Accelerated optimization methods for large-scale medical image reconstruction

Jeffrey A. Fessler

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

web.eecs.umich.edu/~fessler

CIMI Workshop, Toulouse

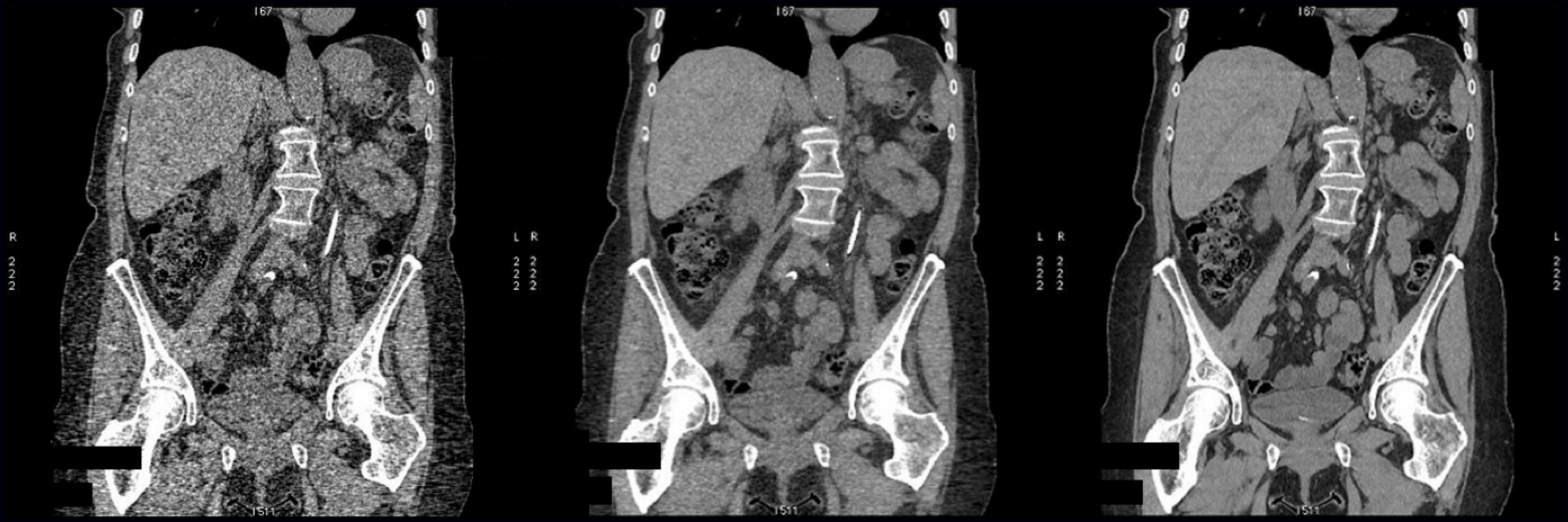
25 June 2013

Disclosure

- Research support from GE Healthcare
- Research support to GE Global Research
- Supported in part by NIH grants R01 HL-098686 and P01 CA-87634
- Equipment support from Intel

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

- Image denoising (review)
- Image restoration
  Antonios Matakos, Sathish Ramani, JF, IEEE T-IP, May 2013
  Accelerated edge-preserving image restoration without boundary artifacts
- Low-dose X-ray CT image reconstruction
  Sathish Ramani & JF, IEEE T-MI, Mar. 2012
  A splitting-based iterative algorithm for accelerated statistical X-ray CT reconstruction
  Donghwan Kim, Sathish Ramani, JF, Fully3D June 2013
  Accelerating X-ray CT ordered subsets image reconstruction with Nesterov's first-order methods
- Model-based MR image reconstruction
  Sathish Ramani & JF, IEEE T-MI, Mar. 2011
  Parallel MR image reconstruction using augmented Lagrangian methods
- Image in-painting (*e.g.*, from cutset sampling) using sparsity

Image denoising

Denoising using sparsity

Measurement model:

$$\underbrace{\boldsymbol{y}}_{\text{observed}} = \underbrace{\boldsymbol{x}}_{\text{unknown}} + \underbrace{\boldsymbol{\varepsilon}}_{\text{noise}}$$

Object model: assume $\boldsymbol{Q}\boldsymbol{x}$ is sparse (compressible) for some orthogonal sparsifying transform $\boldsymbol{Q}$, such as an orthogonal wavelet transform (OWT).

Sparsity regularized estimator:

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

Regularization parameter β determines trade-off.

Equivalently (because $\boldsymbol{Q}^{-1} = \boldsymbol{Q}'$ is an orthonormal matrix):

$$\hat{\boldsymbol{x}} = \boldsymbol{Q}'\hat{\boldsymbol{\theta}}, \quad \hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \frac{1}{2}\|\boldsymbol{Q}\boldsymbol{y}-\boldsymbol{\theta}\|_2^2 + \beta\,\|\boldsymbol{\theta}\|_p = \operatorname{shrink}(\boldsymbol{Q}\boldsymbol{y} : \beta, p)$$

Non-iterative solution!

Orthogonal transform thresholding

Equation:

$$\hat{\boldsymbol{x}} = \boldsymbol{Q}' \operatorname{shrink}(\boldsymbol{Q}\boldsymbol{y} : \beta, p)$$

Block diagram:

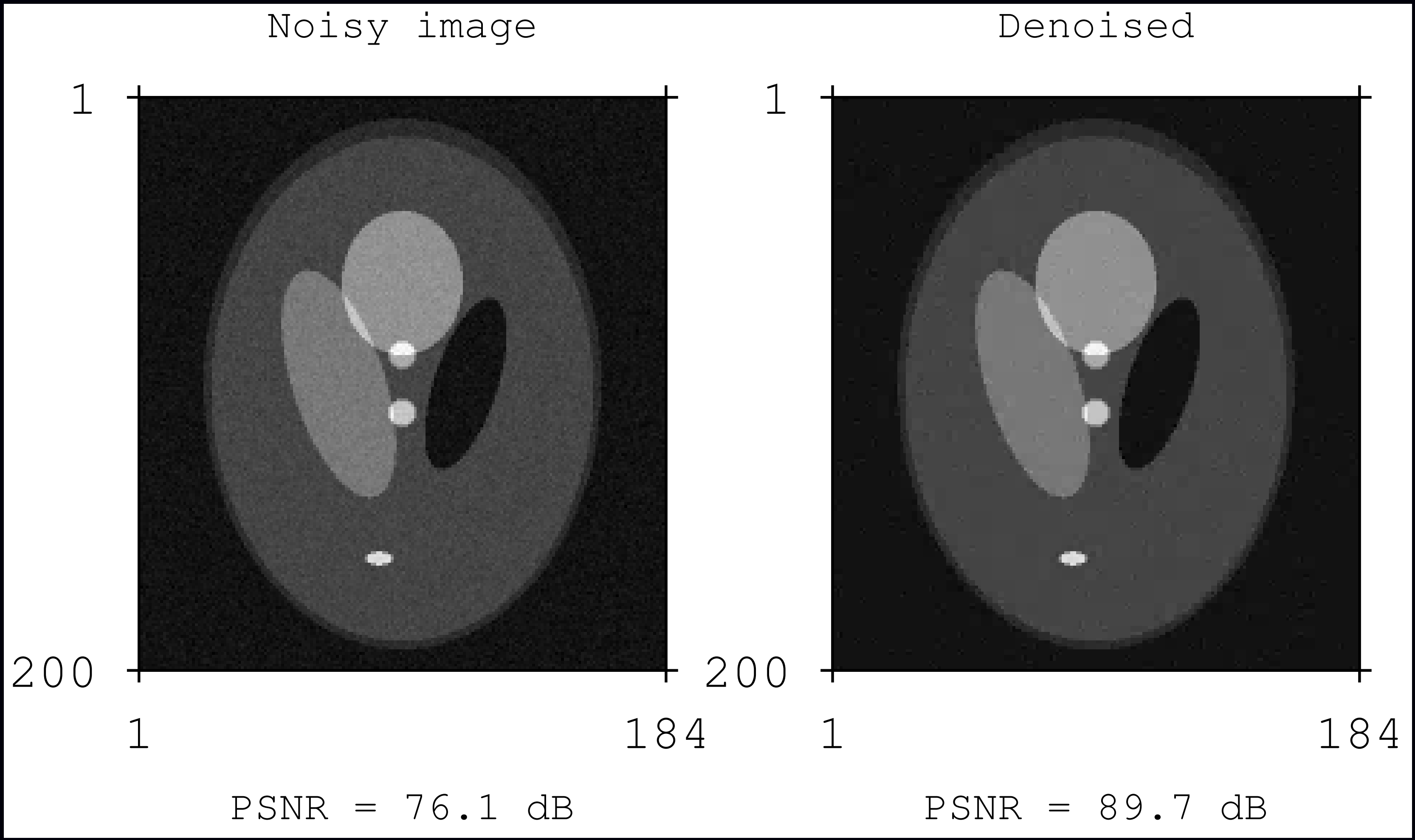
Noisy image $\boldsymbol{y}$ $\rightarrow$ $\boxed{\text{Analysis Transform } \boldsymbol{Q}}$ $\rightarrow$ $\boxed{\text{Shrink } \beta, p}$ $\rightarrow \hat{\boldsymbol{\theta}} \rightarrow$ $\boxed{\text{Synthesis Transform } \boldsymbol{Q}'}$ $\rightarrow$ Denoised image $\hat{\boldsymbol{x}}$

todo: show shrink function for $p = 1$ and $p = 0$

But sparsity in orthogonal transforms often yields artifacts.

Spin cycling...

Hard thresholding example



$p = 0$, orthonormal Haar wavelets

Sparsity using shift-invariant models

Analysis form:
Assume $\boldsymbol{Rx}$ is sparse for some sparsifying transform $\boldsymbol{R}$.
Often $\boldsymbol{R}$ is a "tall" matrix, *e.g.*, finite differences along horizontal and vertical directions, *i.e.*, anisotropic total variation (TV).
Often $\boldsymbol{R}$ is shift invariant: $\|\boldsymbol{Rx}\|_p = \|\boldsymbol{R}\,\mathrm{circshift}(\boldsymbol{x})\|_p$ and $\boldsymbol{R}'\boldsymbol{R}$ is circulant.

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{y}-\boldsymbol{x}\|_2^2 + \beta \underbrace{\|\boldsymbol{Rx}\|_p}_{\text{transform sparsity}}.$$

Synthesis form
Assume $\boldsymbol{x} = \boldsymbol{S\theta}$ where coefficient vector $\boldsymbol{\theta}$ is sparse.
Often $\boldsymbol{S}$ is a "fat" matrix (over-complete dictionary) and $\boldsymbol{S}'\boldsymbol{S}$ is circulant.

$$\hat{\boldsymbol{x}} = \boldsymbol{S}\hat{\boldsymbol{\theta}}, \quad \hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \frac{1}{2}\|\boldsymbol{y}-\boldsymbol{S\theta}\|_2^2 + \beta \underbrace{\|\boldsymbol{\theta}\|_p}_{\text{sparse coefficients}}$$

Analysis form preferable to synthesis form?
(Elad *et al.*, Inv. Prob., June 2007)

Constrained optimization

Unconstrained estimator (analysis form for illustration):

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{x}\|_2^2 + \beta \|\boldsymbol{R}\boldsymbol{x}\|_p .$$

(Nonnegativity constraint or box constraints easily added.)

Equivalent constrained optimization problem:

$$\min_{\boldsymbol{x},\boldsymbol{v}} \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{x}\|_2^2 + \beta \|\boldsymbol{v}\|_p \text{ sub. to } \boldsymbol{v} = \boldsymbol{R}\boldsymbol{x}.$$

(Y. Wang *et al.*, SIAM J. Im. Sci., 2008)

(M Afonso, J Bioucas-Dias, M Figueiredo, IEEE T-IP, Sep. 2010)

(The auxiliary variable $\boldsymbol{v}$ is discarded after optimization; keep only $\hat{\boldsymbol{x}}$.)

Penalty approach:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \min_{\boldsymbol{v}} \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{x}\|_2^2 + \beta \|\boldsymbol{v}\|_p + \frac{\mu}{2} \|\boldsymbol{v} - \boldsymbol{R}\boldsymbol{x}\|_2^2 .$$

Large μ better enforces the constraint $\boldsymbol{v} = \boldsymbol{R}\boldsymbol{x}$, but can worsen conditioning.

Preferable (?) approach: augmented Lagrangian.

Augmented Lagrangian method: V1

General linearly constrained optimization problem:

$$\min_{\boldsymbol{u}} \Psi(\boldsymbol{u}) \text{ sub. to } \boldsymbol{C}\boldsymbol{u} = \boldsymbol{b}.$$

Form *augmented Lagrangian*:

$$L(\boldsymbol{u}, \boldsymbol{\gamma}) \triangleq \Psi(\boldsymbol{u}) + \boldsymbol{\gamma}'(\boldsymbol{C}\boldsymbol{u} - \boldsymbol{b}) + \frac{\rho}{2} \|\boldsymbol{C}\boldsymbol{u} - \boldsymbol{b}\|_2^2$$

where $\boldsymbol{\gamma}$ is the *dual variable* or *Lagrange multiplier vector*.

AL method alternates between minimizing over $\boldsymbol{u}$ and gradient ascent on $\boldsymbol{\gamma}$:

$$\begin{aligned} \boldsymbol{u}^{(n+1)} &= \arg\min_{\boldsymbol{u}} L(\boldsymbol{u}, \boldsymbol{\gamma}^{(n)}) \\ \boldsymbol{\gamma}^{(n+1)} &= \boldsymbol{\gamma}^{(n)} + \rho \left(\boldsymbol{C}\boldsymbol{u}^{(n+1)} - \boldsymbol{b}\right). \end{aligned}$$

Desirable convergence properties.
AL penalty parameter ρ affects convergence *rate*, not solution!

Unfortunately, minimizing over $\boldsymbol{u}$ is impractical here:

$\boldsymbol{v} = \boldsymbol{R}\boldsymbol{x}$ equivalent to $\boldsymbol{C}\boldsymbol{u} = \boldsymbol{b}$, $\boldsymbol{C} = [\boldsymbol{R} \quad -\boldsymbol{I}]$, $\boldsymbol{u} = \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{v} \end{bmatrix}$, $\boldsymbol{b} = \boldsymbol{0}$.

Augmented Lagrangian method: V2

General linearly constrained optimization problem:

$$\min_{\boldsymbol{u}} \Psi(\boldsymbol{u}) \ \text{sub. to } \boldsymbol{C}\boldsymbol{u} = \boldsymbol{b}.$$

Form (modified) *augmented Lagrangian* by completing the square:

$$L(\boldsymbol{u}, \boldsymbol{\eta}) \triangleq \Psi(\boldsymbol{u}) + \frac{\rho}{2} \|\boldsymbol{C}\boldsymbol{u} - \boldsymbol{\eta}\|_2^2 + C_{\boldsymbol{\eta}},$$

where $\boldsymbol{\eta} \triangleq \boldsymbol{b} - \frac{1}{\rho}\boldsymbol{\gamma}$ is a modified *dual variable* or *Lagrange multiplier vector*.

AL method alternates between minimizing over $\boldsymbol{u}$ and gradient ascent on $\boldsymbol{\eta}$:

$$\begin{aligned} \boldsymbol{u}^{(n+1)} &= \arg\min_{\boldsymbol{u}} L(\boldsymbol{u}, \boldsymbol{\gamma}^{(n)}) \\ \boldsymbol{\eta}^{(n+1)} &= \boldsymbol{\eta}^{(n)} - (\boldsymbol{C}\boldsymbol{u}^{(n+1)} - \boldsymbol{b}). \end{aligned}$$

Desirable convergence properties.
AL penalty parameter ρ affects convergence *rate*, not solution!

