

# Iterative Methods for Image Reconstruction

Jeffrey A. Fessler

EECS Department  
The University of Michigan

ISBI Tutorial

May 14, 2008

# Image Reconstruction Methods (Simplified View)

**Analytical**

(FBP)

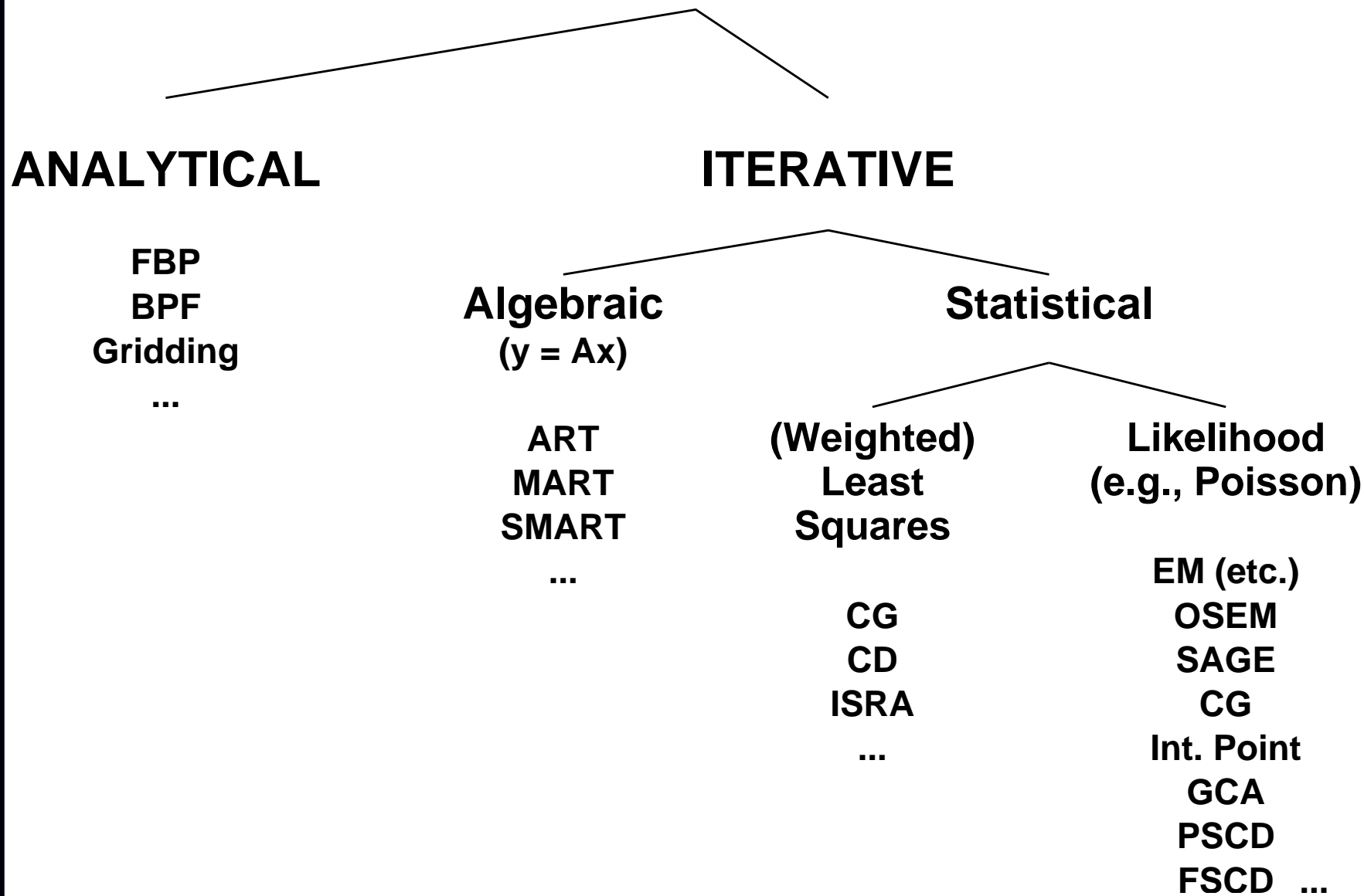
(MR: iFFT)

**Iterative**

(OSEM?)

(MR: CG?)

# Image Reconstruction Methods / Algorithms



# Outline :

## Part 0: Introduction / Overview / Examples

## Part 1: Problem Statements

- Continuous-discrete vs continuous-continuous vs discrete-discrete

## Part 2: Four of Five Choices for Statistical Image Reconstruction

- Object parameterization
- System physical modeling
- Statistical modeling of measurements
- Cost functions and regularization

## Part 3: Fifth Choice: Iterative algorithms

- Classical optimization methods
- Considerations: nonnegativity, convergence rate, ...
- Optimization transfer: EM etc.
- Ordered subsets / block iterative / incremental gradient methods

## Part 4: Performance Analysis

- Spatial resolution properties
- Noise properties
- Detection performance

# History

- Successive substitution method vs direct Fourier (Bracewell, 1956)
- Iterative method for emission tomography (Kuhl, 1963)
- Iterative method for X-ray CT (Hounsfield, 1968)
- ART for tomography (Gordon, Bender, Herman, JTB, 1970)
- **Weighted least squares** for 3D SPECT (Goitein, NIM, 1972)
- Richardson/Lucy iteration for image restoration (1972, 1974)
- Proposals to use Poisson likelihood for emission and transmission tomography (Rockmore and Macovski, TNS, 1976, 1977)
- Expectation-maximization (EM) algorithms for Poisson model  
Emission: (Shepp and Vardi, TMI, 1982)  
Transmission: (Lange and Carson, JCAT, 1984)
- Regularized (aka Bayesian) Poisson emission reconstruction (Geman and McClure, ASA, 1985)
- Ordered-subsets EM algorithm (Hudson and Larkin, TMI, 1994)
- Commercial introduction of OSEM for PET scanners circa 1997

# Why Statistical Methods?

- Object constraints (e.g., nonnegativity, object support)
- Accurate physical models (less bias  $\implies$  improved quantitative accuracy) (e.g., nonuniform attenuation in SPECT)  
improved spatial resolution?
- Appropriate statistical models (less variance  $\implies$  lower image noise) (FBP treats all rays equally)
- Side information (e.g., MRI or CT boundaries)
- Nonstandard geometries (e.g., irregular sampling or “missing” data)

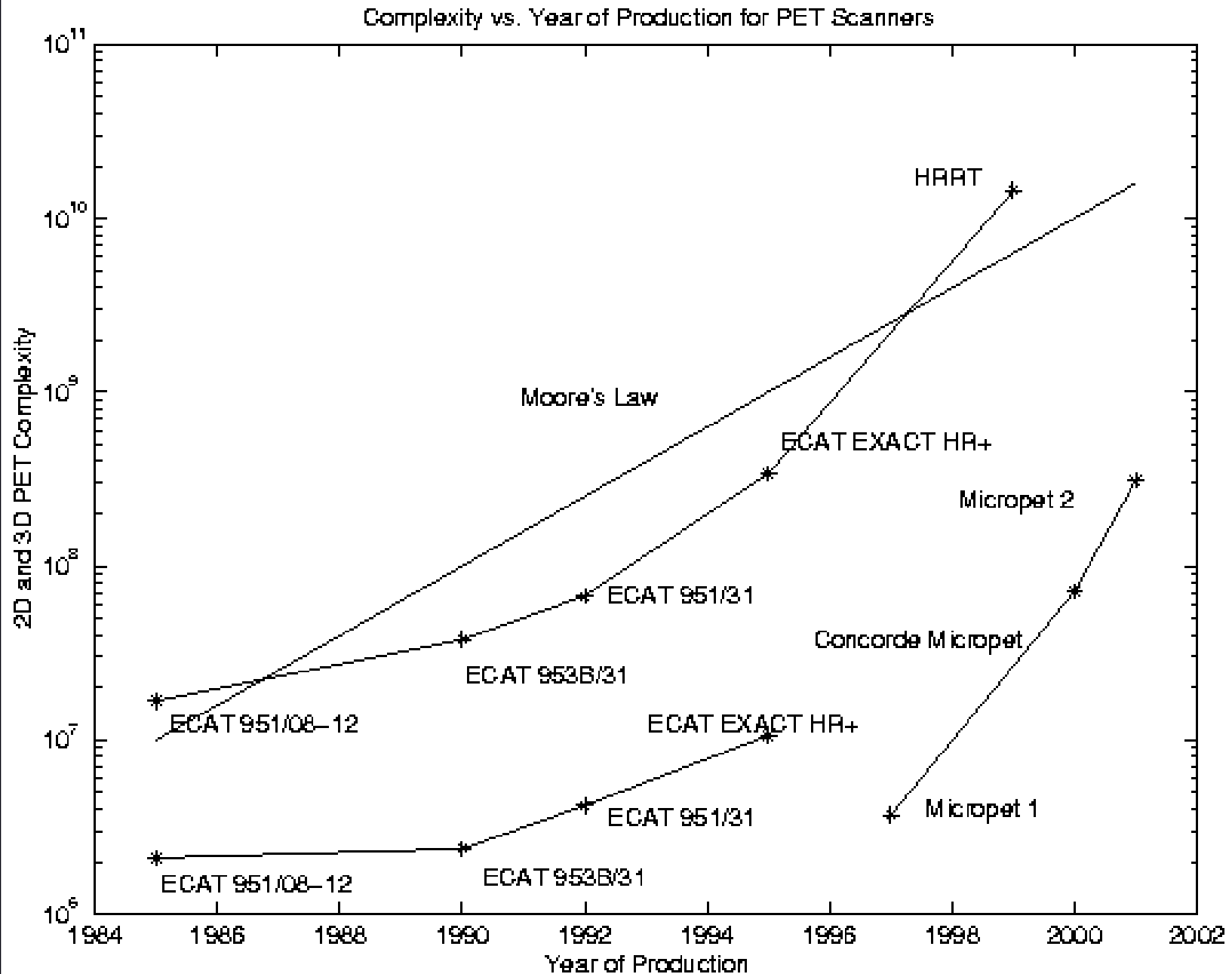
## Disadvantages?

- Computation time
- Model complexity
- Software complexity

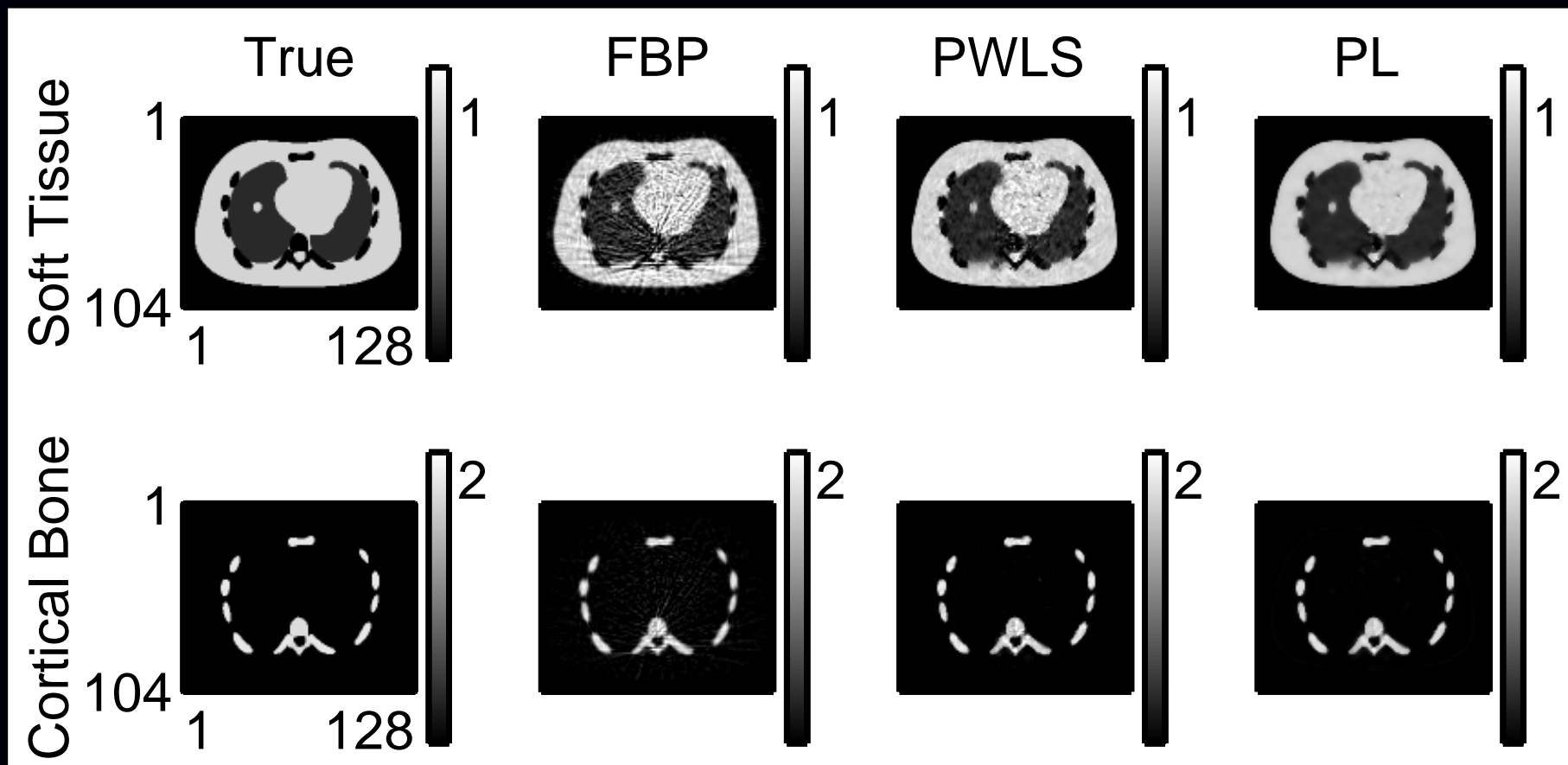
## Analytical methods (a different short course!)

- Idealized mathematical model
  - Usually geometry only, greatly over-simplified physics
  - Continuum measurements (discretize/sample *after* solving)
- No statistical model
- Easier analysis of properties (due to linearity)  
e.g., Huesman (1984) FBP ROI variance for kinetic fitting

# What about Moore's Law?



## Benefit Example: Statistical Models

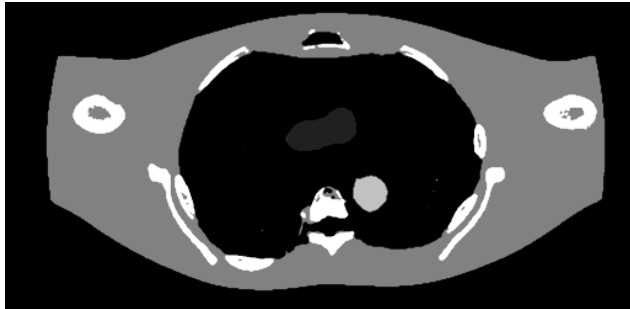


Method	NRMS Error	
	Soft Tissue	Cortical Bone
FBP	22.7%	29.6%
PWLS	13.6%	16.2%
PL	11.8%	15.8%

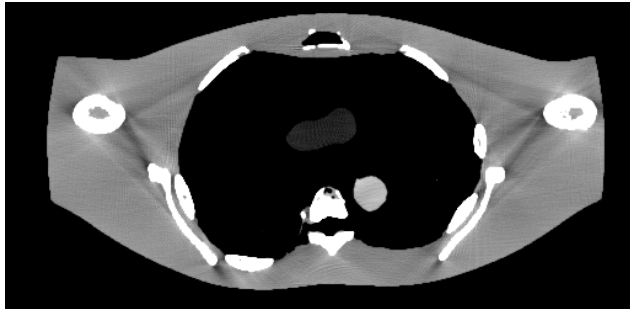


# Benefit Example: Physical Models

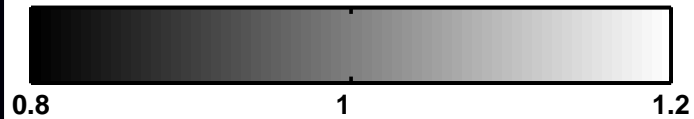
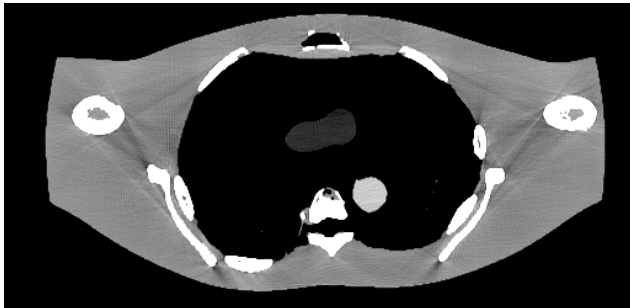
a. True object



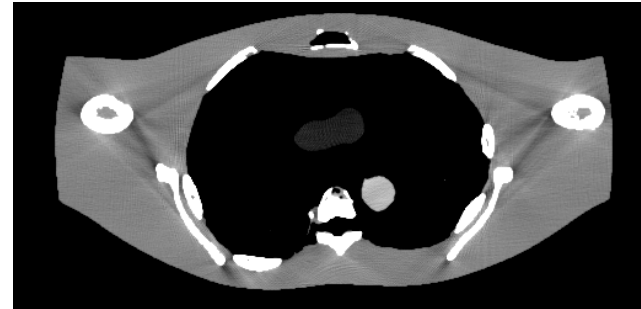
b. Uncorrected FBP



c. Monoenergetic statistical reconstruction



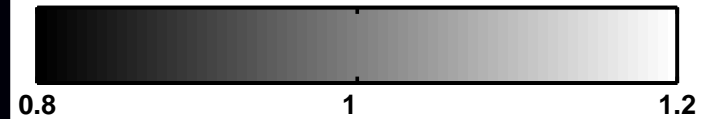
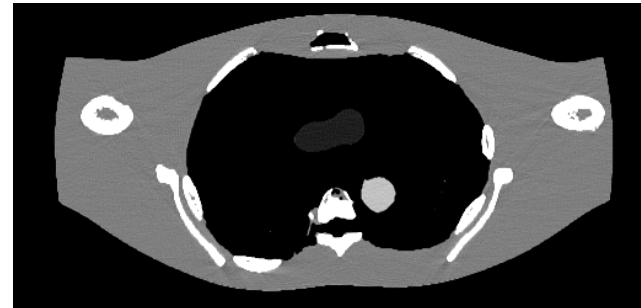
a. Soft-tissue corrected FBP



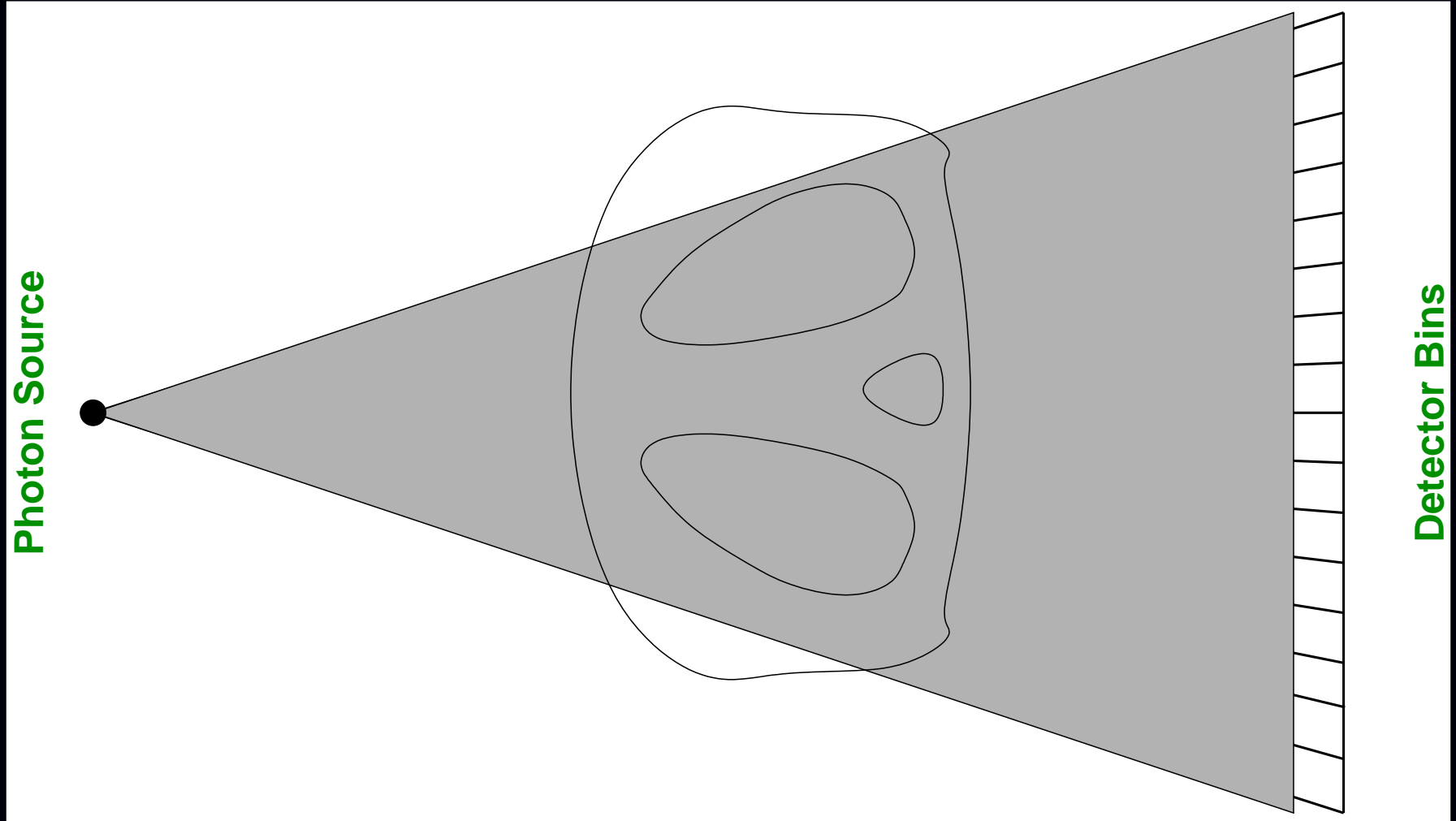
b. JS corrected FBP



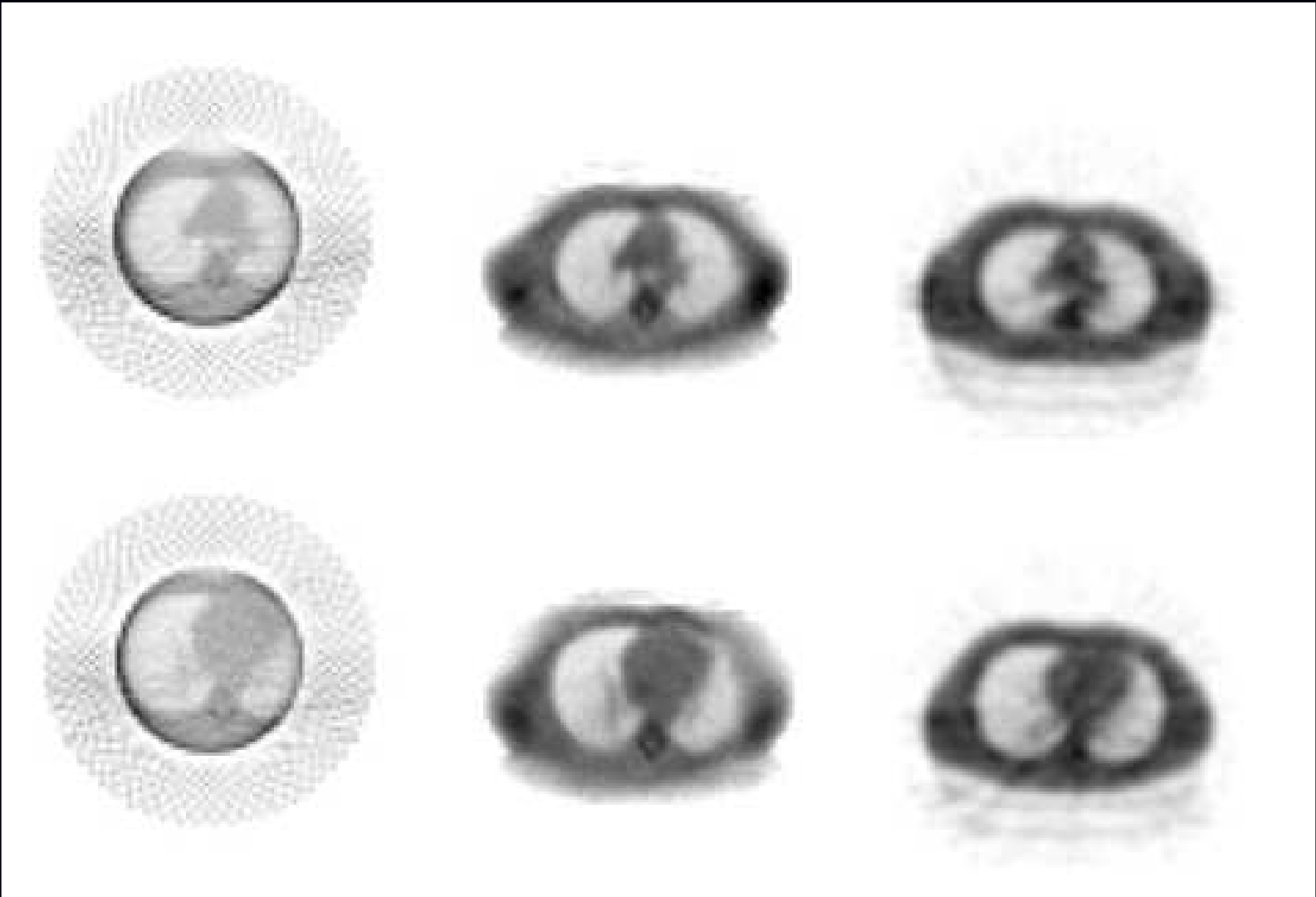
c. Polyenergetic Statistical Reconstruction



# Benefit Example: Nonstandard Geometries



# Truncated Fan-Beam SPECT Transmission Scan



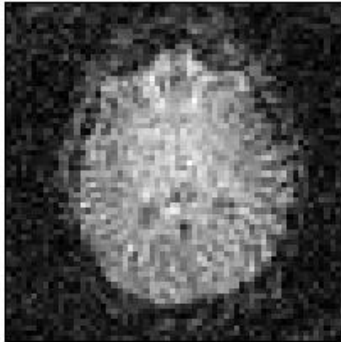
Truncated  
FBP

Truncated  
PWLS

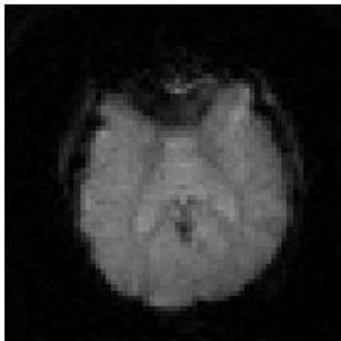
Untruncated  
FBP

# One Final Advertisement: Iterative MR Reconstruction

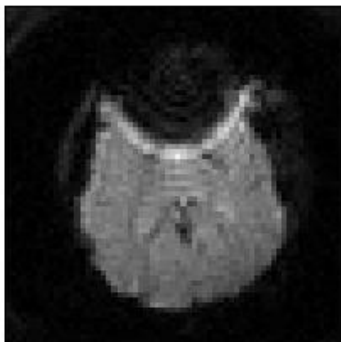
Spin Echo



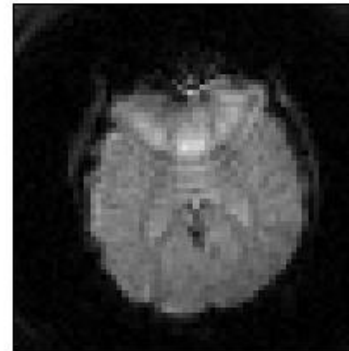
Iterative NUFFT  
with min-max



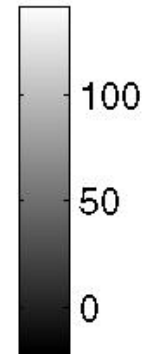
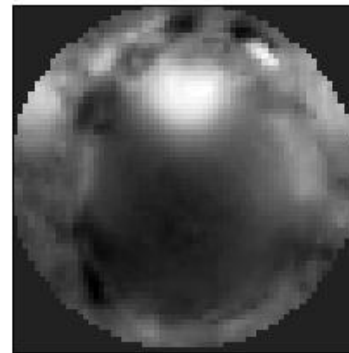
Uncorrected



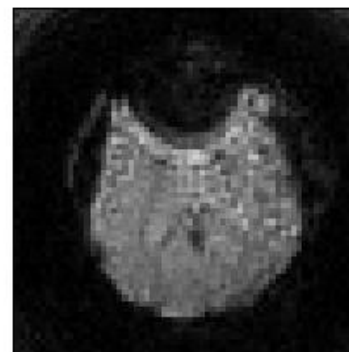
Conjugate Phase



Field Map in Hz



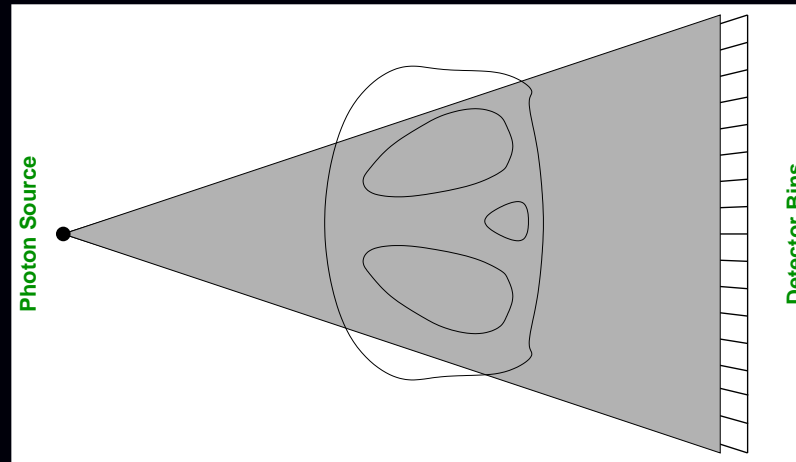
SPHERE



# Part 1: Problem Statement(s)

Example:

in monoenergetic transmission tomography with photon counting detectors, the goal is to reconstruct the *attenuation map*  $\mu(\vec{x})$  from transmission measurements  $\{y_i\}_{i=1}^{n_d}$ , given the system response  $s_i(\vec{x})$ ,  $i = 1, \dots, n_d$ , for each detector element.



Statistical model:  $y_i \sim \text{Poisson}\{b_i \exp(-\int \mu(\vec{x}) s_i(\vec{x}) d\vec{x}) + r_i\}$

- $b_i$ : blank/air scan
- $s_i(\vec{x})$ : line impulse associated with line integral for  $i$ th ray, possibly including detector blur and finite source size (approximation)
- $r_i$ : background due to Compton scatter

# Continuous-Discrete Models

Emission tomography:  $y_i \sim \text{Poisson}\left\{\int \lambda(\vec{x}) s_i(\vec{x}) d\vec{x} + r_i\right\}$

Transmission tomography (monoenergetic):  $y_i \sim \text{Poisson}\left\{b_i \exp\left(-\int_{L_i} \mu(\vec{x}) d\ell\right) + r_i\right\}$

Transmission (polyenergetic):  $y_i \sim \text{Poisson}\left\{\int I_i(\mathcal{E}) \exp\left(-\int_{L_i} \mu(\vec{x}, \mathcal{E}) d\ell\right) d\mathcal{E} + r_i\right\}$

Magnetic resonance imaging:  $y_i = \int f(\vec{x}) s_i(\vec{x}) d\vec{x} + \varepsilon_i$

Discrete measurements  $\mathbf{y} = (y_1, \dots, y_{n_d})$

Continuous-space unknowns:  $\lambda(\vec{x}), \mu(\vec{x}), f(\vec{x})$

Goal: estimate  $f(\vec{x})$  given  $\mathbf{y}$

## Solution options:

- Continuous-continuous formulations (“analytical,” *cf.* FBP for tomography)
- Continuous-discrete formulations  
Usually  $\hat{f}(\vec{x}) = \sum_{i=1}^{n_d} c_i s_i(\vec{x})$
- Discrete-discrete formulations  $f(\vec{x}) \approx \sum_{j=1}^{n_p} x_j b_j(\vec{x})$

# Textbook MRI Measurement Model

Ignoring *lots* of things, the standard measurement model is:

$$y_i = s(t_i) + \text{noise}_i, \quad i = 1, \dots, n_d$$

$$s(t) = \int f(\vec{x}) e^{-i2\pi\vec{k}(t)\cdot\vec{x}} d\vec{x} = F(\vec{k}(t)).$$

$\vec{x}$ : spatial coordinates

$\vec{k}(t)$ : k-space trajectory of the MR pulse sequence

$f(\vec{x})$ : object's unknown **transverse magnetization**

$F(\vec{k})$ : Fourier transform of  $f(\vec{x})$ . We get noisy samples of this!

$e^{-i2\pi\vec{k}(t)\cdot\vec{x}}$  provides spatial information  $\implies$  Nobel Prize

Goal of image reconstruction: find  $f(\vec{x})$  from measurements  $\{y_i\}_{i=1}^{n_d}$ .

The unknown object  $f(\vec{x})$  is a continuous-space function, but the recorded measurements  $\mathbf{y} = (y_1, \dots, y_{n_d})$  are finite.

Under-determined (ill posed) problem  $\implies$  no canonical solution.

*All MR scans provide only “partial” k-space data.*

# Image Reconstruction Strategies

- Continuous-continuous formulation

Pretend that a continuum of measurements are available:

$$F(\vec{\mathbf{k}}) = \int f(\vec{\mathbf{x}}) e^{-i2\pi\vec{\mathbf{k}}\cdot\vec{\mathbf{x}}} d\vec{\mathbf{x}}.$$

The “solution” is an inverse Fourier transform:

$$f(\vec{\mathbf{x}}) = \int F(\vec{\mathbf{k}}) e^{i2\pi\vec{\mathbf{k}}\cdot\vec{\mathbf{x}}} d\vec{\mathbf{k}}.$$

Now discretize the integral solution:

$$\hat{f}(\vec{\mathbf{x}}) = \sum_{i=1}^{n_d} F(\vec{\mathbf{k}}_i) e^{i2\pi\vec{\mathbf{k}}_i\cdot\vec{\mathbf{x}}} w_i \approx \sum_{i=1}^{n_d} y_i w_i e^{i2\pi\vec{\mathbf{k}}_i\cdot\vec{\mathbf{x}}},$$

where  $w_i$  values are “sampling density compensation factors.” Numerous methods for choosing  $w_i$  values in the literature.

For Cartesian sampling, using  $w_i = 1/N$  suffices, and the summation is an inverse FFT.

For non-Cartesian sampling, replace summation with **gridding**.



- Continuous-discrete formulation

Use many-to-one linear model:

$$\mathbf{y} = \mathcal{A} f + \boldsymbol{\varepsilon}, \text{ where } \mathcal{A} : \mathcal{L}_2(\mathbb{R}^{\bar{d}}) \rightarrow \mathbb{C}^{n_d}.$$

Minimum norm solution (cf. “natural pixels”):

$$\min_{\hat{f}} \|\hat{f}\|_2 \text{ subject to } \mathbf{y} = \mathcal{A} \hat{f}$$

$$\hat{f} = \mathcal{A}^* (\mathcal{A} \mathcal{A}^*)^{-1} \mathbf{y} = \sum_{i=1}^{n_d} c_i e^{-i2\pi \vec{k}_i \cdot \vec{x}}, \text{ where } \mathcal{A} \mathcal{A}^* \mathbf{c} = \mathbf{y}.$$

- Discrete-discrete formulation

Assume parametric model for object:

$$f(\vec{x}) = \sum_{j=1}^{n_p} x_j b_j(\vec{x}).$$

Estimate parameter vector  $\mathbf{x} = (x_1, \dots, x_{n_p})$  from data vector  $\mathbf{y}$ .

## Part 2: Five Categories of Choices

- Object parameterization: function  $f(\vec{r})$  vs finite coefficient vector  $x$
- System physical model:  $\{s_i(\vec{r})\}$
- Measurement statistical model  $y_i \sim \boxed{?}$
- Cost function: data-mismatch and regularization
- Algorithm / initialization

No perfect choices - one can critique all approaches!

# Choice 1. Object Parameterization

Finite measurements:  $\{y_i\}_{i=1}^{n_d}$ .

Continuous object:  $f(\vec{r})$ .

Hopeless?

“All models are wrong but some models are useful.”

Linear *series expansion* approach. Replace  $f(\vec{r})$  by  $\mathbf{x} = (x_1, \dots, x_{n_p})$  where

$$f(\vec{r}) \approx \tilde{f}(\vec{r}) = \sum_{j=1}^{n_p} x_j b_j(\vec{r}) \leftarrow \text{“basis functions”}$$

Forward projection:

$$\begin{aligned} \int s_i(\vec{r}) f(\vec{r}) d\vec{r} &= \int s_i(\vec{r}) \left[ \sum_{j=1}^{n_p} x_j b_j(\vec{r}) \right] d\vec{r} = \sum_{j=1}^{n_p} \left[ \int s_i(\vec{r}) b_j(\vec{r}) d\vec{r} \right] x_j \\ &= \sum_{j=1}^{n_p} a_{ij} x_j = [\mathbf{Ax}]_i, \quad \text{where } a_{ij} \triangleq \int s_i(\vec{r}) b_j(\vec{r}) d\vec{r} \end{aligned}$$

- Projection integrals become finite summations.
- $a_{ij}$  is contribution of  $j$ th basis function (e.g., voxel) to  $i$ th measurement.
- The units of  $a_{ij}$  and  $x_j$  depend on the user-selected units of  $b_j(\vec{r})$ .
- The  $n_d \times n_p$  matrix  $\mathbf{A} = \{a_{ij}\}$  is called the *system matrix*.

## (Linear) Basis Function Choices

- Fourier series (complex / not sparse)
- Circular harmonics (complex / not sparse)
- Wavelets (negative values / not sparse)
- Kaiser-Bessel window functions (blobs)
- Overlapping circles (disks) or spheres (balls)
- Polar grids, logarithmic polar grids
- “Natural pixels”  $\{s_i(\vec{r})\}$
- B-splines (pyramids)
- Rectangular **pixels** / **voxels** (rect functions)
- Point masses / bed-of-nails / lattice of points / “comb” function
- Organ-based voxels (e.g., from CT in PET/CT systems)
- ...

# Basis Function Considerations

## Mathematical

- Represent  $f(\vec{r})$  “well” with moderate  $n_p$  (approximation accuracy)
- e.g., represent a constant (uniform) function
- Orthogonality? (not essential)
- Linear independence (ensures uniqueness of expansion)
- Insensitivity to shift of basis-function grid (approximate shift invariance)
- Rotation invariance

## Computational

- “Easy” to compute  $a_{ij}$  values and/or  $\mathbf{Ax}$
- If stored, the system matrix  $\mathbf{A}$  should be sparse (mostly zeros).
- Easy to represent nonnegative functions e.g., if  $x_j \geq 0$ , then  $f(\vec{r}) \geq 0$ .  
A sufficient condition is  $b_j(\vec{r}) \geq 0$ .

# Nonlinear Object Parameterizations

Estimation of intensity *and* shape (e.g., location, radius, etc.)

Surface-based (homogeneous) models

- Circles / spheres
- Ellipses / ellipsoids
- Superquadrics
- Polygons
- Bi-quadratic triangular Bezier patches, ...

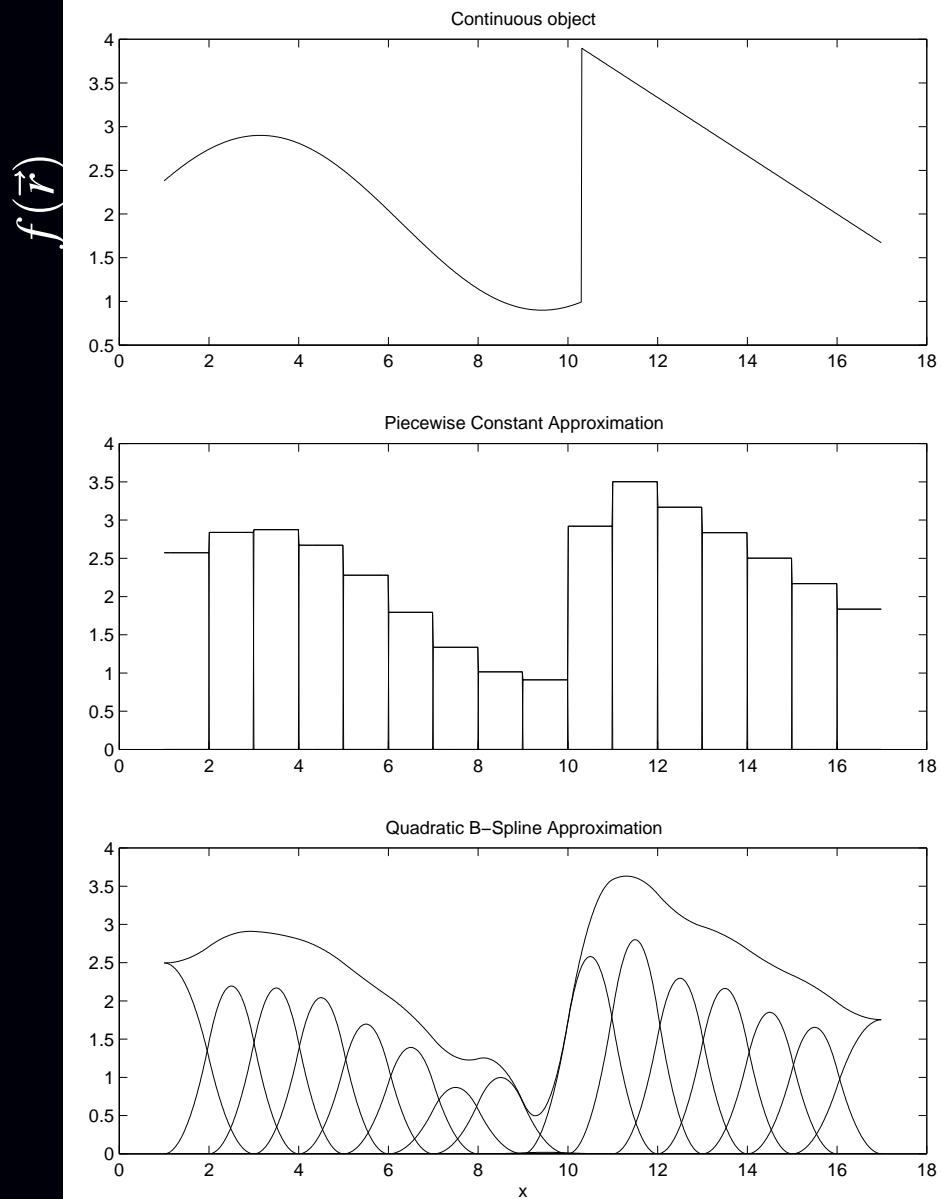
Other models

- Generalized series  $f(\vec{r}) = \sum_j x_j b_j(\vec{r}, \boldsymbol{\theta})$
- Deformable templates  $f(\vec{r}) = b(T_{\boldsymbol{\theta}}(\vec{r}))$
- ...

