Accelerating image reconstruction for low-dose X-ray CT and MRI

Jeffrey A. Fessler

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

web.eecs.umich.edu/~fessler

Work with Donghwan Kim, Madison McGaffin, Matt Muckley (and others)

CSP Seminar

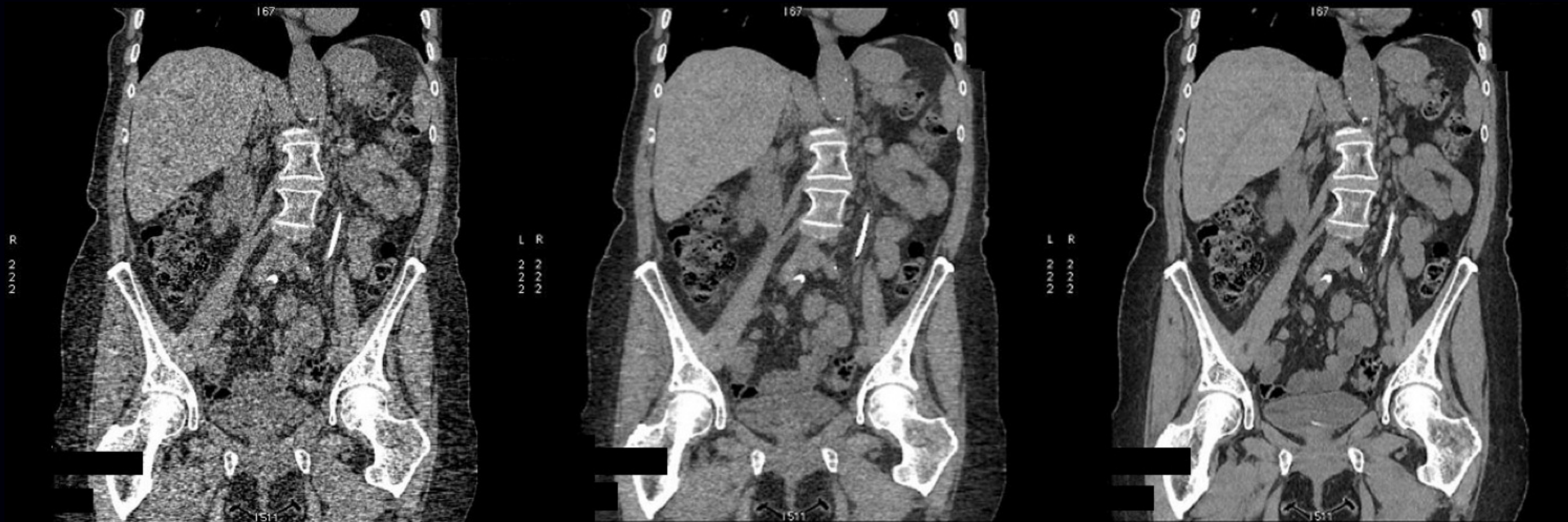
23 Apr. 2015

Disclosure

- Research support from GE Healthcare
- Supported in part by NIH grants P01 CA-87634, U01 EB018753
- Equipment support from Intel Corporation

Statistical image reconstruction: a 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

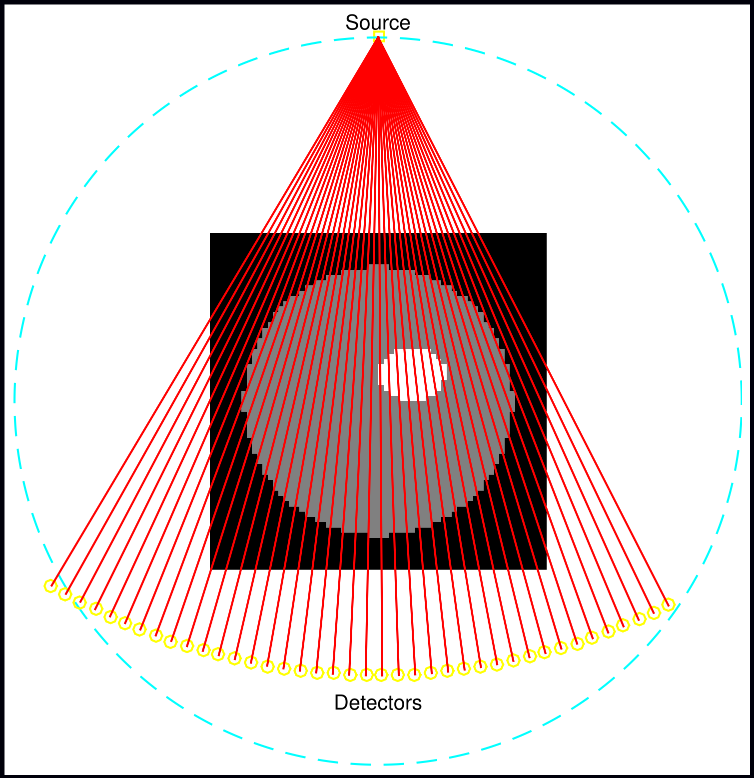
(Same sinogram, so all at same dose)

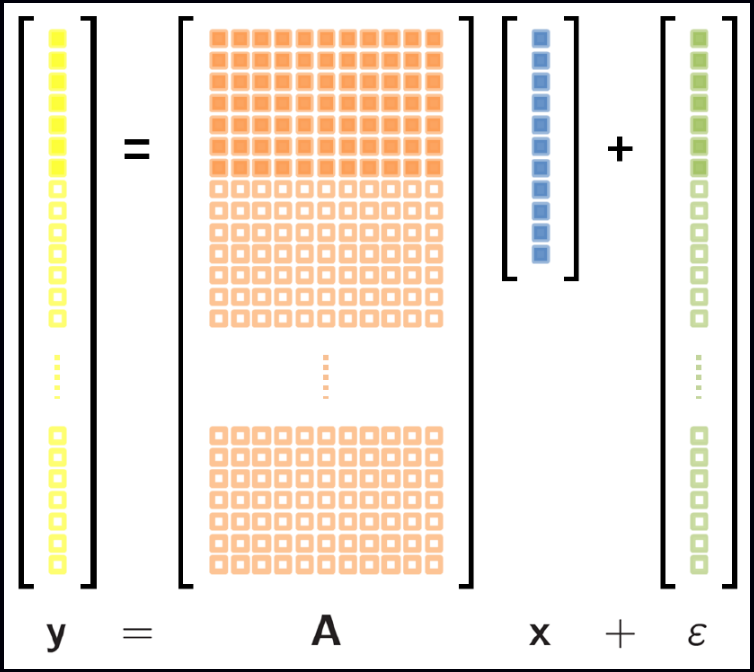
Outline

- Model-based image reconstruction
  - Low-dose X-ray CT
  - MRI
- Accelerating low-dose X-ray CT image reconstruction
  - Optimized first-order methods
    Donghwan Kim, JF; ArXiv 2014 Math. Prog., in review; ICIP 2015, submitted
  - Ordered-subsets + momentum
    Donghwan Kim, Sathish Ramani, JF; IEEE T-MI, Jan. 2015.
  - Distributed block-separable ordered subsets
    Donghwan Kim, JF; Fully 3D, 2015, to appear
  - Duality-based approach using GPU
    Madison G McGaffin, JF; Fully 3D, 2015, to appear
- Accelerating model-based MR image reconstruction
  - BARISTA (B1-based, adaptive restart, iterative soft thresholding algorithm)
    M. J. Muckley, D. C. Noll, JF; IEEE T-MI, Feb. 2015.

Conspicuously down-played: Alternating direction method of multipliers (ADMM)

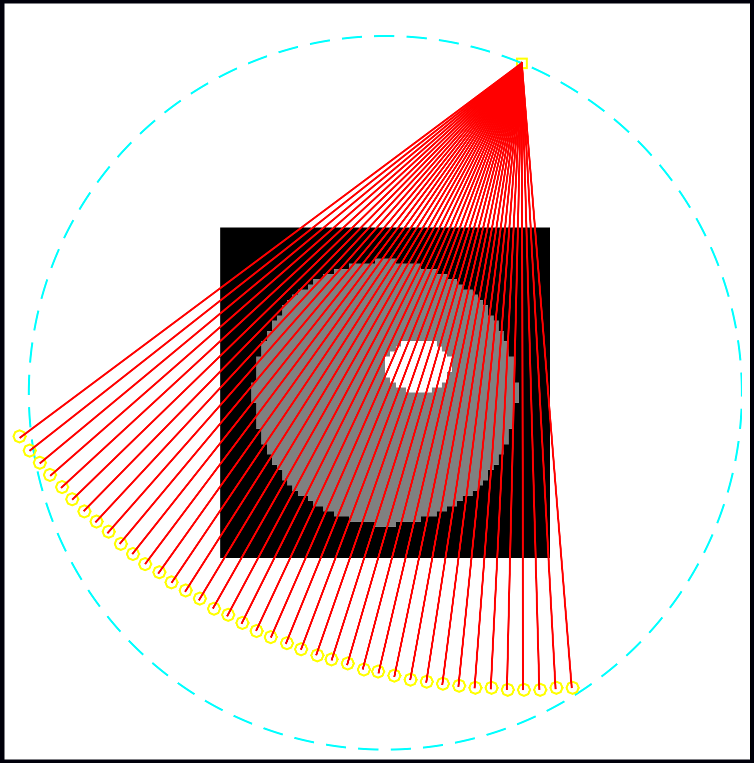
X-ray CT scans

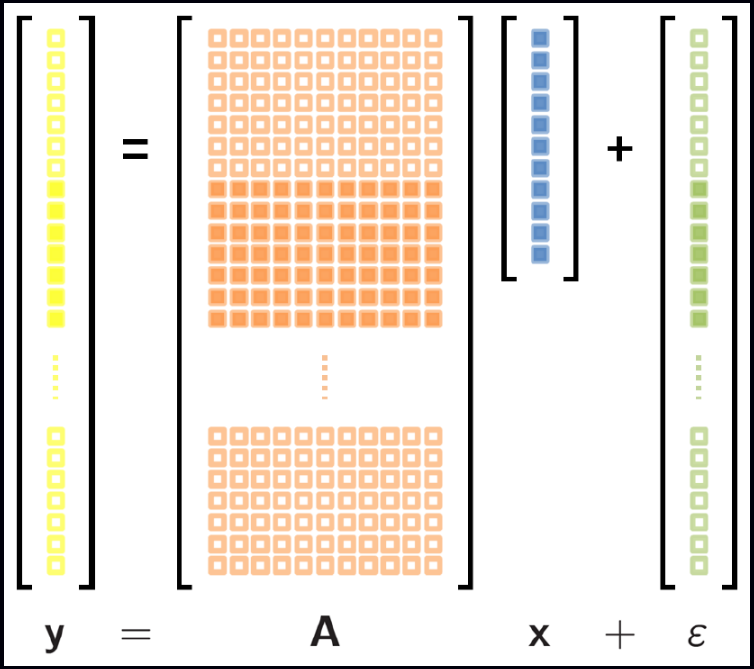




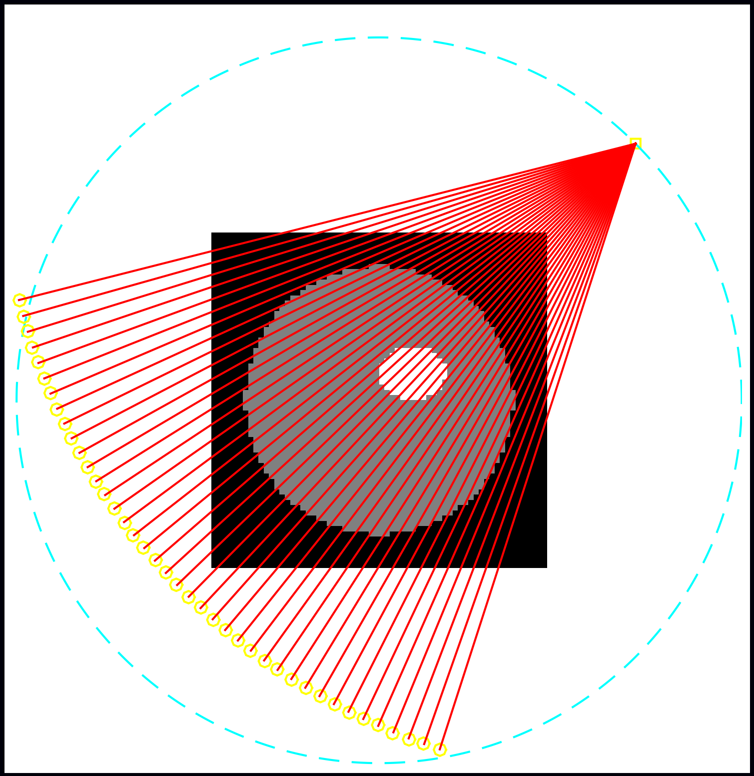
$\boldsymbol{y}$: measured data (sinogram)
$\boldsymbol{A}$: system matrix
$\boldsymbol{x}$: unknown image (attenuation map)

