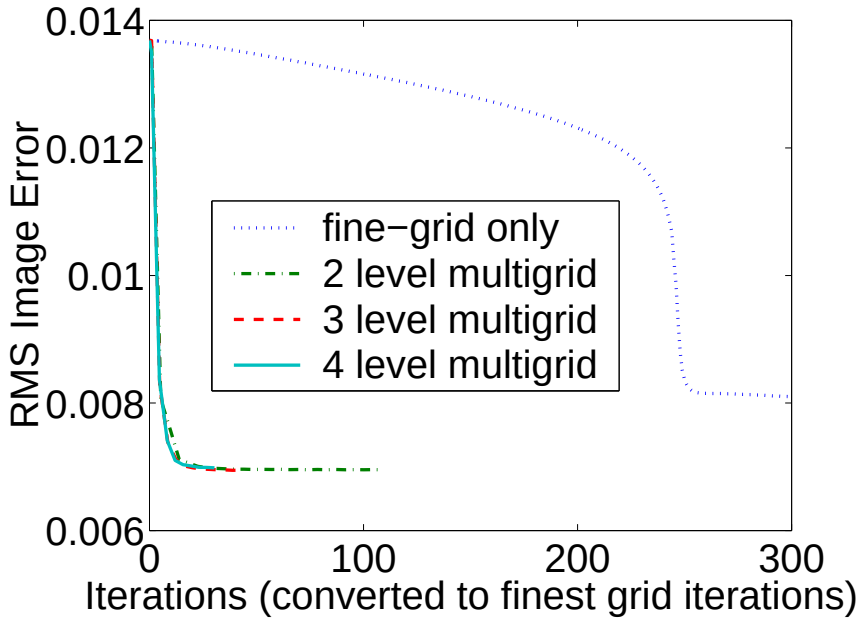
- Nonlinear forward model: $E[Y|X = x] = f(x)$
  - Obeys the frequency-domain diffusion equation

$$\nabla \cdot [D(r)\nabla\phi_k(r)] + [-\mu_a(r) - j\omega/c]\,\phi_k(r) = -\delta(r - a_k)$$

Convergence with Good Initial Condition



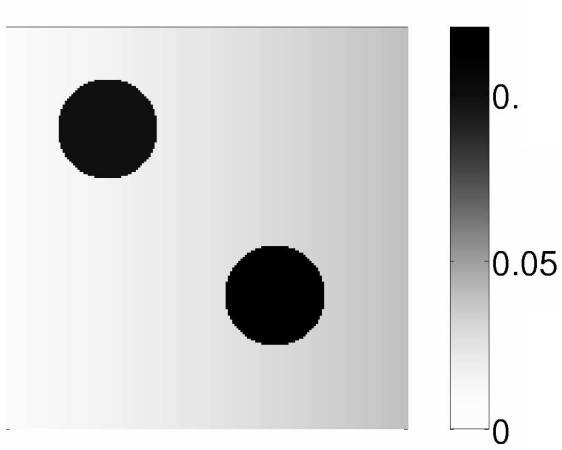
Cost



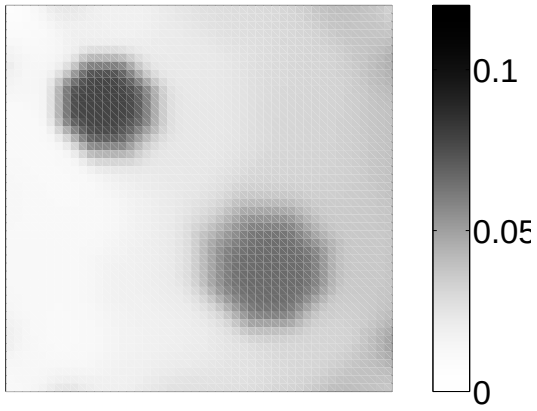
RMS image error

- Initialized with true phantom's average value
- All iterations in units of a single fixed grid iteration

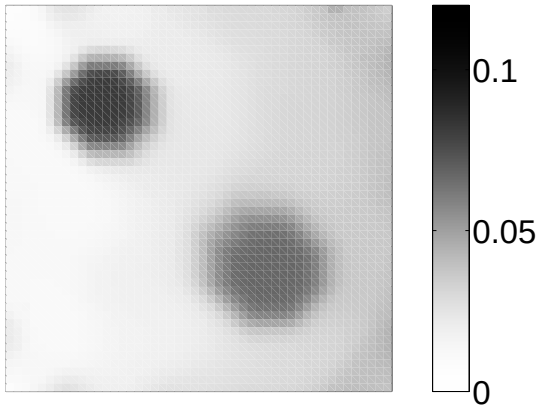
Reconstructions



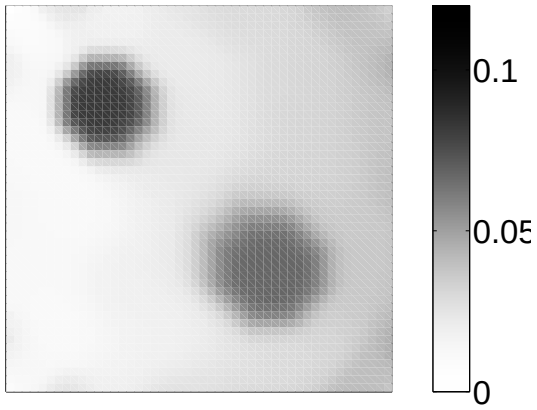
True phantom



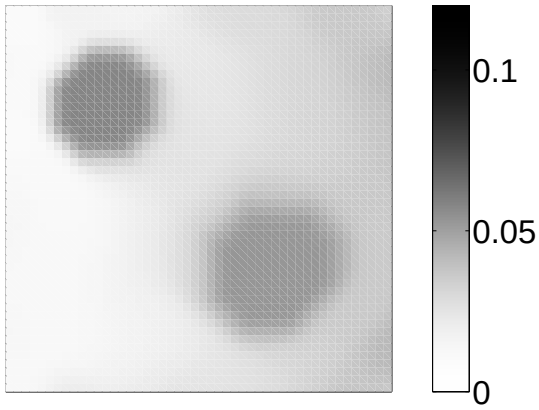
4 level mg (19.35 iter)



3 level mg (19.95 iter)