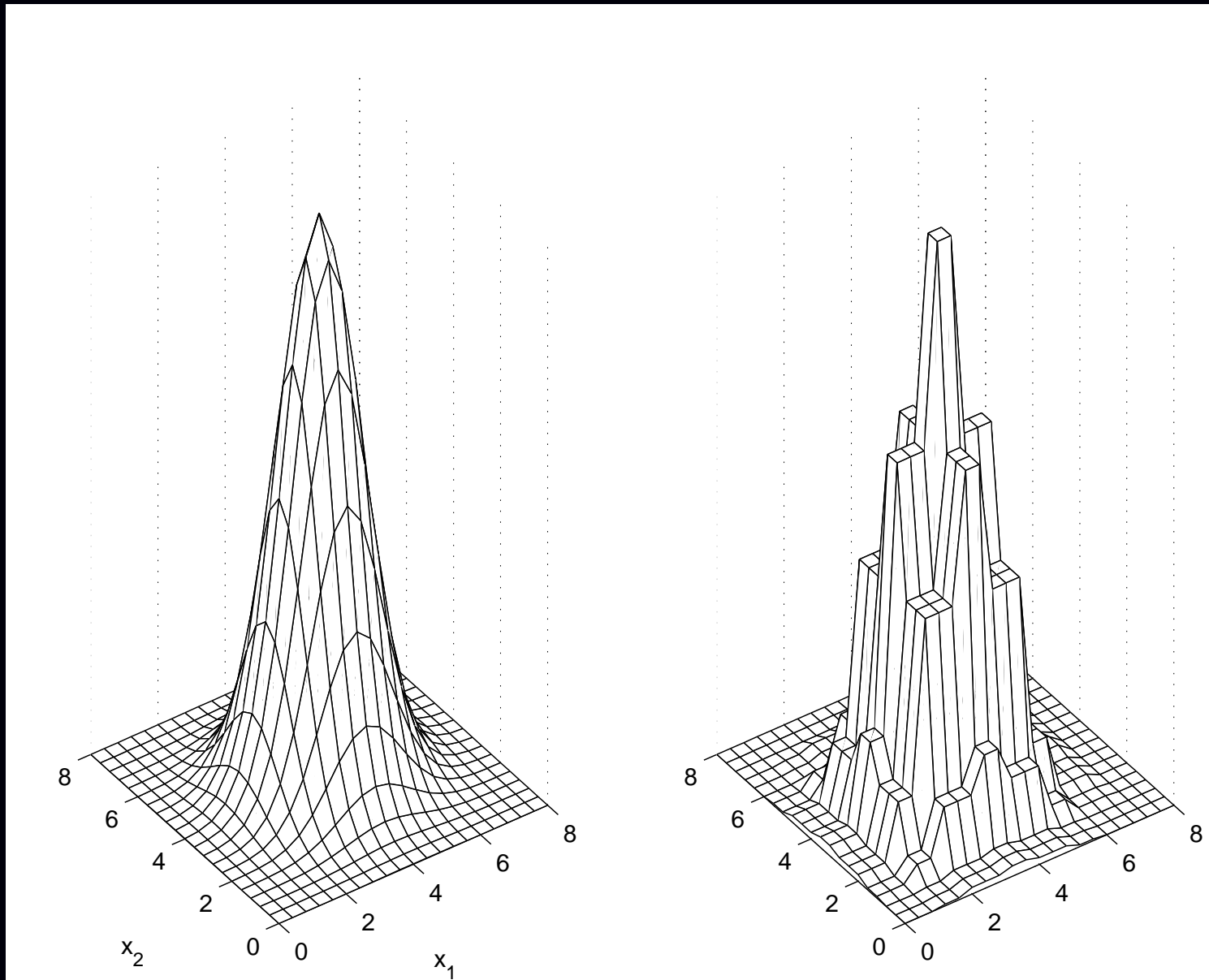
## Considerations

- Can be considerably more parsimonious
- If correct, yield greatly reduced estimation error
- Particularly compelling in limited-data problems
- Often oversimplified (all models are wrong but...)
- Nonlinear dependence on location induces non-convex cost functions, complicating optimization

# Example Basis Functions - 1D



# Pixel Basis Functions - 2D



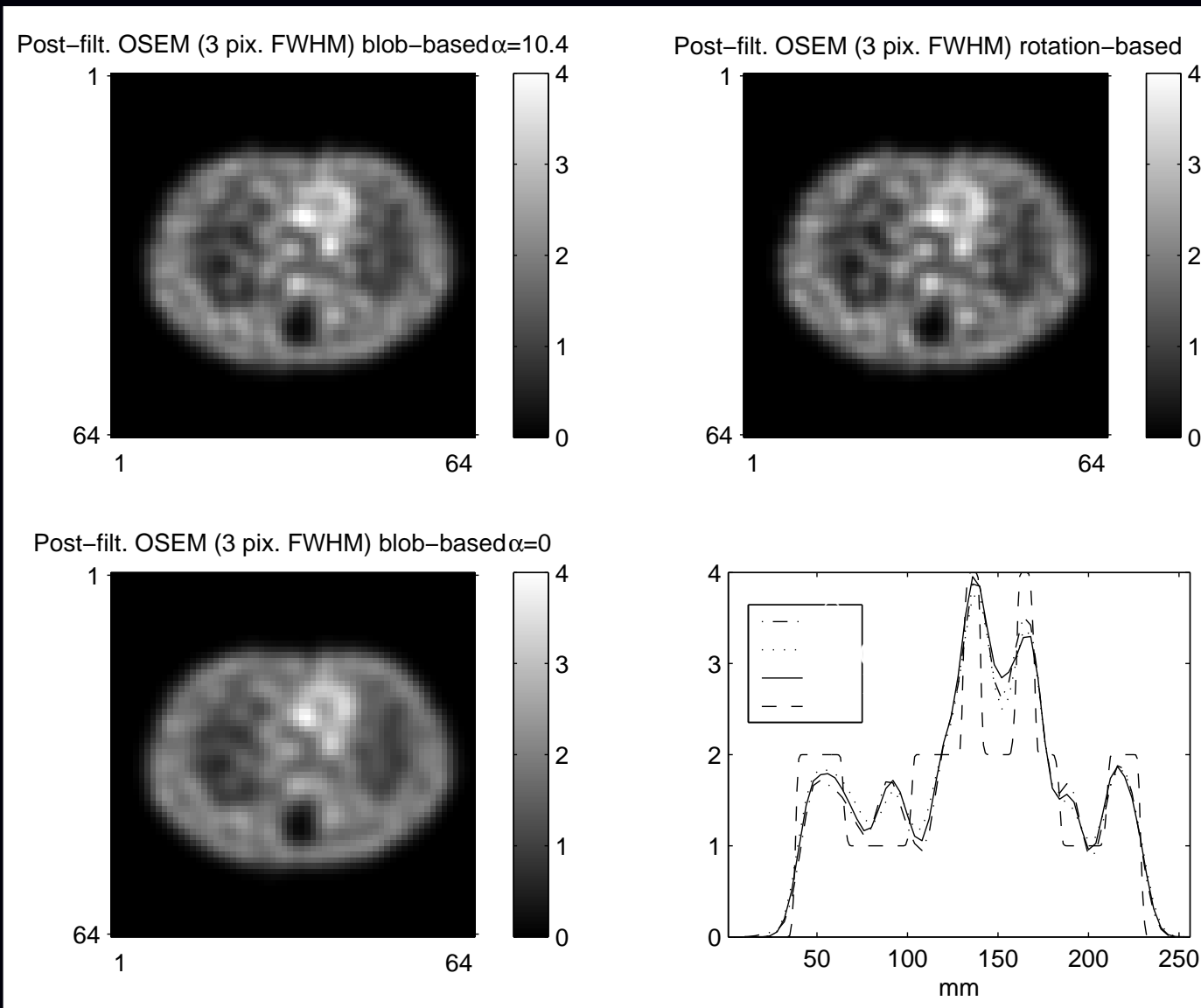
Continuous image  $f(\vec{r})$

Pixel basis approximation

$$\sum_{j=1}^{n_p} x_j b_j(\vec{r})$$

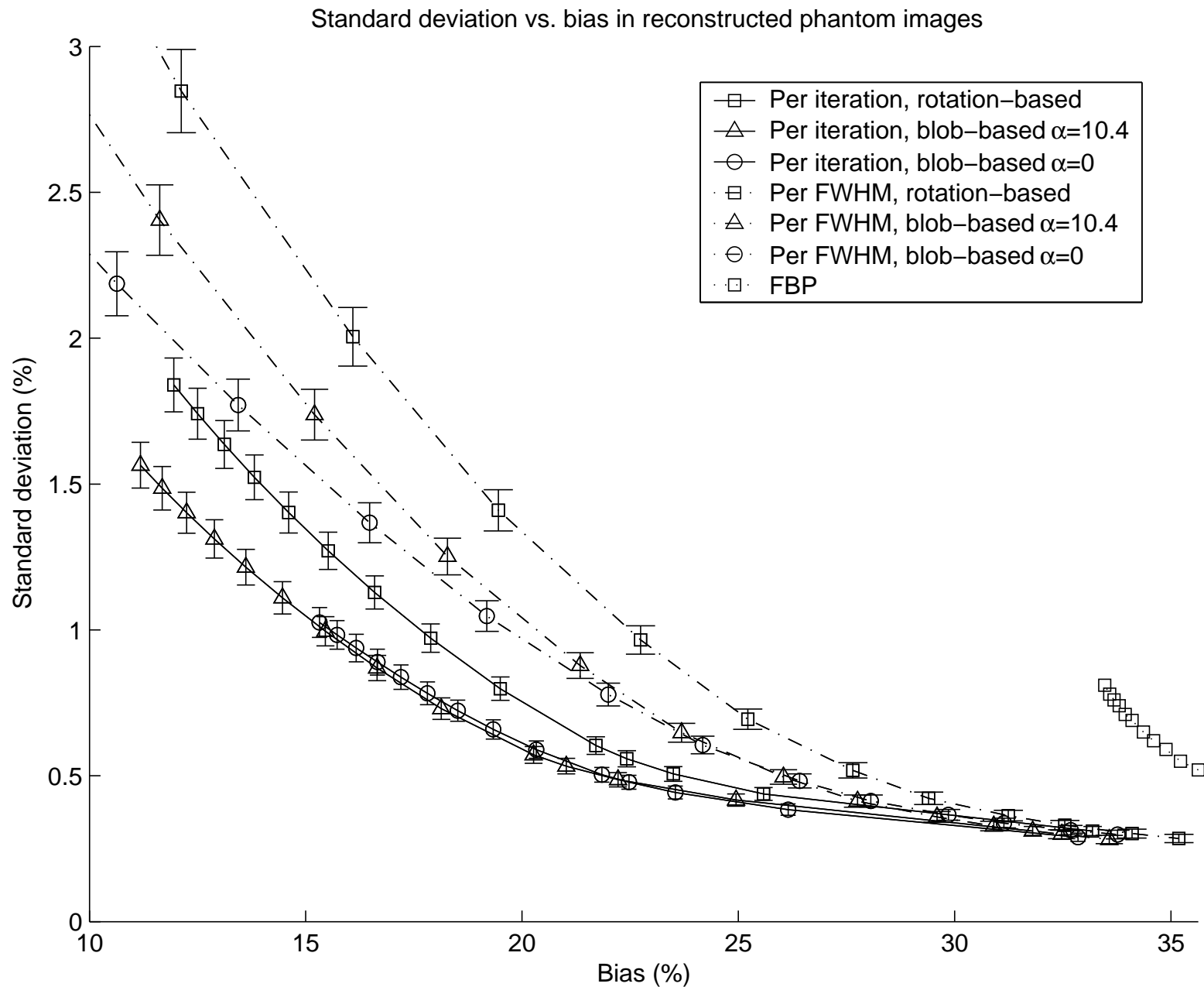


# Blobs in SPECT: Qualitative



(2D SPECT thorax phantom simulations)

# Blobs in SPECT: Quantitative



# Discrete-Discrete Emission Reconstruction Problem

Having chosen a basis and *linearly* parameterized the emission density...

Estimate the emission density coefficient vector  $\mathbf{x} = (x_1, \dots, x_{n_p})$  (aka “image”) using (something like) this statistical model:

$$y_i \sim \text{Poisson} \left\{ \sum_{j=1}^{n_p} a_{ij} x_j + r_i \right\}, \quad i = 1, \dots, n_d.$$

- $\{y_i\}_{i=1}^{n_d}$  : observed counts from each detector unit
- $\mathbf{A} = \{a_{ij}\}$  : system matrix (determined by system models)
- $r_i$  values : background contributions (determined separately)

---

Many image reconstruction problems are “find  $\mathbf{x}$  given  $\mathbf{y}$ ” where

$$y_i = g_i([\mathbf{A}\mathbf{x}]_i) + \varepsilon_i, \quad i = 1, \dots, n_d.$$

## Choice 2. System Model, aka Physics

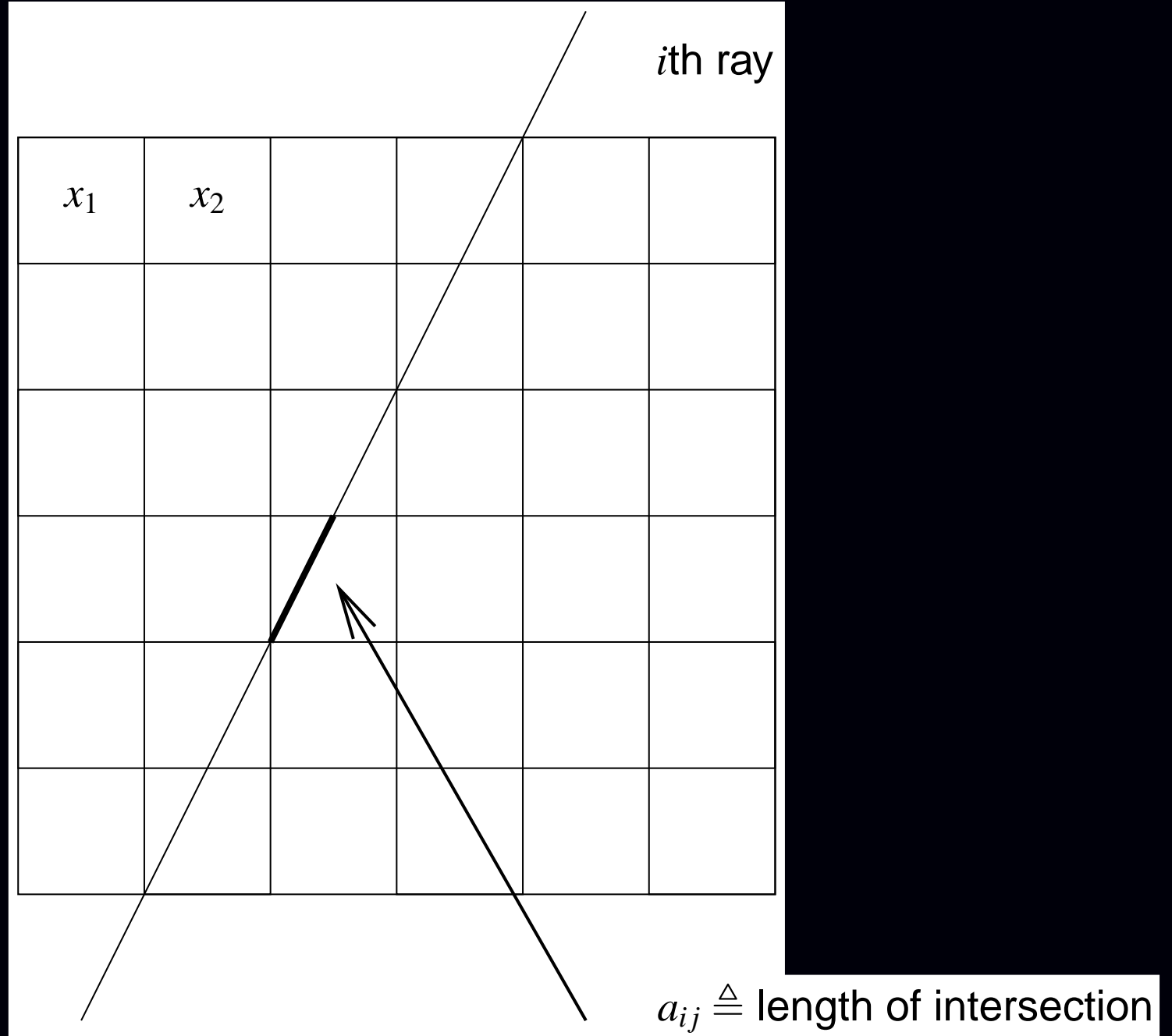
$$\text{System matrix elements: } a_{ij} = \int s_i(\vec{r}) b_j(\vec{r}) d\vec{r}$$

- scan geometry
- collimator/detector response
- attenuation
- scatter (object, collimator, scintillator)
- duty cycle (dwell time at each angle)
- detector efficiency / dead-time losses
- positron range, noncollinearity, crystal penetration, ...
- ...

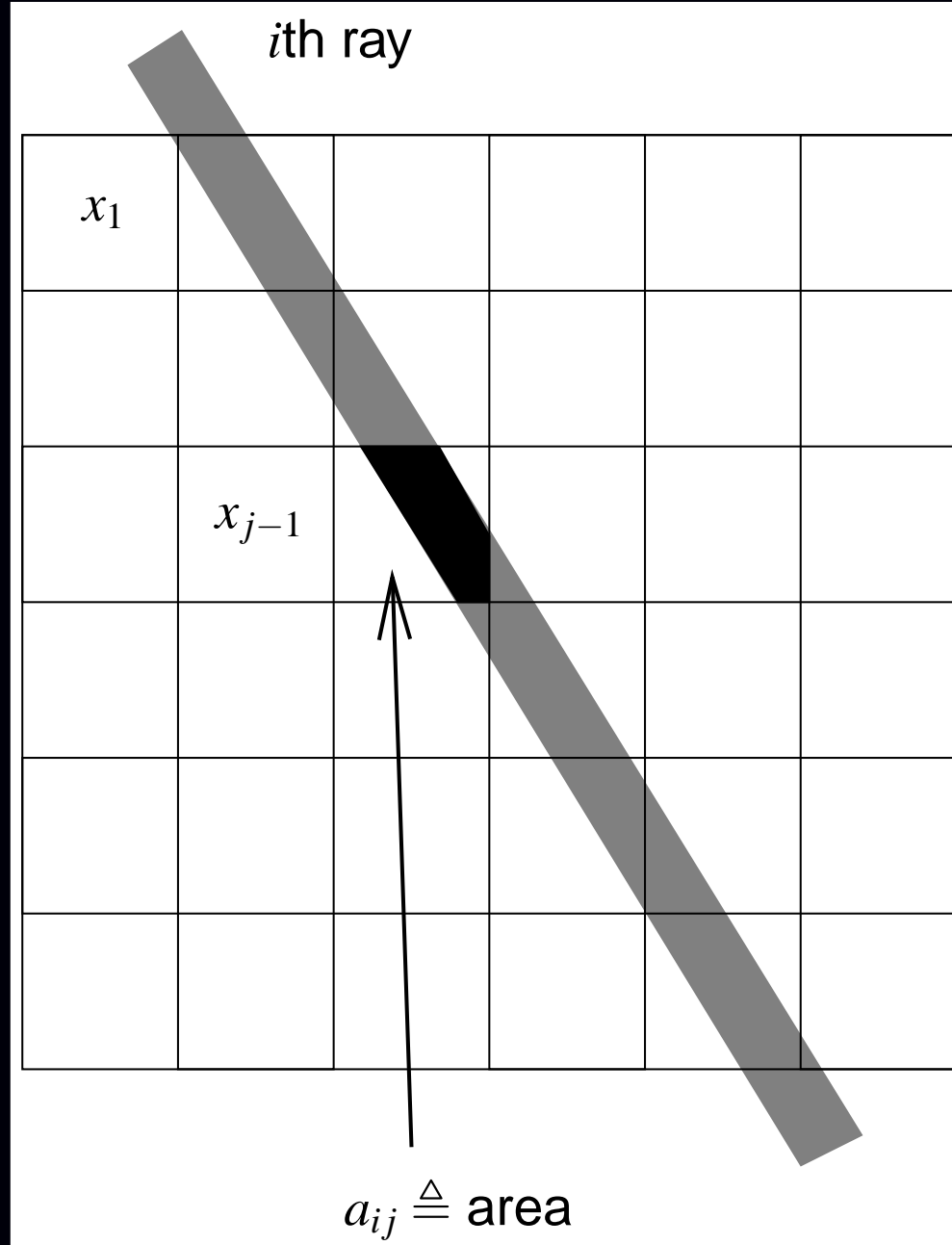
### Considerations

- Improving system model can improve
  - Quantitative accuracy
  - Spatial resolution
  - Contrast, SNR, detectability
- Computation time (and storage vs compute-on-fly)
- Model uncertainties  
(e.g., calculated scatter probabilities based on noisy attenuation map)
- Artifacts due to over-simplifications

# “Line Length” System Model for Tomography

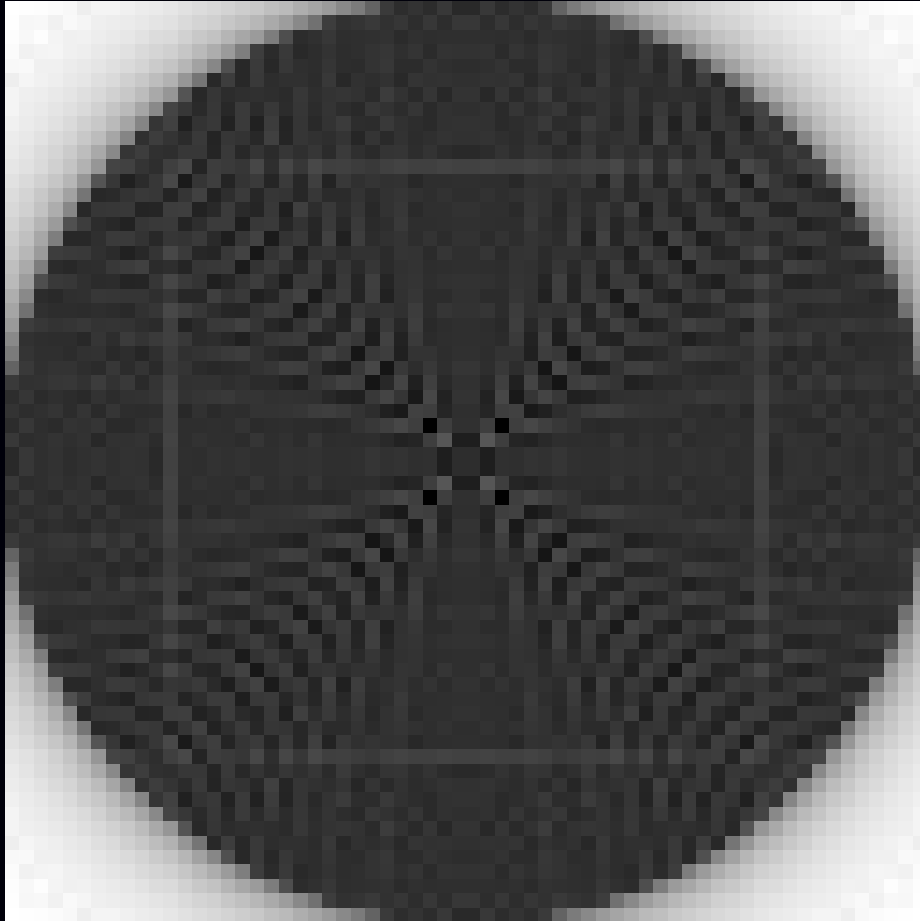


# “Strip Area” System Model for Tomography

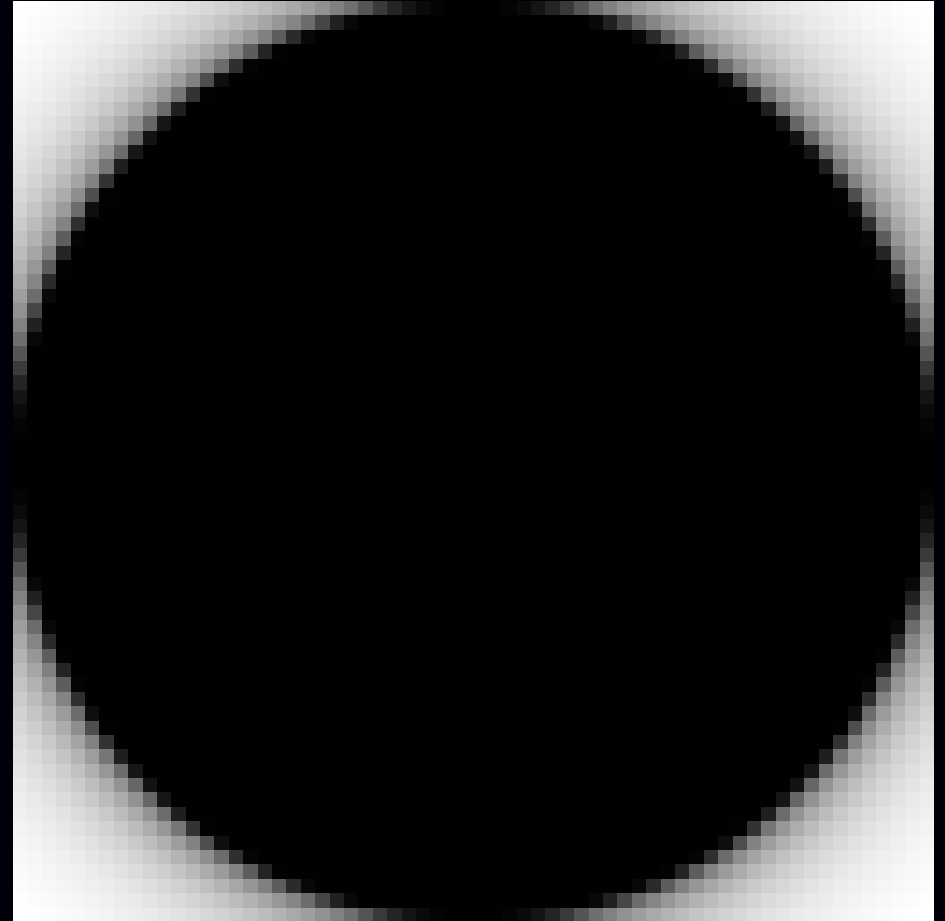


# (Implicit) System Sensitivity Patterns

$$\sum_{i=1}^{n_d} a_{ij} \approx s(\vec{r}_j) = \sum_{i=1}^{n_d} s_i(\vec{r}_j)$$



Line Length



Strip Area

## Forward- / Back-projector “Pairs”

Forward projection (image domain to projection domain):

$$\bar{y}_i = \int s_i(\vec{r}) f(\vec{r}) d\vec{r} = \sum_{j=1}^{n_p} a_{ij} x_j = [\mathbf{Ax}]_i, \quad \text{or} \quad \bar{\mathbf{y}} = \mathbf{Ax}$$

Backprojection (projection domain to image domain):

$$\mathbf{A}'\mathbf{y} = \left\{ \sum_{i=1}^{n_d} a_{ij} y_i \right\}_{j=1}^{n_p}$$

The term “forward/backprojection pair” often corresponds to an implicit choice for the object basis and the system model.

Sometimes  $\mathbf{A}'\mathbf{y}$  is implemented as  $\mathbf{By}$  for some “backprojector”  $\mathbf{B} \neq \mathbf{A}'$

Least-squares solutions (for example):

$$\hat{\mathbf{x}} = [\mathbf{A}'\mathbf{A}]^{-1} \mathbf{A}'\mathbf{y} \neq [\mathbf{BA}]^{-1} \mathbf{By}$$

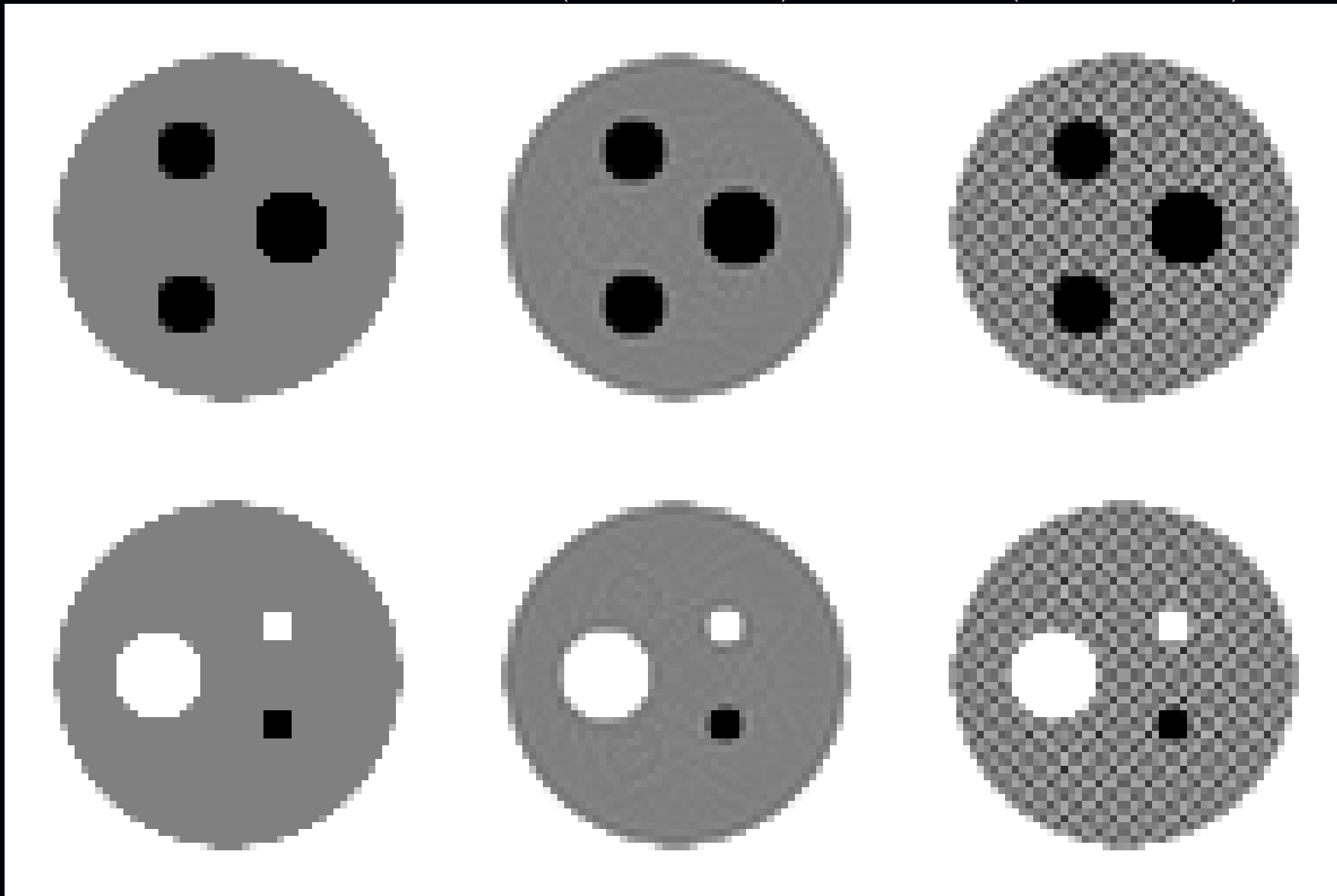


# Mismatched Backprojector $B \neq A'$

$x$

$\hat{x}(PWLS - CG)$

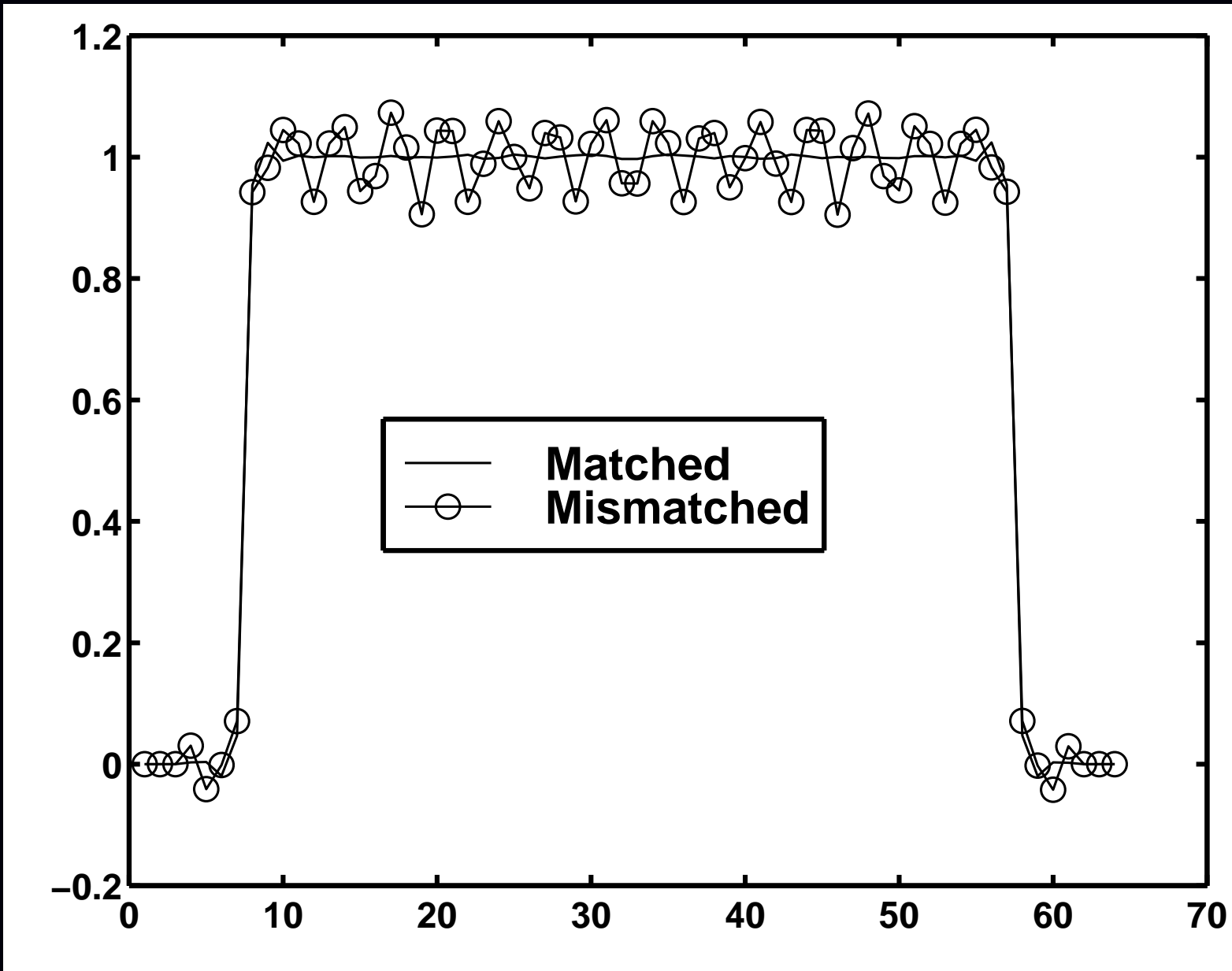
$\hat{x}(PWLS - CG)$



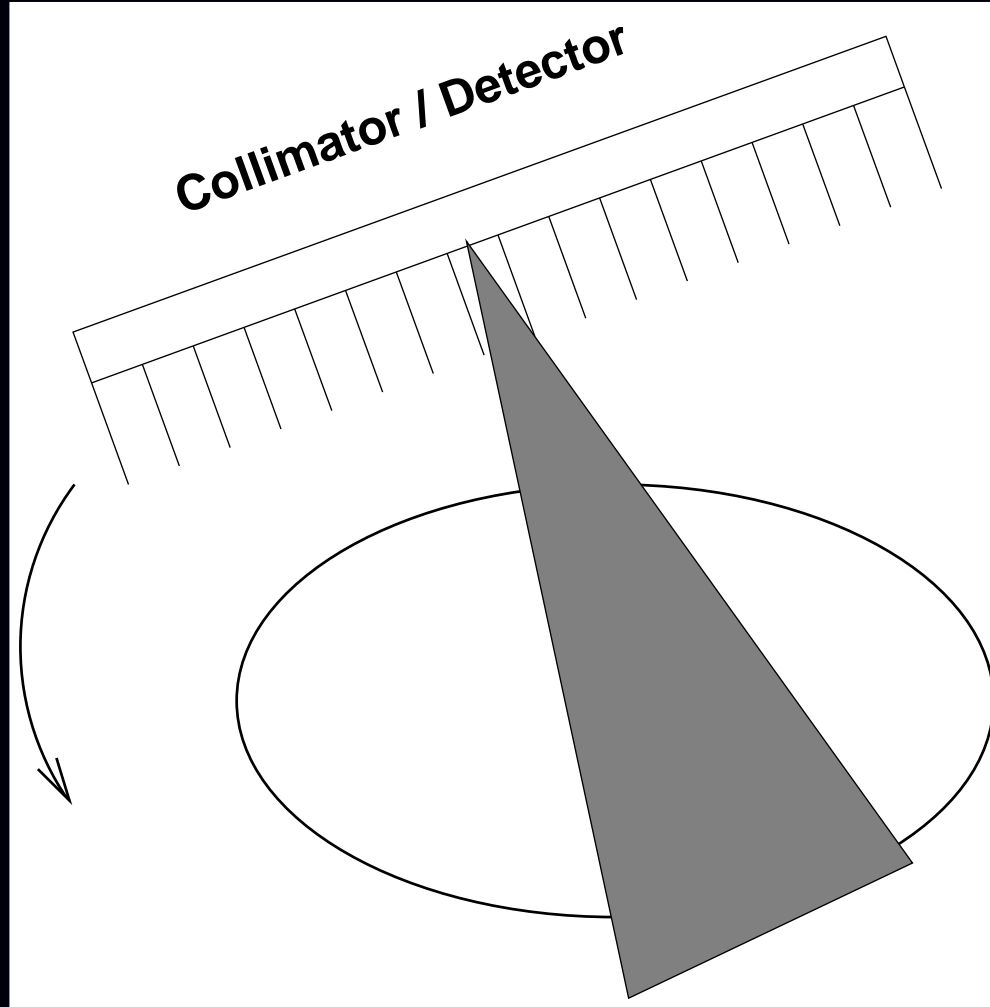
Matched

Mismatched

# Horizontal Profiles



# SPECT System Modeling



Complications: nonuniform attenuation, depth-dependent PSF, Compton scatter

## Choice 3. Statistical Models

After modeling the system physics, we have a deterministic “model:”

$$y_i \approx g_i([\mathbf{A}\mathbf{x}]_i)$$

for some functions  $g_i$ , e.g.,  $g_i(l) = l + r_i$  for emission tomography.

Statistical modeling is concerned with the “ $\approx$ ” aspect.

### Considerations

- More accurate models:
  - can lead to lower variance images,
  - may incur additional computation,
  - may involve additional algorithm complexity  
(e.g., proper transmission Poisson model has nonconcave log-likelihood)
- Statistical model errors (e.g., deadtime)
- Incorrect models (e.g., log-processed transmission data)

# Statistical Model Choices for Emission Tomography

- “None.” Assume  $\mathbf{y} - \mathbf{r} = \mathbf{Ax}$ . “Solve algebraically” to find  $\mathbf{x}$ .
- White Gaussian noise. Ordinary least squares: minimize  $\|\mathbf{y} - \mathbf{Ax}\|^2$   
(This is the appropriate statistical model for MR.)
- Non-white Gaussian noise. Weighted least squares: minimize

$$\|\mathbf{y} - \mathbf{Ax}\|_{\mathbf{w}}^2 = \sum_{i=1}^{n_d} w_i (y_i - [\mathbf{Ax}]_i)^2, \quad \text{where} \quad [\mathbf{Ax}]_i \triangleq \sum_{j=1}^{n_p} a_{ij} x_j$$

(e.g., for Fourier rebinned (FORE) PET data)

- Ordinary Poisson model (ignoring or precorrecting for background)

$$y_i \sim \text{Poisson}\{[\mathbf{Ax}]_i\}$$

- Poisson model

$$y_i \sim \text{Poisson}\{[\mathbf{Ax}]_i + r_i\}$$

- Shifted Poisson model (for randoms precorrected PET)

$$y_i = y_i^{\text{prompt}} - y_i^{\text{delay}} \sim \text{Poisson}\{[\mathbf{Ax}]_i + 2r_i\} - 2r_i$$

## Shifted-Poisson Model for X-ray CT

A model that includes both photon variability and electronic readout noise:

$$y_i \sim \text{Poisson}\{\bar{y}_i(\boldsymbol{\mu})\} + N(0, \sigma^2)$$

Shifted Poisson approximation (matches first two moments):

$$[y_i + \sigma^2]_+ \sim \text{Poisson}\{\bar{y}_i(\boldsymbol{\mu}) + \sigma^2\}$$

or just use WLS...

Complications:

- Intractability of likelihood for Poisson+Gaussian
- Compound Poisson distribution due to photon-energy-dependent detector signal.

X-ray statistical modeling is a current research area in several groups!

## Choice 4. Cost Functions

Components:

- *Data-mismatch* term
- *Regularization* term (and regularization parameter  $\beta$ )
- Constraints (e.g., nonnegativity)

Cost function:

$$\Psi(\mathbf{x}) = \text{DataMismatch}(\mathbf{y}, \mathbf{Ax}) + \beta \text{Roughness}(\mathbf{x})$$

Reconstruct image  $\hat{\mathbf{x}}$  by minimization:

$$\hat{\mathbf{x}} \triangleq \arg \min_{\mathbf{x} \geq 0} \Psi(\mathbf{x})$$

Actually *several* sub-choices to make for Choice 4 ...

Distinguishes “statistical methods” from “algebraic methods” for “ $\mathbf{y} = \mathbf{Ax}$ .”

# Why Cost Functions?

(vs “procedure” e.g., adaptive neural net with wavelet denoising)

## Theoretical reasons

ML is based on minimizing a cost function: the negative log-likelihood

- ML is asymptotically consistent
- ML is asymptotically unbiased
- ML is asymptotically efficient (under true statistical model...)
- **Estimation**: Penalized-likelihood achieves uniform CR bound asymptotically
- **Detection**: Qi and Huesman showed analytically that MAP reconstruction outperforms FBP for SKE/BKE lesion detection (T-MI, Aug. 2001)

## Practical reasons

- Stability of estimates (if  $\Psi$  and algorithm chosen properly)
- Predictability of properties (despite nonlinearities)
- Empirical evidence (?)



# Bayesian Framework

Given a prior distribution  $p(\mathbf{x})$  for image vectors  $\mathbf{x}$ , by Bayes' rule:

$$\text{posterior: } p(\mathbf{x}|\mathbf{y}) = p(\mathbf{y}|\mathbf{x}) p(\mathbf{x}) / p(\mathbf{y})$$

so

$$\log p(\mathbf{x}|\mathbf{y}) = \log p(\mathbf{y}|\mathbf{x}) + \log p(\mathbf{x}) - \log p(\mathbf{y})$$

- $-\log p(\mathbf{y}|\mathbf{x})$  corresponds to data mismatch term (negative log-likelihood)
- $-\log p(\mathbf{x})$  corresponds to regularizing penalty function

**Maximum a posteriori (MAP) estimator:**

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} \log p(\mathbf{x}|\mathbf{y}) = \arg \max_{\mathbf{x}} \log p(\mathbf{y}|\mathbf{x}) + \log p(\mathbf{x})$$

- Has certain optimality properties (provided  $p(\mathbf{y}|\mathbf{x})$  and  $p(\mathbf{x})$  are correct).
- Same form as  $\Psi$

## Choice 4.1: Data-Mismatch Term

Options (for emission tomography):

- Negative log-likelihood of statistical model. Poisson *emission* case:

$$-L(\mathbf{x}; \mathbf{y}) = -\log p(\mathbf{y}|\mathbf{x}) = \sum_{i=1}^{n_d} ([\mathbf{Ax}]_i + r_i) - y_i \log([\mathbf{Ax}]_i + r_i) + \log y_i!$$

- Ordinary (unweighted) least squares:  $\sum_{i=1}^{n_d} \frac{1}{2} (y_i - \hat{r}_i - [\mathbf{Ax}]_i)^2$
- Data-weighted least squares:  $\sum_{i=1}^{n_d} \frac{1}{2} (y_i - \hat{r}_i - [\mathbf{Ax}]_i)^2 / \hat{\sigma}_i^2$ ,  $\hat{\sigma}_i^2 = \max(y_i + \hat{r}_i, \sigma_{\min}^2)$ , (causes bias due to data-weighting).
- Reweighted least-squares:  $\hat{\sigma}_i^2 = [\mathbf{A} \hat{\mathbf{x}}]_i + \hat{r}_i$
- Model-weighted least-squares (nonquadratic, but convex!)

$$\sum_{i=1}^{n_d} \frac{1}{2} (y_i - \hat{r}_i - [\mathbf{Ax}]_i)^2 / ([\mathbf{Ax}]_i + \hat{r}_i)$$

- Nonquadratic cost-functions that are robust to outliers
- ...

### Considerations

- Faithfulness to statistical model vs computation
- Ease of optimization (convex?, quadratic?)
- Effect of statistical modeling errors

## Choice 4.2: Regularization

Forcing too much “data fit” gives noisy images

Ill-conditioned problems: small data noise causes large image noise

### Solutions:

- Noise-reduction methods
- True regularization methods

### Noise-reduction methods

- Modify the *data*
  - Prefilter or “denoise” the sinogram measurements
  - Extrapolate missing (e.g., truncated) data
- Modify an *algorithm* derived for an ill-conditioned problem
  - Stop algorithm before convergence
  - Run to convergence, post-filter
  - Toss in a filtering step every iteration or couple iterations
  - Modify update to “dampen” high-spatial frequencies

# Noise-Reduction vs True Regularization

## Advantages of **noise-reduction** methods

- Simplicity (?)
- Familiarity
- Appear less subjective than using penalty functions or priors
- Only fiddle factors are # of iterations, or amount of smoothing
- Resolution/noise tradeoff usually varies with iteration (stop when image looks good - in principle)
- Changing post-smoothing does not require re-iterating

## Advantages of **true regularization** methods

- Stability (unique minimizer & convergence  $\implies$  initialization independence)
- Faster convergence
- Predictability
- Resolution can be made object independent
- Controlled resolution (e.g., spatially uniform, edge preserving)
- Start with reasonable image (e.g., FBP)  $\implies$  reach solution faster.

# True Regularization Methods

Redefine the *problem* to eliminate ill-conditioning, rather than patching the data or algorithm!

## Options

- Use bigger pixels (fewer basis functions)
  - Visually unappealing
  - Can only preserve edges coincident with pixel edges
  - Results become even less invariant to translations
- Method of sieves (constrain image roughness)
  - Condition number for “pre-emission space” can be even worse
  - Lots of iterations
  - Commutability condition rarely holds exactly in practice
  - Degenerates to post-filtering in some cases

- Change cost function by adding a roughness penalty / prior

$$\hat{\mathbf{x}} = \arg \min_x \Psi(\mathbf{x}), \quad \Psi(\mathbf{x}) = \mathcal{L}(\mathbf{x}) + \beta R(\mathbf{x})$$

- Disadvantage: apparently subjective choice of penalty
- Apparent difficulty in choosing penalty parameter(s), e.g.,  $\beta$  (cf. apodizing filter / cutoff frequency in FBP)

# Penalty Function Considerations

- Computation
- Algorithm complexity
- Uniqueness of minimizer of  $\Psi(\mathbf{x})$
- Resolution properties (edge preserving?)
- # of adjustable parameters
- Predictability of properties (resolution and noise)

## Choices

- separable vs nonseparable
- quadratic vs nonquadratic
- convex vs nonconvex

# Penalty Functions: Separable vs Nonseparable

## Separable

- Identity norm:  $R(\mathbf{x}) = \frac{1}{2}\mathbf{x}'\mathbf{I}\mathbf{x} = \sum_{j=1}^{n_p} x_j^2/2$   
penalizes large values of  $\mathbf{x}$ , but causes “squashing bias”
- Entropy:  $R(\mathbf{x}) = \sum_{j=1}^{n_p} x_j \log x_j$
- Gaussian prior with mean  $\mu_j$ , variance  $\sigma_j^2$ :  $R(\mathbf{x}) = \sum_{j=1}^{n_p} \frac{(x_j - \mu_j)^2}{2\sigma_j^2}$
- Gamma prior  $R(\mathbf{x}) = \sum_{j=1}^{n_p} p(x_j, \mu_j, \sigma_j)$  where  $p(x, \mu, \sigma)$  is Gamma pdf

The first two basically keep pixel values from “blowing up.”

The last two encourage pixels values to be close to prior means  $\mu_j$ .

$$\text{General separable form: } R(\mathbf{x}) = \sum_{j=1}^{n_p} f_j(x_j)$$

Slightly simpler for minimization, but these do not explicitly enforce smoothness. The simplicity advantage has been overcome in newer algorithms.

# Penalty Functions: Separable vs Nonseparable

**Nonseparable** (partially couple pixel values) to penalize *roughness*

$x_1$	$x_2$	$x_3$
$x_4$	$x_5$	

Example

$$R(\mathbf{x}) = (x_2 - x_1)^2 + (x_3 - x_2)^2 + (x_5 - x_4)^2 \\ + (x_4 - x_1)^2 + (x_5 - x_2)^2$$

2	2	2
2	1	

$$R(\mathbf{x}) = 1$$

3	3	1
2	2	

$$R(\mathbf{x}) = 6$$

1	3	1
2	2	

$$R(\mathbf{x}) = 10$$

Rougher images  $\implies$  larger  $R(\mathbf{x})$  values



# Roughness Penalty Functions

First-order neighborhood and pairwise pixel differences:

$$R(\mathbf{x}) = \sum_{j=1}^{n_p} \frac{1}{2} \sum_{k \in \mathcal{N}_j} \psi(x_j - x_k)$$

$\mathcal{N}_j \triangleq$  *neighborhood* of  $j$ th pixel (e.g., left, right, up, down)  
 $\psi$  called the *potential function*

Finite-difference approximation to continuous roughness measure:

$$R(f(\cdot)) = \int \|\nabla f(\vec{r})\|^2 d\vec{r} = \int \left| \frac{\partial}{\partial x} f(\vec{r}) \right|^2 + \left| \frac{\partial}{\partial y} f(\vec{r}) \right|^2 + \left| \frac{\partial}{\partial z} f(\vec{r}) \right|^2 d\vec{r}.$$

Second derivatives also useful:  
(More choices!)

$$\left. \frac{\partial^2}{\partial x^2} f(\vec{r}) \right|_{\vec{r}=\vec{r}_j} \approx f(\vec{r}_{j+1}) - 2f(\vec{r}_j) + f(\vec{r}_{j-1})$$

$$R(\mathbf{x}) = \sum_{j=1}^{n_p} \psi(x_{j+1} - 2x_j + x_{j-1}) + \dots$$

# Penalty Functions: General Form

$$R(\mathbf{x}) = \sum_k \psi_k([\mathbf{C}\mathbf{x}]_k) \quad \text{where} \quad [\mathbf{C}\mathbf{x}]_k = \sum_{j=1}^{n_p} c_{kj}x_j$$

**Example:**

$x_1$	$x_2$	$x_3$
$x_4$	$x_5$	

$$\mathbf{C}\mathbf{x} = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 \\ -1 & 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} x_2 - x_1 \\ x_3 - x_2 \\ x_5 - x_4 \\ x_4 - x_1 \\ x_5 - x_2 \end{bmatrix}$$

$$\begin{aligned} R(\mathbf{x}) &= \sum_{k=1}^5 \psi_k([\mathbf{C}\mathbf{x}]_k) \\ &= \psi_1(x_2 - x_1) + \psi_2(x_3 - x_2) + \psi_3(x_5 - x_4) + \psi_4(x_4 - x_1) + \psi_5(x_5 - x_2) \end{aligned}$$

# Penalty Functions: Quadratic vs Nonquadratic

$$R(\mathbf{x}) = \sum_k \psi_k([\mathbf{C}\mathbf{x}]_k)$$

## Quadratic $\psi_k$

If  $\psi_k(t) = t^2/2$ , then  $R(\mathbf{x}) = \frac{1}{2}\mathbf{x}'\mathbf{C}'\mathbf{C}\mathbf{x}$ , a quadratic form.

