Iterative image reconstruction for CT

Jeffrey A. Fessler

EECS Dept., BME Dept., Dept. of Radiology
University of Michigan

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

AAPM Image Educational Course - Image Reconstruction II

Aug. 2, 2011

Full disclosure

- Research support from GE Healthcare
- Research support to GE Global Research
- Work supported in part by NIH grant R01-HL-098686
- Research support from Intel

Credits

Current students / post-docs
- Jang Hwan Cho
- Se Young Chun
- Donghwan Kim
- Yong Long
- Madison McGaffin
- Sathish Ramani
- Stephen Schmitt

GE collaborators
- Jiang Hsieh
- Jean-Baptiste Thibault
- Bruno De Man

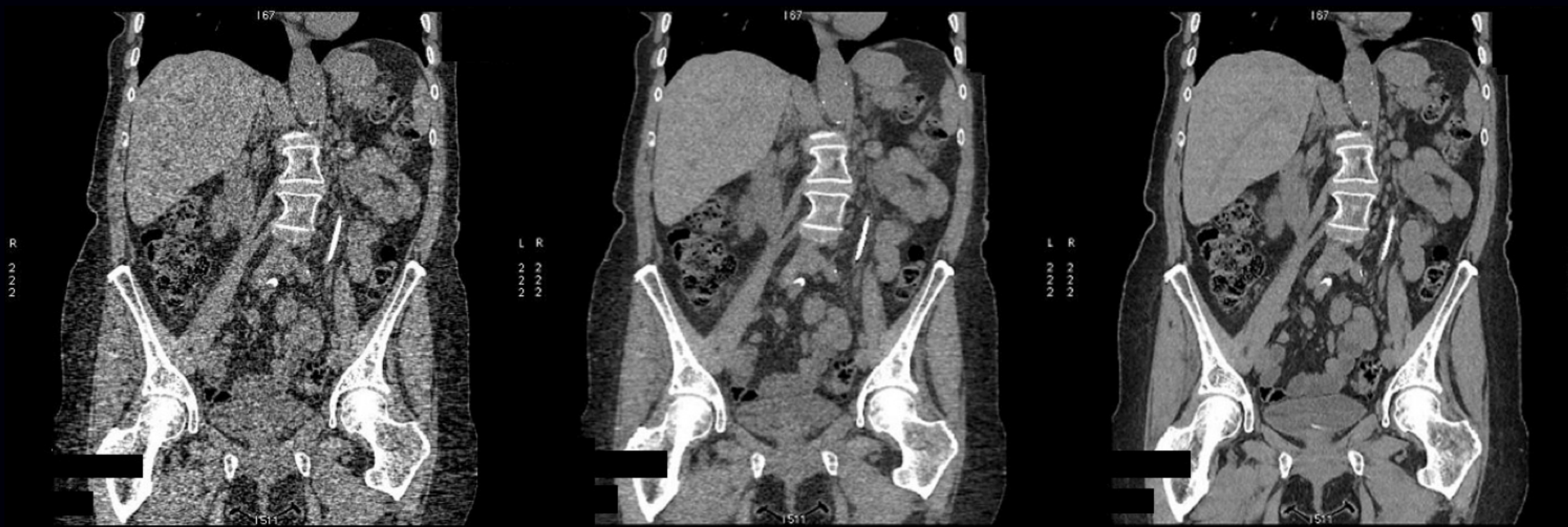
CT collaborators
- Mitch Goodsitt, UM
- Ella Kazerooni, UM
- Neal Clinthorne, UM
- Paul Kinahan, UW

Former PhD students (who did/do CT)
- Wonseok Huh, Bain & Company
- Hugo Shi, Enthought
- Joonki Noh, Emory
- Somesh Srivastava, JHU
- Rongping Zeng, FDA
- Yingying, Zhang-O'Connor, RGM Advisors
- Matthew Jacobson, Xoran
- Sangtae Ahn, GE
- Idris Elbakri, CancerCare / Univ. of Manitoba
- Saowapak Sotthivirat, NSTDA Thailand
- Web Stayman, JHU
- Feng Yu, Univ. Bristol
- Mehmet Yavuz, Qualcomm
- Hakan Erdoğan, Sabanci University

Former MS / undegraduate students
- Kevin Brown, Philips
- Meng Wu, Stanford
- ...

Statistical image reconstruction: CT revolution

- A picture is worth 1000 words
- (and perhaps several 1000 seconds of computation?)



Thin-slice FBP	ASIR	Statistical
Seconds	A bit longer	Much longer

Why statistical/iterative methods for CT?

- Accurate physics models
  - X-ray spectrum, beam-hardening, scatter, ...
    $\Longrightarrow$ reduced artifacts? quantitative CT?
  - X-ray detector spatial response, focal spot size, ...
    $\Longrightarrow$ improved spatial resolution?
  - detector spectral response (*e.g.*, photon-counting detectors)
    $\Longrightarrow$ improved contrast?

- Nonstandard geometries
  - transaxial truncation (wide patients)
  - long-object problem in helical CT
  - irregular sampling in "next-generation" geometries
  - coarse angular sampling in image-guidance applications
  - limited angular range (tomosynthesis)
  - "missing" data, *e.g.*, bad pixels in flat-panel systems

- Appropriate models of measurement statistics
  - weighting reduces influence of photon-starved rays (*cf.* FBP)
    $\Longrightarrow$ reducing image noise or X-ray dose

and more...

- Object constraints
  - nonnegativity
  - object support
  - piecewise smoothness
  - object sparsity (*e.g.*, angiography)
  - sparsity in some basis
  - motion models
  - dynamic models
  - ...

Disadvantages?

- Computation time (super computer)
- Must reconstruct entire FOV
- Model complexity
- Software complexity
- Algorithm nonlinearities
  - Difficult to analyze resolution/noise properties (*cf.* FBP)
  - Tuning parameters
  - Challenging to characterize performance

“Iterative” vs “Statistical”

- Traditional *successive substitutions* iterations
  - *e.g.*, Joseph and Spital (JCAT, 1978) bone correction
  - usually only one or two “iterations”
  - not statistical

- Algebraic reconstruction methods
  - Given sinogram data $\boldsymbol{y}$ and system model $\boldsymbol{A}$, reconstruct object $\boldsymbol{x}$ by “solving” $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$
  - ART, SIRT, SART, ...
  - iterative, but typically not statistical
  - Iterative filtered back-projection (FBP):

$$\boldsymbol{x}^{(n+1)} = \boldsymbol{x}^{(n)} + \underbrace{\alpha}_{\substack{\text{step}\\\text{size}}} \mathrm{FBP}(\underbrace{\boldsymbol{y}}_{\text{data}} - \underbrace{\boldsymbol{A}\boldsymbol{x}^{(n)}}_{\substack{\text{forward}\\\text{project}}})$$

- Statistical reconstruction methods
  - Image domain
  - Sinogram domain
  - Fully statistical (both)
  - Hybrid methods (*e.g.*, AIR, SPIE 7961-18, Bruder *et al.*)

"Statistical" methods: Image domain

- Denoising methods

$$\text{sinogram } \boldsymbol{y} \rightarrow \boxed{\text{FBP}} \rightarrow \text{noisy reconstruction } \tilde{\boldsymbol{x}} \rightarrow \boxed{\text{iterative denoiser}} \rightarrow \text{final image } \hat{\boldsymbol{x}}$$

  - Relatively fast, even if iterative
  - Remarkable advances in denoising methods in last decade


Zhu & Milanfar, T-IP, Dec. 2010, using "steering kernel regression" (SKR) method

Challenges:

  - Typically assume *white noise*
  - Streaks in low-dose FBP appear like edges (highly correlated noise)

- Image denoising methods "guided by data statistics"

$$\begin{tikzcd} \text{sinogram } \boldsymbol{y} \arrow[r] & \boxed{\text{FBP}} \arrow[r] & \text{noisy reconstruction } \tilde{\boldsymbol{x}} \arrow[r] & \boxed{\text{magical iterative denoiser}} \arrow[r] & \text{final image } \hat{\boldsymbol{x}} \\ & & & \text{sinogram statistics?} \arrow[u] & \end{tikzcd}$$

  - Image-domain methods are fast (thus very practical)
  - ASIR? IRIS? ...
  - The technical details are often a mystery...

Challenges:

  - FBP often does not use all data efficiently (*e.g.*, Parker weighting)
  - Low-dose CT statistics most naturally expressed in sinogram domain

“Statistical” methods: Sinogram domain

- Sinogram restoration methods

noisy sinogram $\boldsymbol{y}$ $\rightarrow$ $\boxed{\text{adaptive or iterative denoiser}}$ $\rightarrow$ cleaned sinogram $\hat{\boldsymbol{y}}$ $\rightarrow$ $\boxed{\text{FBP}}$ $\rightarrow$ final image $\hat{\boldsymbol{x}}$

  - Adaptive: J. Hsieh, Med. Phys., 1998; Kachelrieß, Med. Phys., 2001, ...
  - Iterative: P. La Riviere, IEEE T-MI, 2000, 2005, 2006, 2008
  - Relatively fast even if iterative

Challenges:

  - Limited denoising without resolution loss
  - Difficult to “preserve edges” in sinograms

FBP, 10 mA

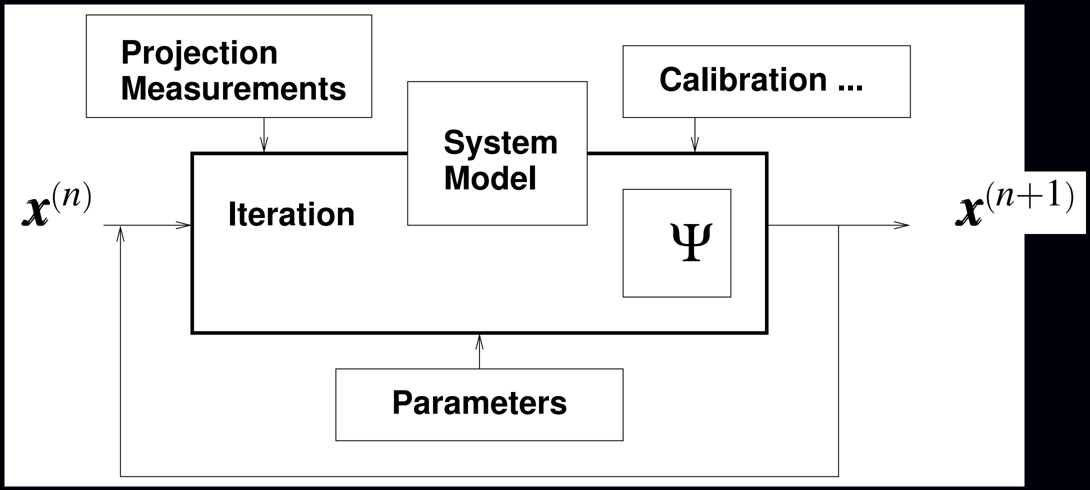
FBP from denoised sinogram

Wang *et al.*, T-MI, Oct. 2006, using PWLS-GS on sinogram

(True? Fully? Slow?) Statistical image reconstruction

- Object model
- Physics/system model
- Statistical model
- Cost function (log-likelihood + regularization)
- Iterative algorithm for minimization

"Find the image $\hat{\boldsymbol{x}}$ that best fits the sinogram data $\boldsymbol{y}$ according to the physics model, the statistical model and prior information about the object"



- Repeatedly revisiting the sinogram data can use statistics fully
- Repeatedly updating the image can exploit object properties
- $\therefore$ greatest potential dose reduction, but repetition is expensive...

History: Statistical reconstruction for PET

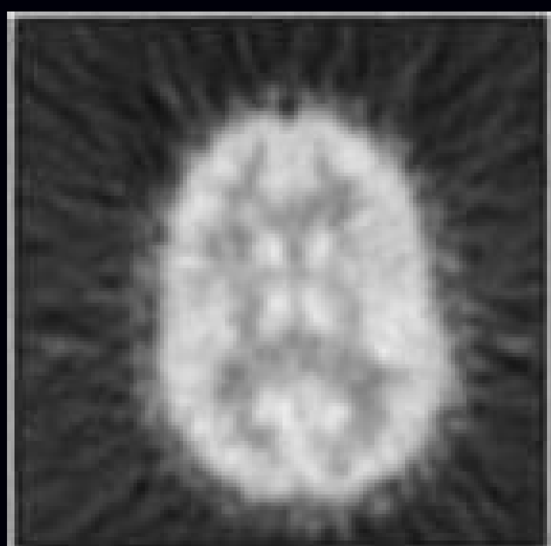
- Iterative method for emission tomography (Kuhl, 1963)
- FBP for PET (Chesler, 1971)
- Weighted least squares for 3D SPECT (Goitein, NIM, 1972)
- Richardson/Lucy iteration for image restoration (1972, 1974)
- Poisson likelihood (emission) (Rockmore and Macovski, TNS, 1976)
- Expectation-maximization (EM) algorithm (Shepp and Vardi, TMI, 1982)
- Regularized (aka Bayesian) Poisson emission reconstruction (Geman and McClure, ASA, 1985)
- Ordered-subsets EM (OSEM) algorithm (Hudson and Larkin, TMI, 1994)
- Commercial release of OSEM for PET scanners circa 1997

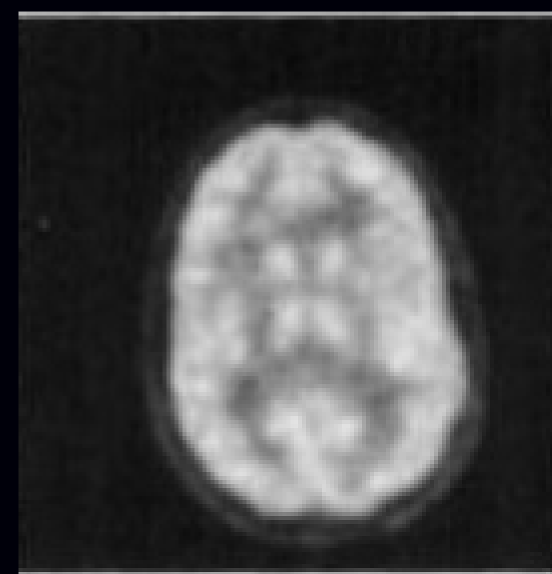
Today, most (all?) commercial PET systems include *unregularized* OSEM.