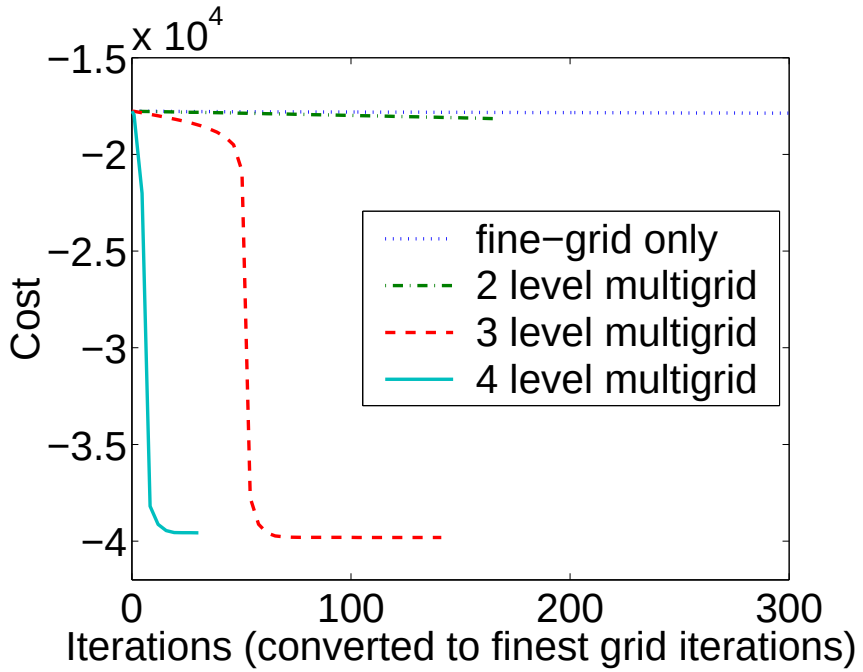


2 level mg (18.24 iter)

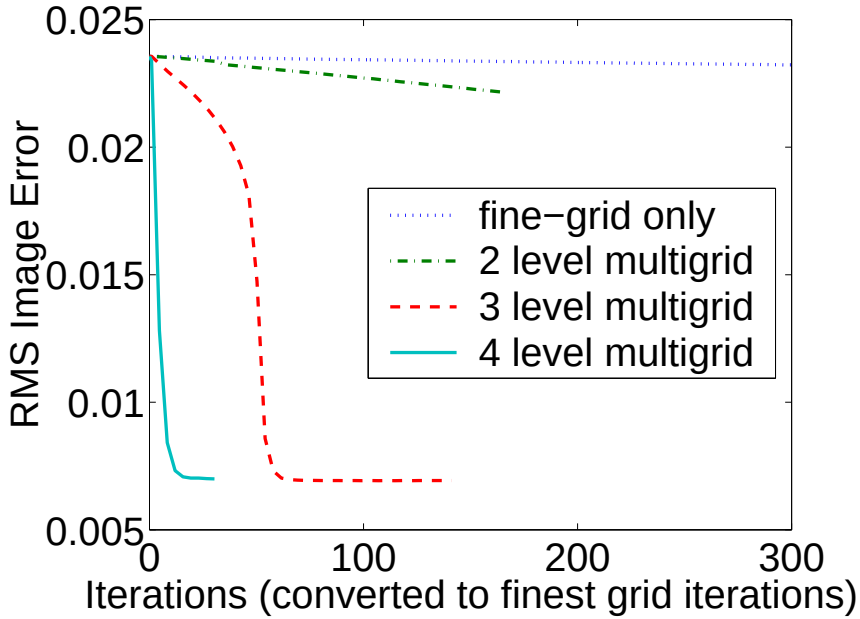


Fixed grid (270 iter.)

Convergence with Poor Initial Condition



Cost



RMS image error

- Initialized with a value of 1.75 $\times$ (true phantom's average value)

Concluding Remarks

- Multigrid Inversion
  - Formulates a set of consistent optimization functions at different scales
  - Changes grid resolution for both forward and inverse problems
  - Preserves the nonlinear characteristics of forward models

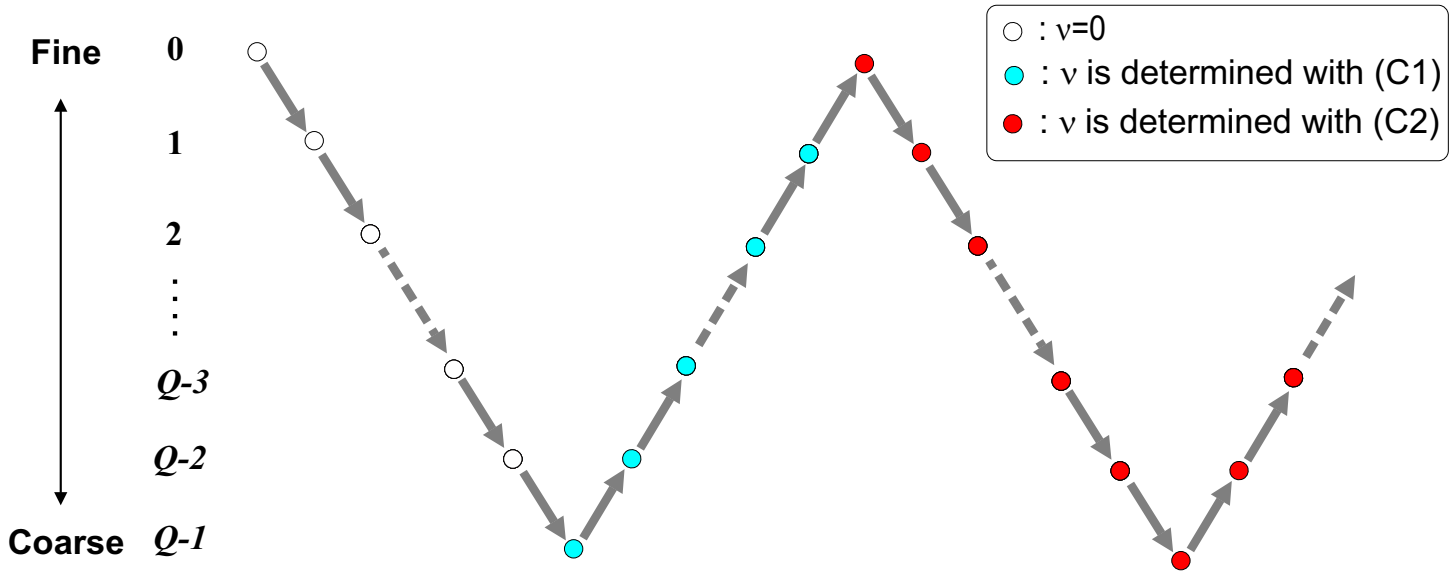
- Advantages
  - Fast and robust convergence
  - Less sensitive to initial condition
  - Widely applicable