Unfortunately, minimizing over $\boldsymbol{u}$ is impractical here:

$\boldsymbol{v} = \boldsymbol{R}\boldsymbol{x}$ equivalent to $\boldsymbol{C}\boldsymbol{u} = \boldsymbol{b}, \quad \boldsymbol{C} = [\boldsymbol{R} \quad -\boldsymbol{I}], \quad \boldsymbol{u} = \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{v} \end{bmatrix}, \quad \boldsymbol{b} = \boldsymbol{0}.$

Alternating direction method of multipliers (ADMM)

When $\boldsymbol{u}$ has multiple component vectors, *e.g.*, $\boldsymbol{u} = \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{v} \end{bmatrix}$,

rewrite (modified) augmented Lagrangian in terms of all component vectors:

$$\begin{aligned} L(\boldsymbol{x}, \boldsymbol{v}; \boldsymbol{\eta}) &= \Psi(\boldsymbol{x}, \boldsymbol{v}) + \frac{\rho}{2} \|\boldsymbol{R}\boldsymbol{x} - \boldsymbol{v} - \boldsymbol{\eta}\|_2^2 \\ &= \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{x}\|_2^2 + \beta \|\boldsymbol{v}\|_p + \underbrace{\frac{\rho}{2} \|\boldsymbol{R}\boldsymbol{x} - \boldsymbol{v} - \boldsymbol{\eta}\|_2^2}_{\textit{cf.}\text{ penalty!}} \end{aligned}$$

because here $\boldsymbol{C}\boldsymbol{u} = \boldsymbol{R}\boldsymbol{x} - \boldsymbol{v}$.

Alternate between minimizing over each *component* vector:

$$\begin{aligned} \boldsymbol{x}^{(n+1)} &= \arg\min_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{v}^{(n)}, \boldsymbol{\eta}^{(n)}) \\ \boldsymbol{v}^{(n+1)} &= \arg\min_{\boldsymbol{v}} L(\boldsymbol{x}^{(n+1)}, \boldsymbol{v}, \boldsymbol{\eta}^{(n)}) \\ \boldsymbol{\eta}^{(n+1)} &= \boldsymbol{\eta}^{(n)} + \left(\boldsymbol{R}\boldsymbol{x}^{(n+1)} - \boldsymbol{v}^{(n+1)}\right). \end{aligned}$$

Reasonably desirable convergence properties. (Inexact inner minimizations!)
Sufficient conditions on matrix $\boldsymbol{C}$.

(Eckstein & Bertsekas, Math. Prog., Apr. 1992)

(Douglas and Rachford, Tr. Am. Math. Soc., 1956, heat conduction problems)

ADMM for image denoising

Augmented Lagrangian:

$$L(\boldsymbol{x}, \boldsymbol{v}; \boldsymbol{\eta}) = \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{x}\|_2^2 + \beta \|\boldsymbol{v}\|_p + \frac{\rho}{2} \|\boldsymbol{R}\boldsymbol{x} - \boldsymbol{v} - \boldsymbol{\eta}\|_2^2$$

Update of primal variable (unknown image):

$$\boldsymbol{x}^{(n+1)} = \arg\min_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{v}^{(n)}, \boldsymbol{\eta}^{(n)}) = \underbrace{[\boldsymbol{I} + \rho \boldsymbol{R}'\boldsymbol{R}]^{-1}}_{\text{Wiener filter}} \left(\boldsymbol{y} + \rho \boldsymbol{R}' \left(\boldsymbol{v}^{(n)} + \boldsymbol{\eta}^{(n)}\right)\right)$$

Update of auxiliary variable: (No "corner rounding" needed for ℓ_1.)

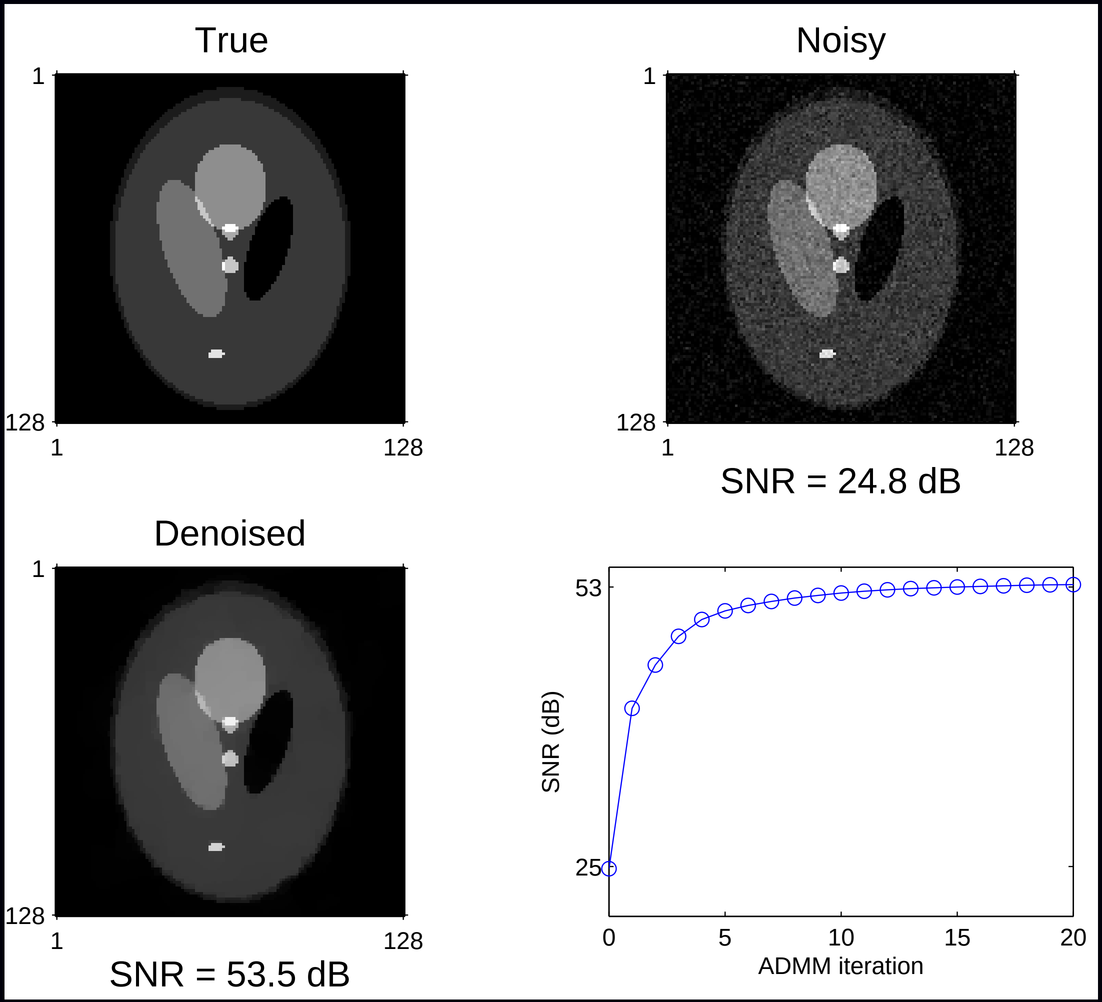
$$\boldsymbol{v}^{(n+1)} = \arg\min_{\boldsymbol{v}} L(\boldsymbol{x}^{(n+1)}, \boldsymbol{v}, \boldsymbol{\eta}^{(n)}) = \operatorname{shrink}\left(\boldsymbol{R}\boldsymbol{x}^{(n+1)} - \boldsymbol{\eta}^{(n)}; \beta/\rho, p\right)$$

Update of multiplier: $\boldsymbol{\eta}^{(n+1)} = \boldsymbol{\eta}^{(n)} + \left(\boldsymbol{R}\boldsymbol{x}^{(n+1)} - \boldsymbol{v}^{(n+1)}\right)$

Equivalent to "*split Bregman*" approach.
(Goldstein & Osher, SIAM J. Im. Sci. 2009)

Each update is simple and exact (non-iterative) if $[\boldsymbol{I} + \rho \boldsymbol{R}'\boldsymbol{R}]^{-1}$ is easy.

ADMM image denoising example



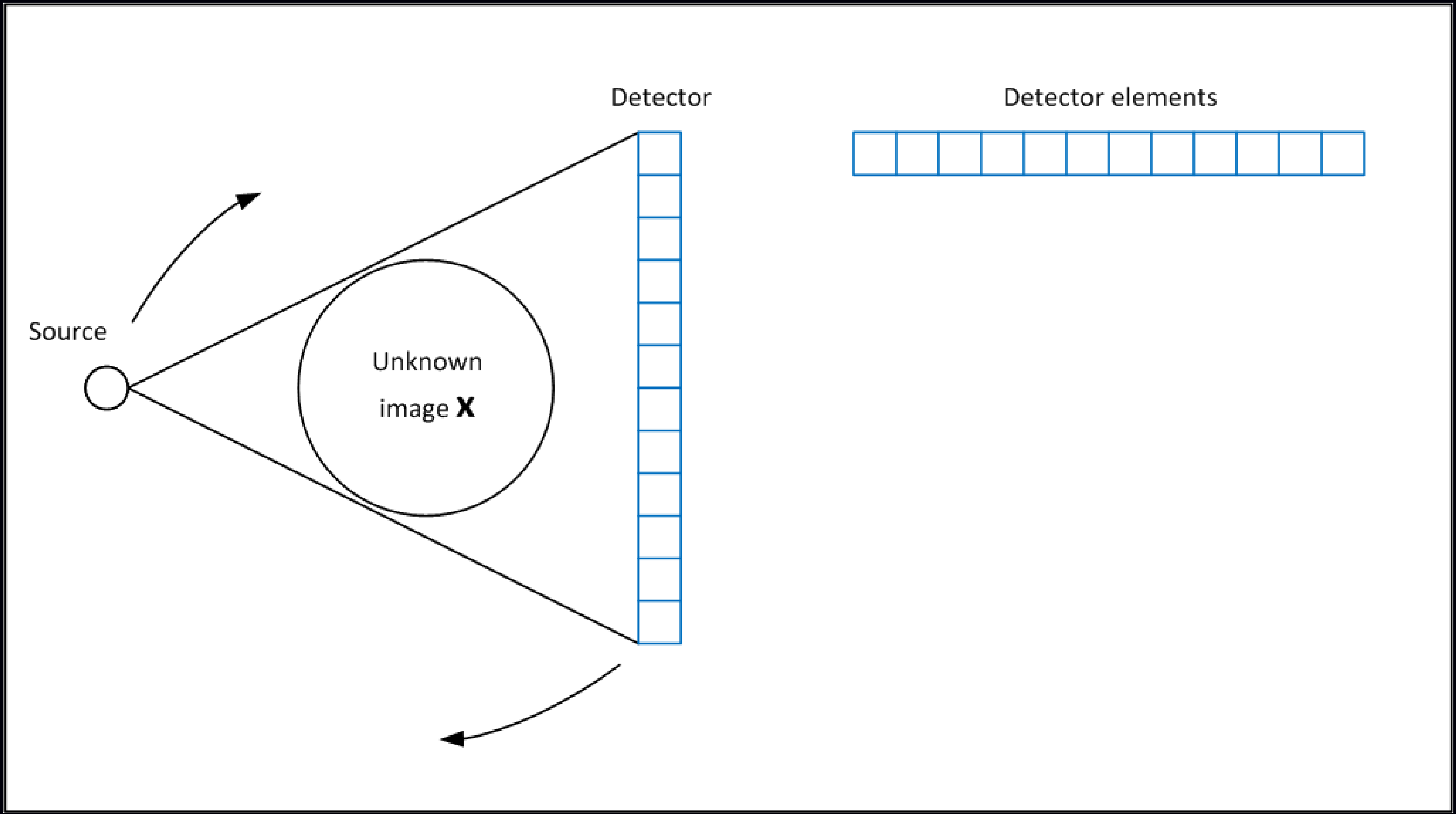
$\boldsymbol{R}$: horizontal and vertical finite differences (anisotropic TV),
$p = 1$ (*i.e.*, ℓ_1), $\beta = 1/2$, $\rho = 1$ (condition number of $(\boldsymbol{I} + \rho \boldsymbol{R}'\boldsymbol{R})$ is 9)

ADMM image denoising iterates

admm denoising movie

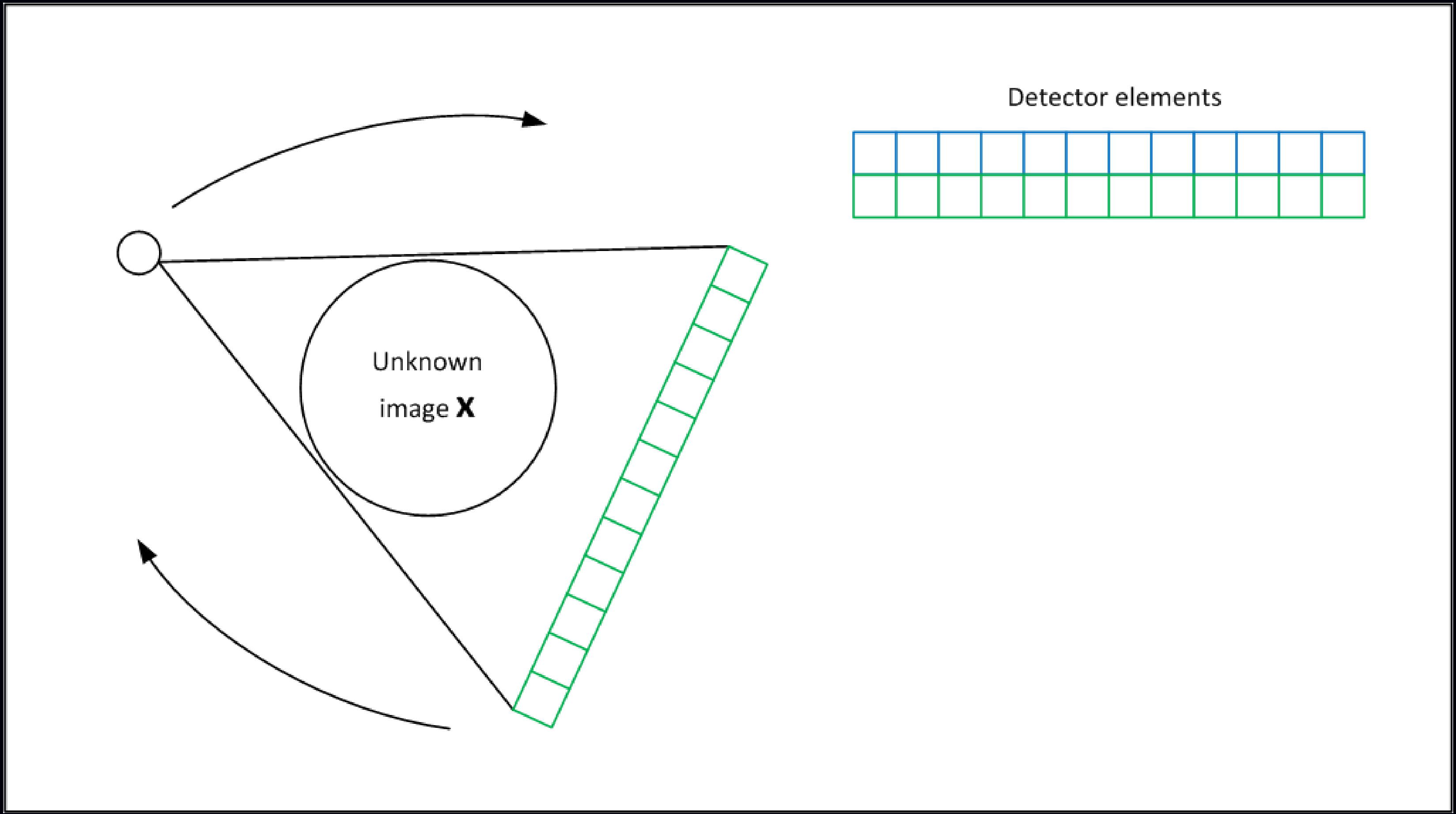
X-ray CT image reconstruction
Part 1: ADMM