15 years between key EM paper (1982) and commercial adoption (1997)
(25 years if you count the R/L paper in 1972 which is the same as EM)

Key factors in PET

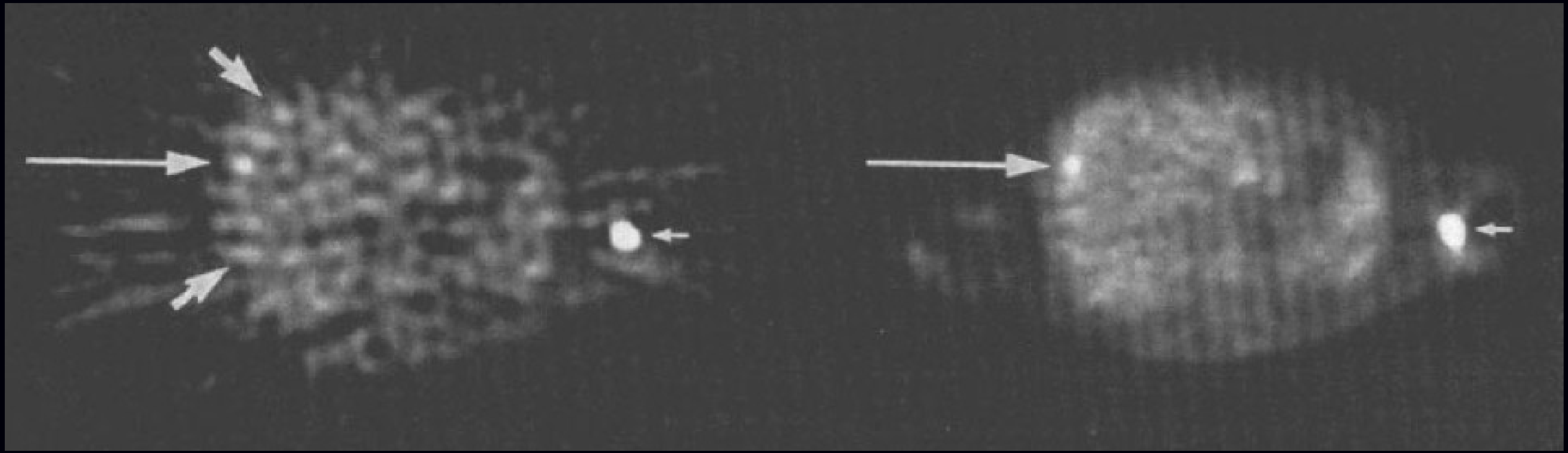
- OS algorithm accelerated convergence by order of magnitude
- Computers got faster (but problem size grew too)
- Key clinical validation papers?
- Key numerical observer studies?
- Nuclear medicine physicians grew accustomed to appearance of images reconstructed using statistical methods

FBP: 

ML-EM: 

Llacer *et al.*, 1993

Whole-body PET example



FBP ML-OSEM

Meikle *et al.*, 1994

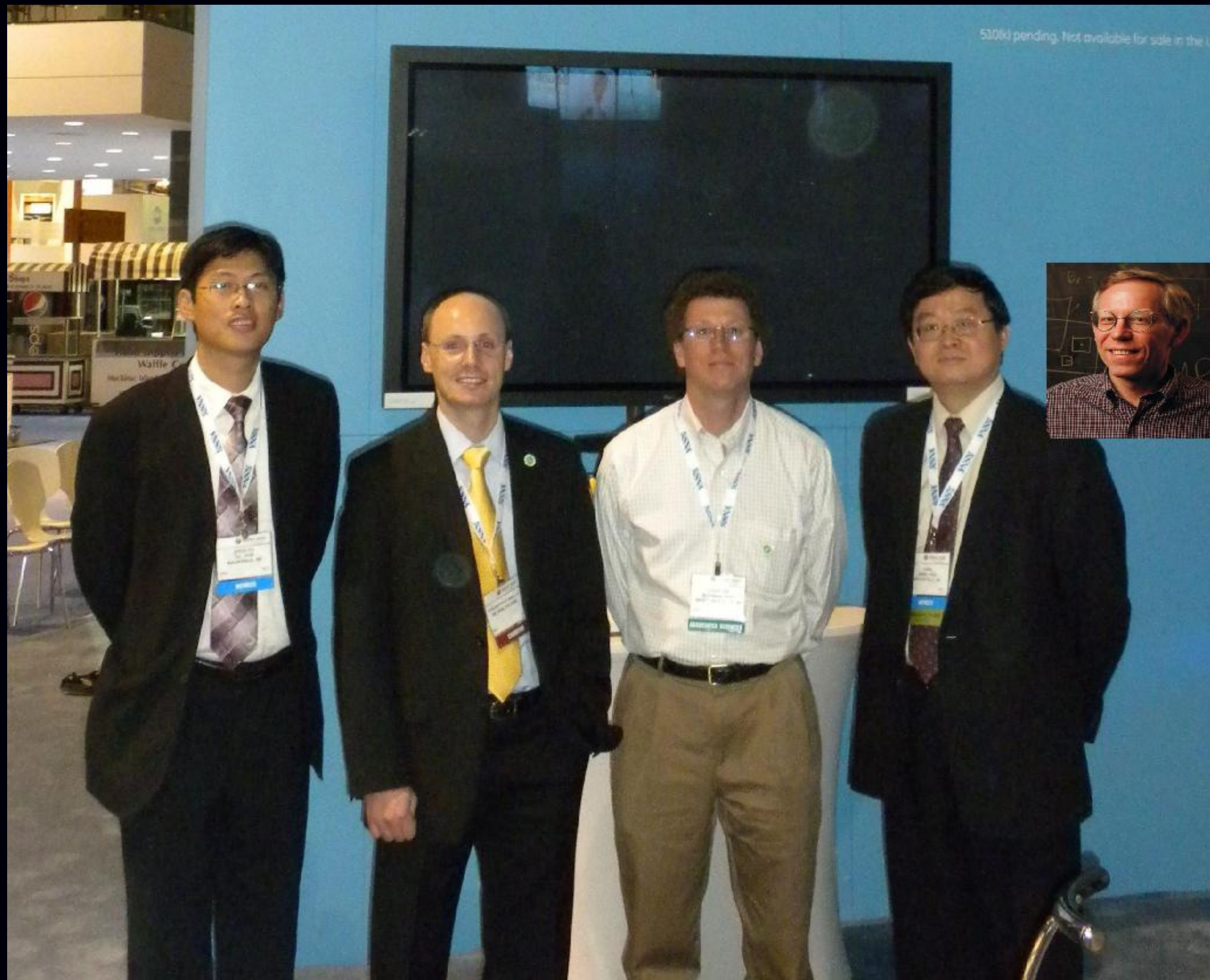
Key factor in PET: modeling measurement statistics

History: Statistical reconstruction for CT*

- Iterative method for X-ray CT (Hounsfield, 1968)
- ART for tomography (Gordon, Bender, Herman, JTB, 1970)
- ...
- Roughness regularized LS for tomography (Kashyap & Mittal, 1975)
- Poisson likelihood (transmission) (Rockmore and Macovski, TNS, 1977)
- EM algorithm for Poisson transmission (Lange and Carson, JCAT, 1984)
- Iterative coordinate descent (ICD) (Sauer and Bouman, T-SP, 1993)
- Ordered-subsets algorithms
  (Manglos *et al.*, PMB 1995)
  (Kamphuis & Beekman, T-MI, 1998)
  (Erdoğan & Fessler, PMB, 1999)
- ...
- Commercial introduction of ICD for CT scanners circa 2010

(* numerous omissions, including the many denoising methods)

RSNA 2010



Zhou Yu, Jean-Baptiste Thibault, Charles Bouman, Jiang Hsieh, Ken Sauer

https://engineering.purdue.edu/BME/AboutUs/News/HomepageFeatures/ResultsofPurdueResearchUnveiledatRSNA

MBIR example: Routine chest CT

Helical chest CT study with dose = 0.09 mSv.
Typical CXR effective dose is about 0.06 mSv. Source: Health Physics Society.
`http://www.hps.org/publicinformation/ate/q2372.html`

FBP

MBIR

Veo (MBIR) is 510(k) pending. Not available for sale in the U.S.
Images courtesy of Jiang Hsieh, GE Healthcare

Five Choices for Statistical Image Reconstruction

1. Object model
2. System physical model
3. Measurement statistical model
4. Cost function: data-mismatch and regularization
5. Algorithm / initialization

No perfect choices - one can critique all approaches!

Historically these choices are often left implicit in publications,
but being explicit facilitates reproducibility.

Choice 1. Object Parameterization

Finite measurements: $\{y_i\}_{i=1}^{M}$. Continuous object: $f(\vec{r}) = \mu(\vec{r})$.

"All models are wrong but some models are useful."

Linear *series expansion* approach. Represent $f(\vec{r})$ by $\boldsymbol{x} = (x_1, \ldots, x_N)$ where

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

Reconstruction problem becomes "discrete-discrete:" estimate $\boldsymbol{x}$ from $\boldsymbol{y}$

Numerous basis functions in literature. Two primary contenders:

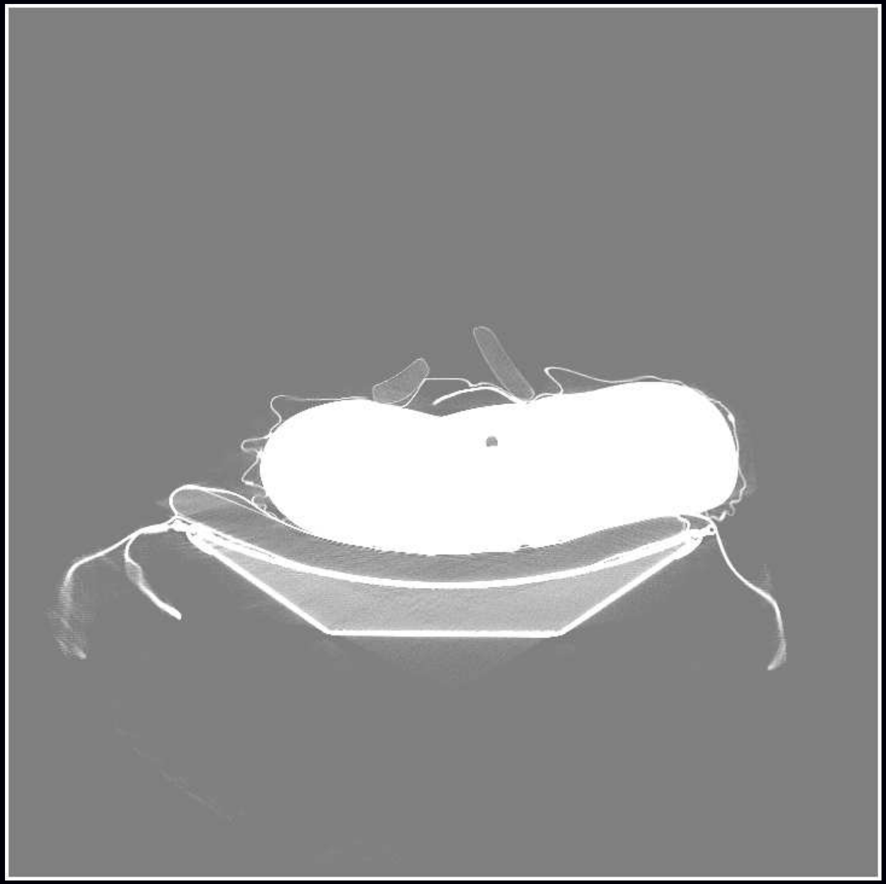
- voxels
- blobs (Kaiser-Bessel functions)
  - \+ Blobs are approximately band-limited (reduced aliasing?)
  - – Blobs have larger footprints, increasing computation.

