Accelerating image reconstruction methods

Jeffrey A. Fessler

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

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

Shanghai United Imaging Healthcare Co. (UIH)

19 Sep. 2016

Disclosure

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

Acknowledgment:
many collaborators and many students and post-docs,
particularly Donghwan Kim and Madison McGaffin

Outline

- What
  - CT
  - MRI
- Why
  - Why CT iterative
  - Why MRI iterative
- How
  - Optimization transfer
  - Separable quadratic surrogates
  - Momentum
  - Ordered subsets
- Parallelization
- Summary / open problems

Outline

What
- CT
- MRI

Why
- Why CT iterative
- Why MRI iterative

How
- Optimization transfer
- Separable quadratic surrogates
- Momentum
- Ordered subsets

Parallelization

Summary / open problems

X-ray CT scans

CT image reconstruction problem:
Determine unknown attenuation map $\boldsymbol{x}$ given sinogram data $\boldsymbol{y}$ using system matrix $\boldsymbol{A}$.

Defer motion hereafter...

MRI scans

(No moving parts to animate)

MR image reconstruction problem:
Determine unknown magnetization image $\boldsymbol{x}$ given k-space data $\boldsymbol{y}$ using system matrix $\boldsymbol{A}$

Defer motion for now...

Inverse problems

Unknown object $\boldsymbol{x}$ $\rightarrow$ $\boxed{\text{Imaging system}}$ $\rightarrow$ Data $\boldsymbol{y}$ $\rightarrow$ $\boxed{\text{Recon}}$ $\rightarrow$ Image $\hat{\boldsymbol{x}}$

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

Non-iterative methods:

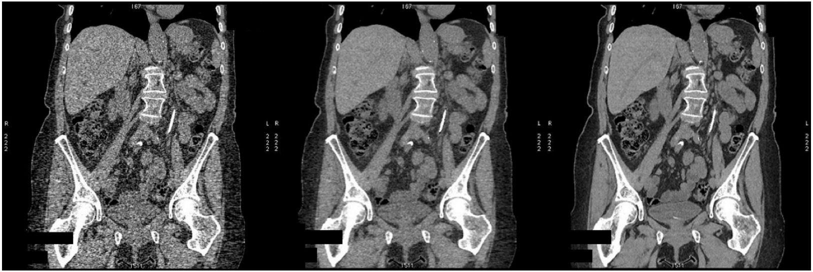
- analytical / direct
  - Filtered back-projection (FBP) for CT (textbook: Radon transform)
  - Inverse FFT for MRI (textbook: FFT)
- idealized description of the system ("textbook model")
  - geometry / sampling
  - disregards noise and simplifies physics
- typically fast

Iterative methods:

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

Statistical image reconstruction: CT example

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



Thin-slice FBP	ASIR (denoise)	Statistical
Seconds	A bit longer	Much longer

(Same sinogram, so all at same dose)

Outline

What
- CT
- MRI

Why
- Why CT iterative
- Why MRI iterative

How
- Optimization transfer
- Separable quadratic surrogates
- Momentum
- Ordered subsets

Parallelization

Summary / open problems

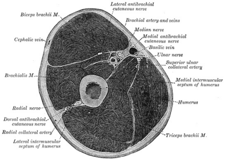
Why statistical/iterative methods for CT?

- 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

Why iterative for CT ... continued

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

Constraints may help reduce image artifacts or noise or dose.

Similar motivations/benefits in PET and SPECT.

Disadvantages of iterative methods for CT?

- Computation time
- 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 IQ

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

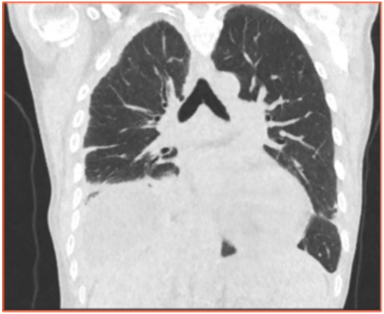
Statistical

MBIR example: Chest CT

Helical chest CT study with dose = 0.09 mSv.
Typical CXR effective dose is about 0.06 mSv.

(Health Physics Soc.: `http://www.hps.org/publicinformation/ate/q2372.html`)

FBP



MBIR

Veo (MBIR) images courtesy of Jiang Hsieh, GE Healthcare

History: Statistical reconstruction for X-ray CT*

- Iterative method for X-ray CT (Hounsfield, 1968)
- ART (Kaczmarz) 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)
  (Erdoνgan & Fessler, PMB, 1999)
- ...
- Commercial OS for Philips BrightView SPECT-CT (2010)
- Commercial ICD for GE CT scanners (Veo) (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)

Optimization problem formulation: $\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x} \geq \boldsymbol{0}} \Psi(\boldsymbol{x})$

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

$\boldsymbol{y}$: measured data (sinogram)
$\boldsymbol{A}$: system matrix (physics / geometry)
$\boldsymbol{W}$: weighting matrix (statistics)
$\boldsymbol{x}$: unknown image (attenuation map)
β : regularization parameter(s)
$\mathcal{N}_j$: neighborhood of jth voxel
ψ : edge-preserving potential function
(piece-wise smoothness / gradient sparsity)

Statistical image reconstruction for CT: Research

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x} \geq \boldsymbol{0}} \Psi(\boldsymbol{x}), \quad \Psi(\boldsymbol{x}) \triangleq \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_{\boldsymbol{W}}^2 + \sum_j \sum_k \beta_{j,k}\, \psi(x_j - x_k)$$

Apparent topics:

- regularization design / parameter selection ψ, β_{jk}
- statistical modeling $\boldsymbol{W}$, $\|\cdot\|$
- system modeling $\boldsymbol{A}$
- optimization algorithms (arg min)
- assessing IQ of $\hat{\boldsymbol{x}}$

Other topics:

- system design
- motion
- spectral
- dose ...

MRI: Why iterative reconstruction?

Inverse FFT is fast (like FBP). Why change?

(Joint work with D. Noll, J. Nielsen, ...)

Recall rationale for CT/PET/SPECT:

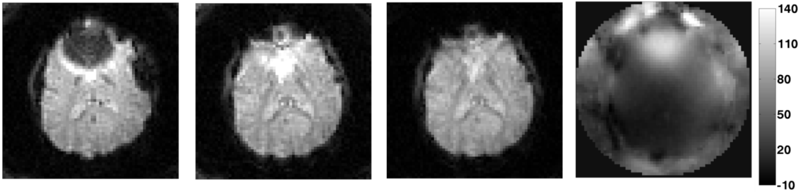
- physics modeling
  - reduce artifacts
  - improve resolution
  - improve contrast
- noise modeling: (dose, variability)
- sampling: non-standard geometries
- constraints on object

Which of these matter for MRI?

