Innovations Required in Data Reconstruction

Jeffrey A. Fessler

EECS Department
University of Michigan

Summit on Management of Radiation Dose in Computerized Tomography:
Toward the Sub-mSv Exam

Feb. 15, 2011

http://www.eecs.umich.edu/∼fessler

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
- Wonseok Huh
- Donghwan Kim
- Yong Long
- Madison McGaffin
- Sathish Ramani
- Stephen Schmitt
- Meng Wu

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

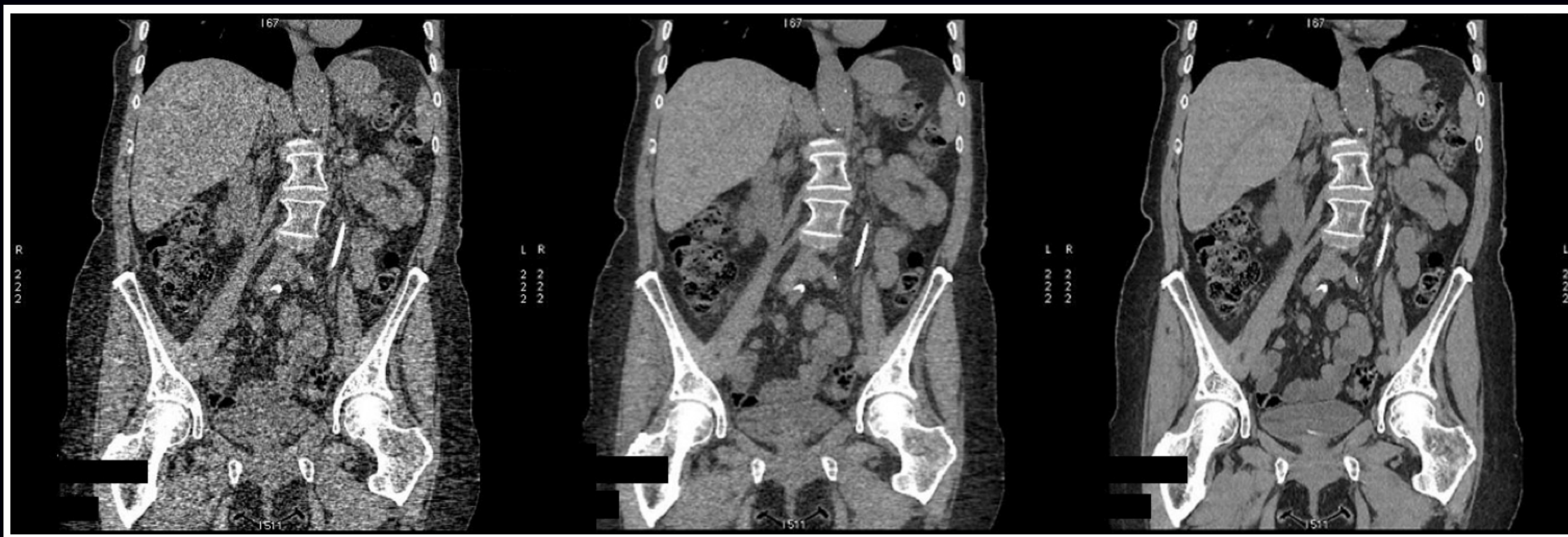
CT collaborators
- Mitch Goodsitt, UM
- Ella Kazerooni, UM
- Neal Clinthorne, UM
- Paul Kinahan, UW

Former PhD students (who did/do CT)
- Se Young Chun, Harvard / Brigham
- 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 students
- Kevin Brown, Philips
- ...

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 methods for CT?

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

- Nonstandard geometries
  - transaxial truncation (big 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 statistical models
  - weighting reduces influence of photon-starved rays
    (FBP treats all rays equally)
  - reducing image noise or 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

Flavors of "Statistical" reconstruction

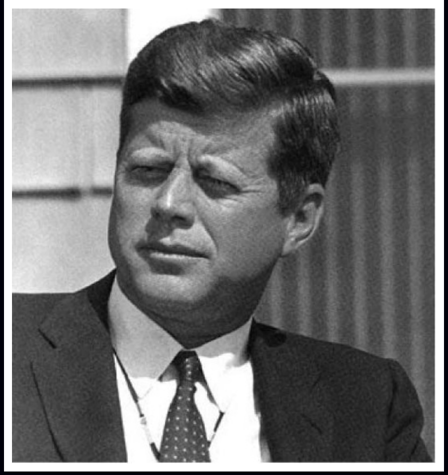
- Image domain
- Sinogram domain
- Fully statistical (both)
- Hybrid methods

"Statistical" methods: Image domain

- Denoising methods

$$\begin{array}{c}\text{sinogram}\\ \boldsymbol{y}\end{array} \longrightarrow \boxed{\text{FBP}} \longrightarrow \begin{array}{c}\text{noisy}\\ \text{reconstruction}\\ \tilde{\boldsymbol{x}}\end{array} \longrightarrow \boxed{\begin{array}{c}\text{iterative}\\ \text{denoiser}\end{array}} \longrightarrow \begin{array}{c}\text{final}\\ \text{image}\\ \hat{\boldsymbol{x}}\end{array}$$

  - 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”

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

$$\uparrow$$

sinogram
statistics?