X-ray CT review 1



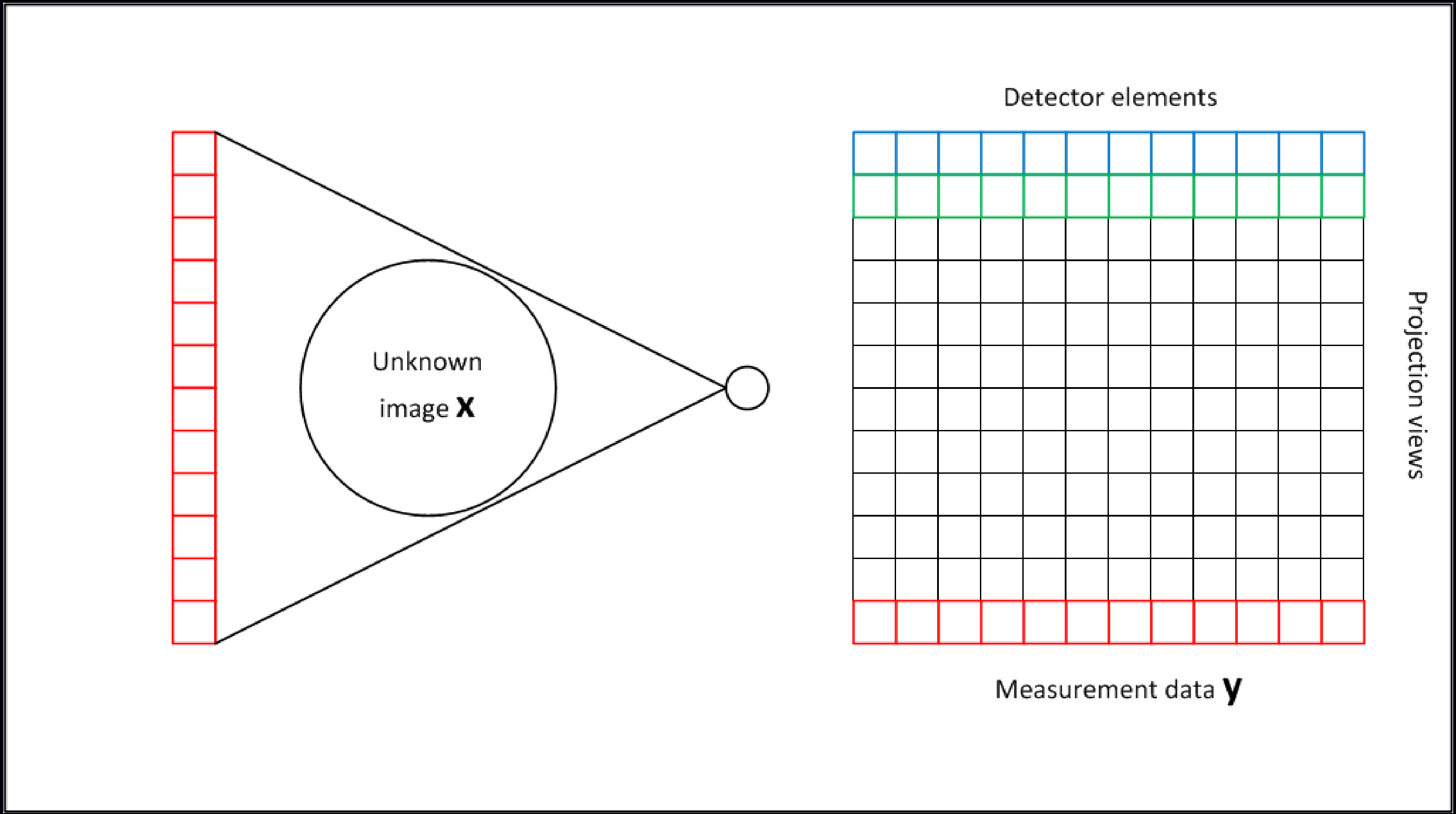
X-ray source transmits X-ray photons through object.
Recorded signal relates to line integral of attenuation along photon path.

X-ray CT review 2



X-ray source and detector rotate around object.

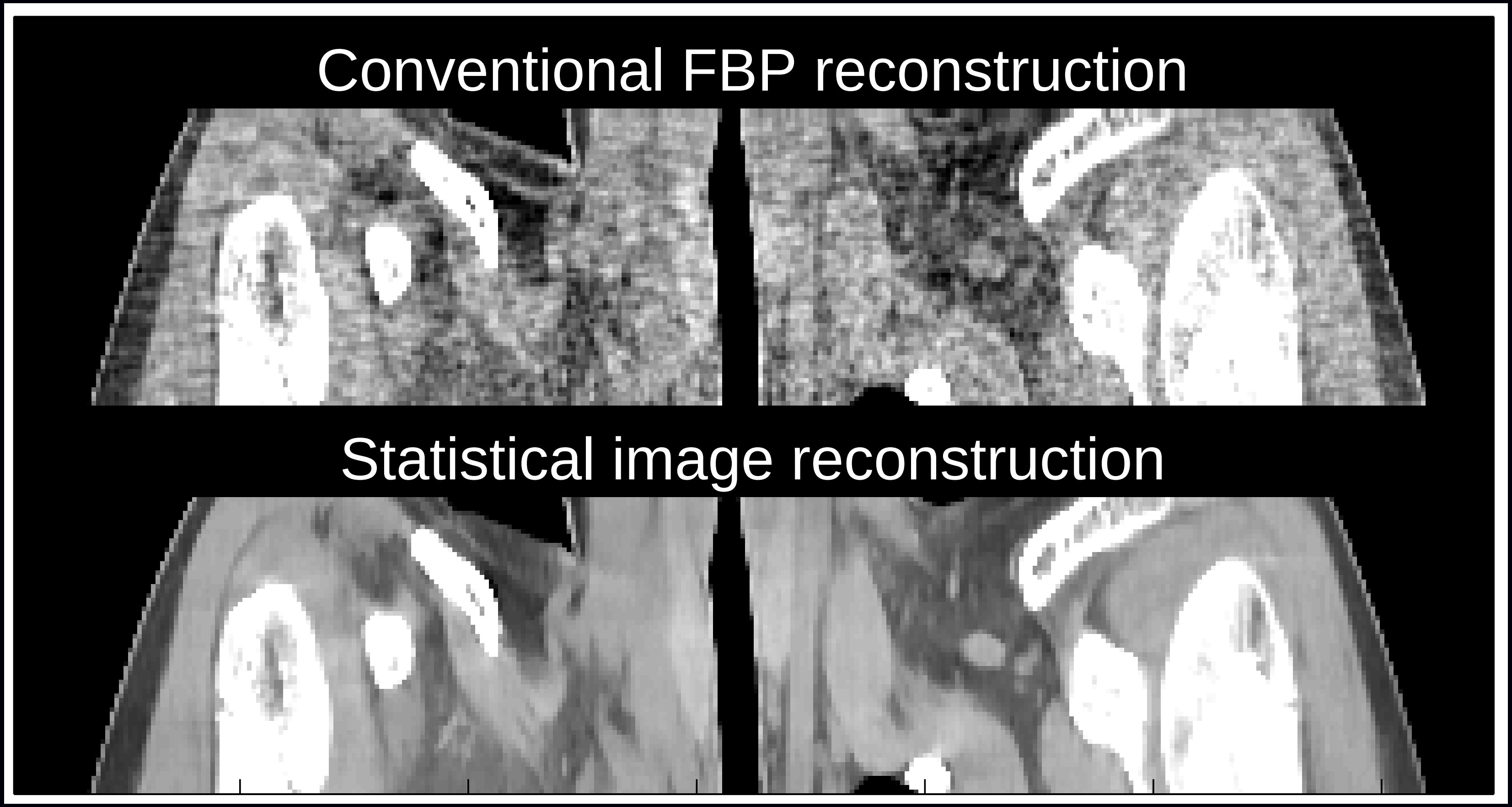
X-ray CT review 3



Collection of recorded views called a sinogram.
Goal is to reconstruct (3D) image of object attenuation from sinogram.

Lower-dose X-ray CT

Radiation dose proportional to X-ray source intensity.
Reducing dose $\Longrightarrow$ fewer recorded photons $\Longrightarrow$ lower SNR
Conventional filter back-project (FBP) method derived for noiseless data



Low-dose X-ray CT image reconstruction

Regularized estimator:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x} \succeq \boldsymbol{0}} \underbrace{\frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_{\boldsymbol{W}}^2}_{\text{data fit}} + \beta \underbrace{\|\boldsymbol{R}\boldsymbol{x}\|_p}_{\text{sparsity}}.$$

Complications:

- Large problem size
  - $\boldsymbol{x}$: $512 \times 512 \times 800 \approx 2 \cdot 10^8$ unknown image volume
  - $\boldsymbol{y}$: $888 \times 64 \times 7000 \approx 4 \cdot 10^8$ measured sinogram
  - $\boldsymbol{A}$: $(4 \cdot 10^8) \times (2 \cdot 10^8)$ system matrix
  - $\boldsymbol{A}$ is sparse but still too large to store
  - Projection $\boldsymbol{A}\boldsymbol{x}$ and back-projection $\boldsymbol{A}'\boldsymbol{r}$ operations computed on the fly
  - Computing gradient $\nabla \Psi(\boldsymbol{x}) = \boldsymbol{A}'\boldsymbol{W}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}) + \beta \nabla \mathsf{R}(\boldsymbol{x})$ requires projection and back-projection operations that dominate computation
- $\boldsymbol{A}'\boldsymbol{A}$ is not circulant (but “approximately Toeplitz” in 2D)
- $\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}$ is highly shift variant due to huge dynamic range of weighting $\boldsymbol{W}$
- Non-quadratic (edge-preserving) regularizer *e.g.*, $\mathsf{R}(\boldsymbol{x}) = \|\boldsymbol{R}\boldsymbol{x}\|_p$
- Nonnegativity constraint
- Goal: fast parallelizable algorithms that “converge” in a few iterations

Basic ADMM for X-ray CT

Basic equivalent constrained optimization problem (*cf.* split Bregman):

$$\min_{\boldsymbol{x} \succeq \mathbf{0}, \boldsymbol{v}} \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_{\boldsymbol{W}}^2 + \beta \|\boldsymbol{v}\|_p \text{ sub. to } \boldsymbol{v} = \boldsymbol{R}\boldsymbol{x}.$$

Corresponding (modified) augmented Lagrangian (*cf.* “split Bregman”):

$$L(\boldsymbol{x}, \boldsymbol{v}; \boldsymbol{\eta}) = \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_{\boldsymbol{W}}^2 + \beta \|\boldsymbol{v}\|_p + \frac{\rho}{2} \|\boldsymbol{R}\boldsymbol{x} - \boldsymbol{v} - \boldsymbol{\eta}\|_2^2$$

ADMM update of primal variable (unknown image):

$$\boldsymbol{x}^{(n+1)} = \arg\min_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{v}^{(n)}, \boldsymbol{\eta}^{(n)}) = \left[\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} + \rho \boldsymbol{R}'\boldsymbol{R}\right]^{-1} \left(\boldsymbol{A}'\boldsymbol{W}'\boldsymbol{y} + \rho \boldsymbol{R}' \left(\boldsymbol{v}^{(n)} + \boldsymbol{\eta}^{(n)}\right)\right)$$

Drawbacks:

- Ignores nonnegativity constraint
- $\left[\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} + \rho \boldsymbol{R}'\boldsymbol{R}\right]^{-1}$ requires iteration (*e.g.*, PCG) but hard to precondition. “second order method”
- Auxiliary variable $\boldsymbol{v} = \boldsymbol{R}\boldsymbol{x}$ is enormous in 3D CT

Improved ADMM for X-ray CT

$$\min_{\boldsymbol{x}\succeq\boldsymbol{0},\boldsymbol{u},\boldsymbol{v}} \frac{1}{2}\|\boldsymbol{y}-\boldsymbol{u}\|_{\boldsymbol{W}}^2 + \beta\|\boldsymbol{v}\|_p \text{ sub. to } \boldsymbol{v}=\boldsymbol{R}\boldsymbol{x}, \quad \boldsymbol{u}=\boldsymbol{A}\boldsymbol{x}.$$

Corresponding (modified) augmented Lagrangian:

$$L(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v};\boldsymbol{\eta}_1,\boldsymbol{\eta}_2) = \frac{1}{2}\|\boldsymbol{y}-\boldsymbol{u}\|_{\boldsymbol{W}}^2 + \beta\|\boldsymbol{v}\|_p + \frac{\rho_1}{2}\|\boldsymbol{R}\boldsymbol{x}-\boldsymbol{v}-\boldsymbol{\eta}_1\|_2^2 + \frac{\rho_2}{2}\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{u}-\boldsymbol{\eta}_2\|_2^2$$

ADMM update of primal variable (ignoring nonnegativity):

$$\arg\min_{\boldsymbol{x}} L(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v},\boldsymbol{\eta}_1,\boldsymbol{\eta}_2) = \left[\rho_2\boldsymbol{A}'\boldsymbol{A}+\rho_1\boldsymbol{R}'\boldsymbol{R}\right]^{-1}\left(\rho_1\boldsymbol{R}'(\boldsymbol{v}+\boldsymbol{\eta}_1)+\rho_2\boldsymbol{A}'(\boldsymbol{u}+\boldsymbol{\eta}_2)\right)$$

For 2D CT, $\left[\rho_2\boldsymbol{A}'\boldsymbol{A}+\rho_1\boldsymbol{R}'\boldsymbol{R}\right]^{-1}$ is approximately Toeplitz so a circulant preconditioner is very effective.

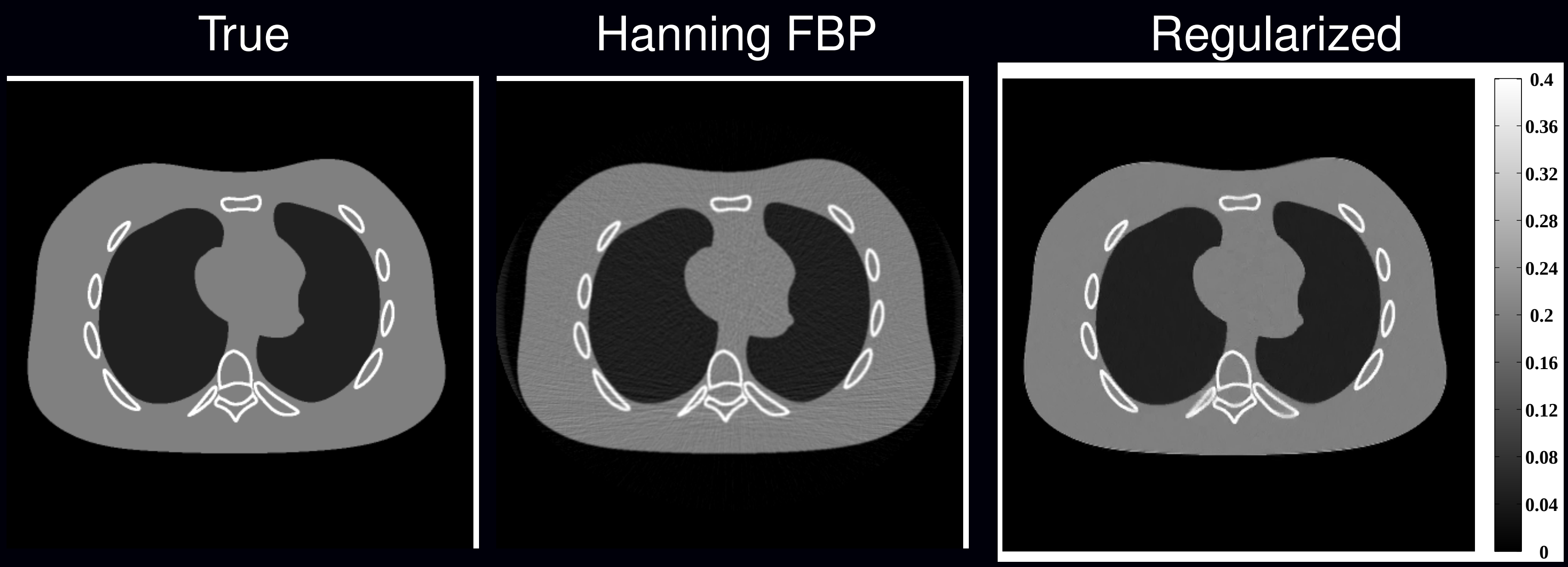
ADMM update of auxiliary variable $\boldsymbol{u}$:

$$\arg\min_{\boldsymbol{u}} L(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v},\boldsymbol{\eta}_1,\boldsymbol{\eta}_2) = \underbrace{\left[\boldsymbol{W}+\rho_2\boldsymbol{I}\right]^{-1}}_{\text{diagonal}}(\boldsymbol{W}\boldsymbol{y}+\rho_2(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{\eta}_2))$$

$\boldsymbol{v}$ update is shrinkage again. Reasonably simple to code.