MRI why iterative: Physics

Physics modeling (*e.g.*, field inhomogeneity) $\Longrightarrow$ reduced artifacts

Example: T2*-weighted imaging (Sutton et al., IEEE T-MI, 03)



uncorrected traditional iterative field map

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \beta \, \mathsf{R}(\boldsymbol{x})$$

System matrix $\boldsymbol{A}$ depends on (measured) field map:

$$a_{ij} = \mathrm{e}^{-\imath \omega_j t_i} \, \mathrm{e}^{-\imath 2\pi \vec{\nu}_i \cdot \vec{r}_j}$$

No analytical inverse of $\boldsymbol{A}$. *cf.* nonuniform attenuation correction in SPECT

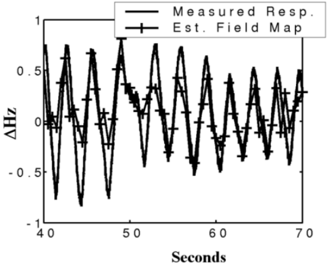
MRI why iterative: Physics

Joint estimation of field map $\boldsymbol{\omega}$ and magnetization image $\boldsymbol{x}$:

$$(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\omega}}) = \arg\min_{\boldsymbol{x},\boldsymbol{\omega}} \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}(\boldsymbol{\omega})\boldsymbol{x}\|_2^2 + \beta_1 \mathsf{R}_1(\boldsymbol{x}) + \beta_2 \mathsf{R}_2(\boldsymbol{\omega})$$

Useful when field map drifts in dynamic imaging.

(Sutton et al., MRM 04) (Olafsson et al., T-MI 08)



cf. joint estimation of attenuation map $\boldsymbol{\mu}$ and activity image $\boldsymbol{\lambda}$ in SPECT, PET and TOF-PET.

(Censor et al., T-NS 79) (Clinthorne et al., NSS 91) (Rezaei, Defrise, Nuyts, T-MI 14)

RF pulse design

RF pulse $\boldsymbol{b}$ $\rightarrow$ $\boxed{\text{Bloch Eqn}}$ $\rightarrow$ Excited magnetization $\boldsymbol{m}$

Small-tip approximation: $\boldsymbol{m} \approx \boldsymbol{A}\boldsymbol{b}$

Iterative RF pulse design (with RF power regularization):

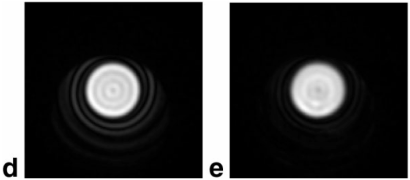
$$\arg\min_{\boldsymbol{b}} \|\boldsymbol{m} - \boldsymbol{A}\boldsymbol{b}\|_2^2 + \beta \|\boldsymbol{b}\|_2^2$$

Minimize using CG.

(Yip et al., MRM, Oct. 2005)

d. Non-iterative:

e. Iterative:

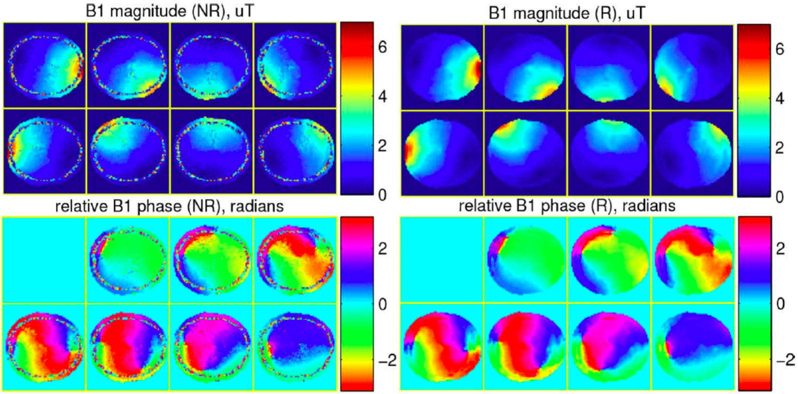


MRI why iterative: Noise

- MRI measurements: (complex) AWGN $\Longrightarrow$ easy !?

MRI why iterative: Noise

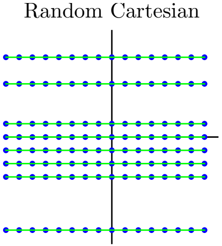
- MRI measurements: (complex) AWGN $\Longrightarrow$ easy !?
- Variance of image *phase* depends on image magnitude.
- Image phase useful in some applications, *e.g.*, B1 mapping:

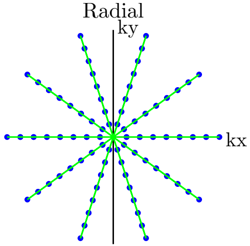


Unregularized vs regularized phase estimate. (Zhao et al., T-MI 14)

MRI why iterative: Sampling

- Reducing k-space sampling $\Longrightarrow$ reduced scan time
- Especially compelling for dynamic imaging (*cf.* CT and SPECT)
- Popular "under-sampled" patterns: (*cf.* sparse-view CT)



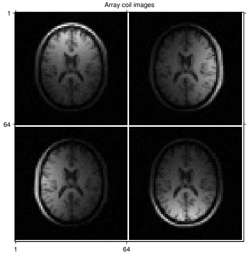


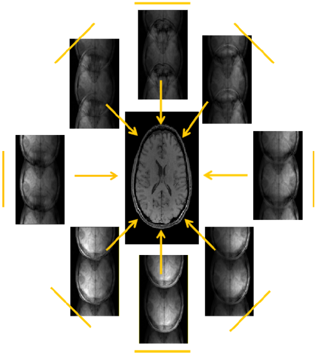
- Solution strategies
  - Multiple receive coils
  - Object model assumptions (*e.g.*, sparsity)
  - iterative reconstruction ("compressed sensing")

  Used clinically

Parallel MRI

Under-sampled Cartesian k-space: use multiple receive coils with individual spatial sensitivity patterns. (Pruessmann et al., MRM, 1999)





$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{F}\boldsymbol{S}_1 \\ \boldsymbol{F}\boldsymbol{S}_2 \\ \vdots \end{bmatrix}$$

Compressed sensing parallel MRI $\equiv$ (random) under-sampling

Lustig et al., IEEE Sig. Proc. Mag., Mar. 2008

cf. multiple-source CT (speed) or multi-camera SPECT (counts)

Model-based image reconstruction in parallel MRI

Regularized estimator:

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

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

Features:

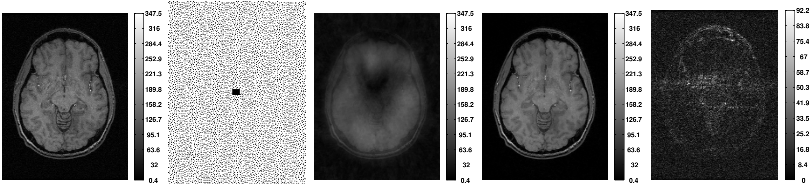
- coil sensitivity matrix $\boldsymbol{S}$ is block diagonal
- $\boldsymbol{F}'\boldsymbol{F}$ is circulant (for Cartesian sampling)

Challenges:

- 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 $\|\cdot\|_1$ (*cf.* sparse view CT)
- Complex quantities
- Large problem size (if 3D or dynamic or many coils)

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 (144 $\times$ 128)
- 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
- 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)

Summary of “What” and “Why”

- CT and MRI both involve inverse problems
- Some similarities in motivations and formulations
- Some similarities in computation challenges
- Some opportunities for cross-fertilization
- Caution: MRI reconstruction field is crowded!