- Simpler optimization
- Global smoothing

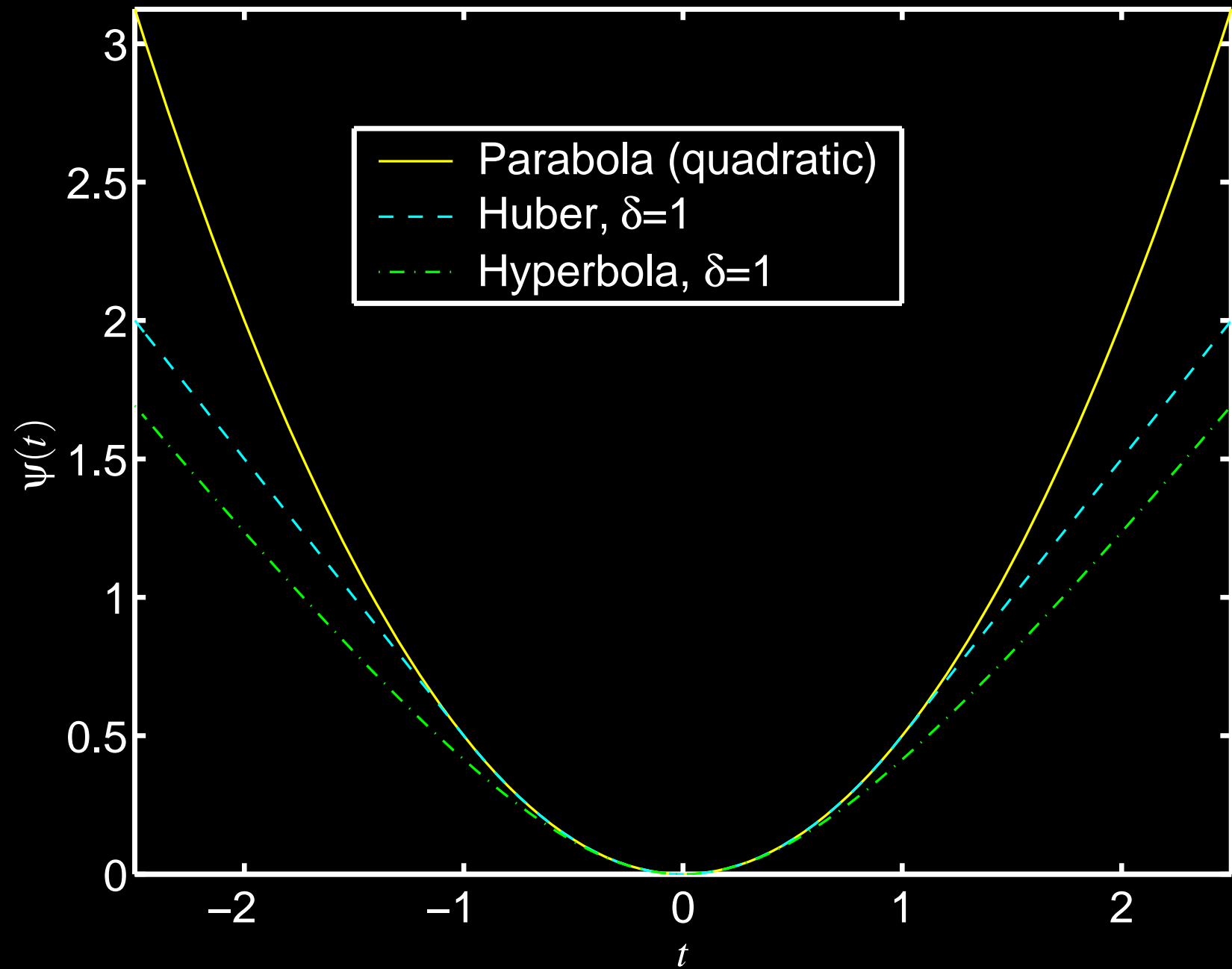
## Nonquadratic $\psi_k$

- Edge preserving
- More complicated optimization. (This is essentially solved in convex case.)
- Unusual noise properties
- Analysis/prediction of resolution and noise properties is difficult
- More adjustable parameters (e.g.,  $\delta$ )

Example: Huber function.  $\psi(t) \triangleq \begin{cases} t^2/2, & |t| \leq \delta \\ \delta|t| - \delta^2/2, & |t| > \delta \end{cases}$

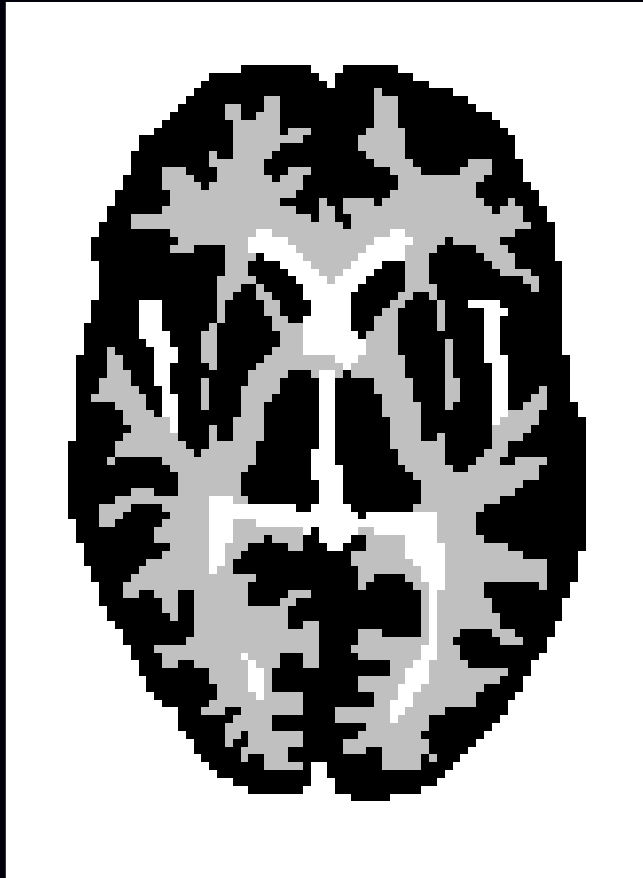
Example: Hyperbola function.  $\psi(t) \triangleq \delta^2 \left( \sqrt{1 + (t/\delta)^2} - 1 \right)$

# Quadratic vs Non-quadratic Potential Functions

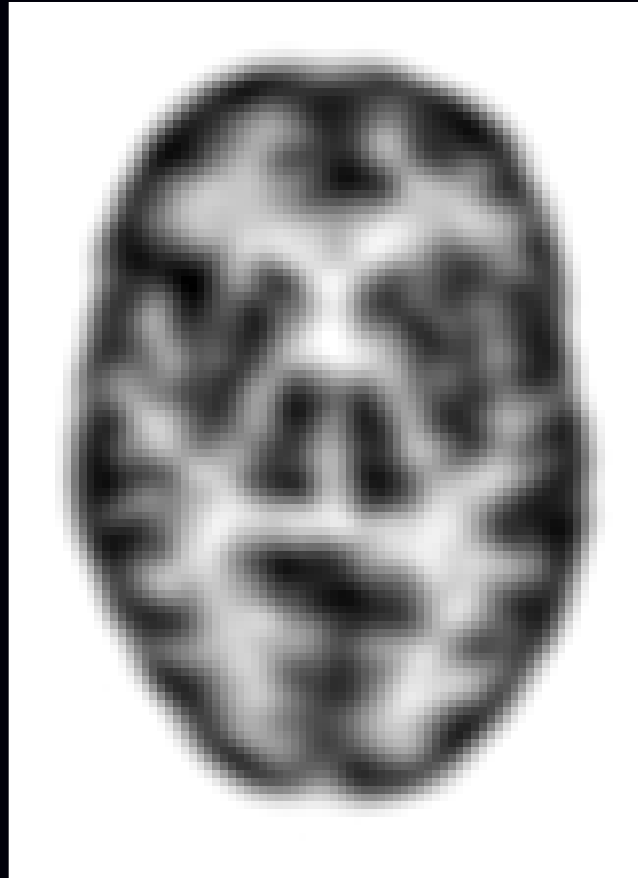


Lower cost for large differences  $\implies$  edge preservation

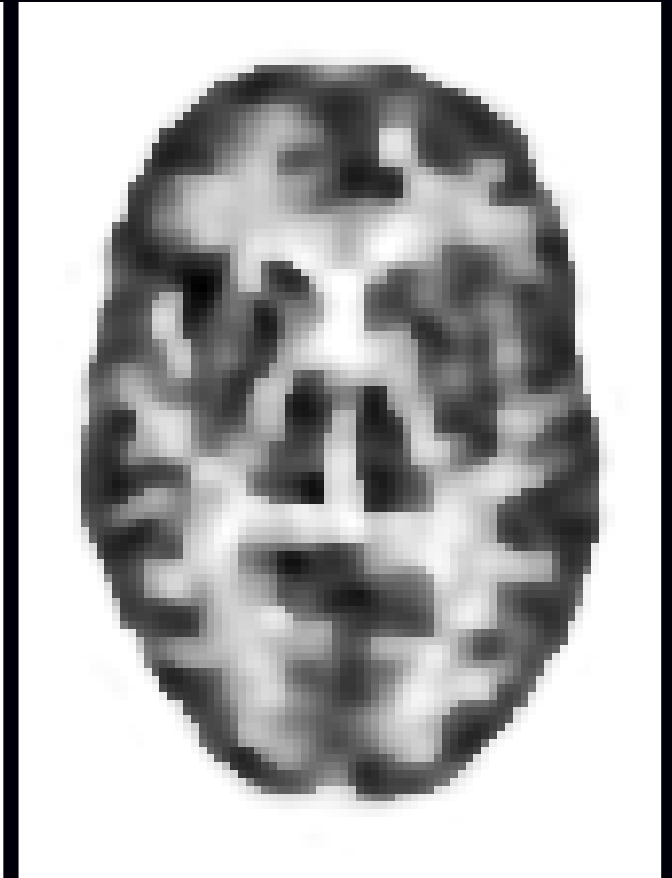
## Edge-Preserving Reconstruction Example



Phantom

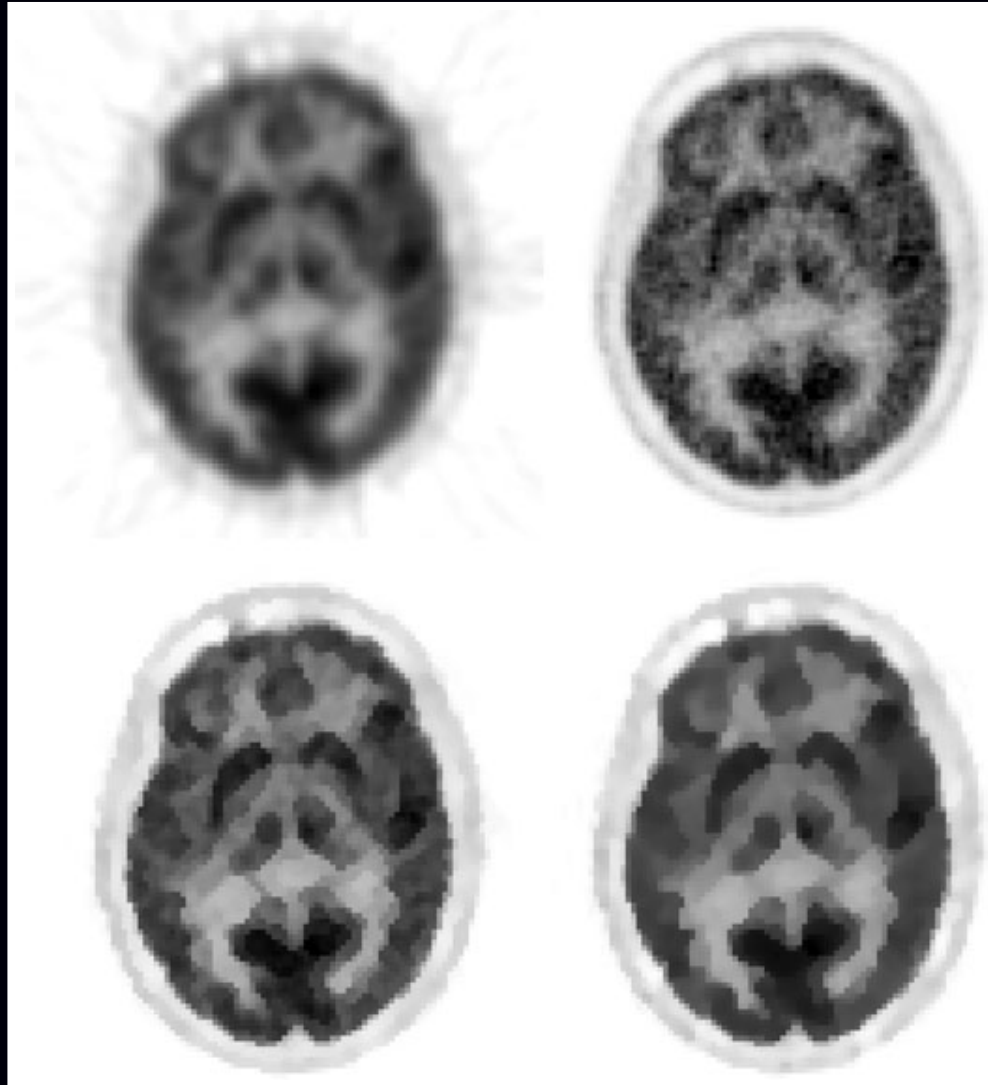


Quadratic Penalty



Huber Penalty

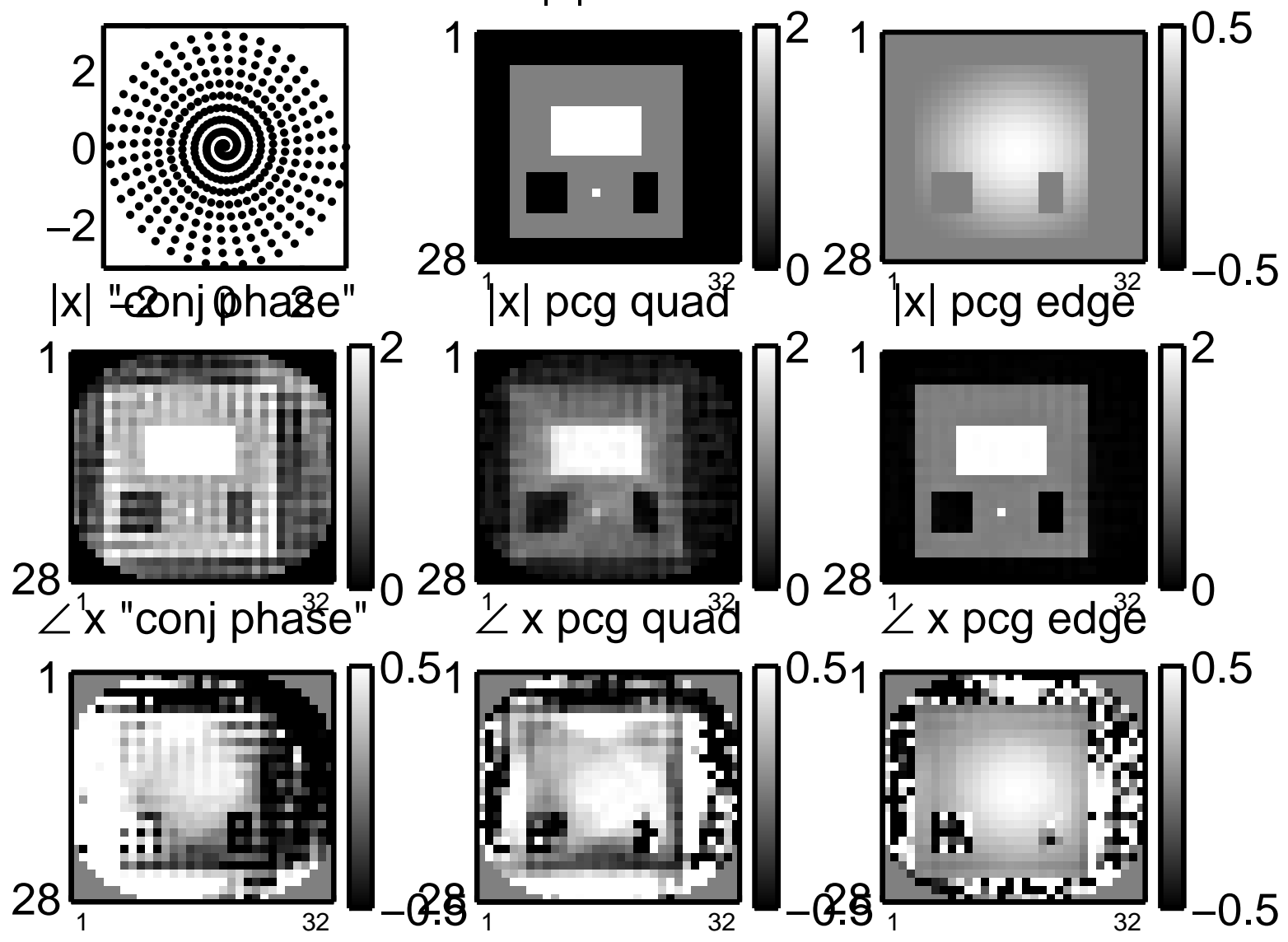
## More “Edge Preserving” Regularization



Chlewicki *et al.*, PMB, Oct. 2004: “Noise reduction and convergence of Bayesian algorithms with blobs based on the Huber function and median root prior”

# Piecewise Constant "Cartoon" Objects

400 k-space samples



# Total Variation Regularization

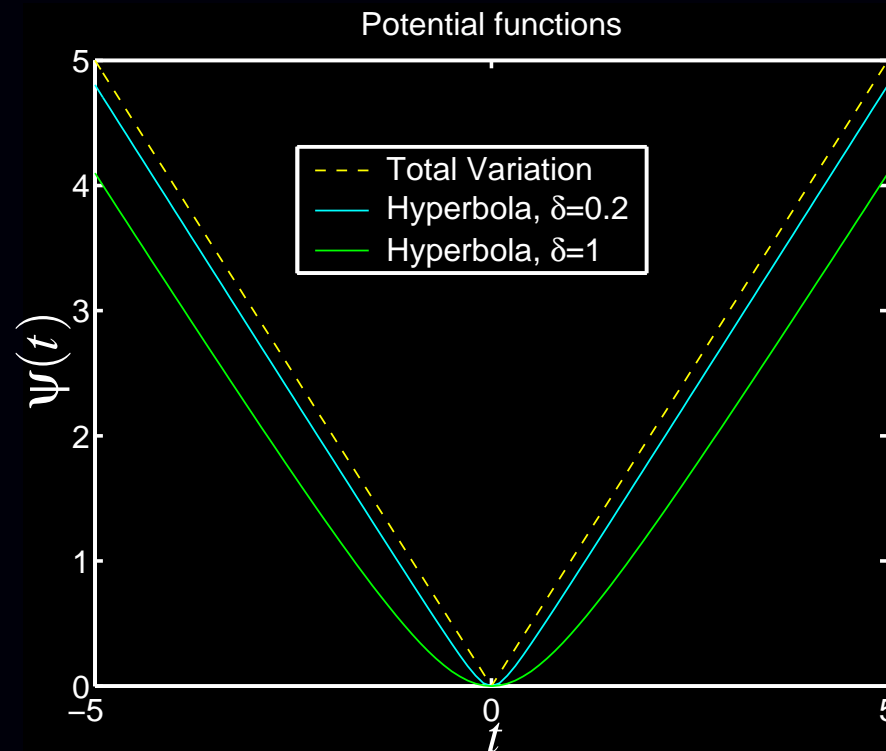
Non-quadratic roughness penalty:

$$\int \|\nabla f(\vec{r})\| d\vec{r} \approx \sum_k |[\mathbf{C}\mathbf{x}]_k|$$

Uses *magnitude* instead of *squared magnitude* of gradient.

Problem:  $|\cdot|$  is not differentiable.

Practical solution:  $|t| \approx \delta \left( \sqrt{1 + (t/\delta)^2} - 1 \right)$  (hyperbola!)





# Total Variation Example

todo: example showing blocky reconstruction with TV

# Compressed Sensing

aka compressive sampling or sparsity regularization

Idea: find a basis  $\mathbf{B}$  for representing  $\mathbf{x}$  in terms of coefficients  $\boldsymbol{\theta}$ :  
 $\mathbf{x} = \mathbf{B}\boldsymbol{\theta}$ , where only a “small number” of  $\theta_j$  values are nonzero.

Previous cost function:  $\Psi(\mathbf{x}) = \text{DataMismatch}(\mathbf{y}, \mathbf{A}\mathbf{x}) + \beta \text{Roughness}(\mathbf{x})$

New cost function with sparsity regularization:

$$\Psi(\boldsymbol{\theta}) = \text{DataMismatch}(\mathbf{y}, \mathbf{A}\mathbf{B}\boldsymbol{\theta}) + \beta \|\boldsymbol{\theta}\|_0$$

$$\|\boldsymbol{\theta}\|_p \triangleq (\sum_j |\theta_j|^p)^{1/p}$$

Recall:  $\|\boldsymbol{\theta}\|_\infty \triangleq \lim_{p \rightarrow \infty} \|\boldsymbol{\theta}\|_p = \max_j |\theta_j|$

$$\|\boldsymbol{\theta}\|_0 \triangleq \lim_{p \rightarrow 0} \|\boldsymbol{\theta}\|_p^p = \sum_j 1_{\{\theta_j \neq 0\}} \quad (\text{not a norm in the Banach sense})$$

Because  $\|\boldsymbol{\theta}\|_0$  is nonconvex, it usually is replaced with  $\|\boldsymbol{\theta}\|_1$ .

Because  $\|\boldsymbol{\theta}\|_1$  is nondifferentiable, the corner is often rounded (hyperbola).

If  $\mathbf{B}$  is the Harr wavelet basis, then  $\|\boldsymbol{\theta}\|_1 = \|\mathbf{B}^{-1}\mathbf{x}\|_1$  is similar to TV regularization.

For certain types of under-sampled measurements  $\mathbf{A}$ , “good” reconstructions are possible!  
Example: radial k-space sampling for Shepp-Logan.

# Penalty Functions: Convex vs Nonconvex

## Convex

- Easier to optimize
- Guaranteed unique minimizer of  $\Psi$  (for convex negative log-likelihood)

## Nonconvex

- Greater degree of edge preservation
- Nice images for piecewise-constant phantoms!
- Even more unusual noise properties
- Multiple extrema
- More complicated optimization (simulated / deterministic annealing)
- Estimator  $\hat{\mathbf{x}}$  becomes a discontinuous function of data  $\mathbf{Y}$

## Nonconvex examples

- “broken parabola”

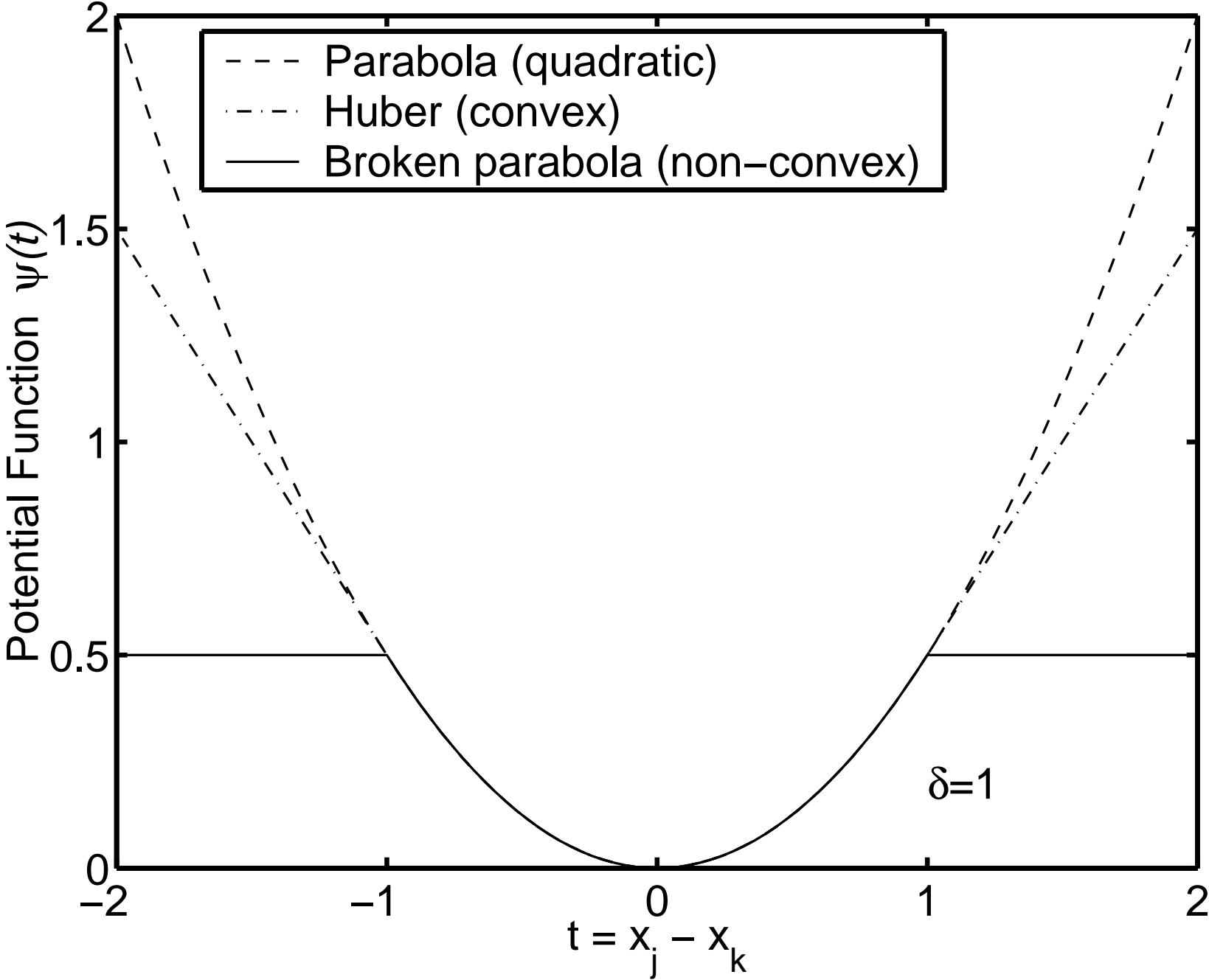
$$\psi(t) = \min(t^2, t_{\max}^2)$$

- true median root prior:

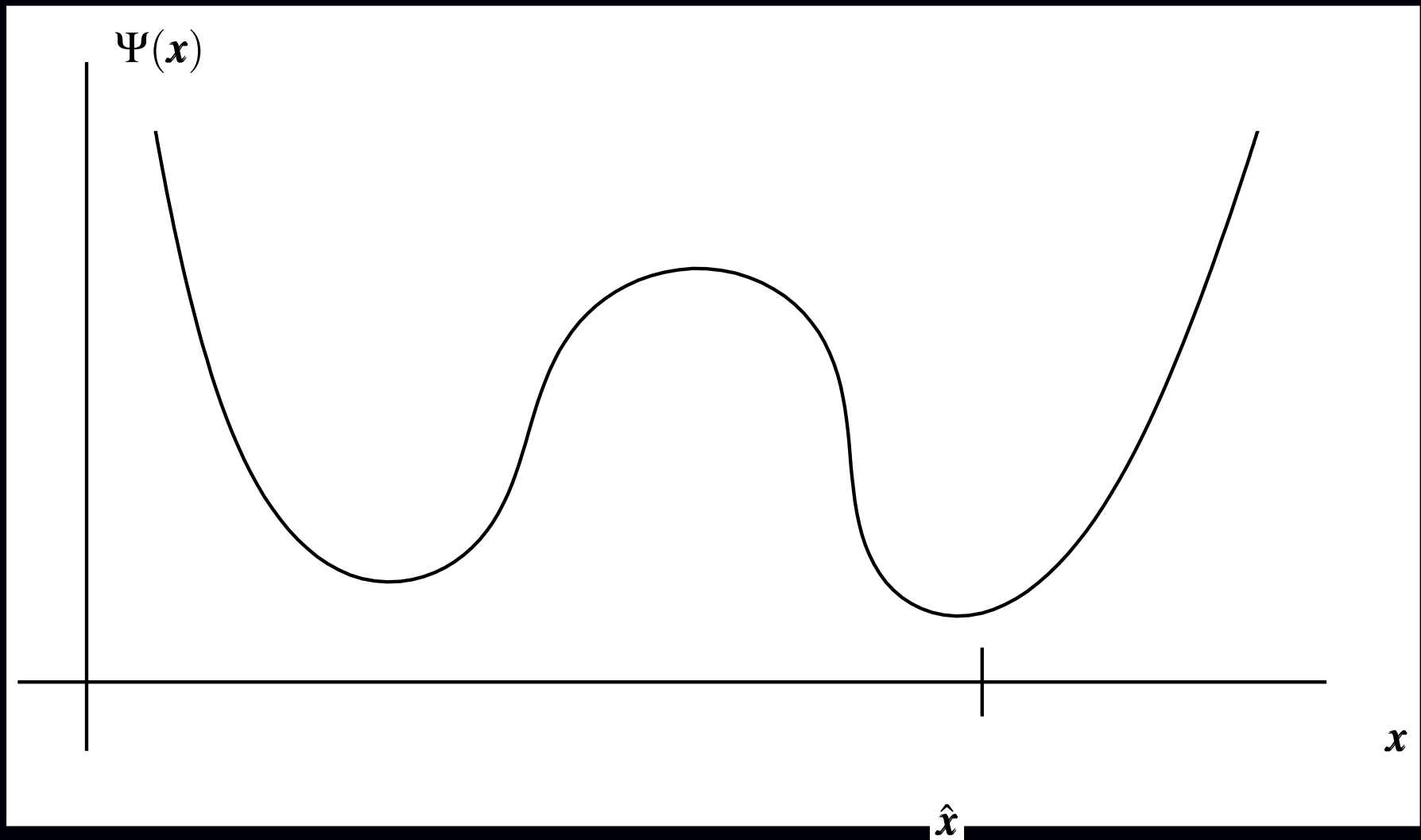
$$R(\mathbf{x}) = \sum_{j=1}^{n_p} \frac{(x_j - \text{median}_j(\mathbf{x}))^2}{\text{median}_j(\mathbf{x})} \quad \text{where } \text{median}_j(\mathbf{x}) \text{ is local median}$$

Exception: orthonormal wavelet threshold *denoising* via nonconvex potentials!

# Potential Functions



# Local Extrema and Discontinuous Estimators



Small change in data  $\implies$  large change in minimizer  $\hat{x}$ .  
Using convex penalty functions obviates this problem.

# Augmented Regularization Functions

Replace roughness penalty  $R(\mathbf{x})$  with  $R(\mathbf{x}|\mathbf{b}) + \alpha R(\mathbf{b})$ , where the elements of  $\mathbf{b}$  (often binary) indicate boundary locations.

- Line-site methods
- Level-set methods

Joint estimation problem:

$$(\hat{\mathbf{x}}, \hat{\mathbf{b}}) = \arg \min_{\mathbf{x}, \mathbf{b}} \Psi(\mathbf{x}, \mathbf{b}), \quad \Psi(\mathbf{x}, \mathbf{b}) = \ell(\mathbf{x}; \mathbf{y}) + \beta R(\mathbf{x}|\mathbf{b}) + \alpha R(\mathbf{b}).$$

**Example:**  $b_{jk}$  indicates the presence of edge between pixels  $j$  and  $k$ :

$$R(\mathbf{x}|\mathbf{b}) = \sum_{j=1}^{n_p} \sum_{k \in \mathcal{N}_j} (1 - b_{jk}) \frac{1}{2} (x_j - x_k)^2$$

Penalty to discourage too many edges (e.g.):

$$R(\mathbf{b}) = \sum_{jk} b_{jk}.$$

- Can encourage local edge continuity
- May require annealing methods for minimization

# Modified Penalty Functions

$$R(\mathbf{x}) = \sum_{j=1}^{n_p} \frac{1}{2} \sum_{k \in \mathcal{N}_j} w_{jk} \Psi(x_j - x_k)$$

Adjust weights  $\{w_{jk}\}$  to

- Control resolution properties
- Incorporate anatomical side information (MR/CT)  
(avoid smoothing across anatomical boundaries)

## Recommendations

- Emission tomography:
  - Begin with quadratic (nonseparable) penalty functions
  - Consider modified penalty for resolution control and choice of  $\beta$
  - Use modest regularization and post-filter more if desired
- Transmission tomography (attenuation maps), X-ray CT
  - consider convex nonquadratic (e.g., Huber) penalty functions
  - choose  $\delta$  based on attenuation map **units** (water, bone, etc.)
  - choice of regularization parameter  $\beta$  remains nontrivial,  
learn appropriate values by experience for given study type

## Choice 4.3: Constraints

- Nonnegativity
- Known support
- Count preserving
- Upper bounds on values  
*e.g.*, maximum  $\mu$  of attenuation map in transmission case

### Considerations

- Algorithm complexity
- Computation
- Convergence rate
- Bias (in low-count regions)
- ...



## Open Problems

- Performance prediction for nonquadratic regularization
- Effect of nonquadratic regularization on detection tasks
- Choice of regularization parameters for nonquadratic regularization

# Summary

- 1. Object parameterization: function  $f(\vec{r})$  vs vector  $x$
- 2. System physical model:  $s_i(\vec{r})$
- 3. Measurement statistical model  $Y_i \sim \boxed{?}$
- 4. Cost function: data-mismatch / regularization / constraints

**Reconstruction Method  $\triangleq$  Cost Function + Algorithm**

Naming convention: “criterion”-“algorithm”:

- ML-EM, MAP-OSL, PL-SAGE, PWLS+SOR, PWLS-CG, ...

# Part 3. Algorithms

**Method = Cost Function + Algorithm**

## Outline

- Ideal algorithm
- Classical general-purpose algorithms
- Considerations:
  - nonnegativity
  - parallelization
  - convergence rate
  - monotonicity
- Algorithms tailored to cost functions for imaging
  - Optimization transfer
  - EM-type methods
  - Poisson emission problem
  - Poisson transmission problem
- Ordered-subsets / block-iterative algorithms
  - Recent convergent versions (relaxation, incrementalism)

## Why iterative algorithms?

- For nonquadratic  $\Psi$ , no closed-form solution for minimizer.
- For quadratic  $\Psi$  with nonnegativity constraints, no closed-form solution.
- For quadratic  $\Psi$  without constraints, closed-form solutions:

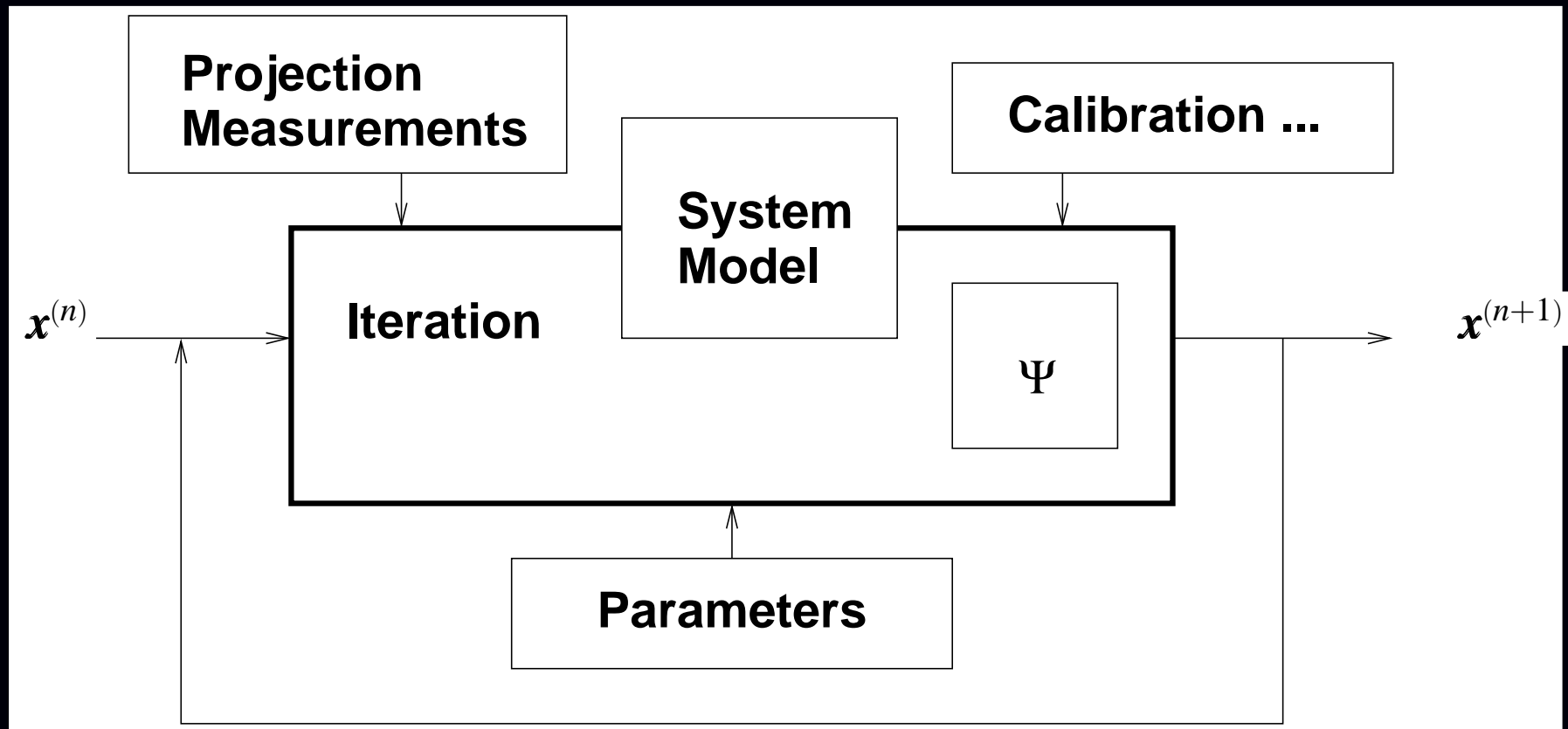
$$\text{PWLS: } \hat{\mathbf{x}} = \arg \min_x \|\mathbf{y} - \mathbf{Ax}\|_{\mathbf{W}^{1/2}}^2 + \mathbf{x}'\mathbf{Rx} = [\mathbf{A}'\mathbf{WA} + \mathbf{R}]^{-1}\mathbf{A}'\mathbf{W}\mathbf{y}$$

$$\text{OLS: } \hat{\mathbf{x}} = \arg \min_x \|\mathbf{y} - \mathbf{Ax}\|^2 = [\mathbf{A}'\mathbf{A}]^{-1}\mathbf{A}'\mathbf{y}$$

Impractical (memory and computation) for realistic problem sizes.  
 $\mathbf{A}$  is sparse, but  $\mathbf{A}'\mathbf{A}$  is not.

All algorithms are imperfect. No single best solution.

# General Iteration



Deterministic iterative mapping:  $\mathbf{x}^{(n+1)} = \mathcal{M}(\mathbf{x}^{(n)})$

# Ideal Algorithm

$$\mathbf{x}^* \triangleq \arg \min_{\mathbf{x} \geq \mathbf{0}} \Psi(\mathbf{x}) \quad (\text{global minimizer})$$

## Properties

stable and convergent

converges quickly

globally convergent

fast

robust

user friendly

parallelizable

simple

flexible

(matrix stored by row or column, or factored, or projector/backprojector)

$\{\mathbf{x}^{(n)}\}$  converges to  $\mathbf{x}^*$  if run indefinitely

$\{\mathbf{x}^{(n)}\}$  gets “close” to  $\mathbf{x}^*$  in just a few iterations

$\lim_n \mathbf{x}^{(n)}$  independent of starting image  $\mathbf{x}^{(0)}$

requires minimal computation per iteration

insensitive to finite numerical precision

nothing to adjust (e.g., acceleration factors)

(when necessary)

easy to program and debug

accommodates any type of system model

Choices: forgo one or more of the above

# Classic Algorithms

## Non-gradient based

- Exhaustive search
- Nelder-Mead simplex (amoeba)

Converge very slowly, but work with nondifferentiable cost functions.

## Gradient based

- Gradient descent

$$\mathbf{x}^{(n+1)} \triangleq \mathbf{x}^{(n)} - \alpha \nabla \Psi(\mathbf{x}^{(n)})$$

Choosing  $\alpha$  to ensure convergence is nontrivial.

- Steepest descent

$$\mathbf{x}^{(n+1)} \triangleq \mathbf{x}^{(n)} - \alpha_n \nabla \Psi(\mathbf{x}^{(n)}) \quad \text{where} \quad \alpha_n \triangleq \arg \min_{\alpha} \Psi(\mathbf{x}^{(n)} - \alpha \nabla \Psi(\mathbf{x}^{(n)}))$$

Computing stepsize  $\alpha_n$  can be expensive or inconvenient.

## Limitations

- Converge slowly.
- Do not easily accommodate nonnegativity constraint.

# Gradients & Nonnegativity - A Mixed Blessing

**Unconstrained optimization** of differentiable cost functions:

$$\nabla \Psi(\mathbf{x}) = \mathbf{0} \text{ when } \mathbf{x} = \mathbf{x}^*$$

- A necessary condition always.
- A sufficient condition for strictly convex cost functions.
- Iterations search for zero of gradient.

**Nonnegativity-constrained minimization:**

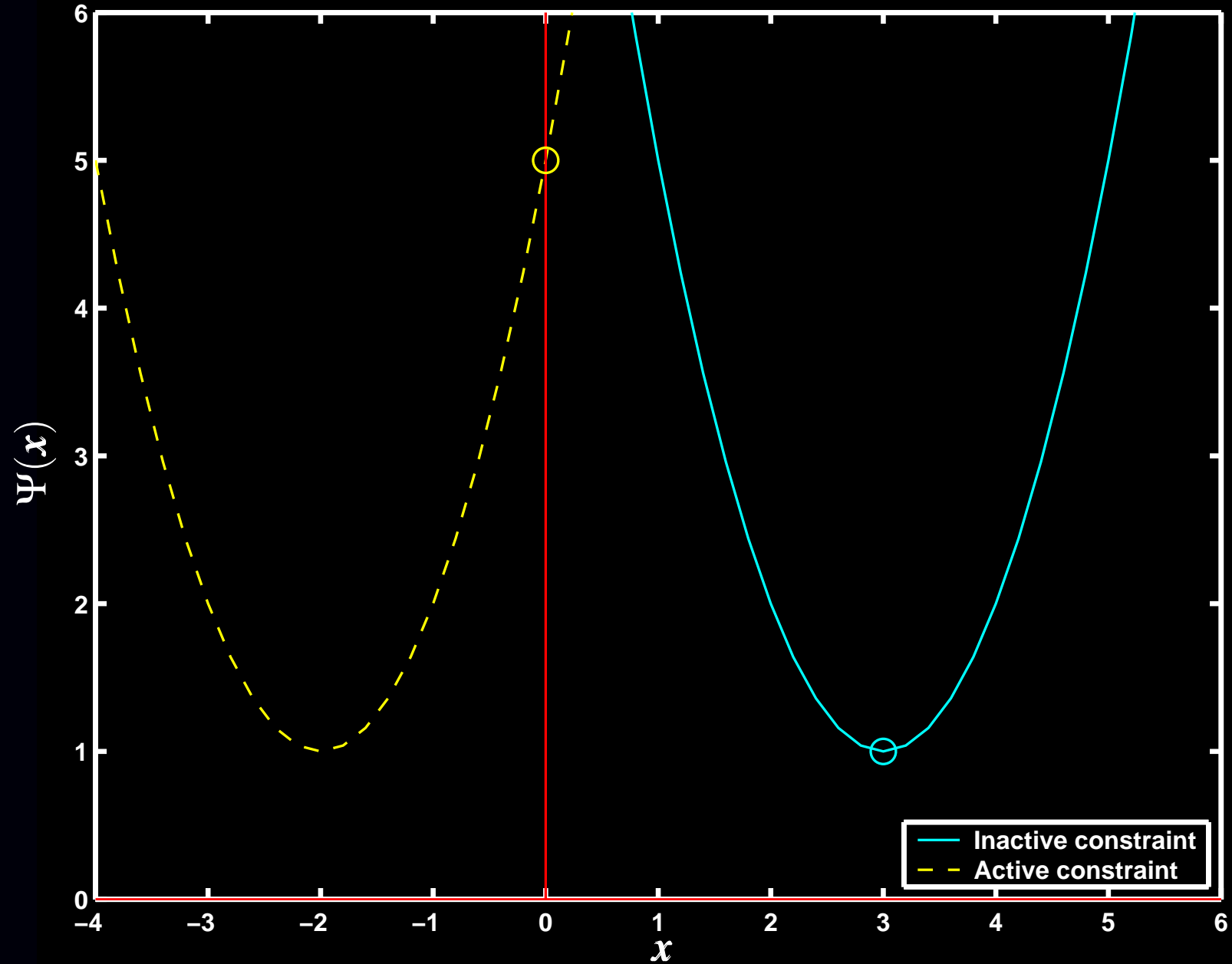
Karush-Kuhn-Tucker conditions

$$\left. \frac{\partial}{\partial x_j} \Psi(\mathbf{x}) \right|_{\mathbf{x}=\mathbf{x}^*} \text{ is } \begin{cases} = 0, & x_j^* > 0 \\ \geq 0, & x_j^* = 0 \end{cases}$$

- A necessary condition always.
- A sufficient condition for strictly convex cost functions.
- Iterations search for ???
- $0 = x_j^* \frac{\partial}{\partial x_j} \Psi(\mathbf{x}^*)$  is a necessary condition, but never sufficient condition.

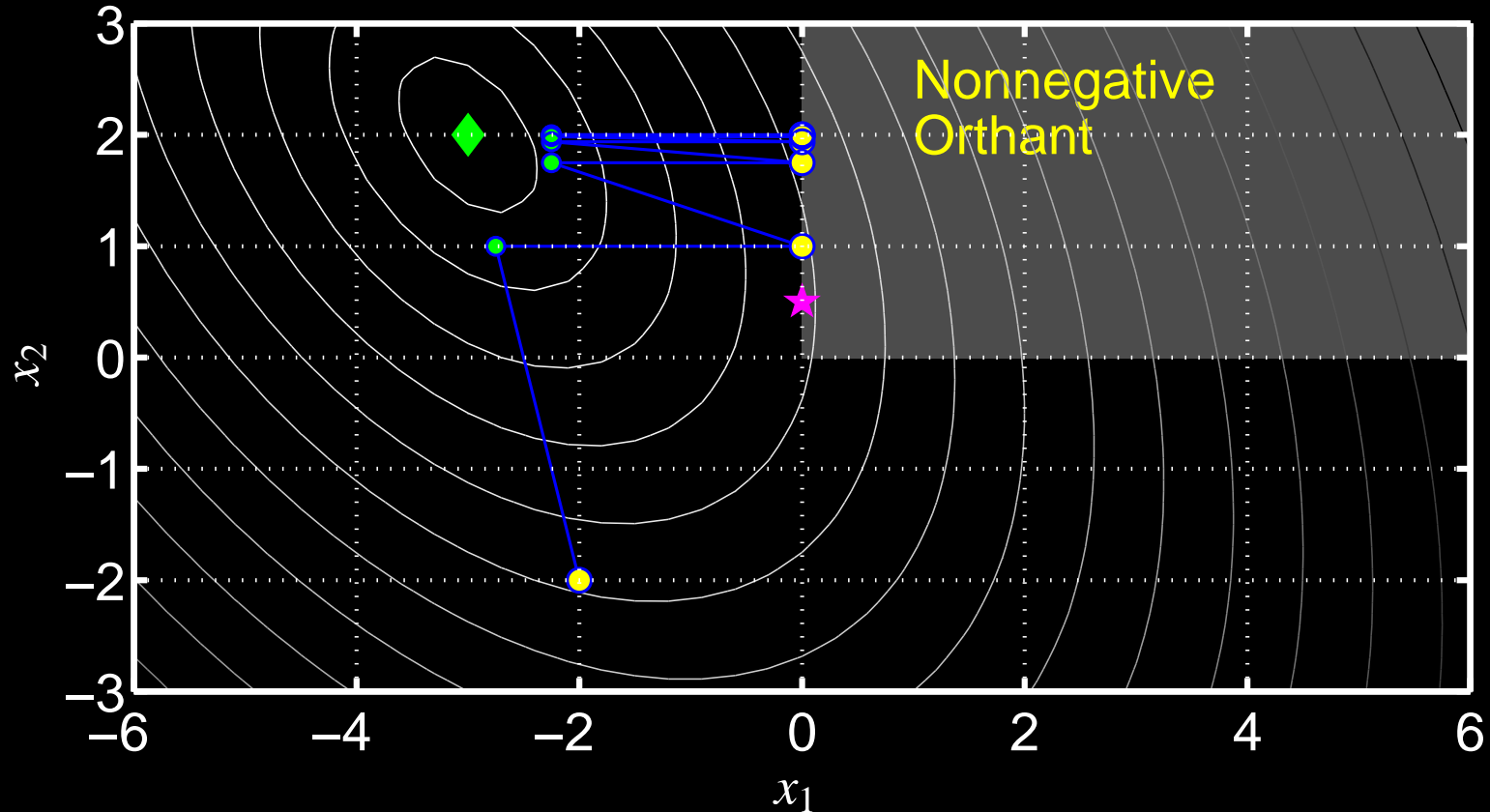


# Karush-Kuhn-Tucker Illustrated



# Why Not Clip Negatives?

WLS with Clipped Newton–Raphson



Newton-Raphson with negatives set to zero each iteration.  
Fixed-point of iteration is not the constrained minimizer!