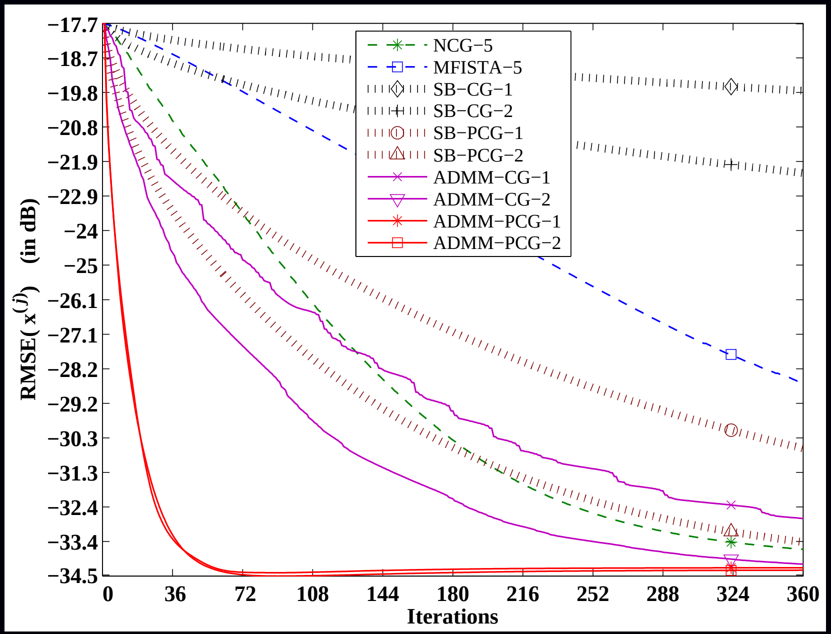
(Sathish Ramani & JF, IEEE T-MI, Mar. 2012)

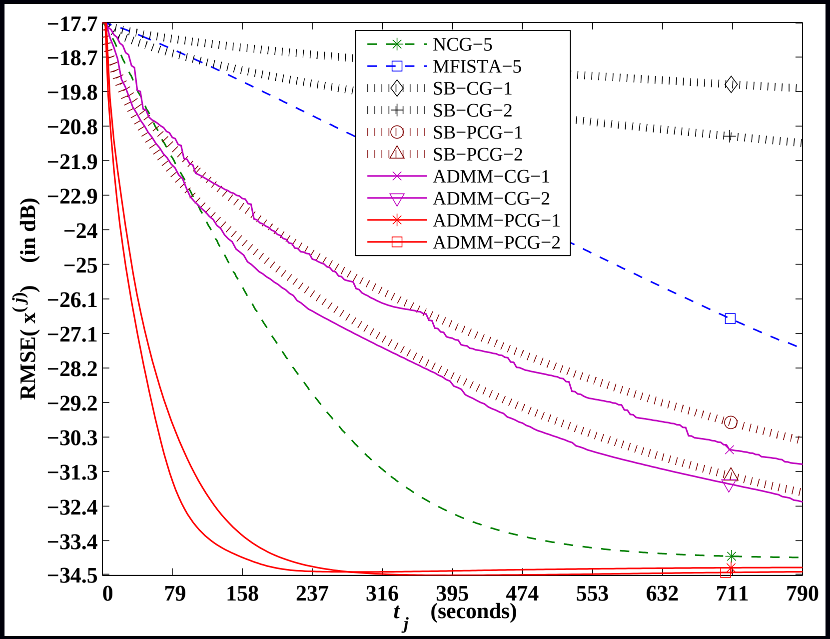
2D X-ray CT image reconstruction results: quality



PWLS with ℓ_1 regularization of shift-invariant Haar wavelet transform.
No nonnegativity constraint, but probably unimportant if well-regularized.

2D X-ray CT image reconstruction results: speed





Circulant preconditioner for $\left[\rho_2 \boldsymbol{A}'\boldsymbol{A} + \rho_1 \boldsymbol{R}'\boldsymbol{R}\right]^{-1}$ is crucial to acceleration.

Similar results for real head CT scan in paper.

Lower-memory ADMM for X-ray CT

$$\min_{x, u, z \succeq 0} \frac{1}{2} \|y - u\|_W^2 + \beta \|Rz\|_p \text{ sub. to } z = x, \quad u = Ax.$$

(M McGaffin, S Ramani, JF, SPIE 2012)

Corresponding (modified) augmented Lagrangian:

$$L(x, u, z; \eta_1, \eta_2) = \frac{1}{2} \|y - u\|_W^2 + \beta \|Rz\|_p + \frac{\rho_1}{2} \|x - z - \eta_1\|_2^2 + \frac{\rho_2}{2} \|Ax - u - \eta_2\|_2^2$$

ADMM update of primal variable (nonnegativity not required, use PCG):

$$\arg\min_x L(x, u, z, \eta_1, \eta_2) = \left[\rho_2 A'A + \rho_1 I\right]^{-1} \left(\rho_1 (z + \eta_1) + \rho_2 A'(u + \eta_2)\right).$$

ADMM update of auxiliary variable z:

$$\arg\min_{z \succeq 0} L(x, u, z, \eta_1, \eta_2) = \arg\min_{z \succeq 0} \frac{\rho_1}{2} \|x - z - \eta_1\|_2^2 + \beta \|Rz\|_p.$$

Use nonnegatively constrained, edge-preserving image denoising.

ADMM updates of auxiliary variables u and v same as before.
Variations...

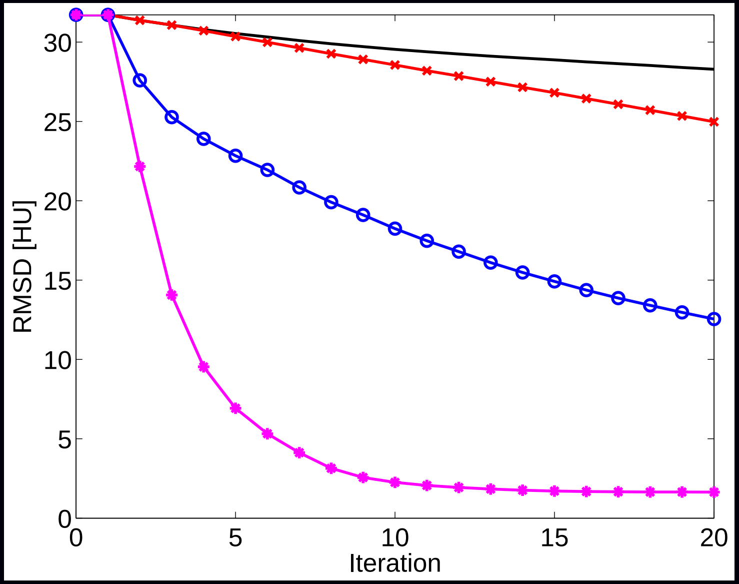
3D X-ray CT image reconstruction results

Awaiting better preconditioner for $\left[\rho_2 \boldsymbol{A}'\boldsymbol{A} + \rho_1 \boldsymbol{I}\right]^{-1}$ in 3D CT...
This is not an “easy problem” like in (idealized) image restoration...

X-ray CT image reconstruction
Part 2: OS+Momentum

OS+Momentum preview

- Optimization transfer (aka majorize-minimize, half-quadratic)
- Ordered-subsets (OS) acceleration (aka incremental gradient, block iterative)
- Nesterov's momentum-based acceleration
- Proposed OS+Momentum approach combines both



Donghwan Kim, Sathish Ramani, JF, Fully3D June 2013

Accelerating X-ray CT ordered subsets image reconstruction with Nesterov's first-order methods

Optimization transfer method

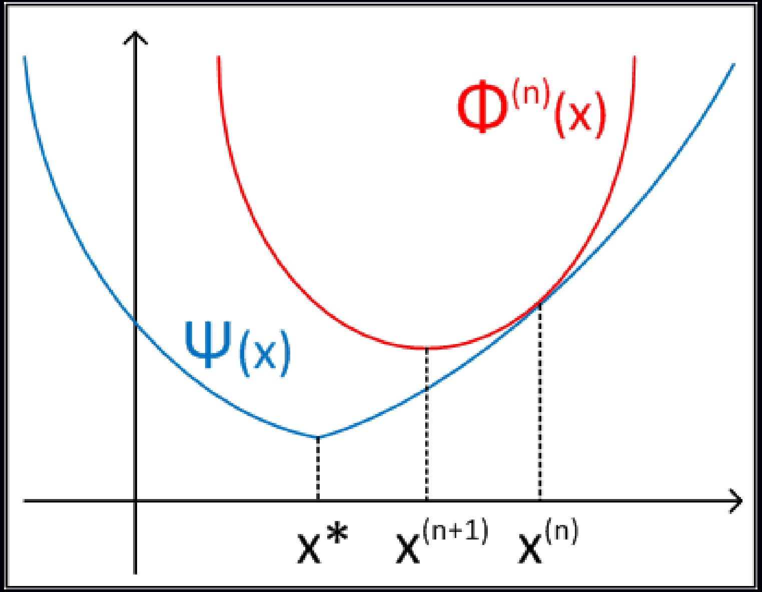
(aka majorization-minimization or half-quadratic)

- At nth iteration, replace original cost function $\Psi(\boldsymbol{x})$ by a surrogate function $\phi^{(n)}(\boldsymbol{x})$ that is easier to minimize:

$$\boldsymbol{x}^{(n+1)} = \underset{\boldsymbol{x} \succeq \boldsymbol{0}}{\arg\min}\, \phi^{(n)}(\boldsymbol{x})$$

- To monotonically decrease $\Psi(\boldsymbol{x})$, *i.e.*, $\Psi\left(\boldsymbol{x}^{(n+1)}\right) \leq \Psi(\boldsymbol{x}^{(n)})$, surrogate should satisfy the following majorization conditions:

$$\begin{aligned} \phi^{(n)}\left(\boldsymbol{x}^{(n)}\right) &= \Psi\left(\boldsymbol{x}^{(n)}\right) \\ \phi^{(n)}(\boldsymbol{x}) &\geq \Psi(\boldsymbol{x}), \quad \forall \boldsymbol{x} \succeq \boldsymbol{0} \end{aligned}$$



Separable quadratic surrogate (SQS)

Quadratic surrogate functions are particularly convenient:

$$\Psi(\boldsymbol{x}) \leq \phi^{(n)}(\boldsymbol{x}) \triangleq \Psi\left(\boldsymbol{x}^{(n)}\right) + \nabla\Psi\left(\boldsymbol{x}^{(n)}\right)\left(\boldsymbol{x} - \boldsymbol{x}^{(n)}\right) + \frac{1}{2}\left\|\boldsymbol{x} - \boldsymbol{x}^{(n)}\right\|_{\boldsymbol{D}}^{2},$$

where $\boldsymbol{D}$ is a specially designed diagonal matrix:

$$\boldsymbol{D} = \boldsymbol{D}_L + \beta\boldsymbol{D}_R, \qquad \boldsymbol{D}_L \triangleq \mathrm{diag}\left\{\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}\boldsymbol{1}\right\}, \qquad \boldsymbol{D}_R \triangleq \lambda_{\max}\left(\nabla^2 \mathsf{R}(\boldsymbol{x})\right)\boldsymbol{I}.$$

(Erdoğan and Fessler, PMB, 1999)

(Easier to compute $\boldsymbol{D}_L$ than to find Lipschitz constant of $\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}$.)

SQS leads to trivial M-step:

$$\boldsymbol{x}^{(n+1)} = \arg\min_{\boldsymbol{x}\succeq\boldsymbol{0}} \phi^{(n)}(\boldsymbol{x}) = \left[\boldsymbol{x}^{(n)} - \boldsymbol{D}^{-1}\nabla\Psi\left(\boldsymbol{x}^{(n)}\right)\right]_{+}.$$

“diagonally preconditioned gradient projection method”

Convergence rate of SQS method

Asymptotic convergence rate:

$$\rho\left(\boldsymbol{I} - \boldsymbol{D}^{-1} \nabla^2 \Psi(\hat{\boldsymbol{x}})\right)$$

Slightly generalizing Theorem 3.1 of Beck and Teboulle:

$$\Psi\left(\boldsymbol{x}^{(n)}\right) - \Psi(\hat{\boldsymbol{x}}) \leq \frac{\left\|\boldsymbol{x}^{(0)} - \hat{\boldsymbol{x}}\right\|_{\boldsymbol{D}}^2}{2n}.$$

(Beck and Teboulle, SIAM J Im. Sci., 2009)

Pro: easily parallelized
Con: very slow convergence

Accelerating SQS using Nesterov's momemtum

SQS+Momentum Algorithm:

- Initialize image $\boldsymbol{x}^{(0)}$ and $\boldsymbol{z}^{(0)}$
- for $n = 0, 1, \ldots$

$$\begin{aligned}
\boldsymbol{x}^{(n+1)} &= \operatorname*{arg\,min}_{\boldsymbol{x} \succeq \boldsymbol{0}} \phi^{(n)}\left(\boldsymbol{z}^{(n)}\right) = \left[\boldsymbol{z}^{(n)} - \boldsymbol{D}^{-1} \nabla \Psi\left(\boldsymbol{z}^{(n)}\right)\right]_+ \\
t_{n+1} &= \left(1 + \sqrt{1 + 4t_n^2}\right)/2 \\
\boldsymbol{z}^{(n+1)} &= \boldsymbol{x}^{(n+1)} + \frac{t_n - 1}{t_{n+1}} \left(\boldsymbol{x}^{(n+1)} - \boldsymbol{x}^{(n)}\right)
\end{aligned}$$

Convergence rate of SQS+Momentum method:

$$\Psi\left(\boldsymbol{x}^{(n)}\right) - \Psi(\hat{\boldsymbol{x}}) \le \frac{2\left\|\boldsymbol{x}^{(0)} - \hat{\boldsymbol{x}}\right\|_{\boldsymbol{D}}^2}{(n+1)^2}.$$

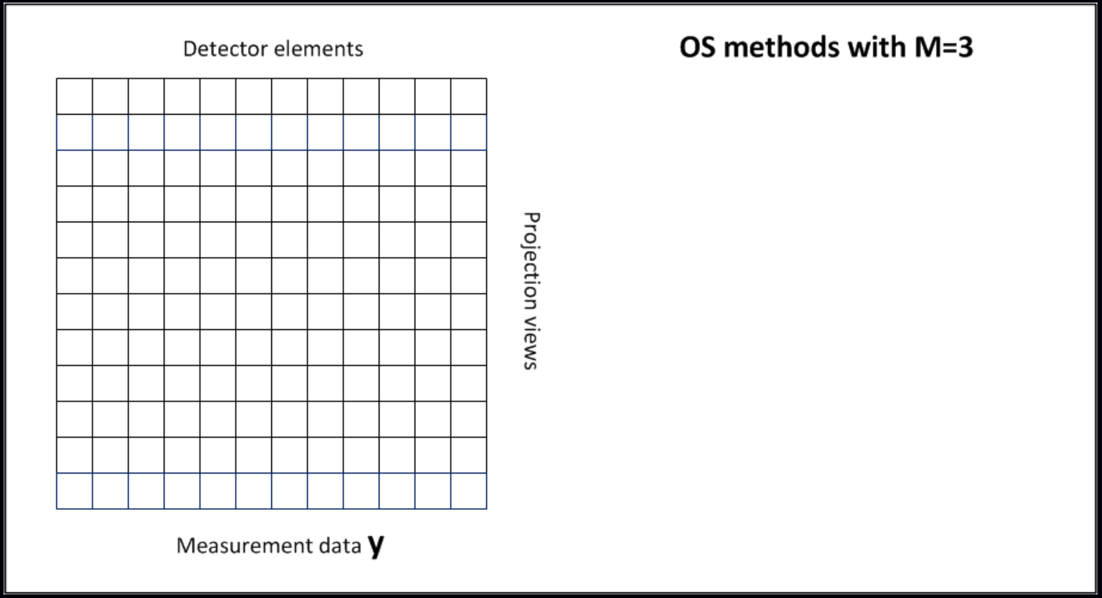
Simple generalization of Thm. 4.4 for FISTA of
(Beck and Teboulle, SIAM J Im. Sci., 2009)