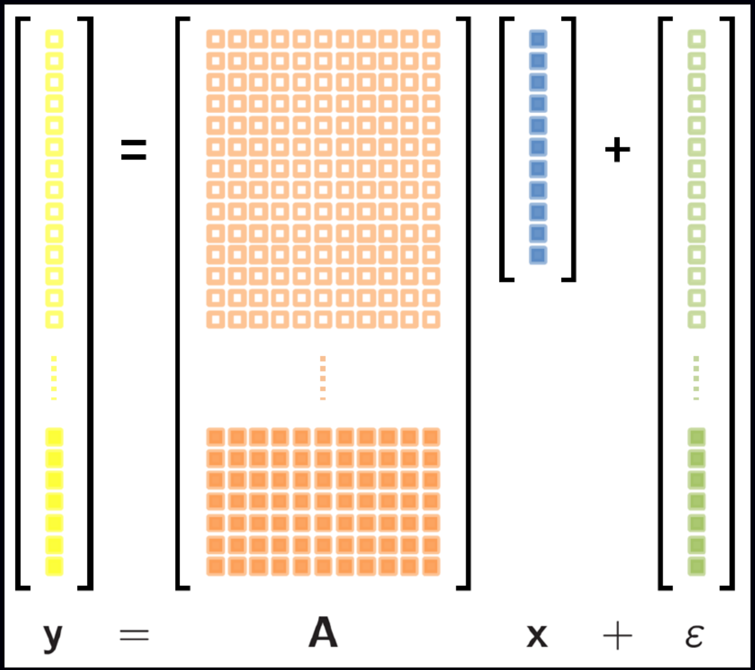
X-ray CT scans





X-ray CT scans





X-ray CT scans

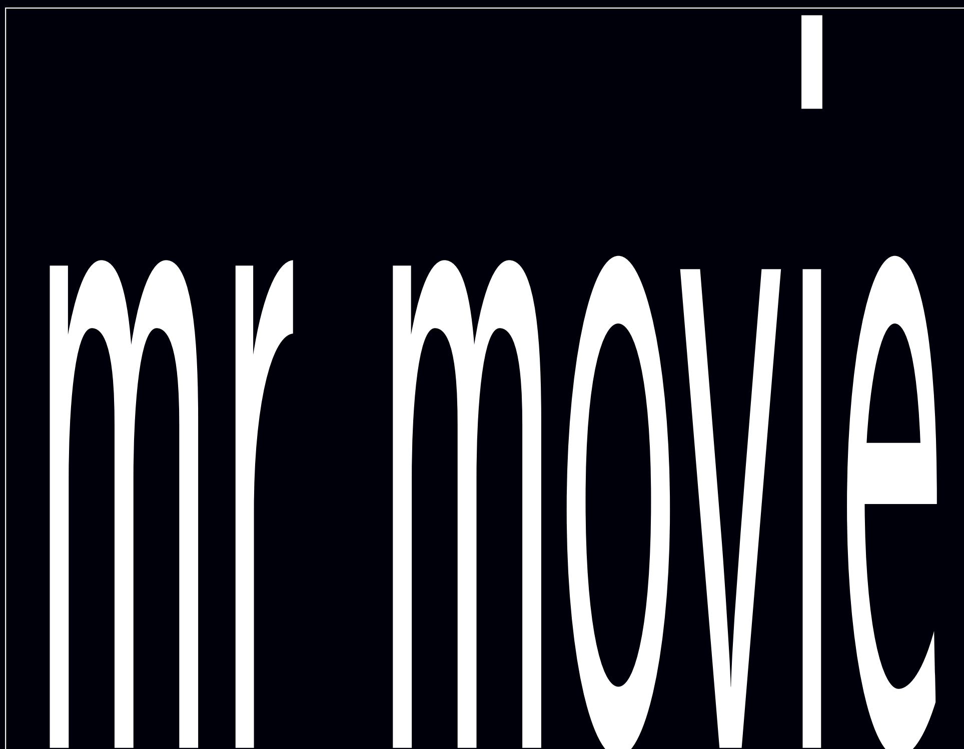
ct movie

CT image reconstruction problem:
Determine attenuation map $\boldsymbol{x}$ from sinogram data $\boldsymbol{y}$

Ignoring motion hereafter...

MRI scans

No moving parts to animate...



MR image reconstruction problem:
Determine magnetization image $\boldsymbol{x}$ from k-space data $\boldsymbol{y}$

Inverse problems

Unknown object $\boldsymbol{x}$ $\rightarrow$ [Imaging system $\boldsymbol{A}$ physics, noise, ...] $\rightarrow$ Measurements $\boldsymbol{y}$

How to reconstruct object $\boldsymbol{x}$ from data $\boldsymbol{y}$?

Classical approach:

- analytical / direct / non-iterative
  - Filtered back-projection (FBP) for CT
  - Inverse FFT for MRI
- idealized description of the system
  - geometry / sampling
  - disregards noise and simplifies physics
- typically fast

Contemporary approach:

- model-based / statistical / iterative
- based on "reasonably accurate" models for physics and statistics
- usually much slower

Why statistical/iterative methods for CT?

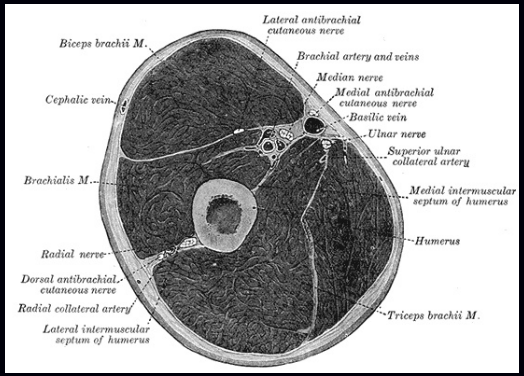
- Accurate physics models
  - X-ray spectrum, beam-hardening, scatter, ...
    $\Longrightarrow$ reduced artifacts? quantitative CT?
  - X-ray detector spatial response, focal spot size, ...
    $\Longrightarrow$ improved spatial resolution?
  - detector spectral response (*e.g.*, photon-counting detectors)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

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

and more...

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



Henry Gray, Anatomy of the Human Body, 1918, Fig. 413.

These constraints may help reduce image artifacts or noise or dose.

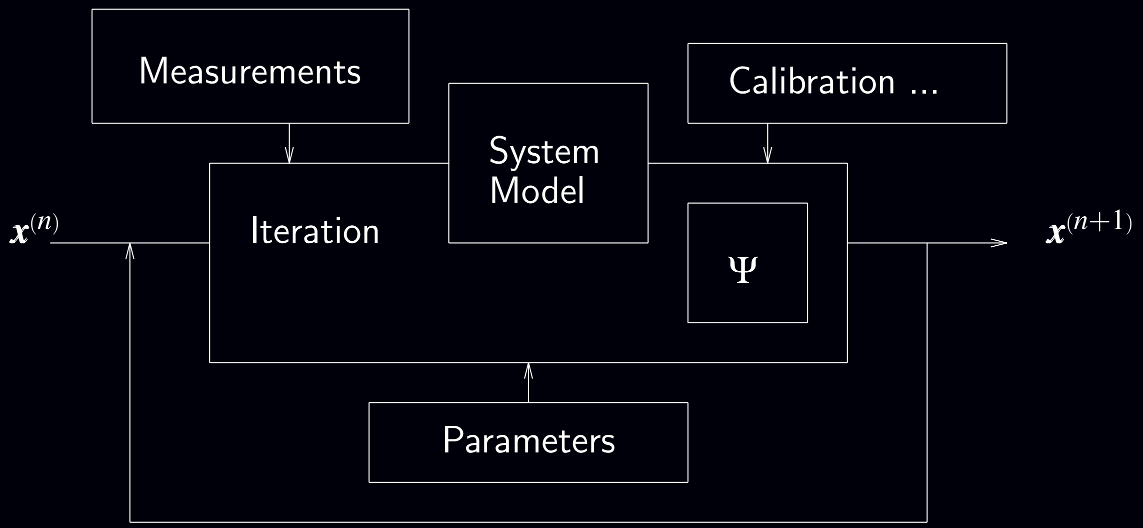
Disadvantages?

- Computation time (super computer)
- Must reconstruct entire FOV
- Complexity of models and software
- Algorithm nonlinearities
  - Difficult to analyze resolution/noise properties (*cf.* FBP)
  - Tuning parameters
  - Challenging to characterize performance / assess image quality

Statistical image reconstruction overview

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

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



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

Sub-mSv example

3D helical X-ray CT scan of abdomen/pelvis:
100 kVp, 25-38 mA, 0.4 second rotation, 0.625 mm slice, 0.6 mSv.

FBP ASIR Statistical

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) images courtesy of Jiang Hsieh, GE Healthcare

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

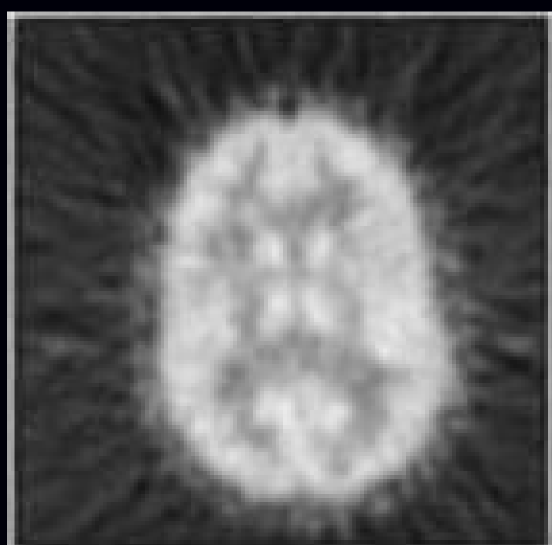
Today, most (all?) commercial PET systems include *unregularized* OSEM, and recently possibly some regularized version.

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

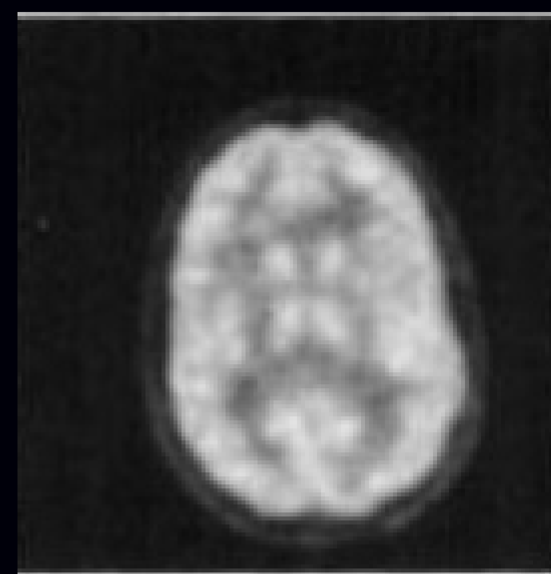
Key factors in PET

- 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:

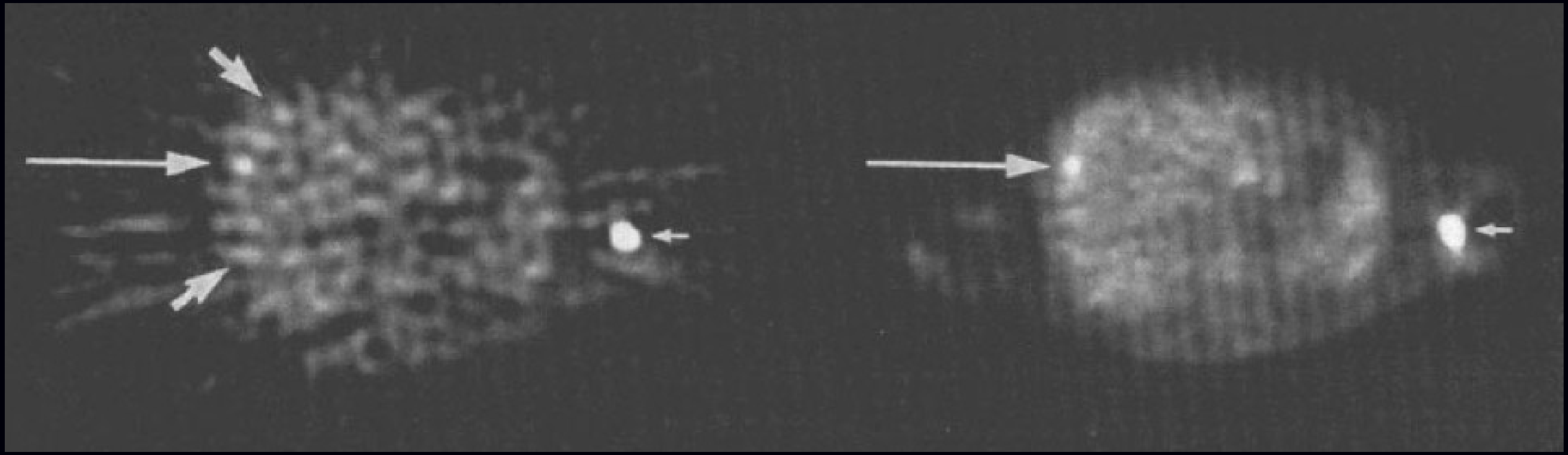


ML-EM:



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 X-ray 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 OS for Philips BrightView SPECT-CT 2010
- Commercial introduction of ICD for CT scanners circa 2010
- FDA 510(k) clearance of Veo Sep. 2011
- First Veo installation in USA (at UM) Jan. 2012

(* numerous omissions, including many denoising methods)

Statistical image reconstruction for low-dose CT

Optimization problem formulation:

$$\hat{\boldsymbol{x}} = \underbrace{\arg\min_{\boldsymbol{x} \geq \boldsymbol{0}}}_{\substack{\text{optimization}\\ \text{algorithm}}} \Psi(\boldsymbol{x}), \qquad \underbrace{\Psi(\boldsymbol{x})}_{\substack{\text{cost}\\ \text{function}}} \triangleq \underbrace{\frac{1}{2}\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_{\boldsymbol{W}}^2}_{\substack{\text{data-fit term}\\ \text{physics \& statistics}}} \underbrace{+ \beta \sum_{j=1}^{N} \sum_{k \in \mathcal{N}_j} \psi(x_j - x_k)}_{\substack{\text{regularizer}\\ \text{prior models}}}$$

$\boldsymbol{y}$: measured data (sinogram)
$\boldsymbol{A}$: system matrix (physics / geometry)
$\boldsymbol{W}$: weighting matrix (statistics)
$\boldsymbol{x}$: unknown image (attenuation map)
ψ : edge-preserving potential function (piece-wise smoothness / sparse gradients)

Optimization challenges:

- large problem size: $\boldsymbol{x} \in \mathbb{R}^{512 \times 512 \times 600}$, $\boldsymbol{y} \in \mathbb{R}^{888 \times 64 \times 7000}$
- $\boldsymbol{A}$ is sparse but still too large to store; compute $\boldsymbol{A}\boldsymbol{x}$ on-the-fly
- $\boldsymbol{W}$ has enormous dynamic range (1 to $\exp(-9) \approx 1.2 \cdot 10^{-4}$)
- Gram matrix $\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}$ highly shift variant
- Ψ is non-quadratic but convex (and often smooth)

Optimization algorithms for X-ray CT

Classical gradient descent (GD)

Assumptions:

- Ψ is convex (need not be strictly convex)
- Ψ has non-empty set of global minimizers
  $\hat{\boldsymbol{x}} \in \mathscr{X}^* = \left\{ \boldsymbol{x}^{(\star)} \in \mathbb{R}^N : \Psi(\boldsymbol{x}^{(\star)}) \leq \Psi(\boldsymbol{x}),\ \forall \boldsymbol{x} \in \mathbb{R}^N \right\}$
- Ψ is smooth (differentiable with L-Lipshitz gradient)
  $\|\nabla \Psi(\boldsymbol{x}) - \nabla \Psi(\boldsymbol{z})\|_2 \leq L \|\boldsymbol{x} - \boldsymbol{z}\|_2, \quad \forall \boldsymbol{x}, \boldsymbol{z} \in \mathbb{R}^N$

Gradient descent (GD) with step size $1/L$ ensures monotonic descent of Ψ:

$$\boldsymbol{x}^{(n+1)} = \boldsymbol{x}^{(n)} - \frac{1}{L} \nabla \Psi\left(\boldsymbol{x}^{(n)}\right).$$

Drori & Teboulle (2014) derive tightest "inaccuracy" bound:

$$\underbrace{\Psi\left(\boldsymbol{x}^{(n)}\right) - \Psi\left(\boldsymbol{x}^{(\star)}\right)}_{\text{inaccuracy}} \leq \frac{L \left\|\boldsymbol{x}^{(0)} - \boldsymbol{x}^{(\star)}\right\|_2^2}{4n+2}.$$

They construct a Huber-like function Ψ for which GD achieves that (tight) bound.