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

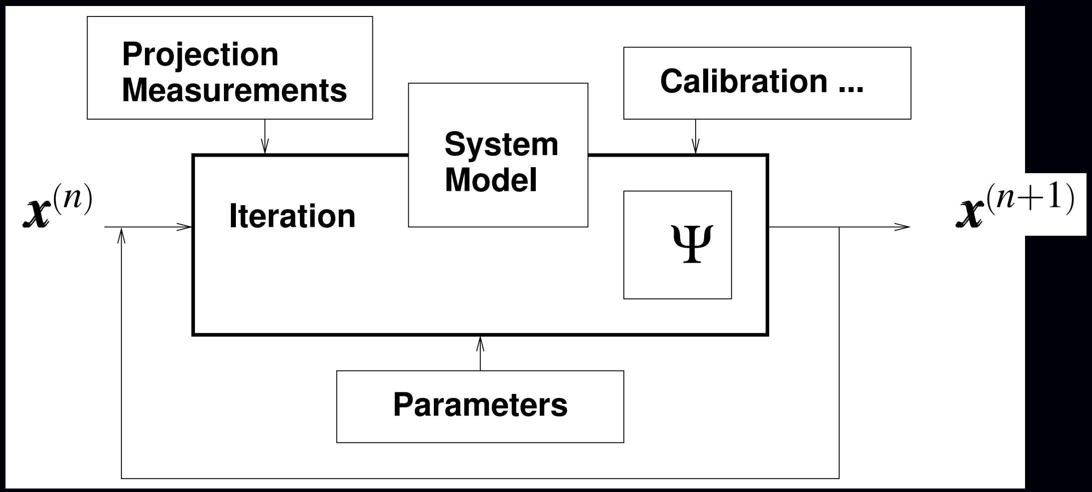
FBP from denoised sinogram

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

(True? Fully? Slow?) Statistical 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)
- 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

Key factors

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

Five Choices for Statistical Reconstruction

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

There are challenges with each choice.