Extension to Adaptive Scheme

- Baseline algorithm
  - At each scale q, apply $\nu^{(q)}$ fixed-grid update iterations
  - $\nu^{(q)}$ is a pre-determined value

- Adaptive algorithm
  - Adaptively allocate computation to the scale at which the algorithm can best reduce the cost
  - Adaptively choose $\nu^{(q)}$ at each scale
  - Speedup with only minor code changes

Adaptive Scheme Implementation



- At each scale q, apply fixed-grid update iterations as long as one of the following condition is satisfied:

(C1) Fixed-grid update at the current scale did not converge yet:

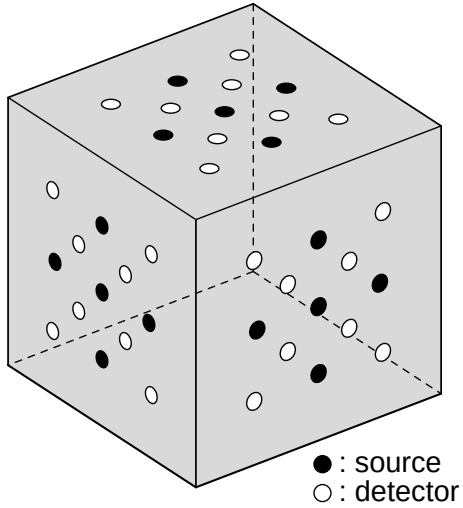
$$\Delta c^{(q)} \geq \Delta_{max} c^{(q)} \times Threshold$$

(C2) Computation decreases cost more efficiently than at the next scale:

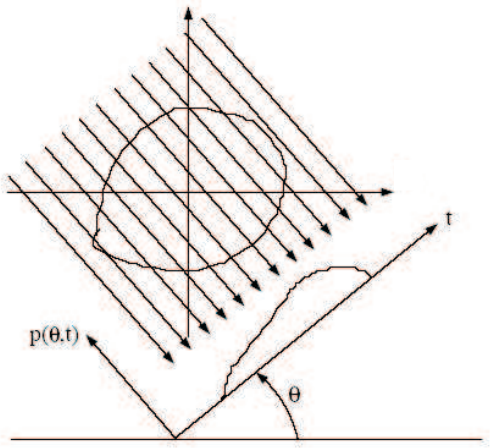
$$\frac{\Delta c^{(q)}}{computation^{(q)}} \geq \frac{\Delta c^{(q_{next})}}{computation^{(q_{next})}}$$

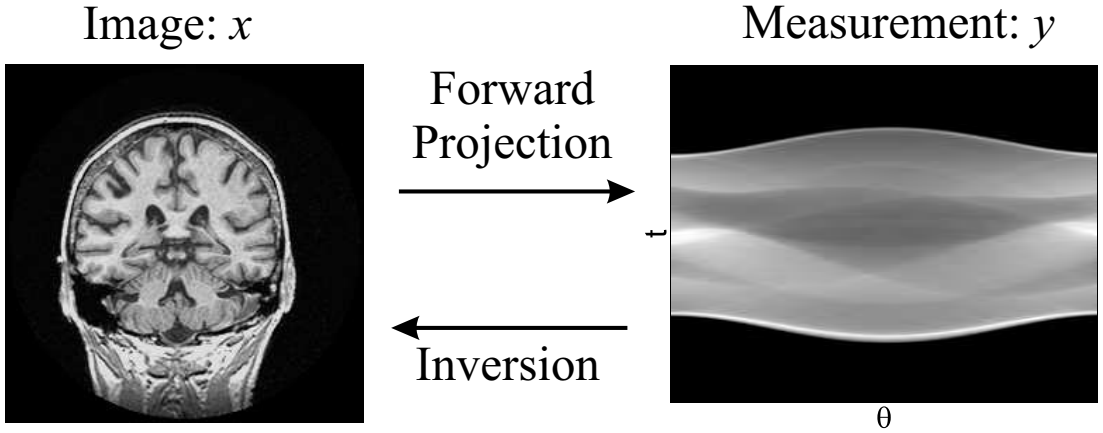
Motivation of Variable Data Resolution

- ODT: sparse data from point sources and detectors



- CT and PET: rich data





The Issue: Variable Data Resolution?

- Multigrid inversion
  - Uses **all** data at each resolution
  - Works well when data is sparse, for example, ODT
- Key issue
  - Coarse resolution $\Rightarrow$ lower resolution images
  - **Same resolution of data at every resolution?**
  - Potentially important for data rich problems, for example, CT

Multigrid Inversion with Variable Data and Image Resolutions

- Extension to variable data resolution
  - Change resolution of both **data** and image at each scale
  - Decimation operator for data may be different than image
    - $J_{(q)}^{(l)}$: decimator/interpolator for **data**
    - $I_{(q)}^{(l)}$: decimator/interpolator for image

- Advantages
  - Reduce computation at coarse scales
  - Each iteration does not require the use of all the data.