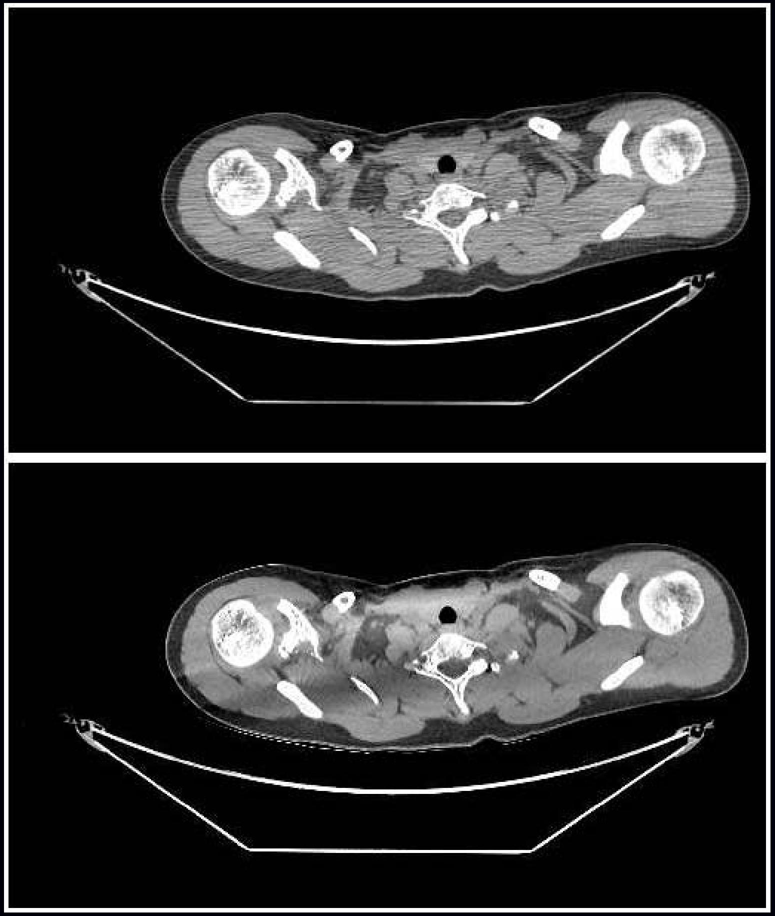
Open question: how small should the voxels be?

One practical compromise: wide FOV coarse-grid reconstruction followed by fine-grid refinement over ROI, *e.g.*, Ziegler *et al.*, Med. Phys., Apr. 2008

Global reconstruction: An inconvenient truth

70-cm FOV reconstruction

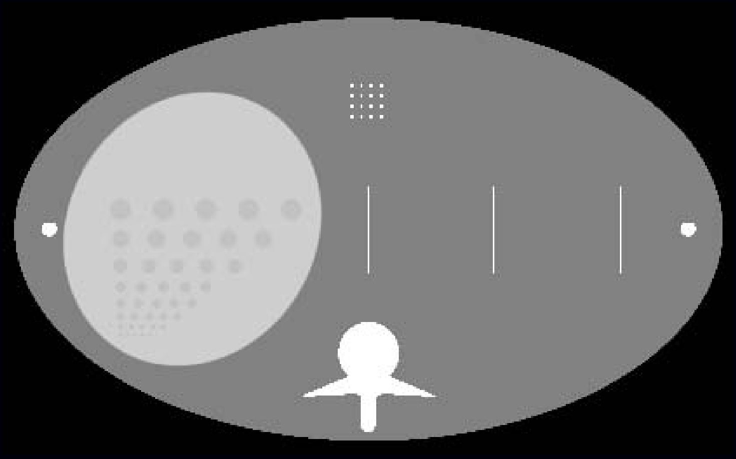




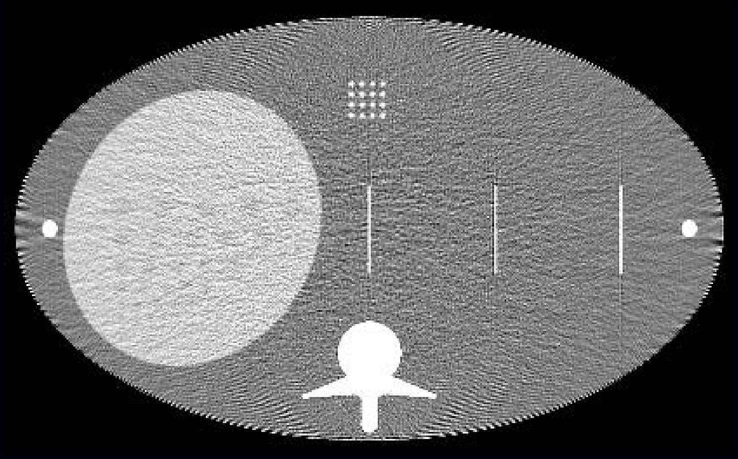
Thibault *et al.*, Fully3D, 2007

For a statistical approach to interior tomography, see Xu *et al.*, IEEE T-MI, May 2011.

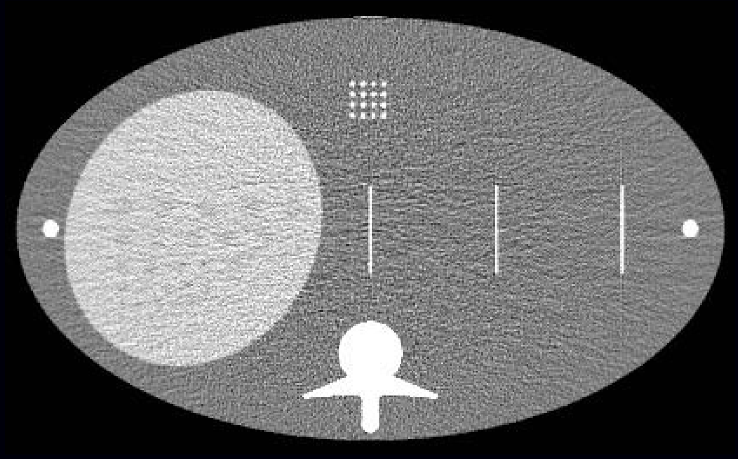
Voxel size matters?



digital phantom



512^2 grid



1024^2 grid

Unregularized OS reconstructions. Zbijewski & Beekman, PMB, Jan. 2004

Choice 2. System model / Physics model

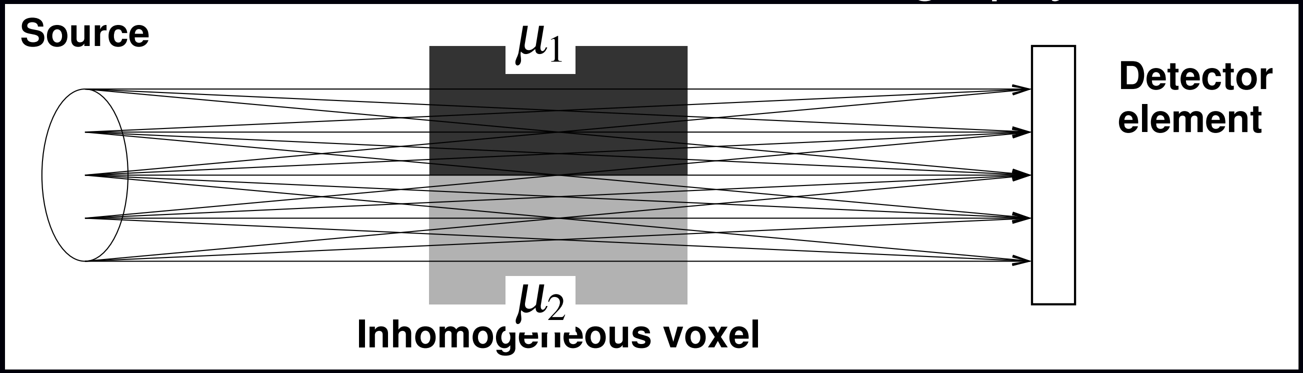
- scan geometry
- source intensity I_0
  - spatial variations (air scan)
  - intensity fluctuations
- resolution effects
  - finite detector size / detector spatial response
  - finite X-ray spot size / anode angulation
  - detector afterglow / gantry rotation
- spectral effects
  - X-ray source spectrum
  - bowtie filters
  - detector spectra response
- scatter
- ...

Challenges / trade-offs

- computation time
- accuracy/artifacts/resolution/contrast
- dose?

Exponential edge-gradient effect

Fundamental difference between emission tomography and CT:



Recorded intensity for ith ray: (Joseph and Spital, PMB, May 1981)

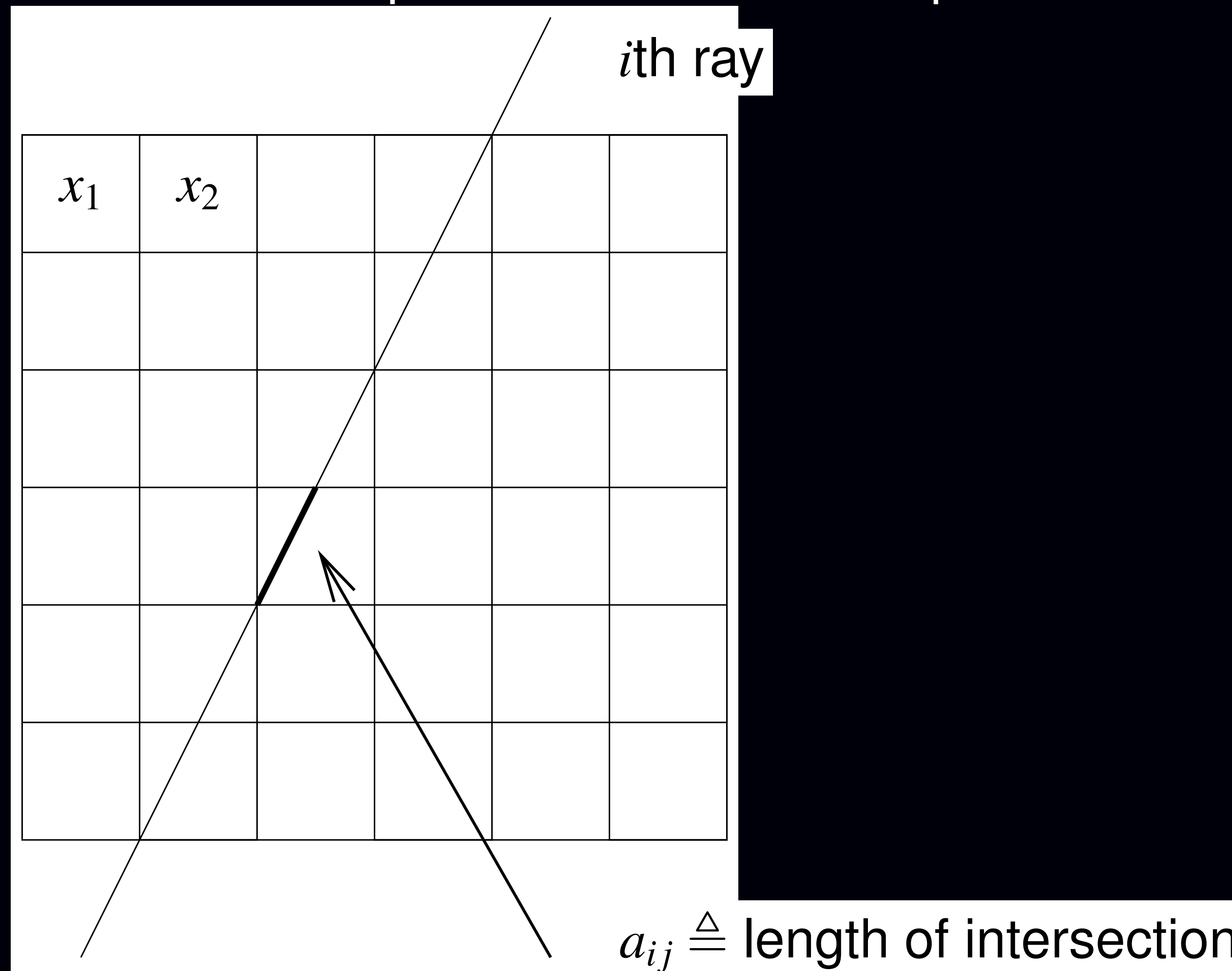
$$I_i = \int_{\text{source}} \int_{\text{detector}} I_0(\vec{p}_s, \vec{p}_d) \exp\left(- \int_{\mathscr{L}(\vec{p}_s,\vec{p}_d)} \mu(\vec{r}) \, d\ell \right) d\vec{p}_d \, d\vec{p}_s$$
$$\neq I_0 \exp\left(- \int_{\text{source}} \int_{\text{detector}} \int_{\mathscr{L}(\vec{p}_s,\vec{p}_d)} \mu(\vec{r}) \, d\ell \, d\vec{p}_d \, d\vec{p}_s \right).$$

Usual "linear" approximation:

$$I_i \approx I_0 \exp\left(- \sum_{j=1}^{N} a_{ij} x_j \right), \qquad \underbrace{a_{ij} \triangleq \int_{\text{source}} \int_{\text{detector}} \int_{\mathscr{L}(\vec{p}_s,\vec{p}_d)} b_j(\vec{r}) \, d\ell \, d\vec{p}_d \, d\vec{p}_s}_{\text{elements of system matrix } \boldsymbol{A}}$$

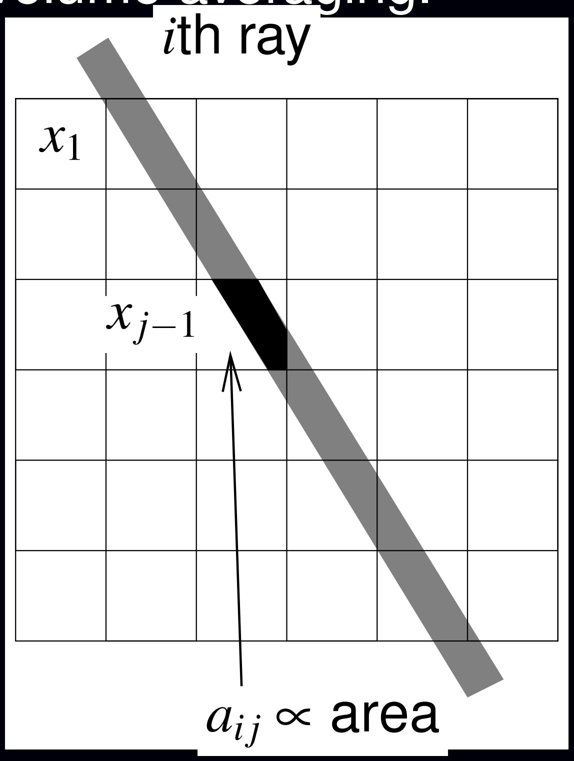
"Line Length" System Model

Assumes (implicitly?) that source is a point and detector is a point.