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 $\boldsymbol{y}$

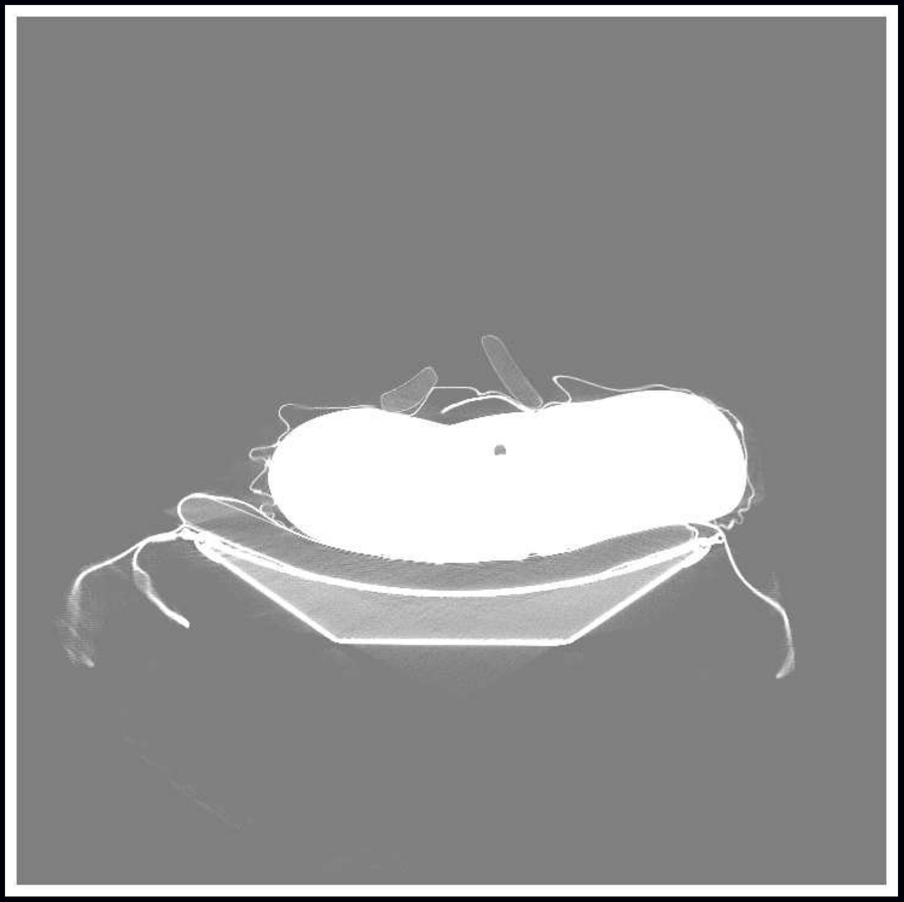
Challenges

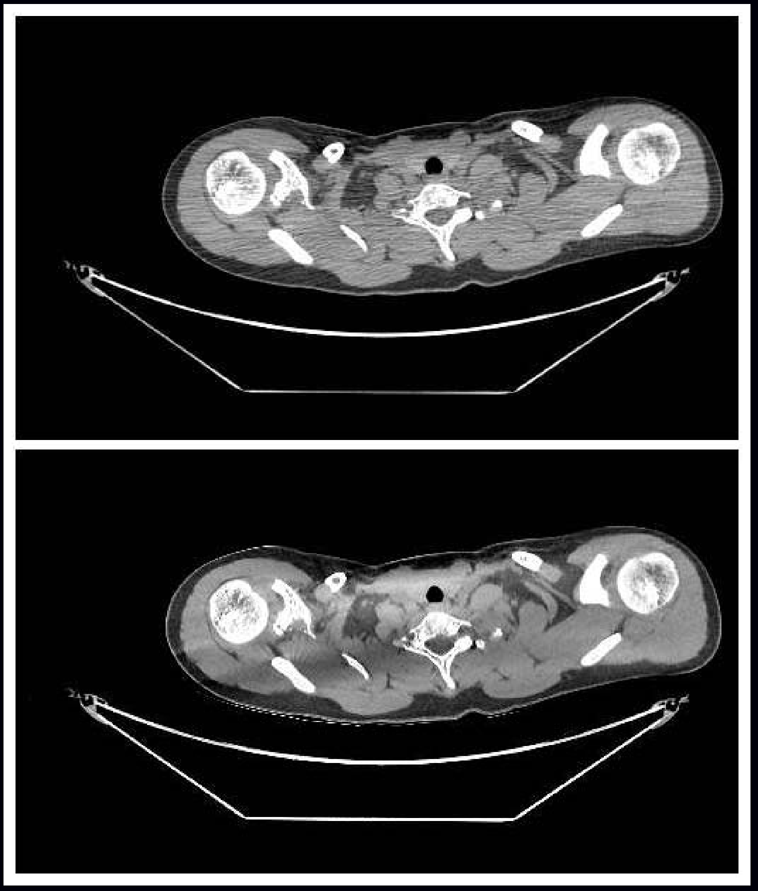
- Compact (voxels) versus band-limited (but bigger) blobs
- Choice of voxel size (resolution vs computation)
- Must cover entire object