# Newton-Raphson Algorithm

$$\mathbf{x}^{(n+1)} = \mathbf{x}^{(n)} - [\nabla^2 \Psi(\mathbf{x}^{(n)})]^{-1} \nabla \Psi(\mathbf{x}^{(n)})$$

## Advantage:

- Super-linear convergence rate (if convergent)

## Disadvantages:

- Requires twice-differentiable  $\Psi$
- Not guaranteed to converge
- Not guaranteed to monotonically decrease  $\Psi$
- Does not enforce nonnegativity constraint
- Computing Hessian  $\nabla^2 \Psi$  often expensive
- Impractical for image recovery due to matrix inverse

General purpose remedy: bound-constrained Quasi-Newton algorithms

# Newton's Quadratic Approximation

2nd-order Taylor series:

$$\Psi(\mathbf{x}) \approx \phi(\mathbf{x}; \mathbf{x}^{(n)}) \triangleq \Psi(\mathbf{x}^{(n)}) + \nabla \Psi(\mathbf{x}^{(n)}) (\mathbf{x} - \mathbf{x}^{(n)}) + \frac{1}{2} (\mathbf{x} - \mathbf{x}^{(n)})^T \nabla^2 \Psi(\mathbf{x}^{(n)}) (\mathbf{x} - \mathbf{x}^{(n)})$$

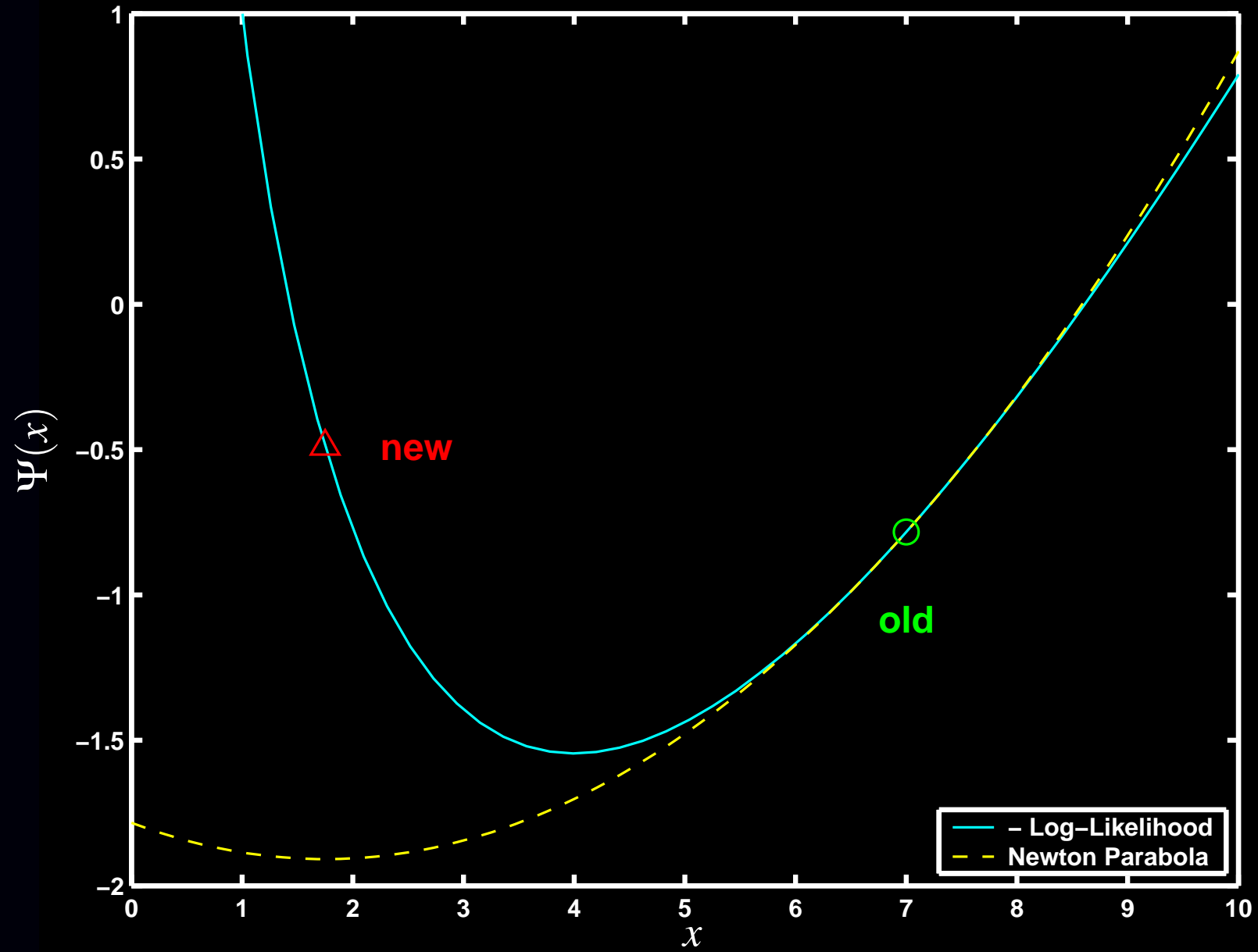
Set  $\mathbf{x}^{(n+1)}$  to the (“easily” found) minimizer of this quadratic approximation:

$$\begin{aligned} \mathbf{x}^{(n+1)} &\triangleq \arg \min_{\mathbf{x}} \phi(\mathbf{x}; \mathbf{x}^{(n)}) \\ &= \mathbf{x}^{(n)} - [\nabla^2 \Psi(\mathbf{x}^{(n)})]^{-1} \nabla \Psi(\mathbf{x}^{(n)}) \end{aligned}$$

Can be nonmonotone for Poisson emission tomography log-likelihood, even for a single pixel and single ray:

$$\Psi(x) = (x + r) - y \log(x + r).$$

# Nonmonotonicity of Newton-Raphson



## Consideration: Monotonicity

An algorithm is monotonic if

$$\Psi(\mathbf{x}^{(n+1)}) \leq \Psi(\mathbf{x}^{(n)}), \quad \forall \mathbf{x}^{(n)}.$$

Three categories of algorithms:

- Nonmonotonic (or unknown)
- Forced monotonic (e.g., by line search)
- **Intrinsically monotonic** (by design, simplest to implement)

---

### Forced monotonicity

Most nonmonotonic algorithms can be converted to forced monotonic algorithms by adding a line-search step:

$$\mathbf{x}^{\text{temp}} \triangleq \mathcal{M}(\mathbf{x}^{(n)}), \quad \mathbf{d} = \mathbf{x}^{\text{temp}} - \mathbf{x}^{(n)}$$
$$\mathbf{x}^{(n+1)} \triangleq \mathbf{x}^{(n)} - \alpha_n \mathbf{d}^{(n)} \quad \text{where} \quad \alpha_n \triangleq \arg \min_{\alpha} \Psi(\mathbf{x}^{(n)} - \alpha \mathbf{d}^{(n)})$$

Inconvenient, sometimes expensive, nonnegativity problematic.

# Conjugate Gradient (CG) Algorithm

## Advantages:

- Fast converging (if suitably preconditioned) (in unconstrained case)
- Monotonic (forced by line search in nonquadratic case)
- Global convergence (unconstrained case)
- Flexible use of system matrix  $A$  and tricks
- Easy to implement in unconstrained quadratic case
- Highly parallelizable

## Disadvantages:

- Nonnegativity constraint awkward (slows convergence?)
- Line-search somewhat awkward in nonquadratic cases
- Possible need to “restart” after many iterations

Highly recommended for unconstrained quadratic problems (e.g., PWLS without nonnegativity). Useful (but perhaps not ideal) for Poisson case too.

## Consideration: Parallelization

### **Simultaneous** (fully parallelizable)

update all pixels simultaneously using all data  
EM, Conjugate gradient, ISRA, OSL, SIRT, MART, ...

### **Block iterative** (ordered subsets)

update (nearly) all pixels using one subset of the data at a time  
OSEM, RBBI, ...

### **Row action**

update many pixels using a single ray at a time  
ART, RAMLA

### **Pixel grouped** (multiple column action)

update some (but not all) pixels simultaneously a time, using all data  
Grouped coordinate descent, multi-pixel SAGE  
(Perhaps the most nontrivial to implement)

### **Sequential** (column action)

update one pixel at a time, using all (relevant) data  
Coordinate descent, SAGE



# Coordinate Descent Algorithm

aka Gauss-Siedel, successive over-relaxation (SOR), iterated conditional modes (ICM)

Update one pixel at a time, holding others fixed to their most recent values:

$$x_j^{\text{new}} = \arg \min_{x_j \geq 0} \Psi \left( x_1^{\text{new}}, \dots, x_{j-1}^{\text{new}}, x_j, x_{j+1}^{\text{old}}, \dots, x_{n_p}^{\text{old}} \right), \quad j = 1, \dots, n_p$$

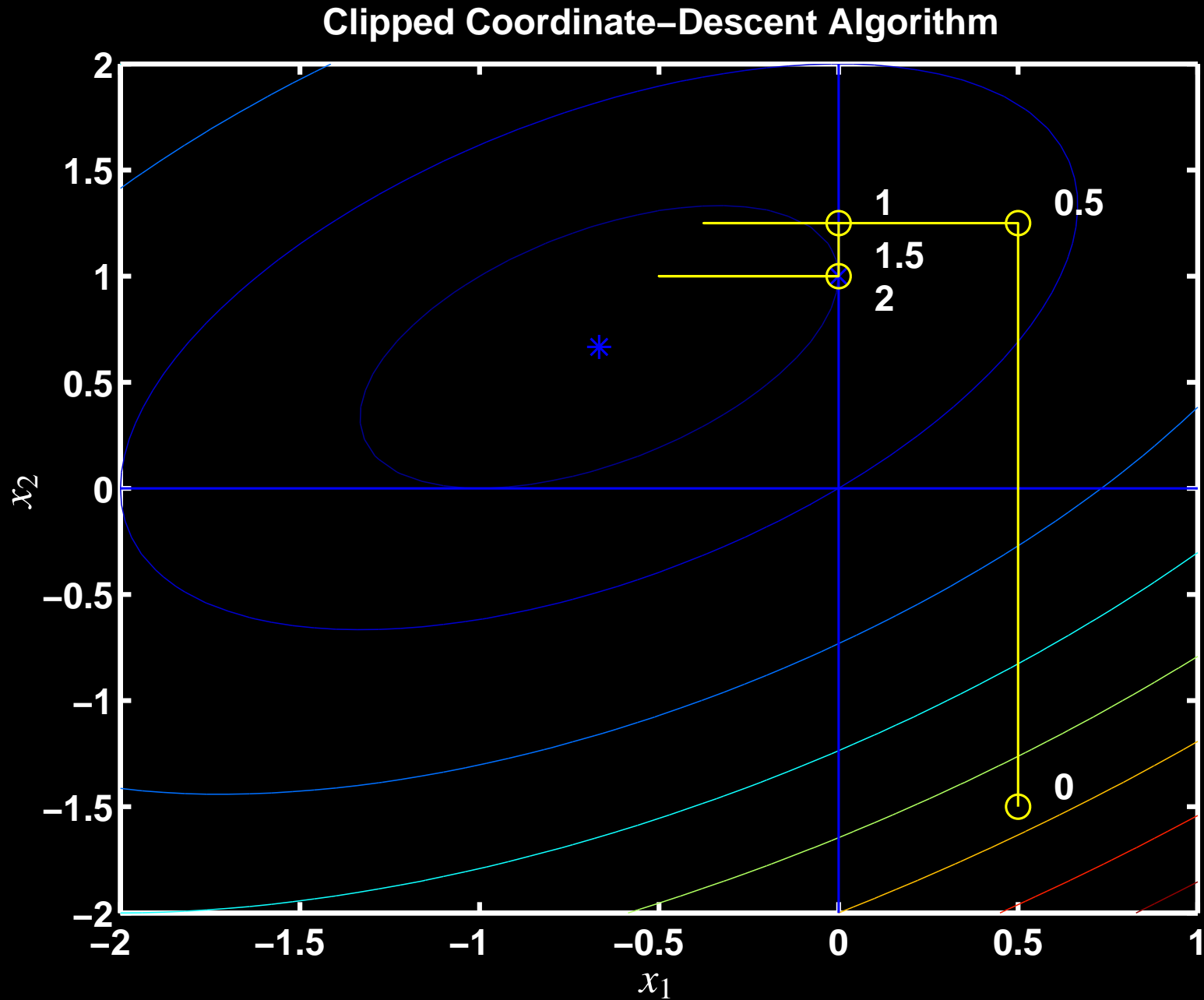
## Advantages:

- Intrinsically monotonic
- Fast converging (from good initial image)
- Global convergence
- Nonnegativity constraint trivial

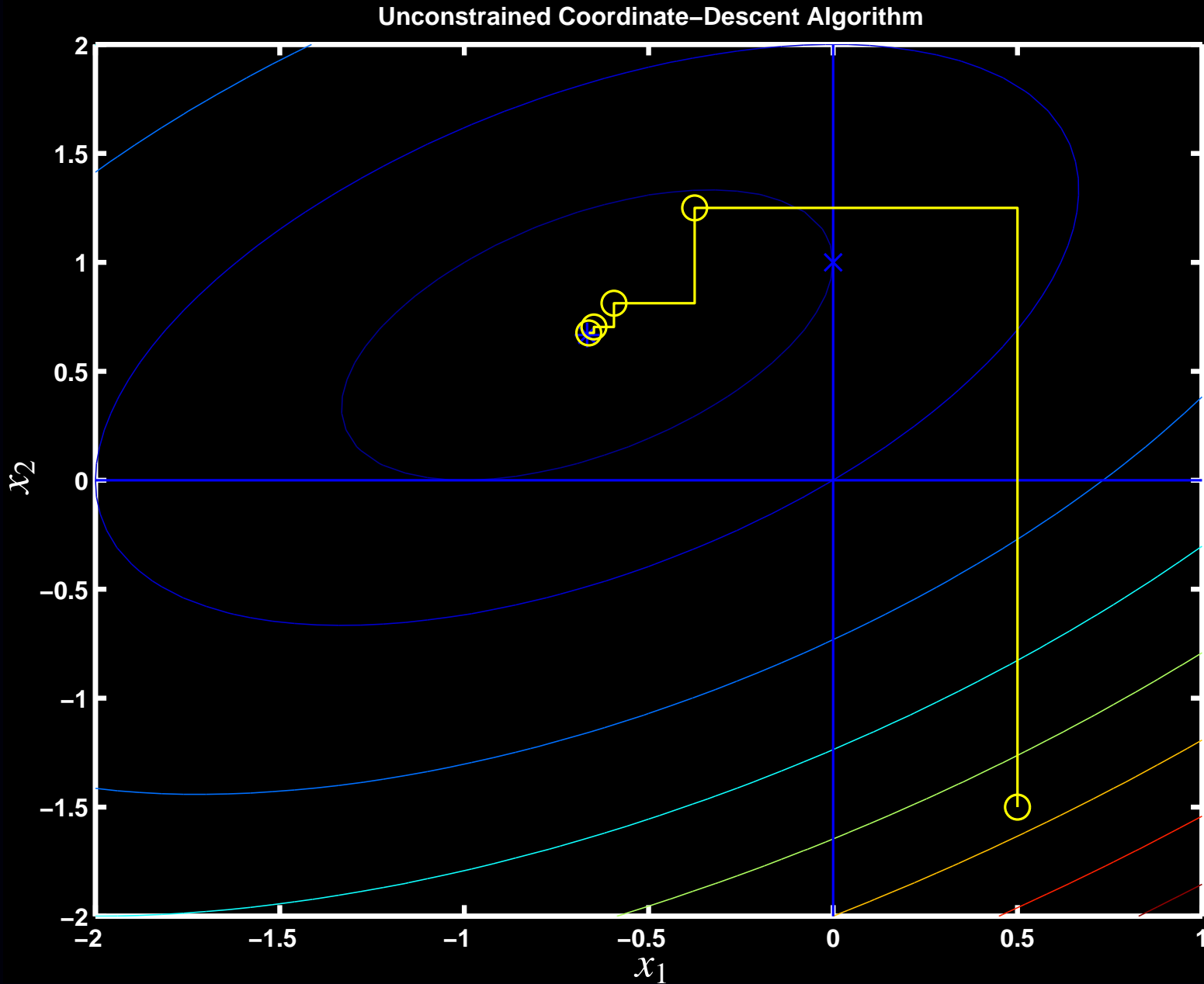
## Disadvantages:

- Requires column access of system matrix  $A$
- Cannot exploit some “tricks” for  $A$ , e.g., factorizations
- Expensive “arg min” for nonquadratic problems
- Poorly parallelizable

# Constrained Coordinate Descent Illustrated



# Coordinate Descent - Unconstrained



# Coordinate-Descent Algorithm Summary

Recommended when all of the following apply:

- quadratic or nearly-quadratic convex cost function
- nonnegativity constraint desired
- precomputed and stored system matrix  $A$  with column access
- parallelization not needed (standard workstation)

Cautions:

- Good initialization (e.g., properly scaled FBP) essential.  
(Uniform image or zero image cause slow initial convergence.)
- Must be programmed carefully to be efficient.  
(Standard Gauss-Siedel implementation is suboptimal.)
- Updates high-frequencies fastest  $\implies$  poorly suited to unregularized case

Used daily in UM clinic for 2D SPECT / PWLS / nonuniform attenuation

Under investigation for 3D helical CT reconstruction by Thibault *et al.*

# Summary of General-Purpose Algorithms

## Gradient-based

- Fully parallelizable
- Inconvenient line-searches for nonquadratic cost functions
- Fast converging in unconstrained case
- Nonnegativity constraint inconvenient

## Coordinate-descent

- Very fast converging
- Nonnegativity constraint trivial
- Poorly parallelizable
- Requires precomputed/stored system matrix

CD is well-suited to moderate-sized 2D problem (e.g., 2D PET), but challenging for large 2D problems (X-ray CT) and fully 3D problems

Neither is ideal.

∴ need *special-purpose algorithms* for image reconstruction!

# Data-Mismatch Functions Revisited

For fast converging, intrinsically monotone algorithms, consider the form of  $\Psi$ .

**WLS:**

$$\mathbf{t}(\mathbf{x}) = \sum_{i=1}^{n_d} \frac{1}{2} w_i (y_i - [\mathbf{Ax}]_i)^2 = \sum_{i=1}^{n_d} h_i([\mathbf{Ax}]_i), \quad \text{where } h_i(l) \triangleq \frac{1}{2} w_i (y_i - l)^2.$$

**Emission Poisson (negative) log-likelihood:**

$$\mathbf{t}(\mathbf{x}) = \sum_{i=1}^{n_d} ([\mathbf{Ax}]_i + r_i) - y_i \log([\mathbf{Ax}]_i + r_i) = \sum_{i=1}^{n_d} h_i([\mathbf{Ax}]_i)$$

where  $h_i(l) \triangleq (l + r_i) - y_i \log(l + r_i)$ .

**Transmission Poisson log-likelihood:**

$$\mathbf{t}(\mathbf{x}) = \sum_{i=1}^{n_d} \left( b_i e^{-[\mathbf{Ax}]_i} + r_i \right) - y_i \log \left( b_i e^{-[\mathbf{Ax}]_i} + r_i \right) = \sum_{i=1}^{n_d} h_i([\mathbf{Ax}]_i)$$

where  $h_i(l) \triangleq (b_i e^{-l} + r_i) - y_i \log(b_i e^{-l} + r_i)$ .

MRI, polyenergetic X-ray CT, confocal microscopy, image restoration, ...  
All have same *partially separable* form.

# General Imaging Cost Function

General form for data-mismatch function:

$$\ell(\mathbf{x}) = \sum_{i=1}^{n_d} h_i([\mathbf{A}\mathbf{x}]_i)$$

General form for regularizing penalty function:

$$R(\mathbf{x}) = \sum_k \psi_k([\mathbf{C}\mathbf{x}]_k)$$

General form for cost function:

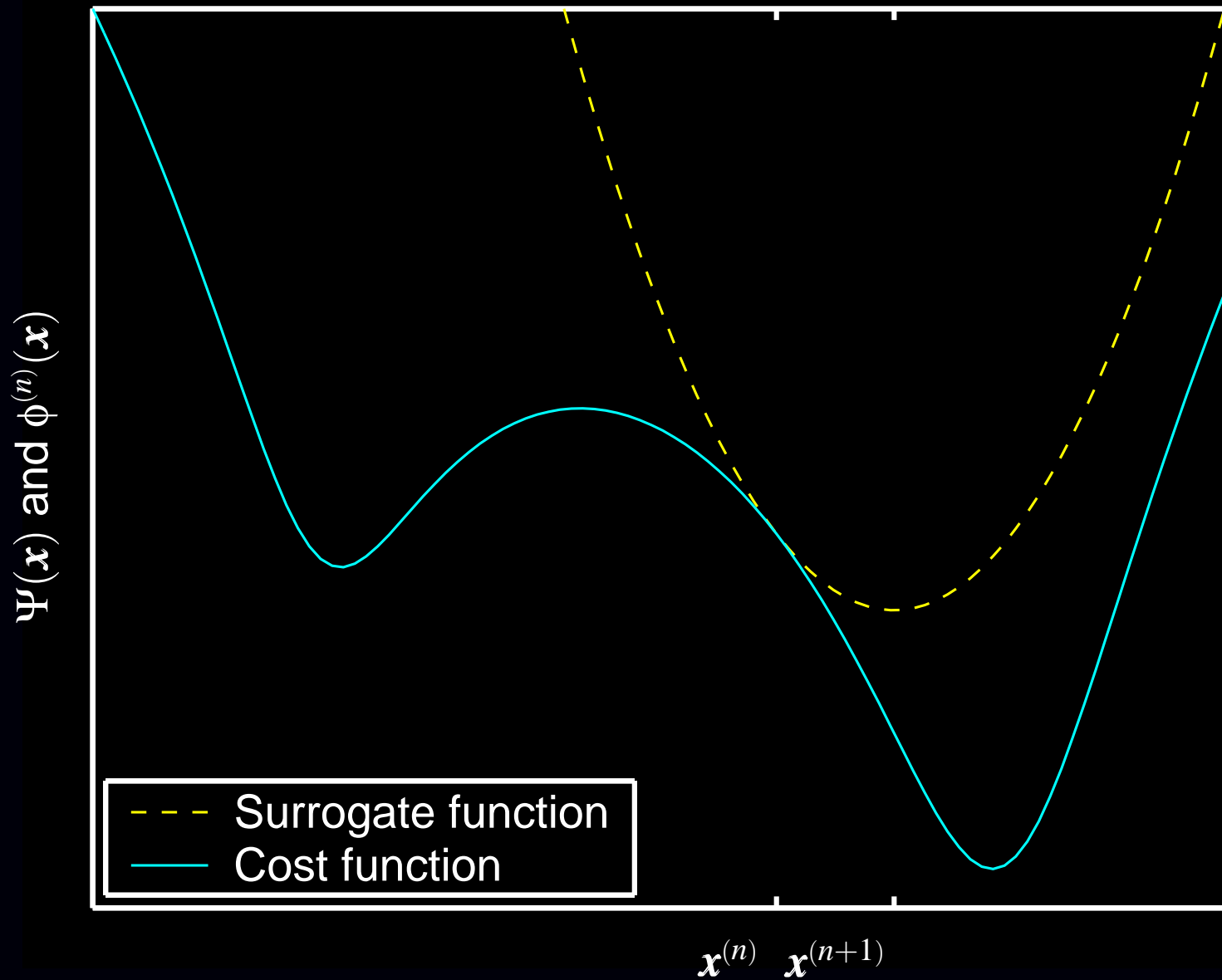
$$\Psi(\mathbf{x}) = \ell(\mathbf{x}) + \beta R(\mathbf{x}) = \sum_{i=1}^{n_d} h_i([\mathbf{A}\mathbf{x}]_i) + \beta \sum_k \psi_k([\mathbf{C}\mathbf{x}]_k)$$

Properties of  $\Psi$  we can exploit:

- summation form (due to independence of measurements)
- convexity of  $h_i$  functions (usually)
- summation argument (inner product of  $\mathbf{x}$  with  $i$ th row of  $\mathbf{A}$ )

Most methods that use these properties are forms of *optimization transfer*.

# Optimization Transfer Illustrated





# Optimization Transfer

General iteration:

$$\mathbf{x}^{(n+1)} = \arg \min_{\mathbf{x} \geq \mathbf{0}} \phi(\mathbf{x}; \mathbf{x}^{(n)})$$

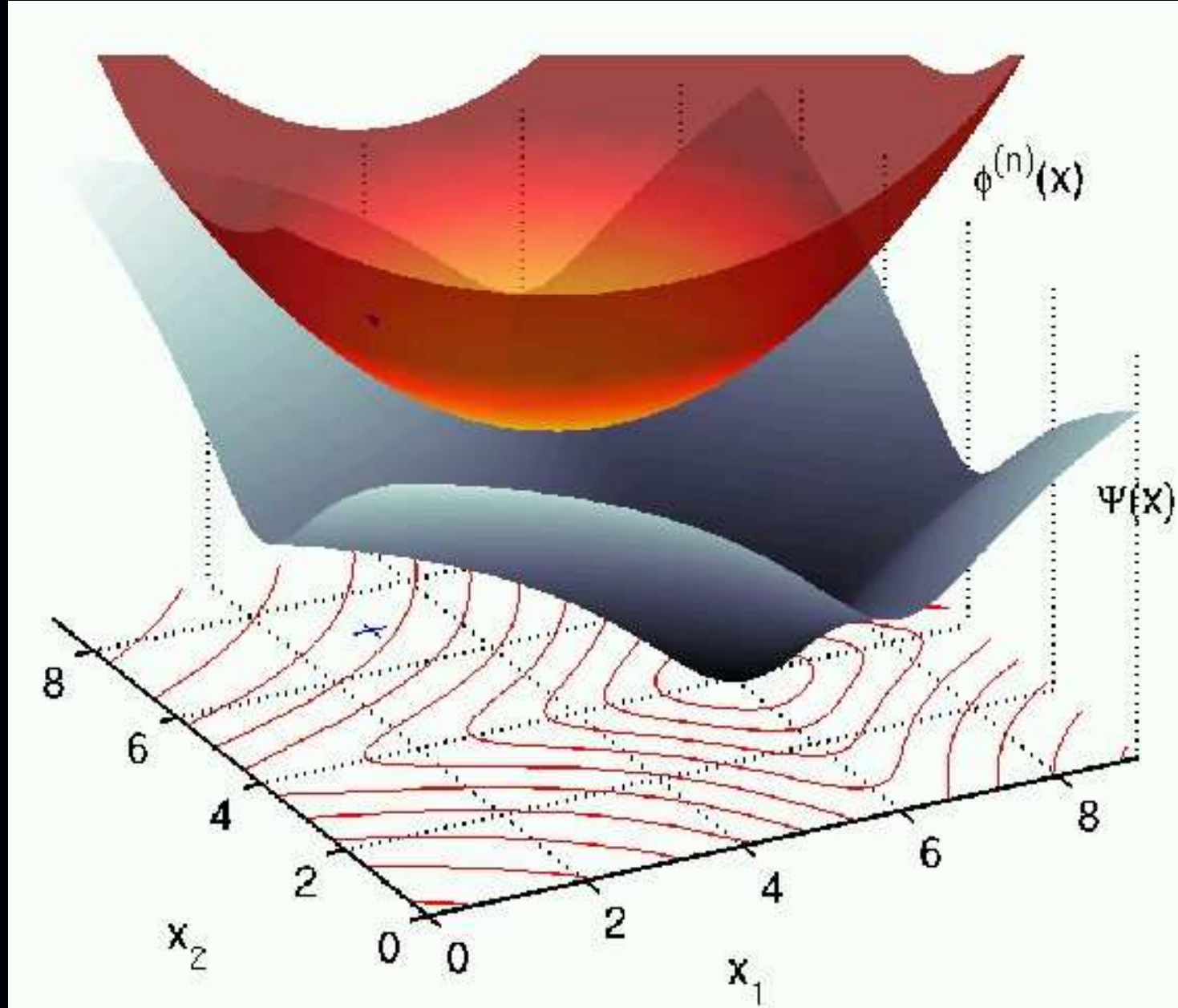
Monotonicity conditions (cost function  $\Psi$  decreases provided these hold):

- $\phi(\mathbf{x}^{(n)}; \mathbf{x}^{(n)}) = \Psi(\mathbf{x}^{(n)})$  (matched current value)
- $\nabla_{\mathbf{x}} \phi(\mathbf{x}; \mathbf{x}^{(n)}) \Big|_{\mathbf{x}=\mathbf{x}^{(n)}} = \nabla \Psi(\mathbf{x}) \Big|_{\mathbf{x}=\mathbf{x}^{(n)}}$  (matched gradient)
- $\phi(\mathbf{x}; \mathbf{x}^{(n)}) \geq \Psi(\mathbf{x}) \quad \forall \mathbf{x} \geq \mathbf{0}$  (lies above)

These 3 (sufficient) conditions are satisfied by the  $Q$  function of the EM algorithm (and its relatives like SAGE).

The 3rd condition is *not* satisfied by the Newton-Raphson quadratic approximation, which leads to its nonmonotonicity.

# Optimization Transfer in 2d



# Optimization Transfer of EM Algorithm

E-step: choose surrogate function  $\phi(\mathbf{x}; \mathbf{x}^{(n)})$

M-step: minimize surrogate function

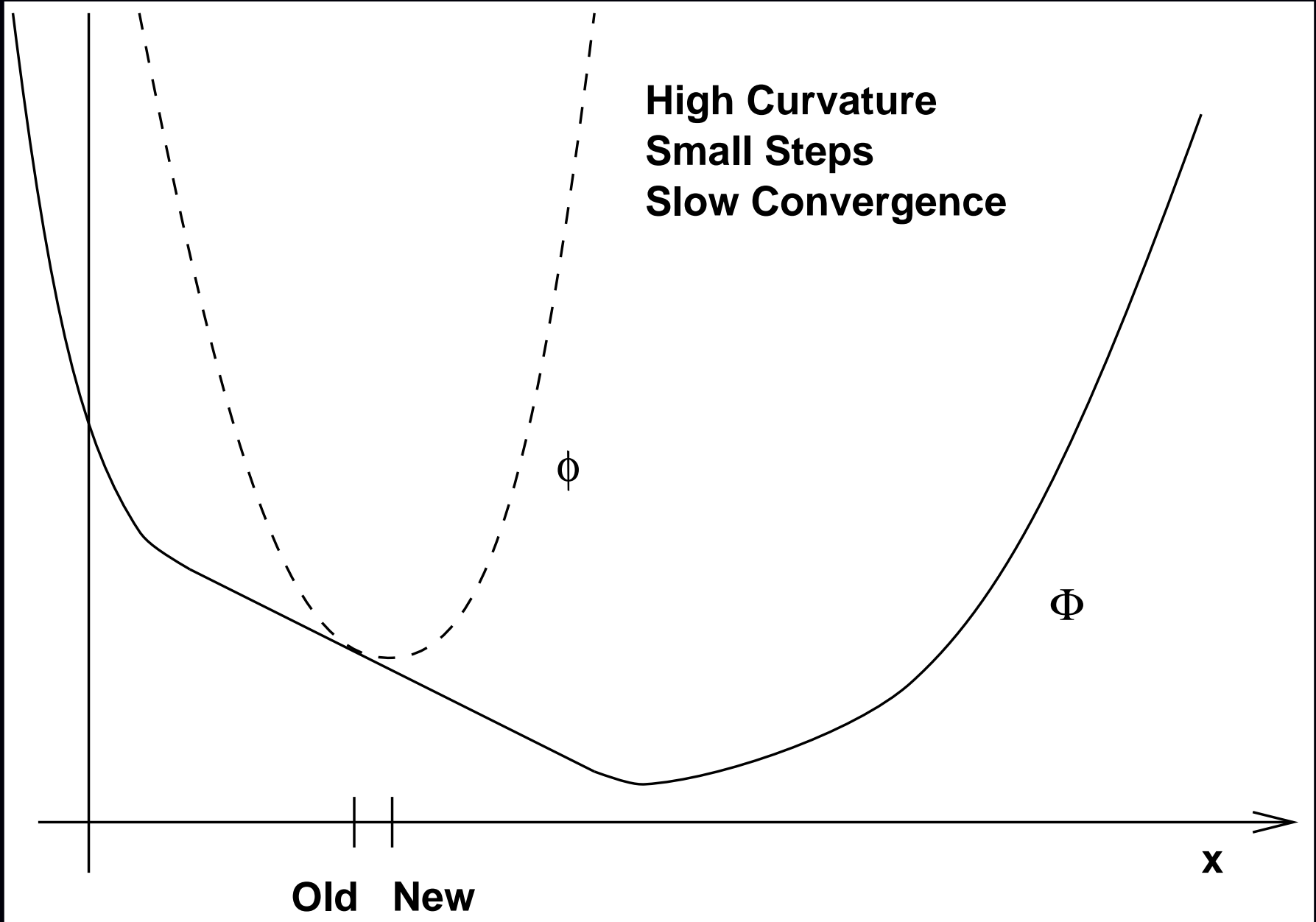
$$\mathbf{x}^{(n+1)} = \arg \min_{\mathbf{x} \geq \mathbf{0}} \phi(\mathbf{x}; \mathbf{x}^{(n)})$$

Designing surrogate functions

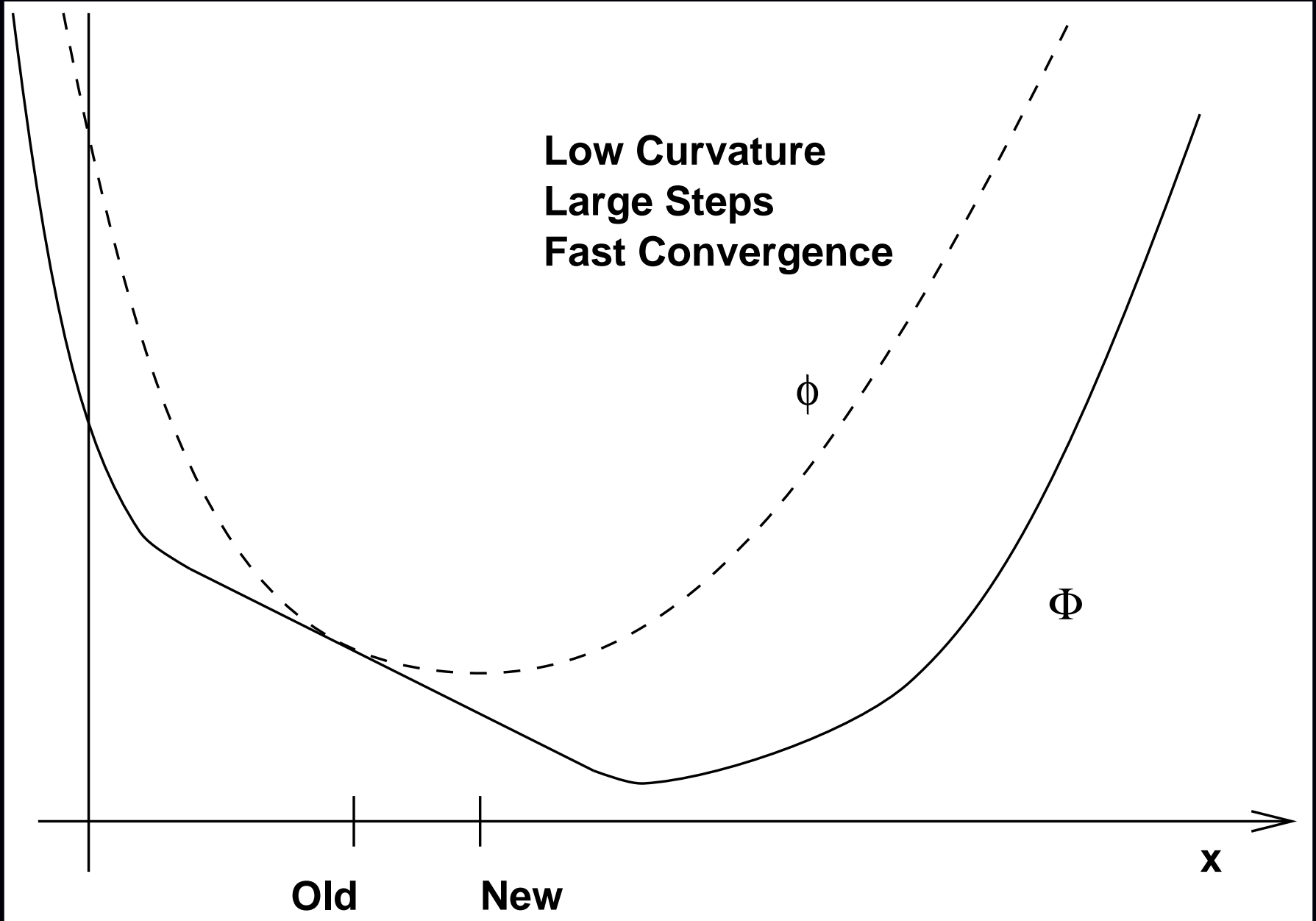
- Easy to “compute”
- Easy to minimize
- Fast convergence rate

Often mutually incompatible goals  $\therefore$  compromises

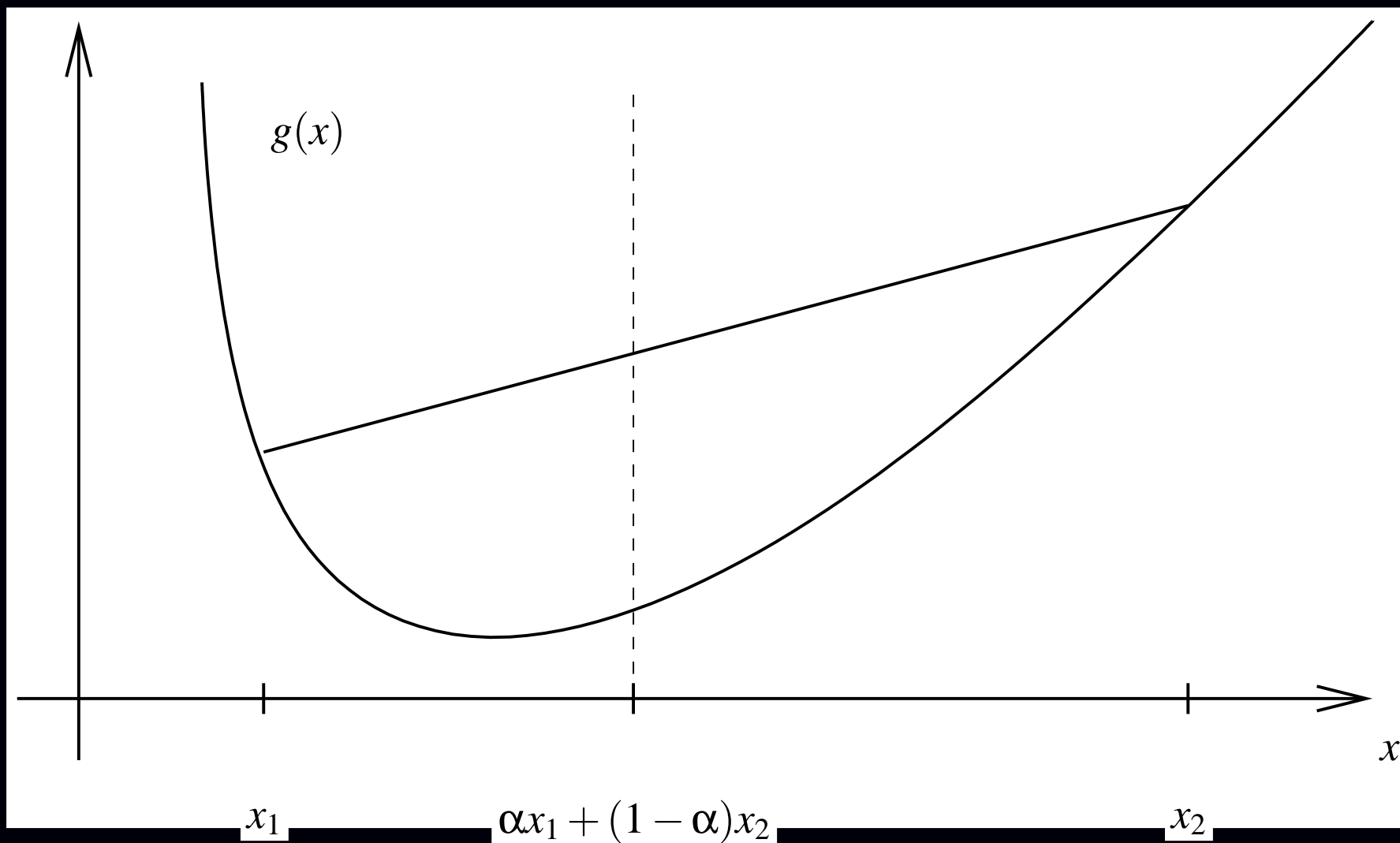
# Convergence Rate: Slow



# Convergence Rate: Fast



## Tool: Convexity Inequality



$g$  convex  $\implies g(\alpha x_1 + (1 - \alpha)x_2) \leq \alpha g(x_1) + (1 - \alpha)g(x_2)$  for  $\alpha \in [0, 1]$

More generally:  $\alpha_k \geq 0$  and  $\sum_k \alpha_k = 1 \implies g(\sum_k \alpha_k x_k) \leq \sum_k \alpha_k g(x_k)$ . Sum outside!

## Example 1: Classical ML-EM Algorithm

Negative Poisson log-likelihood cost function (unregularized):

$$\Psi(\mathbf{x}) = \sum_{i=1}^{n_d} h_i([\mathbf{Ax}]_i), \quad h_i(l) = (l + r_i) - y_i \log(l + r_i).$$

Intractable to minimize directly due to summation within logarithm.

Clever trick due to De Pierro (let  $\bar{y}_i^{(n)} = [\mathbf{Ax}^{(n)}]_i + r_i$ ):

$$[\mathbf{Ax}]_i = \sum_{j=1}^{n_p} a_{ij}x_j = \sum_{j=1}^{n_p} \left[ \frac{a_{ij}x_j^{(n)}}{\bar{y}_i^{(n)}} \right] \left( \frac{x_j^{(n)}}{x_j^{(n)}} \bar{y}_i^{(n)} \right).$$

Since the  $h_i$ 's are **convex** in Poisson emission model:

$$h_i([\mathbf{Ax}]_i) = h_i \left( \sum_{j=1}^{n_p} \left[ \frac{a_{ij}x_j^{(n)}}{\bar{y}_i^{(n)}} \right] \left( \frac{x_j^{(n)}}{x_j^{(n)}} \bar{y}_i^{(n)} \right) \right) \leq \sum_{j=1}^{n_p} \left[ \frac{a_{ij}x_j^{(n)}}{\bar{y}_i^{(n)}} \right] h_i \left( \frac{x_j^{(n)}}{x_j^{(n)}} \bar{y}_i^{(n)} \right)$$

$$\Psi(\mathbf{x}) = \sum_{i=1}^{n_d} h_i([\mathbf{Ax}]_i) \leq \phi(\mathbf{x}; \mathbf{x}^{(n)}) \triangleq \sum_{i=1}^{n_d} \sum_{j=1}^{n_p} \left[ \frac{a_{ij}x_j^{(n)}}{\bar{y}_i^{(n)}} \right] h_i \left( \frac{x_j^{(n)}}{x_j^{(n)}} \bar{y}_i^{(n)} \right)$$

Replace convex cost function  $\Psi(\mathbf{x})$  with **separable** surrogate function  $\phi(\mathbf{x}; \mathbf{x}^{(n)})$ .

## “ML-EM Algorithm” M-step

E-step gave separable surrogate function:

$$\phi(\mathbf{x}; \mathbf{x}^{(n)}) = \sum_{j=1}^{n_p} \phi_j(x_j; \mathbf{x}^{(n)}), \quad \text{where} \quad \phi_j(x_j; \mathbf{x}^{(n)}) \triangleq \sum_{i=1}^{n_d} \left[ \frac{a_{ij} x_j^{(n)}}{\bar{y}_i^{(n)}} \right] h_i \left( \frac{x_j}{x_j^{(n)}} \bar{y}_i^{(n)} \right).$$

M-step separates:

$$\mathbf{x}^{(n+1)} = \arg \min_{\mathbf{x} \geq \mathbf{0}} \phi(\mathbf{x}; \mathbf{x}^{(n)}) \implies x_j^{(n+1)} = \arg \min_{x_j \geq 0} \phi_j(x_j; \mathbf{x}^{(n)}), \quad j = 1, \dots, n_p$$

Minimizing:

$$\frac{\partial}{\partial x_j} \phi_j(x_j; \mathbf{x}^{(n)}) = \sum_{i=1}^{n_d} a_{ij} h_i \left( \bar{y}_i^{(n)} x_j / x_j^{(n)} \right) = \sum_{i=1}^{n_d} a_{ij} \left[ 1 - \frac{y_i}{\bar{y}_i^{(n)} x_j / x_j^{(n)}} \right] \Big|_{x_j = x_j^{(n+1)}} = 0.$$

Solving (in case  $r_i = 0$ ):

$$x_j^{(n+1)} = x_j^{(n)} \left[ \sum_{i=1}^{n_d} a_{ij} \frac{y_i}{[\mathbf{A} \mathbf{x}^{(n)}]_i} \right] / \left( \sum_{i=1}^{n_d} a_{ij} \right), \quad j = 1, \dots, n_p$$

- Derived without any statistical considerations, unlike classical EM formulation.
- Uses only convexity and algebra.
- Guaranteed monotonic: surrogate function  $\phi$  satisfies the 3 required properties.
- M-step trivial due to *separable surrogate*.