Outline

What
- CT
- MRI

Why
- Why CT iterative
- Why MRI iterative

How
- Optimization transfer
- Separable quadratic surrogates
- Momentum
- Ordered subsets

Parallelization

Summary / open problems

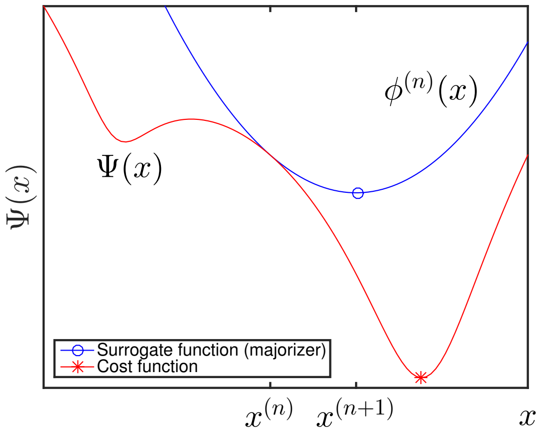
SIR for CT: Optimization challenges

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x} \geq \boldsymbol{0}}{\arg\min}\, \Psi(\boldsymbol{x}), \quad \Psi(\boldsymbol{x}) \triangleq \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_{\boldsymbol{W}}^2 + \sum_{j=1}^{N} \sum_{k} \beta_{j,k}\, \psi(x_j - x_k)$$

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)
- nonnegativity constraint
- data size grows: dual-source CT, spectral CT, wide-cone CT, ...
- Moore's law insufficient
  latest GPU clocks slower, but more threads

Optimization transfer (Majorize-Minimize) methods: 1D

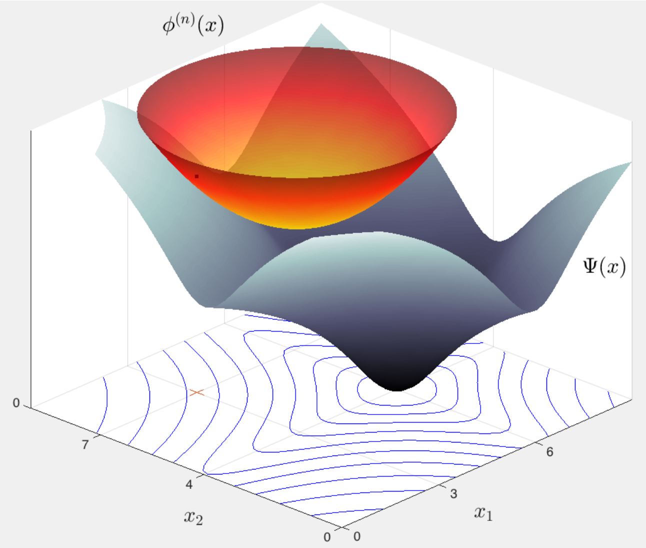


$$\phi^{(n)}(\boldsymbol{x}^{(n)}) = \Psi(\boldsymbol{x}^{(n)})$$
$$\phi^{(n)}(\boldsymbol{x}) \geq \Psi(\boldsymbol{x})$$

cf. ML-EM

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

Optimization transfer (Majorize-Minimize) methods: 2D



Outline

What
- CT
- MRI

Why
- Why CT iterative
- Why MRI iterative

How
- Optimization transfer
- Separable quadratic surrogates
- Momentum
- Ordered subsets

Parallelization

Summary / open problems

Separable Quadratic Surrogates (SQS): Math

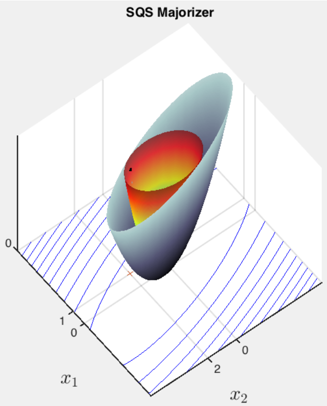
$$
\begin{aligned}
\mathsf{L}(\boldsymbol{x}) &= \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_{\boldsymbol{W}}^2 \\
&= \mathsf{L}(\boldsymbol{x}^{(n)}) + \nabla \mathsf{L}(\boldsymbol{x}^{(n)})(\boldsymbol{x} - \boldsymbol{x}^{(n)}) + \frac{1}{2} \underbrace{(\boldsymbol{x} - \boldsymbol{x}^{(n)})' \boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} (\boldsymbol{x} - \boldsymbol{x}^{(n)})}_{\text{non-separable}} \\
&\leq \mathsf{L}(\boldsymbol{x}^{(n)}) + \nabla \mathsf{L}(\boldsymbol{x}^{(n)})(\boldsymbol{x} - \boldsymbol{x}^{(n)}) + \frac{1}{2} \underbrace{(\boldsymbol{x} - \boldsymbol{x}^{(n)})' \boldsymbol{D} (\boldsymbol{x} - \boldsymbol{x}^{(n)})}_{\text{separable}} \\
&\triangleq \phi_{\mathsf{L}}^{(n)}(\boldsymbol{x}), \qquad \text{a “SQS”},
\end{aligned}
$$

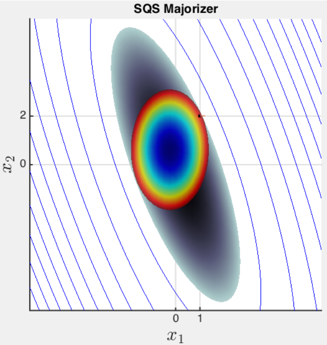
where $\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} \preceq \boldsymbol{D} = \mathrm{diag}\{\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}\mathbf{1}\}$. (De Pierro, T-MI, Mar. 1995)

Proofs:

- Convexity of x^2
- Geršgorin disk theorem ($\boldsymbol{D} - \boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}$ is diagonally dominant)
- Cauchy-Schwarz inequality

Separable Quadratic Surrogates (SQS): Pictures





- Find minimizer of $\mathsf{L}(\boldsymbol{x})$: challenging
- Find minimizer of $\phi_{\mathsf{L}}^{(n)}(\boldsymbol{x})$: easy (separate 1D problems)

WLS-SQS: Iteration

General optimization transfer (majorize-minimize) method:

$$\boldsymbol{x}^{(n+1)} = \arg\min_{\boldsymbol{x}} \phi_{\mathsf{L}}^{(n)}(\boldsymbol{x})$$