Pro: Almost same computation per iteration; slightly more memory needed.
Con: still converges too slowly for X-ray CT

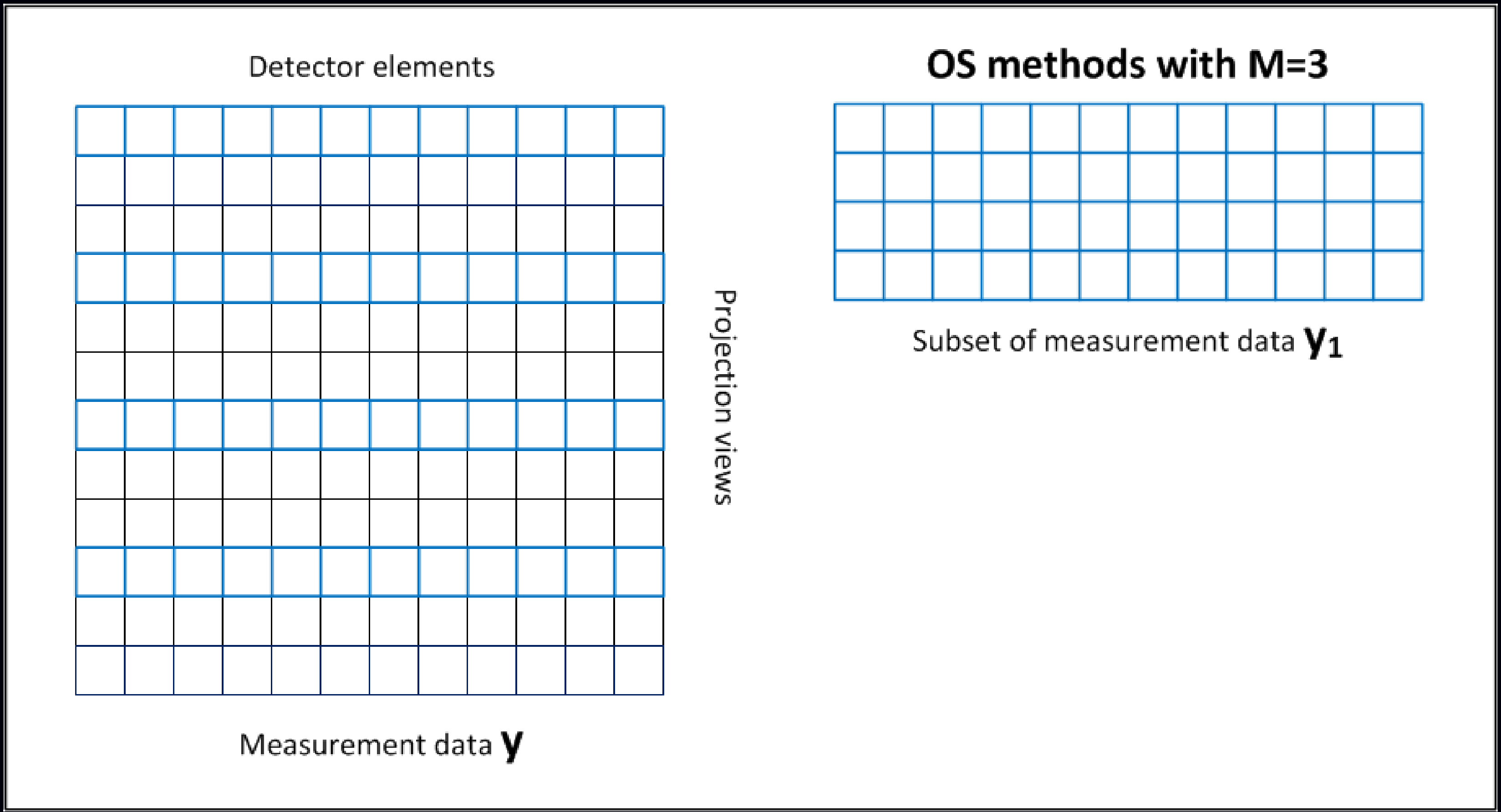
Ordered subsets (OS) methods

- **Recall:** Projection operator A is computationally expensive.
-
- OS methods group projection views into M subsets, and use each subset per each update, instead of using all measurement data.
  (Hudson and Larkin, IEEE T-MI, 1994)
  (Erdoğan and Fessler, PMB, 1999)

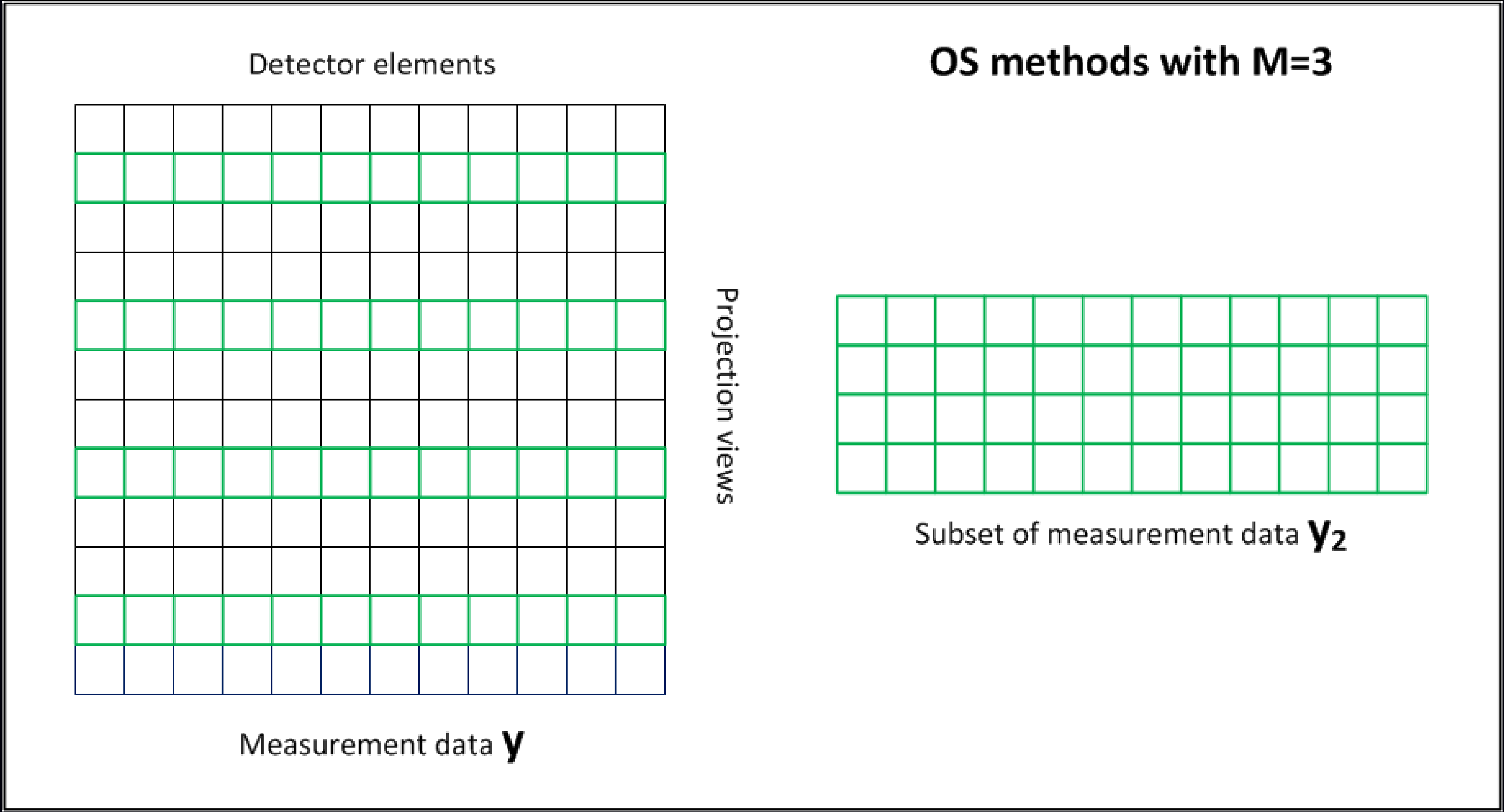
cf block-iterative incremental sub-gradient for machine learning



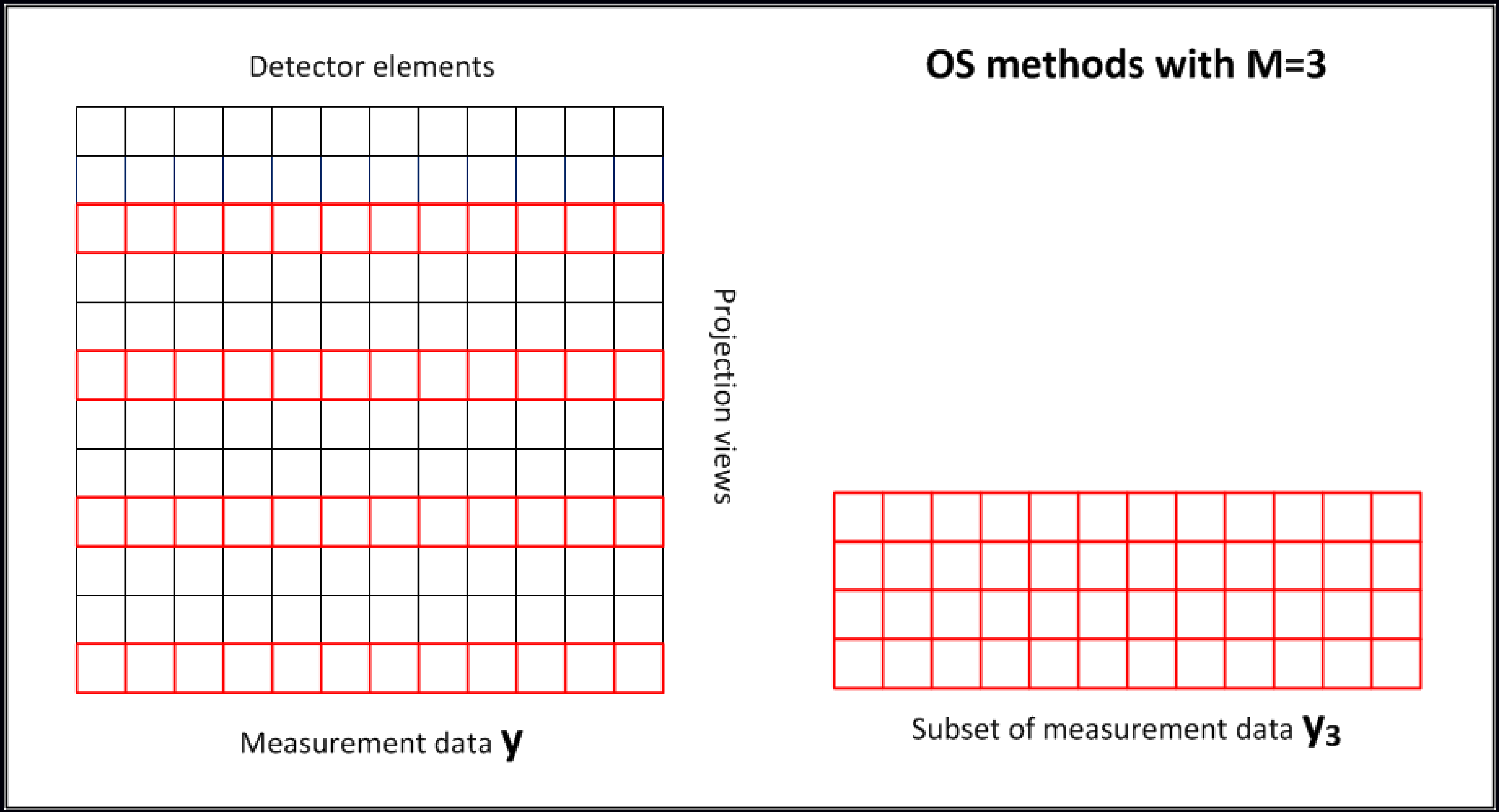
OS projection view grouping



OS projection view grouping



OS projection view grouping



OS algorithm

Cost function decomposition:

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

$\boldsymbol{y}_m$, $\boldsymbol{A}_m$, $\boldsymbol{W}_m$: sinogram rows, system matrix rows, weighting elements for mth subset of projection views

Intuition: in early iterations (when $\boldsymbol{x}^{(n)}$ is far from $\hat{\boldsymbol{x}}$):

$$M \nabla \Psi_m\left(\boldsymbol{x}^{(n)}\right) \approx \nabla \Psi\left(\boldsymbol{x}^{(n)}\right).$$

OS-SQS Algorithm (Erdoğan and Fessler, PMB, 1999)

- Initialize image $\boldsymbol{x}^{(0)}$
- for $n = 0, 1, \ldots$
- for $m = 0, \ldots, M-1$

$$\boldsymbol{x}^{(n+(m+1)/M)} = \arg\min_{\boldsymbol{x} \succeq \boldsymbol{0}} \left[\boldsymbol{x}^{(n+m/M)} - \boldsymbol{D}^{-1} M \nabla \Psi_m\left(\boldsymbol{x}^{(n+m/M)}\right)\right]_+$$

OS-SQS algorithm properties

- One iteration corresponds to updating all M subsets.
  Computation cost similar to original SQS
  (one full forward $\boldsymbol{A}$ and back-projection $\boldsymbol{A}'$ per iteration)
- + Highly parallelizable
- + In early iterations, “we expect” the sequence $\{\boldsymbol{x}^{(n)}\}$ to satisfy

$$\Psi\left(\boldsymbol{x}^{(n)}\right) - \Psi(\hat{\boldsymbol{x}}) \lessapprox \frac{\left\|\boldsymbol{x}^{(0)} - \hat{\boldsymbol{x}}\right\|_{\boldsymbol{D}}^2}{2nM}.$$

  M times acceleration!
- - Does not converge to $\hat{\boldsymbol{x}}$
  Approaches a limit cycle, size related to M
  Luo, Neural Computation, June 1991
- - Computing $\nabla \mathsf{R}(\boldsymbol{x})$ for each of M subsets $\Longrightarrow$ prefer small M
- Since about 1997, OS methods have been used for (unregularized) PET reconstruction in nearly every PET scanner sold.
- Still undesirably slow (for small M) or unstable (for large M) in X-ray CT.