Multigrid Cost Function: Gaussian Noise Case

- Original cost function (We will drop $S(x)$ for simplicity)

$$c(x) = ||y - f(x)||^2_\Lambda + S(x)$$

- $J^{(0)}_{(q)} f^{(q)}(x^{(q)})$ - Data domain interpolation of forward model
- Cost function at scale q

$$c^{(q)}(x^{(q)}) = ||y - J^{(0)}_{(q)} f^{(q)}(x^{(q)})||^2_\Lambda - r^{(q)} x^{(q)}$$

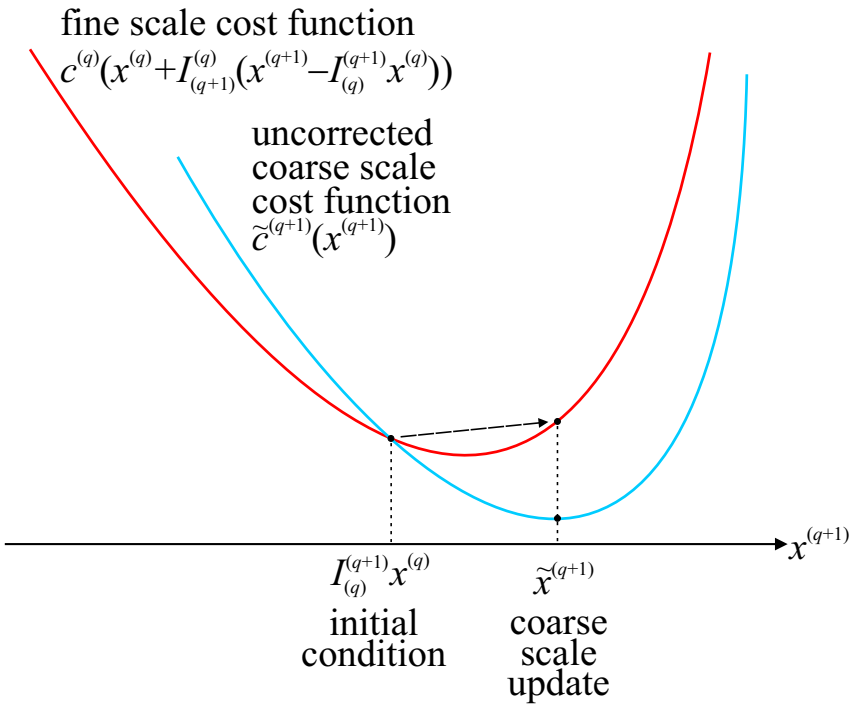
$$= ||J^{(0)}_{(q)} f^{(q)}(x^{(q)})||^2_\Lambda + 2a^T J^{(0)}_{(q)} f^{(q)}(x^{(q)}) - r^{(q)} x^{(q)} + const$$

$$= ||f^{(q)}(x^{(q)})||^2_{\Lambda^{(q)}} + 2{a^{(q)}}^T f^{(q)}(x^{(q)}) - r^{(q)} x^{(q)} + const$$

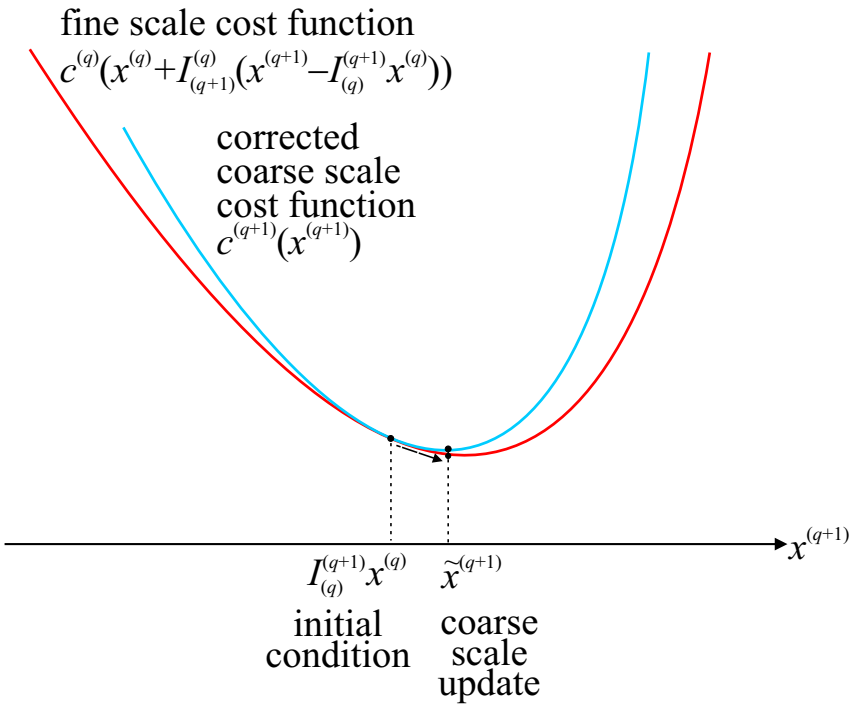
where

$$\Lambda^{(q)} = [J^{(0)}_{(q)}]^T \Lambda^{(0)} J^{(0)}_{(q)}$$
$$a^{(q)} = [J^{(0)}_{(q)}]^T a^{(0)}$$

Choosing $r^{(q)}$



Inconsistent Update



Consistent Update

- Choose $r^{(q)}$ to match gradients between scales

Multigrid Cost Function: Poisson Noise Case

- Original cost function

$$c(x) = \sum_{m=1}^{M} \{f_m(x) - y_m \log f_m(x)\}$$

- Coarse scale cost function

$$c^{(q)}(x^{(q)}) = \frac{M^{(0)}}{M^{(q)}} \sum_{m=1}^{M^{(q)}} \left\{ f_m^{(q)}(x^{(q)}) - y_m^{(q)} \log f_m^{(q)}(x^{(q)}) \right\} - r^{(q)} x^{(q)}$$