For SQS:

$$\begin{aligned}
\phi_{\mathsf{L}}^{(n)}(\boldsymbol{x}) &= \mathsf{L}(\boldsymbol{x}^{(n)}) + \nabla \mathsf{L}(\boldsymbol{x}^{(n)})(\boldsymbol{x} - \boldsymbol{x}^{(n)}) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^{(n)})' \boldsymbol{D} (\boldsymbol{x} - \boldsymbol{x}^{(n)}) \\
\nabla \phi_{\mathsf{L}}^{(n)}(\boldsymbol{x}) &= \nabla \mathsf{L}(\boldsymbol{x}^{(n)}) + \boldsymbol{D}(\boldsymbol{x} - \boldsymbol{x}^{(n)}) \\
\boldsymbol{0} = \nabla \phi_{\mathsf{L}}^{(n)}\left(\boldsymbol{x}^{(n+1)}\right) &= \nabla \mathsf{L}(\boldsymbol{x}^{(n)}) + \boldsymbol{D}\left(\boldsymbol{x}^{(n+1)} - \boldsymbol{x}^{(n)}\right)
\end{aligned}$$

$$\boxed{\boldsymbol{x}^{(n+1)} = \boldsymbol{x}^{(n)} - \boldsymbol{D}^{-1} \nabla \mathsf{L}(\boldsymbol{x}^{(n)})}$$

“diagonally preconditioned gradient descent”

(Erdoνgan & JF, PMB, 1999)

SQS versus GD: Math

Ordinary gradient descent (GD) for WLS:

$$\boldsymbol{x}^{(n+1)} = \boldsymbol{x}^{(n)} - \alpha \nabla \mathsf{L}(\boldsymbol{x}^{(n)}) = \boldsymbol{x}^{(n)} - \alpha \boldsymbol{A}' \boldsymbol{W} (\boldsymbol{A} \boldsymbol{x}^{(n)} - \boldsymbol{y}),$$

where textbook step size is reciprocal of Lipschitz constant:

$$\alpha = \frac{1}{\lambda_{\max}(\boldsymbol{A}' \boldsymbol{W} \boldsymbol{A})}.$$

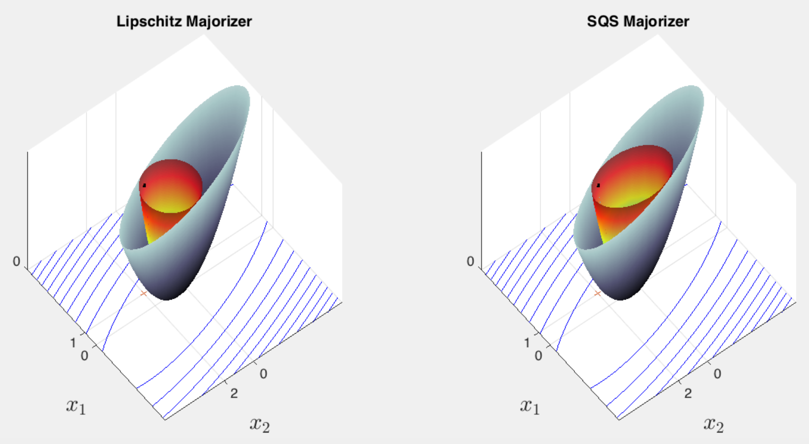
WLS-GD is equivalent to WLS-SQS with "isotropic" majorizer Hessian:

$$\boldsymbol{D} = \lambda_{\max}(\boldsymbol{A}' \boldsymbol{W} \boldsymbol{A}) \boldsymbol{I}.$$

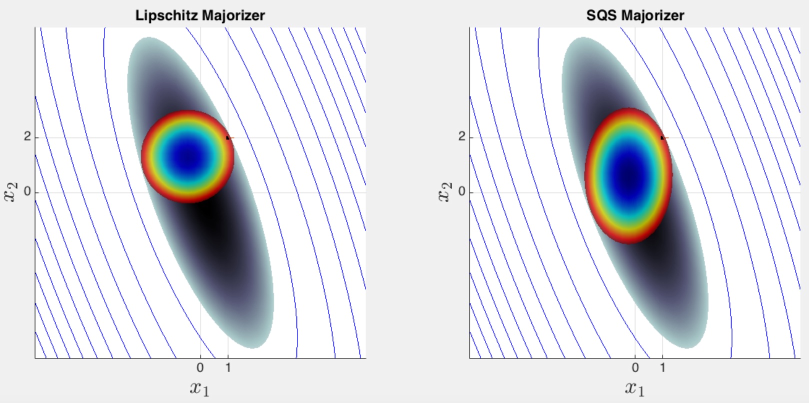
Drawbacks:

- $\lambda_{\max}(\boldsymbol{A}' \boldsymbol{W} \boldsymbol{A})$ usually impractical to compute (in CT) (power iteration?)
- GD usually converges slower than SQS due to smaller step sizes

SQS versus GD: Pictures



SQS versus GD: Pictures



Outline

What
- CT
- MRI

Why
- Why CT iterative
- Why MRI iterative

How
- Optimization transfer
- Separable quadratic surrogates
- Momentum
- Ordered subsets

Parallelization

Summary / open problems

Classical gradient descent (GD)

Assumptions:

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

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:

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

For a Huber-like function Ψ_0, GD achieves that (tight) bound.
$O(1/n)$ 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 $t_n = 1, \forall n$.
- Comparable computation as GD
- Drawbacks?
  - Store one additional image-sized vector $\boldsymbol{z}^{(n)}$
  - Ψ need not decrease monotonically

FGM1 properties

FGM1 shown by Nesterov to be $O(1/n^2)$ for "primary" sequence:

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

Nesterov constructed a function Ψ_1 such that any first-order method converges no faster than

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

Thus $O(1/n^2)$ rate of FGM1 is optimal.
Donghwan Kim (2014) analyzed "secondary" sequence:

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

SQS plus momentum for parallel MRI

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

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

  $\lambda_{\max}$ is maximum sum-of-squares value of sensitivity maps.
- Augmented Lagrangian (AL) methods converge faster than ISTA, FISTA, MFISTA (Ramani & JF, T-MI, 2011)
- BARISTA (B1-based, adaptive restart, ISTA) (Muckley, Noll, JF, T-MI, 2015)

  For synthesis operator $\boldsymbol{x} = \boldsymbol{Q}\boldsymbol{z}$ with $\boldsymbol{z}$ sparse:

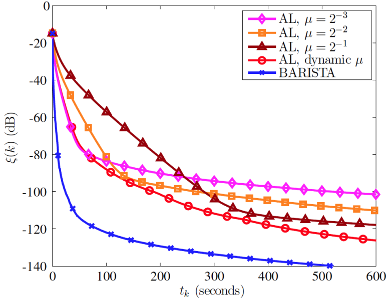
$$\nabla^2 \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{F}\boldsymbol{S}\boldsymbol{Q}\|_2^2 = \boldsymbol{Q}'\boldsymbol{S}'\boldsymbol{F}'\boldsymbol{F}\boldsymbol{S}\boldsymbol{Q} \preceq \boldsymbol{Q}'\boldsymbol{S}'\boldsymbol{S}\boldsymbol{Q} \preceq \boldsymbol{D}$$

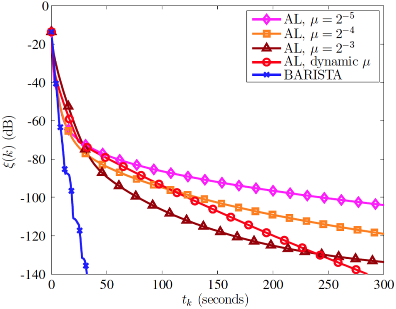
  for a suitable diagonal matrix $\boldsymbol{D}$. (*cf.*, SQS)
- $\boldsymbol{D}^{-1}$ becomes voxel-dependent step size, akin to SQS in CT

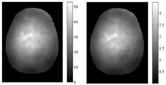
BARISTA convergence rates

"Compressed sensing" MRI reconstruction:

Total variation (TV) regularizer | Undecimated Haar Wavelets







Corresponding ***D*** for each of the two cases:

BARISTA requires no algorithm parameter tuning, unlike AL.