“Strip Area” System Model

Account for finite detector width.
Ignores nonlinear partial-volume averaging.



Practical (?) implementations in 3D include

- Distance-driven method (De Man and Basu, PMB, Jun. 2004)
- Separable-footprint method (Long *et al.*, T-MI, Nov. 2010)
- Further comparisons needed...

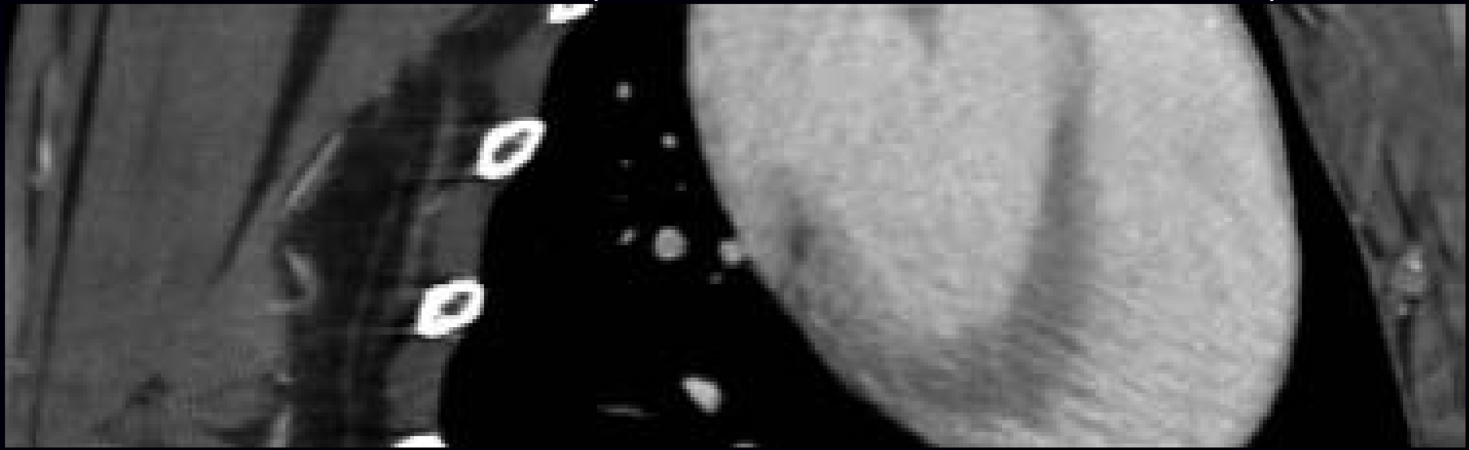
Lines versus strips

From (De Man and Basu, PMB, Jun. 2004)

MLTR of rabbit heart

Ray-driven (idealized point detector)

Distance-driven (models finite detector width)



Forward- / Back-projector "Pairs"

Typically iterative algorithms require two key steps.

- forward projection (image domain to projection domain):

$$\bar{\boldsymbol{y}} = \boldsymbol{A}\boldsymbol{x}, \qquad \bar{y}_i = \sum_{j=1}^{N} a_{ij} x_j = [\boldsymbol{A}\boldsymbol{x}]_i$$

- backprojection (projection domain to image domain):

$$\boldsymbol{z} = \boldsymbol{A}'\boldsymbol{y}, \qquad z_j = \sum_{i=1}^{M} a_{ij} y_i$$

The term "forward/backprojection pair" often refers to some implicit choices for the object basis and the system model.

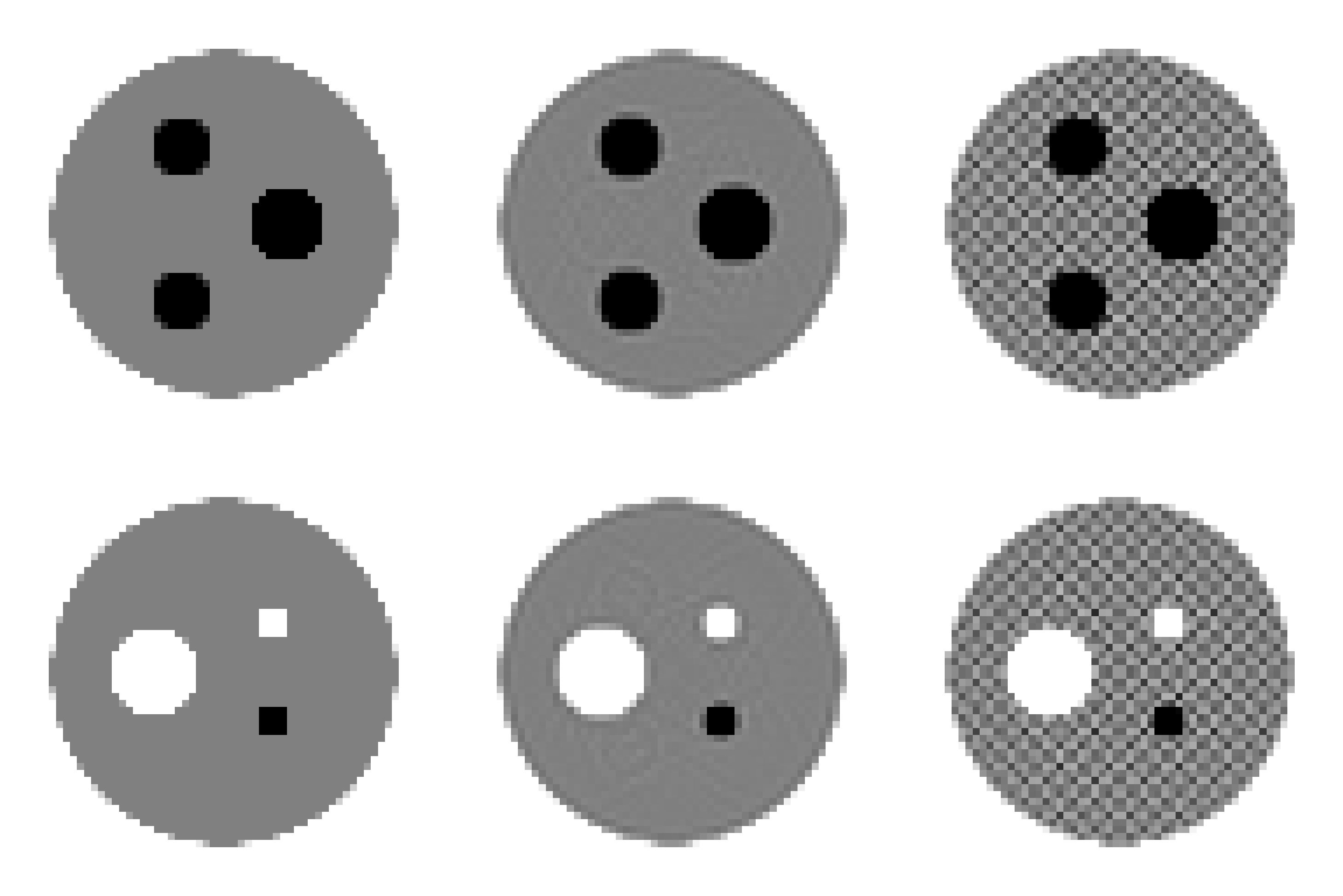
Sometimes $\boldsymbol{A}'\boldsymbol{y}$ is implemented as $\boldsymbol{B}\boldsymbol{y}$ for some "backprojector" $\boldsymbol{B} \neq \boldsymbol{A}'$.
Especially in SPECT and sometimes in PET.

Least-squares solutions (for example):

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|^2 = \left[\boldsymbol{A}'\boldsymbol{A}\right]^{-1} \boldsymbol{A}'\boldsymbol{y} \neq [\boldsymbol{B}\boldsymbol{A}]^{-1} \boldsymbol{B}\boldsymbol{y}$$

Mismatched Backprojector $B \neq A'$

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



Matched | Mismatched

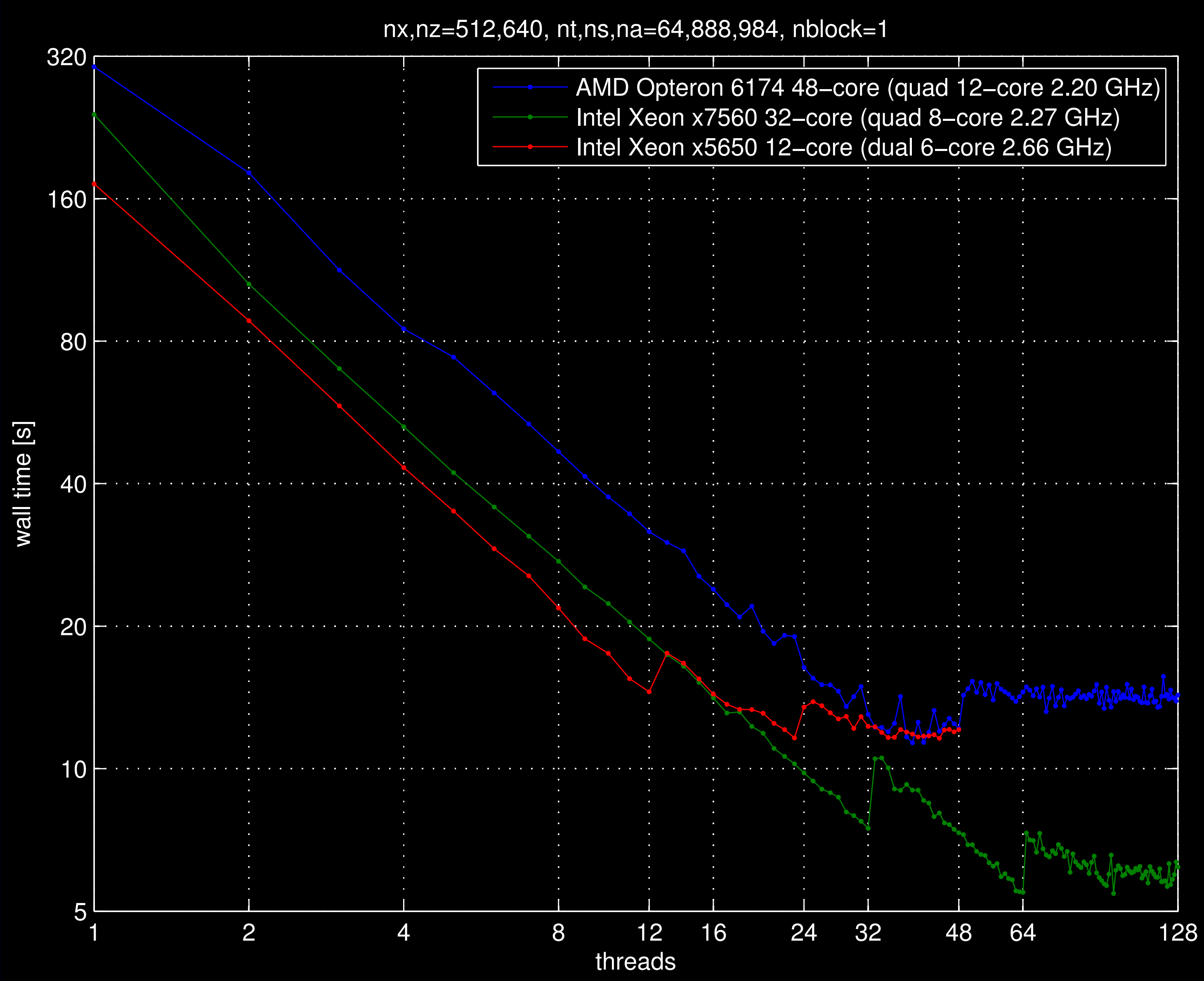
cf. SPECT/PET reconstruction – usually unregularized

Projector/back-projector bottleneck

Challenges

- Projector/backprojector algorithm design
  - Approximations (*e.g.*, transaxial/axial separability)
  - Symmetry
- Hardware / software implementation
  - GPU, CUDA, OpenCL, FPGA, SIMD, pthread, OpenMP, MPI, ...
- Further "wholistic" approaches?
  e.g., Basu & De Man, "Branchless distance driven projection ...," SPIE 2006
- ...