  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

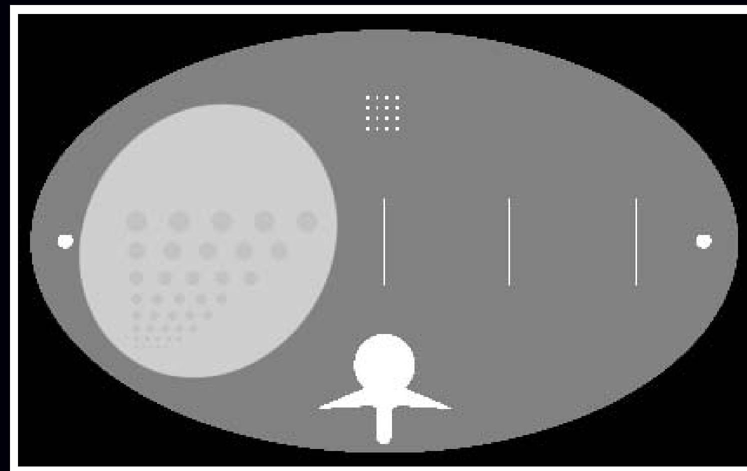
70-cm FOV reconstruction

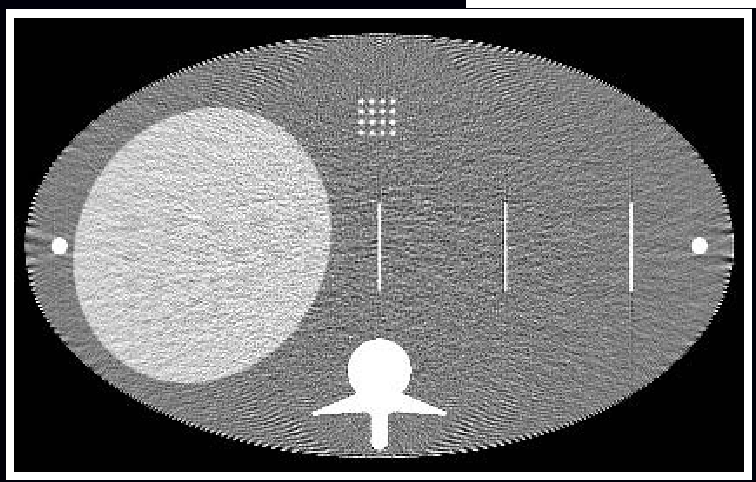




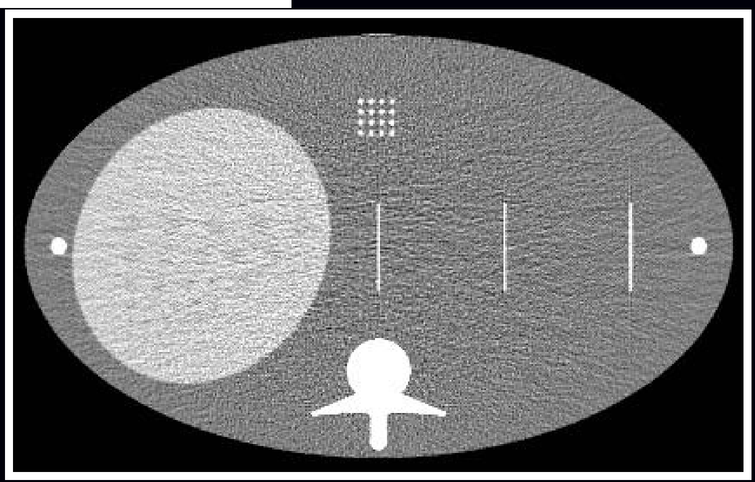
Thibault *et al.*, Fully3D, 2007

Voxel size matters?





512 grid



1024 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 Inhomogeneous
  - detector afterglow
- spectral effects
  - X-ray source spectrum
  - bowtie filters
  - detector spectra response
- scatter
- ...

Challenges

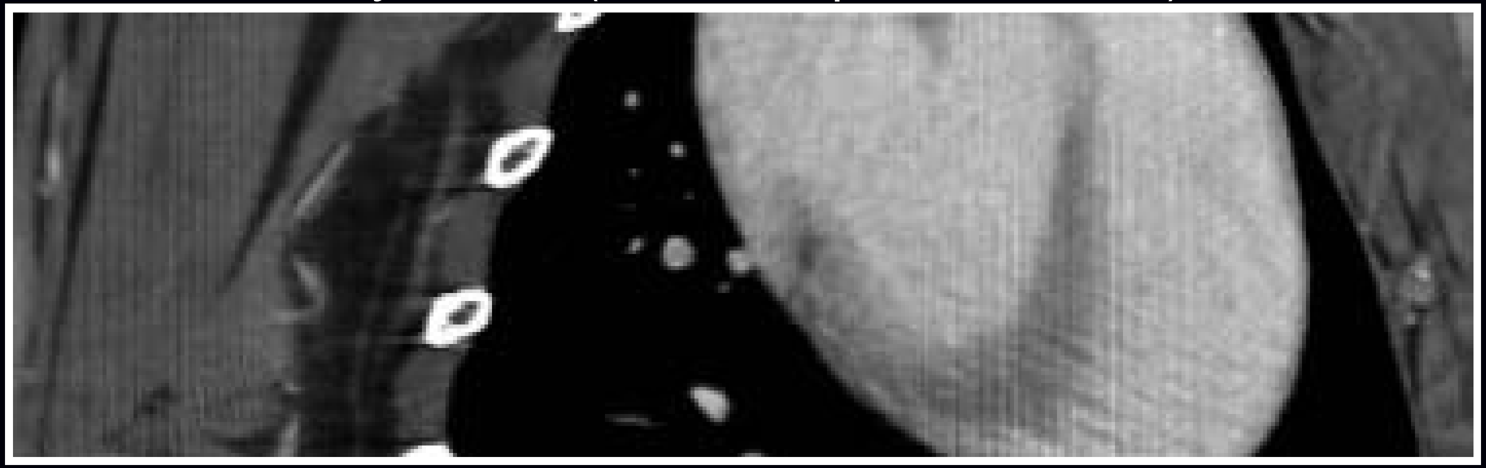
- computation time versus accuracy/artifacts/resolution/contrast
- dose?

Lines versus strips

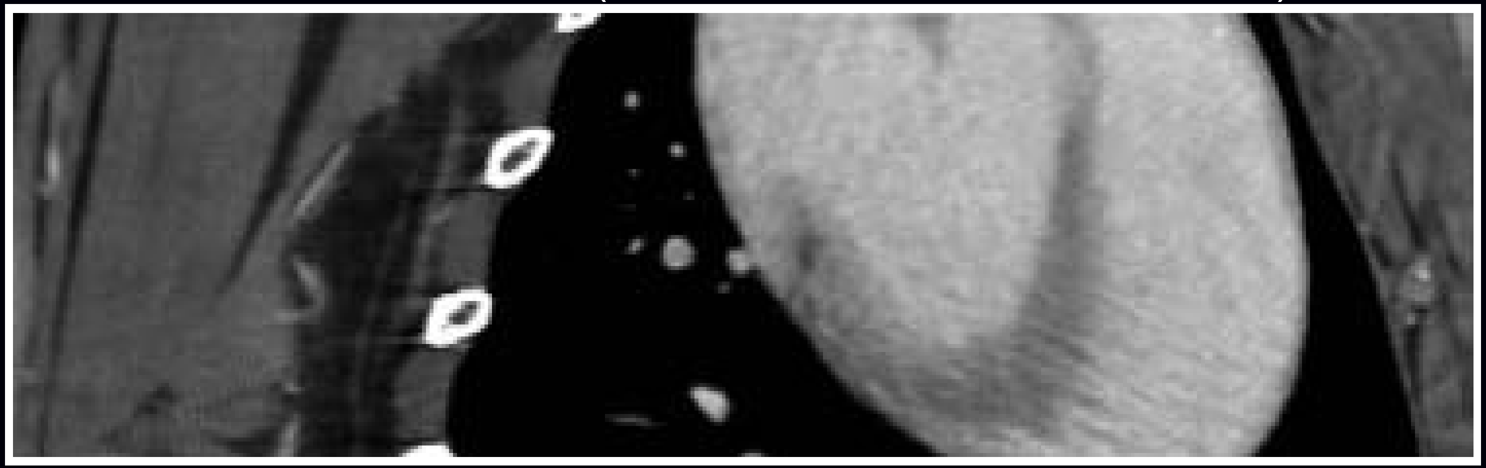
From (De Man and Basu, PMB, Jun. 2004)