Includes momentum with adaptive restart of O'Donoghue and Candès (2015).

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

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & \\ 0 & 1.25 & 0 & 0 & 0 & 0 & \\ 0 & 0.10 & 1.40 & 0 & 0 & 0 & \\ 0 & 0.05 & 0.20 & 1.50 & 0 & 0 & \\ 0 & 0.03 & 0.11 & 0.29 & 1.57 & 0 & \\ \vdots & & & & & & \ddots \end{bmatrix}$$

Use of previous gradients $\Longrightarrow$ "momentum"
Is this the optimal choice for $\{h_{n,k}\}$?
Can we improve on the constant 2 in worst-case convergence rate?
Drori & Teboulle (2014) numerically found $2\times$ better $\{h_{n,k}\}$

Optimized gradient method (OGM1)

New approach by optimizing step-sizes $\{h_{n,k}\}$ analytically

Initialize: $t_0 = 1$, $\boldsymbol{z}^{(0)} = \boldsymbol{x}^{(0)}$ (Donghwan Kim and JF; 2014, 2015)

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

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

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

Smaller (worst-case) convergence bound than Nesterov by $2\times$:

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

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

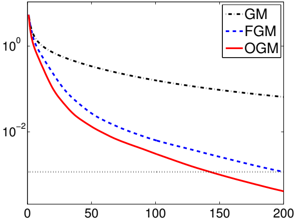
Example: Image restoration (!?)

True
$\boldsymbol{x}$

Blurry
$\boldsymbol{y}$

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

Rate



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

$$\arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + R(\boldsymbol{x})$$

Outline

What
- CT
- MRI

Why
- Why CT iterative
- Why MRI iterative

How
- Optimization transfer
- Separable quadratic surrogates
- Momentum
- Ordered subsets

Parallelization

Summary / open problems

Ordered subsets approximation

- Data decomposition (aka incremental gradients, *cf.* stochastic GD):

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

- Key idea. For $\boldsymbol{x}$ far from minimizer: $\nabla\Psi(\boldsymbol{x}) \approx M\,\nabla\Psi_m(\boldsymbol{x})$
- SQS:

$$\boldsymbol{x}^{(n+1)} = \boldsymbol{x}^{(n)} - \boldsymbol{D}^{-1}\,\nabla\Psi\left(\boldsymbol{x}^{(n)}\right)$$

- OS-SQS:

  for $n = 0, 1, \ldots$ (iteration)

  for $m = 1, \ldots, M$ (subset)

$$k = nM + m \text{ (subiteration)}$$

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^{k} - \boldsymbol{D}^{-1} M \underbrace{\nabla\Psi_m\left(\boldsymbol{x}^{k}\right)}_{\text{less work}}$$

- Applied coil-wise in parallel MRI (Muckley, Noll, JF, ISMRM 2014)

For more acceleration, combine OGM1 with ordered subsets (OS).

OS-OGM1:
Initialize: $t_0 = 1$, $\boldsymbol{z}^{(0)} = \boldsymbol{x}^{(0)}$
for $n = 0, 1, \ldots$ (iteration)
for $m = 1, \ldots, M$ (subset)

$$k = nM + m \text{ (subiteration)}$$

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

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

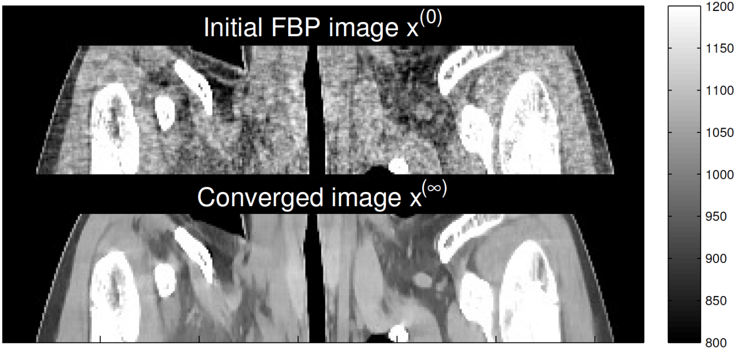
$$\boldsymbol{x}^{k+1} = \boldsymbol{z}^{k+1} + \frac{t_k - 1}{t_{k+1}}\left(\boldsymbol{z}^{k+1} - \boldsymbol{z}^k\right) + \frac{t_k}{t_{k+1}}\left(\boldsymbol{z}^{k+1} - \boldsymbol{x}^k\right)$$

OS-OGM1 properties

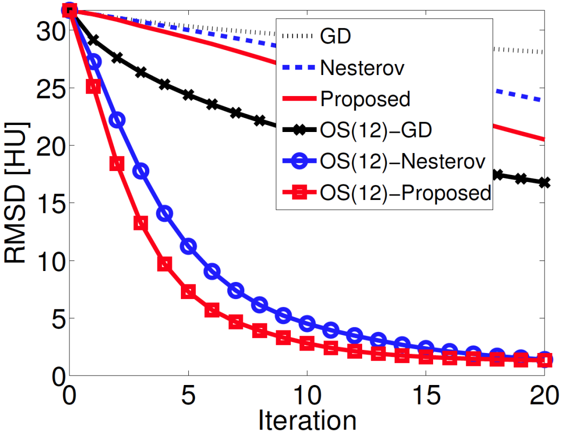
- Approximate convergence rate for Ψ: $O\left(\frac{1}{n^2 M^2}\right)$
  (Donghwan Kim and JF; CT 2014)
- Same compute per iteration as other OS methods
  (One forward / backward projection and M regularizer gradients per iteration)
- Same memory as OGM1 (two more images than OS-SQS)
- Guaranteed convergence for $M = 1$
- No convergence theory for $M > 1$
  - unstable for large M
  - small M preferable for parallelization
- Now fast enough to show X-ray CT examples...