## ML-EM is Scaled Gradient Descent

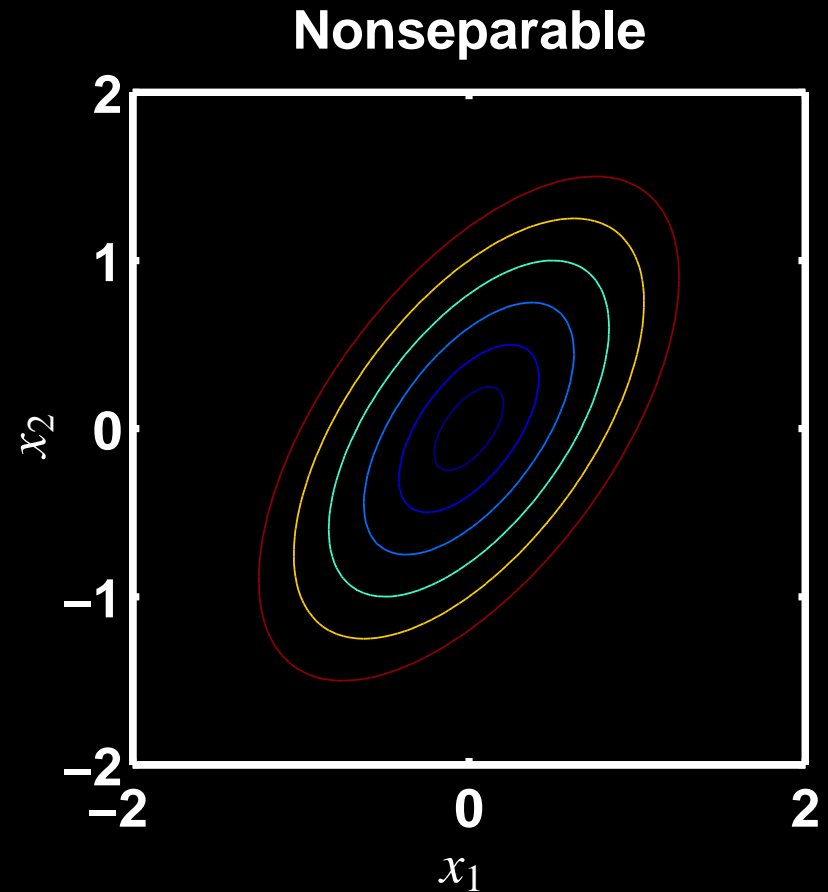
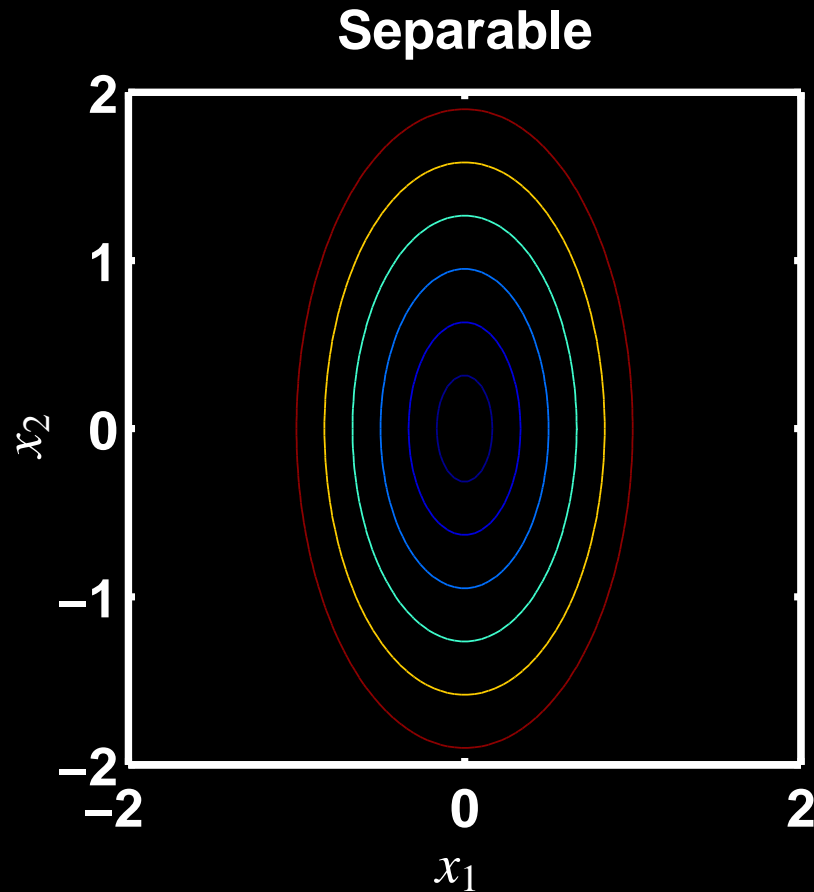
$$\begin{aligned}x_j^{(n+1)} &= x_j^{(n)} \left[ \sum_{i=1}^{n_d} a_{ij} \frac{y_i}{\bar{y}_i^{(n)}} \right] / \left( \sum_{i=1}^{n_d} a_{ij} \right) \\&= x_j^{(n)} + x_j^{(n)} \left[ \sum_{i=1}^{n_d} a_{ij} \left( \frac{y_i}{\bar{y}_i^{(n)}} - 1 \right) \right] / \left( \sum_{i=1}^{n_d} a_{ij} \right) \\&= \boxed{x_j^{(n)} - \left( \frac{x_j^{(n)}}{\sum_{i=1}^{n_d} a_{ij}} \right) \frac{\partial}{\partial x_j} \Psi(\mathbf{x}^{(n)})}, \quad j = 1, \dots, n_p\end{aligned}$$

$$\mathbf{x}^{(n+1)} = \mathbf{x}^{(n)} + \mathbf{D}(\mathbf{x}^{(n)}) \nabla \Psi(\mathbf{x}^{(n)})$$

This particular diagonal scaling matrix remarkably

- ensures monotonicity,
- ensures nonnegativity.

## Consideration: Separable vs Nonseparable



Contour plots: loci of equal function values.

Uncoupled vs coupled minimization.

## Separable Surrogate Functions (Easy M-step)

The preceding EM derivation structure applies to *any* cost function of the form

$$\Psi(\mathbf{x}) = \sum_{i=1}^{n_d} h_i([\mathbf{A}\mathbf{x}]_i).$$

cf ISRA (for nonnegative LS), “convex algorithm” for transmission reconstruction

Derivation yields a separable surrogate function

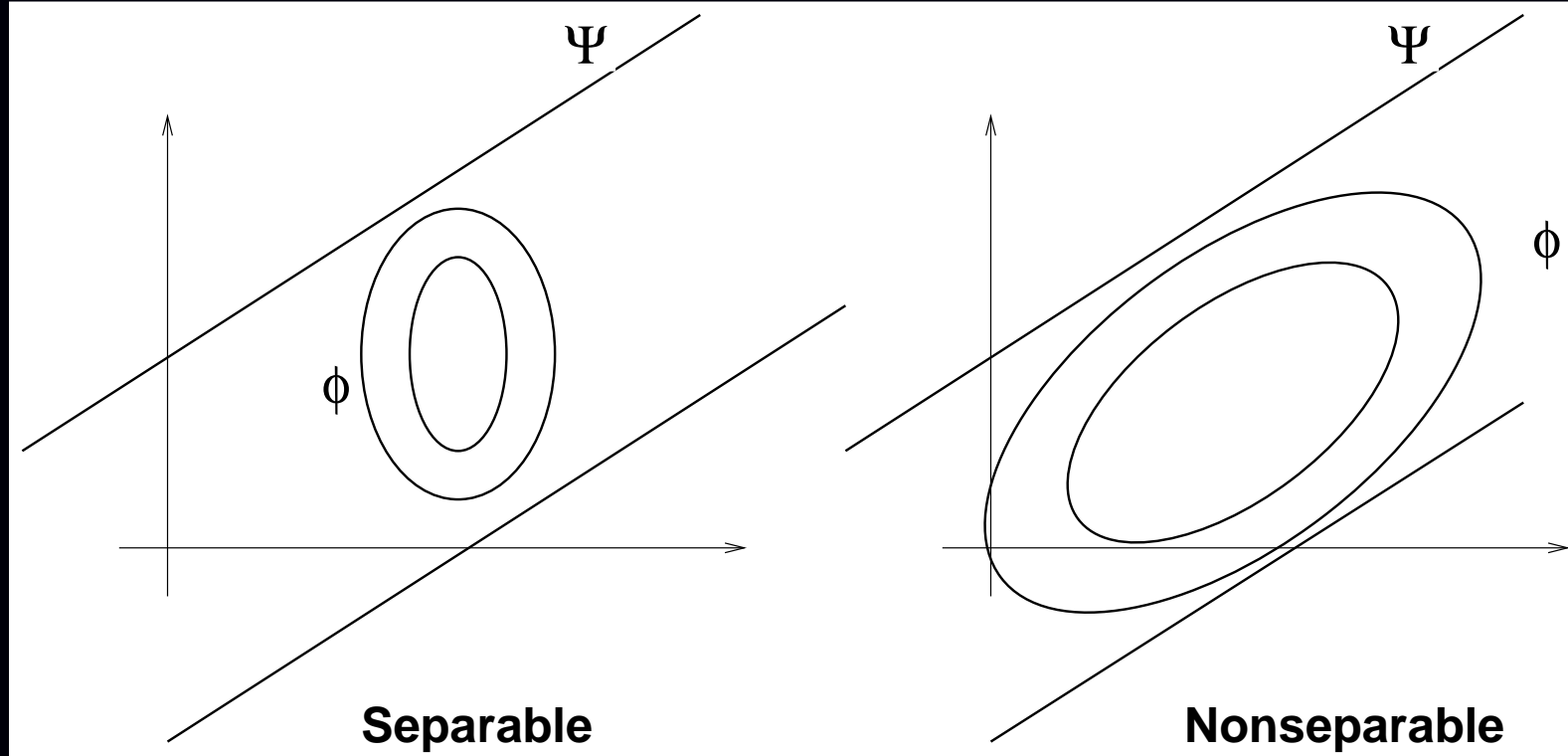
$$\Psi(\mathbf{x}) \leq \phi(\mathbf{x}; \mathbf{x}^{(n)}), \quad \text{where} \quad \phi(\mathbf{x}; \mathbf{x}^{(n)}) = \sum_{j=1}^{n_p} \phi_j(x_j; \mathbf{x}^{(n)})$$

M-step separates into 1D minimization problems (fully parallelizable):

$$\mathbf{x}^{(n+1)} = \arg \min_{\mathbf{x} \geq \mathbf{0}} \phi(\mathbf{x}; \mathbf{x}^{(n)}) \implies x_j^{(n+1)} = \arg \min_{x_j \geq 0} \phi_j(x_j; \mathbf{x}^{(n)}), \quad j = 1, \dots, n_p$$

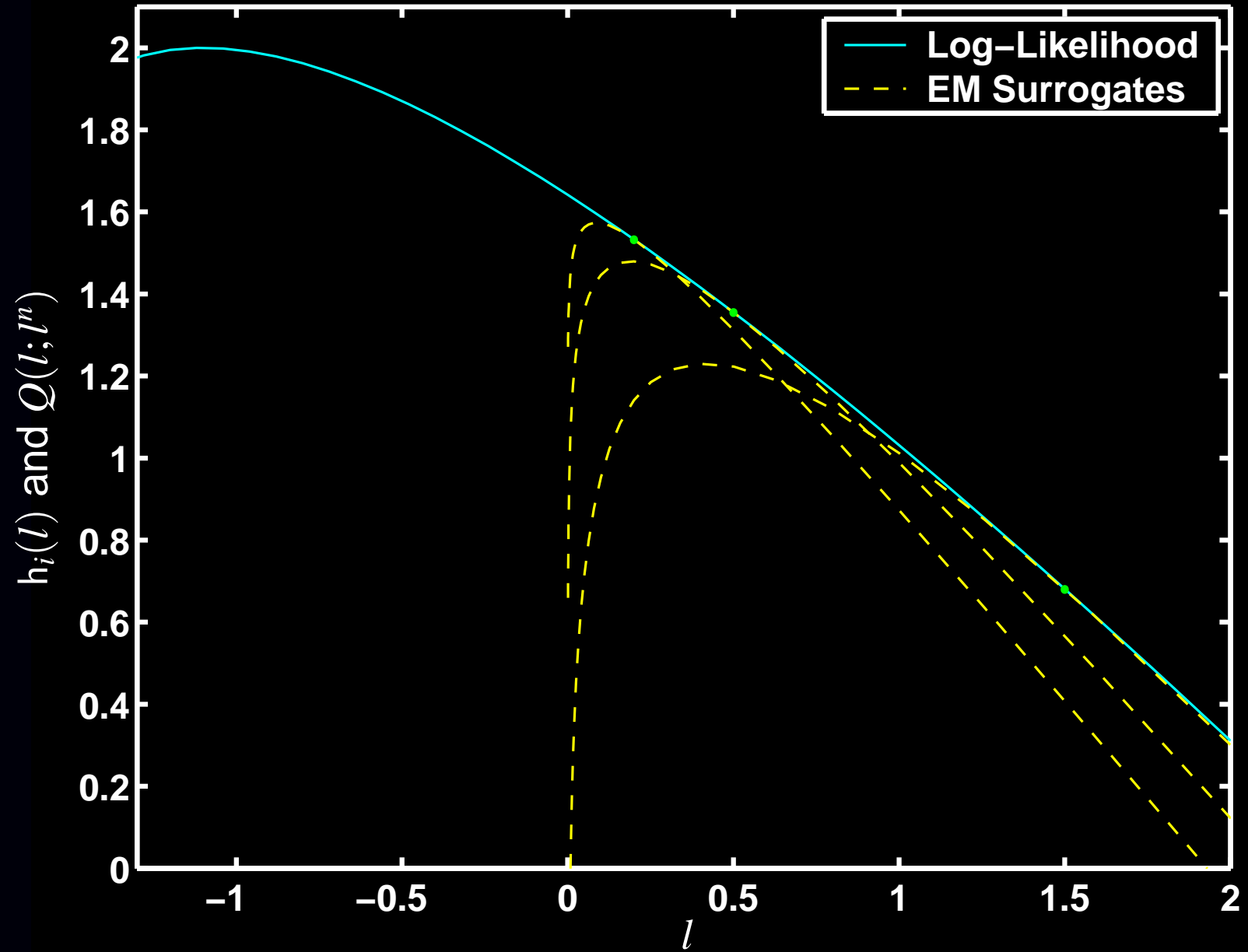
Why do EM / ISRA / convex-algorithm / etc. converge so slowly?

# Separable vs Nonseparable



Separable surrogates (e.g., EM) have high curvature  $\therefore$  slow convergence.  
Nonseparable surrogates can have lower curvature  $\therefore$  faster convergence.  
Harder to minimize? Use paraboloids (quadratic surrogates).

# High Curvature of EM Surrogate



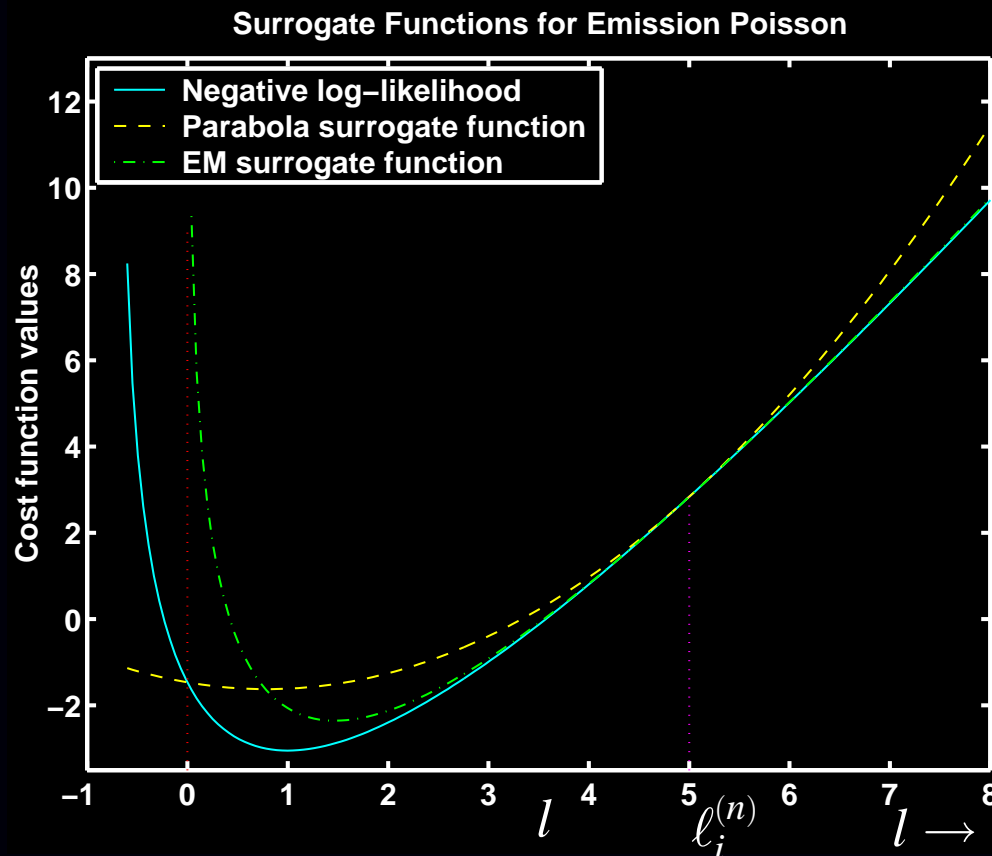
# 1D Parabola Surrogate Function

Find parabola  $q_i^{(n)}(l)$  of the form:

$$q_i^{(n)}(l) = h_i(\ell_i^{(n)}) + \dot{h}_i(\ell_i^{(n)})(l - \ell_i^{(n)}) + c_i^{(n)} \frac{1}{2}(l - \ell_i^{(n)})^2, \quad \text{where } \ell_i^{(n)} \triangleq [\mathbf{Ax}^{(n)}]_i$$

Satisfies tangent condition. Choose curvature to ensure “lies above” condition:

$$c_i^{(n)} \triangleq \min \left\{ c \geq 0 : q_i^{(n)}(l) \geq h_i(l), \quad \forall l \geq 0 \right\}.$$



Lower  
curvature!

# Paraboloidal Surrogate

Combining 1D parabola surrogates yields *paraboloidal surrogate*:

$$\Psi(\mathbf{x}) = \sum_{i=1}^{n_d} h_i([\mathbf{A}\mathbf{x}]_i) \leq \phi(\mathbf{x}; \mathbf{x}^{(n)}) = \sum_{i=1}^{n_d} q_i^{(n)}([\mathbf{A}\mathbf{x}]_i)$$

$$\text{Rewriting: } \phi(\boldsymbol{\delta} + \mathbf{x}^{(n)}; \mathbf{x}^{(n)}) = \Psi(\mathbf{x}^{(n)}) + \nabla \Psi(\mathbf{x}^{(n)}) \boldsymbol{\delta} + \frac{1}{2} \boldsymbol{\delta}' \mathbf{A}' \text{diag}\{c_i^{(n)}\} \mathbf{A} \boldsymbol{\delta}$$

## Advantages

- Surrogate  $\phi(\mathbf{x}; \mathbf{x}^{(n)})$  is *quadratic*, unlike Poisson log-likelihood  
⇒ easier to minimize
- Not separable (unlike EM surrogate)
- Not self-similar (unlike EM surrogate)
- Small curvatures ⇒ fast convergence
- Intrinsically monotone global convergence
- Fairly simple to derive / implement

## Quadratic minimization

- Coordinate descent
  - + fast converging
  - + Nonnegativity easy
  - precomputed column-stored system matrix
- Gradient-based quadratic minimization methods
  - Nonnegativity inconvenient

## Example: PSCD for PET Transmission Scans

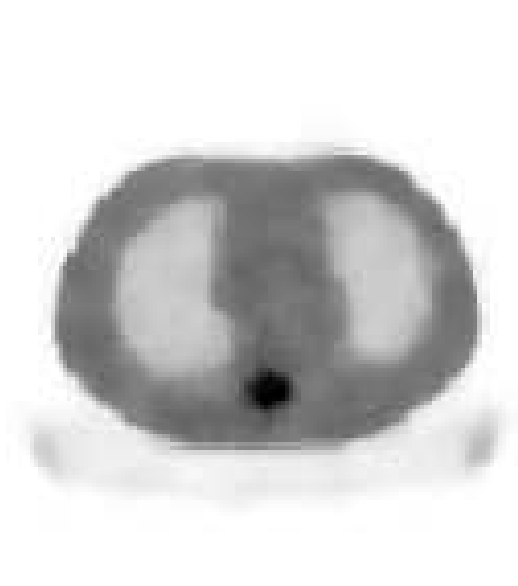
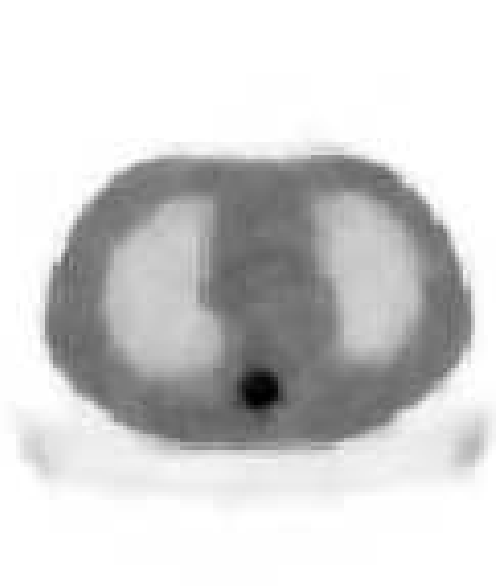
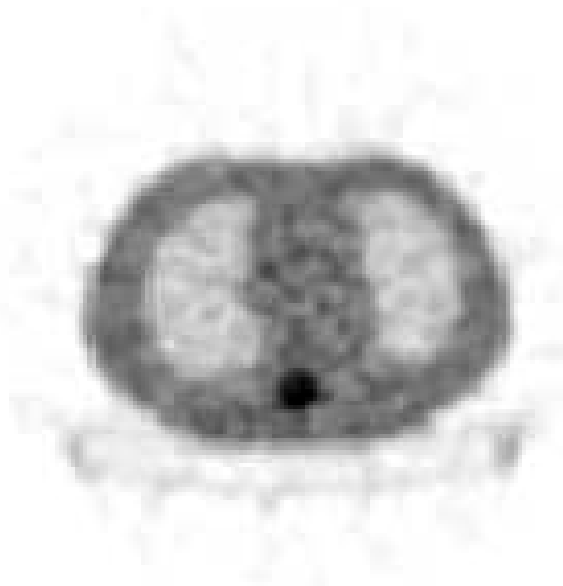
**FBP**

**PL-OSTR-16**

**PL-PSCD**

**4 iterations**

**10 iterations**



- square-pixel basis
- strip-integral system model
- shifted-Poisson statistical model
- edge-preserving convex regularization (Huber)
- nonnegativity constraint
- inscribed circle support constraint
- paraboloidal surrogate coordinate descent (PSCD) algorithm



## Separable Paraboloidal Surrogate

To derive a parallelizable algorithm apply another De Pierro trick:

$$[\mathbf{Ax}]_i = \sum_{j=1}^{n_p} \pi_{ij} \left[ \frac{a_{ij}}{\pi_{ij}} (x_j - x_j^{(n)}) + \ell_i^{(n)} \right], \quad \ell_i^{(n)} = [\mathbf{Ax}^{(n)}]_i.$$

Provided  $\pi_{ij} \geq 0$  and  $\sum_{j=1}^{n_p} \pi_{ij} = 1$ , since parabola  $q_i$  is convex:

$$q_i^{(n)}([\mathbf{Ax}]_i) = q_i^{(n)} \left( \sum_{j=1}^{n_p} \pi_{ij} \left[ \frac{a_{ij}}{\pi_{ij}} (x_j - x_j^{(n)}) + \ell_i^{(n)} \right] \right) \leq \sum_{j=1}^{n_p} \pi_{ij} q_i^{(n)} \left( \frac{a_{ij}}{\pi_{ij}} (x_j - x_j^{(n)}) + \ell_i^{(n)} \right)$$

$$\therefore \phi(\mathbf{x}; \mathbf{x}^{(n)}) = \sum_{i=1}^{n_d} q_i^{(n)}([\mathbf{Ax}]_i) \leq \tilde{\phi}(\mathbf{x}; \mathbf{x}^{(n)}) \triangleq \sum_{i=1}^{n_d} \sum_{j=1}^{n_p} \pi_{ij} q_i^{(n)} \left( \frac{a_{ij}}{\pi_{ij}} (x_j - x_j^{(n)}) + \ell_i^{(n)} \right)$$

**Separable Paraboloidal Surrogate:**

$$\tilde{\phi}(\mathbf{x}; \mathbf{x}^{(n)}) = \sum_{j=1}^{n_p} \phi_j(x_j; \mathbf{x}^{(n)}), \quad \phi_j(x_j; \mathbf{x}^{(n)}) \triangleq \sum_{i=1}^{n_d} \pi_{ij} q_i^{(n)} \left( \frac{a_{ij}}{\pi_{ij}} (x_j - x_j^{(n)}) + \ell_i^{(n)} \right)$$

Parallelizable M-step (cf gradient descent!):

$$x_j^{(n+1)} = \arg \min_{x_j \geq 0} \phi_j(x_j; \mathbf{x}^{(n)}) = \left[ x_j^{(n)} - \frac{1}{d_j^{(n)}} \frac{\partial}{\partial x_j} \Psi(\mathbf{x}^{(n)}) \right]_+, \quad d_j^{(n)} = \sum_{i=1}^{n_d} \frac{a_{ij}^2}{\pi_{ij}} c_i^{(n)}$$

Natural choice is  $\pi_{ij} = |a_{ij}|/|a|_i$ ,  $|a|_i = \sum_{j=1}^{n_p} |a_{ij}|$

## Example: Poisson ML Transmission Problem

Transmission negative log-likelihood (for  $i$ th ray):

$$h_i(l) = (b_i e^{-l} + r_i) - y_i \log(b_i e^{-l} + r_i).$$

Optimal (smallest) parabola surrogate curvature (Erdođan, T-MI, Sep. 1999):

$$c_i^{(n)} = c(\ell_i^{(n)}, h_i), \quad c(l, h) = \begin{cases} \left[ 2 \frac{h(0) - h(l) + \dot{h}(l)l}{l^2} \right]_+, & l > 0 \\ [\ddot{h}(l)]_+, & l = 0. \end{cases}$$

### Separable Paraboloidal Surrogate (SPS) Algorithm:

Precompute  $|a|_i = \sum_{j=1}^{n_p} a_{ij}$ ,  $i = 1, \dots, n_d$

$$\ell_i^{(n)} = [\mathbf{A}\mathbf{x}^{(n)}]_i, \quad (\text{forward projection})$$

$$\bar{y}_i^{(n)} = b_i e^{-\ell_i^{(n)}} + r_i \quad (\text{predicted means})$$

$$\dot{h}_i^{(n)} = 1 - y_i / \bar{y}_i^{(n)} \quad (\text{slopes})$$

$$c_i^{(n)} = c(\ell_i^{(n)}, h_i) \quad (\text{curvatures})$$

$$x_j^{(n+1)} = \left[ x_j^{(n)} - \frac{1}{d_j^{(n)}} \frac{\partial}{\partial x_j} \Psi(\mathbf{x}^{(n)}) \right]_+ = \left[ x_j^{(n)} - \frac{\sum_{i=1}^{n_d} a_{ij} \dot{h}_i^{(n)}}{\sum_{i=1}^{n_d} a_{ij} |a|_i c_i^{(n)}} \right]_+, \quad j = 1, \dots, n_p$$

Monotonically decreases cost function each iteration.

**No logarithm!**

# The MAP-EM M-step “Problem”

Add a penalty function to our surrogate for the negative log-likelihood:

$$\Psi(\mathbf{x}) = \ell(\mathbf{x}) + \beta R(\mathbf{x})$$
$$\phi(\mathbf{x}; \mathbf{x}^{(n)}) = \sum_{j=1}^{n_p} \phi_j(x_j; \mathbf{x}^{(n)}) + \beta R(\mathbf{x})$$

$$\text{M-step: } \mathbf{x}^{(n+1)} = \arg \min_{\mathbf{x} \geq \mathbf{0}} \phi(\mathbf{x}; \mathbf{x}^{(n)}) = \arg \min_{\mathbf{x} \geq \mathbf{0}} \sum_{j=1}^{n_p} \phi_j(x_j; \mathbf{x}^{(n)}) + \beta R(\mathbf{x}) = ?$$

For nonseparable penalty functions, the M-step is coupled  $\therefore$  difficult.

## Suboptimal solutions

- Generalized EM (GEM) algorithm (coordinate descent on  $\phi$ )  
Monotonic, but inherits slow convergence of EM.
- One-step late (OSL) algorithm (use outdated gradients) (Green, T-MI, 1990)

$$\frac{\partial}{\partial x_j} \phi(\mathbf{x}; \mathbf{x}^{(n)}) = \frac{\partial}{\partial x_j} \phi_j(x_j; \mathbf{x}^{(n)}) + \beta \frac{\partial}{\partial x_j} R(\mathbf{x}) \stackrel{?}{\approx} \frac{\partial}{\partial x_j} \phi_j(x_j; \mathbf{x}^{(n)}) + \beta \frac{\partial}{\partial x_j} R(\mathbf{x}^{(n)})$$

Nonmonotonic. Known to diverge, depending on  $\beta$ .

Temporarily simple, but *avoid!*

## Contemporary solution

- Use separable surrogate for penalty function too (De Pierro, T-MI, Dec. 1995)  
Ensures monotonicity. Obviates all reasons for using OSL!

## De Pierro's MAP-EM Algorithm

Apply separable paraboloidal surrogates to penalty function:

$$R(\mathbf{x}) \leq R_{\text{SPS}}(\mathbf{x}; \mathbf{x}^{(n)}) = \sum_{j=1}^{n_p} R_j(x_j; \mathbf{x}^{(n)})$$

$$\text{Overall separable surrogate: } \phi(\mathbf{x}; \mathbf{x}^{(n)}) = \sum_{j=1}^{n_p} \phi_j(x_j; \mathbf{x}^{(n)}) + \beta \sum_{j=1}^{n_p} R_j(x_j; \mathbf{x}^{(n)})$$

The M-step becomes fully parallelizable:

$$x_j^{(n+1)} = \arg \min_{x_j \geq 0} \phi_j(x_j; \mathbf{x}^{(n)}) - \beta R_j(x_j; \mathbf{x}^{(n)}), \quad j = 1, \dots, n_p.$$

Consider quadratic penalty  $R(\mathbf{x}) = \sum_k \psi([\mathbf{C}\mathbf{x}]_k)$ , where  $\psi(t) = t^2/2$ .

If  $\gamma_{kj} \geq 0$  and  $\sum_{j=1}^{n_p} \gamma_{kj} = 1$  then

$$[\mathbf{C}\mathbf{x}]_k = \sum_{j=1}^{n_p} \gamma_{kj} \left[ \frac{c_{kj}}{\gamma_{kj}} (x_j - x_j^{(n)}) + [\mathbf{C}\mathbf{x}^{(n)}]_k \right].$$

Since  $\psi$  is convex:

$$\begin{aligned} \psi([\mathbf{C}\mathbf{x}]_k) &= \psi \left( \sum_{j=1}^{n_p} \gamma_{kj} \left[ \frac{c_{kj}}{\gamma_{kj}} (x_j - x_j^{(n)}) + [\mathbf{C}\mathbf{x}^{(n)}]_k \right] \right) \\ &\leq \sum_{j=1}^{n_p} \gamma_{kj} \psi \left( \frac{c_{kj}}{\gamma_{kj}} (x_j - x_j^{(n)}) + [\mathbf{C}\mathbf{x}^{(n)}]_k \right) \end{aligned}$$

## De Pierro's Algorithm Continued

So  $R(\mathbf{x}) \leq R(\mathbf{x}; \mathbf{x}^{(n)}) \triangleq \sum_{j=1}^{n_p} R_j(x_j; \mathbf{x}^{(n)})$  where

$$R_j(x_j; \mathbf{x}^{(n)}) \triangleq \sum_k \gamma_{kj} \Psi \left( \frac{c_{kj}}{\gamma_{kj}} (x_j - x_j^{(n)}) + [\mathbf{C}\mathbf{x}^{(n)}]_k \right)$$

M-step: Minimizing  $\phi_j(x_j; \mathbf{x}^{(n)}) + \beta R_j(x_j; \mathbf{x}^{(n)})$  yields the iteration:

$$x_j^{(n+1)} = \frac{x_j^{(n)} \sum_{i=1}^{n_d} a_{ij} y_i / \bar{y}_i^{(n)}}{B_j + \sqrt{B_j^2 + \left( x_j^{(n)} \sum_{i=1}^{n_d} a_{ij} y_i / \bar{y}_i^{(n)} \right) \left( \beta \sum_k c_{kj}^2 / \gamma_{kj} \right)}}$$

$$\text{where } B_j \triangleq \frac{1}{2} \left[ \sum_{i=1}^{n_d} a_{ij} + \beta \sum_k \left( c_{kj} [\mathbf{C}\mathbf{x}^{(n)}]_k - \frac{c_{kj}^2}{\gamma_{kj}} x_j^{(n)} \right) \right], \quad j = 1, \dots, n_p$$

and  $\bar{y}_i^{(n)} = [\mathbf{A}\mathbf{x}^{(n)}]_i + r_i$ .

Advantages: Intrinsically monotone, nonnegativity, fully parallelizable.  
Requires only a couple % more computation per iteration than ML-EM

Disadvantages: Slow convergence (like EM) due to separable surrogate

# Ordered Subsets Algorithms

aka *block iterative* or *incremental gradient* algorithms

The gradient appears in essentially every algorithm:

$$\ell(\mathbf{x}) = \sum_{i=1}^{n_d} h_i([\mathbf{A}\mathbf{x}]_i) \implies \frac{\partial}{\partial x_j} \ell(\mathbf{x}) = \sum_{i=1}^{n_d} a_{ij} \dot{h}_i([\mathbf{A}\mathbf{x}]_i).$$

This is a *backprojection* of a sinogram of the derivatives  $\{\dot{h}_i([\mathbf{A}\mathbf{x}]_i)\}$ .

Intuition: with half the angular sampling, this backprojection would be fairly similar

$$\frac{1}{n_d} \sum_{i=1}^{n_d} a_{ij} \dot{h}_i(\cdot) \approx \frac{1}{|\mathcal{S}|} \sum_{i \in \mathcal{S}} a_{ij} \dot{h}_i(\cdot),$$

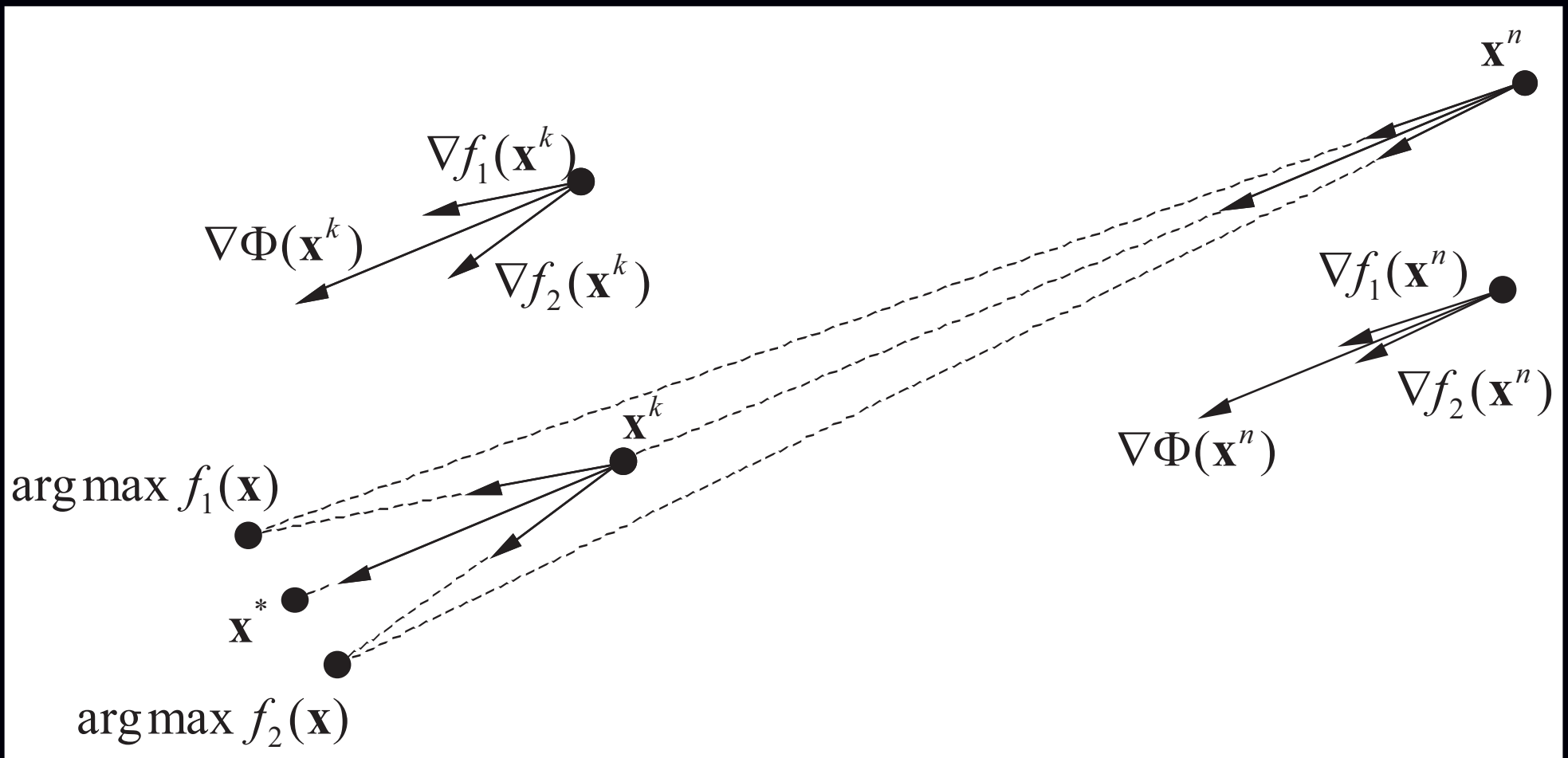
where  $\mathcal{S}$  is a subset of the rays.

To “OS-ize” an algorithm, replace all backprojections with partial sums.

Recall typical iteration:

$$\mathbf{x}^{(n+1)} = \mathbf{x}^{(n)} - \mathbf{D}(\mathbf{x}^{(n)}) \nabla \Psi(\mathbf{x}^{(n)}).$$

## Geometric View of Ordered Subsets



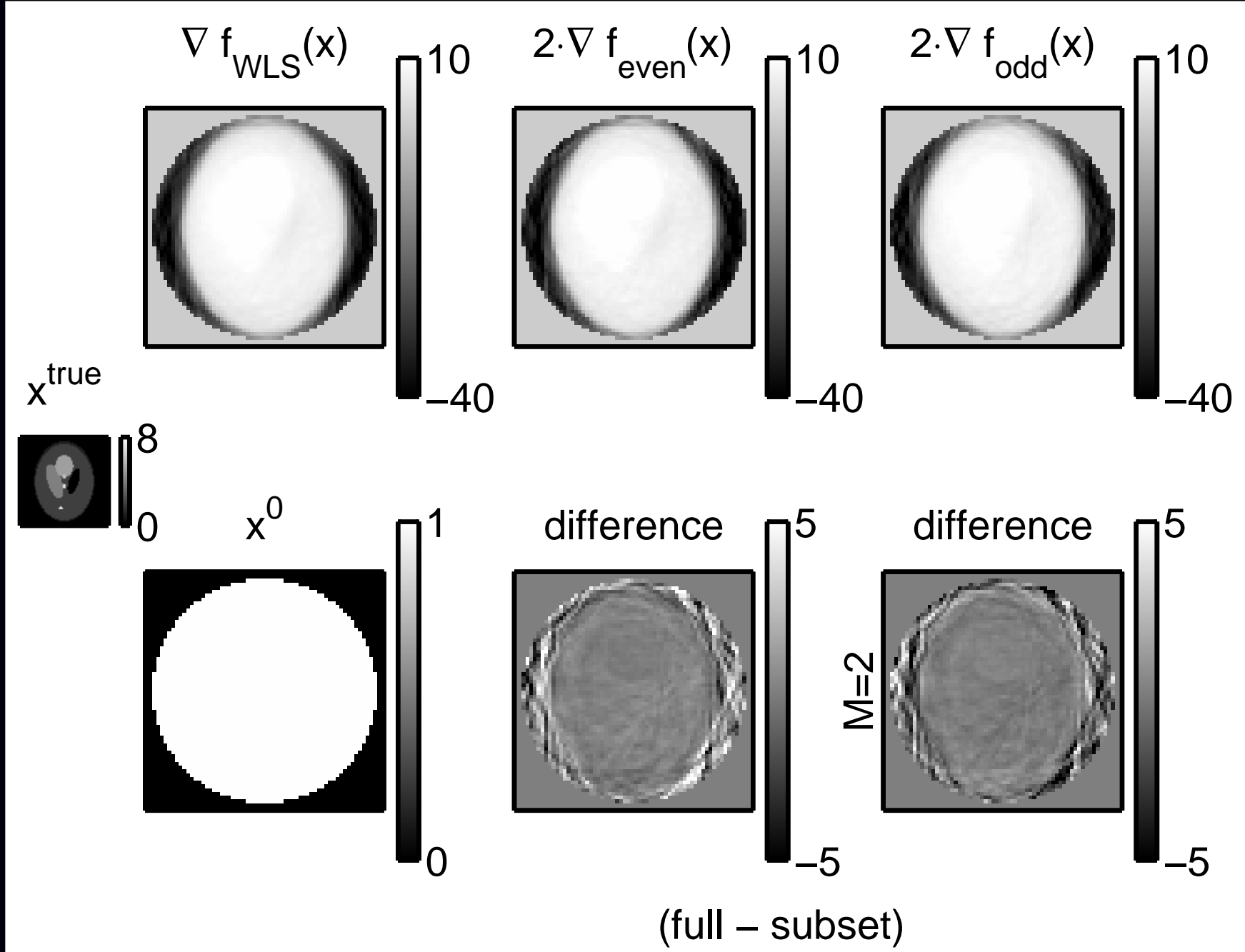
Two subset case:  $\Psi(\mathbf{x}) = f_1(\mathbf{x}) + f_2(\mathbf{x})$  (e.g., odd and even projection views).

For  $\mathbf{x}^{(n)}$  far from  $\mathbf{x}^*$ , even partial gradients should point roughly towards  $\mathbf{x}^*$ .

For  $\mathbf{x}^{(n)}$  near  $\mathbf{x}^*$ , however,  $\nabla\Psi(\mathbf{x}) \approx \mathbf{0}$ , so  $\nabla f_1(\mathbf{x}) \approx -\nabla f_2(\mathbf{x}) \implies$  cycles!

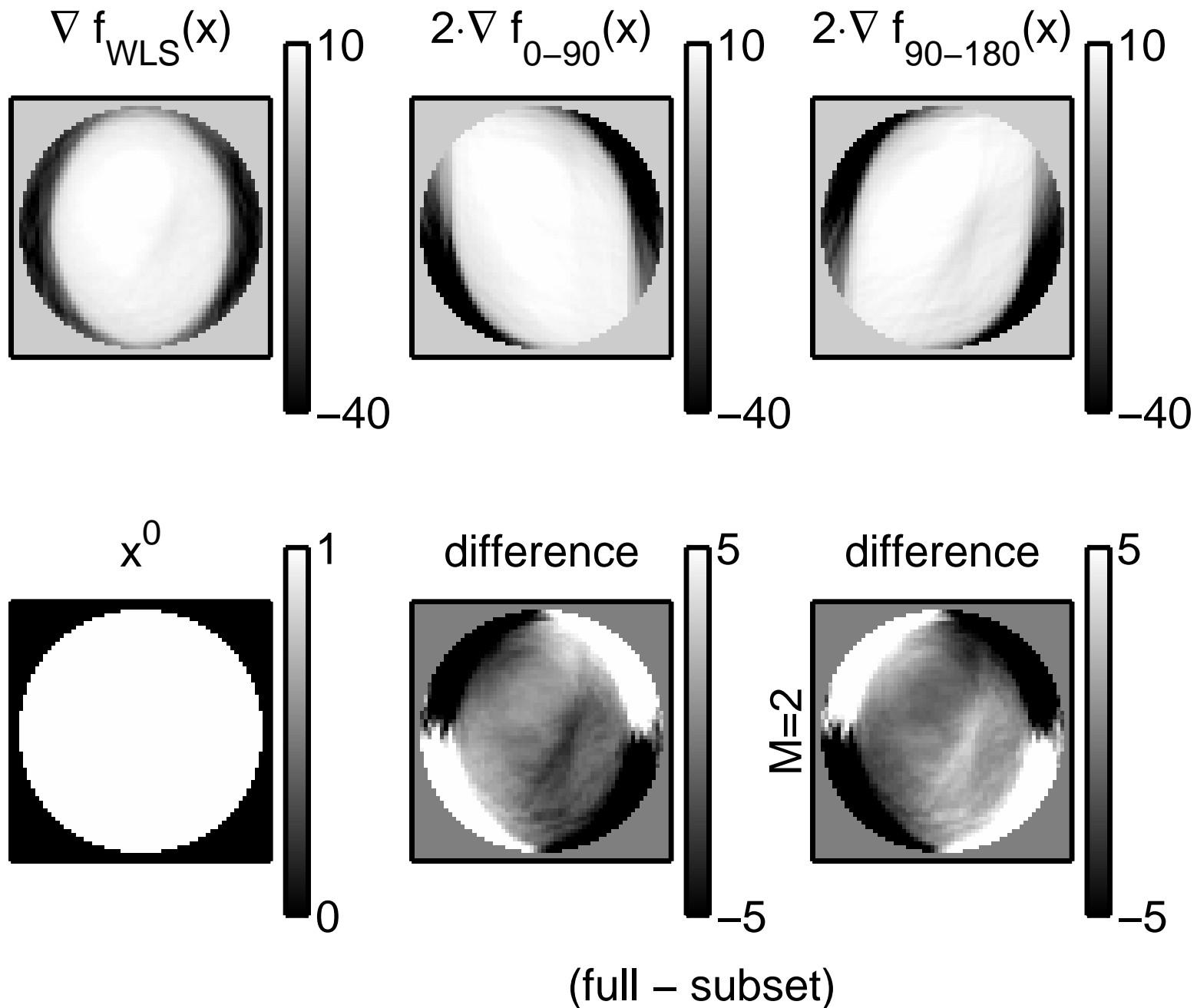
Issues. "Subset gradient balance":  $\nabla\Psi(\mathbf{x}) \approx M\nabla f_k(\mathbf{x})$ . Choice of ordering.