MLTR of rabbit heart

Ray-driven (idealized point detector)



Distance-driven (models finite detector width)



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
- ...

Choice 3. Statistical Model

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

$$\bar{I}_i([\boldsymbol{A}\boldsymbol{x}]_i) = I_0 \, \mathrm{e}^{-\sum_{j=1}^{N} a_{ij} x_j}.$$

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

Challenges

- Trade-off between using more accurate statistical models (less noise?) and computation / complexity
- CT measurement statistics are very complicated, more so 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):

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

  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: $\bar{I}_i = \mathrm{e}^{-[\boldsymbol{A}\boldsymbol{x}]_i}$

After taking the log()

Taking the log leads to a 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 noise ε_i 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: $\sigma_i^2 = \frac{I_i + \sigma^2}{I_i^2}$
  Thibault *et al.*, SPIE 6065, 2006
- consider just Poisson effect: $\sigma_i^2 = \frac{1}{I_i}$ (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 minimizing a cost function:

$$\begin{aligned} \hat{\boldsymbol{x}} &\triangleq \arg\min_{\boldsymbol{x} \geq \boldsymbol{0}} \Psi(\boldsymbol{x}) \\ \Psi(\boldsymbol{x}) &= \text{DataMismatch}(\boldsymbol{y}, \boldsymbol{A}\boldsymbol{x}) + \beta\, \text{Regularizer}(\boldsymbol{x}) \end{aligned}$$

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

Controlling noise due to ill-conditioning via imposing constraints

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

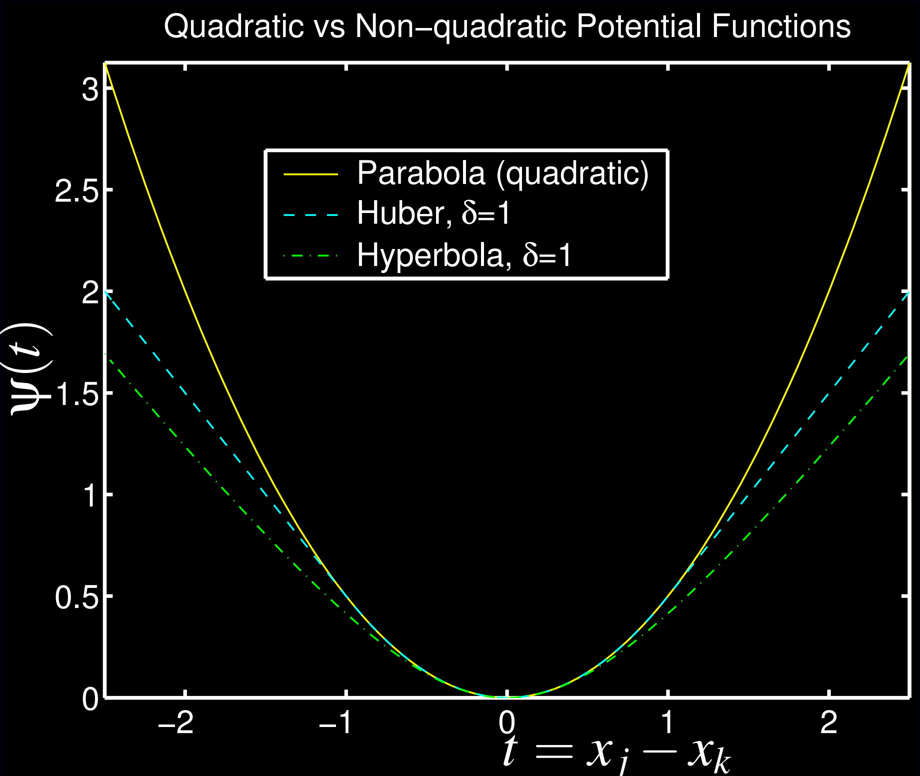
- Reducing noise without degrading spatial resolution
- Balancing regularization strength between and within slices
- Parameter selection
- Computational complexity (voxels have 26 neighbors in 3D)
- Preserving "familiar" noise texture
- Optimizing clinical task performance