But $O(1/n)$ rate is undesirably slow.

Nesterov's fast gradient method (FGM1)

Nesterov (1983) iteration: Initialize: $t_0 = 1$, $\boldsymbol{z}^{(0)} = \boldsymbol{x}^{(0)}$

$$\boldsymbol{z}^{(n+1)} = \boldsymbol{x}^{(n)} - \frac{1}{L}\nabla\Psi\left(\boldsymbol{x}^{(n)}\right) \qquad \text{(usual GD update)}$$

$$t_{n+1} = \frac{1}{2}\left(1 + \sqrt{1 + 4t_n^2}\right) \qquad \text{(magic momentum factors)}$$

$$\boldsymbol{x}^{(n+1)} = \boldsymbol{z}^{(n+1)} + \frac{t_n - 1}{t_{n+1}}\left(\boldsymbol{z}^{(n+1)} - \boldsymbol{z}^{(n)}\right) \qquad \text{(update with momentum)}\ .$$

Reverts to GD if $t_n = 1, \forall n$.

Shown by Nesterov to be $O(1/n^2)$ for "primary" sequence:

$$\Psi\left(\boldsymbol{z}^{(n)}\right) - \Psi\left(\boldsymbol{x}^{(\star)}\right) \leq \frac{2L\left\|\boldsymbol{x}^{(0)} - \boldsymbol{x}^{(\star)}\right\|_2^2}{(n+1)^2}.$$

Nesterov constructed a function Ψ such that any first-order method achieves

$$\frac{\frac{3}{32}L\left\|\boldsymbol{x}^{(0)} - \boldsymbol{x}^{(\star)}\right\|_2^2}{(n+1)^2} \leq \Psi\left(\boldsymbol{x}^{(n)}\right) - \Psi\left(\boldsymbol{x}^{(\star)}\right).$$

Thus $O(1/n^2)$ rate of FGM1 is optimal.

Donghwan Kim (2014) analyzed "secondary" sequence: $\Psi(\boldsymbol{x}^{(n)}) - \Psi(\boldsymbol{x}^{(\star)}) \leq \dfrac{2L\left\|\boldsymbol{x}^{(0)} - \boldsymbol{x}^{(\star)}\right\|_2^2}{(n+2)^2}$.

Generalizing Nesterov's FGM

FGM1 is in the general class of first-order methods:

$$\boldsymbol{x}^{(n+1)} = \boldsymbol{x}^{(n)} - \frac{1}{L}\sum_{k=0}^{n} h_{n+1,k} \nabla\Psi\left(\boldsymbol{x}^{(k)}\right)$$

where the step-size factors $\{h_{n,k}\}$ are given by:

$$h_{n+1,k} = \begin{cases} \frac{t_n - 1}{t_{n+1}} h_{n,k}, & k = 0,\ldots,n-2 \\ \frac{t_n - 1}{t_{n+1}} \left(h_{n,n-1} - 1\right), & k = n-1 \\ 1 + \frac{t_n - 1}{t_{n+1}}, & k = n. \end{cases}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.25 & 0 & 0 & 0 & 0 \\ 0 & 0.10 & 1.40 & 0 & 0 & 0 \\ 0 & 0.05 & 0.20 & 1.50 & 0 & 0 \\ 0 & 0.03 & 0.11 & 0.29 & 1.57 & 0 \\ 0 & 0.02 & 0.07 & 0.18 & 0.36 & 1.62 \end{bmatrix}$$

Note use of previous gradients $\Longrightarrow$ "momentum"

Is this the optimal choice for $\{h_{n,k}\}$?
Can we do better than the constant 2 in worst-case convergence rate?

Drori & Teboulle (2014) numerically found $\{h_{n,k}\}$ that are factor of two better.
(Factors of two matter practically.)

Optimized gradient method (OGM1)

New approach by optimizing $\{h_{n,k}\}$ analytically (Donghwan Kim and JF; 2014, 2015):

Initialize: $t_0 = 1$, $\boldsymbol{z}^{(0)} = \boldsymbol{x}^{(0)}$

$$\boldsymbol{z}^{(n+1)} = \boldsymbol{x}^{(n)} - \frac{1}{L}\nabla\Psi\left(\boldsymbol{x}^{(n)}\right) \qquad \text{(usual GD update)}$$

$$t_{n+1} = \frac{1}{2}\left(1 + \sqrt{1 + 4t_n^2}\right) \qquad \text{(momentum factors)}$$

$$\boldsymbol{x}^{(n+1)} = \boldsymbol{z}^{(n+1)} + \frac{t_n - 1}{t_{n+1}}\left(\boldsymbol{z}^{(n+1)} - \boldsymbol{z}^{(n)}\right) + \underbrace{\frac{t_n}{t_{n+1}}\left(\boldsymbol{z}^{(n+1)} - \boldsymbol{x}^{(n)}\right)}_{\text{new momentum}} .$$

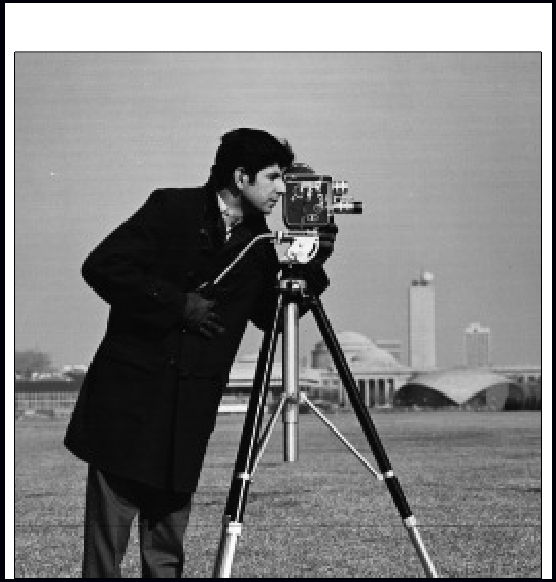
Smaller (worst-case) convergence bound than Nesterov by factor of 2:

$$\Psi\left(\boldsymbol{z}^{(n)}\right) - \Psi\left(\boldsymbol{x}^{(\star)}\right) \le \frac{1L\left\|\boldsymbol{x}^{(0)} - \boldsymbol{x}^{(\star)}\right\|_2^2}{(n+1)^2}.$$

Recently (very) DK found a Huber-like function for which OGM1 achieves that upper bound (thus tight), inspired by numerical work of Taylor *et al.* (2015).

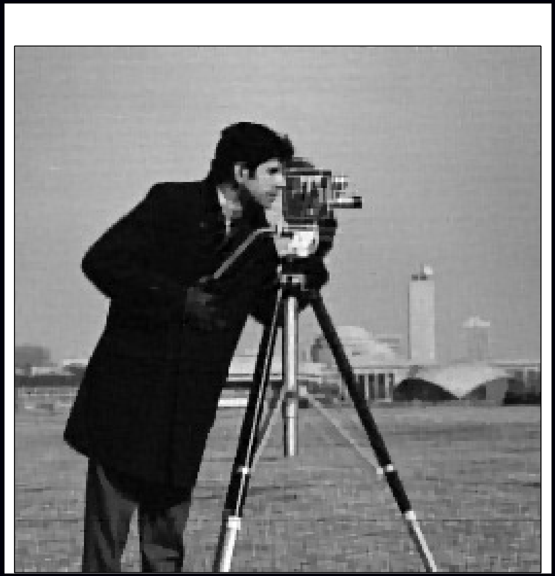
Example: Image restoration (!?)

True $\boldsymbol{x}$:

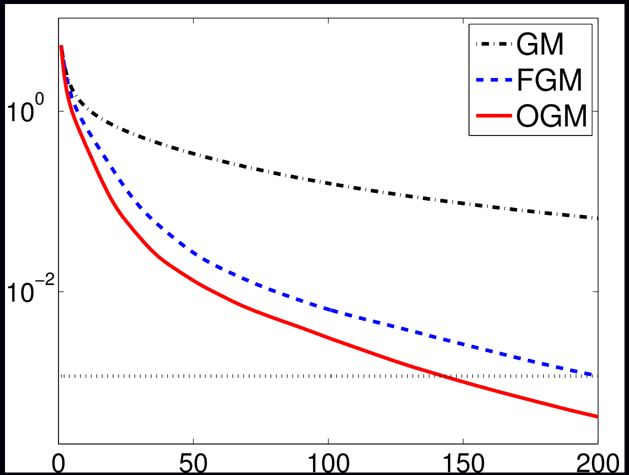


Blurred $\boldsymbol{y}$:

Restored
$\hat{\boldsymbol{x}}$



Rate:
$\Psi(\boldsymbol{x}^{(n)}) - \Psi(\hat{\boldsymbol{x}})$
vs iteration n



Ordered subsets version of OGM1

For further acceleration, combine OGM with ordered subsets (OS),

$$\Psi(\boldsymbol{x}) = \sum_{m=1}^{M} \Psi_m(\boldsymbol{x}), \quad \Psi_m(\boldsymbol{x}) \triangleq \underbrace{\frac{1}{2} \|\boldsymbol{y}_m - \boldsymbol{A}_m \boldsymbol{x}\|^2_{\boldsymbol{W}_m}}_{1/M\text{th of measurements}} + \frac{1}{M} \mathsf{R}(\boldsymbol{x})$$

(aka incremental gradients, *cf.* stochastic gradient descent)

Initialize: $t_0 = 1$, $\boldsymbol{z}^{(0)} = \boldsymbol{x}^{(0)}$
for $n = 0, 1, \ldots$ (iteration)
  for $m = 1, \ldots, M$ (subset)
    $k = nM + m$ (subiteration)

$$\boldsymbol{z}^{k+1} = \left[\boldsymbol{x}^k - \boldsymbol{D} M \nabla \Psi_m\left(\boldsymbol{x}^k\right)\right]_+ \text{ (typical OS-SQS)}$$

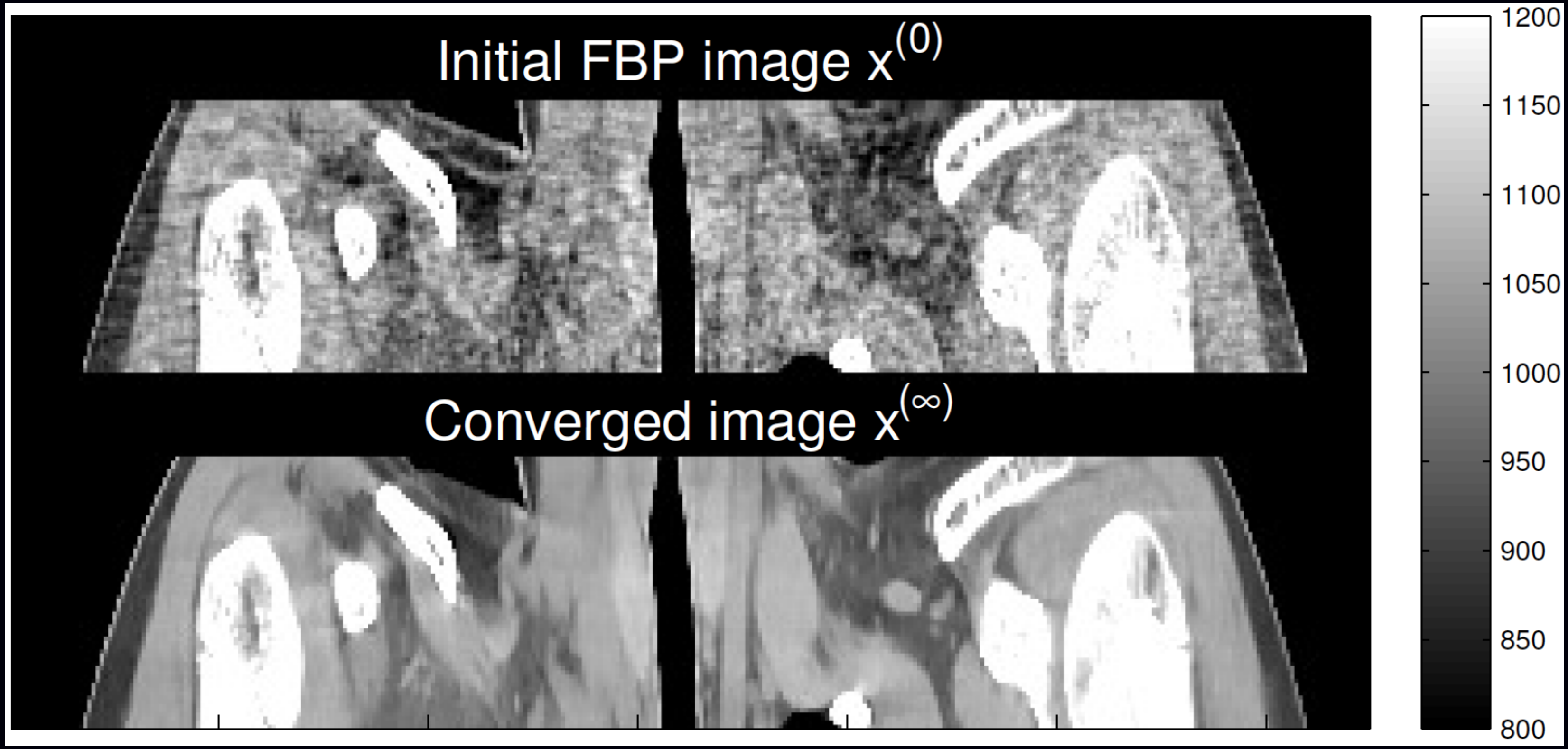
$$t_{k+1} = \frac{1}{2}\left(1 + \sqrt{1 + 4t_k^2}\right)$$

$$\boldsymbol{x}^{k+1} = \boldsymbol{z}^{k+1} + \frac{t_k - 1}{t_{k+1}}\left(\boldsymbol{z}^{k+1} - \boldsymbol{z}^k\right) + \frac{t_k}{t_{k+1}}\left(\boldsymbol{z}^{k+1} - \boldsymbol{x}^k\right) \text{ (momentum)}$$