OS+Momentum algorithm

- Initialize image $\boldsymbol{x}^{(0)}$ and $\boldsymbol{z}^{(0)}$
- for $n = 0, 1, \ldots$
- for $m = 0, 1, \ldots$

$$
\begin{aligned}
\boldsymbol{x}^{(n+(m+1)/M)} &= \left[\boldsymbol{z}^{(n+m/M)} - \boldsymbol{D}^{-1} M \nabla \Psi_m\left(\boldsymbol{z}^{(n+m/M)}\right)\right]_+ \\
t_{\text{new}} &= \left(1 + \sqrt{1 + 4t_{\text{old}}^2}\right)/2 \\
\boldsymbol{z}^{(n+(m+1)/M)} &= \boldsymbol{x}^{(n+(m+1)/M)} + \frac{t_{\text{old}} - 1}{t_{\text{new}}} \left(\boldsymbol{x}^{(n+(m+1)/M)} - \boldsymbol{x}^{(n+m/M)}\right) \\
t_{\text{old}} &:= t_{\text{new}}
\end{aligned}
$$

- \+ In early iterations, “we expect” the sequence $\{\boldsymbol{x}^{(n)}\}$ to satisfy

$$
\Psi\left(\boldsymbol{x}^{(n)}\right) - \Psi(\hat{\boldsymbol{x}}) \lessapprox \frac{\left\|\boldsymbol{x}^{(0)} - \hat{\boldsymbol{x}}\right\|_{\boldsymbol{D}}^2}{2(nM)^2}.
$$

  M^2 times acceleration!
- \+ Very similar computation as OS-SQS
- \+ Easily implemented
- \- Unknown convergence properties

Summary of convergence rates

SQS (optimization transfer) methods:

- Convergence rate
  - SQS: $O\left(\frac{1}{n}\right)$
  - SQS+Momentum: $O\left(\frac{1}{n^2}\right)$
- Expected convergence rate with OS method in early iterations
  - OS-SQS: $O\left(\frac{1}{nM}\right)$
  - Proposed OS-SQS+Momentum: $O\left(\frac{1}{(nM)^2}\right)$

- Pros: Owing to M^2 times acceleration from OS methods, we can use small M, improving stability and reducing regularizer computation.
- Cons: Behavior of OS methods with momentum is unknown, while ordinary OS methods approach a limit-cycle. (Luo, Neural Comp., Jun. 1991)

Patient 3D helical CT scan results

- 3D cone-beam helical CT scan with pitch 1.0
- 3D image $\boldsymbol{x}$: $512 \times 512 \times 109$
- voxel size: 1.369 mm $\times$ 1.369 mm $\times$ 0.625 mm
- measured sinogram data $\boldsymbol{y}$: $888 \times 32 \times 7146$
  (detector columns $\times$ detector rows $\times$ projection views)

Convergence rates (empirical)

- Root mean square difference (RMSD) between current $\boldsymbol{x}^{(n)}$ and converged image $\hat{\boldsymbol{x}}$

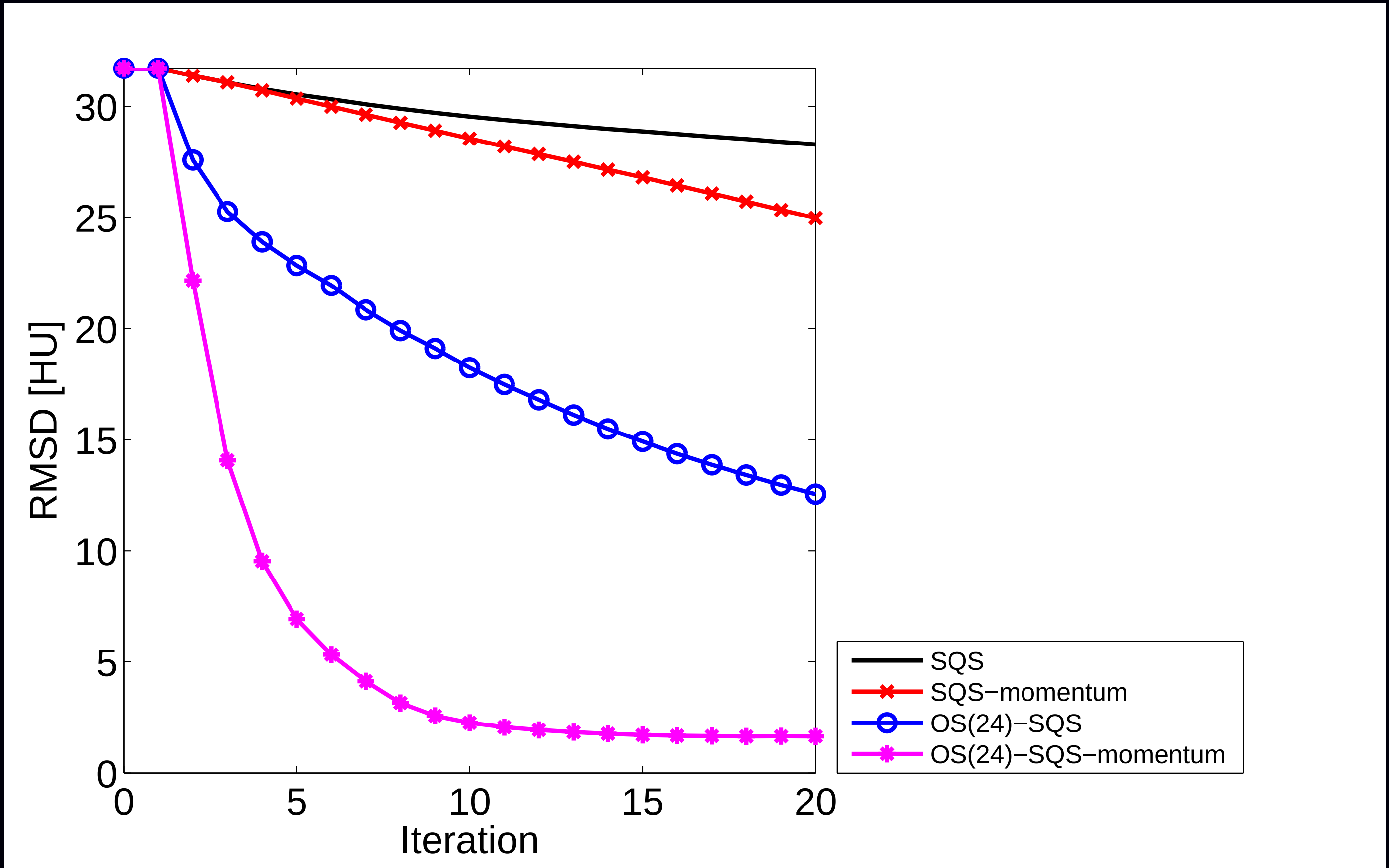
$$\mathrm{RMSD} \triangleq \frac{\|\boldsymbol{x}^{(n)} - \hat{\boldsymbol{x}}\|_2}{\sqrt{N_p}} \text{ [HU]},$$

  where $N_p = 512 \times 512 \times 109$ is the number of image voxels in $\boldsymbol{x}$.
- Normalized RMSD:

$$\mathrm{NRMSD} \triangleq 20 \log_{10} \left(\frac{\|\boldsymbol{x}^{(n)} - \hat{\boldsymbol{x}}\|_2}{\|\hat{\boldsymbol{x}}\|_2} \right) \text{ [dB]}.$$

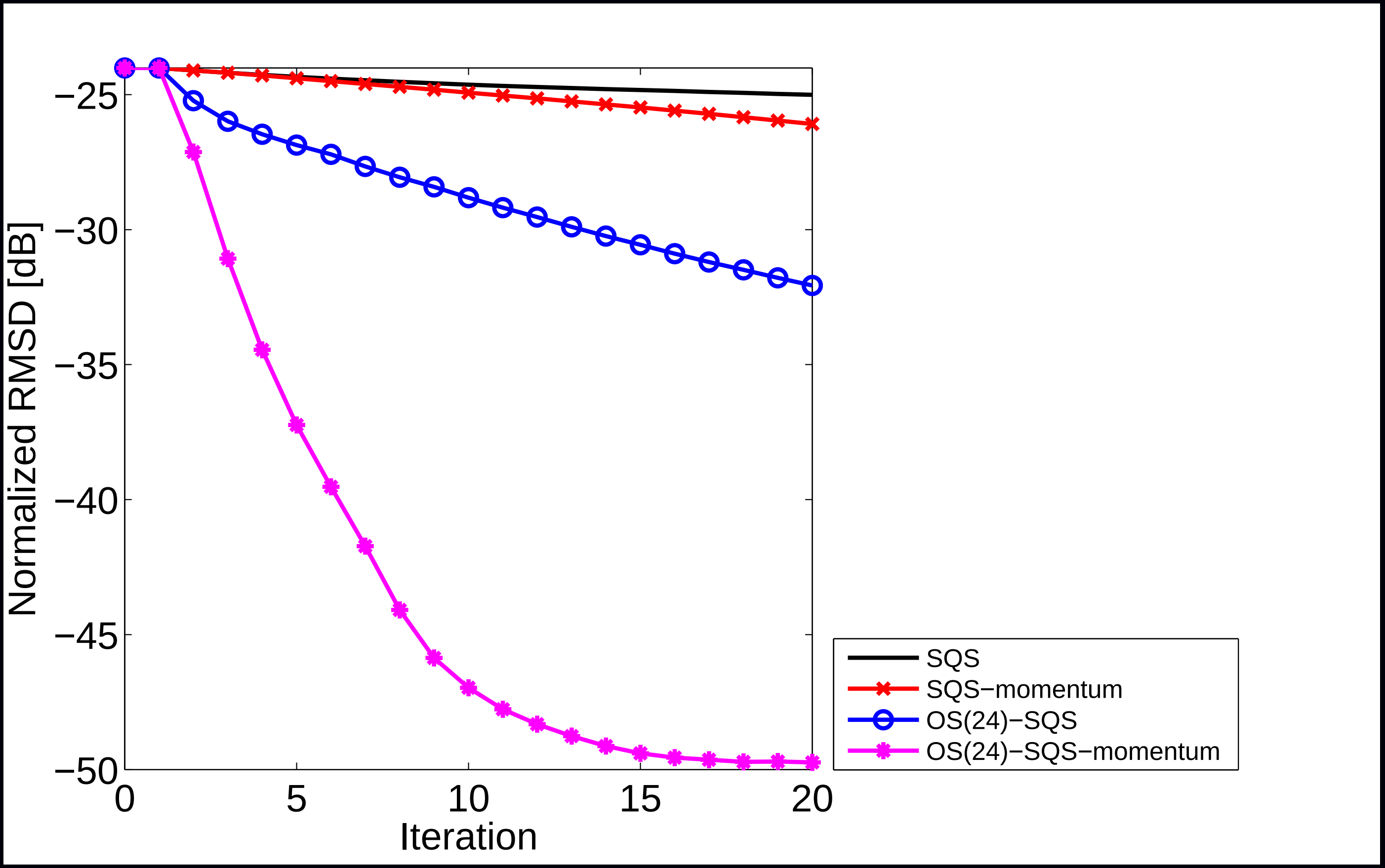
- $\hat{\boldsymbol{x}}$ obtained by *many* iterations of several convergent algorithms

Convergence rate: RMSD [HU]



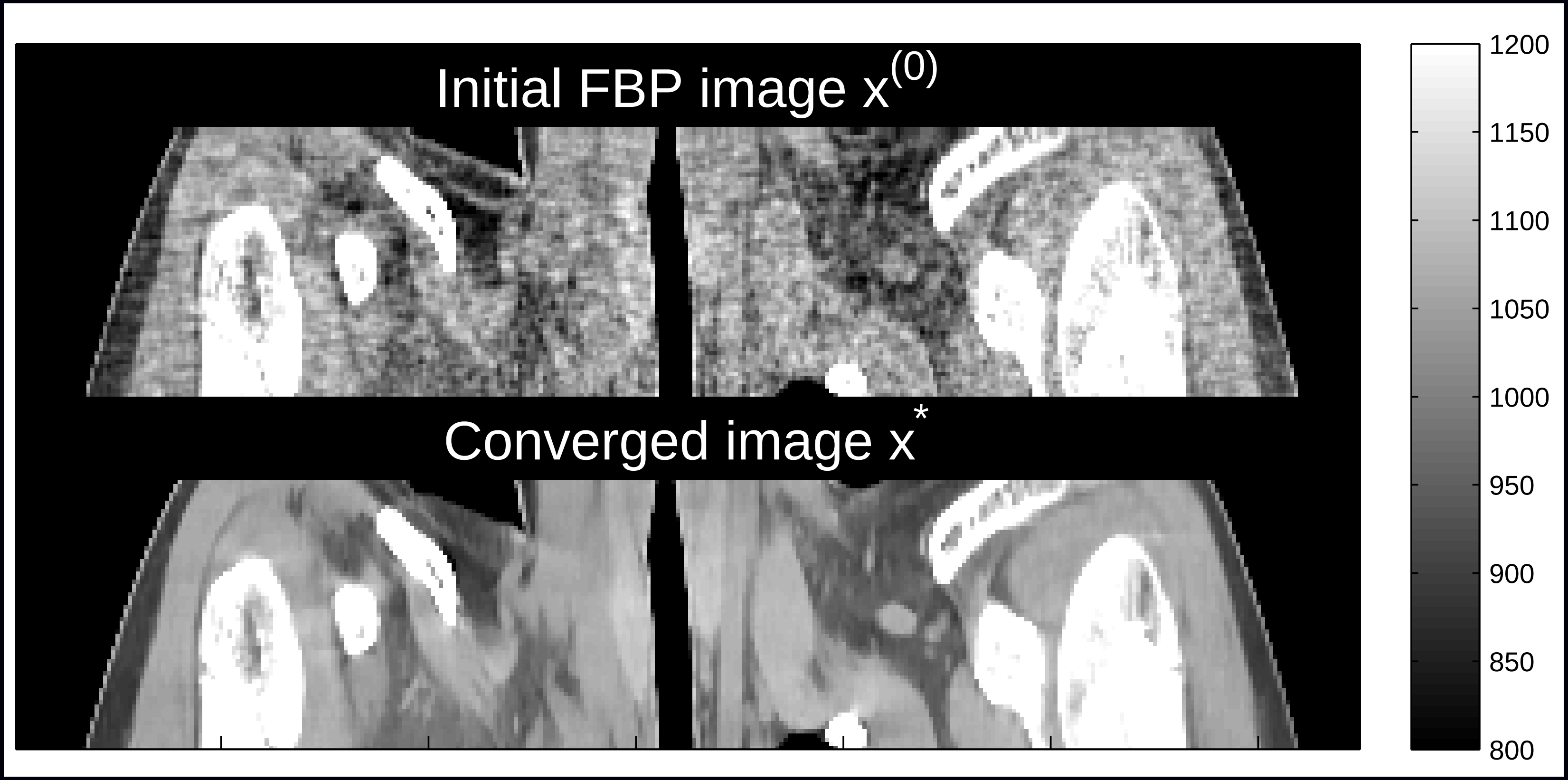
- Slow convergence without OS methods.
- OS methods with $M = 24$ subsets needed 20% extra compute time per iteration due to $\nabla\,\mathrm{R}(\boldsymbol{x})$.
- OS-SQS-Momentum “converges” very rapidly in early iterations!
- Does not reach RMSD=0...

Convergence rate: Normalized RMSD [dB]

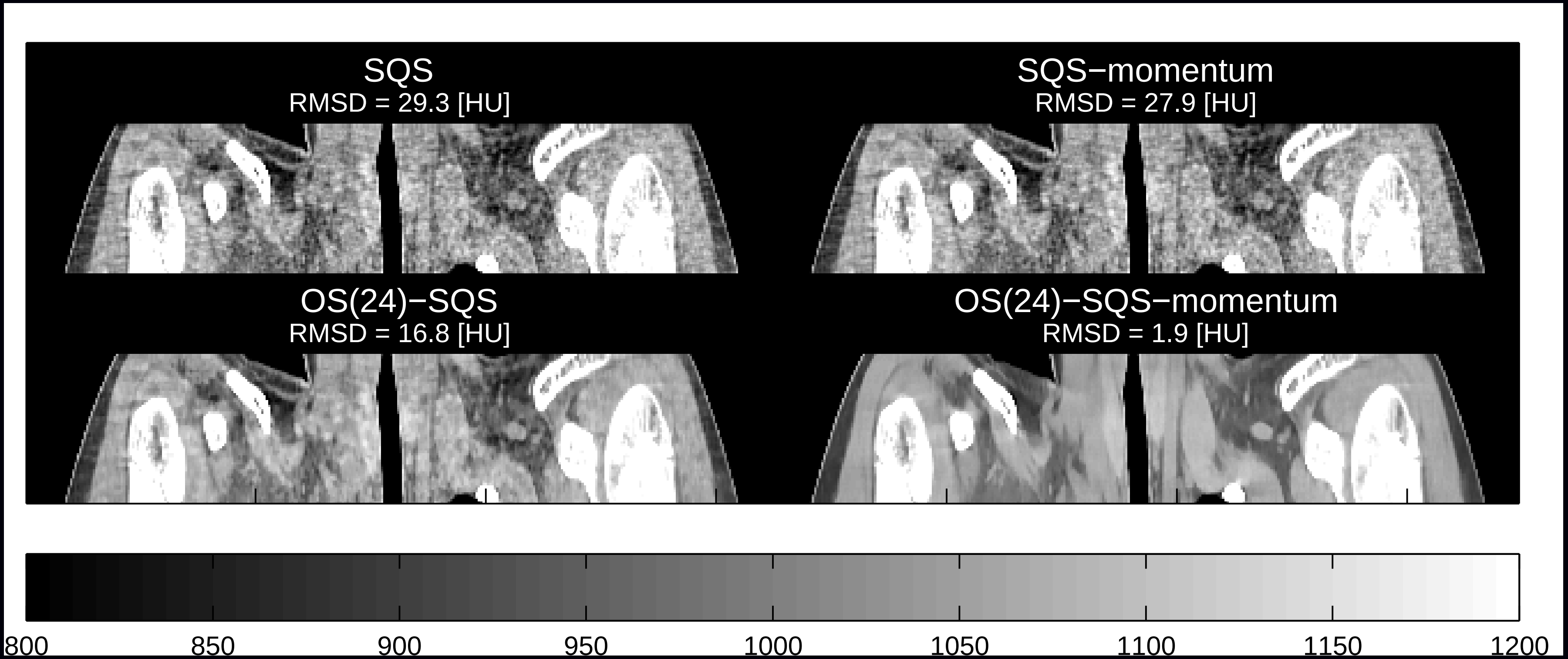


Combining incremental (sub)gradient with Nesterov-type momentum acceleration may help other "big data" estimation problems.