Forward projector parallelization (Fully3D 2011)



Choice 3. Statistical Model

The physical model describes measurement mean,
e.g., for a monoenergetic X-ray source and ignoring scatter etc.:

$$\bar{I}_i = I_0 \, \mathrm{e}^{-\sum_{j=1}^{N} a_{ij} x_j}.$$

The raw noisy measurements $\{I_i\}$ are distributed around those means.
Statistical reconstruction methods require a model for that distribution.

Challenges / Trade offs: using more accurate statistical models
- *may* lead to less noisy images
- may incur additional computation
- may involve higher algorithm complexity.

CT measurement statistics are very complicated, particularly at low doses.
- incident photon flux variations (Poisson)
- X-ray photon absorption/scattering (Bernoulli)
- energy-dependent light production in scintillator (?)
- shot noise in photodiodes (Poisson?)
- electronic noise in readout electronics (Gaussian?)
  Whiting, SPIE 4682, 2002; Lasio *et al.*, PMB, Apr. 2007
- Inaccessibility of raw sinogram data

To log() or not to log() – That is the question

Models for "raw" data I_i (before logarithm)

- compound Poisson (complicated) Whiting, SPIE 4682, 2002; Elbakri & Fessler, SPIE 5032, 2003; Lasio *et al.*, PMB, Apr. 2007
- Poisson + Gaussian (photon variability and electronic readout noise):

$$I_i \sim \text{Poisson}\{\bar{I}_i\} + \text{N}\left(0, \sigma^2\right)$$

  Snyder *et al.*, JOSAA, May 1993 & Feb. 1995 .

- Shifted Poisson approximation (matches first two moments):

$$\tilde{I}_i \triangleq \left[I_i + \sigma^2\right]_+ \sim \text{Poisson}\left\{\bar{I}_i + \sigma^2\right\}$$

  Yavuz & Fessler, MIA, Dec. 1998

- Ordinary Poisson (ignore electronic noise):

$$I_i \sim \text{Poisson}\{\bar{I}_i\}$$

  Rockmore and Macovski, TNS, Jun. 1977; Lange and Carson, JCAT, Apr. 1984

- Photon-counting detectors would simplify statistical modeling

All are somewhat complicated by the nonlinearity of the physics: $\bar{I}_i = e^{-[\boldsymbol{A}\boldsymbol{x}]_i}$

After taking the log()

Taking the log leads to a simpler linear model (ignoring beam hardening):

$$y_i \triangleq -\log\left(\frac{I_i}{I_0}\right) \approx [\boldsymbol{A}\boldsymbol{x}]_i + \varepsilon_i$$

Drawbacks:

- Undefined if $I_i \leq 0$ (*e.g.*, due to electronic noise)
- It is *biased* (by Jensen's inequality): $\mathsf{E}[y_i] \geq -\log(\bar{I}_i / I_0) = [\boldsymbol{A}\boldsymbol{x}]_i$
- Exact distribution of log-domain noise ε_i is intractable.

Practical approach: assume Gaussian noise model: $\varepsilon_i \sim \mathsf{N}\left(0, \sigma_i^2\right)$

Options for modeling noise variance $\sigma_i^2 = \mathsf{Var}\{\varepsilon_i\}$

- consider both Poisson and Gaussian noise effects: $\sigma_i^2 = \frac{\bar{I}_i + \sigma^2}{\bar{I}_i^2}$
  (Thibault *et al.*, SPIE 6065, 2006)
- consider just Poisson effect: $\sigma_i^2 = \frac{1}{\bar{I}_i}$ (Sauer & Bouman, T-SP, Feb. 1993)
- pretend it is white noise: $\sigma_i^2 = \sigma_0^2$
- ignore noise altogether and "solve" $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$

Whether using pre-log data is better than post-log data is an open question.

Choice 4. Cost Functions

Components:

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

Reconstruct image $\hat{\boldsymbol{x}}$ by finding minimizer of a cost function:

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

$$\Psi(\boldsymbol{x}) = \text{DataMismatch}(\boldsymbol{y}, \boldsymbol{A}\boldsymbol{x}) + \beta \, \text{Regularizer}(\boldsymbol{x})$$

Forcing too much “data fit” alone would give noisy images.

Equivalent to a Bayesian MAP (maximum *a posteriori*) estimator.

Distinguishes “statistical methods” from “algebraic methods” for “$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$.”

Choice 4.1: Data-Mismatch Term

Standard choice is the negative log-likelihood of statistical model:

$$\text{DataMismatch} = -L(\boldsymbol{x};\boldsymbol{y}) = -\log \mathrm{p}(\boldsymbol{y}|\boldsymbol{x}) = \sum_{i=1}^{M} -\log \mathrm{p}(y_i|\boldsymbol{x}) \,.$$

- For pre-log data $\boldsymbol{I}$ with shifted Poisson model:

$$-L(\boldsymbol{x};\boldsymbol{I}) = \sum_{i=1}^{M} \left(\bar{I}_i + \sigma^2\right) - \left[I_i + \sigma^2\right]_+ \log\left(\bar{I}_i + \sigma^2\right), \qquad \bar{I}_i = I_0 \,\mathrm{e}^{-[\boldsymbol{A}\boldsymbol{x}]_i}$$

  This can be non-convex if $\sigma^2 > 0$;
  it is convex if we ignore electronic noise $\sigma^2 = 0$. Trade-off ...

- For post-log data $\boldsymbol{y}$ with Gaussian model:

$$-L(\boldsymbol{x};\boldsymbol{y}) = \sum_{i=1}^{M} w_i \frac{1}{2}(y_i - [\boldsymbol{A}\boldsymbol{x}]_i)^2 = \frac{1}{2}(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x})'\boldsymbol{W}(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}), \qquad w_i = 1/\sigma_i^2$$

  This is a kind of (data-based) weighted least squares (WLS).
  It is always convex in $\boldsymbol{x}$. Quadratic functions are “easy” to minimize.

- ...

Choice 4.2: Regularization

How to control noise due to ill-conditioning?

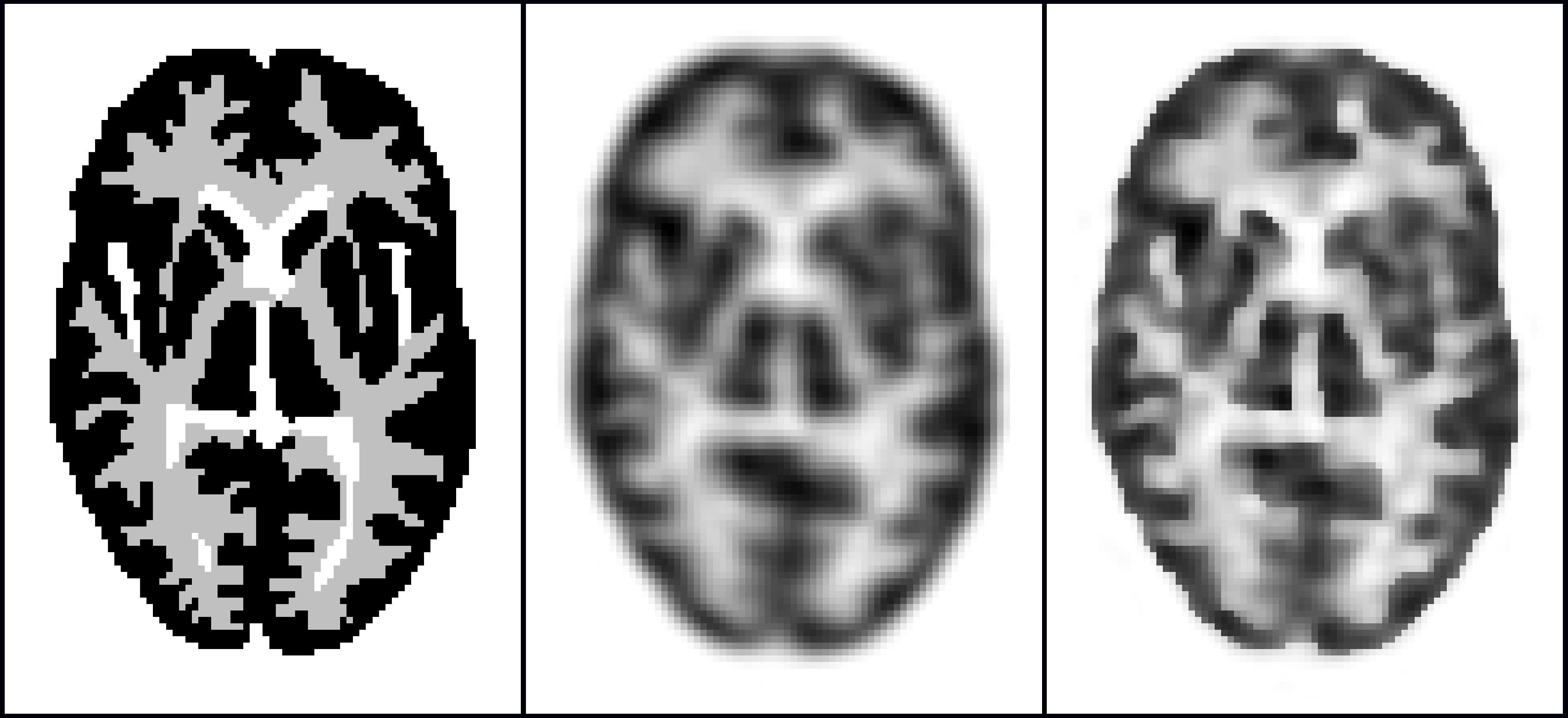
Noise-control methods in clinical use in PET reconstruction today:
- Stop an unregularized algorithm before convergence
- Over-iterate an unregularized algorithm then post-filter

Other possible "simple" solutions:
- Modify the raw data (pre-filter / denoise)
- Filter between iterations
- ...

Appeal:
- simple / familiar
- filter parameters have intuitive units (*e.g.*, FWHM), unlike a regularization parameter β
- Changing a post-filter does not require re-iterating, unlike changing a regularization parameter β

Dozens of papers on regularized methods for PET, but little clinical impact.
(USC MAP method is available in mouse scanners.)

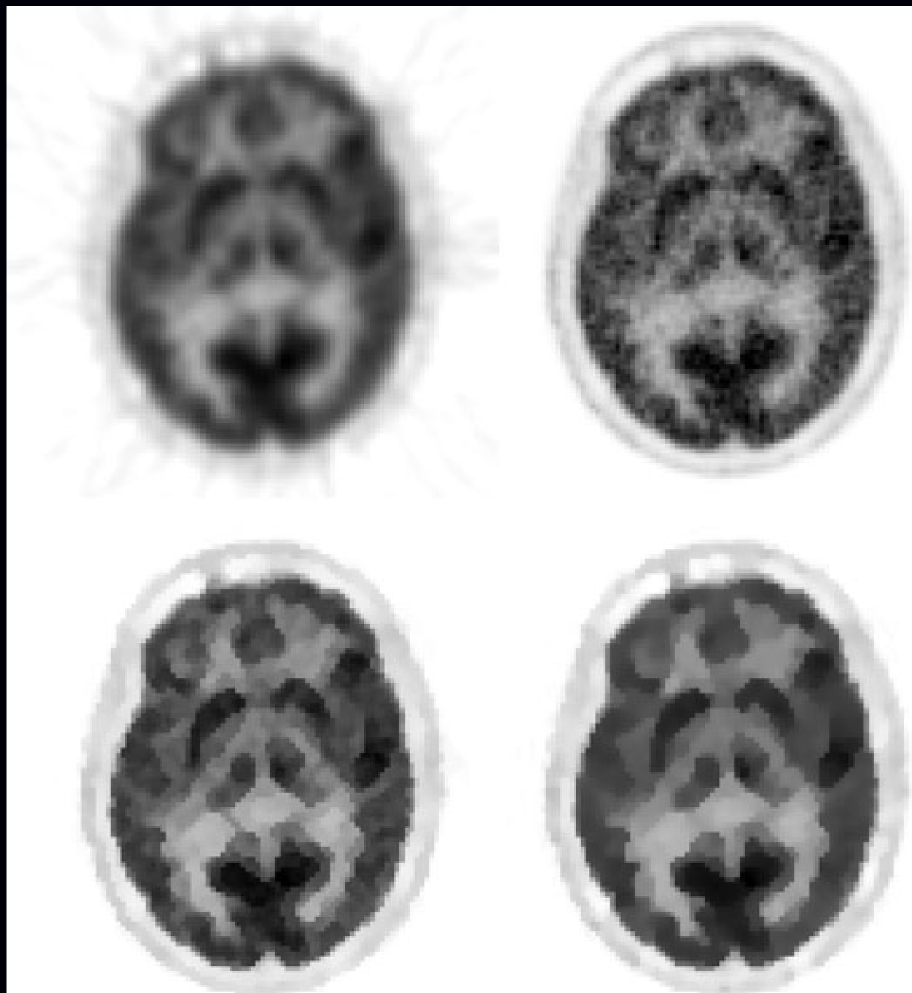
Edge-Preserving Reconstruction: PET Example



Phantom Quadratic regularizer Huber regularizer

Quantification vs qualitative vs tasks...