OS-OGM1 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



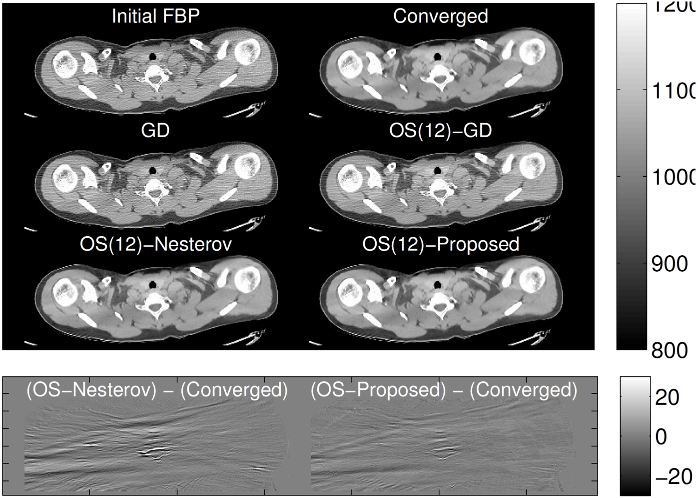
OS-OGM1 results: convergence rate



Root mean square difference (RMSD) between $\boldsymbol{x}^{(n)}$ and $\boldsymbol{x}^{(\infty)}$ over ROI (in HU), versus iteration. ("Proposed" = OGM1.)

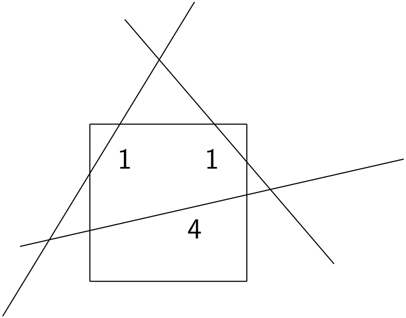
(Compute times per iteration are very similar.)

OS-OGM1 results: images



At iteration $n = 10$ with $M = 12$ subsets.

OS divergence example



- one-pixel image
- three intersecting rays
- $\boldsymbol{A} = \begin{bmatrix} 1 \\ 1 \\ 4 \end{bmatrix}$
- $\boldsymbol{x} = 2$, $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} = \begin{bmatrix} 2 \\ 2 \\ 8 \end{bmatrix}$
- condition number of $\boldsymbol{A}'\boldsymbol{A} = 1$
- consistent system of eqns.

OS divergence example

OS-SQS-LS for $M = 3$ subsets:

$$\boldsymbol{x}^{\text{new}} = \boldsymbol{x}^{\text{old}} - \boldsymbol{D}^{-1} 3\, \nabla \Psi_m\left(\boldsymbol{x}^{\text{old}}\right) = \boldsymbol{x}^{\text{old}} - \boldsymbol{D}^{-1} 3 \boldsymbol{A}_m' (\boldsymbol{A}_m \boldsymbol{x}^{\text{old}} - \boldsymbol{y}_m)$$

$\boldsymbol{D} = \operatorname{diag}\{\boldsymbol{A}'\boldsymbol{A}\mathbf{1}\} = 1^2 + 1^2 + 4^2 = 18$
After 3 updates:

$$\boldsymbol{x}^{(n+1)} - \boldsymbol{x} = \left(1 - \frac{3}{18} 1^2\right) \left(1 - \frac{3}{18} 1^2\right) \left(1 - \frac{3}{18} 4^2\right) (\boldsymbol{x}^{(n)} - \boldsymbol{x})$$
$$= -2(15/18)^3 (\boldsymbol{x}^{(n)} - \boldsymbol{x}) = -\frac{125}{108} (\boldsymbol{x}^{(n)} - \boldsymbol{x})$$

Divergence of OS-SQS-LS is possible even in well-conditioned, consistent case

Outline

What
- CT
- MRI

Why
- Why CT iterative
- Why MRI iterative

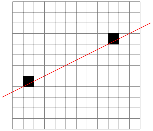
How
- Optimization transfer
- Separable quadratic surrogates
- Momentum
- Ordered subsets

Parallelization

Summary / open problems

Parallelization challenges in CT

- CT is not “embarrassingly parallel” (except across patients)
- In 2D, Hessian $\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}$ is not only dense, but completely full

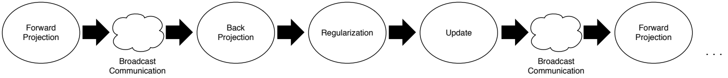


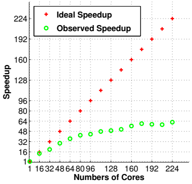
- In 3D, Hessian $\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}$ is dense and almost full

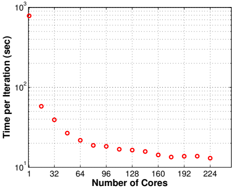
Amazon Cloud version of OS-OGM

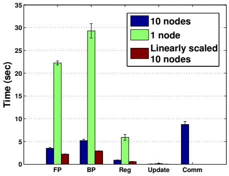
Distribute long object (320 useful slices) into (overlapping) slabs (128 slices each) across 5 separate clusters, each with 10 nodes having 16 cores.

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

Block-separable surrogates for distributed reconstruction

Conventional OS approach uses a voxel-wise SQS:

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

Diagonal matrix $\boldsymbol{D}$ majorizes the Hessian of Ψ: $\nabla^2 \Psi(\boldsymbol{x}) \preceq \boldsymbol{D}$.
Distributed computing alternative: slab-separable surrogate:

$$\Psi(\boldsymbol{x}) - \Psi\left(\boldsymbol{x}^{(n)}\right) \leq \sum_{b=1}^{B} \Psi_b(\boldsymbol{x}_b)$$

$$\Psi_b(\boldsymbol{x}_b) \triangleq \nabla_{\boldsymbol{x}_b} \Psi\left(\boldsymbol{x}^{(n)}\right)(\boldsymbol{x}_b - \boldsymbol{x}_b^{(n)}) + \frac{1}{2} \left(\boldsymbol{x}_b - \boldsymbol{x}_b^{(n)}\right)' \boldsymbol{H}_b \left(\boldsymbol{x}_b - \boldsymbol{x}_b^{(n)}\right)$$

Block diagonal matrix $\boldsymbol{H} = \operatorname{diag}\{\boldsymbol{H}_1, \ldots, \boldsymbol{H}_B\}$ majorizes $\nabla^2 \Psi(\boldsymbol{x})$.

BSS continued

$$\Psi_b(\boldsymbol{x}_b) \triangleq \nabla_{\boldsymbol{x}_b} \Psi(\boldsymbol{x}^{(n)})(\boldsymbol{x}_b - \boldsymbol{x}_b^{(n)}) + \frac{1}{2}\left(\boldsymbol{x}_b - \boldsymbol{x}_b^{(n)}\right)' \boldsymbol{H}_b \left(\boldsymbol{x}_b - \boldsymbol{x}_b^{(n)}\right)$$

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

Updates parallelizable across blocks (slabs):

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

- Reduces communication.
- (Apply favorite optimization method within slab.)
- (Donghwan Kim and JF; Fully 3D, 2015)

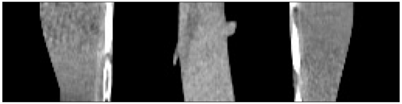
Block-separable surrogate (BSS) OS-OGM

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 $\phi_{\text{BSS}}(\boldsymbol{x}; \tilde{\boldsymbol{x}}^{(n)})$ 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) $\boldsymbol{g}_{m,b}^{(l)} = M \nabla_b \phi_{\text{BSS},m}(\boldsymbol{z}^{(\frac{l}{M})}; \boldsymbol{z}^{(0)})$ [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**