Images


Initial FBP image $x^{(0)}$
Converged image x^*
1200
1150
1100
1050
1000
950
900
850
800

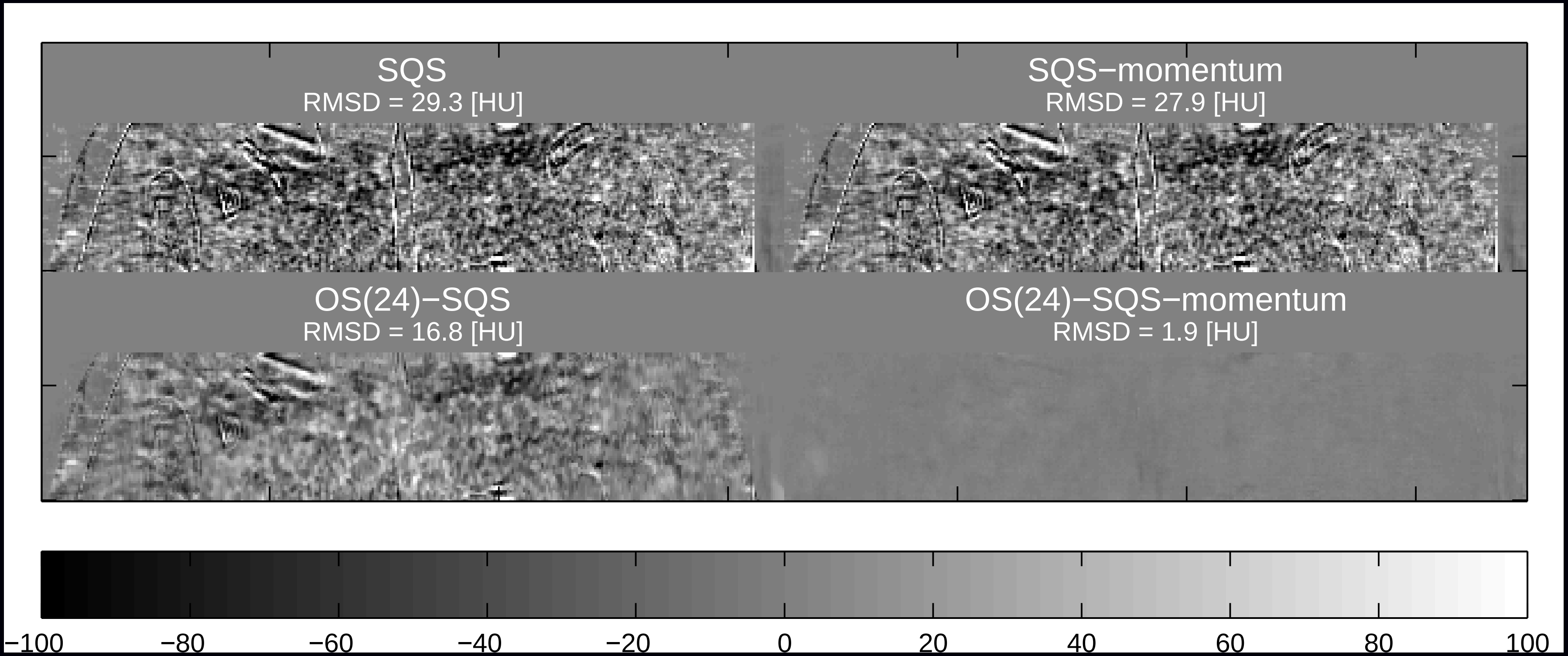
Images



Reconstructed images at 12th iteration. ([800 1200] HU)

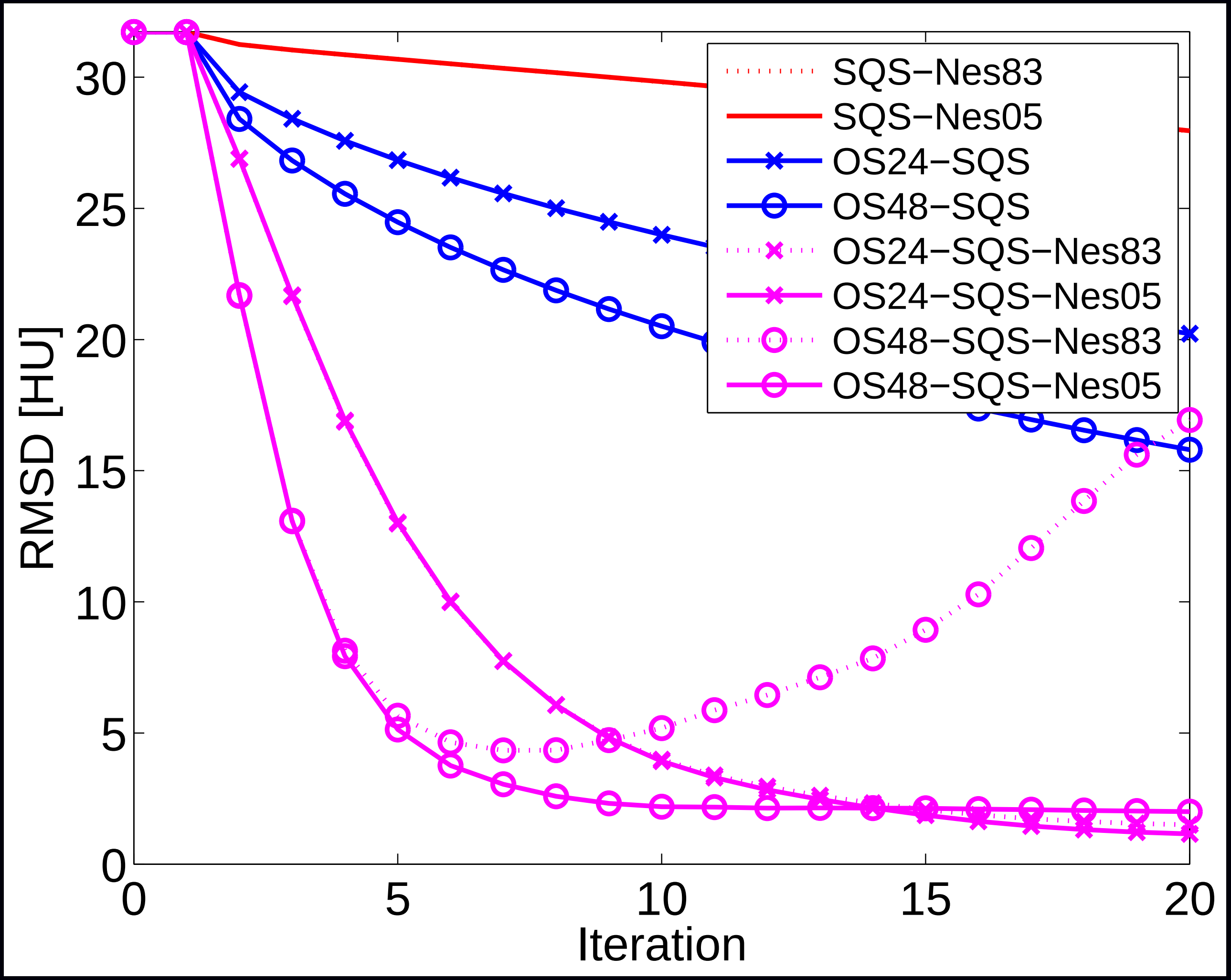
OS-SQS+Momentum with $M = 24$ subsets much closer to minimizer $\hat{\boldsymbol{x}}$

Difference images



Difference between reconstructed images at 12th iteration and converged image $\hat{x}$. ([-100 100] HU)

Newer Nesterov method



Remains stable even for $M = 48$ subsets. (Nesterov, Math. Prog., May 2005)

Some research problems in CT

$$\hat{\boldsymbol{x}} = \operatorname*{arg\,min}_{\boldsymbol{x} \succeq \boldsymbol{0}} \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_{\boldsymbol{W}}^2 + \beta \mathsf{R}(\boldsymbol{x}).$$

Reasonably mature research areas

- Design and implementation of system model $\boldsymbol{A}$
- Statistical modeling $\boldsymbol{W}$

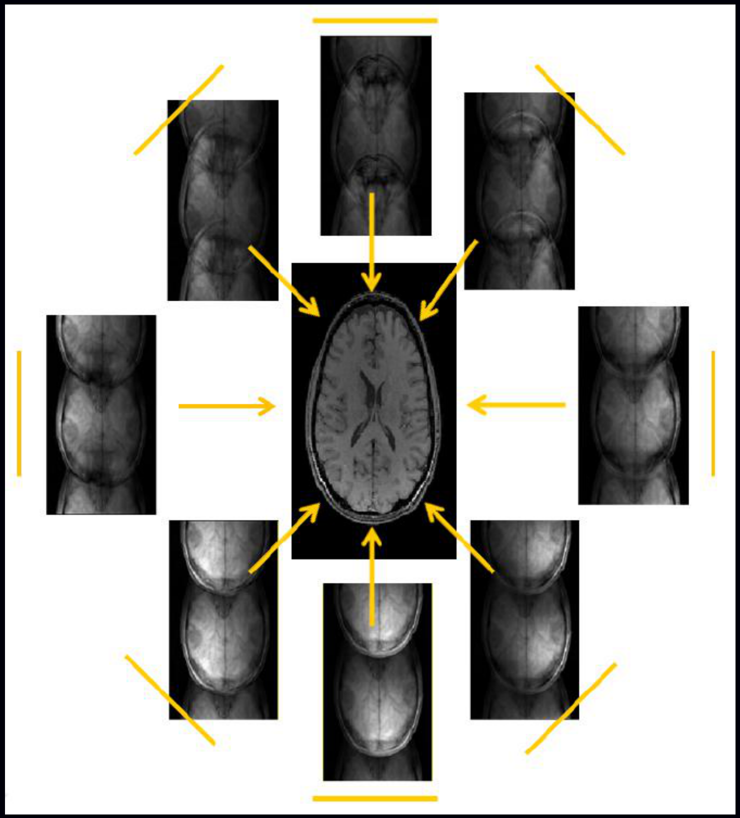
Open problems

- Design of regularizer $\mathsf{R}(\boldsymbol{x})$ to maximize radiologist performance
- Faster parallelizable algorithms (argmin) with global convergence
- Distributed computation – reducing communication
- Algorithms for more complete/complicated physical models (*e.g.*, dual energy or spectral CT)
- Dynamic imaging / motion compensated image reconstruction
- Analysis of statistical properties of (highly nonlinear) estimator $\hat{\boldsymbol{x}}$

Image reconstruction for parallel MRI

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

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

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 $\|\cdot\|_p$
- Complex quantities
- Large problem size (if 3D)

Basic ADMM for parallel MRI

Basic equivalent constrained optimization problem (*cf.* split Bregman):

$$\min_{\boldsymbol{x},\boldsymbol{v}} \frac{1}{2}\|\boldsymbol{y}-\boldsymbol{F}\boldsymbol{S}\boldsymbol{x}\|_2^2+\beta\|\boldsymbol{v}\|_p \text{ sub. to } \boldsymbol{v}=\boldsymbol{R}\boldsymbol{x}.$$

Corresponding (modified) augmented Lagrangian (*cf.* “split Bregman”):

$$L(\boldsymbol{x},\boldsymbol{v};\boldsymbol{\eta})=\frac{1}{2}\|\boldsymbol{y}-\boldsymbol{F}\boldsymbol{S}\boldsymbol{x}\|_2^2+\beta\|\boldsymbol{v}\|_p+\frac{\rho}{2}\|\boldsymbol{R}\boldsymbol{x}-\boldsymbol{v}-\boldsymbol{\eta}\|_2^2$$

(Skipping technical details about complex vectors.)

ADMM update of primal variable (unknown image):

$$\boldsymbol{x}^{(n+1)}=\arg\min_{\boldsymbol{x}} L(\boldsymbol{x},\boldsymbol{v}^{(n)},\boldsymbol{\eta}^{(n)})=\left[\boldsymbol{S}'\boldsymbol{F}'\boldsymbol{F}\boldsymbol{S}+\rho\boldsymbol{R}'\boldsymbol{R}\right]^{-1}\left(\boldsymbol{S}'\boldsymbol{F}'\boldsymbol{y}+\rho\boldsymbol{R}'\left(\boldsymbol{v}^{(n)}+\boldsymbol{\eta}^{(n)}\right)\right)$$

- $\left[\boldsymbol{S}'\boldsymbol{F}'\boldsymbol{F}\boldsymbol{S}+\rho\boldsymbol{R}'\boldsymbol{R}\right]^{-1}$ requires iteration (*e.g.*, PCG) but hard to precondition
- (Trivial for single coil case with $\boldsymbol{S}=\boldsymbol{I}$.)
- The “problem” matrix is on opposite side:
  - MRI: $\boldsymbol{F}\boldsymbol{S}$
  - Restoration: $\boldsymbol{T}\boldsymbol{A}$

Improved ADMM for parallel MRI

$$\min_{x,u,v,z} \frac{1}{2}\|y - F\,u\|_2^2 + \beta\,\|v\|_p \text{ sub. to } v = Rz, \quad u = Sx, \quad z = x.$$

Corresponding (modified) augmented Lagrangian:

$$\frac{1}{2}\|y - F\,u\|_2^2 + \beta\,\|v\|_p + \frac{\rho_1}{2}\|Rz - v - \eta_1\|_2^2 + \frac{\rho_2}{2}\|Sx - u - \eta_2\|_2^2 + \frac{\rho_3}{2}\|x - z - \eta_3\|_2^2$$

ADMM update of primal variable

$$\arg\min_x L(x, u, v, z; \eta_1, \eta_2, \eta_3) = \underbrace{\left[\rho_2 S'S + \rho_3 I\right]^{-1}}_{\text{diagonal}} \left(\rho_2 S'(u + \eta_2) + \rho_3(z + \eta_3)\right)$$

ADMM update of auxiliary variables:

$$\arg\min_u L(x, u, v, z; \eta_1, \eta_2, \eta_3) = \underbrace{\left[F'F + \rho_2 I\right]^{-1}}_{\text{circulant}} (F'y + \rho_2(Sx - \eta_2))$$

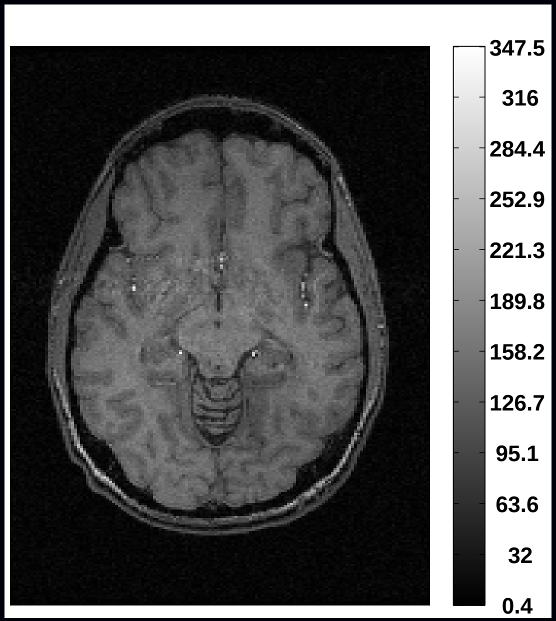
$$\arg\min_z L(x, u, v, z; \eta_1, \eta_2, \eta_3) = \underbrace{\left[\rho_1 R'R + \rho_3 I\right]^{-1}}_{\text{circulant}} (\rho_1 R'(v + \eta_1) + \rho_3(x - \eta_3))$$

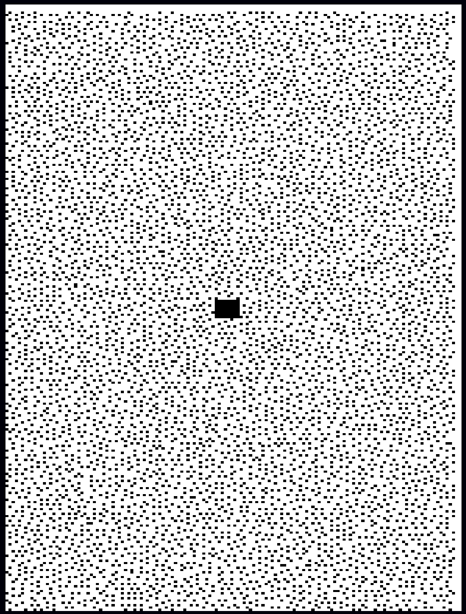
v update is shrinkage again.