More "Edge Preserving" PET Regularization



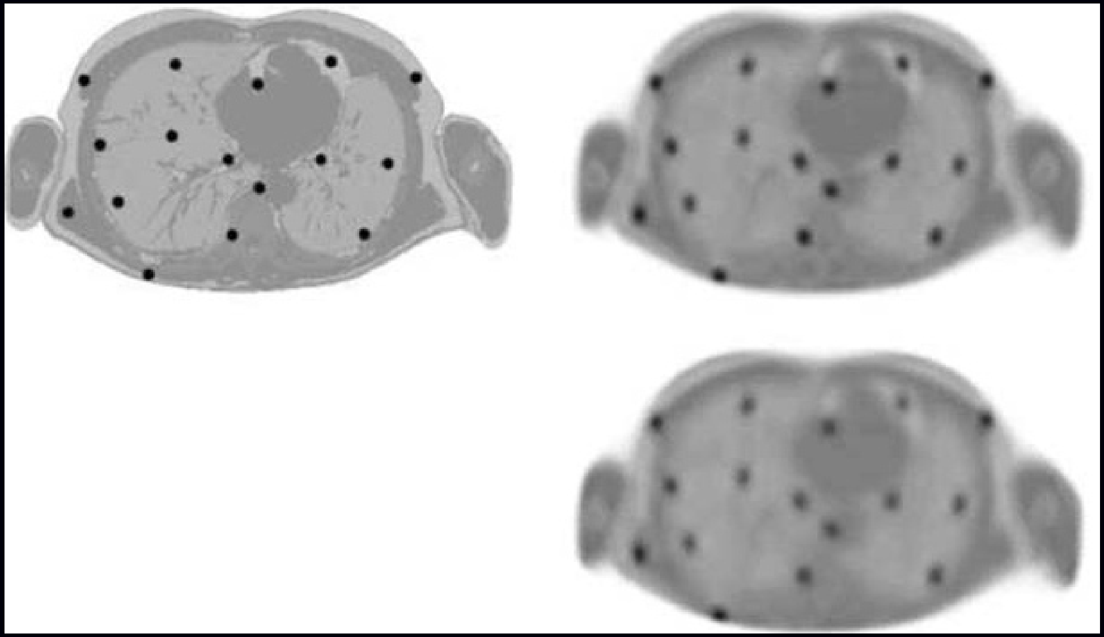
FBP	ML-EM
Median-root prior	Huber regularizer

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

Regularization in PET

Nuyts *et al.*, T-MI, Jan. 2009:
MAP method outperformed post-filtered ML for lesion detection in simulation

Noiseless images:



Phantom	ML-EM filtered
	Regularized

Regularization options

Options for regularizer $\mathsf{R}(\boldsymbol{x})$ in increasing complexity:

- quadratic roughness
- convex, non-quadratic roughness
- non-convex roughness
- total variation
- convex sparsity
- non-convex sparsity

Challenges

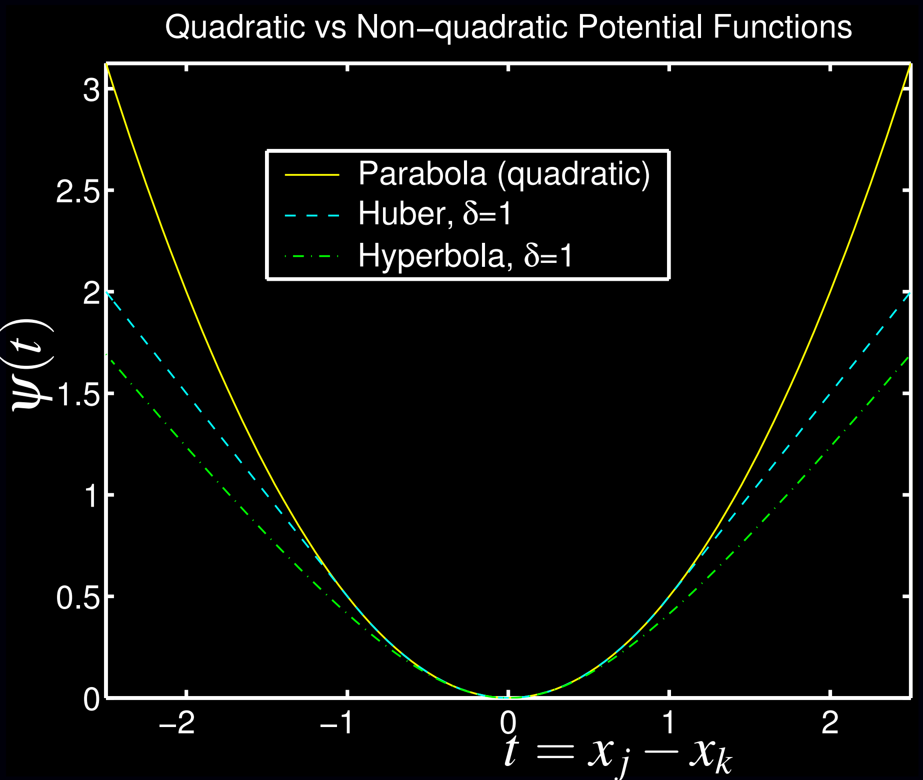
- Reducing noise without degrading spatial resolution
- Balancing regularization strength between and within slices
- Parameter selection
- Computational complexity (voxels have 26 neighbors in 3D)
- Preserving “familiar” noise texture
- Optimizing clinical task performance

Many open questions...

Roughness Penalty Functions

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

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



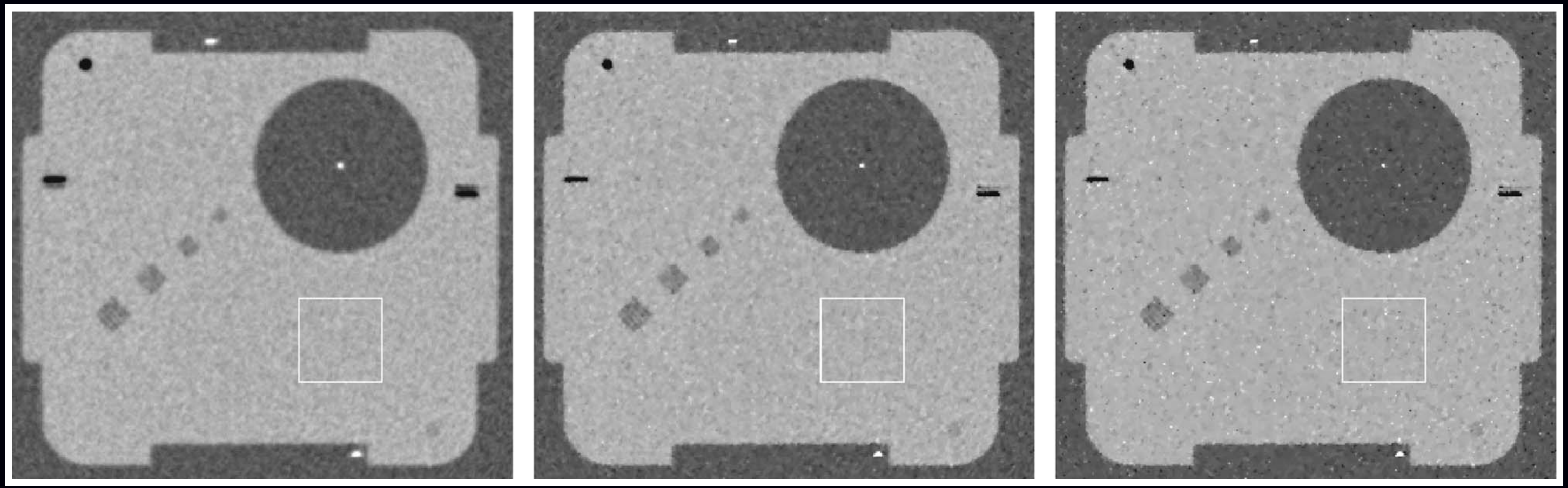
quadratic: $\psi(t) = t^2$
hyperbola: $\psi(t) = \sqrt{1 + (t/\delta)^2}$
(edge preservation)

Regularization parameters: Dramatic effects

Thibault *et al.*, Med. Phys., Nov. 2007

"q generalized gaussian" potential function with tuning parameters: β, δ, p, q:

$$\beta\, \psi(t) = \beta \frac{\frac{1}{2}|t|^p}{1 + |t/\delta|^{p-q}}$$

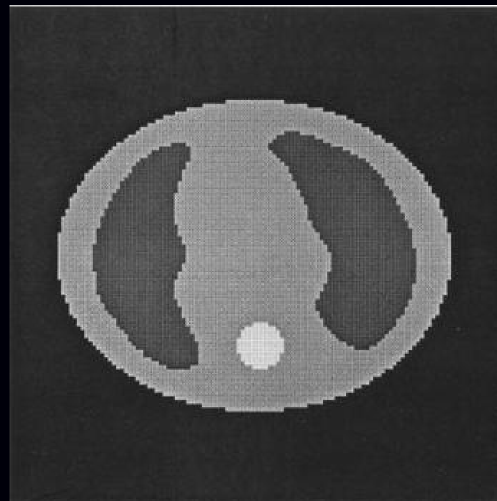


$p = q = 2$ | $p = 2,\ q = 1.2,\ \delta = 10$ HU | $p = q = 1.1$

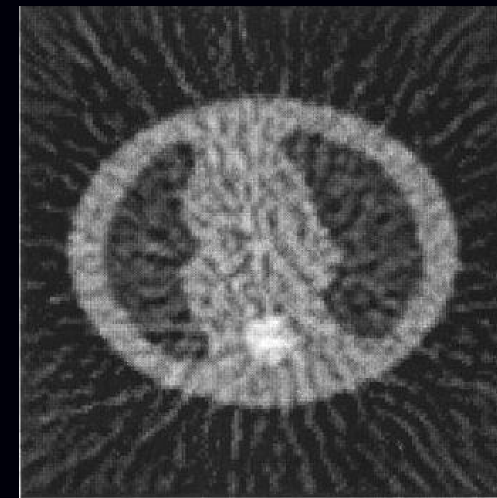
noise:	11.1	10.9	10.8
(#lp/cm):	4.2	7.2	8.2

Piecewise constant phantoms

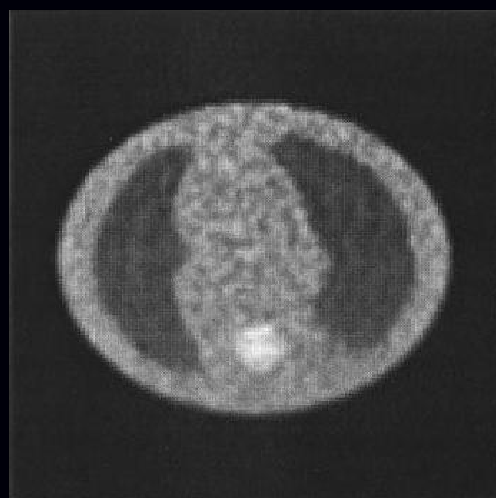
Phantom:



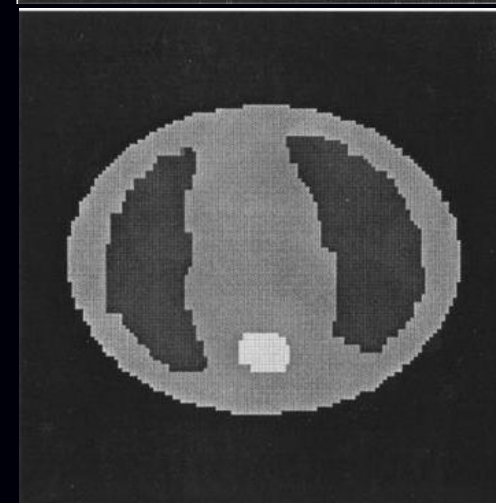
FBP:



MLEM:



MAP:



Lee *et al.*, IEEE T-NS, 2002, 300K counts

non-convex “broken parabola” potential function and deterministic annealing

Summary thus far

1. Object parameterization
2. System physical model
3. Measurement statistical model
4. Cost function: data-mismatch / regularization / constraints