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*



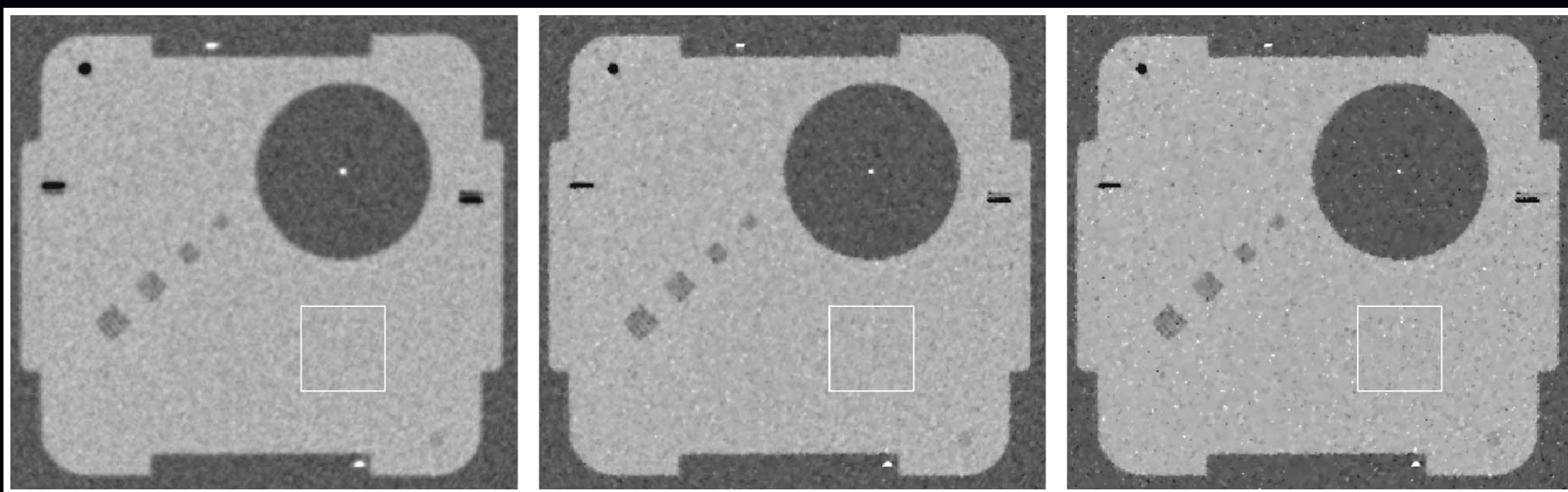
quadratic: $\psi(t) = t^2$
hyperbola: $\psi(t) = \sqrt{1 + (t/\delta)^2}$
(edge preservation)

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

Summary thus far

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

$$\boxed{\textbf{Reconstruction Method} \triangleq \textbf{Models + Cost Function + Algorithm}}$$

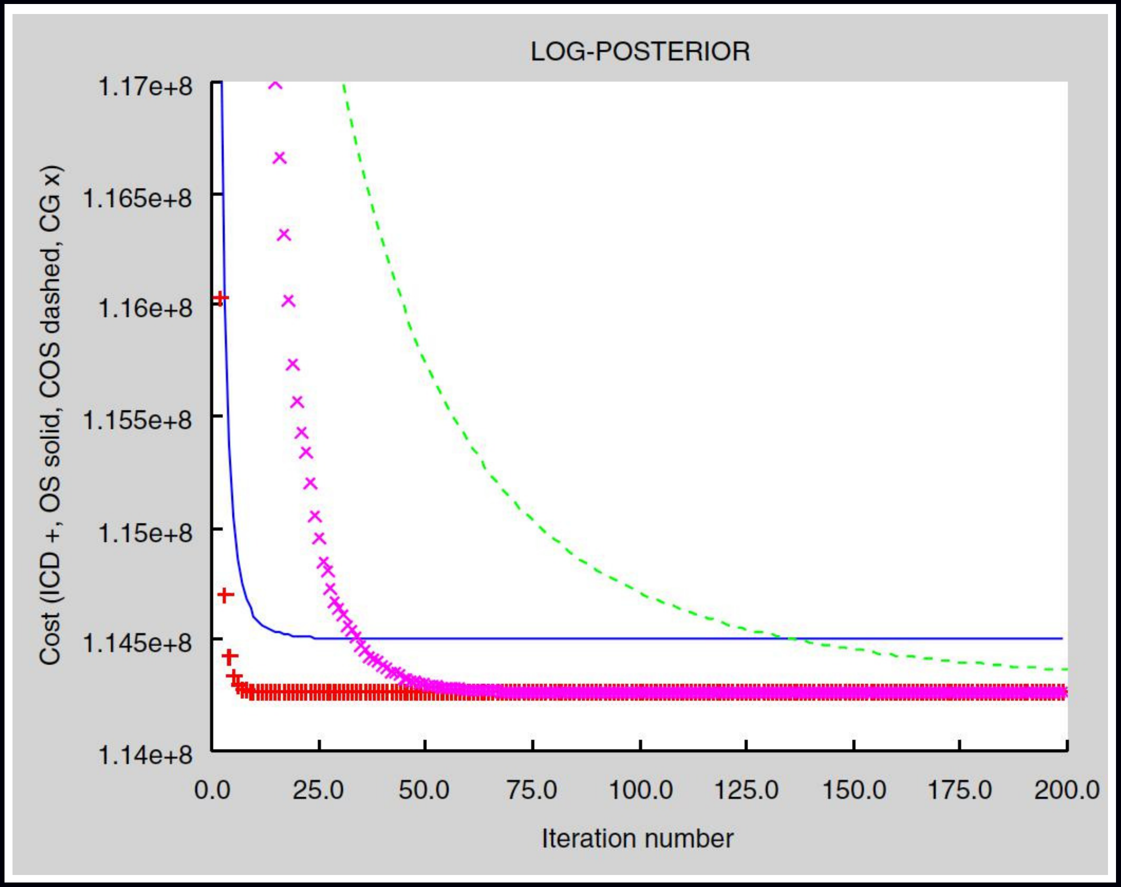
5. Minimization algorithms:

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

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

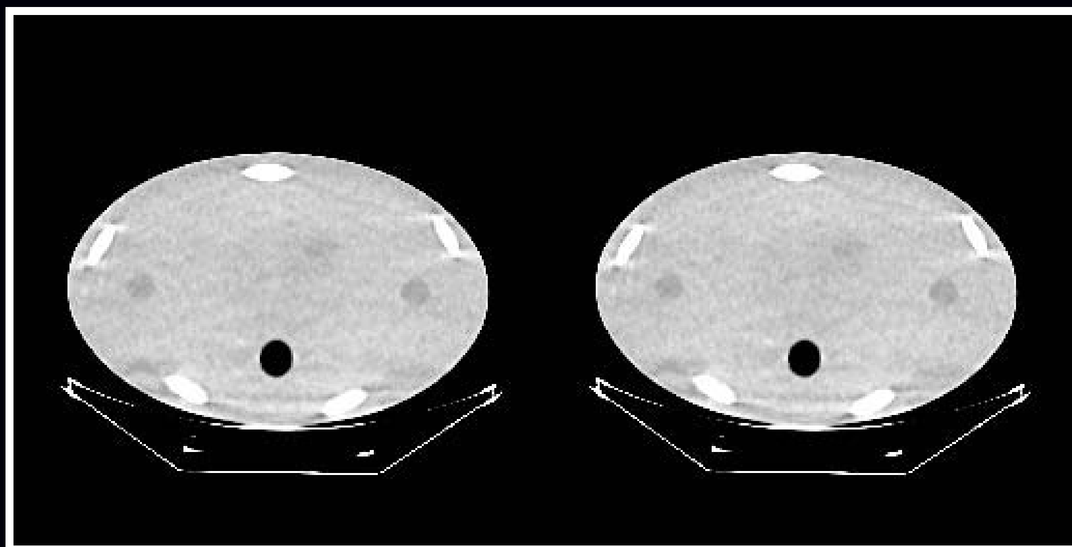


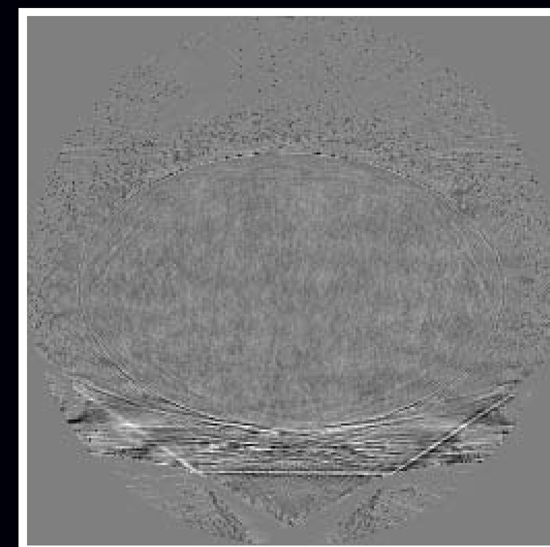
(De Man *et al.*, NSS/MIC 2005)

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

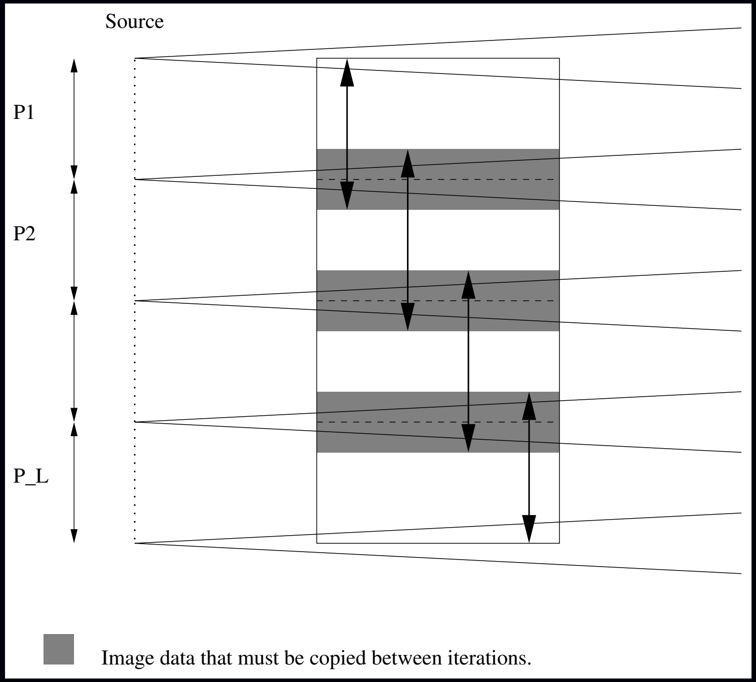
(De Man *et al.*, NSS/MIC 2005)