  where coarse scale measurement data is computed by data domain decimation

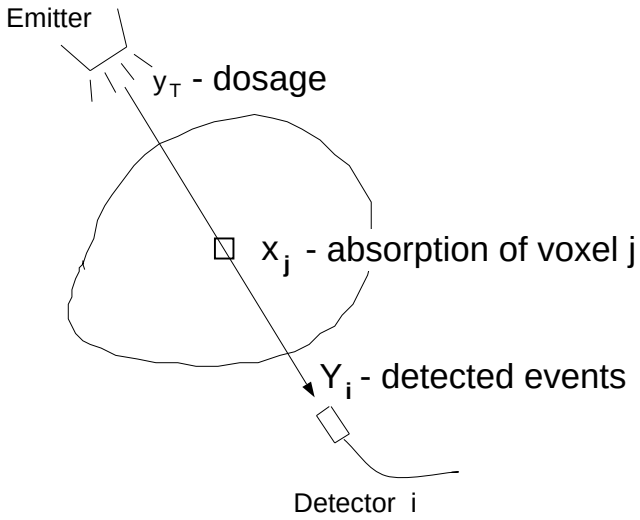
$$y^{(q)} = J_{(0)}^{(q)} y^{(0)}$$

- $r^{(q)}$ is chosen to match gradients between scales

Applications to Transmission / Emission Tomography

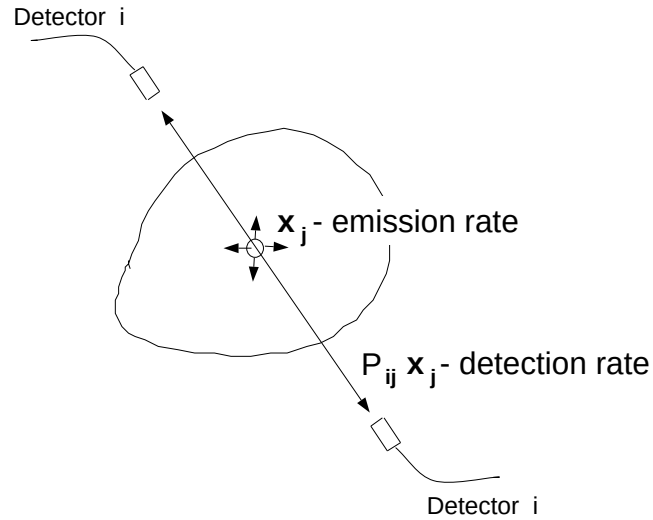
- Determine unknown absorption or emission rate cross-section of material
- Poisson measurement noise model with the mean $f(x)$

Transmission Problem



$$f_i(x) = y_T \exp(-P_{i*}x)$$

Emission Problem



$$f_i(x) = P_{i*}x$$

Multigrid Poisson Mean Vector

- Emission tomography
  - Original Poisson mean

$$f(x) = Px$$

  - Coarse scale Poisson mean

$$f^{(q)}(x^{(q)}) = P^{(q)} x^{(q)}$$

    where

$$P^{(q)} = J^{(q)}_{(0)} P^{(0)} I^{(0)}_{(q)}$$

- Transmission tomography
  - Original Poisson mean

$$f_m(x) = y_T \exp(-P_{m,*} x)$$

  - Coarse scale Poisson mean

$$f^{(q)}_m(x^{(q)}) = y_T \exp(-P^{(q)}_{m,*} x^{(q)})$$

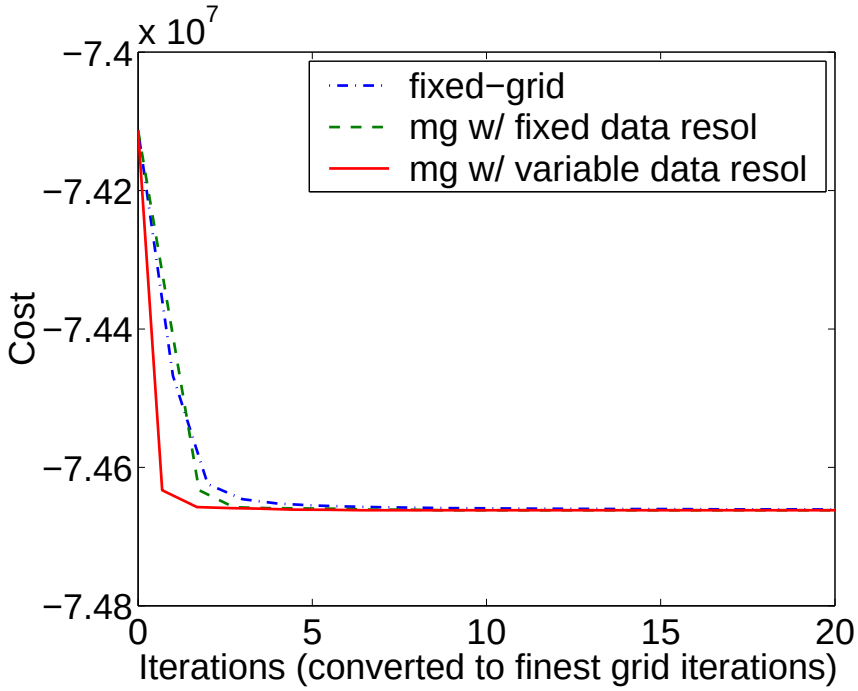
Simulation Experiment

- Measurement
  - 180 angles $\times$ 512 rays/angle
  - Poisson noise was added
  - $y_T = 800$ for transmission tomography



- Reconstruction
  - Image is reconstructed on 513×513 grid points
  - Generalized Gaussian Markov random field image prior is used
  - Iterative coordinate descent (ICD) method is used as fixed-grid update
  - Multigrid algorithms use 3 level multigrid-V recursion

Convergence: Emission Case with Poisson Model



Cost

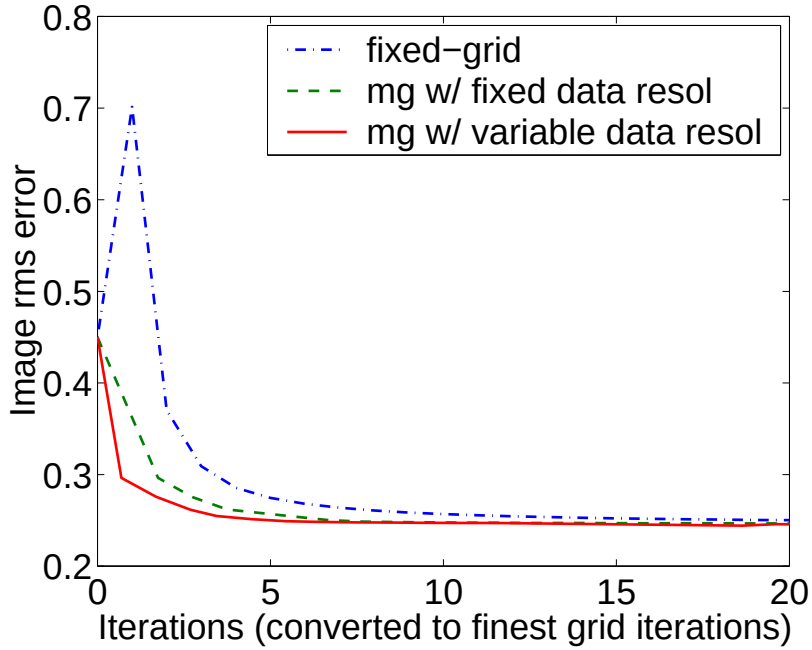
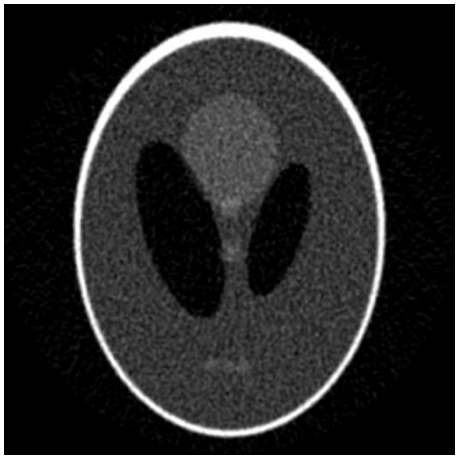