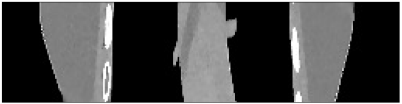
BSS OS-OGM: data

- $256 \times 256 \times 160$ 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

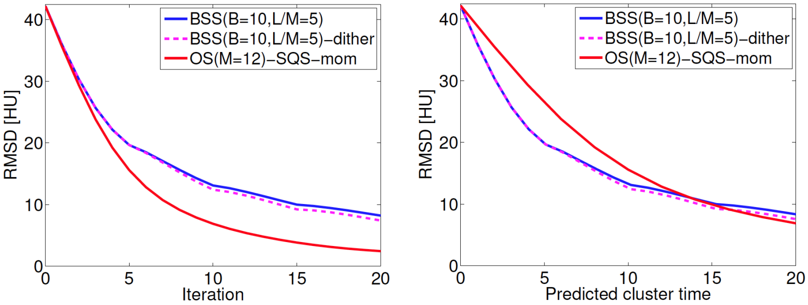
FBP initializer $\boldsymbol{x}^{(0)}$



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



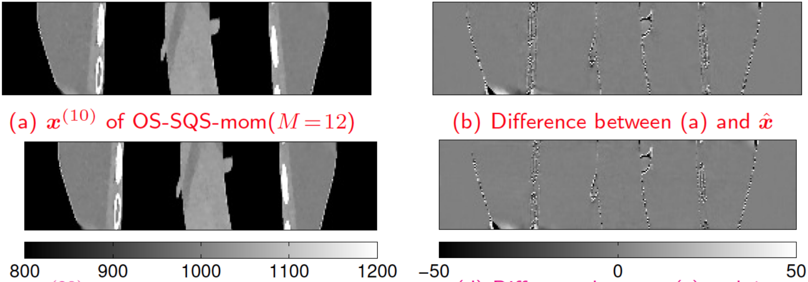
BSS OS-OGM: rates



- 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

Duality approach for using GPU

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

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

- Write cost function $\Psi(\boldsymbol{x})$ in terms of dual variables $\boldsymbol{v}$ and $\boldsymbol{u}$ for data-fit and regularizer: $\Psi(\boldsymbol{x}) = \sum_{i=1}^{M} \mathrm{h}_i([\boldsymbol{A}\boldsymbol{x}]_i) + \sum_k \psi([\boldsymbol{C}\boldsymbol{x}]_k)$

$$\boldsymbol{x}^{(n+1)} = \arg\min_{\boldsymbol{x}} \sup_{\boldsymbol{u},\boldsymbol{v}}$$

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

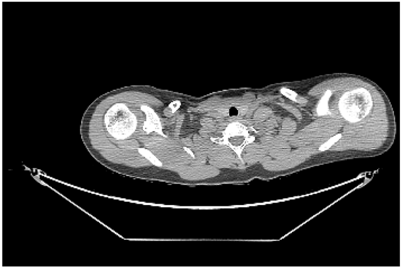
h_i^* and ψ^* denote convex conjugates of h_i and ψ

- Alternate between updating
  - several projection view dual variables $\{u_i\}$
  - dual variables for one regularization direction $\{v_k\}$
- Using dual variables "decouples" regularizer and data terms
- OS-like method with convergence theorem!

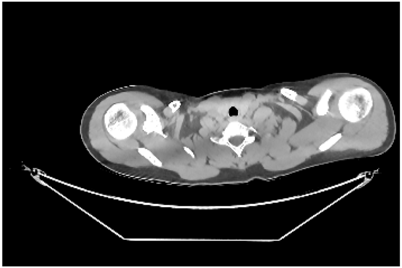
Duality-GPU: data

- 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 RAM

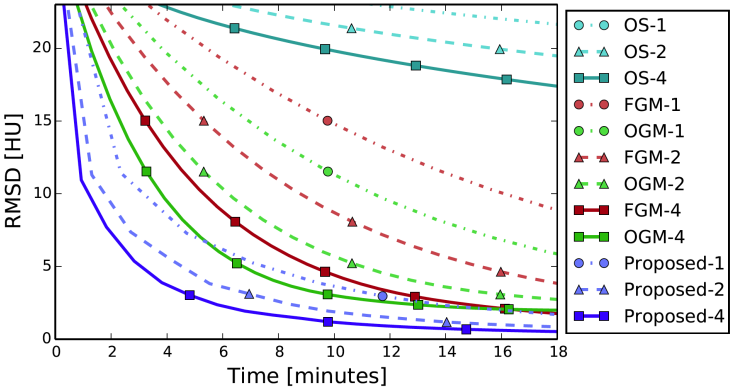
FBP initializer $\boldsymbol{x}^{(0)}$



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

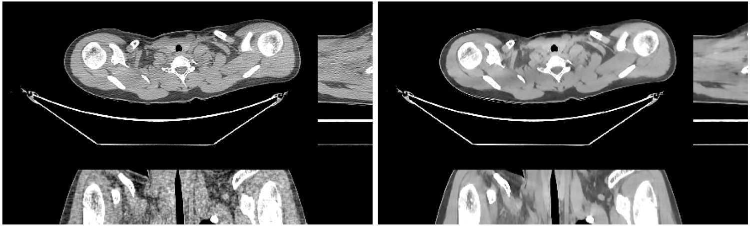


Duality-GPU: timing results

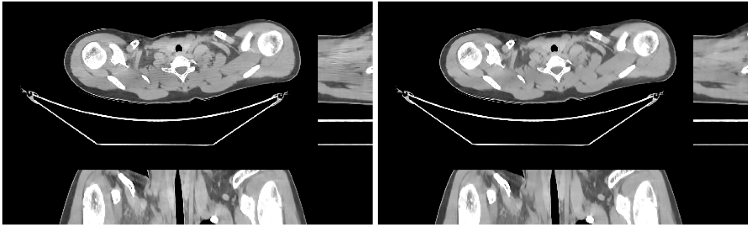


- Algorithm design for specific GPU architecture characteristics
- Future work: combine with BSS for multiple nodes ?

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)

Summary

- Model-based image reconstruction can
  - improve image quality for low-dose X-ray CT
  - enable faster MRI scans via under-sampling
- Much more: dynamic image reconstruction, motion compensation, ...
- Computation time remains a significant challenge
- Moore's law alone will not solve the computation problem
- Algorithms designed for distributed computation are essential
  - Block-separable surrogates to reduce communication
    (Donghwan Kim and JF; Fully 3D, 2015)
  - Duality approach to overlap communication with computation
    Also provides a OS-like algorithm with convergence theory
    (Madison McGaffin and JF; Fully 3D, 2015)

Bibliography I

[1] G. Hounsfield, *A method of apparatus for examination of a body by radiation such as x-ray or gamma radiation*, US Patent 1283915. British patent 1283915, London., 1972.

[2] S. Kaczmarz, "Angenaherte auflosung von systemen linearer gleichungen," *Bull. Acad. Polon. Sci. Lett. A*, vol. 35, 355–7, 1937, Approximate solution to systems of linear equations.