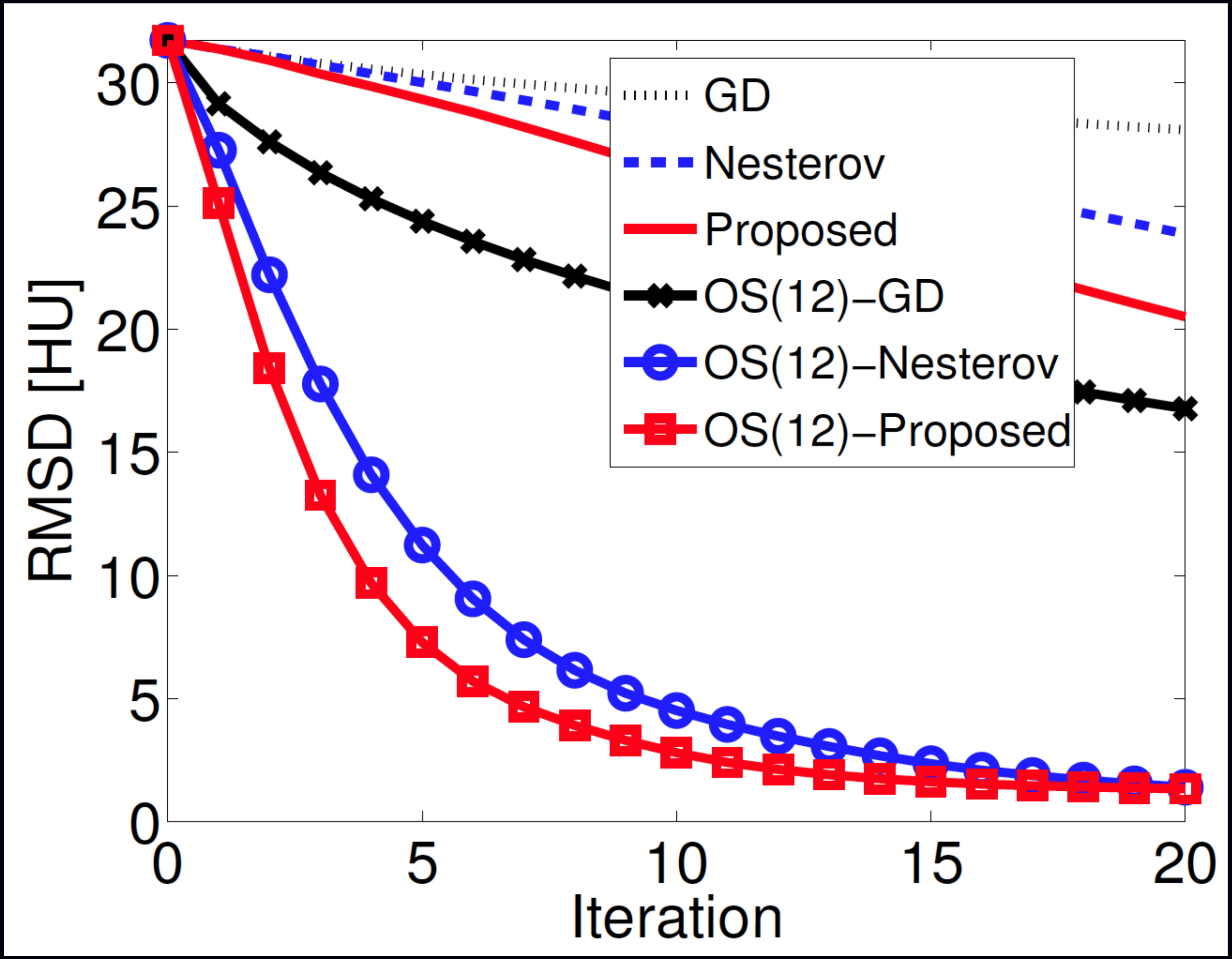
Approximate convergence rate for Ψ: $O\left(\frac{1}{n^2 M^2}\right)$ (Donghwan Kim and JF; CT 2014)
Now fast enough to show an X-ray CT example...

OS+OGM results: data

- 3D cone-beam helical X-ray CT scan
- pitch 0.5
- image $\boldsymbol{x}$: $512 \times 512 \times 109$ with 70 cm FOV and 0.625 mm slices
- sinogram : $\boldsymbol{y}$ 888 detectors $\times$ 32 rows $\times$ 7146 views

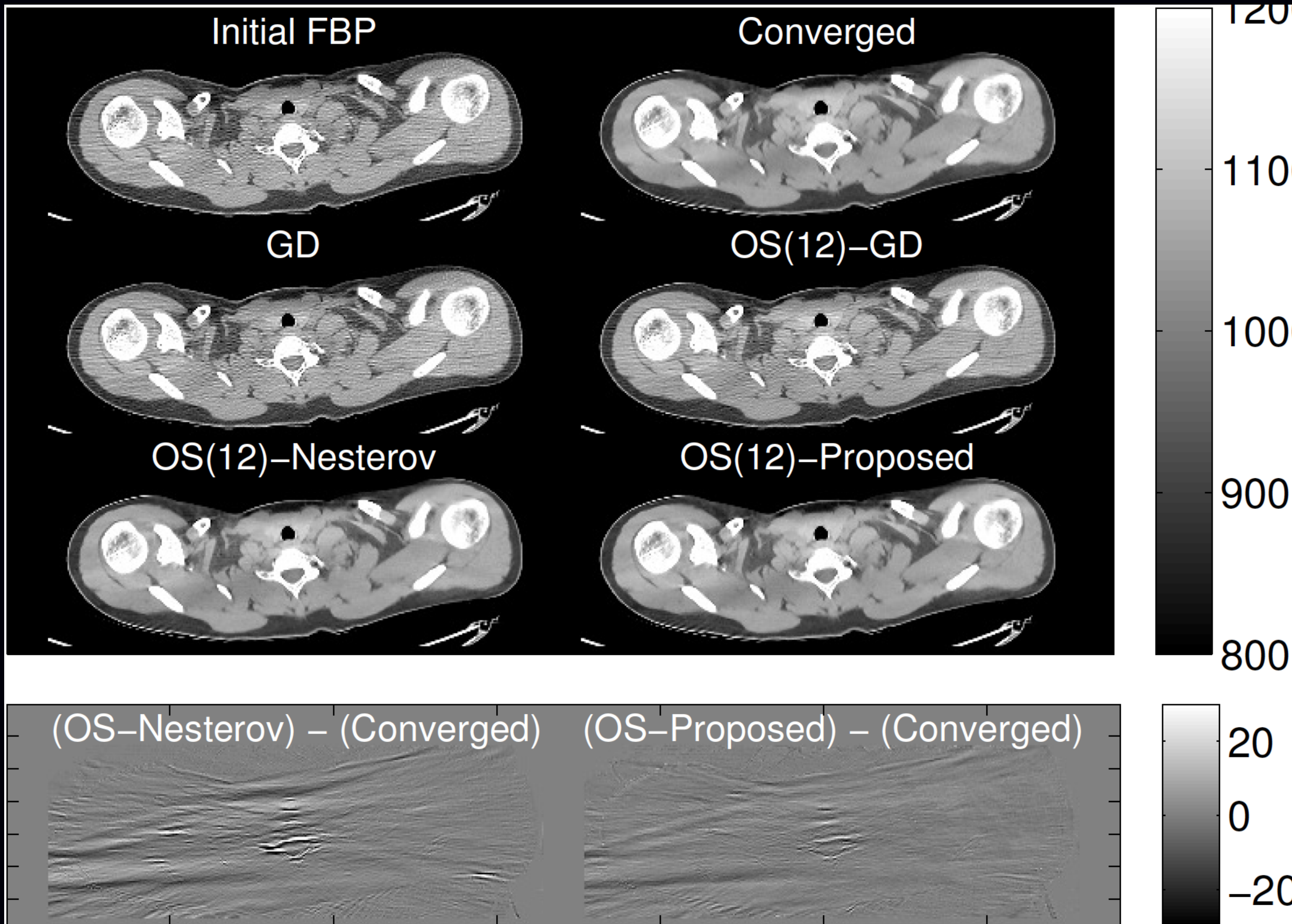


OS+OGM results: convergence rate



Root mean square difference (RMSD) between $\boldsymbol{x}^{(n)}$ and $\boldsymbol{x}^{(\infty)}$ over ROI (in HU), versus iteration. (Compute time per iteration very similar.)

OS+OGM results: images



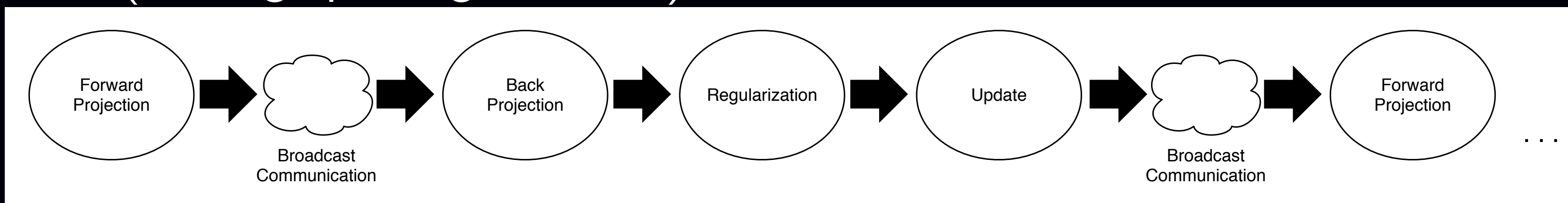
At iteration 10 with $M = 12$ subsets.

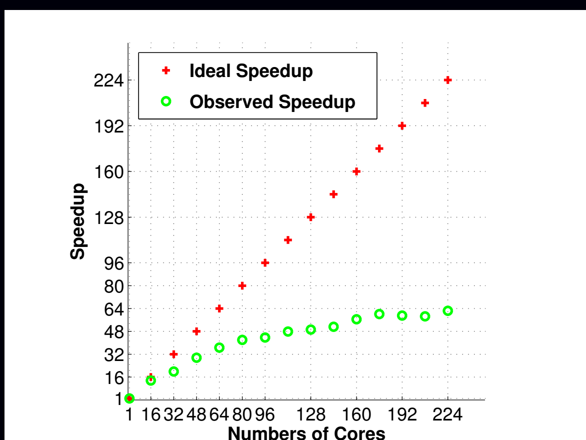
Towards parallel computing

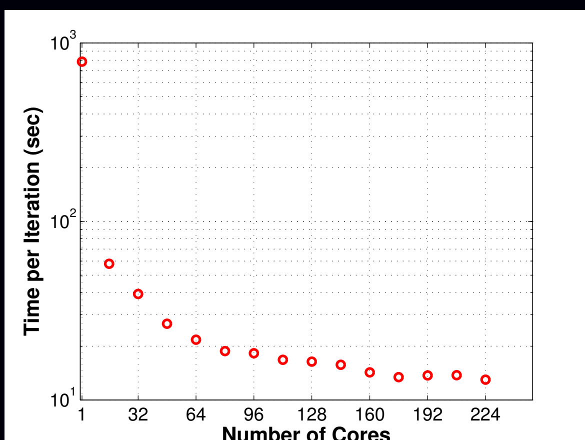
Amazon Cloud version of OS+OGM

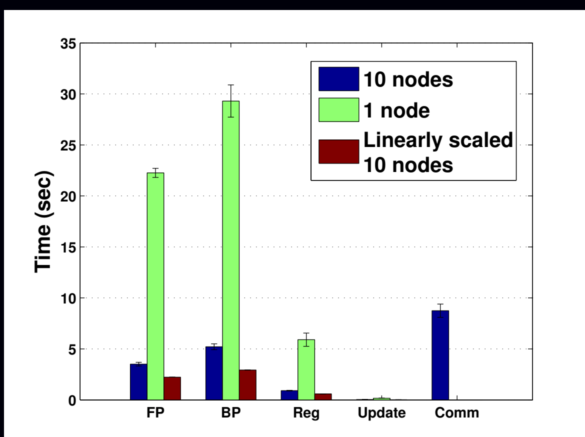
Distribute long object (320 useful slices) into (overlapping) slabs (128 slices each) across 5 separate clusters, each with 10 nodes having 16 cores.