Image root-mean-square error

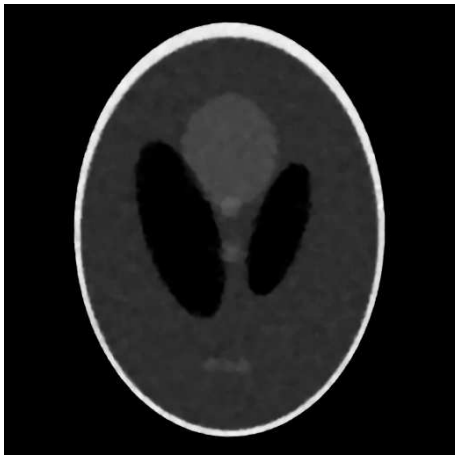
Reconstructions



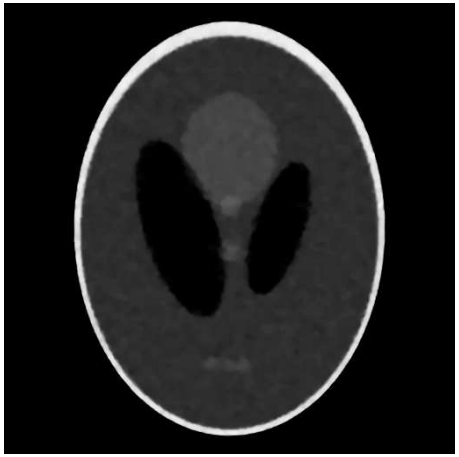
True



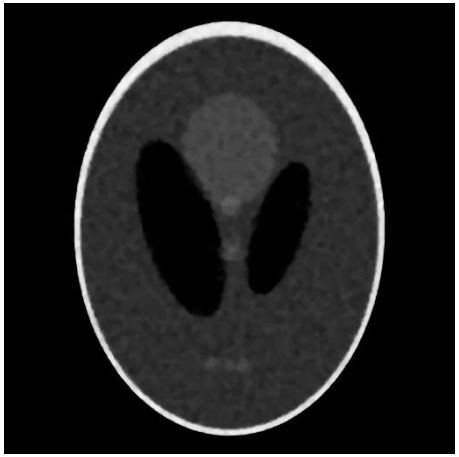
CBP initialization



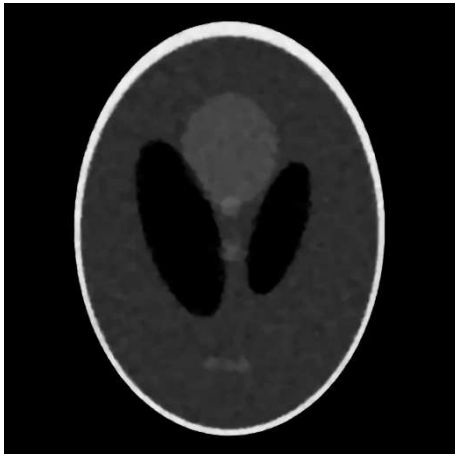
MG w/ var. data resol. (5.31 iter)



MG w/ fi xed data resol. (8.06 iter)



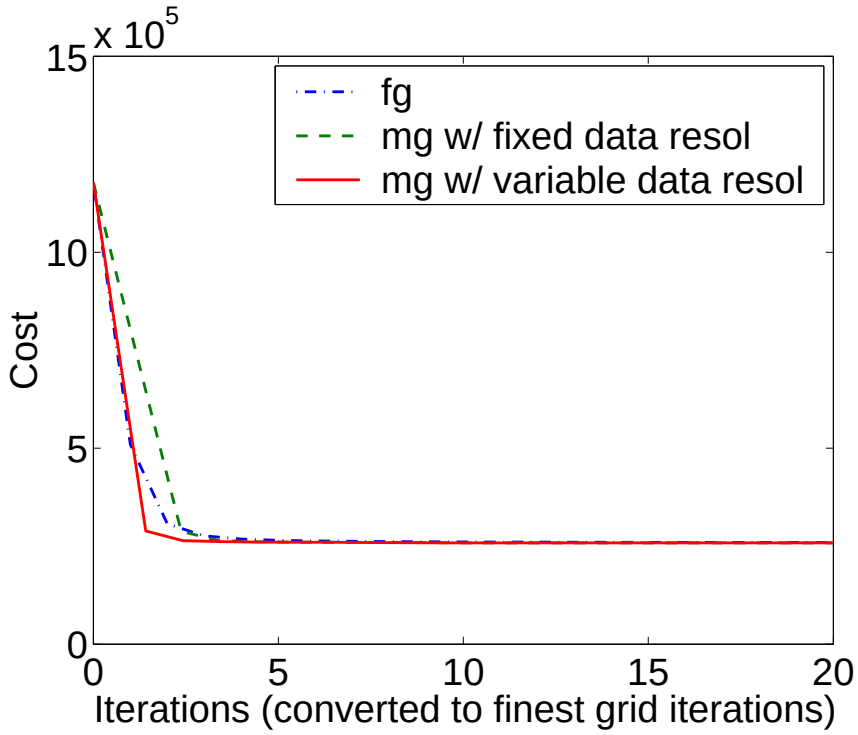
Fixed-grid (14 iter)



Fixed-grid (50 iter)