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

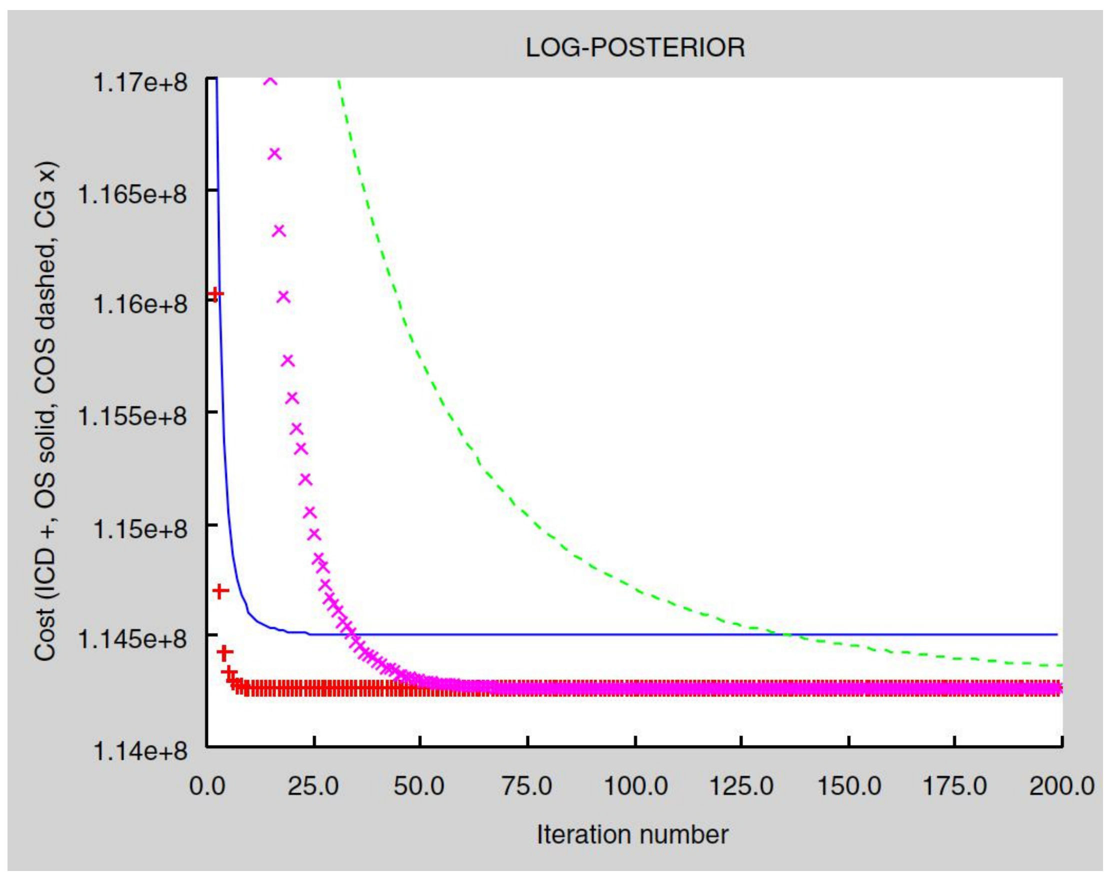
5. Minimization algorithms:

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

Choice 5: Minimization algorithms

- Conjugate gradients
  - Converges slowly for CT
  - Difficult to precondition due to weighting and regularization
  - Difficult to enforce nonnegativity constraint
  - Very easily parallelized
- Ordered subsets
  - Initially converges faster than CG if many subsets used
  - Does not converge without relaxation etc., but those slow it down
  - Computes regularizer gradient $\nabla \mathsf{R}(\boldsymbol{x})$ for every subset - expensive?
  - Easily enforces nonnegativity constraint
  - Easily parallelized
- Coordinate descent (Sauer and Bouman, T-SP, 1993)
  - Converges high spatial frequencies rapidly, but low frequencies slowly
  - Easily enforces nonnegativity constraint
  - Challenging to parallelize
- Block coordinate descent (Benson *et al.*, NSS/MIC, 2010)
  - Spatial frequency convergence properties depend...
  - Easily enforces nonnegativity constraint
  - More opportunity to parallelize than CD

Convergence rates

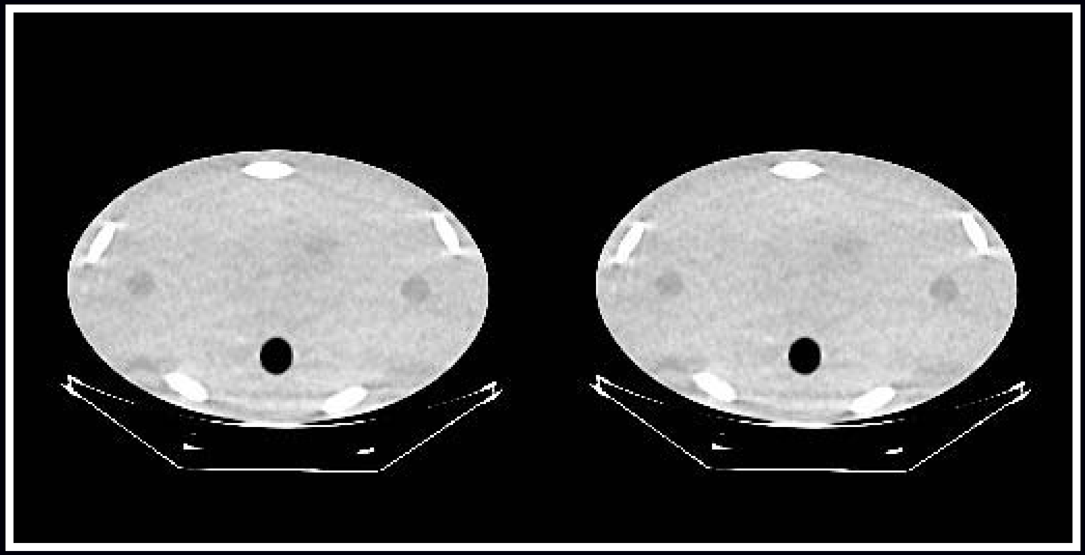


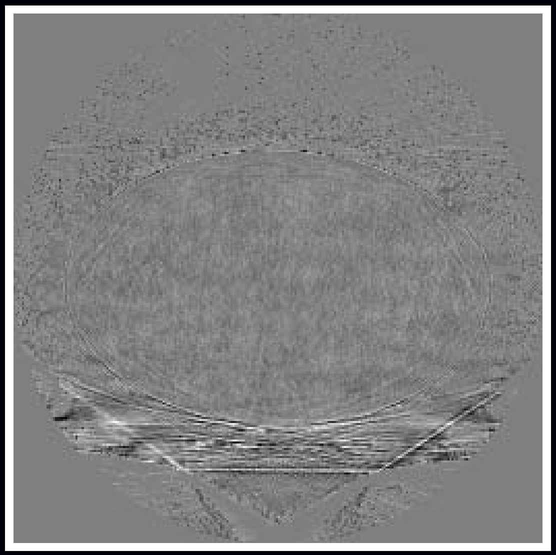
(De Man *et al.*, NSS/MIC 2005)

In terms of iterations: CD $<$ OS $<$ CG $<$ Convergent OS
In terms of compute time? (it depends...)

Ordered subsets convergence

Theoretically OS does not converge, but it may get "close enough," even with regularization.





CD
200 iter

OS
41 subsets
200 iter

difference
0 ± 10HU

display: 930 HU $\pm$ 58 HU

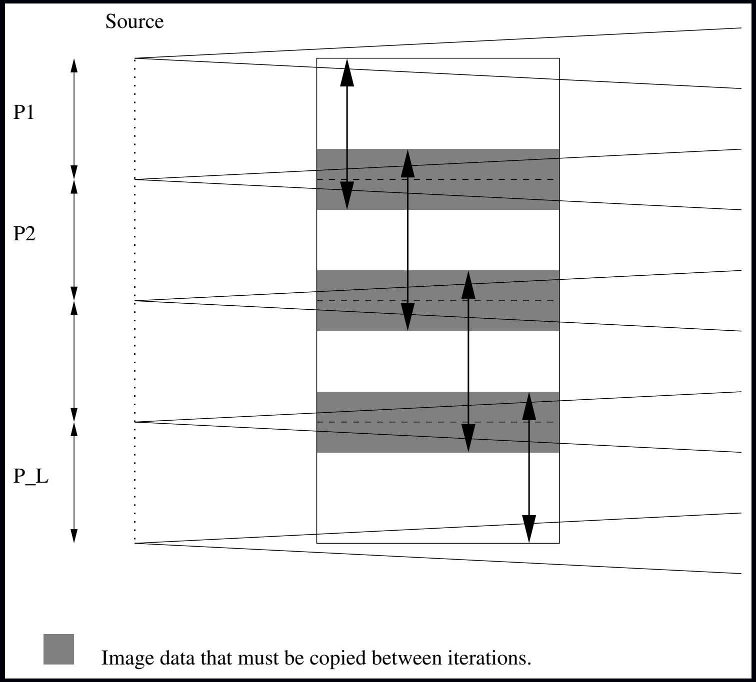
(De Man *et al.*, NSS/MIC 2005)

Ongoing saga...

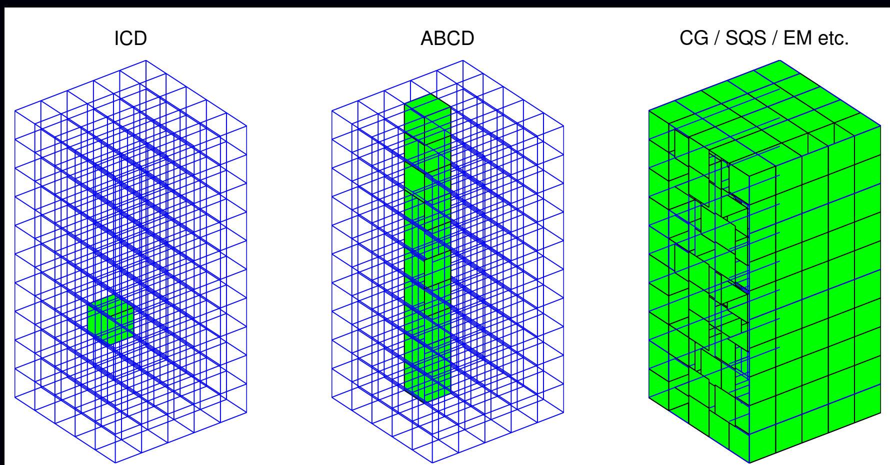
(SPIE, ISBI, Fully 3D, ...)

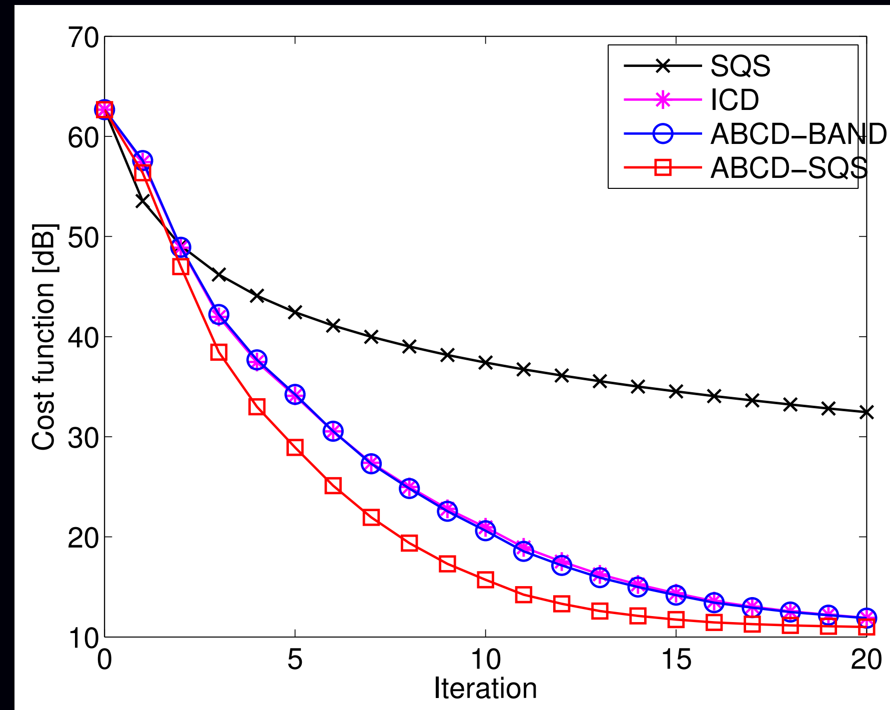
Optimization algorithms

Challenges:
- theoretical convergence (to establish gold standards)
- practical: near convergence in few iterations
- highly parallelizable
- efficient use of hardware: memory bandwidth, cache, ...
- predictable stopping rules
- partitioning of helical CT data across multiple compute nodes



Axial block coordinate descent (ABCD) (Fully3D 2011)





Optimizing non-differentiable functions using constraints

Especially for angularly under-sampled problems,
"strong" regularizers, like *total variation* (TV), may be needed, *e.g.*,

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \beta \|\boldsymbol{C}\boldsymbol{x}\|_1,$$

where $\boldsymbol{C}$ is a wavelet transform or finite-differencing operator.