Use MPI (message passing interface) for within-cluster communication:





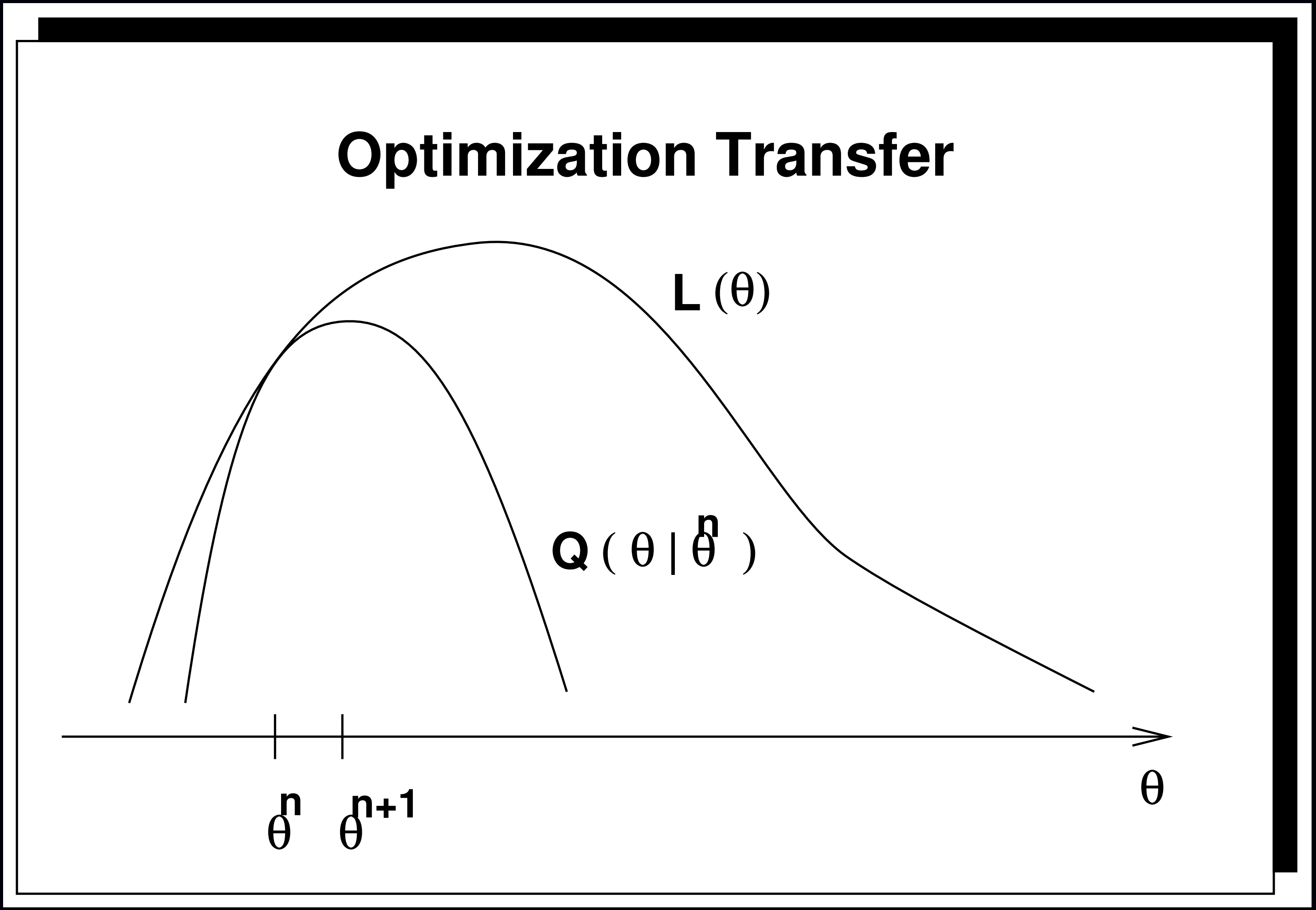




Rosen, Wu, Wenisch, JF (Fully 3D, 2013)

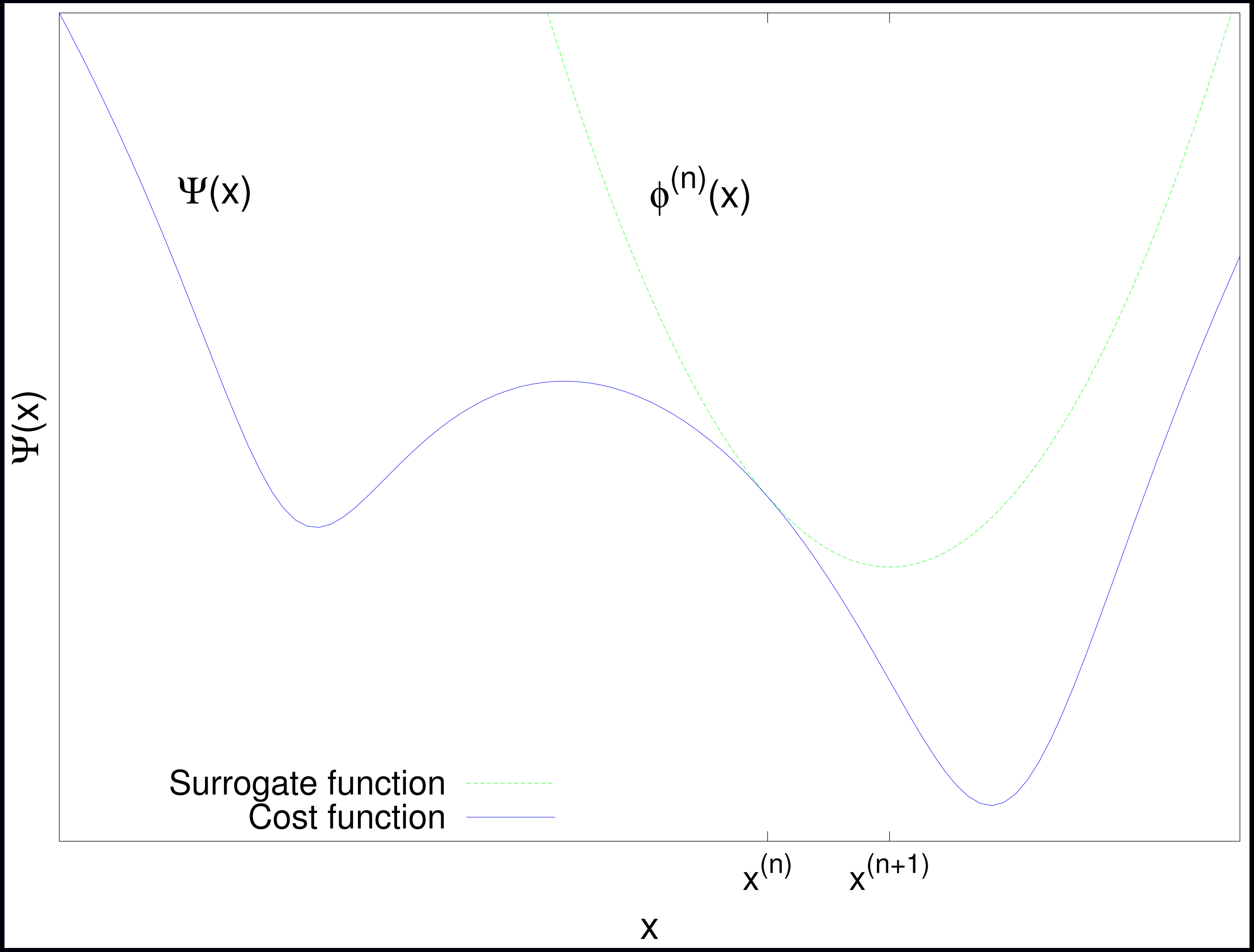
- Overlapping slabs is inefficient
- Communication time (within cluster, after *every subset*) is serious bottleneck

Optimization transfer (Majorize-Minimize) methods



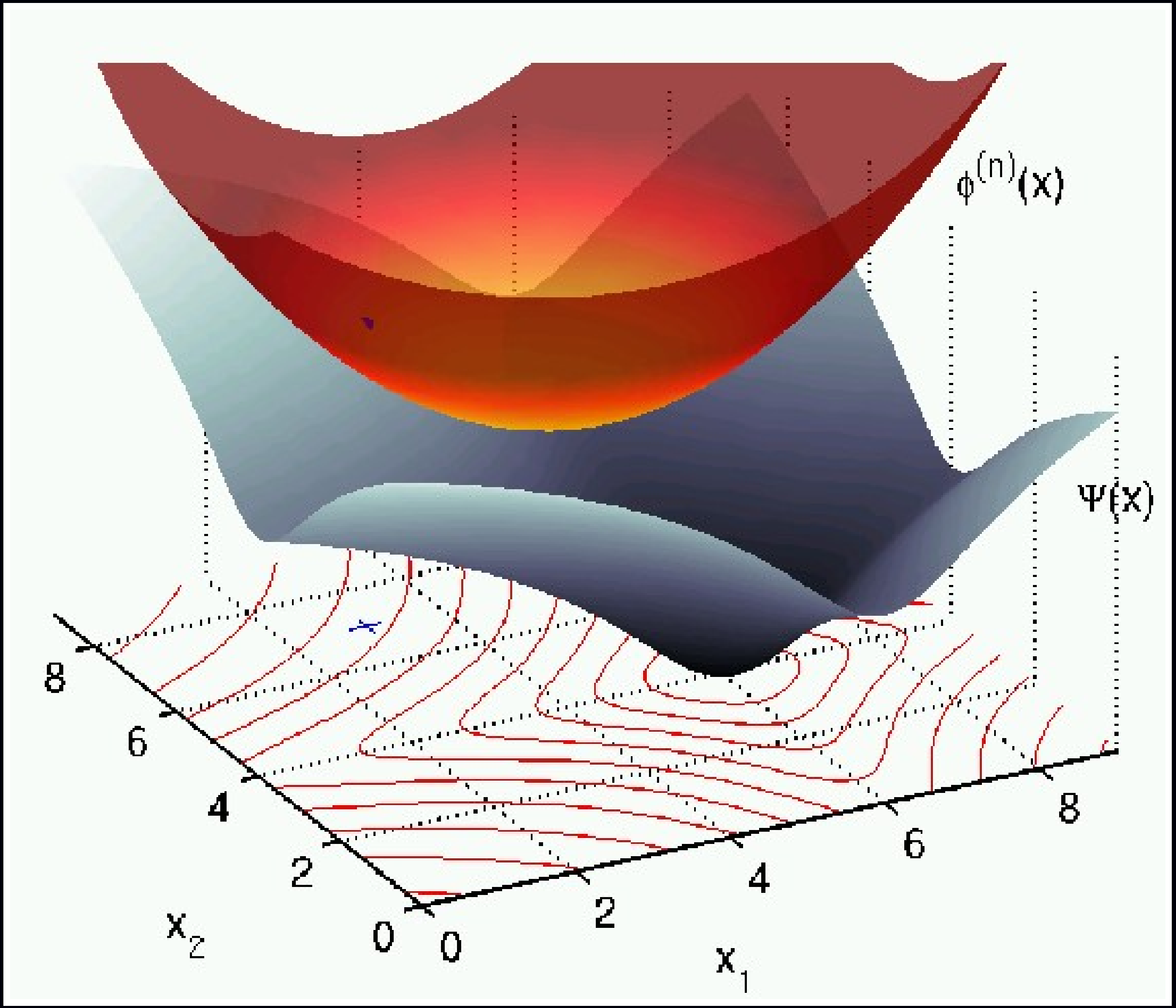
(March 1995 version)

Optimization transfer (Majorize-Minimize) methods: 1D



$$\boldsymbol{x}^{(n+1)} = \arg\min_{\boldsymbol{x}} \phi^{(n)}(\boldsymbol{x})$$

Optimization transfer (Majorize-Minimize) methods: 2D



Block-separable surrogates for distributed reconstruction

Conventional OS approach uses a (voxel) separable quadratic surrogate (SQS):

$$\begin{aligned}\Psi(\boldsymbol{x}) &\leq \Psi\left(\boldsymbol{x}^{(n)}\right) + \nabla\Psi\left(\boldsymbol{x}^{(n)}\right)(\boldsymbol{x}-\boldsymbol{x}^{(n)}) + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}^{(n)})'\boldsymbol{D}(\boldsymbol{x}-\boldsymbol{x}^{(n)}) \\ &= \Psi\left(\boldsymbol{x}^{(n)}\right) + \sum_{j=1}^{N} \frac{\partial}{\partial x_j}\Psi\left(\boldsymbol{x}^{(n)}\right)(x_j - x_j^{(n)}) + \frac{1}{2} d_j \left(x_j - x_j^{(n)}\right)^2,\end{aligned}$$

where diagonal matrix $\boldsymbol{D}$ majorizes the Hessian of Ψ: $\nabla^2\Psi(\boldsymbol{x}) \preceq \boldsymbol{D}$.

Distributed computing alternative: derive slab-separable surrogate instead:

$$\Psi(\boldsymbol{x}) - \Psi\left(\boldsymbol{x}^{(n)}\right) \leq \sum_{b=1}^{B} \Psi_b(\boldsymbol{x}_b), \quad \Psi_b(\boldsymbol{x}_b) \triangleq \nabla_{\boldsymbol{x}_b}\Psi\left(\boldsymbol{x}^{(n)}\right)(\boldsymbol{x}_b - \boldsymbol{x}_b^{(n)}) + \frac{1}{2}\left(\boldsymbol{x}_b - \boldsymbol{x}_b^{(n)}\right)' \boldsymbol{H}_b \left(\boldsymbol{x}_b - \boldsymbol{x}_b^{(n)}\right),$$

where *block* diagonal matrix $\boldsymbol{H} = \operatorname{diag}\{\boldsymbol{H}_1, \ldots, \boldsymbol{H}_B\}$ majorizes the Hessian of Ψ.

$$\boldsymbol{H}_b \triangleq \boldsymbol{A}_b'\boldsymbol{W}\Lambda_b\boldsymbol{A}_b, \quad \Lambda_b \triangleq \operatorname{diag}\{\boldsymbol{A}\boldsymbol{1} \oslash \boldsymbol{A}_b\boldsymbol{1}_b\}$$

Updates parallelizable across blocks (slabs): (Donghwan Kim and JF; Fully 3D, 2015)

$$\boldsymbol{x}_b^{(n+1)} \triangleq \underset{\boldsymbol{x}_b \succeq \boldsymbol{0}}{\arg\min}\, \Psi_b(\boldsymbol{x}_b).$$

Reduces communication. (Apply favorite optimization method within slab.)

Block-separable surrogate (BSS) OS+OGM

1: Initialize $\tilde{\boldsymbol{x}}^{(0)}$ by FBP, and compute $\boldsymbol{D}$.
2: Distribute image $\tilde{\boldsymbol{x}}^{(0)}$ and data $\boldsymbol{y}$ into B nodes.
3: **for** $n = 0, 1, \ldots$
4: Minimize $\phi_{\mathrm{BSS}}(\boldsymbol{x}; \tilde{\boldsymbol{x}}^{(n)})$ using L sub-iterations of OS-SQS-mom.