# Incremental Gradients (WLS, 2 Subsets)





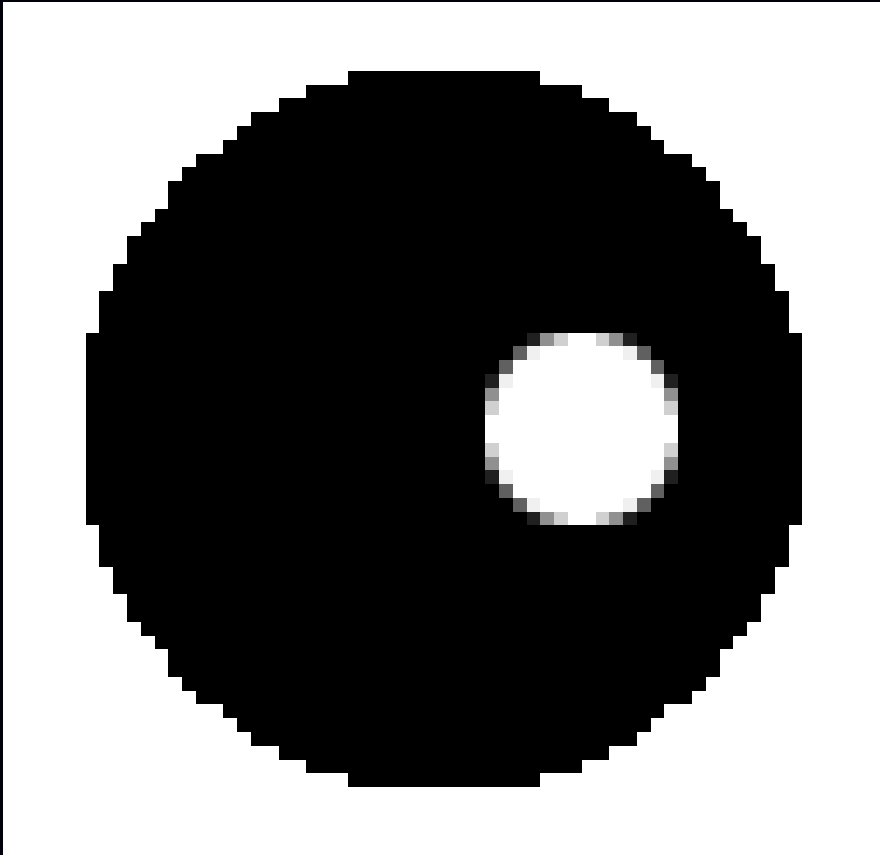
# Subset Gradient Imbalance



## Problems with OS-EM

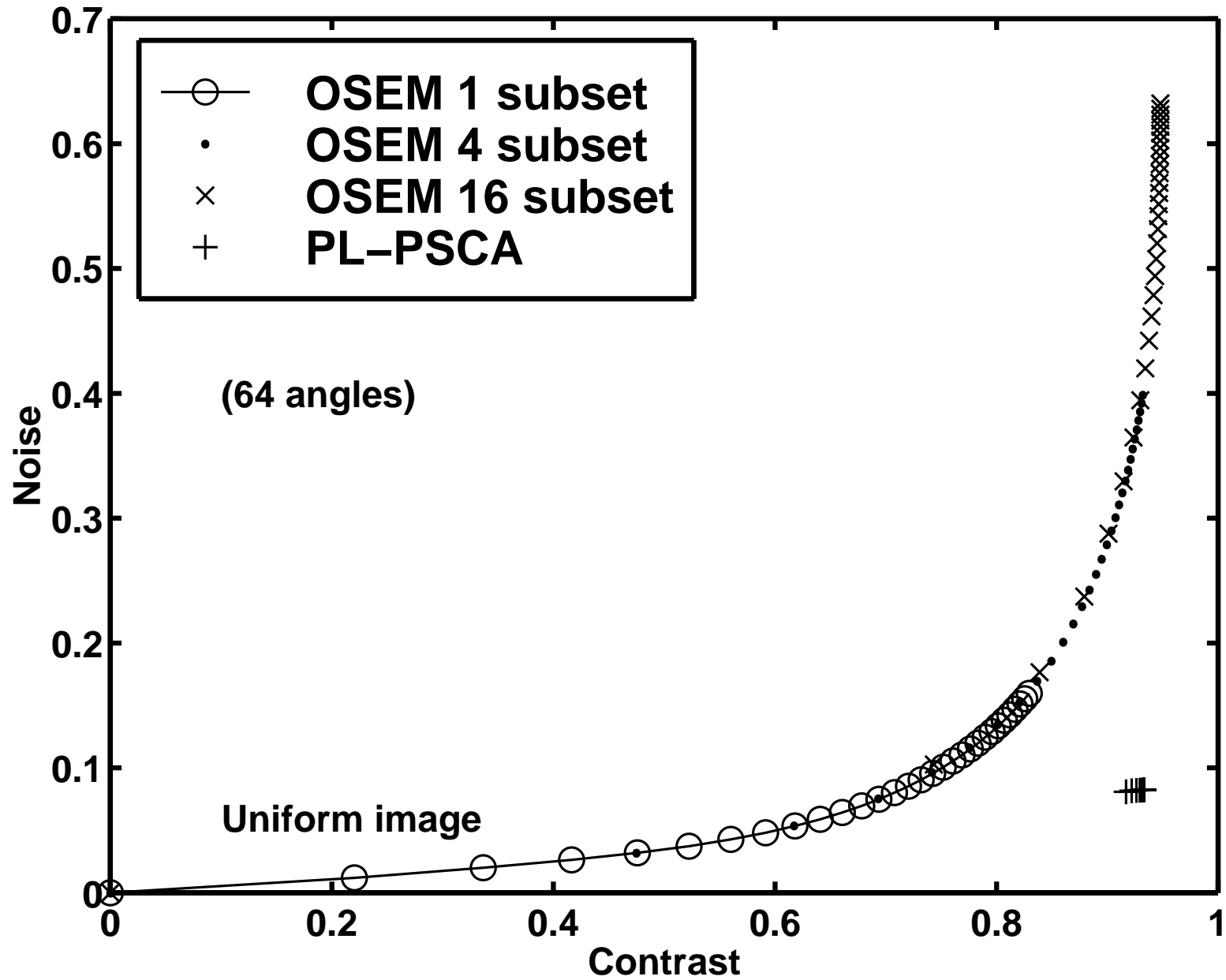
- Non-monotone
- Does not converge (may cycle)
- Byrne's "rescaled block iterative" (RBI) approach converges only for consistent (noiseless) data
- $\therefore$  unpredictable
  - What resolution after  $n$  iterations?  
Object-dependent, spatially nonuniform
  - What variance after  $n$  iterations?
  - ROI variance? (e.g., for Huesman's WLS kinetics)

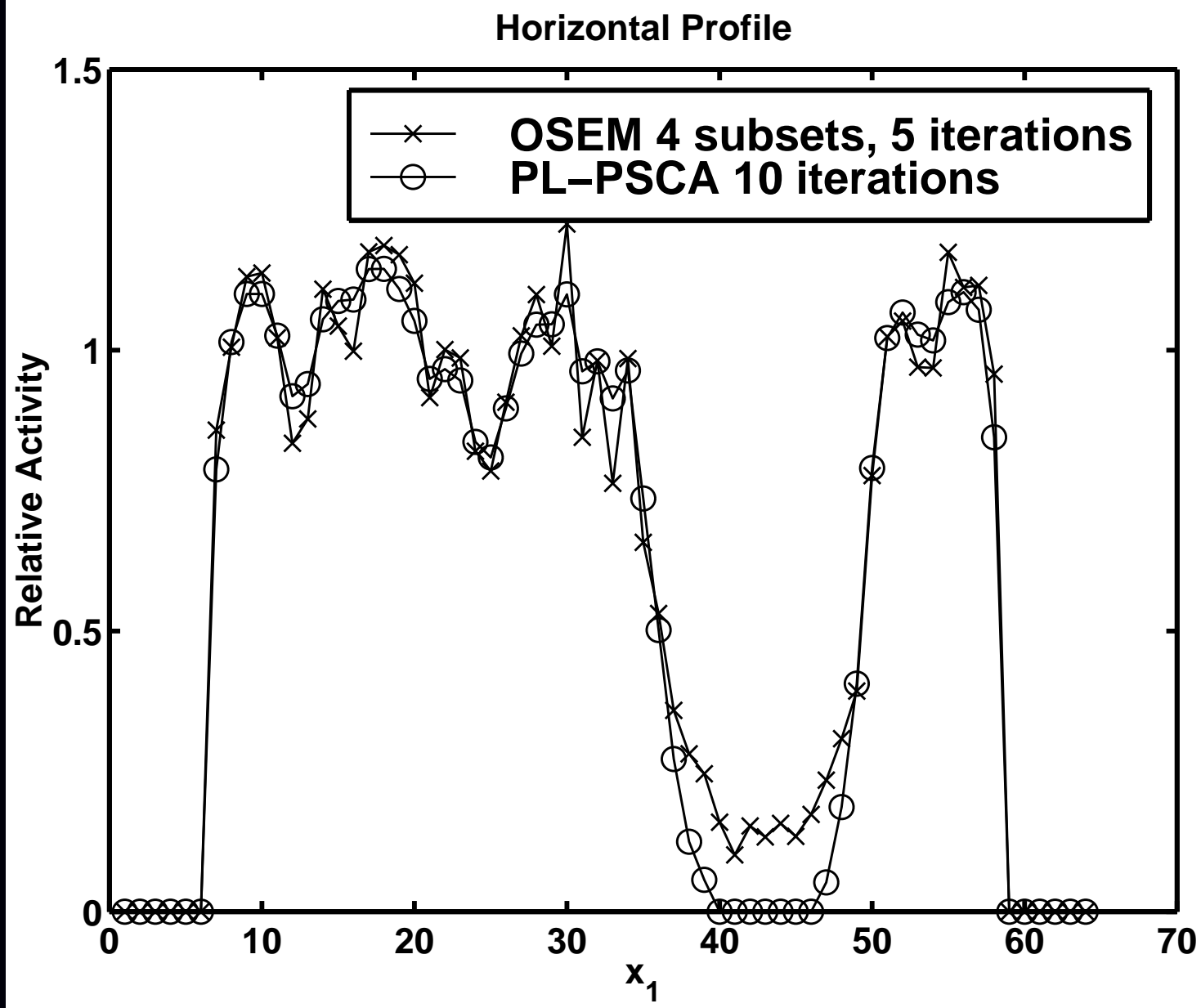
## OSEM vs Penalized Likelihood



- $64 \times 62$  image
- $66 \times 60$  sinogram
- $10^6$  counts
- 15% randoms/scatter
- uniform attenuation
- contrast in cold region
- within-region  $\sigma$  opposite side

# Contrast-Noise Results





# Making OS Methods Converge

- Relaxation
- Incrementalism

## Relaxed block-iterative methods

$$\Psi(\mathbf{x}) = \sum_{m=1}^M \Psi_m(\mathbf{x})$$

$$\mathbf{x}^{(n+(m+1)/M)} = \mathbf{x}^{(n+m/M)} - \alpha_n D(\mathbf{x}^{(n+m/M)}) \nabla \Psi_m(\mathbf{x}^{(n+m/M)}), \quad m = 0, \dots, M-1$$

Relaxation of step sizes:

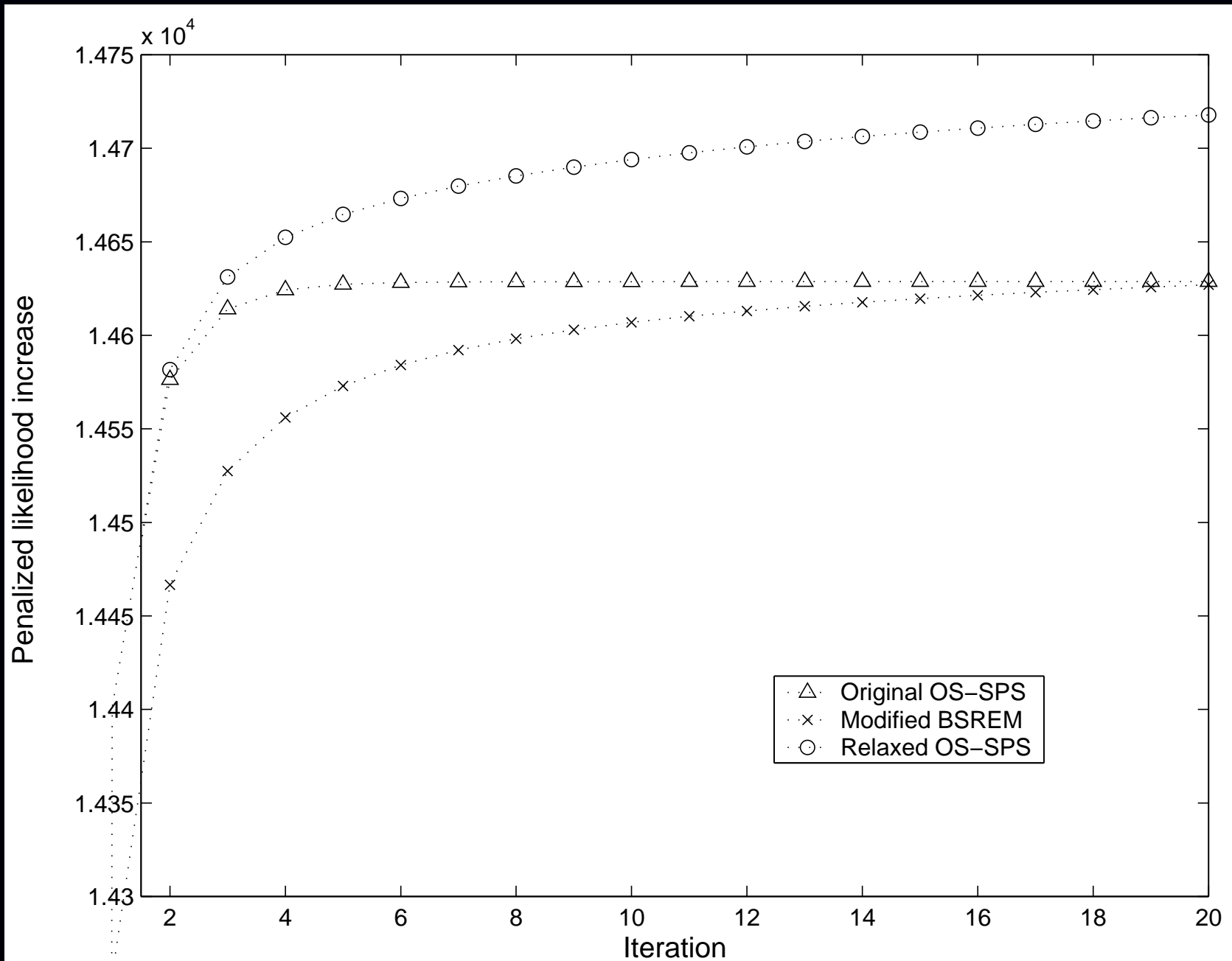
$$\alpha_n \rightarrow 0 \text{ as } n \rightarrow \infty, \quad \sum_n \alpha_n = \infty, \quad \sum_n \alpha_n^2 < \infty$$

- ART
- RAMLA, BSREM (De Pierro, T-MI, 1997, 2001)
- Ahn and Fessler, NSS/MIC 2001, T-MI 2003

## Considerations

- Proper relaxation can induce convergence, *but* still lacks monotonicity.
- Choice of relaxation schedule requires experimentation.
- $\Psi_m(\mathbf{x}) = \ell_m(\mathbf{x}) + \frac{1}{M} R(\mathbf{x})$ , so each  $\Psi_m$  includes part of the likelihood yet all of  $R$

# Relaxed OS-SPS



## Incremental Methods

Incremental EM applied to emission tomography by Hsiao *et al.* as C-OSEM

Incremental optimization transfer (Ahn & Fessler, MIC 2004)

Find majorizing surrogate for each sub-objective function:

$$\begin{aligned}\phi_m(\mathbf{x}; \mathbf{x}) &= \Psi_m(\mathbf{x}), & \forall \mathbf{x} \\ \phi_m(\mathbf{x}; \bar{\mathbf{x}}) &\geq \Psi_m(\mathbf{x}), & \forall \mathbf{x}, \bar{\mathbf{x}}\end{aligned}$$

Define the following augmented cost function:  $F(\mathbf{x}; \bar{\mathbf{x}}_1, \dots, \bar{\mathbf{x}}_M) = \sum_{m=1}^M \phi_m(\mathbf{x}; \bar{\mathbf{x}}_m)$ .

Fact: by construction  $\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \Psi(\mathbf{x}) = \arg \min_{\mathbf{x}} \min_{\bar{\mathbf{x}}_1, \dots, \bar{\mathbf{x}}_M} F(\mathbf{x}; \bar{\mathbf{x}}_1, \dots, \bar{\mathbf{x}}_M)$ .

Alternating minimization: for  $m = 1, \dots, M$ :

$$\mathbf{x}^{\text{new}} = \arg \min_{\mathbf{x}} F\left(\mathbf{x}; \bar{\mathbf{x}}_1^{(n+1)}, \dots, \bar{\mathbf{x}}_{m-1}^{(n+1)}, \bar{\mathbf{x}}_m^{(n)}, \bar{\mathbf{x}}_{m+1}^{(n)}, \dots, \bar{\mathbf{x}}_M^{(n)}\right)$$

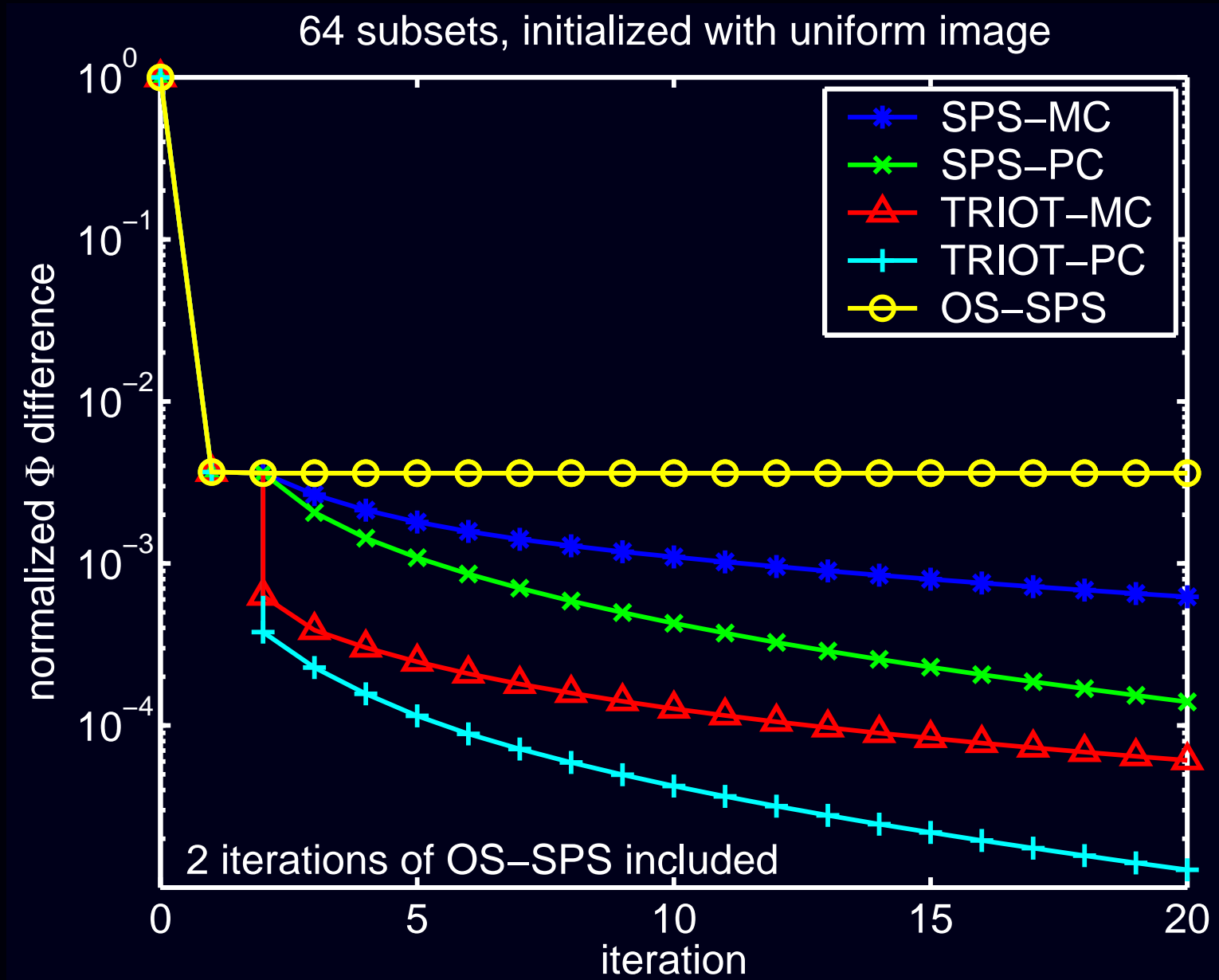
$$\bar{\mathbf{x}}_m^{(n+1)} = \arg \min_{\bar{\mathbf{x}}_m} F\left(\mathbf{x}^{\text{new}}; \bar{\mathbf{x}}_1^{(n+1)}, \dots, \bar{\mathbf{x}}_{m-1}^{(n+1)}, \bar{\mathbf{x}}_m, \bar{\mathbf{x}}_{m+1}^{(n)}, \dots, \bar{\mathbf{x}}_M^{(n)}\right) = \mathbf{x}^{\text{new}}.$$

- Use all current information, but increment the surrogate for only one subset.
- Monotone in  $F$ , converges under reasonable assumptions on  $\Psi$
- In contrast, OS-EM uses the information from *only* one subset at a time

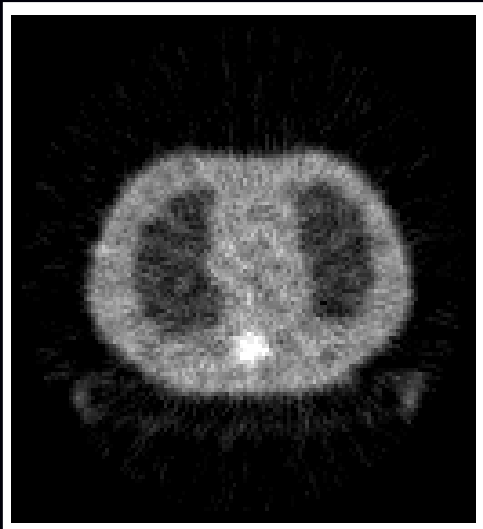


# TRIOT Example: Convergence Rate

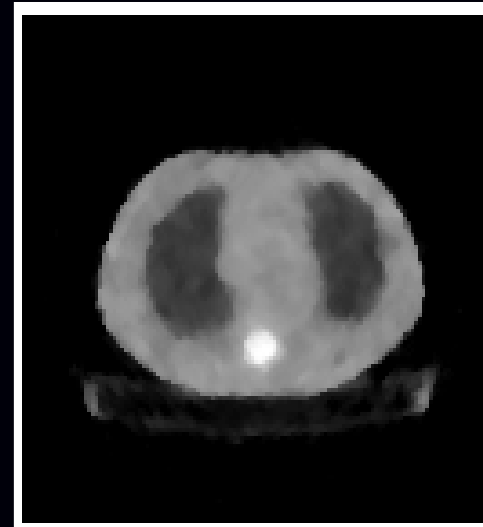
Transmission incremental optimization transfer (TRIOT)



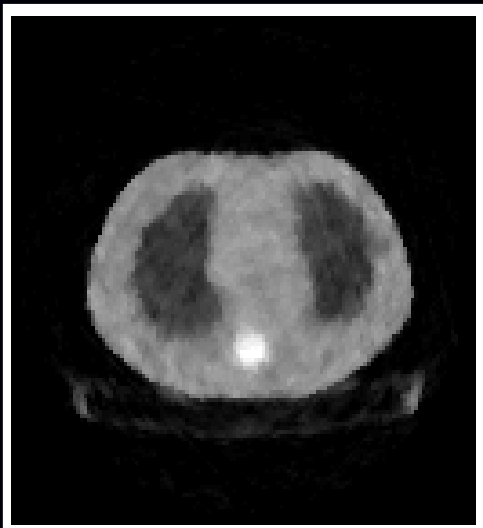
## TRIOT Example: Attenuation Map Images



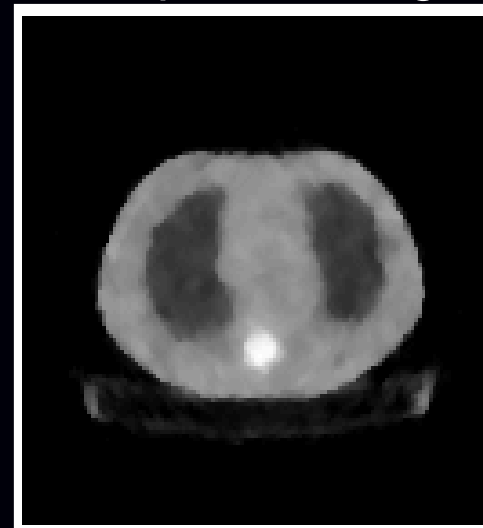
FBP



PL optimal image



OS-SPS

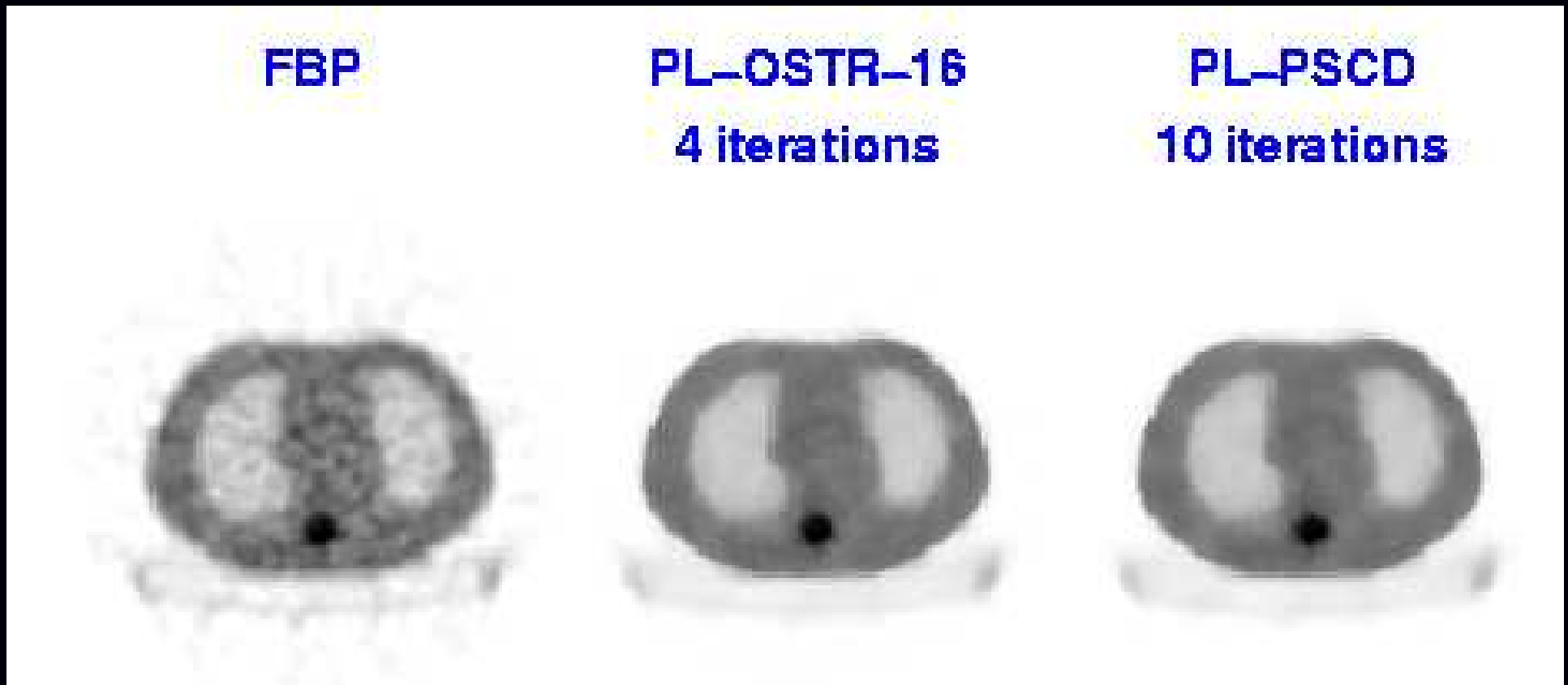


TRIOT-PC

OS-SPS: 64 subsets, 20 iterations, one point of the limit cycle

TRIOT-PC: 64 subsets, 20 iterations, after 2 iterations of OS-SPS)

## OSTR aka Transmission OS-SPS



Ordered subsets version of separable paraboloidal surrogates for PET transmission problem with nonquadratic convex *regularization*

Matlab m-file [http://www.eecs.umich.edu/~fessler/irt/irt/transmission/tpl\\_os\\_sps.m](http://www.eecs.umich.edu/~fessler/irt/irt/transmission/tpl_os_sps.m)

# Precomputed curvatures for OS-SPS

**Separable Paraboloidal Surrogate (SPS) Algorithm:**

$$x_j^{(n+1)} = \left[ x_j^{(n)} - \frac{\sum_{i=1}^{n_d} a_{ij} \dot{h}_i([\mathbf{A}\mathbf{x}^{(n)}]_i)}{\sum_{i=1}^{n_d} a_{ij} |a|_i c_i^{(n)}} \right]_+, \quad j = 1, \dots, n_p$$

Ordered-subsets abandons monotonicity, so why use optimal curvatures  $c_i^{(n)}$ ?

Precomputed curvature:

$$c_i = \ddot{h}_i(\hat{l}_i), \quad \hat{l}_i = \arg \min_l h_i(l)$$

Precomputed denominator (saves one backprojection each iteration!):

$$d_j = \sum_{i=1}^{n_d} a_{ij} |a|_i c_i, \quad j = 1, \dots, n_p.$$

OS-SPS algorithm with  $M$  subsets:

$$x_j^{(n+1)} = \left[ x_j^{(n)} - \frac{\sum_{i \in \mathcal{S}^{(n)}} a_{ij} \dot{h}_i([\mathbf{A}\mathbf{x}^{(n)}]_i)}{d_j / M} \right]_+, \quad j = 1, \dots, n_p$$

# Summary of Algorithms

- General-purpose optimization algorithms
- Optimization transfer for image reconstruction algorithms
- Separable surrogates  $\implies$  high curvatures  $\implies$  slow convergence
- Ordered subsets accelerate *initial* convergence  
require relaxation or incrementalism for true convergence
- Principles apply to emission and transmission reconstruction
- Still work to be done...

Matlab/Freemat “image reconstruction toolbox” online:

<http://www.eecs.umich.edu/~fessler/code>

## An Open Problem

Still no algorithm with all of the following properties:

- Nonnegativity easy
- Fast converging
- Intrinsically monotone global convergence
- Accepts any type of system matrix
- Parallelizable

## Part 4. Performance Characteristics

Easy case: MRI with quadratic regularization

- Spatial resolution properties
- Noise properties

General case

- Spatial resolution properties
- Noise properties
- Detection properties

# Regularized Least-Squares Estimation

Estimate object by minimizing a *regularized* cost function:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathbb{C}^{np}} \Psi(\mathbf{x}), \quad \Psi(\mathbf{x}) = \|\mathbf{y} - \mathbf{Ax}\|^2 + \alpha R(\mathbf{x})$$

- **data fit** term  $\|\mathbf{y} - \mathbf{Ax}\|^2$   
corresponds to negative log-likelihood of Gaussian distribution
- **regularizing** term  $R(\mathbf{x})$  controls noise by penalizing roughness,

$$\text{e.g. : } R(\mathbf{x}) \approx \int \|\nabla f\|^2 d\vec{r}$$

- **regularization parameter**  $\alpha > 0$   
controls tradeoff between spatial resolution and noise
- Equivalent to Bayesian MAP estimation with prior  $\propto e^{-\alpha R(\mathbf{x})}$

Complexities:

- choosing  $R(f)$
- choosing  $\alpha$
- computing minimizer rapidly.

## Quadratic regularization

1D example: squared differences between neighboring pixel values:

$$R(f) = \sum_{j=2}^{n_p} \frac{1}{2} |f_j - f_{j-1}|^2.$$

In matrix-vector notation,  $R(\mathbf{x}) = \frac{1}{2} \|\mathbf{C}\mathbf{x}\|^2$  where

$$\mathbf{C} = \begin{bmatrix} -1 & 1 & 0 & 0 & \dots & 0 \\ 0 & -1 & 1 & 0 & \dots & 0 \\ & & & \ddots & \ddots & \\ 0 & \dots & 0 & 0 & -1 & 1 \end{bmatrix}, \text{ so } \mathbf{C}\mathbf{x} = \begin{bmatrix} x_2 - x_1 \\ \vdots \\ x_N - x_{N-1} \end{bmatrix}.$$

For 2D and higher-order differences, modify differencing matrix  $\mathbf{C}$ .

Leads to closed-form solution:

$$\begin{aligned} \hat{\mathbf{x}} &= \arg \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 + \alpha \|\mathbf{C}\mathbf{x}\|^2 \\ &= [\mathbf{A}'\mathbf{A} + \alpha \mathbf{C}'\mathbf{C}]^{-1} \mathbf{A}'\mathbf{y}. \end{aligned}$$

(a formula of limited practical use for computing  $\hat{\mathbf{x}}$ )



# Choosing the Regularization Parameter

Spatial resolution analysis (Fessler & Rogers, IEEE T-IP, 1996):

$$\hat{x} = [A'A + \alpha C'C]^{-1} A'y$$
$$E[\hat{x}] = [A'A + \alpha C'C]^{-1} A'E[y]$$
$$E[\hat{x}] = \underbrace{[A'A + \alpha C'C]^{-1} A'A}_{\text{blur}} x$$

$A'A$  and  $C'C$  are Toeplitz  $\implies$  blur is approximately shift-invariant.

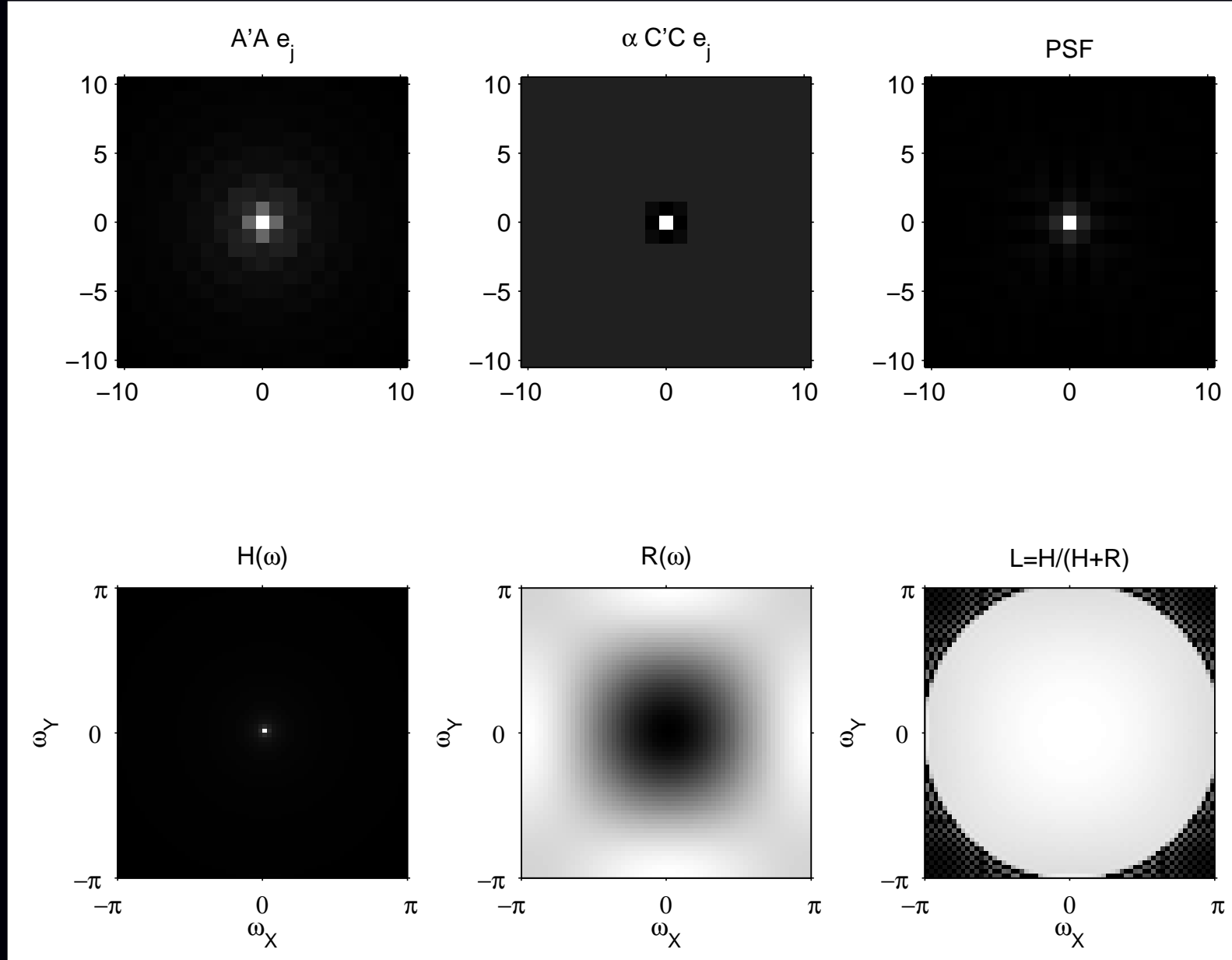
Frequency response of blur:

$$L(\omega) = \frac{H(\omega)}{H(\omega) + \alpha R(\omega)}$$

- $H(\omega_k) = \text{FFT}(A'A e_j)$  (lowpass)
- $R(\omega_k) = \text{FFT}(C'C e_j)$  (highpass)

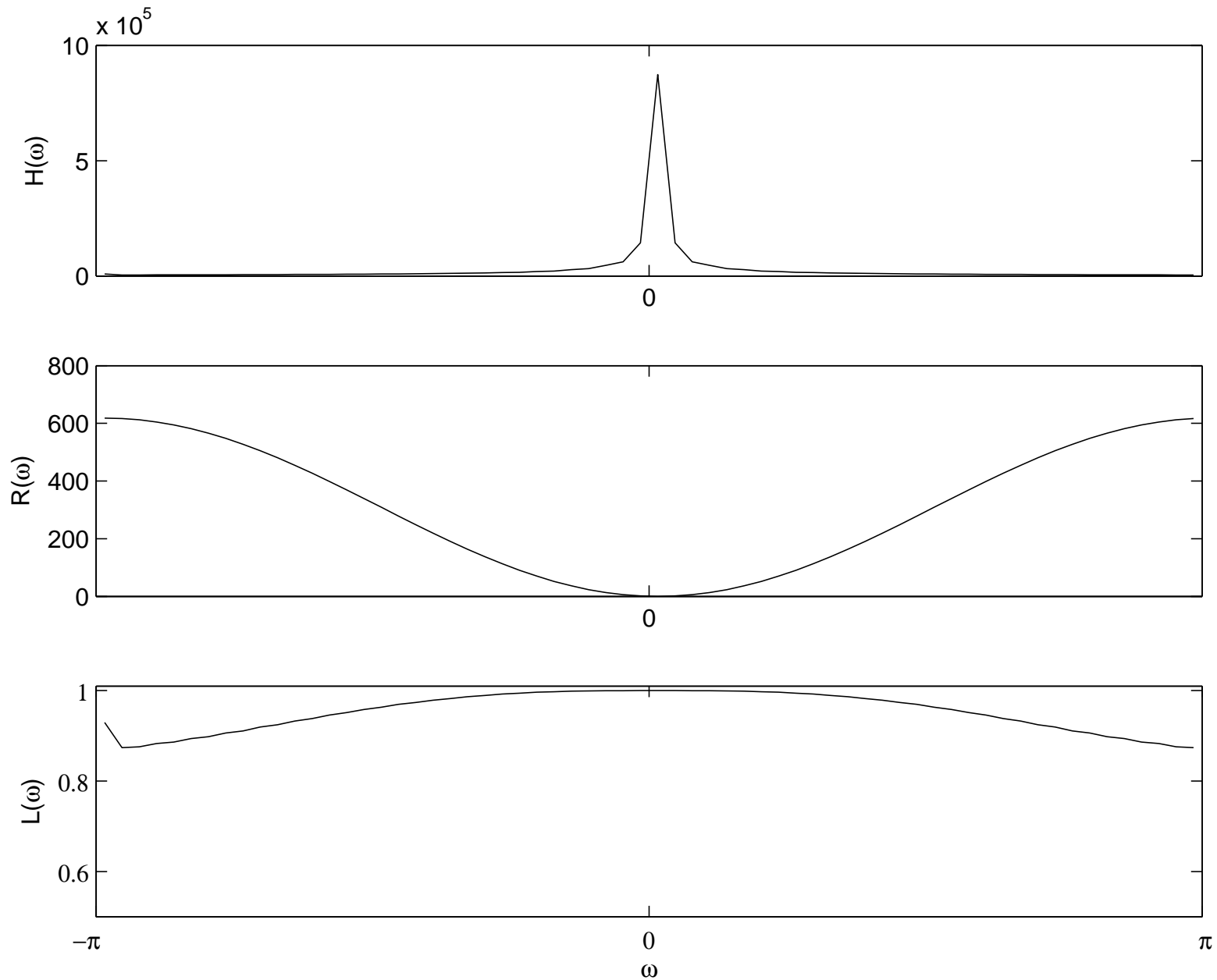
Adjust  $\alpha$  to achieve desired spatial resolution.

# Spatial Resolution Example

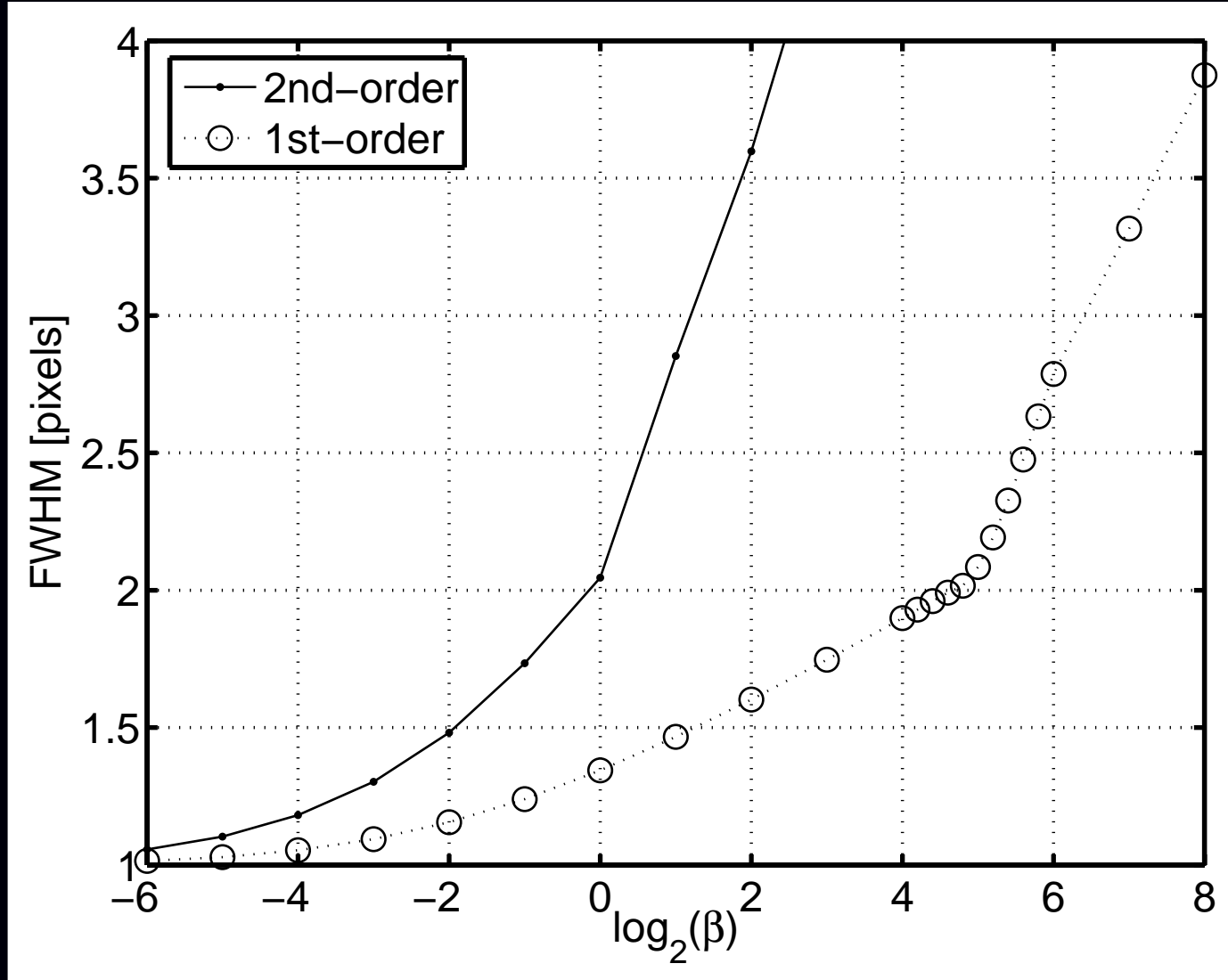


Radial k-space trajectory, FWHM of PSF is 1.2 pixels

# Spatial Resolution Example: Profiles



# Tabulating Spatial Resolution vs Regularization



Trajectory specific, but easily computed using a few FFTs  
Works only for quadratic regularization

## Resolution/noise tradeoffs

Noise analysis:

$$\text{Cov}\{\hat{\mathbf{x}}\} = [\mathbf{A}'\mathbf{A} + \alpha\mathbf{C}'\mathbf{C}]^{-1} \mathbf{A}' \text{Cov}\{\mathbf{y}\} \mathbf{A} [\mathbf{A}'\mathbf{A} + \alpha\mathbf{C}'\mathbf{C}]^{-1}$$

Using circulant approximations to  $\mathbf{A}'\mathbf{A}$  and  $\mathbf{C}'\mathbf{C}$  yields:

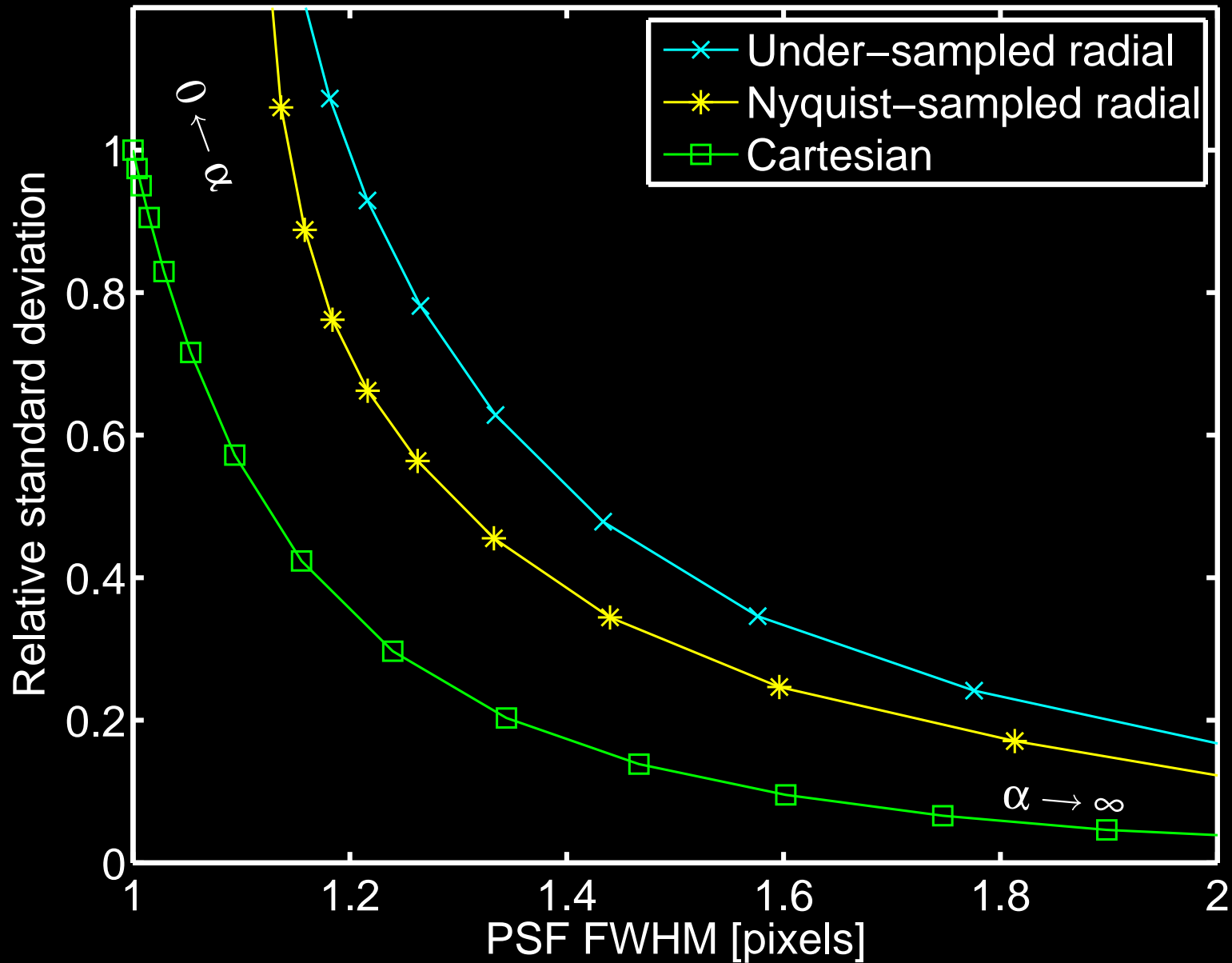
$$\text{Var}\{\hat{x}_j\} \approx \sigma_\varepsilon^2 \sum_k \frac{H(\omega_k)}{(H(\omega_k) + \alpha R(\omega_k))^2}$$

- $H(\omega_k) = \text{FFT}(\mathbf{A}'\mathbf{A} e_j)$  (lowpass)
- $R(\omega_k) = \text{FFT}(\mathbf{C}'\mathbf{C} e_j)$  (highpass)

⇒ Predicting reconstructed image noise requires just 2 FFTs.  
(cf. gridding approach?)

Adjust  $\alpha$  to achieve desired spatial resolution / noise tradeoff.

# Resolution/Noise Tradeoff Example



In short: one can choose  $\alpha$  rapidly and predictably for quadratic regularization.

# Part 4. Performance Characteristics

## (General case)

- Spatial resolution properties
- Noise properties
- Detection properties

# Spatial Resolution Properties

Choosing  $\beta$  can be painful, so ...

For true minimization methods:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \Psi(\mathbf{x})$$

the *local impulse response* is approximately (Fessler and Rogers, T-MI, 1996):

$$\mathbf{l}_j(\mathbf{x}) = \lim_{\delta \rightarrow 0} \frac{E[\hat{\mathbf{x}} | \mathbf{x} + \delta \mathbf{e}_j] - E[\hat{\mathbf{x}} | \mathbf{x}]}{\delta} \approx [-\nabla^{20} \Psi]^{-1} \nabla^{11} \Psi \frac{\partial}{\partial x_j} \bar{\mathbf{y}}(\mathbf{x}).$$

Depends only on chosen **cost function** and **statistical model**.

Independent of optimization algorithm (if iterated “to convergence”).