- Example of emission tomography with Poisson model

Convergence: Transmission Case with Poisson Model



Cost

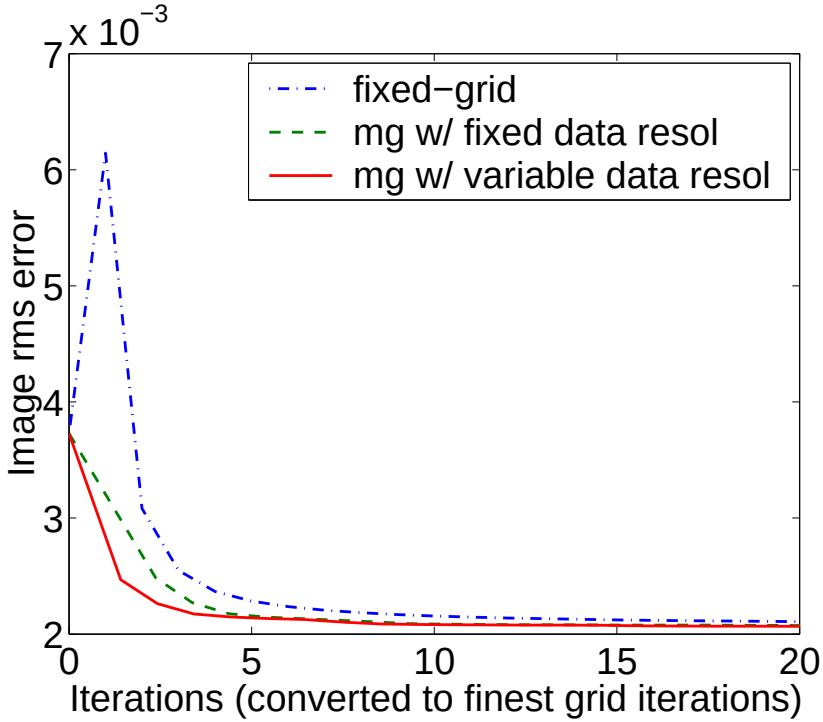


Image root-mean-square error

Quadratic Approximation of Poisson Log Likelihood

- Results in quadratic data term

$$c(x) = ||\gamma - Px||^2_\Lambda$$

  - Emission tomography

$$\begin{aligned} \gamma_m &= y_m \\ \Lambda &= \frac{1}{2}\,\text{diag}\left\{\frac{1}{y_1}, \frac{1}{y_2}, \ldots, \frac{1}{y_M}\right\} \end{aligned}$$

  - Transmission tomography

$$\begin{aligned} \gamma_m &= \log\frac{y_T}{y_m} \\ \Lambda &= \frac{1}{2}\,\text{diag}\{y_1, y_2, \ldots, y_M\} \end{aligned}$$

$\Rightarrow$ A special case of Gaussian noise with the forward model $f(x) = Px$

Coarse Scale Forward Model for Quadratic Approximation of Poisson Log Likelihood

- Quadratic approximation of Poisson log likelihood is equivalent to a special case of Gaussian noise with the forward model $f(x) = Px$

- Apply multigrid inversion developed for Gaussian noise with

$$f^{(q)}(x^{(q)}) = P^{(q)}x^{(q)}$$

- Coarse scale projection matrix is chosen by

$$P^{(q)} = J_{(0)}^{(q)} P^{(0)} I_{(q)}^{(0)}$$

Data Interpolation/Decimation Operators

- Cost function

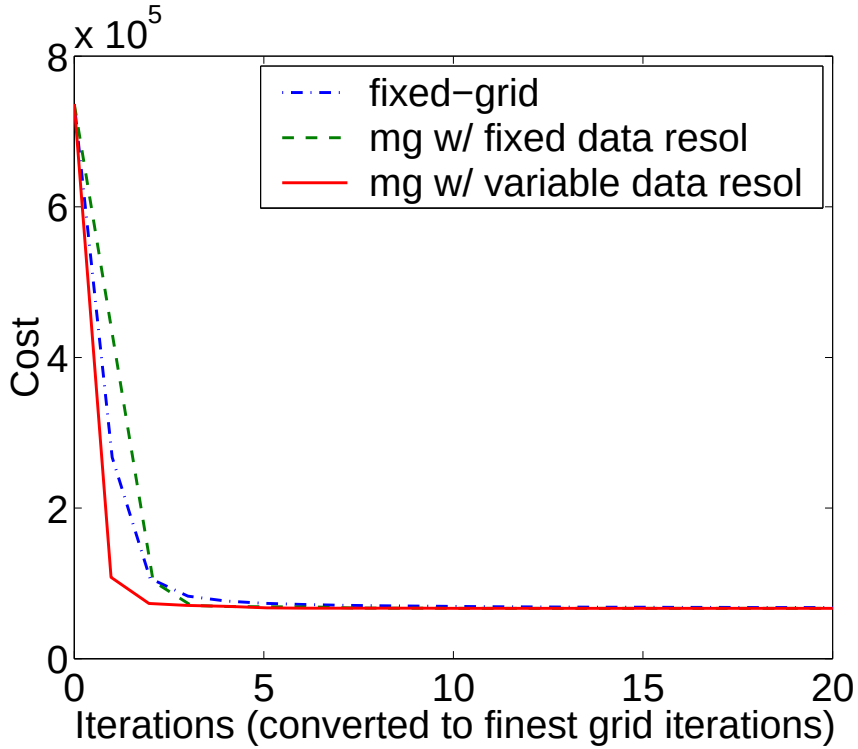
$$c(x) = ||\gamma - Px||^2_\Lambda$$

- Interpolation operator
  - No angle interpolation
  - Interpolation on displacement axis is pixel replication

$$\begin{aligned} f^{(q-1)}_{2m} &\leftarrow f^{(q)}_m \\ f^{(q-1)}_{2m+1} &\leftarrow f^{(q)}_m \end{aligned}$$

$\Rightarrow$ This choice preserves sparsity of weight matrix $\Lambda^{(q)}$ at coarse scales

Convergence: Emission Case with with Quadratic Approximation of Poisson Log Likelihood