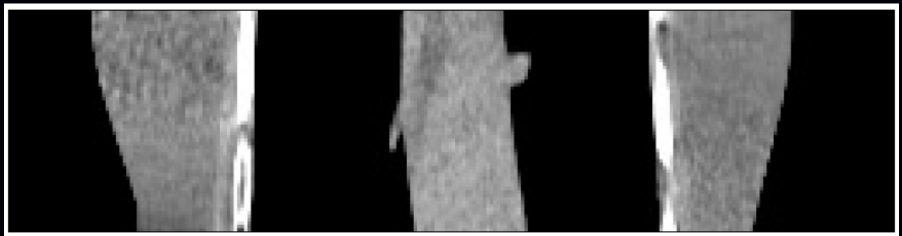
1) Initialize $\boldsymbol{x}^{(0)} = \boldsymbol{z}^{(0)}$ by $\tilde{\boldsymbol{x}}^{(n)}$, and $t^{(0)} = 1$.
2) **for** $l = 0, 1, \ldots, L-1$
3) $m = l \bmod M$
4) $t^{(l+1)} = \frac{1}{2}\left(1 + \sqrt{1 + 4\left[t^{(l)}\right]^2}\right)$
5) **for** $b = 1, \ldots, B$ **simultaneously**
6) $\boldsymbol{g}_{m,b}^{(l)} = M \nabla_b \phi_{\mathrm{BSS},m}(\boldsymbol{z}^{(\frac{l}{M})}; \boldsymbol{z}^{(0)})$ [subset gradient]
7) $\boldsymbol{x}_b^{(\frac{l+1}{M})} = \left[\boldsymbol{z}_b^{(\frac{l}{M})} - \boldsymbol{D}_b^{-1} \boldsymbol{g}_{m,b}^{(l)}\right]_+$ [OS-SQS update]
8) $\boldsymbol{z}_b^{(\frac{l+1}{M})} = \boldsymbol{x}_b^{(\frac{l+1}{M})} + \frac{t^{(l)}-1}{t^{(l+1)}}\left(\boldsymbol{x}_b^{(\frac{l+1}{M})} - \boldsymbol{x}_b^{(\frac{l}{M})}\right)$ [momentum]
9) **end for**
10) **end for**
11) $\tilde{\boldsymbol{x}}^{(n+1)} = \boldsymbol{x}^{(\frac{L}{M})}$

5: Communicate $\tilde{\boldsymbol{x}}^{(n+1)}$.
6: **end for**

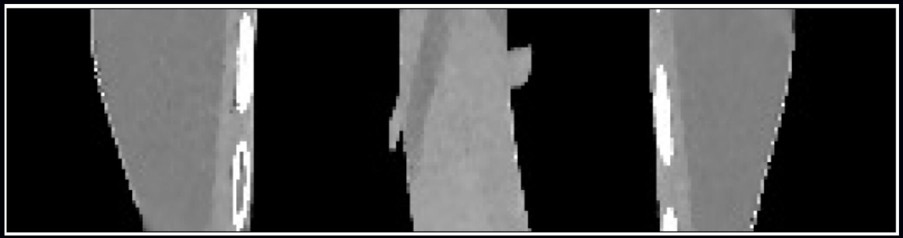
BSS OS+OGM: data

- $256 \times 256 \times 160$ XCAT phantom (Segars *et al.*, 2008)
- Simulated helical CT, $444 \times 32 \times 492$
- $M = 12$ subsets, $B = 10$ blocks, $L = 5$ inner iterations
- Matlab emulation

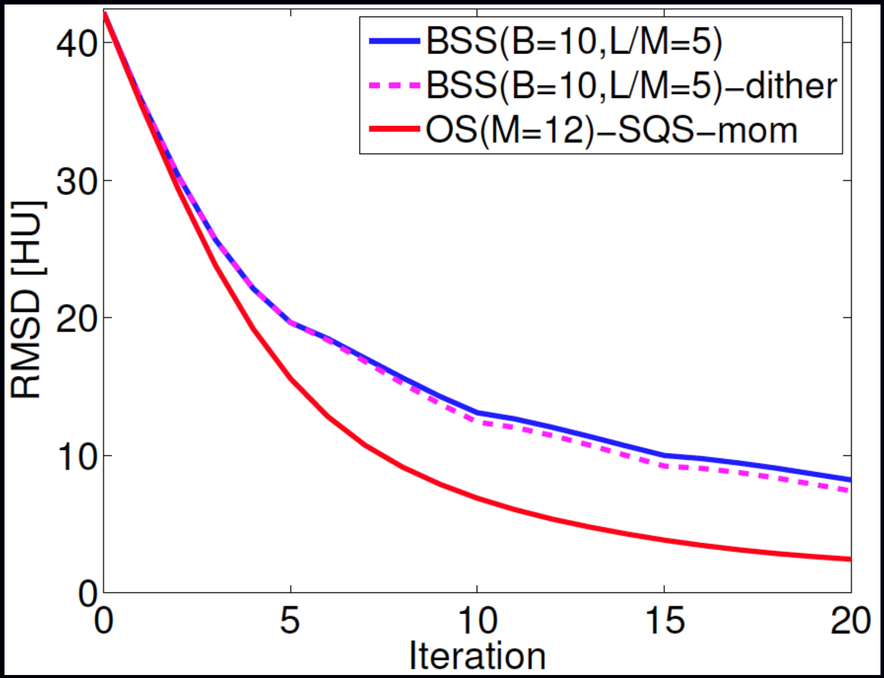
FBP initializer $\boldsymbol{x}^{(0)}$

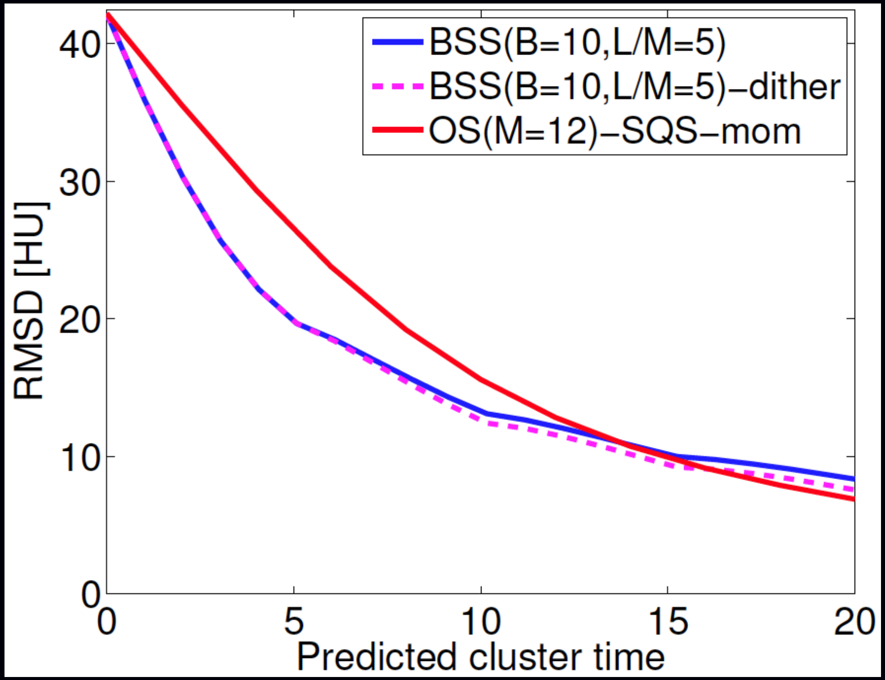


Converged $\boldsymbol{x}^{(\infty)}$



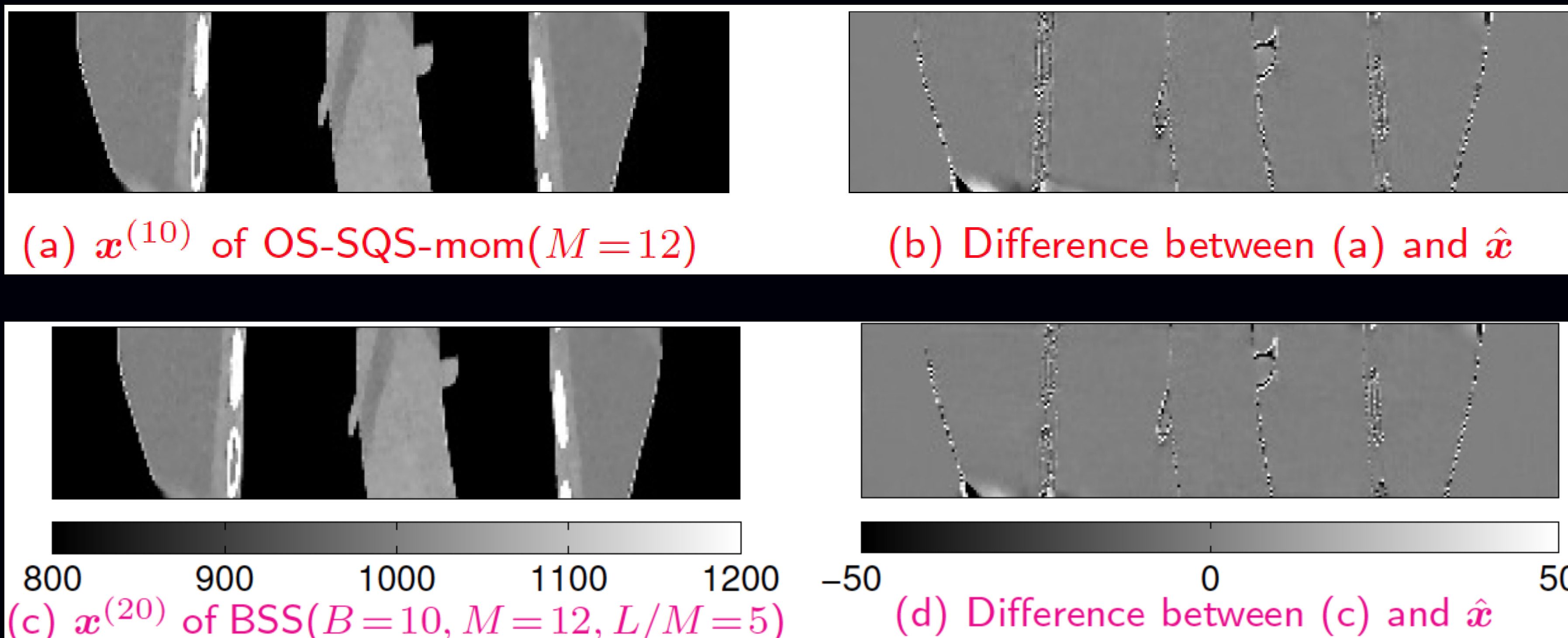
BSS OS+OGM: rates





- Outer loop interrupts momentum $\Longrightarrow$ BSS is slower per iteration than OS+OGM
- Reduced communication reduces overall time

BSS OS+OGM: images



(a) $\boldsymbol{x}^{(10)}$ of OS-SQS-mom($M=12$)

(b) Difference between (a) and $\hat{\boldsymbol{x}}$

(c) $\boldsymbol{x}^{(20)}$ of BSS($B=10, M=12, L/M=5$)

(d) Difference between (c) and $\hat{\boldsymbol{x}}$

- Comparable images
- Algorithm designed for distributed computation
- More results by Fully 3D conference in June...

Duality approach for using GPU

- Data transfer between system RAM and GPU can be bottleneck
- Want to "hide" communication time by overlapping with computation

Algorithm synopsis: Madison McGaffin and JF; Fully 3D, 2015 (to appear)

- Write cost function $\Psi(\boldsymbol{x})$ in terms of dual variables $\boldsymbol{v}$ and $\boldsymbol{u}$ for data-fit and regularizer:

$$\Psi(\boldsymbol{x}) = \sum_{i=1}^{M} \mathsf{h}_i([\boldsymbol{A}\boldsymbol{x}]_i) + \sum_k \psi([\boldsymbol{C}\boldsymbol{x}]_k)$$

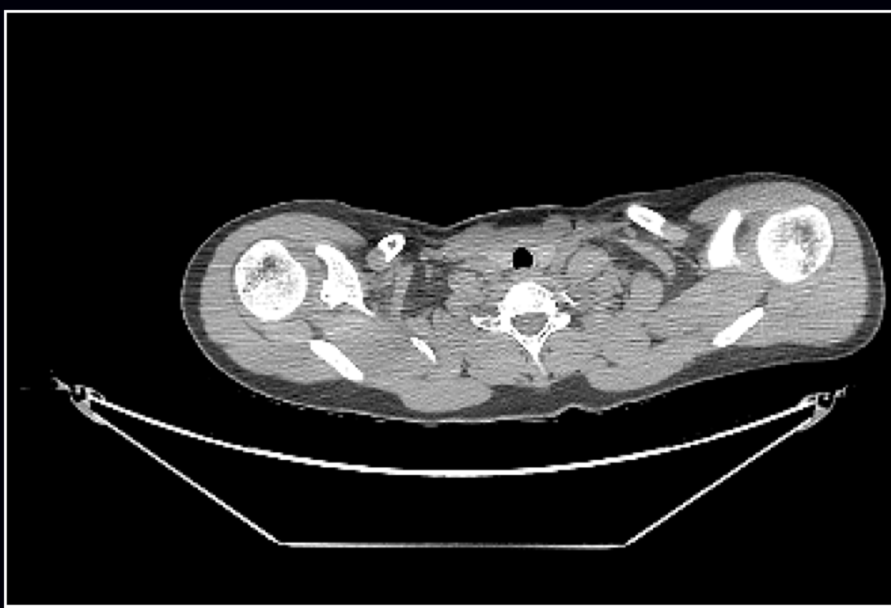
$$\boldsymbol{x}^{(n+1)} = \arg\min_{\boldsymbol{x}} \sup_{\boldsymbol{u},\boldsymbol{v}} \left(\boldsymbol{A}' v + \boldsymbol{C}' \boldsymbol{v}\right)' \boldsymbol{x} - \sum_{i=1}^{M} \mathsf{h}_i^*(u_i) - \sum_k \psi^*(v_k) + \frac{\mu}{2} \left\|\boldsymbol{x} - \boldsymbol{x}^{(n)}\right\|_2^2$$

  h_i^* and ψ^* denote convex conjugates of h_i and ψ
- Alternate between updating several projection view dual variables $\{u_i\}$ and dual variables for one regularization direction $\{v_k\}$
- Using dual variables "decouples" regularizer and data terms
- More details at Fully 3D ...