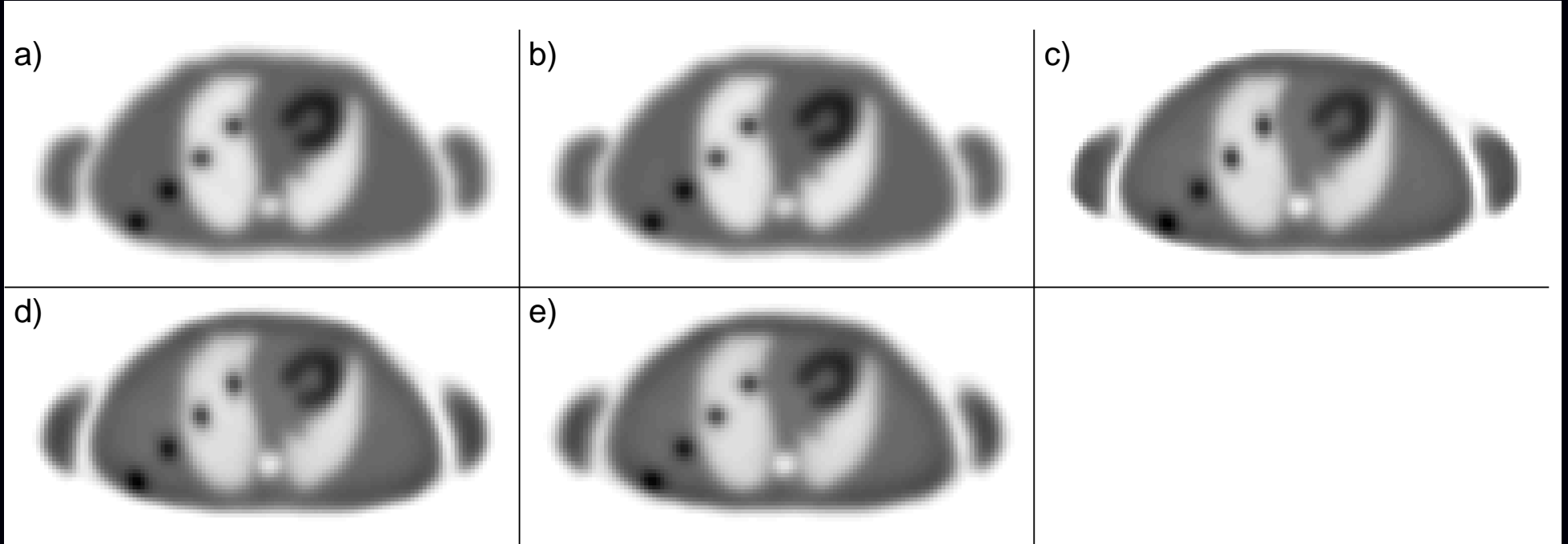
- Enables prediction of resolution properties (provided  $\Psi$  is minimized)
- Useful for designing regularization penalty functions with desired resolution properties. For penalized likelihood:

$$\mathbf{l}_j(\mathbf{x}) \approx [\mathbf{A}'\mathbf{W}\mathbf{A} + \beta\mathbf{R}]^{-1} \mathbf{A}'\mathbf{W}\mathbf{A}\mathbf{x}^{\text{true}}.$$

- Helps choose  $\beta$  for desired spatial resolution



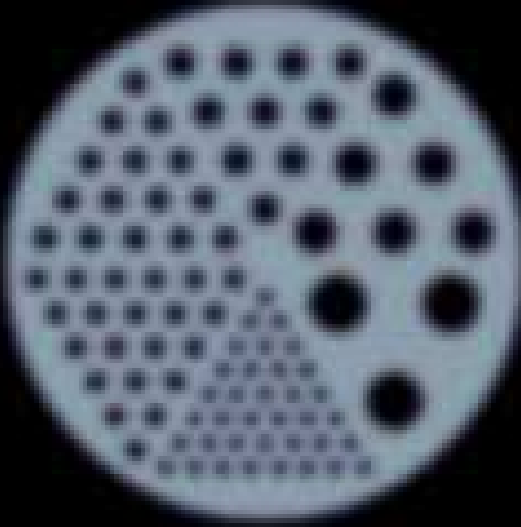
## Modified Penalty Example, PET



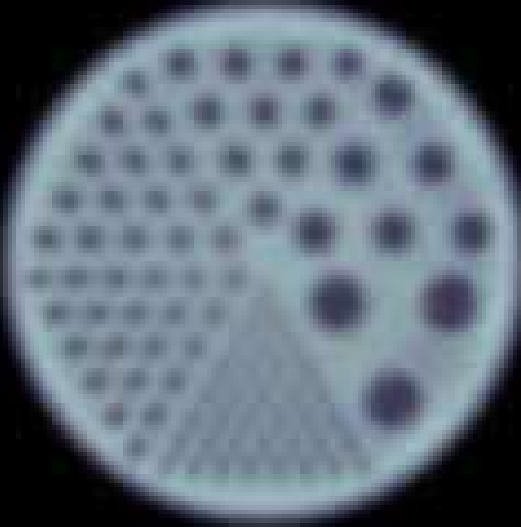
- a) filtered backprojection
- b) Penalized unweighted least-squares
- c) PWLS with conventional regularization
- d) PWLS with certainty-based penalty (Fessler & Rogers, 1996, T-MI)
- e) PWLS with modified penalty (Stayman & Fessler, 2000, T-MI)

# Modified Penalty Example, SPECT - Noiseless

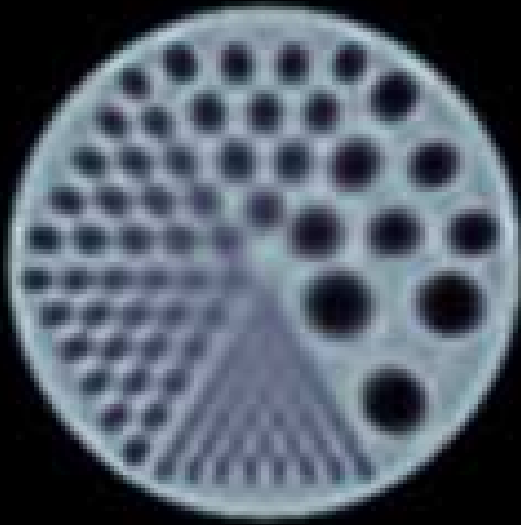
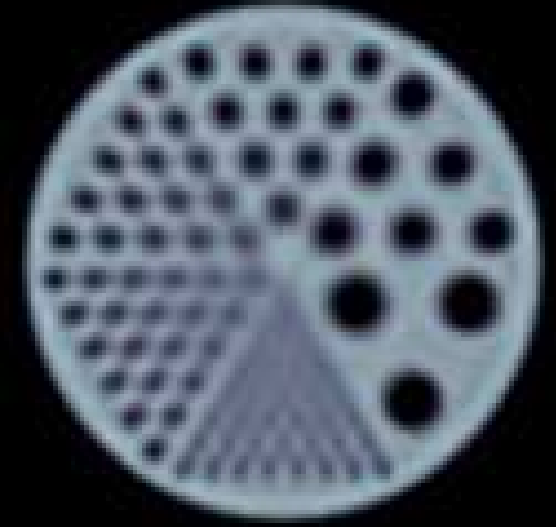
Target filtered object



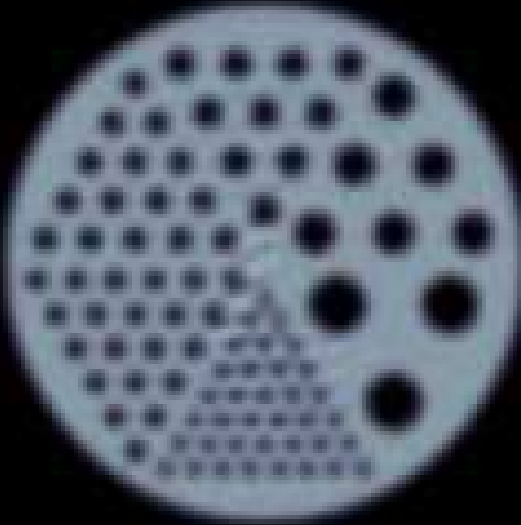
FBP



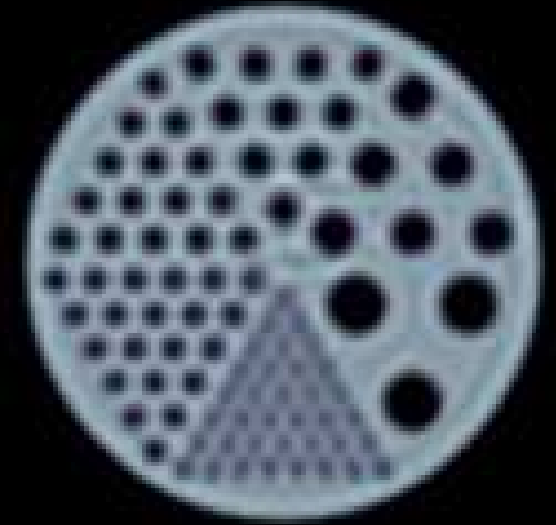
Conventional PWLS



Truncated EM



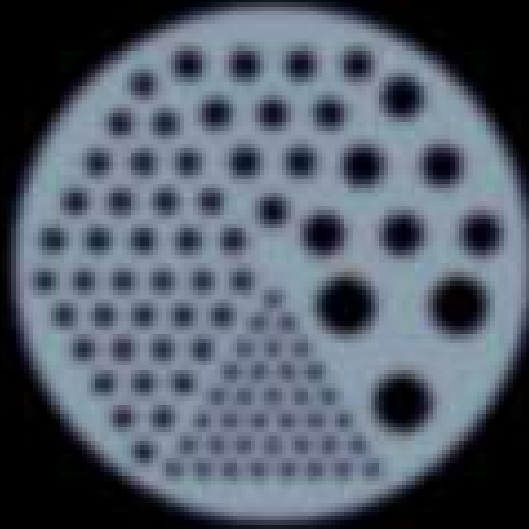
Post-filtered EM



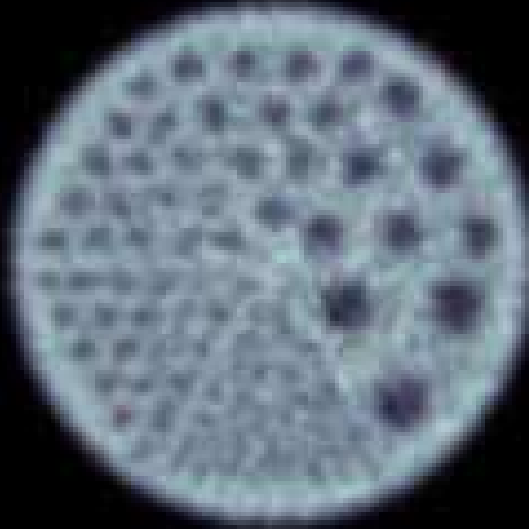
Modified Regularization

# Modified Penalty Example, SPECT - Noisy

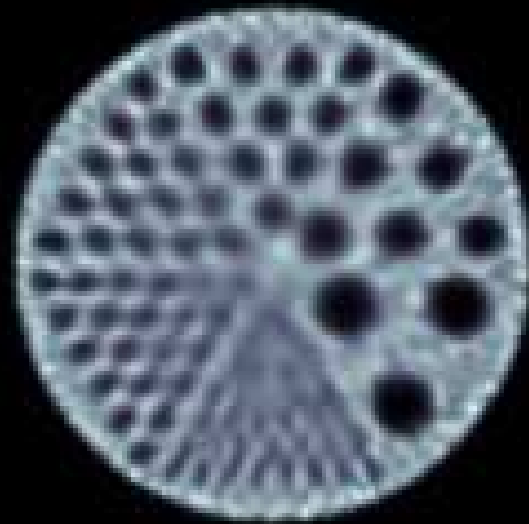
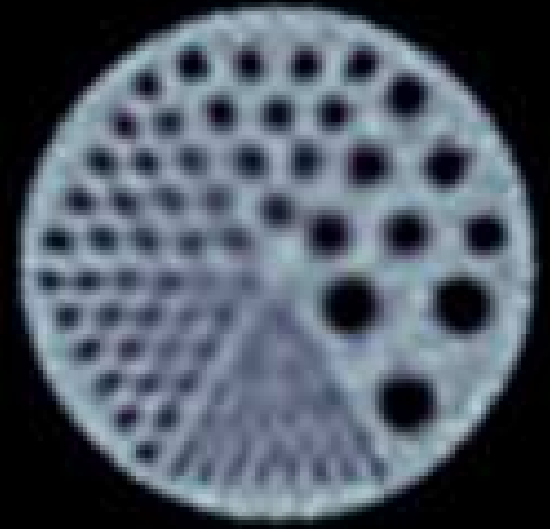
Target filtered object



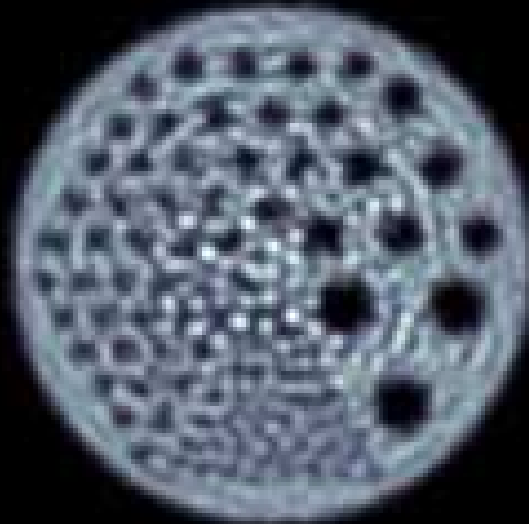
FBP



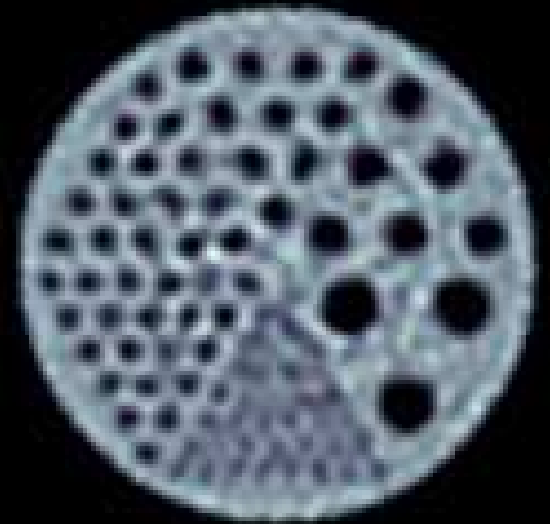
Conventional PWLS



Truncated EM

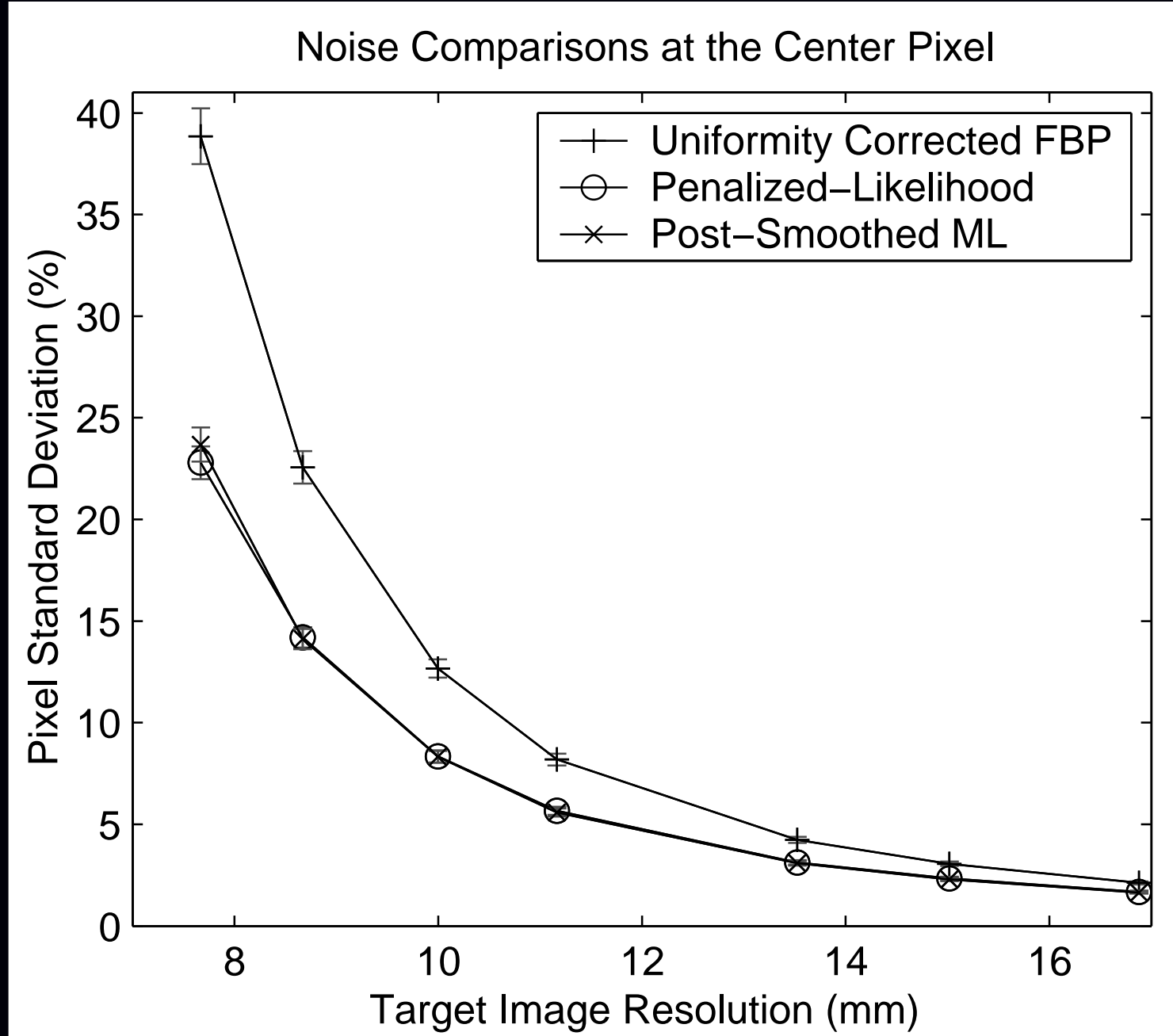


Post-filtered EM



Modified Regularization

# Regularized vs Post-filtered, with Matched PSF



# Reconstruction Noise Properties

For unconstrained (converged) minimization methods, the estimator is *implicit*:

$$\hat{\mathbf{x}} = \hat{\mathbf{x}}(\mathbf{y}) = \arg \min_x \Psi(\mathbf{x}, \mathbf{y}).$$

What is  $\text{Cov}\{\hat{\mathbf{x}}\}$ ?

New simpler derivation.

Denote the column gradient by  $g(\mathbf{x}, \mathbf{y}) \triangleq \nabla_x \Psi(\mathbf{x}, \mathbf{y})$ .

Ignoring constraints, the gradient is zero at the minimizer:  $g(\hat{\mathbf{x}}(\mathbf{y}), \mathbf{y}) = \mathbf{0}$ .

First-order Taylor series expansion:

$$\begin{aligned} g(\hat{\mathbf{x}}, \mathbf{y}) &\approx g(\mathbf{x}^{\text{true}}, \mathbf{y}) + \nabla_x g(\mathbf{x}^{\text{true}}, \mathbf{y})(\hat{\mathbf{x}} - \mathbf{x}^{\text{true}}) \\ &= g(\mathbf{x}^{\text{true}}, \mathbf{y}) + \nabla_x^2 \Psi(\mathbf{x}^{\text{true}}, \mathbf{y})(\hat{\mathbf{x}} - \mathbf{x}^{\text{true}}). \end{aligned}$$

Equating to zero:

$$\hat{\mathbf{x}} \approx \mathbf{x}^{\text{true}} - [\nabla_x^2 \Psi(\mathbf{x}^{\text{true}}, \mathbf{y})]^{-1} \nabla_x \Psi(\mathbf{x}^{\text{true}}, \mathbf{y}).$$

If the Hessian  $\nabla^2 \Psi$  is weakly dependent on  $\mathbf{y}$ , then

$$\boxed{\text{Cov}\{\hat{\mathbf{x}}\} \approx [\nabla_x^2 \Psi(\mathbf{x}^{\text{true}}, \bar{\mathbf{y}})]^{-1} \text{Cov}\{\nabla_x \Psi(\mathbf{x}^{\text{true}}, \mathbf{y})\} [\nabla_x^2 \Psi(\mathbf{x}^{\text{true}}, \bar{\mathbf{y}})]^{-1} .}$$

If we further linearize w.r.t. the data:  $g(\mathbf{x}, \mathbf{y}) \approx g(\mathbf{x}, \bar{\mathbf{y}}) + \nabla_y g(\mathbf{x}, \bar{\mathbf{y}})(\mathbf{y} - \bar{\mathbf{y}})$ , then

$$\text{Cov}\{\hat{\mathbf{x}}\} \approx [\nabla_x^2 \Psi]^{-1} (\nabla_x \nabla_y \Psi) \text{Cov}\{\mathbf{y}\} (\nabla_x \nabla_y \Psi)' [\nabla_x^2 \Psi]^{-1} .$$

## Covariance Continued

Covariance approximation:

$$\text{Cov}\{\hat{\mathbf{x}}\} \approx [\nabla_x^2 \Psi(\mathbf{x}^{\text{true}}, \bar{\mathbf{y}})]^{-1} \text{Cov}\{\nabla_x \Psi(\mathbf{x}^{\text{true}}, \mathbf{y})\} [\nabla_x^2 \Psi(\mathbf{x}^{\text{true}}, \bar{\mathbf{y}})]^{-1}$$

Depends only on chosen **cost function** and **statistical model**.  
Independent of optimization algorithm.

- Enables prediction of noise properties
- Can make variance images
- Useful for computing ROI variance (e.g., for weighted kinetic fitting)
- Good variance prediction for quadratic regularization in nonzero regions
- Inaccurate for nonquadratic penalties, or in nearly-zero regions

## Detection Analysis

Qi and Huesman (IEEE T-MI, Aug. 2001) showed analytically:

SNR of MAP reconstruction  $>$  SNR of FBP reconstruction

quadratic regularization

SKE/BKE task

prewhitened observer

non-prewhitened observer

## Open issues

Choice of regularizer to optimize detectability?

Active work in several groups.

# Choosing $\beta$ : Unknown location

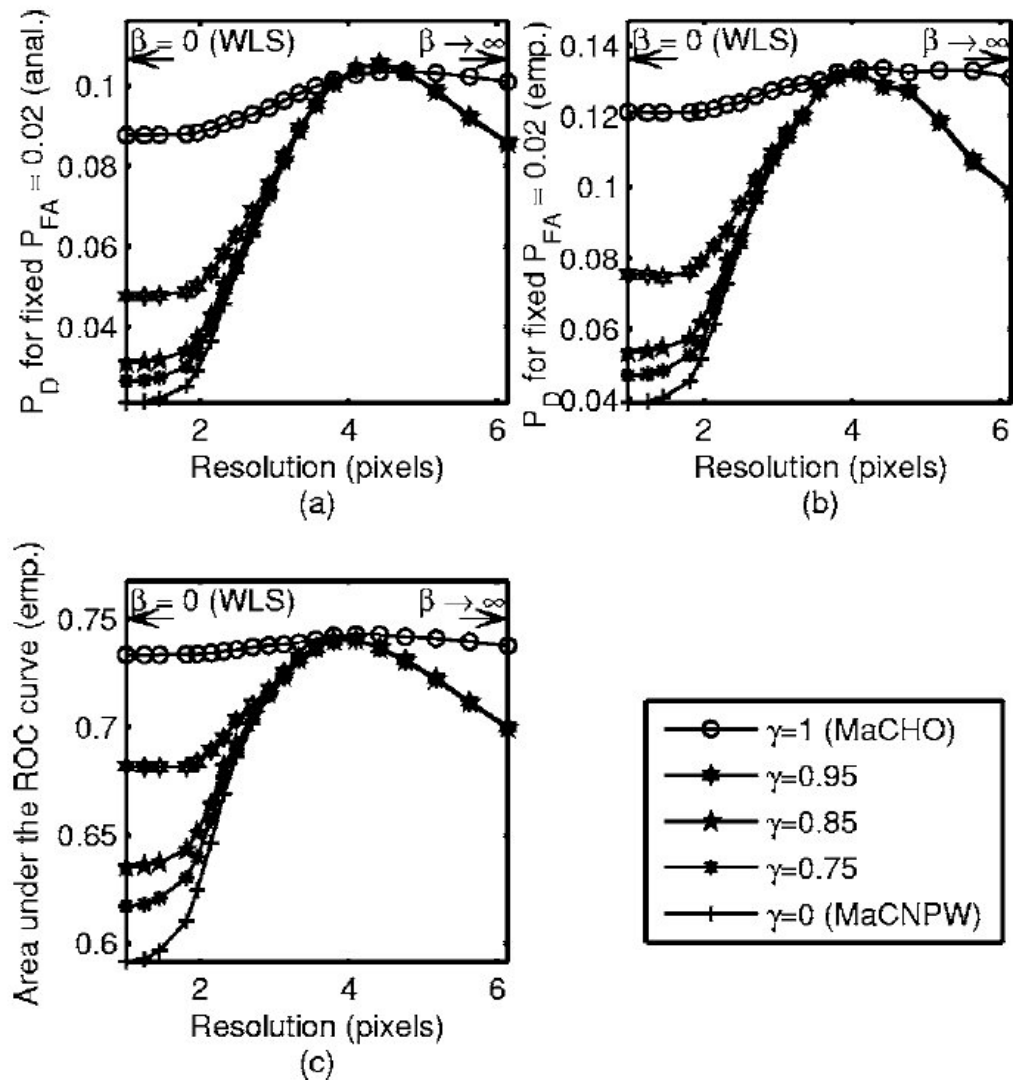


Fig. 2. Detection performance of MaCPPW observers versus QPWLS reconstruction resolution:  $P_D$  obtained analytically (a),  $P_D$  obtained empirically (b), and AUC obtained empirically (c). Results are shown for five different degrees of prewhitening accuracy. The search area is a disk with a diameter of 9 pixels.

AUC for signal detection with unknown location task.

Yendiki & Fessler, JOSA-A 24(12):B199, Dec. 2007



## Summary of Performance Analysis

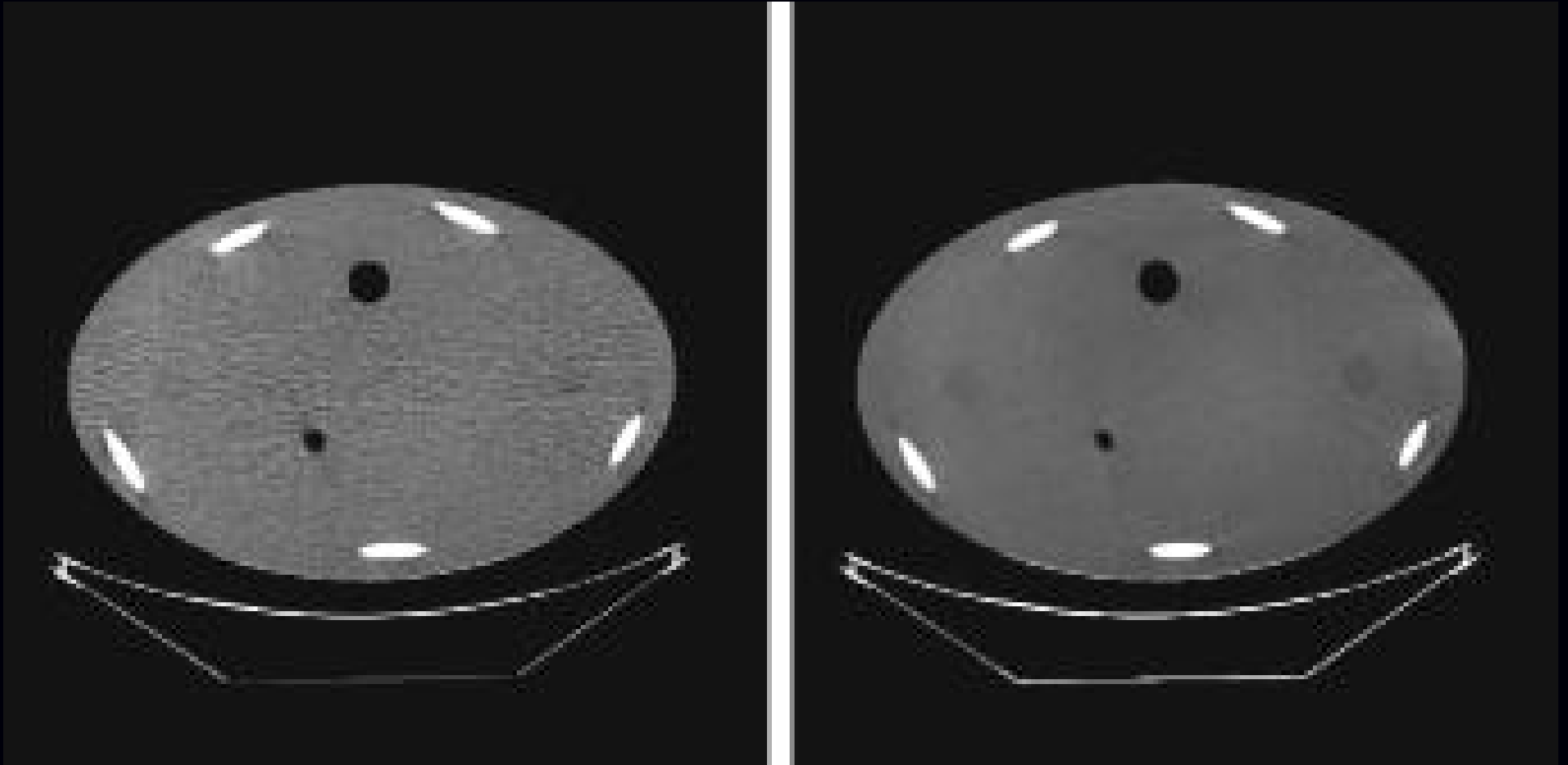
Spatial resolution / noise variance and covariance / AUC for signal detection:  
all (somewhat) predictable based on properties of *cost function*  $\Psi$ .  
(Provided an iterative algorithm is run “to convergence” to find minimizer of  $\Psi$ .)

This predictability also motivates regularized cost functions.  
(*cf.* unregularized cost function with a stopping rule.)

## Part: Application examples

- X-ray CT
- MRI

## Example: X-ray Helical CT

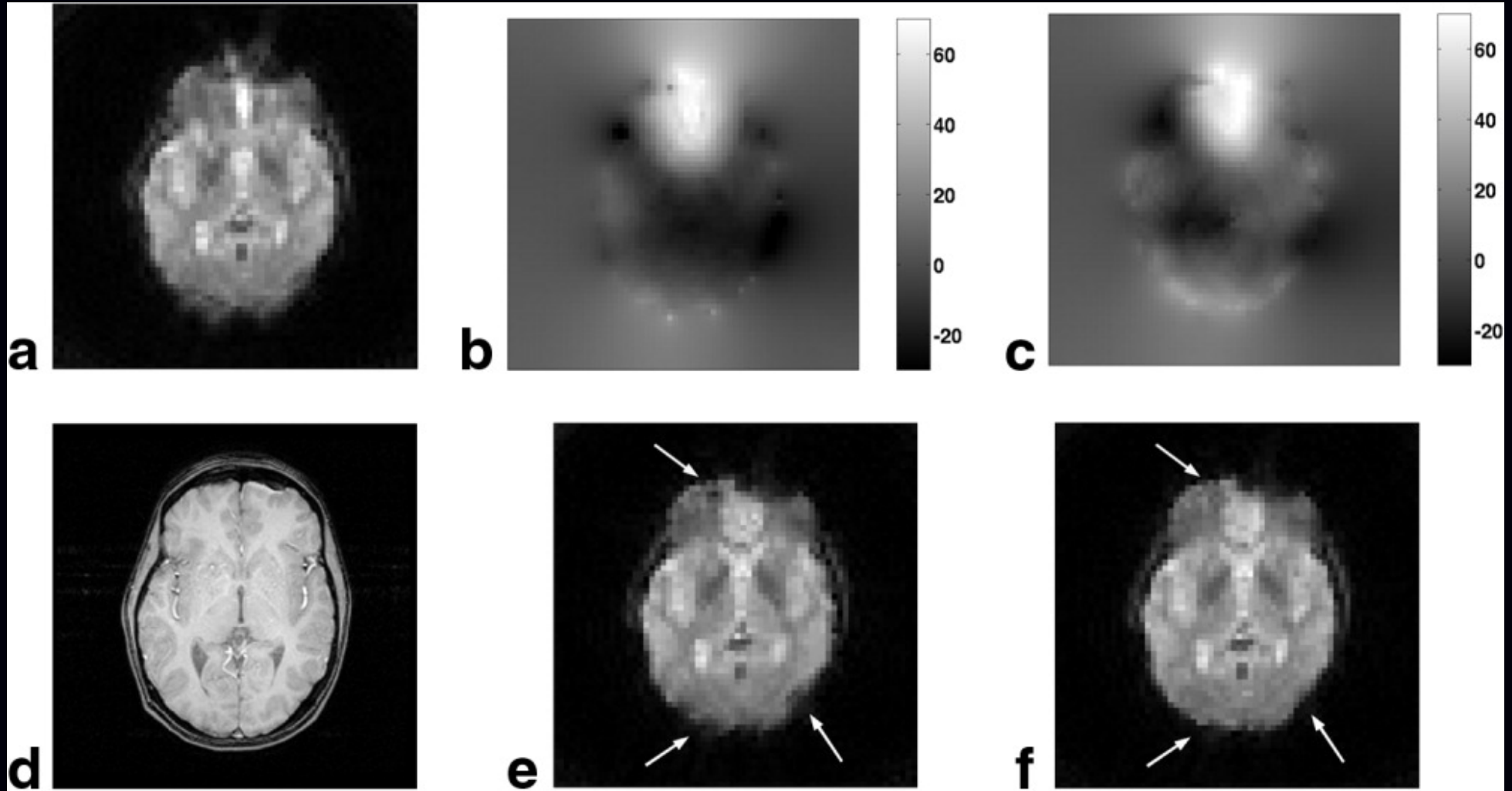


Left: FBP

Right: PWLS-ICD, edge-preserving

Thibault *et al.*, Med Phys. 34(11):4526, Nov. 2007

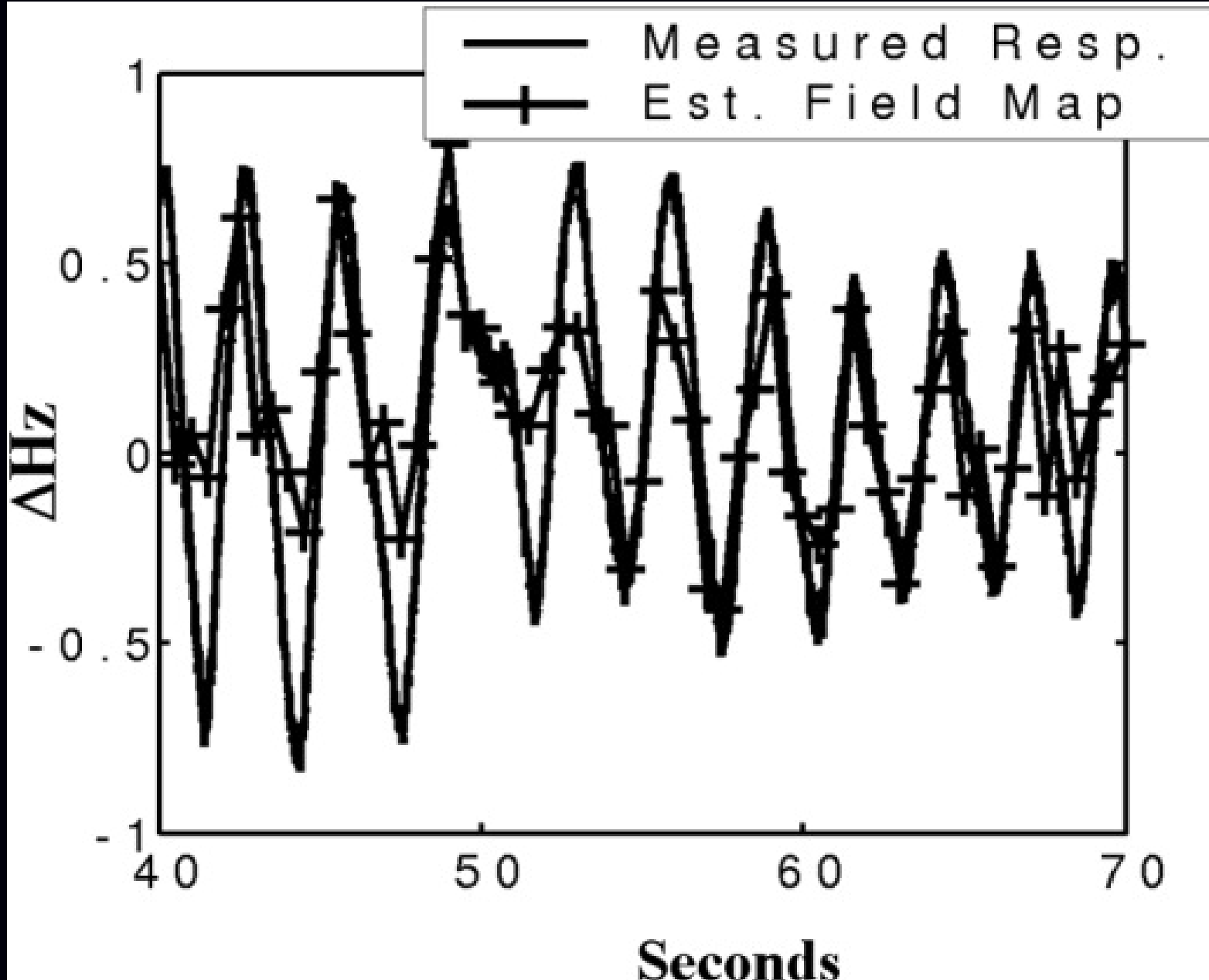
## Example: fMRI with Joint Estimation of Fieldmap



(a) uncorr., (b) std. map, (c) joint map, (d) T1 ref, (e) using std, (f) using joint.

PWLS-CG with quadratic regularization.  $\beta$  chosen by PSF analysis.  
Sutton *et al.*, MRM 51(6):1194, Jun. 2004

# Tracking Respiration-Induced Field Changes



## Other Topics

- Dynamic image sequence reconstruction / 4D regularization
- Motion and/or dynamic contrast changes

# Summary

- Iterative reconstruction has had clinical impact in PET and SPECT
- MRI and X-ray CT may be next?
- todo: Still work to be done...



## References

- [1] S. Webb. *From the watching of shadows: the origins of radiological tomography*. A. Hilger, Bristol, 1990.
- [2] H. H. Barrett and K. J. Myers. *Foundations of image science*. Wiley, New York, 2003.
- [3] J. Kay. The EM algorithm in medical imaging. *Stat. Meth. Med. Res.*, 6(1):55–75, January 1997.
- [4] J. A. Fessler. Statistical image reconstruction methods for transmission tomography. In M. Sonka and J. Michael Fitzpatrick, editors, *Handbook of Medical Imaging, Volume 2. Medical Image Processing and Analysis*, pages 1–70. SPIE, Bellingham, 2000.
- [5] R. M. Leahy and J. Qi. Statistical approaches in quantitative positron emission tomography. *Statistics and Computing*, 10(2):147–65, April 2000.
- [6] M. Defrise. A short reader's guide to 3D tomographic reconstruction. *Computerized Medical Imaging and Graphics*, 25(2):113–6, March 2001.
- [7] S. Vandenberghe, Y. D'Asseler, R. V. Walle, T. Kauppinen, M. Koole, L. Bouwens, K. V. Laere, I. Lemahieu, and R. A. Dierckx. Iterative reconstruction algorithms in nuclear medicine. *Computerized Medical Imaging and Graphics*, 25(2):105–11, March 2001.
- [8] G. L. Zeng. Image reconstruction, a tutorial. *Computerized Medical Imaging and Graphics*, 25(2):97–103, March 2001.
- [9] R. M. Lewitt and S. Matej. Overview of methods for image reconstruction from projections in emission computed tomography. *Proc. IEEE*, 91(10):1588–611, October 2003.
- [10] R. N. Bracewell. Strip integration in radio astronomy. *Aust. J. Phys.*, 9:198–217, 1956.
- [11] D. E. Kuhl and R. Q. Edwards. Image separation radioisotope scanning. *Radiology*, 80:653–62, 1963.
- [12] G. Hounsfield. A method of apparatus for examination of a body by radiation such as x-ray or gamma radiation, 1972. US Patent 1283915. British patent 1283915, London.
- [13] D. C. Solmon. The X-ray transform. *J. Math. Anal. Applic.*, 56(1):61–83, October 1976.
- [14] G. Muehlethner and R. A. Wetzel. Section imaging by computer calculation. *J. Nuc. Med.*, 12(2):76–85, February 1971.
- [15] D. E. Kuhl, R. Q. Edwards, A. R. Ricci, and M. Reivich. Quantitative section scanning using orthogonal tangent correction. *J. Nuc. Med.*, 14(4):196–200, April 1973.
- [16] R. Gordon, R. Bender, and G. T. Herman. Algebraic reconstruction techniques (ART) for the three-dimensional electron microscopy and X-ray photography. *J. Theor. Biol.*, 29(3):471–81, December 1970.
- [17] R. Gordon and G. T. Herman. Reconstruction of pictures from their projections. *Comm. ACM*, 14(12):759–68, December 1971.
- [18] G. T. Herman, A. Lent, and S. W. Rowland. ART: mathematics and applications (a report on the mathematical foundations and on the applicability to real data of the algebraic reconstruction techniques). *J. Theor. Biol.*, 42(1):1–32, November 1973.
- [19] R. Gordon. A tutorial on ART (algebraic reconstruction techniques). *IEEE Trans. Nuc. Sci.*, 21(3):78–93, June 1974.
- [20] W. H. Richardson. Bayesian-based iterative method of image restoration. *J. Opt. Soc. Am.*, 62(1):55–9, January 1972.
- [21] L. Lucy. An iterative technique for the rectification of observed distributions. *The Astronomical Journal*, 79(6):745–54, June 1974.
- [22] A. J. Rockmore and A. Macovski. A maximum likelihood approach to emission image reconstruction from projections. *IEEE Trans. Nuc. Sci.*, 23:1428–32, 1976.
- [23] A. J. Rockmore and A. Macovski. A maximum likelihood approach to transmission image reconstruction from projections. *IEEE Trans. Nuc. Sci.*, 24(3):1929–35, June 1977.
- [24] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Royal Stat. Soc. Ser. B*, 39(1):1–38, 1977.
- [25] L. A. Shepp and Y. Vardi. Maximum likelihood reconstruction for emission tomography. *IEEE Trans. Med. Imag.*, 1(2):113–22, October 1982.
- [26] K. Lange and R. Carson. EM reconstruction algorithms for emission and transmission tomography. *J. Comp. Assisted Tomo.*, 8(2):306–16, April 1984.
- [27] S. Geman and D. E. McClure. Bayesian image analysis: an application to single photon emission tomography. In *Proc. of Stat. Comp. Sect. of Amer. Stat. Assoc.*, pages 12–8, 1985.
- [28] H. M. Hudson and R. S. Larkin. Accelerated image reconstruction using ordered subsets of projection data. *IEEE Trans. Med. Imag.*, 13(4):601–9, December 1994.
- [29] M. Goitein. Three-dimensional density reconstruction from a series of two-dimensional projections. *Nucl. Instr. Meth.*, 101(3):509–18, June 1972.
- [30] T. F. Budinger and G. T. Gullberg. Three dimensional reconstruction in nuclear medicine emission imaging. *IEEE Trans. Nuc. Sci.*, 21(3):2–20, June 1974.
- [31] R. H. Huesman, G. T. Gullberg, W. L. Greenberg, and T. F. Budinger. *RECLBL library users manual*. Lawrence Berkeley Laboratory, Berkeley, CA, 1977.
- [32] R. H. Huesman. A new fast algorithm for the evaluation of regions of interest and statistical uncertainty in computed tomography. *Phys. Med. Biol.*, 29(5):543–52,



May 1984.

- [33] D. W. Wilson and B. M. W. Tsui. Noise properties of filtered-backprojection and ML-EM reconstructed emission tomographic images. *IEEE Trans. Nuc. Sci.*, 40(4):1198–1203, August 1993.
- [34] D. W. Wilson and B. M. W. Tsui. Spatial resolution properties of FB and ML-EM reconstruction methods. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 2, pages 1189–93, 1993.
- [35] H. H. Barrett, D. W. Wilson, and B. M. W. Tsui. Noise properties of the EM algorithm: I. Theory. *Phys. Med. Biol.*, 39(5):833–46, May 1994.
- [36] D. W. Wilson, B. M. W. Tsui, and H. H. Barrett. Noise properties of the EM algorithm: II. Monte Carlo simulations. *Phys. Med. Biol.*, 39(5):847–72, May 1994.
- [37] J. A. Fessler. Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): Applications to tomography. *IEEE Trans. Im. Proc.*, 5(3):493–506, March 1996.
- [38] J. A. Fessler and W. L. Rogers. Spatial resolution properties of penalized-likelihood image reconstruction methods: Space-invariant tomographs. *IEEE Trans. Im. Proc.*, 5(9):1346–58, September 1996.
- [39] W. Wang and G. Gindi. Noise analysis of regularized EM SPECT reconstruction. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 3, pages 1933–7, 1996.
- [40] C. K. Abbey, E. Clarkson, H. H. Barrett, S. P. Mueller, and F. J. Rybicki. Approximate distributions for maximum likelihood and maximum a posteriori estimates under a Gaussian noise model. In J. Duncan and G. Gindi, editors, *Information Processing in Medical Im.*, pages 167–75. Springer-Verlag, Berlin, 1997.
- [41] W. Wang and G. Gindi. Noise analysis of MAP-EM algorithms for emission tomography. *Phys. Med. Biol.*, 42(11):2215–32, November 1997.
- [42] S. J. Glick and E. J. Soares. Noise characteristics of SPECT iterative reconstruction with a mis-matched projector-backprojector pair. *IEEE Trans. Nuc. Sci.*, 45(4):2183–8, August 1998.
- [43] E. J. Soares, C. L. Byrne, T-S. Pan, S. J. Glick, and M. A. King. Modeling the population covariance matrices of block-iterative expectation-maximization reconstructed images. In *Proc. SPIE 3034, Med. Im. 1997: Im. Proc.*, pages 415–25, 1997.
- [44] J. Qi and R. H. Huesman. Theoretical study of lesion detectability of MAP reconstruction using computer observers. *IEEE Trans. Med. Imag.*, 20(8):815–22, August 2001.
- [45] D. Brasse, P. E. Kinahan, R. Clackdoyle, M. Defrise, C. Comtat, and D. W. Townsend. Fast fully 3-D image reconstruction in PET using planograms. *IEEE Trans. Med. Imag.*, 23(4):413–25, April 2004.
- [46] J. A. Fessler, I. Elbakri, P. Sukovic, and N. H. Clinthorne. Maximum-likelihood dual-energy tomographic image reconstruction. In *Proc. SPIE 4684, Medical Imaging 2002: Image Proc.*, volume 1, pages 38–49, 2002.
- [47] B. De Man, J. Nuyts, P. Dupont, G. Marchal, and P. Suetens. An iterative maximum-likelihood polychromatic algorithm for CT. *IEEE Trans. Med. Imag.*, 20(10):999–1008, October 2001.
- [48] I. A. Elbakri and J. A. Fessler. Statistical image reconstruction for polyenergetic X-ray computed tomography. *IEEE Trans. Med. Imag.*, 21(2):89–99, February 2002.
- [49] I. A. Elbakri and J. A. Fessler. Segmentation-free statistical image reconstruction for polyenergetic X-ray computed tomography with experimental validation. *Phys. Med. Biol.*, 48(15):2543–78, August 2003.
- [50] P. E. Kinahan, J. A. Fessler, and J. S. Karp. Statistical image reconstruction in PET with compensation for missing data. *IEEE Trans. Nuc. Sci.*, 44(4):1552–7, August 1997.
- [51] J. A. Fessler and B. P. Sutton. Nonuniform fast Fourier transforms using min-max interpolation. *IEEE Trans. Sig. Proc.*, 51(2):560–74, February 2003.
- [52] B. P. Sutton, D. C. Noll, and J. A. Fessler. Fast, iterative image reconstruction for MRI in the presence of field inhomogeneities. *IEEE Trans. Med. Imag.*, 22(2):178–88, February 2003.
- [53] B. P. Sutton, D. C. Noll, and J. A. Fessler. Dynamic field map estimation using a spiral-in / spiral-out acquisition. *Mag. Res. Med.*, 51(6):1194–204, June 2004.
- [54] R. Van de Walle, H. H. Barrett, K. J. Myers, M. I. Altbach, B. Desplanques, A. F. Gmitro, J. Cornelis, and I. Lemahieu. Reconstruction of MR images from data acquired on a general non-regular grid by pseudoinverse calculation. *IEEE Trans. Med. Imag.*, 19(12):1160–7, December 2000.
- [55] B. W. Silverman. Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *J. Royal Stat. Soc. Ser. B*, 47(1):1–52, 1985.
- [56] M. Bertero, C. De Mol, and E. R. Pike. Linear inverse problems with discrete data, I: General formulation and singular system analysis. *Inverse Prob.*, 1(4):301–30, November 1985.
- [57] K. M. Hanson and G. W. Wecksung. Local basis-function approach to computed tomography. *Appl. Optics*, 24(23):4028–39, December 1985.
- [58] E. J. Mazur and R. Gordon. Interpolative algebraic reconstruction techniques without beam partitioning for computed tomography. *Med. Biol. Eng. Comput.*, 33(1):82–6, January 1995.