Optimization trick (synopsis): introduce auxiliary variable $\boldsymbol{z} = \boldsymbol{C}\boldsymbol{x}$:

$$\arg\min_{\boldsymbol{x},\boldsymbol{z}} \Phi(\boldsymbol{x}, \boldsymbol{z}), \qquad \Phi(\boldsymbol{x}, \boldsymbol{z}) \triangleq \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \beta \|\boldsymbol{z}\|_1 + \mu \|\boldsymbol{z} - \boldsymbol{C}\boldsymbol{x}\|_2^2$$

Alternate between updating $\boldsymbol{x}$ and $\boldsymbol{z}$:

$$\boldsymbol{x}^{(n+1)} = \arg\min_{\boldsymbol{x}} \Phi(\boldsymbol{x}, \boldsymbol{z}^{(n)}) = \arg\min_{\boldsymbol{x}} \underbrace{\frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \mu \|\boldsymbol{z} - \boldsymbol{C}\boldsymbol{x}\|_2^2}_{\text{quadratic: CG}}$$

$$\boldsymbol{z}^{(n+1)} = \arg\min_{\boldsymbol{z}} \Phi(\boldsymbol{x}^{(n+1)}, \boldsymbol{z}) = \arg\min_{\boldsymbol{z}} \underbrace{\beta \|\boldsymbol{z}\|_1 + \mu \|\boldsymbol{z} - \boldsymbol{C}\boldsymbol{x}\|_2^2}_{\text{separable: soft thresholding}}$$

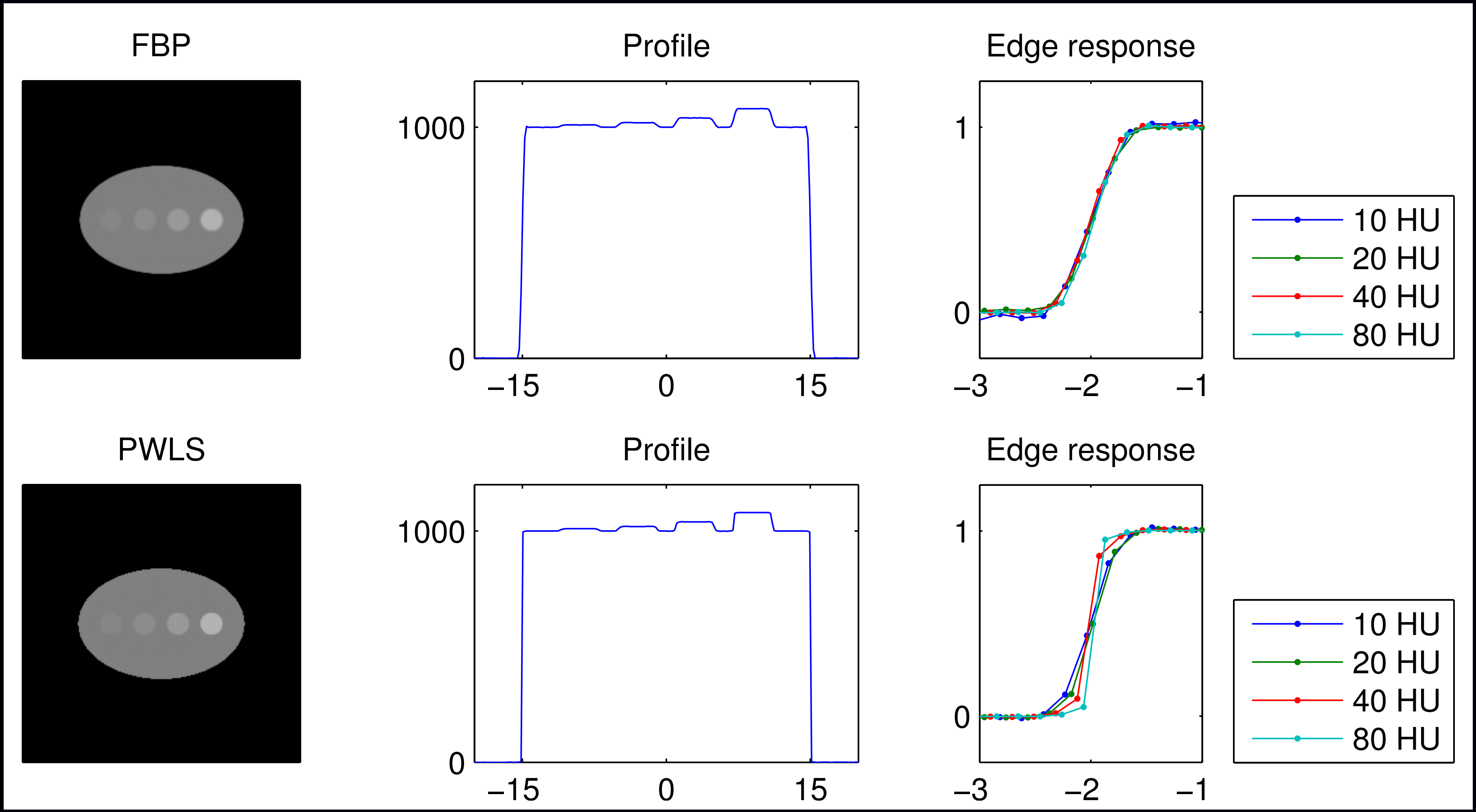
Many more details unfolding rapidly in literature...

Example

(movie in pdf)

82-subset OS with two different (but similar) edge-preserving regularizers.
One frame per every 10th iteration.

Resolution characterization: 2D CT



Challenge:

Shape of edge response depends on contrast for edge-preserving regularization.

Assessing image quality

Challenges:

- Resolution (PSF, edge response, MTF)
- Noise (predictions)
- Task-based performance measures
  Known-location versus unknown-location tasks
- ...

"How low can the dose go" – quite challenging to answer

Some open problems in statistical image reconstruction

- Modeling
  - Statistical modeling for very low-dose CT
  - Resolution effects
  - Spectral CT
  - Object motion
- Parameter selection / performance characterization
  - Performance prediction for nonquadratic regularization
  - Effect of nonquadratic regularization on detection tasks
  - Choice of regularization parameters for nonquadratic regularization
- Algorithms
  - optimization algorithm design
  - software/hardware implementation
  - Moore's law alone will not suffice
    (dual energy, dual source, motion, dynamic, smaller voxels ...)
- Clinical evaluation
- ...

Many research opportunities to aid this CT revolution...

Bibliography

References

[1] P. M. Joseph and R. D. Spital. A method for correcting bone induced artifacts in computed tomography scanners. *J. Comp. Assisted Tomo.*, 2(1):100–8, January 1978.

[2] H. K. Bruder, R. Raupach, M. Sedlmair, J. Sunnegardh, K. Stierstorfer, and T. Flohr. Adaptive iterative reconstruction (AIR). In *spie-7691*, page 76910J, 2011.

[3] X. Zhu and P. Milanfar. Automatic parameter selection for denoising algorithms using a no-reference measure of image content. *IEEE Trans. Im. Proc.*, 19(12):3116–32, December 2010.

[4] D. L. Parker. Optimal short scan convolution reconstruction for fan beam CT. *Med. Phys.*, 9(2):254–7, March 1982.

[5] J. Hsieh. Adaptive streak artifact reduction in computed tomography resulting from excessive x-ray photon noise. *Med. Phys.*, 25(11):2139–47, November 1998.

[6] P. J. La Riviere and X. Pan. Nonparametric regression sinogram smoothing using a roughness-penalized Poisson likelihood objective function. *IEEE Trans. Med. Imag.*, 19(8):773–86, August 2000.

[7] P. J. La Riviere and D. M. Billmire. Reduction of noise-induced streak artifacts in X-ray computed tomography through spline-based penalized-likelihood sinogram smoothing. *IEEE Trans. Med. Imag.*, 24(1):105–11, January 2005.

[8] P. J. La Riviere, J. Bian, and P. A. Vargas. Penalized-likelihood sinogram restoration for computed tomography. *IEEE Trans. Med. Imag.*, 25(8):1022–36, August 2006.

[9] P. J. La Rivière and P. Vargas. Correction for resolution nonuniformities caused by anode angulation in computed tomography. *IEEE Trans. Med. Imag.*, 27(9):1333–41, September 2008.

[10] J. Wang, T. Li, H. Lu, and Z. Liang. Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose X-ray computed tomography. *IEEE Trans. Med. Imag.*, 25(10):1272–83, October 2006.

[11] D. E. Kuhl and R. Q. Edwards. Image separation radioisotope scanning. *Radiology*, 80:653–62, 1963.

[12] D. A. Chesler. Three-dimensional activity distribution from multiple positron scintgraphs. *J. Nuc. Med.*, 12(6):347–8, June 1971.

[13] M. Goitein. Three-dimensional density reconstruction from a series of two-dimensional projections. *Nucl. Instr. Meth.*, 101(3):509–18, June 1972.

[14] W. H. Richardson. Bayesian-based iterative method of image restoration. *J. Opt. Soc. Am.*, 62(1):55–9, January 1972.

[15] L. Lucy. An iterative technique for the rectification of observed distributions. *The Astronomical Journal*, 79(6):745–54, June 1974.

[16] A. J. Rockmore and A. Macovski. A maximum likelihood approach to emission image reconstruction from projections. *IEEE Trans. Nuc. Sci.*, 23:1428–32, 1976.

[17] L. A. Shepp and Y. Vardi. Maximum likelihood reconstruction for emission tomography. *IEEE Trans. Med. Imag.*, 1(2):113–22, October 1982.
[18] 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.
[19] 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.
[20] J. Llacer, E. Veklerov, L. R. Baxter, S. T. Grafton, L. K. Griffeth, R. A. Hawkins, C. K. Hoh, J. C. Mazziotta, E. J. Hoffman, and C. E. Metz. Results of a clinical receiver operating characteristic study comparing filtered backprojection and maximum likelihood estimator images in FDG PET studies. *J. Nuc. Med.*, 34(7):1198–203, July 1993.
[21] S. R. Meikle, B. F. Hutton, D. L. Bailey, P. K. Hooper, and M. J. Fulham. Accelerated EM reconstruction in total-body PET: potential for improving tumour detectability. *Phys. Med. Biol.*, 39(10):1689–794, October 1994.
[22] 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.
[23] 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.
[24] R. Gordon and G. T. Herman. Reconstruction of pictures from their projections. *Comm. ACM*, 14(12):759–68, December 1971.
[25] 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.
[26] R. Gordon. A tutorial on ART (algebraic reconstruction techniques). *IEEE Trans. Nuc. Sci.*, 21(3):78–93, June 1974.
[27] R. L. Kashyap and M. C. Mittal. Picture reconstruction from projections. *IEEE Trans. Comp.*, 24(9):915–23, September 1975.
[28] 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.
[29] K. Lange and R. Carson. EM reconstruction algorithms for emission and transmission tomography. *J. Comp. Assisted Tomo.*, 8(2):306–16, April 1984.
[30] K. Sauer and C. Bouman. A local update strategy for iterative reconstruction from projections. *IEEE Trans. Sig. Proc.*, 41(2):534–48, February 1993.
[31] S. H. Manglos, G. M. Gagne, A. Krol, F. D. Thomas, and R. Narayanaswamy. Transmission maximum-likelihood reconstruction with ordered subsets for cone beam CT. *Phys. Med. Biol.*, 40(7):1225–41, July 1995.
[32] C. Kamphuis and F. J. Beekman. Accelerated iterative transmission CT reconstruction using an ordered subsets convex algorithm. *IEEE Trans. Med. Imag.*, 17(6):1001–5, December 1998.
[33] H. Erdoğan and J. A. Fessler. Ordered subsets algorithms for transmission tomography. *Phys. Med. Biol.*, 44(11):2835–51, November 1999.
[34] Q. Xu, X. Mou, G. Wang, J. Sieren, E. A. Hoffman, and H. Yu. Statistical interior tomography. *IEEE Trans. Med. Imag.*, 30(5):1116–28, May 2011.
[35] W. Zbijewski and F. J. Beekman. Suppression of intensity transition artifacts in statistical x-ray computer tomography recon-

struction through Radon inversion initialization. *Med. Phys.*, 31(1):62–9, January 2004.

[36] P. M. Joseph and R. D. Spital. The exponential edge-gradient effect in x-ray computed tomography. *Phys. Med. Biol.*, 26(3):473–87, May 1981.

[37] B. De Man and S. Basu. Distance-driven projection and backprojection in three dimensions. *Phys. Med. Biol.*, 49(11):2463–75, June 2004.

[38] Y. Long, J. A. Fessler, and J. M. Balter. 3D forward and back-projection for X-ray CT using separable footprints. *IEEE Trans. Med. Imag.*, 29(11):1839–50, November 2010.

[39] Y. Long, J. A. Fessler, and J. M. Balter. 3D forward and back-projection for X-ray CT using separable footprints with trapezoid functions. In *Proc. First Intl. Mtg. on image formation in X-ray computed tomography*, pages 216–9, 2010.

[40] S. Basu and B. De Man. Branchless distance driven projection and backprojection. In *Proc. SPIE 6065, Computational Imaging IV*, page 60650Y, 2006.

[41] B. R. Whiting. Signal statistics in x-ray computed tomography. In *Proc. SPIE 4682, Medical Imaging 2002: Med. Phys.*, pages 53–60, 2002.

[42] 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.

[43] G. M. Lasio, B. R. Whiting, and J. F. Williamson. Statistical reconstruction for x-ray computed tomography using energy-integrating detectors. *Phys. Med. Biol.*, 52(8):2247–66, April 2007.

[44] 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.

[45] 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.

[46] M. Yavuz and J. A. Fessler. Statistical image reconstruction methods for randoms-precorrected PET scans. *Med. Im. Anal.*, 2(4):369–78, December 1998.

[47] J-B. Thibault, C. A. Bouman, K. D. Sauer, and J. Hsieh. A recursive filter for noise reduction in statistical iterative tomographic imaging. In *Proc. SPIE 6065, Computational Imaging IV*, page 60650X, 2006.

[48] 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.

[49] J. Nuyts, C. Michel, L. Brepoels, L. D. Ceuninck, C. Deroose, K. Goffin, F. M. Mottaghy, S. Stroobants, J. V. Riet, and R. Verscuren. Performance of MAP reconstruction for hot lesion detection in whole-body PET/CT: an evaluation with human and numerical observers. *IEEE Trans. Med. Imag.*, 28(1):67–73, January 2009.

[50] 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.

[51] S-J. Lee. Accelerated deterministic annealing algorithms for transmission CT reconstruction using ordered subsets. *IEEE Trans. Nuc. Sci.*, 49(5):2373–80, October 2002.

[52] S. Ramani and J. A. Fessler. Convergent iterative CT reconstruction with sparsity-based regularization. In *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, pages 302–5, 2011.