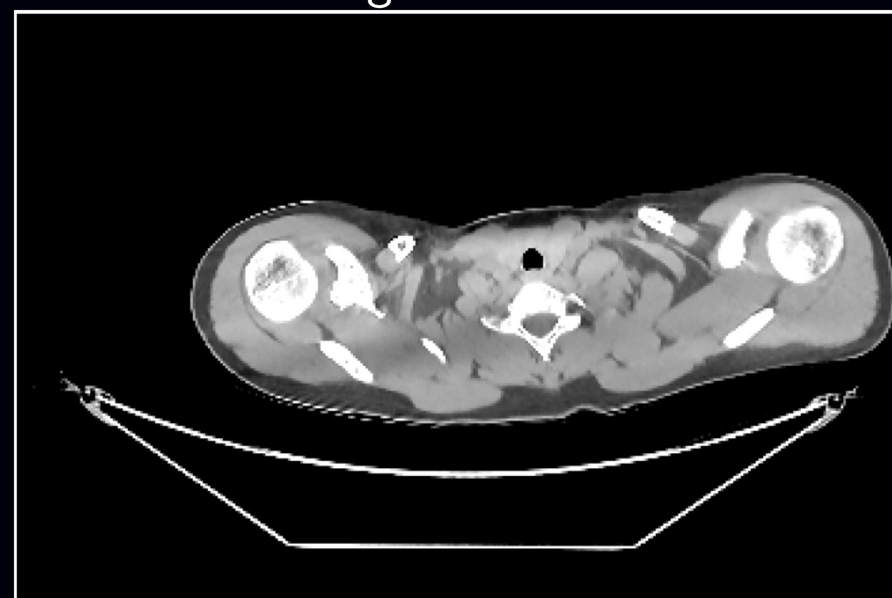
Duality-GPU: data

- 3D cone-beam helical X-ray CT scan
- pitch 0.5
- image $\boldsymbol{x}$: $512 \times 512 \times 109$ with 70 cm FOV and 0.625 mm slices
- sinogram : $\boldsymbol{y}$ 888 detectors $\times$ 32 rows $\times$ 7146 views
- OpenCL on aging NVIDIA GTX 480 GPU with 2.5 GB of memory

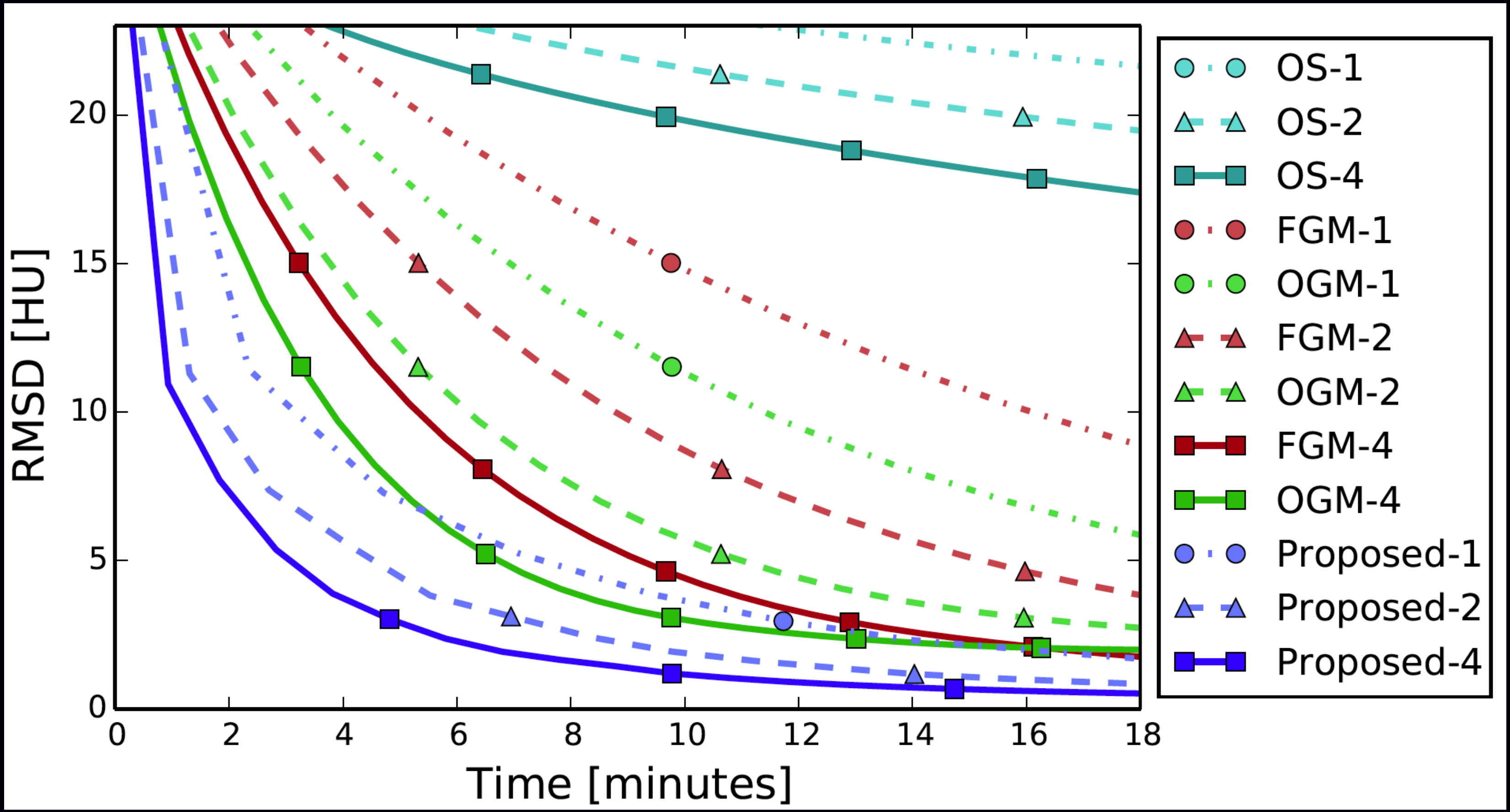
FBP initializer $\boldsymbol{x}^{(0)}$



Converged $\boldsymbol{x}^{(\infty)}$

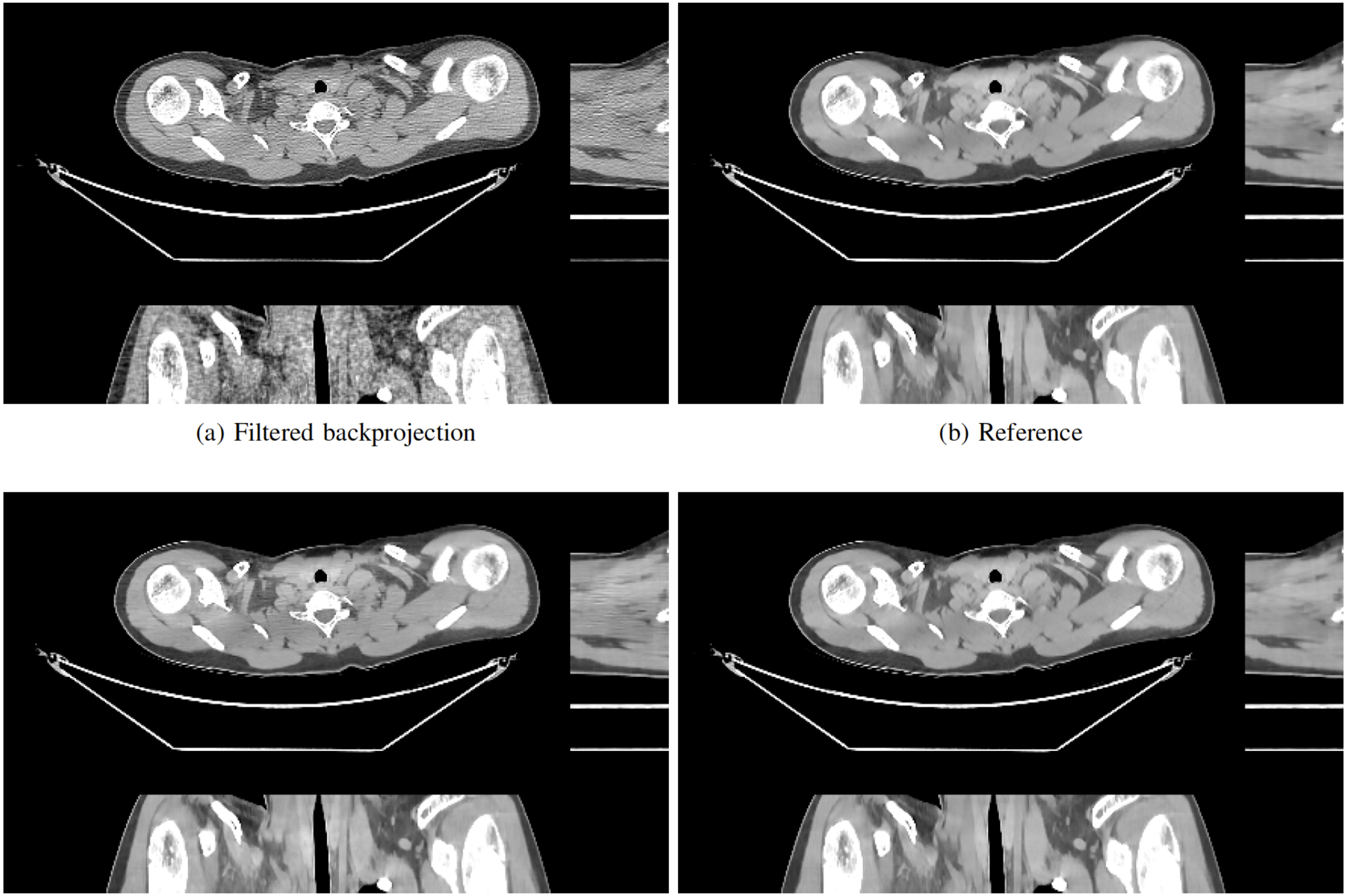


Duality-GPU: timing results



- Algorithm designed specifically for GPU architecture characteristics
- Future work:
  - combine with BSS for multiple nodes ?

Duality-GPU: image results



(a) Filtered backprojection

(b) Reference

(c) OS-OGM with 4 GPUs after 8 iterations (5.2 minutes)

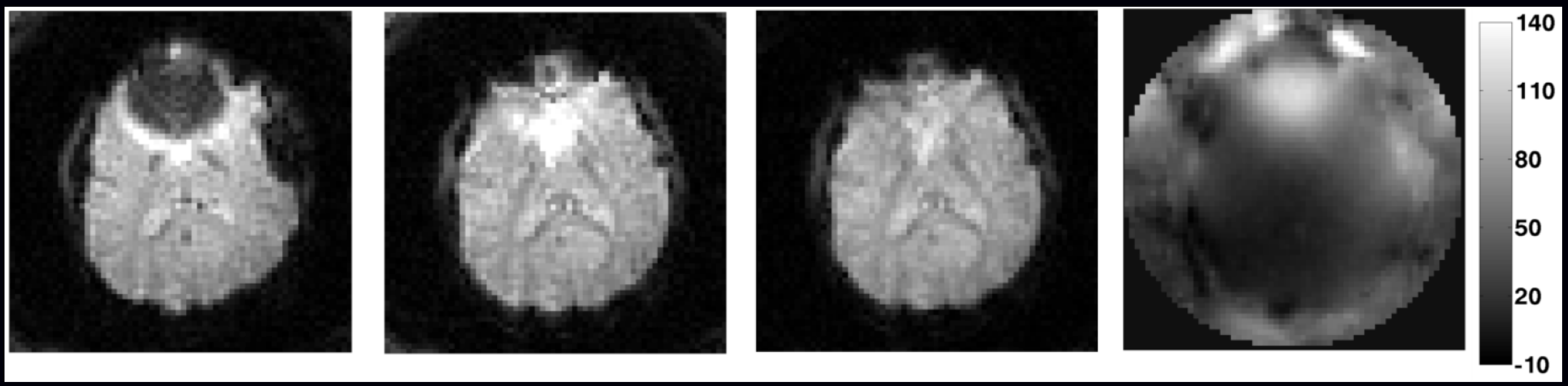
(d) Proposed with 4 GPUs after 5 iterations (4.8 minutes)

MRI image reconstruction

MRI: Why iterative reconstruction?

- Better physics modeling (*e.g.*, field inhomogeneity) $\Longrightarrow$ reduced artifacts

  Example: T2*-weighted imaging Sutton *et al.*, IEEE T-MI, 2003

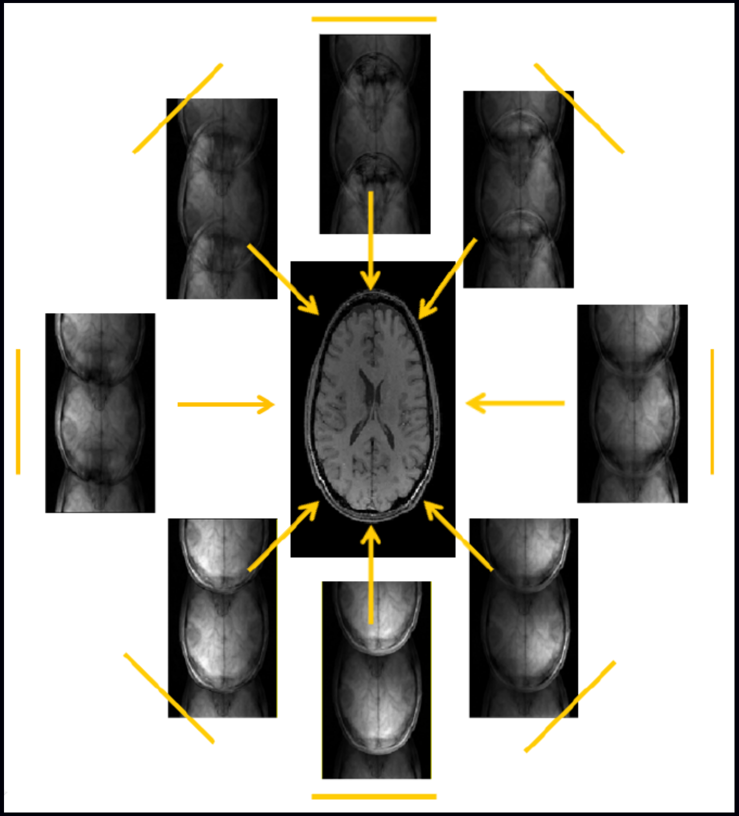


uncorrected traditional model-based field map

- Reducing scan time ("under-sampling")
  - Multiple receive coils
  - Object model assumptions (*e.g.*, sparsity)