- [59] Y. Censor. Finite series expansion reconstruction methods. *Proc. IEEE*, 71(3):409–19, March 1983.
- [60] D. L. Snyder. Utilizing side information in emission tomography. *IEEE Trans. Nuc. Sci.*, 31(1):533–7, February 1984.
- [61] R. E. Carson, M. V. Green, and S. M. Larson. A maximum likelihood method for calculation of tomographic region-of-interest (ROI) values. *J. Nuc. Med. (Abs. Book)*, 26:P20, 1985.
- [62] R. E. Carson and K. Lange. The EM parametric image reconstruction algorithm. *J. Am. Stat. Assoc.*, 80(389):20–2, March 1985.
- [63] R. E. Carson. A maximum likelihood method for region-of-interest evaluation in emission tomography. *J. Comp. Assisted Tomo.*, 10(4):654–63, July 1986.
- [64] A. R. Formiconi. Least squares algorithm for region-of-interest evaluation in emission tomography. *IEEE Trans. Med. Imag.*, 12(1):90–100, March 1993.
- [65] B. W. Reutter, G. T. Gullberg, and R. H. Huesman. Kinetic parameter estimation from attenuated SPECT projection measurements. *IEEE Trans. Nuc. Sci.*, 45(6):3007–13, December 1998.
- [66] D. J. Rossi and A. S. Willsky. Reconstruction from projections based on detection and estimation of objects—Parts I & II: Performance analysis and robustness analysis. *IEEE Trans. Acoust. Sp. Sig. Proc.*, 32(4):886–906, August 1984.
- [67] S. P. Müller, M. F. Kijewski, S. C. Moore, and B. L. Holman. Maximum-likelihood estimation: a mathematical model for quantitation in nuclear medicine. *J. Nuc. Med.*, 31(10):1693–701, October 1990.
- [68] P. C. Chiao, W. L. Rogers, N. H. Clinthorne, J. A. Fessler, and A. O. Hero. Model-based estimation for dynamic cardiac studies using ECT. *IEEE Trans. Med. Imag.*, 13(2):217–26, June 1994.
- [69] Z. P. Liang, F. E. Boada, R. T. Constable, E. M. Haacke, P. C. Lauterbur, and M. R. Smith. Constrained reconstruction methods in MR imaging. *Reviews of Magnetic Resonance in Medicine*, 4:67–185, 1992.
- [70] G. S. Cunningham and A. Lehovitch. 4D reconstructions from low-count SPECT data using deformable models with smooth interior intensity variations. In *Proc. SPIE 3979: Medical Imaging 2000: Image Proc.*, pages 564–74, 2000.
- [71] A. Yendiki and J. A. Fessler. A comparison of rotation- and blob-based system models for 3D SPECT with depth-dependent detector response. *Phys. Med. Biol.*, 49(11):2157–68, June 2004.
- [72] R. M. Lewitt. Multidimensional digital image representations using generalized Kaiser-Bessel window functions. *J. Opt. Soc. Am. A*, 7(10):1834–46, October 1990.
- [73] R. M. Lewitt. Alternatives to voxels for image representation in iterative reconstruction algorithms. *Phys. Med. Biol.*, 37(3):705–16, March 1992.
- [74] G. L. Zeng and G. T. Gullberg. Unmatched projector/backprojector pairs in an iterative reconstruction algorithm. *IEEE Trans. Med. Imag.*, 19(5):548–55, May 2000.
- [75] C. Kamphuis, F. J. Beekman, P. P. van Rijk, and M. A. Viergever. Dual matrix ordered subsets reconstruction for accelerated 3D scatter compensation in single-photon emission tomography. *Eur. J. Nuc. Med.*, 25(1):8–18, January 1998.
- [76] F. J. Beekman, H. W. A. M. de Jong, and S. van Geloven. Efficient fully 3D iterative SPECT reconstruction with Monte Carlo based scatter compensation. *IEEE Trans. Med. Imag.*, 21(8):867–77, August 2002.
- [77] R. Griesse and A. Walther. Evaluating gradients in optimal control: continuous adjoints versus automatic differentiation. *J. Optim. Theory Appl.*, 122(1):63–86, July 2004.
- [78] C. W. Stearns and J. A. Fessler. 3D PET reconstruction with FORE and WLS-OS-EM. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 2, pages 912–5, 2002.
- [79] M. Yavuz and J. A. Fessler. Objective functions for tomographic reconstruction from randoms-precorrected PET scans. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 2, pages 1067–71, 1996.
- [80] M. Yavuz and J. A. Fessler. New statistical models for randoms-precorrected PET scans. In J. Duncan and G. Gindi, editors, *Information Processing in Medical Im.*, volume 1230 of *Lecture Notes in Computer Science*, pages 190–203. Springer-Verlag, Berlin, 1997.
- [81] M. Yavuz and J. A. Fessler. Statistical image reconstruction methods for randoms-precorrected PET scans. *Med. Im. Anal.*, 2(4):369–78, December 1998.
- [82] M. Yavuz and J. A. Fessler. Penalized-likelihood estimators and noise analysis for randoms-precorrected PET transmission scans. *IEEE Trans. Med. Imag.*, 18(8):665–74, August 1999.
- [83] D. L. Snyder, A. M. Hammoud, and R. L. White. Image recovery from data acquired with a charge-coupled-device camera. *J. Opt. Soc. Am. A*, 10(5):1014–23, May 1993.
- [84] D. L. Snyder, C. W. Helstrom, A. D. Lanterman, M. Faisal, and R. L. White. Compensation for readout noise in CCD images. *J. Opt. Soc. Am. A*, 12(2):272–83, February 1995.
- [85] U. Engeland, T. Striker, and H. Luig. Count-rate statistics of the gamma camera. *Phys. Med. Biol.*, 43(10):2939–47, October 1998.

- [86] J. A. Fessler. Penalized weighted least-squares image reconstruction for positron emission tomography. *IEEE Trans. Med. Imag.*, 13(2):290–300, June 1994.
- [87] C. Comtat, P. E. Kinahan, M. Defrise, C. Michel, and D. W. Townsend. Fast reconstruction of 3D PET data with accurate statistical modeling. *IEEE Trans. Nuc. Sci.*, 45(3):1083–9, June 1998.
- [88] B. R. Whiting. Signal statistics in x-ray computed tomography. In *Proc. SPIE 4682, Medical Imaging 2002: Med. Phys.*, pages 53–60, 2002.
- [89] B. R. Whiting, L. J. Montagnino, and D. G. Politte. Modeling X-ray computed tomography sinograms, 2001. submitted to mp.
- [90] I. A. Elbakri and J. A. Fessler. Efficient and accurate likelihood for iterative image reconstruction in X-ray computed tomography. In *Proc. SPIE 5032, Medical Imaging 2003: Image Proc.*, pages 1839–50, 2003.
- [91] A. O. Hero, J. A. Fessler, and M. Usman. Exploring estimator bias-variance tradeoffs using the uniform CR bound. *IEEE Trans. Sig. Proc.*, 44(8):2026–41, August 1996.
- [92] Y. C. Eldar. Minimum variance in biased estimation: bounds and asymptotically optimal estimators. *IEEE Trans. Sig. Proc.*, 52(7):1915–30, July 2004.
- [93] E. Mumcuoglu, R. Leahy, S. Cherry, and E. Hoffman. Accurate geometric and physical response modeling for statistical image reconstruction in high resolution PET scanners. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 3, pages 1569–73, 1996.
- [94] J. Qi, R. M. Leahy, S. R. Cherry, A. Chatziioannou, and T. H. Farquhar. High resolution 3D Bayesian image reconstruction using the microPET small-animal scanner. *Phys. Med. Biol.*, 43(4):1001–14, April 1998.
- [95] G. Christ. Exact treatment of the dual-energy method in CT using polyenergetic X-ray spectra. *Phys. Med. Biol.*, 29(12):1511–25, December 1984.
- [96] Z. Liang. Compensation for attenuation, scatter, and detector response in SPECT reconstruction via iterative FBP methods. *Med. Phys.*, 20(4):1097–106, July 1993.
- [97] X. L. Xu, J. S. Liow, and S. C. Strother. Iterative algebraic reconstruction algorithms for emission computed tomography: a unified framework and its application to positron emission tomography. *Med. Phys.*, 20(6):1675–84, November 1993.
- [98] J. W. Wallis and T. R. Miller. Rapidly converging iterative reconstruction algorithms in single-photon emission computed tomography. *J. Nuc. Med.*, 34(10):1793–800, October 1993.
- [99] P. J. Green. Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives. *J. Royal Stat. Soc. Ser. B*, 46(2):149–92, 1984.
- [100] J. M. M. Anderson, B. A. Mair, M. Rao, and C. H. Wu. A weighted least-squares method for PET. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 2, pages 1292–6, 1995.
- [101] J. M. M. Anderson, B. A. Mair, M. Rao, and C-H. Wu. Weighted least-squares reconstruction methods for positron emission tomography. *IEEE Trans. Med. Imag.*, 16(2):159–65, April 1997.
- [102] P. J. Huber. *Robust statistics*. Wiley, New York, 1981.
- [103] C. Bouman and K. Sauer. A generalized Gaussian image model for edge-preserving MAP estimation. *IEEE Trans. Im. Proc.*, 2(3):296–310, July 1993.
- [104] E. Tanaka. Improved iterative image reconstruction with automatic noise artifact suppression. *IEEE Trans. Med. Imag.*, 11(1):21–7, March 1992.
- [105] B. W. Silverman, M. C. Jones, J. D. Wilson, and D. W. Nychka. A smoothed EM approach to indirect estimation problems, with particular reference to stereology and emission tomography. *J. Royal Stat. Soc. Ser. B*, 52(2):271–324, 1990.
- [106] T. R. Miller, J. W. Wallis, C. S. Butler, M. I. Miller, and D. L. Snyder. Improved brain SPECT by maximum-likelihood reconstruction. *J. Nuc. Med. (Abs. Book)*, 33(5):964, May 1992.
- [107] F. J. Beekman, E. T. P. Slijpen, and W. J. Niessen. Selection of task-dependent diffusion filters for the post-processing of SPECT images. *Phys. Med. Biol.*, 43(6):1713–30, June 1998.
- [108] D. S. Lalush and B. M. W. Tsui. Performance of ordered subset reconstruction algorithms under conditions of extreme attenuation and truncation in myocardial SPECT. *J. Nuc. Med.*, 41(4):737–44, April 2000.
- [109] E. T. P. Slijpen and F. J. Beekman. Comparison of post-filtering and filtering between iterations for SPECT reconstruction. *IEEE Trans. Nuc. Sci.*, 46(6):2233–8, December 1999.
- [110] R. H. Huesman. The effects of a finite number of projection angles and finite lateral sampling of projections on the propagation of statistical errors in transverse section reconstruction. *Phys. Med. Biol.*, 22(3):511–21, May 1977.
- [111] D. L. Snyder and M. I. Miller. The use of sieves to stabilize images produced with the EM algorithm for emission tomography. *IEEE Trans. Nuc. Sci.*, 32(5):3864–71, October 1985.
- [112] D. L. Snyder, M. I. Miller, L. J. Thomas, and D. G. Politte. Noise and edge artifacts in maximum-likelihood reconstructions for emission tomography. *IEEE Trans.*

- Med. Imag.*, 6(3):228–38, September 1987.
- [113] T. R. Miller and J. W. Wallis. Clinically important characteristics of maximum-likelihood reconstruction. *J. Nuc. Med.*, 33(9):1678–84, September 1992.
- [114] A. Tikhonov and V. Arsenin. *Solution of ill-posed problems*. Wiley, New York, 1977.
- [115] I. Csizsár. Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems. *Ann. Stat.*, 19(4):2032–66, 1991.
- [116] D. L. Donoho, I. M. Johnstone, A. S. Stern, and J.C. Hoch. Does the maximum entropy method improve sensitivity. *Proc. Natl. Acad. Sci.*, 87(13):5066–8, July 1990.
- [117] D. L. Donoho, I. M. Johnstone, J. C. Hoch, and A. S. Stern. Maximum entropy and the nearly black object. *J. Royal Stat. Soc. Ser. B*, 54(1):41–81, 1992.
- [118] R. T. Constable and R. M. Henkelman. Why MEM does not work in MR image reconstruction. *Mag. Res. Med.*, 14(1):12–25, April 1990.
- [119] A. R. De Pierro. A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography. *IEEE Trans. Med. Imag.*, 14(1):132–7, March 1995.
- [120] A. H. Delaney and Y. Bresler. A fast and accurate Fourier algorithm for iterative parallel-beam tomography. *IEEE Trans. Im. Proc.*, 5(5):740–53, May 1996.
- [121] S. J. Lee, A. Rangarajan, and G. Gindi. A comparative study of the effects of using higher order mechanical priors in SPECT reconstruction. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 4, pages 1696–1700, 1994.
- [122] S.-J. Lee, A. Rangarajan, and G. Gindi. Bayesian image reconstruction in SPECT using higher order mechanical models as priors. *IEEE Trans. Med. Imag.*, 14(4):669–80, December 1995.
- [123] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and Bayesian restoration of images. *IEEE Trans. Patt. Anal. Mach. Int.*, 6(6):721–41, November 1984.
- [124] B. W. Silverman, C. Jennison, J. Stander, and T. C. Brown. The specification of edge penalties for regular and irregular pixel images. *IEEE Trans. Patt. Anal. Mach. Int.*, 12(10):1017–24, October 1990.
- [125] V. E. Johnson, W. H. Wong, X. Hu, and C. T. Chen. Image restoration using Gibbs priors: Boundary modeling, treatment of blurring, and selection of hyperparameter. *IEEE Trans. Patt. Anal. Mach. Int.*, 13(5):413–25, May 1991.
- [126] V. E. Johnson. A model for segmentation and analysis of noisy images. *J. Am. Stat. Assoc.*, 89(425):230–41, March 1994.
- [127] S. Alenius and U. Ruotsalainen. Bayesian image reconstruction for emission tomography based on median root prior. *Eur. J. Nuc. Med.*, 24(3):258–65, March 1997.
- [128] S. Alenius, U. Ruotsalainen, and J. Astola. Using local median as the location of the prior distribution in iterative emission tomography image reconstruction. *IEEE Trans. Nuc. Sci.*, 45(6):3097–104, December 1998.
- [129] W. Chlewicki, F. Hermansen, and S. B. Hansen. Noise reduction and convergence of Bayesian algorithms with blobs based on the huber function and median root prior. *Phys. Med. Biol.*, 49(20):4717–30, October 2004.
- [130] V. Y. Panin, G. L. Zeng, and G. T. Gullberg. Total variation regulated EM algorithm. *IEEE Trans. Nuc. Sci.*, 46(6):2202–10, December 1999.
- [131] P. Kisilev, M. Zibulevsky, and Y. Zeevi. Wavelet representation and total variation regularization in emission tomography. In *Proc. IEEE Intl. Conf. on Image Processing*, volume 1, pages 702–5, 2001.
- [132] C. R. Vogel and M. E. Oman. Fast numerical methods for total variation minimization in image reconstruction. In *Proc. SPIE 2563, Adv. Signal Proc. Alg.*, pages 359–67, 1995.
- [133] M. Lassas and S. Siltanen. Can one use total variation prior for edge-preserving Bayesian inversion? *Inverse Prob.*, 20(5):1537–1564, October 2004.
- [134] D. Donoho. Superresolution via sparsity constraints. *SIAM J. Math. Anal.*, 23(5):1309–31, 1993.
- [135] G. Harikumar and Y. Bresler. A new algorithm for computing sparse solutions to linear inverse problems. In *Proc. IEEE Conf. Acoust. Speech Sig. Proc.*, volume 3, pages 1331–4, 1996.
- [136] I. F. Gorodnitsky and B. D. Rao. Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm. *IEEE Trans. Sig. Proc.*, 45(3):600–16, March 1997.
- [137] D. Donoho. For most large underdetermined systems of linear equations, the minimal  $\ell_1$  norm solution is also the sparsest solution. *Comm. Pure Appl. Math.*, 59(6):797–829, June 2006.
- [138] E. J. Candés, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Info. Theory*, 52(2):489–509, February 2006.
- [139] D. L. Donoho. Compressed sensing. *IEEE Trans. Info. Theory*, 52(4):1289–1306, April 2006.
- [140] D. L. Donoho, M. Elad, and V. N. Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans. Info. Theory*,

- 52(1):6–18, January 2006.
- [141] M. Figueiredo, R. Nowak, and S. J. Wright. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE J. Sel. Top. Sig. Proc.*, 1(4):586–97, December 2007.
- [142] E. Candès and J. Romberg. Sparsity and incoherence in compressive sampling. *Inverse Prob.*, 23(3):969–86, June 2007.
- [143] R. G. Baraniuk. Compressive sensing. *IEEE Sig. Proc. Mag.*, 24(4):118–21, 2007.
- [144] I-T. Hsiao, A. Rangarajan, and G. Gindi. A new convex edge-preserving median prior with applications to tomography. *IEEE Trans. Med. Imag.*, 22(5):580–5, May 2003.
- [145] M. Nikolova. Thresholding implied by truncated quadratic regularization. *IEEE Trans. Sig. Proc.*, 48(12):3437–50, December 2000.
- [146] A. Antoniadis and J. Fan. Regularization and wavelet approximations. *J. Am. Stat. Assoc.*, 96(455):939–55, September 2001.
- [147] D. F. Yu and J. A. Fessler. Edge-preserving tomographic reconstruction with nonlocal regularization. In *Proc. IEEE Intl. Conf. on Image Processing*, volume 1, pages 29–33, 1998.
- [148] D. F. Yu and J. A. Fessler. Edge-preserving tomographic reconstruction with nonlocal regularization. *IEEE Trans. Med. Imag.*, 21(2):159–73, February 2002.
- [149] J. Ye, Y. Bresler, and P. Moulin. A self-referencing level-set method for image reconstruction from sparse Fourier samples. In *Proc. IEEE Intl. Conf. on Image Processing*, volume 2, pages 33–6, 2001.
- [150] M. J. Black and A. Rangarajan. On the unification of line processes, outlier rejection, and robust statistics with applications in early vision. *Intl. J. Comp. Vision*, 19(1):57–91, July 1996.
- [151] J. W. Stayman and J. A. Fessler. Regularization for uniform spatial resolution properties in penalized-likelihood image reconstruction. *IEEE Trans. Med. Imag.*, 19(6):601–15, June 2000.
- [152] J. W. Stayman and J. A. Fessler. Nonnegative definite quadratic penalty design for penalized-likelihood reconstruction. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 2, pages 1060–3, 2001.
- [153] J. W. Stayman and J. A. Fessler. Compensation for nonuniform resolution using penalized-likelihood reconstruction in space-variant imaging systems. *IEEE Trans. Med. Imag.*, 23(3):269–84, March 2004.
- [154] C. T. Chen, X. Ouyang, W. H. Wong, and X. Hu. Improvement of PET image reconstruction using high-resolution anatomic images. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 3, page 2062, 1991. (Abstract.)
- [155] R. Leahy and X. H. Yan. Statistical models and methods for PET image reconstruction. In *Proc. of Stat. Comp. Sect. of Amer. Stat. Assoc.*, pages 1–10, 1991.
- [156] J. A. Fessler, N. H. Clinthorne, and W. L. Rogers. Regularized emission image reconstruction using imperfect side information. *IEEE Trans. Nuc. Sci.*, 39(5):1464–71, October 1992.
- [157] I. G. Zubal, M. Lee, A. Rangarajan, C. R. Harrell, and G. Gindi. Bayesian reconstruction of SPECT images using registered anatomical images as priors. *J. Nuc. Med. (Abs. Book)*, 33(5):963, May 1992.
- [158] G. Gindi, M. Lee, A. Rangarajan, and I. G. Zubal. Bayesian reconstruction of functional images using anatomical information as priors. *IEEE Trans. Med. Imag.*, 12(4):670–80, December 1993.
- [159] X. Ouyang, W. H. Wong, V. E. Johnson, X. Hu, and C-T. Chen. Incorporation of correlated structural images in PET image reconstruction. *IEEE Trans. Med. Imag.*, 13(4):627–40, December 1994.
- [160] S. J. Lee, G. R. Gindi, I. G. Zubal, and A. Rangarajan. Using ground-truth data to design priors in Bayesian SPECT reconstruction. In Y. Bizais, C. Barillot, and R. D. Paola, editors, *Information Processing in Medical Im.* Kluwer, 1995.
- [161] J. E. Bowsher, V. E. Johnson, T. G. Turkington, R. J. Jaszczak, C. E. Floyd, and R. E. Coleman. Bayesian reconstruction and use of anatomical a priori information for emission tomography. *IEEE Trans. Med. Imag.*, 15(5):673–86, October 1996.
- [162] S. Sastry and R. E. Carson. Multimodality Bayesian algorithm for image reconstruction in positron emission tomography: a tissue composition model. *IEEE Trans. Med. Imag.*, 16(6):750–61, December 1997.
- [163] R. Piramuthu and A. O. Hero. Side information averaging method for PML emission tomography. In *Proc. IEEE Intl. Conf. on Image Processing*, volume 2, pages 671–5, 1998.
- [164] C. Comtat, P. E. Kinahan, J. A. Fessler, T. Beyer, D. W. Townsend, M. Defrise, and C. Michel. Reconstruction of 3d whole-body PET data using blurred anatomical labels. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 3, pages 1651–5, 1998.
- [165] A. O. Hero, R. Piramuthu, S. R. Titus, and J. A. Fessler. Minimax emission computed tomography using high resolution anatomical side information and B-spline models. *IEEE Trans. Info. Theory*, 45(3):920–38, April 1999.

- [166] D. F. Yu and J. A. Fessler. Mean and variance of singles photon counting with deadtime. *Phys. Med. Biol.*, 45(7):2043–56, July 2000.
- [167] D. F. Yu and J. A. Fessler. Mean and variance of coincidence photon counting with deadtime. *Nucl. Instr. Meth. Phys. Res. A.*, 488(1-2):362–74, August 2002.
- [168] J. Qi and R. H. Huesman. Propagation of errors from the sensitivity image in list mode reconstruction. *IEEE Trans. Med. Imag.*, 23(9):1094–9, September 2004.
- [169] J. Qi and R. H. Huesman. Effect of errors in the system matrix on iterative image reconstruction. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 5, pages 2854–8, 2004.
- [170] Y. S. Shim and Z. H. Cho. SVD pseudoinversion image reconstruction. *IEEE Trans. Acoust. Sp. Sig. Proc.*, 29(4):904–9, August 1981.
- [171] U. Raff, D. N. Stroud, and W. R. Hendee. Improvement of lesion detection in scintigraphic images by SVD techniques for resolution recovery. *IEEE Trans. Med. Imag.*, 5(1):35–44, March 1986.
- [172] D. A. Fish, J. Grochmalicki, and E. R. Pike. Scanning SVD method for restoration of images with space-variant blur. *J. Opt. Soc. Am. A*, 13(3):464–9, March 1996.
- [173] A. Caponnetto and M. Bertero. Tomography with a finite set of projections: singular value decomposition and resolution. *Inverse Prob.*, 13(5):1191–1205, October 1997.
- [174] A. K. Louis. Incomplete data problems in x-ray computerized tomography. I. Singular value decomposition of the limited angle transform. *Numerische Mathematik*, 48(3):251–62, May 1986.
- [175] F. Natterer. Numerical treatment of ill-posed problems. In G. Talenti, editor, *Inverse Prob.*, volume 1225, pages 142–67. Berlin, Springer, 1986. Lecture Notes in Math.
- [176] R. C. Liu and L. D. Brown. Nonexistence of informative unbiased estimators in singular problems. *Ann. Stat.*, 21(1):1–13, March 1993.
- [177] J. Ory and R. G. Pratt. Are our parameter estimators biased? The significance of finite-different regularization operators. *Inverse Prob.*, 11(2):397–424, April 1995.
- [178] I. M. Johnstone. On singular value decompositions for the Radon Transform and smoothness classes of functions. Technical Report 310, Dept. of Statistics, Stanford, January 1989.
- [179] M. F. Smith, C. E. Floyd, R. J. Jaszczak, and R. E. Coleman. Reconstruction of SPECT images using generalized matrix inverses. *IEEE Trans. Med. Imag.*, 11(2):165–75, June 1992.
- [180] M. Lavielle. A stochastic algorithm for parametric and non-parametric estimation in the case of incomplete data. *Signal Processing*, 42(1):3–17, 1995.
- [181] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical recipes in C*. Cambridge Univ. Press, New York, 1988.
- [182] K. Sauer and C. Bouman. Bayesian estimation of transmission tomograms using local optimization operations. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 3, pages 2089–93, 1991.
- [183] K. Sauer and C. Bouman. Bayesian estimation of transmission tomograms using segmentation based optimization. *IEEE Trans. Nuc. Sci.*, 39(4):1144–52, August 1992.
- [184] D. P. Bertsekas and S. K. Mitter. A descent numerical method for optimization problems with nondifferentiable cost functionals. *SIAM J. Control*, 11(4):637–52, 1973.
- [185] W. C. Davidon. Variable metric methods for minimization. Technical Report ANL-5990, AEC Research and Development Report, Argonne National Laboratory, USA, 1959.
- [186] H. F. Khalfan, R. H. Byrd, and R. B. Schnabel. A theoretical and experimental study of the symmetric rank-one update. *SIAM J. Optim.*, 3(1):1–24, February 1993.
- [187] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal. Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Trans. Math. Software*, 23(4):550–60, December 1997.
- [188] T. G. Kolda, D. P. O’Leary, and L. Nazareth. BFGS with update skipping and varying memory. *SIAM J. Optim.*, 8(4):1060–83, 1998.
- [189] K. Lange. *Numerical analysis for statisticians*. Springer-Verlag, New York, 1999.
- [190] B. T. Polyak. *Introduction to optimization*. Optimization Software Inc, New York, 1987.
- [191] D. P. Bertsekas. *Constrained optimization and Lagrange multiplier methods*. Academic-Press, New York, 1982.
- [192] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu. A limited memory algorithm for bound constrained optimization. *SIAM J. Sci. Comp.*, 16(5):1190–208, 1995.
- [193] L. Kaufman. Reduced storage, Quasi-Newton trust region approaches to function optimization. *SIAM J. Optim.*, 10(1):56–69, 1999.
- [194] M. Hanke, J. G. Nagy, and C. Vogel. Quasi-Newton approach to nonnegative image restorations. *Linear Algebra and its Applications*, 316(1):223–36, September 2000.
- [195] J. L. Morales and J. Nocedal. Automatic preconditioning by limited memory Quasi-Newton updating. *SIAM J. Optim.*, 10(4):1079–96, 2000.
- [196] R. R. Meyer. Sufficient conditions for the convergence of monotonic mathematical programming algorithms. *J. Comput. System. Sci.*, 12(1):108–21, 1976.

- [197] L. Kaufman. Implementing and accelerating the EM algorithm for positron emission tomography. *IEEE Trans. Med. Imag.*, 6(1):37–51, March 1987.
- [198] N. H. Clinthorne, T. S. Pan, P. C. Chiao, W. L. Rogers, and J. A. Stamos. Preconditioning methods for improved convergence rates in iterative reconstructions. *IEEE Trans. Med. Imag.*, 12(1):78–83, March 1993.
- [199] J. A. Fessler and S. D. Booth. Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction. *IEEE Trans. Im. Proc.*, 8(5):688–99, May 1999.
- [200] E. U. Mumcuoglu, R. Leahy, S. R. Cherry, and Z. Zhou. Fast gradient-based methods for Bayesian reconstruction of transmission and emission PET images. *IEEE Trans. Med. Imag.*, 13(3):687–701, December 1994.
- [201] E. Ü Mumcuoğlu, R. M. Leahy, and S. R. Cherry. Bayesian reconstruction of PET images: methodology and performance analysis. *Phys. Med. Biol.*, 41(9):1777–1807, September 1996.
- [202] J. A. Fessler and A. O. Hero. Space-alternating generalized expectation-maximization algorithm. *IEEE Trans. Sig. Proc.*, 42(10):2664–77, October 1994.
- [203] J. A. Fessler and A. O. Hero. Penalized maximum-likelihood image reconstruction using space-alternating generalized EM algorithms. *IEEE Trans. Im. Proc.*, 4(10):1417–29, October 1995.
- [204] J. A. Fessler, E. P. Ficaro, N. H. Clinthorne, and K. Lange. Grouped-coordinate ascent algorithms for penalized-likelihood transmission image reconstruction. *IEEE Trans. Med. Imag.*, 16(2):166–75, April 1997.
- [205] J. A. Fessler and A. O. Hero. Space-alternating generalized EM algorithms for penalized maximum-likelihood image reconstruction. Technical Report 286, Comm. and Sign. Proc. Lab., Dept. of EECS, Univ. of Michigan, Ann Arbor, MI, 48109-2122, February 1994.
- [206] J. A. Browne and A. R. De Pierro. A row-action alternative to the EM algorithm for maximizing likelihoods in emission tomography. *IEEE Trans. Med. Imag.*, 15(5):687–99, October 1996.
- [207] C. L. Byrne. Block-iterative methods for image reconstruction from projections. *IEEE Trans. Im. Proc.*, 5(5):792–3, May 1996.
- [208] C. L. Byrne. Convergent block-iterative algorithms for image reconstruction from inconsistent data. *IEEE Trans. Im. Proc.*, 6(9):1296–304, September 1997.
- [209] C. L. Byrne. Accelerating the EMMML algorithm and related iterative algorithms by rescaled block-iterative methods. *IEEE Trans. Im. Proc.*, 7(1):100–9, January 1998.
- [210] M. E. Daube-Witherspoon and G. Muehlelehner. An iterative image space reconstruction algorithm suitable for volume ECT. *IEEE Trans. Med. Imag.*, 5(2):61–66, June 1986.
- [211] J. M. Ollinger. Iterative reconstruction-reprojection and the expectation-maximization algorithm. *IEEE Trans. Med. Imag.*, 9(1):94–8, March 1990.
- [212] A. R. De Pierro. On the relation between the ISRA and the EM algorithm for positron emission tomography. *IEEE Trans. Med. Imag.*, 12(2):328–33, June 1993.
- [213] P. J. Green. Bayesian reconstructions from emission tomography data using a modified EM algorithm. *IEEE Trans. Med. Imag.*, 9(1):84–93, March 1990.
- [214] P. J. Green. On use of the EM algorithm for penalized likelihood estimation. *J. Royal Stat. Soc. Ser. B*, 52(3):443–452, 1990.
- [215] K. Sauer and C. Bouman. A local update strategy for iterative reconstruction from projections. *IEEE Trans. Sig. Proc.*, 41(2):534–48, February 1993.
- [216] K. G. Murty. *Linear complementarity, linear and nonlinear programming*. Helderman Verlag, Berlin, 1988.
- [217] C. A. Bouman, K. Sauer, and S. S. Saquib. Tractable models and efficient algorithms for Bayesian tomography. In *Proc. IEEE Conf. Acoust. Speech Sig. Proc.*, volume 5, pages 2907–10, 1995.
- [218] C. A. Bouman and K. Sauer. A unified approach to statistical tomography using coordinate descent optimization. *IEEE Trans. Im. Proc.*, 5(3):480–92, March 1996.
- [219] J. A. Fessler. Hybrid Poisson/polynomial objective functions for tomographic image reconstruction from transmission scans. *IEEE Trans. Im. Proc.*, 4(10):1439–50, October 1995.
- [220] H. Erdoğan and J. A. Fessler. Monotonic algorithms for transmission tomography. *IEEE Trans. Med. Imag.*, 18(9):801–14, September 1999.
- [221] E. P. Ficaro, J. A. Fessler, R. J. Ackerman, W. L. Rogers, J. R. Corbett, and M. Schwaiger. Simultaneous transmission-emission TI-201 cardiac SPECT: Effect of attenuation correction on myocardial tracer distribution. *J. Nuc. Med.*, 36(6):921–31, June 1995.
- [222] E. P. Ficaro, J. A. Fessler, P. D. Shreve, J. N. Kritzman, P. A. Rose, and J. R. Corbett. Simultaneous transmission/emission myocardial perfusion tomography: Diagnostic accuracy of attenuation-corrected 99m-Tc-Sestamibi SPECT. *Circulation*, 93(3):463–73, February 1996.
- [223] J-B. Thibault, K. Sauer, C. Bouman, and J. Hsieh. A three-dimensional statistical approach to improved image quality for multi-slice helical CT. *Med. Phys.*, 34(11):4526–44, November 2007.
- [224] C. A. Johnson, J. Seidel, and A. Sofer. Interior point methodology for 3-D PET reconstruction. *IEEE Trans. Med. Imag.*, 19(4):271–85, April 2000.
- [225] K. Lange and J. A. Fessler. Globally convergent algorithms for maximum a posteriori transmission tomography. *IEEE Trans. Im. Proc.*, 4(10):1430–8, October 1995.

- [226] J. A. Fessler, N. H. Clinthorne, and W. L. Rogers. On complete data spaces for PET reconstruction algorithms. *IEEE Trans. Nuc. Sci.*, 40(4):1055–61, August 1993.
- [227] J. A. Fessler and H. Erdođan. A paraboloidal surrogates algorithm for convergent penalized-likelihood emission image reconstruction. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 2, pages 1132–5, 1998.
- [228] T. Hebert and R. Leahy. A Bayesian reconstruction algorithm for emission tomography using a Markov random field prior. In *Proc. SPIE 1092, Med. Im. III: Im. Proc.*, pages 458–66, 1989.
- [229] T. Hebert and R. Leahy. A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors. *IEEE Trans. Med. Imag.*, 8(2):194–202, June 1989.
- [230] T. J. Hebert and R. Leahy. Statistic-based MAP image reconstruction from Poisson data using Gibbs priors. *IEEE Trans. Sig. Proc.*, 40(9):2290–303, September 1992.
- [231] E. Tanaka. Utilization of non-negativity constraints in reconstruction of emission tomograms. In S L Bacharach, editor, *Information Processing in Medical Im.*, pages 379–93. Martinus-Nijhoff, Boston, 1985.
- [232] R. M. Lewitt and G. Muehlelehner. Accelerated iterative reconstruction for positron emission tomography based on the EM algorithm for maximum likelihood estimation. *IEEE Trans. Med. Imag.*, 5(1):16–22, March 1986.
- [233] T. Hebert, R. Leahy, and M. Singh. Three-dimensional maximum-likelihood reconstruction for an electronically collimated single-photon-emission imaging system. *J. Opt. Soc. Am. A*, 7(7):1305–13, July 1990.
- [234] S. Holte, P. Schmidlin, A. Lindén, G. Rosenqvist, and L. Eriksson. Iterative image reconstruction for emission tomography: A study of convergence and quantitation problems. *IEEE Trans. Nuc. Sci.*, 37(2):629–35, April 1990.
- [235] D. P. Bertsekas. A new class of incremental gradient methods for least squares problems. *SIAM J. Optim.*, 7(4):913–26, November 1997.
- [236] R. Neal and G. E. Hinton. A view of the EM algorithm that justifies incremental, sparse and other variants. In M. I. Jordan, editor, *Learning in Graphical Models*, pages 255–68. Kluwer, Dordrecht, 1998.
- [237] A. Nedic and D. Bertsekas. Convergence rate of incremental subgradient algorithms. In S. Uryasev and P. M. Pardalos, editors, *Stochastic Optimization: Algorithms and Applications*, pages 263–304. Kluwer, New York, 2000.
- [238] A. Nedic, D. Bertsekas, and V. Borkar. Distributed asynchronous incremental subgradient methods. In D. Butnariu, Y. Censor, and S. Reich, editors, *Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications*. Elsevier, Amsterdam, 2000.
- [239] A. Nedic and D. P. Bertsekas. Incremental subgradient methods for nondifferentiable optimization. *SIAM J. Optim.*, 12(1):109–38, 2001.
- [240] V. M. Kibardin. Decomposition into functions in the minimization problem. *Avtomatika i Telemekhanika*, 9:66–79, September 1979. Translation: p. 1311-23 in Plenum Publishing Co. "Adaptive Systems".
- [241] H. Kudo, H. Nakazawa, and T. Saito. Convergent block-iterative method for general convex cost functions. In *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, pages 247–250, 1999.
- [242] A. R. De Pierro and M. E. B. Yamagishi. Fast EM-like methods for maximum 'a posteriori' estimates in emission tomography. *IEEE Trans. Med. Imag.*, 20(4):280–8, April 2001.
- [243] S. Ahn and J. A. Fessler. Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms. *IEEE Trans. Med. Imag.*, 22(5):613–26, May 2003.
- [244] S. Ahn and J. A. Fessler. Globally convergent ordered subsets algorithms: Application to tomography. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 2, pages 1064–8, 2001.
- [245] P. Khurd, I-T. Hsiao, A. Rangarajan, and G. Gindi. A globally convergent regularized ordered-subset EM algorithm for list-mode reconstruction. *IEEE Trans. Nuc. Sci.*, 51(3):719–25, June 2004.
- [246] S. Ahn, J. A. Fessler, D. Blatt, and A. O. Hero. Convergent incremental optimization transfer algorithms: Application to tomography. *IEEE Trans. Med. Imag.*, 25(3):283–96, March 2006.
- [247] S. Ahn, J. A. Fessler, D. Blatt, and A. O. Hero. Incremental surrogates algorithms: Application to transmission tomography. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 5, pages 2835–9, 2004.
- [248] H. Erdođan and J. A. Fessler. Ordered subsets algorithms for transmission tomography. *Phys. Med. Biol.*, 44(11):2835–51, November 1999.
- [249] J. W. Stayman and J. A. Fessler. Spatially-variant roughness penalty design for uniform resolution in penalized-likelihood image reconstruction. In *Proc. IEEE Intl. Conf. on Image Processing*, volume 2, pages 685–9, 1998.



- [250] J. Nuyts and J. A. Fessler. A penalized-likelihood image reconstruction method for emission tomography, compared to post-smoothed maximum-likelihood with matched spatial resolution. *IEEE Trans. Med. Imag.*, 22(9):1042–52, September 2003.
- [251] J. Qi and R. M. Leahy. Resolution and noise properties of MAP reconstruction for fully 3D PET. In *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med.*, pages 35–9, 1999.
- [252] T. H. Farquhar, J. Llacer, C. K. Hoh, J. Czernin, S. S. Gambhir, M. A. Seltzer, D. H. Silverman, J. Qi, C. Hsu, and E. J. Hoffman. ROC and localization ROC analyses of lesion detection in whole-body FDG PET: effects of acquisition mode, attenuation correction and reconstruction algorithm. *J. Nuc. Med.*, 40(12):2043–52, December 1999.
- [253] P. Bonetto, J. Qi, and R. M. Leahy. Covariance approximation for fast and accurate computation of channelized Hotelling observer statistics. *IEEE Trans. Nuc. Sci.*, 47(4):1567–72, August 2000.
- [254] J. Qi. Theoretical evaluation of the detectability of random lesions in Bayesian emission reconstruction. In *Information Processing in Medical Im.*, pages 354–65, 2003.
- [255] P. K. Khurd and G. R. Gindi. LROC model observers for emission tomographic reconstruction. In *Proc. SPIE 5372, Medical Imaging 2004: Image Perception, Observer Performance, and Technology Assessment*, pages 509–20, 2004.
- [256] J. Qi and R. H. Huesman. Fast approach to evaluate MAP reconstruction for lesion detection and localization. In *Proc. SPIE 5372, Medical Imaging 2004: Image Perception, Observer Performance, and Technology Assessment*, pages 273–82, 2004.
- [257] J. Qi. Analysis of lesion detectability in Bayesian emission reconstruction with nonstationary object variability. *IEEE Trans. Med. Imag.*, 23(3):321–9, March 2004.
- [258] A. Yendiki and J. A. Fessler. Analysis of observer performance in unknown-location tasks for tomographic image reconstruction. *J. Opt. Soc. Am. A*, 24(12):B99–109, December 2007. Special issue on Image Quality.
- [259] D. L. Snyder. Parameter estimation for dynamic studies in emission-tomography systems having list-mode data. *IEEE Trans. Nuc. Sci.*, 31(2):925–31, April 1984.
- [260] J. M. Ollinger and D. L. Snyder. A preliminary evaluation of the use of the EM algorithm for estimating parameters in dynamic tracer-studies. *IEEE Trans. Nuc. Sci.*, 32(1):848–54, February 1985.
- [261] J. M. Ollinger and D. L. Snyder. An evaluation of an improved method for computing histograms in dynamic tracer studies using positron-emission tomography. *IEEE Trans. Nuc. Sci.*, 33(1):435–8, February 1986.
- [262] J. M. Ollinger. Estimation algorithms for dynamic tracer studies using positron-emission tomography. *IEEE Trans. Med. Imag.*, 6(2):115–25, June 1987.
- [263] J. M. Ollinger. An evaluation of a practical algorithm for estimating histograms in dynamic tracer studies using positron-emission tomography. *IEEE Trans. Nuc. Sci.*, 34(1):349–53, February 1987.
- [264] F. O’Sullivan. Imaging radiotracer model parameters in PET: a mixture analysis approach. *IEEE Trans. Med. Imag.*, 12(3):399–412, September 1993.
- [265] P. C. Chiao, W. L. Rogers, J. A. Fessler, N. H. Clinthorne, and A. O. Hero. Model-based estimation with boundary side information or boundary regularization. *IEEE Trans. Med. Imag.*, 13(2):227–34, June 1994.
- [266] M. A. Limber, M. N. Limber, A. Celler, J. S. Barney, and J. M. Borwein. Direct reconstruction of functional parameters for dynamic SPECT. *IEEE Trans. Nuc. Sci.*, 42(4):1249–56, August 1995.
- [267] G. L. Zeng, G. T. Gullberg, and R. H. Huesman. Using linear time-invariant system theory to estimate kinetic parameters directly from projection measurements. *IEEE Trans. Nuc. Sci.*, 42(6-2):2339–46, December 1995.
- [268] J. M. Borwein and W. Sun. The stability analysis of dynamic SPECT systems. *Numerische Mathematik*, 77(3):283–98, September 1997.
- [269] R. H. Huesman, B. W. Reutter, G. L. Zeng, and G. T. Gullberg. Kinetic parameter estimation from SPECT cone-beam projection measurements. *Phys. Med. Biol.*, 43(4):973–82, April 1998.
- [270] J. S. Maltz, E. Polak, and T. F. Budinger. Multistart optimization algorithm for joint spatial and kinetic parameter estimation from dynamic ECT projection data. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 3, pages 1567–73, 1998.
- [271] J. S. Maltz. Direct recovery of regional tracer kinetics from temporally inconsistent dynamic ECT projections using dimension-reduced time-activity basis. *Phys. Med. Biol.*, 45(11):3413–29, November 2000.
- [272] E. Hebbler, D. Oldenburg, T. Farncombe, and A. Celler. Direct estimation of dynamic parameters in SPECT tomography. *IEEE Trans. Nuc. Sci.*, 44(6-2):2425–30, December 1997.
- [273] D. S. Lalush and B. M. W. Tsui. Block-iterative techniques for fast 4D reconstruction using a priori motion models in gated cardiac SPECT. *Phys. Med. Biol.*, 43(4):875–86, April 1998.
- [274] H. H. Bauschke, D. Noll, A. Celler, and J. M. Borwein. An EM algorithm for dynamic SPECT. *IEEE Trans. Med. Imag.*, 18(3):252–61, March 1999.

- [275] T. Farncombe, A. Celler, D. Noll, J. Maeght, and R. Harrop. Dynamic SPECT imaging using a single camera rotation (dSPECT). *IEEE Trans. Nuc. Sci.*, 46(4-2):1055–61, August 1999.
- [276] T. E. Nichols, J. Qi, and R. M. Leahy. Continuous time dynamic PET imaging using list mode data. In A. Todd-Pokropek A. Kuba, M. Smal, editor, *Information Processing in Medical Im.*, pages 98–111. Springer, Berlin, 1999.
- [277] E. Asma, T. E. Nichols, J. Qi, and R. M. Leahy. 4D PET image reconstruction from list mode data. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 2, pages 15/57–65, 2000.
- [278] B. W. Reutter, G. T. Gullberg, and R. H. Huesman. Direct least squares estimation of spatiotemporal distributions from dynamic SPECT projections using a spatial segmentation and temporal B-splines. *IEEE Trans. Med. Imag.*, 19(5):434–50, May 2000.
- [279] T. E. Nichols, J. Qi, E. Asma, and R. M. Leahy. Spatiotemporal reconstruction of list mode PET data. *IEEE Trans. Med. Imag.*, 21(4):396–404, April 2002.
- [280] U. Schmitt and A. K. Louis. Efficient algorithms for the regularization of dynamic inverse problems: I. Theory. *Inverse Prob.*, 18(3):645–58, June 2002.
- [281] U. Schmitt, A. K. Louis, C. Wolters, and M. Vauhkonen. Efficient algorithms for the regularization of dynamic inverse problems: II. Applications. *Inverse Prob.*, 18(3):659–76, June 2002.
- [282] C-M. Kao, J. T. Yap, J. Mukherjee, and M. N. Wernick. Image reconstruction for dynamic PET based on low-order approximation and restoration of the sinogram. *IEEE Trans. Med. Imag.*, 16(6):727–37, December 1997.
- [283] J. Matthews, D. Bailey, P. Price, and V. Cunningham. The direct calculation of parametric images from dynamic PET data using maximum-likelihood iterative reconstruction. *Phys. Med. Biol.*, 42(6):1155–73, June 1997.
- [284] S. R. Meikle, J. C. Matthews, V. J. Cunningham, D. L. Bailey, L. Livieratos, T. Jones, and P. Price. Parametric image reconstruction using spectral analysis of PET projection data. *Phys. Med. Biol.*, 43(3):651–66, March 1998.
- [285] M. V. Narayanan, M. A. King, E. J. Soares, C. L. Byrne, P. H. Pretorius, and M. N. Wernick. Application of the Karhunen-Loeve transform to 4D reconstruction of cardiac gated SPECT images. *IEEE Trans. Nuc. Sci.*, 46(4-2):1001–8, August 1999.
- [286] M. N. Wernick, E. J. Infusino, and M. Milosevic. Fast spatio-temporal image reconstruction for dynamic PET. *IEEE Trans. Med. Imag.*, 18(3):185–95, March 1999.
- [287] M. V. Narayanan, M. A. King, M. N. Wernick, C. L. Byrne, E. J. Soares, and P. H. Pretorius. Improved image quality and computation reduction in 4-D reconstruction of cardiac-gated SPECT images. *IEEE Trans. Med. Imag.*, 19(5):423–33, May 2000.
- [288] J. S. Maltz. Optimal time-activity basis selection for exponential spectral analysis: application to the solution of large dynamic emission tomographic reconstruction problems. *IEEE Trans. Nuc. Sci.*, 48(4-2):1452–64, August 2001.
- [289] J. E. Koss, D. L. Kirch, E. P. Little, T. K. Johnson, and P. P. Steele. Advantages of list-mode acquisition of dynamic cardiac data. *IEEE Trans. Nuc. Sci.*, 44(6-2):2431–8, December 1997.

The literature on image reconstruction is enormous and growing. Many valuable publications are not included in this list, which is not intended to be comprehensive.

Slides and lecture notes available from:

<http://www.eecs.umich.edu/~fessler>