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

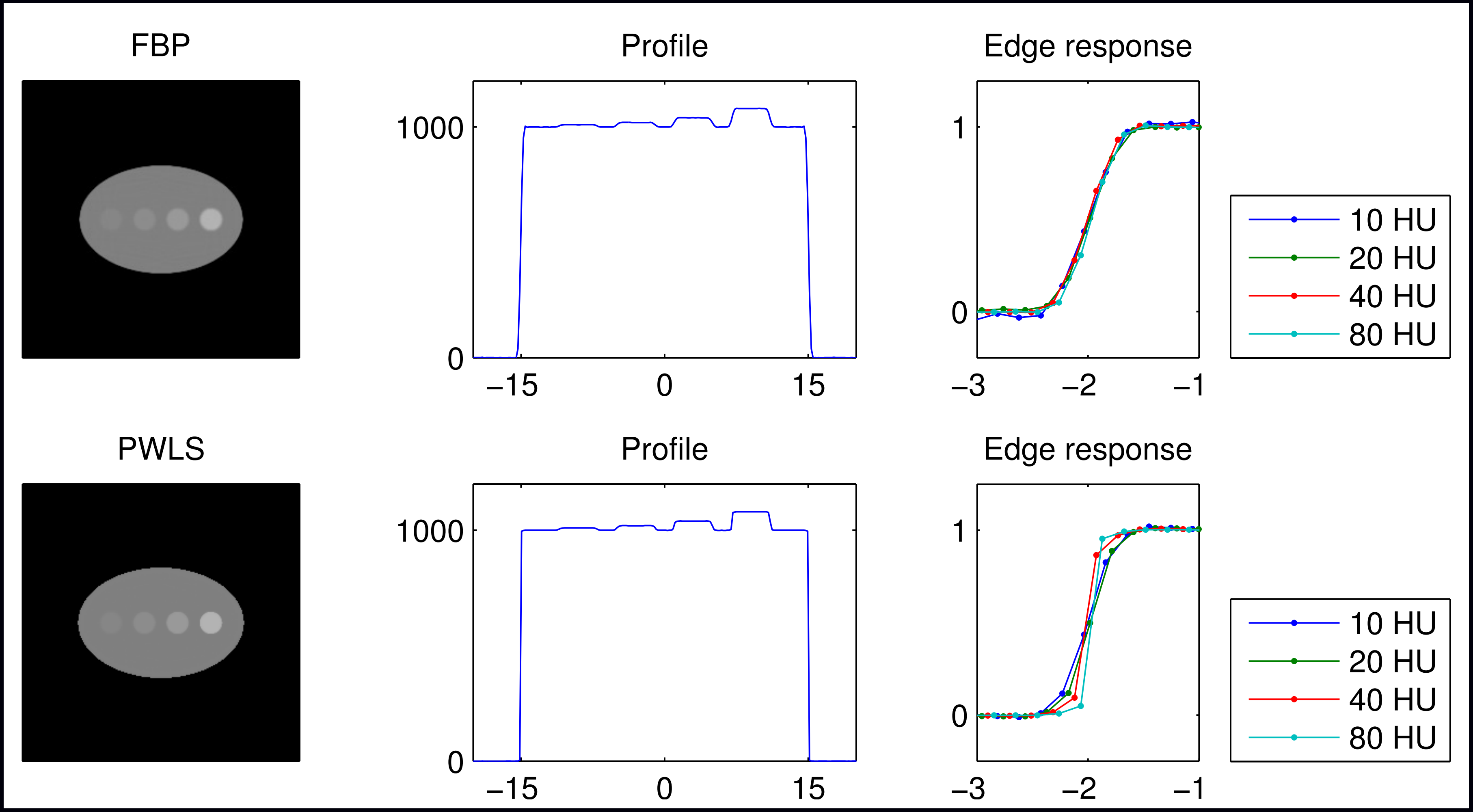
Optimization algorithms

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



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
- Task-based performance measures
  Known-location versus unknown-location tasks
- ...

“How low can the dose go” – quite challenging to answer

Some open problems

- 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
- ...

The CT reconstruction research environment

Challenges

- raw sinogram access is very limited (*cf.* PET, MRI)
- CT preprocessing steps are considered highly proprietary
- No standard databases of sinogram test data sets (*cf.* post-processing)
- Trainees perceived to be data processors, not CT scientists (*cf.* MRI)
- Commercial solutions largely will have predetermined parameters

Nevertheless, the CT reconstruction revolution will occur, like in PET...

Dr. Thrall: statistical image reconstruction will have the single largest bang for impacting dose.