Parallel MRI

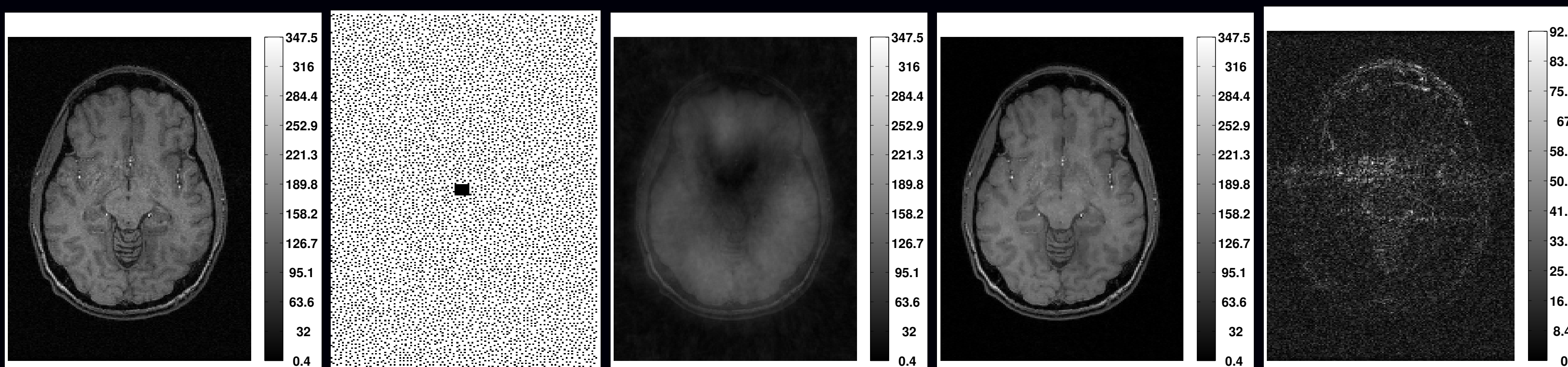
Undersampled Cartesian k-space, multiple receive coils, ...
(Pruessmann *et al.*, MRM, Nov. 1999)



Compressed sensing parallel MRI $\equiv$ further (random) under-sampling
Lustig *et al.*, IEEE Sig. Proc. Mag., Mar. 2008

2.5D parallel MR image reconstruction

Example of "compressed sensing" MRI reconstruction:



original k-space IFFT iterative difference

- Fully sampled body coil image of human brain
- Poisson-disk-based k-space sampling, 16% sampling (acceleration 6.25)
- Square-root of sum-of-squares inverse FFT of zero-filled k-space data for 8 coils (144×128)
- Regularized reconstruction $\boldsymbol{x}^{(\infty)}$
  combined TV and ℓ_1 norm of two-level undecimated Haar wavelets
- Difference image magnitude

(Sathish Ramani & JF, IEEE T-MI, Mar. 2011)

Model-based image reconstruction in parallel MRI

Regularized estimator:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \underbrace{\frac{1}{2} \|\boldsymbol{y} - \boldsymbol{F}\boldsymbol{S}\boldsymbol{x}\|_2^2}_{\text{data fit}} + \beta \underbrace{\|\boldsymbol{R}\boldsymbol{x}\|_p}_{\text{sparsity}}.$$

$\boldsymbol{F}$ is under-sampled DFT matrix (fat)

Features:

- coil sensitivity matrix $\boldsymbol{S}$ is block diagonal (Pruessmann *et al.*, MRM, Nov. 1999)
- $\boldsymbol{F}'\boldsymbol{F}$ is circulant (for Cartesian sampling)

Complications:

- Data-fit Hessian $\boldsymbol{S}'\boldsymbol{F}'\boldsymbol{F}\boldsymbol{S}$ is highly shift variant due to coil sensitivity maps
- Non-quadratic (edge-preserving) regularization $\|\cdot\|_p$
- Non-smooth regularization $\|\cdot\|_1$
- Complex quantities
- Large problem size (if 3D or dynamic or many coils)

ISTA methods for parallel MRI

"Traditional" iterative soft thresholding algorithm (ISTA) for sparsity regularized problems uses (global) Lipschitz constant of data-fit term:

$$\nabla^2 \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{F}\boldsymbol{S}\|_2^2 = \boldsymbol{S}'\boldsymbol{F}'\boldsymbol{F}\boldsymbol{S} \leq \boldsymbol{S}'\boldsymbol{S} \leq \lambda_{\max}\boldsymbol{I}, \quad \lambda_{\max} = \max_j \left[\boldsymbol{S}'\boldsymbol{S}\right]_{j,j}$$

$\lambda_{\max}$ is maximum sum-of-squares value of sensitivity maps; step size is $1/\lambda_{\max}$

Augmented Lagrangian (AL) methods converge faster than ISTA, FISTA, MFISTA

BARISTA: B1-based, adaptive restart, iterative soft thresholding algorithms
For synthesis operator $\boldsymbol{x} = \boldsymbol{Q}\boldsymbol{z}$ with $\boldsymbol{z}$ sparse:

$$\nabla^2 \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{F}\boldsymbol{S}\boldsymbol{Q}\|_2^2 = \boldsymbol{Q}'\boldsymbol{S}'\boldsymbol{F}'\boldsymbol{F}\boldsymbol{S}\boldsymbol{Q} \leq \boldsymbol{Q}'\boldsymbol{S}'\boldsymbol{S}\boldsymbol{Q} \leq \boldsymbol{D}$$

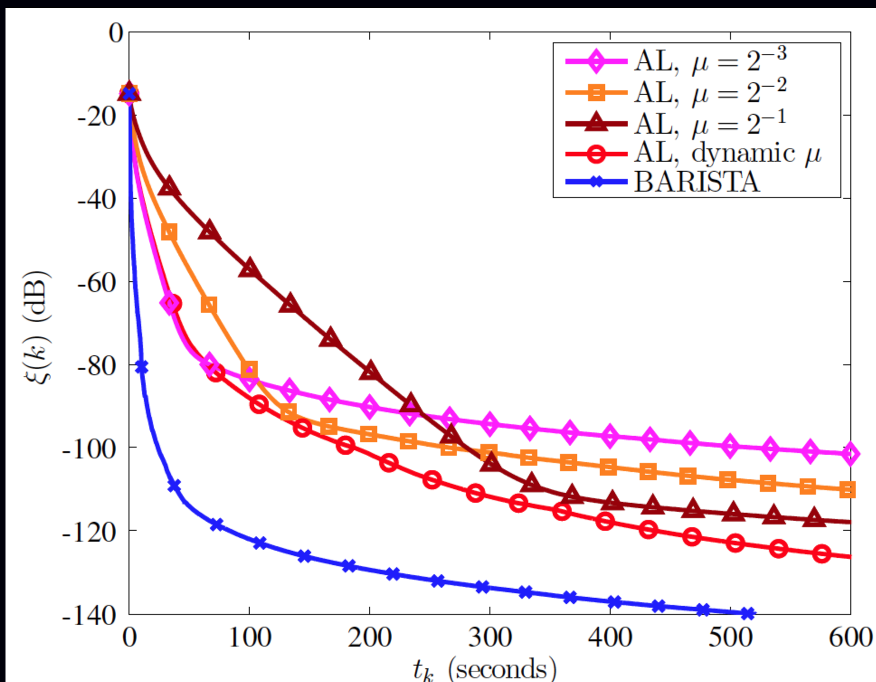
for a suitable diagonal matrix $\boldsymbol{D}$. (*cf.*, SQS) (Muckley *et al.*, IEEE T-MI, Feb. 2015)

$\boldsymbol{D}^{-1}$ becomes voxel-dependent step size, akin to that in CT
Include momentum and adaptive restart of O'Donoghue and Candès (2014).

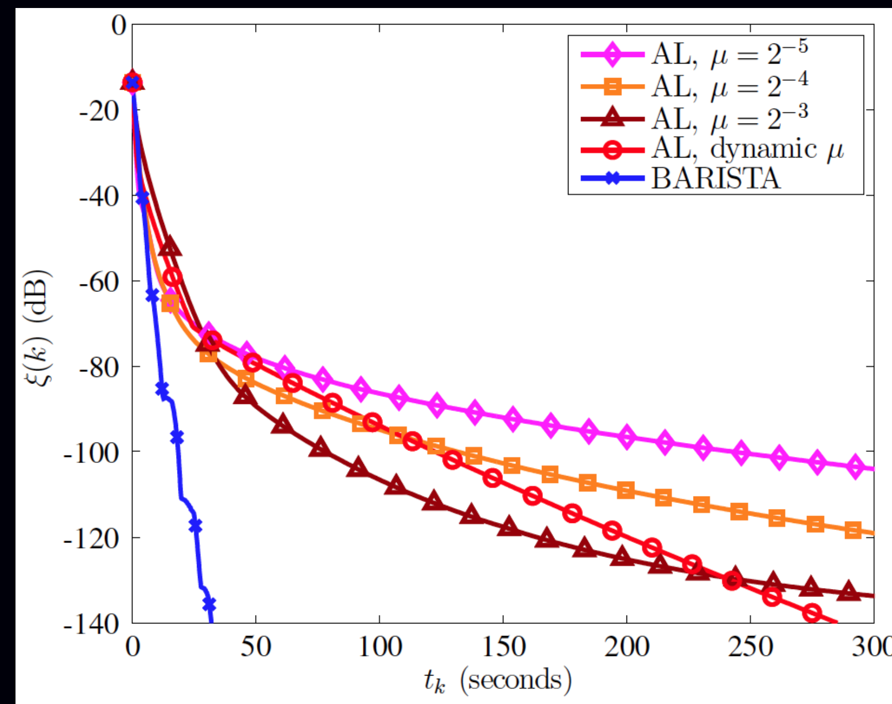
BARISTA convergence rates

Example of "compressed sensing" MRI reconstruction:

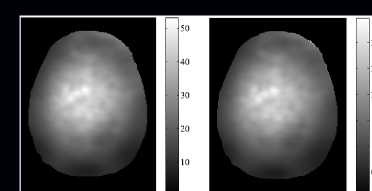
Total variation (TV) regularizer



Undecimated Haar Wavelets



Corresponding $\boldsymbol{D}$ for each of the two cases:



BARISTA requires no algorithm parameter tuning, unlike AL.

Summary

Model-based image reconstruction can

- improve image quality for low-dose X-ray CT
- enable faster MRI scans via under-sampling

Computation time remains a significant challenge
Moore's law will not solve the problem
Algorithms designed for distributed computation are essential

Bibliography

[1] D. Kim and J. A. Fessler. Optimized first-order methods for smooth convex minimization. *Mathematical Programming*, 2015. Submitted.

[2] D. Kim and J. A. Fessler. Optimized first-order methods for smooth convex minimization, 2014. arxiv 1406.5468.

[3] D. Kim and J. A. Fessler. An optimized first-order method for image restoration. In *Proc. IEEE Intl. Conf. on Image Processing*, 2015. Submitted.

[4] D. Kim, S. Ramani, and J. A. Fessler. Combining ordered subsets and momentum for accelerated X-ray CT image reconstruction. *IEEE Trans. Med. Imag.*, 34(1):167–78, January 2015.

[5] D. Kim and J. A. Fessler. Distributed block-separable ordered subsets for helical X-ray CT image reconstruction. In *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, 2015. To appear.

[6] M. G. McGaffin and J. A. Fessler. Fast GPU-driven model-based X-ray CT image reconstruction via alternating dual updates. In *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, 2015. To appear.

[7] M. J. Muckley, D. C. Noll, and J. A. Fessler. Fast parallel MR image reconstruction via B1-based, adaptive restart, iterative soft thresholding algorithms (BARISTA). *IEEE Trans. Med. Imag.*, 34(2):578–88, February 2015.

[8] D. E. Kuhl and R. Q. Edwards. Image separation radioisotope scanning. *Radiology*, 80(4):653–62, April 1963.

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

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

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

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

[13] A. J. Rockmore and A. Macovski. A maximum likelihood approach to emission image reconstruction from projections. *IEEE Trans. Nuc. Sci.*, 23(4):1428–32, August 1976.

[14] L. A. Shepp and Y. Vardi. Maximum likelihood reconstruction for emission tomography. *IEEE Trans. Med. Imag.*, 1(2):113–22, October 1982.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.
[21] R. Gordon and G. T. Herman. Reconstruction of pictures from their projections. *Comm. ACM*, 14(12):759–68, December 1971.
[22] 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.
[23] R. Gordon. A tutorial on ART (algebraic reconstruction techniques). *IEEE Trans. Nuc. Sci.*, 21(3):78–93, June 1974.
[24] R. L. Kashyap and M. C. Mittal. Picture reconstruction from projections. *IEEE Trans. Comp.*, 24(9):915–23, September 1975.
[25] 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.
[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] K. Sauer and C. Bouman. A local update strategy for iterative reconstruction from projections. *IEEE Trans. Sig. Proc.*, 41(2):534–48, February 1993.
[28] 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.
[29] 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.
[30] H. Erdoğan and J. A. Fessler. Ordered subsets algorithms for transmission tomography. *Phys. Med. Biol.*, 44(11):2835–51, November 1999.
[31] E. Hansis, J. Bredno, D. Sowards-Emmerd, and L. Shao. Iterative reconstruction for circular cone-beam CT with an offset flat-panel detector. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, pages 2228–31, 2010.
[32] Y. Drori and M. Teboulle. Performance of first-order methods for smooth convex minimization: A novel approach. *Mathematical Programming*, 145(1-2):451–82, June 2014.
[33] Y. Nesterov. A method for unconstrained convex minimization problem with the rate of convergence $O(1/k^2)$. *Dokl. Akad. Nauk. USSR*, 269(3):543–7, 1983.
[34] Y. Nesterov. Smooth minimization of non-smooth functions. *Mathematical Programming*, 103(1):127–52, May 2005.
[35] A. B. Taylor, J. M. Hendrickx, and François Glineur. Smooth strongly convex interpolation and exact worst-case performance of first-order methods, 2015. arxiv 1502.05666.
[36] D. Kim and J. A. Fessler. Optimized momentum steps for accelerating X-ray CT ordered subsets image reconstruction. In *Proc. 3rd Intl. Mtg. on image formation in X-ray CT*, pages 103–6, 2014.
[37] J. M. Rosen, J. Wu, T. F. Wenisch, and J. A. Fessler. Iterative helical CT reconstruction in the cloud for ten dollars in five minutes. In *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, pages 241–4, 2013.
[38] W. P. Segars, M. Mahesh, T. J. Beck, E. C. Frey, and B. M. W. Tsui. Realistic CT simulation using the 4D XCAT phantom. *Med. Phys.*, 35(8):3800–8, August 2008.
[39] 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.
[40] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger. SENSE: sensitivity encoding for fast MRI. *Mag. Res. Med.*,

42(5):952–62, November 1999.

[41] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly. Compressed sensing MRI. *IEEE Sig. Proc. Mag.*, 25(2):72–82, March 2008.

[42] S. Ramani and J. A. Fessler. Parallel MR image reconstruction using augmented Lagrangian methods. *IEEE Trans. Med. Imag.*, 30(3):694–706, March 2011.

[43] B. O'Donoghue and E. Candès. Adaptive restart for accelerated gradient schemes. *Found. Computational Math.*, 2014.