[3] 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.*, vol. 29, no. 3, 471–81, Dec. 1970.

[4] R. Gordon and G. T. Herman, "Reconstruction of pictures from their projections," *Comm. ACM*, vol. 14, no. 12, 759–68, Dec. 1971.

[5] 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.*, vol. 42, no. 1, 1–32, Nov. 1973.

[6] R. Gordon, "A tutorial on ART (algebraic reconstruction techniques)," *IEEE Trans. Nuc. Sci.*, vol. 21, no. 3, 78–93, Jun. 1974.

[7] R. L. Kashyap and M. C. Mittal, "Picture reconstruction from projections," *IEEE Trans. Comp.*, vol. 24, no. 9, 915–23, Sep. 1975.

[8] A. J. Rockmore and A. Macovski, "A maximum likelihood approach to transmission image reconstruction from projections," *IEEE Trans. Nuc. Sci.*, vol. 24, no. 3, 1929–35, Jun. 1977.

[9] K. Lange and R. Carson, "EM reconstruction algorithms for emission and transmission tomography," *J. Comp. Assisted Tomo.*, vol. 8, no. 2, 306–16, Apr. 1984.

[10] K. Sauer and C. Bouman, "A local update strategy for iterative reconstruction from projections," *IEEE Trans. Sig. Proc.*, vol. 41, no. 2, 534–48, Feb. 1993.

[11] 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.*, vol. 40, no. 7, 1225–41, Jul. 1995.

[12] C. Kamphuis and F. J. Beekman, "Accelerated iterative transmission CT reconstruction using an ordered subsets convex algorithm," *IEEE Trans. Med. Imag.*, vol. 17, no. 6, 1001–5, Dec. 1998.

[13] H. Erdoνgan and J. A. Fessler, "Ordered subsets algorithms for transmission tomography," *Phys. Med. Biol.*, vol. 44, no. 11, 2835–51, Nov. 1999.

[14] 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.*, 2010, 2228–31.

[15] 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.*, vol. 22, no. 2, 178–88, Feb. 2003.

[16] ——,"Dynamic field map estimation using a spiral-in / spiral-out acquisition," *Mag. Res. Med.*, vol. 51, no. 6, 1194–204, Jun. 2004.

[17] V. T. Olafsson, D. C. Noll, and J. A. Fessler, "Fast joint reconstruction of dynamic R_2^* and field maps in functional MRI," *IEEE Trans. Med. Imag.*, vol. 27, no. 9, 1177–88, Sep. 2008.

[18] Y. Censor, D. E. Gustafson, A. Lent, and H. Tuy, "A new approach to the emission computerized tomography problem: Simultaneous calculation of attenuation and activity coefficients," *IEEE Trans. Nuc. Sci.*, vol. 26, no. 2, 2775–9, Apr. 1979.

[19] N. H. Clinthorne, J. A. Fessler, G. D. Hutchins, and W. L. Rogers, "Joint maximum likelihood estimation of emission and attenuation densities in PET," in *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, vol. 3, 1991, 1927–32.

[20] H. Erdoνgan and J. A. Fessler, "Algorithms for joint estimation of attenuation and emission images in PET," in *Proc. IEEE Conf. Acoust. Speech Sig. Proc.*, vol. 6, 2000, 3783–6.

[21] A. Rezaei, M. Defrise, and J. Nuyts, “ML-reconstruction for TOF-PET with simultaneous estimation of the attenuation factors,” *IEEE Trans. Med. Imag.*, vol. 33, no. 7, 1563–72, Jul. 2014.

[22] C. Yip, J. A. Fessler, and D. C. Noll, “Iterative RF pulse design for multidimensional, small-tip-angle selective excitation,” *Mag. Res. Med.*, vol. 54, no. 4, 908–17, Oct. 2005.

[23] F. Zhao, J. A. Fessler, S. M. Wright, and D. C. Noll, “Regularized estimation of magnitude and phase of multi-coil B1 field via Bloch-Siegert B1 mapping and coil combination optimizations,” *IEEE Trans. Med. Imag.*, vol. 33, no. 10, 2020–30, Oct. 2014.

[24] S. S. Vasanawala, M. T. Alley, B. A. Hargreaves, R. A. Barth, J. M. Pauly, and M. Lustig, “Improved pediatric MR imaging with compressed sensing,” *Radiology*, vol. 256, 607–16, 2010.

[25] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, “SENSE: sensitivity encoding for fast MRI,” *Mag. Res. Med.*, vol. 42, no. 5, 952–62, Nov. 1999.

[26] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, “Compressed sensing MRI,” *IEEE Sig. Proc. Mag.*, vol. 25, no. 2, 72–82, Mar. 2008.

[27] S. Ramani and J. A. Fessler, “Parallel MR image reconstruction using augmented Lagrangian methods,” *IEEE Trans. Med. Imag.*, vol. 30, no. 3, 694–706, Mar. 2011.

[28] A. R. De Pierro, “A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography,” *IEEE Trans. Med. Imag.*, vol. 14, no. 1, 132–7, Mar. 1995.

[29] Y. Drori and M. Teboulle, “Performance of first-order methods for smooth convex minimization: A novel approach,” *Mathematical Programming*, vol. 145, no. 1-2, 451–82, Jun. 2014.

[30] Y. Nesterov, “A method for unconstrained convex minimization problem with the rate of convergence $O(1/k^2)$,” *Dokl. Akad. Nauk. USSR*, vol. 269, no. 3, 543–7, 1983.

[31] ——,“Smooth minimization of non-smooth functions,” *Mathematical Programming*, vol. 103, no. 1, 127–52, May 2005.

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

[33] ——,“Optimized first-order methods for smooth convex minimization,” *Mathematical Programming*, vol. 159, no. 1, 81–107, Sep. 2016.

[34] 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.*, vol. 34, no. 2, 578–88, Feb. 2015.

[35] B. O’Donoghue and E. Candès, “Adaptive restart for accelerated gradient schemes,” *Found. Comp. Math.*, vol. 15, no. 3, 715–32, Jun. 2015.

[36] D. Kim and J. A. Fessler, “An optimized first-order method for image restoration,” in *Proc. IEEE Intl. Conf. on Image Processing*, 2015, 3675–9.

[37] A. B. Taylor, J. M. Hendrickx, and François. Glineur, “Smooth strongly convex interpolation and exact worst-case performance of first- order methods,” *Mathematical Programming*, 2016.

[38] M. Muckley, D. C. Noll, and J. A. Fessler, “Accelerating SENSE-type MR image reconstruction algorithms with incremental gradients,” in *Proc. Intl. Soc. Mag. Res. Med.*, 2014, p. 4400.

[39] 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*, 2014, 103–6.

[40] 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*, 2013, 241–4.

[41] 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, 138–41.

[42] 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.*, vol. 35, no. 8, 3800–8, Aug. 2008.

Bibliography V

[43] 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, 312–5.

[44] M. McGaffin and J. A. Fessler, "Alternating dual updates algorithm for X-ray CT reconstruction on the GPU," *IEEE Trans. Computational Imaging*, vol. 1, no. 3, 186–99, Sep. 2015.