Cost

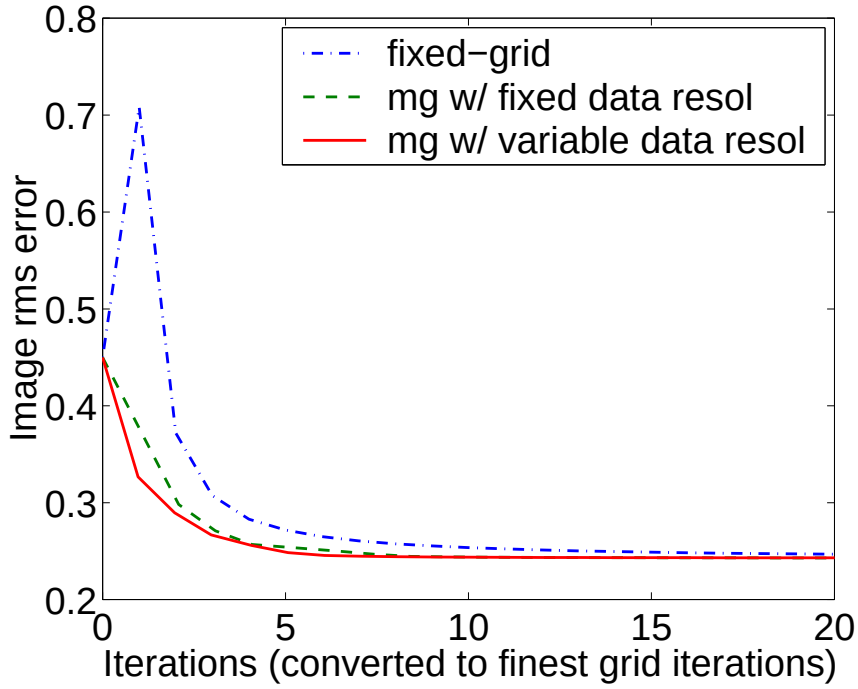
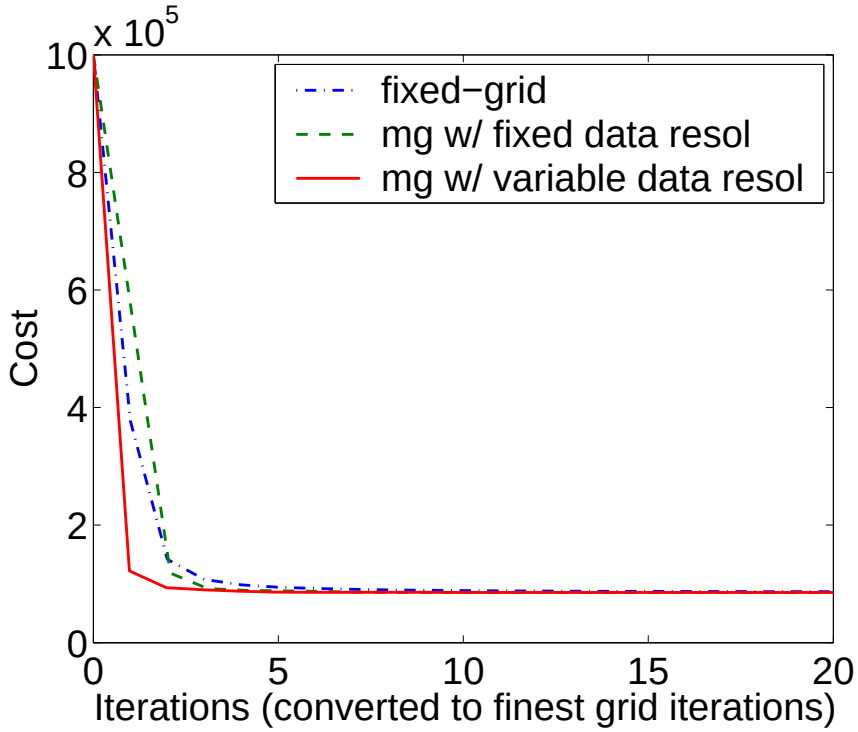


Image root-mean-square error

Convergence: Transmission Case with Quadratic Approximation of Poisson Log Likelihood



Cost

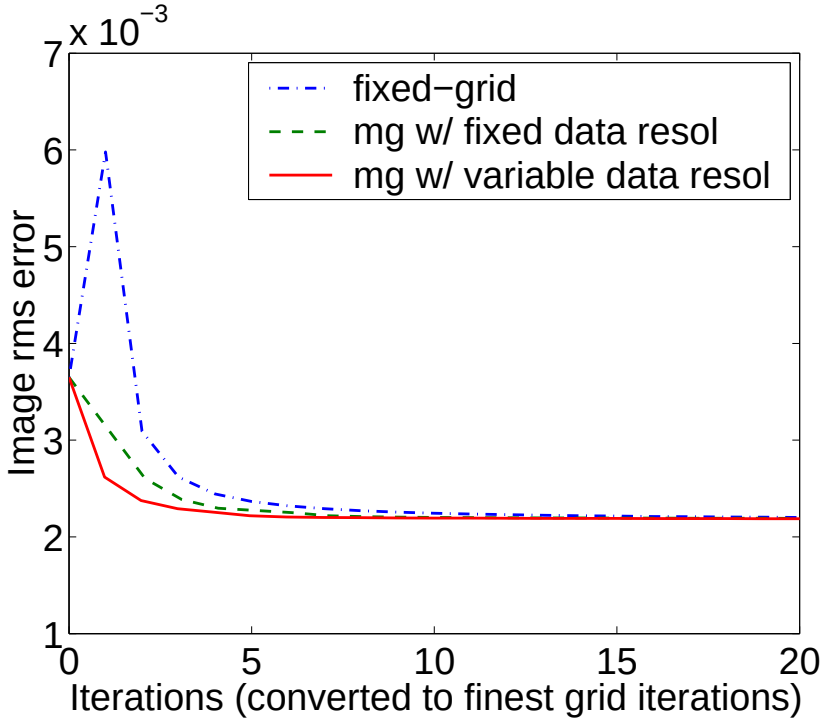


Image root-mean-square error

Concluding Remarks

- Extension of baseline multigrid inversion
  - Adaptively allocates computation to the scale at which the algorithm can best reduce the cost
  - Changes grid resolution of both image space and measurement data space

  $\Rightarrow$ Improves computational efficiency
- Results
  - Algorithms are formulated for Poisson model-based and its quadratic approximation-based reconstruction problems in transmission / emission tomography
  - Substantially reduced computation relative to ICD

Thank you!!!