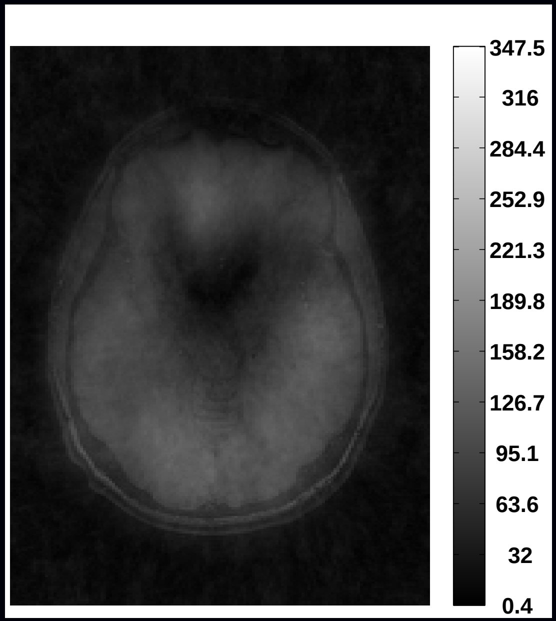
Simple, but does not satisfy sufficient conditions.

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

2.5D parallel MR image reconstruction results: data

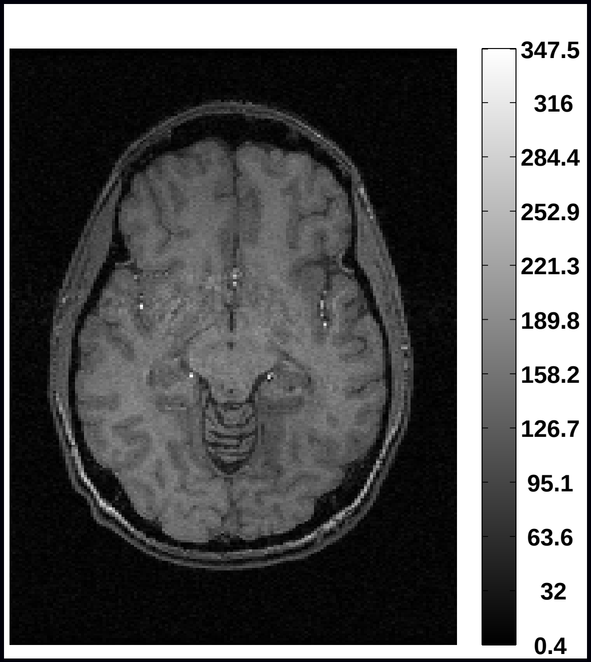


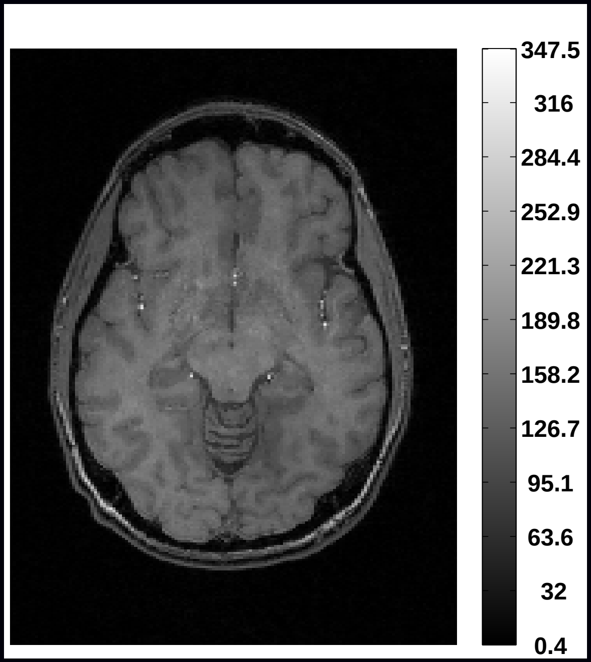


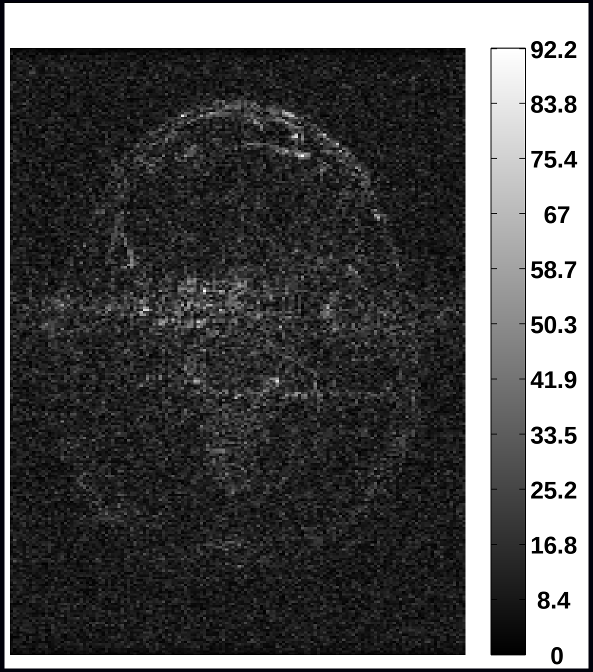


Fully sampled body coil image of human brain
Poisson-disk-based k-space sampling, 16% sampling (acceleration 6.25)
Square-root of sum-of-squares inverse FFT of zero-filled k-space data

2.5D parallel MR image reconstruction results: IQ

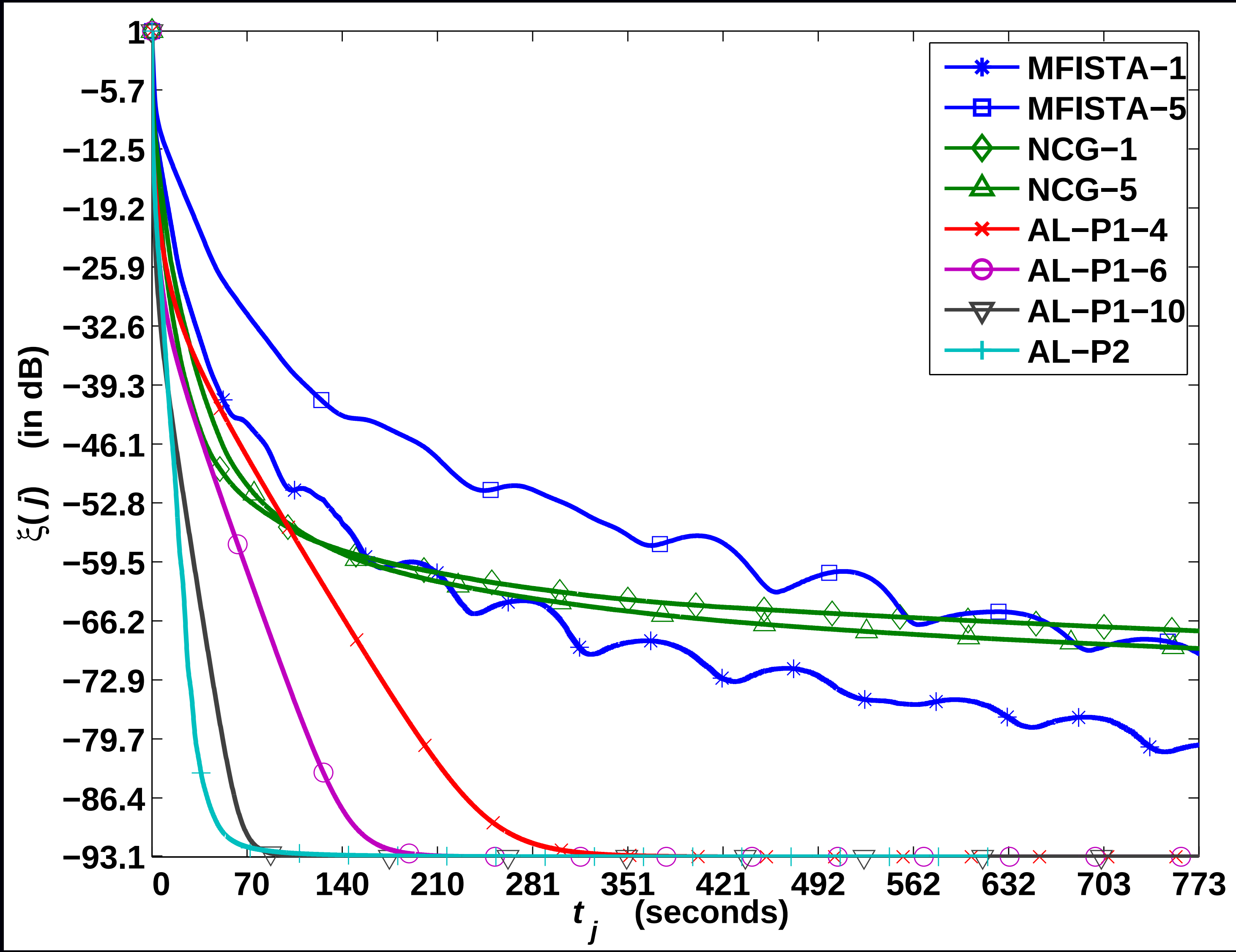






- Fully sampled body coil image of human brain
- Regularized reconstruction $\boldsymbol{x}^{(\infty)}$ (1000s of iterations of MFISTA)
  (A Beck & M Teboulle, SIAM J. Im. Sci, 2009)
  Combined TV and ℓ_1 norm of two-level undecimated Haar wavelets
- Difference image magnitude

2.5D parallel MR image reconstruction results: speed



AL approach converges to $\boldsymbol{x}^{(\infty)}$ much faster than MFISTA and CG

Current and future directions with ADMM

- Motion-compensated image reconstruction: $y = AT(\alpha)x + \varepsilon$
  (J H Cho, S Ramani, JF, 2nd CT meeting, 2012)
  (J H Cho, S Ramani, JF, IEEE Stat. Sig. Proc. W., 2012)
- Dynamic image reconstruction
- Improved preconditioners for ADMM for 3D CT
  (M McGaffin and JF, Submitted to Fully 3D 2013)
- Combining ADMM with ordered subsets (OS) methods
  (H Nien and JF, Submitted to Fully 3D 2013)
- Generalize parallel MRI algorithm to include spatial support constraint
  (M Le, S Ramani, JF, To appear at ISMRM 2013)
- Non-Cartesian MRI (combine optimization transfer and variable splitting)
  (S Ramani and JF, ISBI 2013, to appear.)
- SPECT-CT reconstruction with non-local means regularizer
  (S Y Chun, Y K Dewaraja, JF, Submitted to Fully 3D 2013)
- Estimation of coil sensitivity maps (quadratic problem!)
  (M J Allison, S Ramani, JF, IEEE T-MI, Mar. 2013)
- L1-SPIRiT for non-Cartesian parallel MRI (D S Weller, S Ramani, JF, IEEE T-MI, 2013, submitted)
- Multi-frame super-resolution
- Selection of AL penalty parameter ρ to optimize convergence rate

- Other non-ADMM methods...

Acknowledgements

CT group

- Jang Hwan Cho
- Donghwan Kim
- Jungkuk Kim
- Madison McGaffin
- Hung Nien
- Stephen Schmitt

MR group

- Michael Allison
- Mai Le
- Antonis Matakos
- Matthew Muckley

Post-doctoral fellows

- Se Young Chun
- Sathish Ramani
- Daniel Weller

UM collaborators

- Doug Noll
- Jon-Fredrik Nielsen
- Mitch Goodsitt
- Ella Kazerooni
- Tom Chenevert
- Charles Meyer

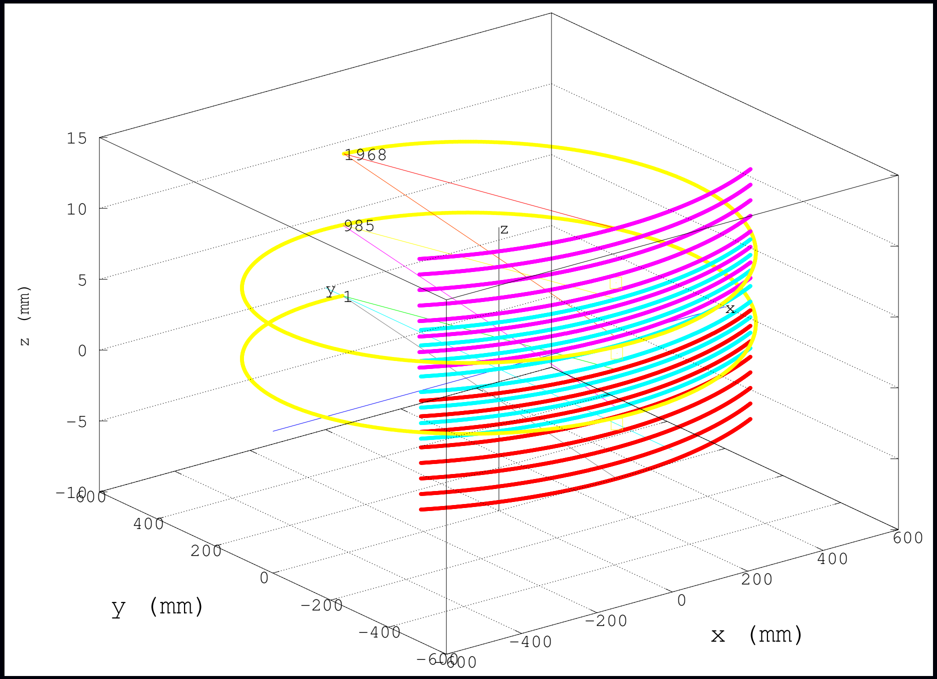
GE collaborators

- Bruno De Man
- Jean-Baptiste Thibault

Image reconstruction toolbox

MATLAB (and increasingly Octave) toolbox for imaging inverse problems (MRI, CT, PET, SPECT, Deblurring)

`web.eecs.umich.edu/~fessler`



Bibliography

[1] A. Matakos, S. Ramani, and J. A. Fessler. Accelerated edge-preserving image restoration without boundary artifacts. *IEEE Trans. Im. Proc.*, 22(5):2019–29, May 2013.

[2] S. Ramani and J. A. Fessler. A splitting-based iterative algorithm for accelerated statistical X-ray CT reconstruction. *IEEE Trans. Med. Imag.*, 31(3):677–88, March 2012.

[3] D. Kim, S. Ramani, and J. A. Fessler. Accelerating X-ray CT ordered subsets image reconstruction with Nesterov's first-order methods. In *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, pages 22–5, 2013.

[4] D. Kim, S. Ramani, and J. A. Fessler. Ordered subsets with momentum for accelerated X-ray CT image reconstruction. In *Proc. IEEE Conf. Acoust. Speech Sig. Proc.*, pages 920–3, 2013.

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

[6] M. Elad, P. Milanfar, and R. Rubinstein. Analysis versus synthesis in signal priors. *Inverse Prob.*, 23(3):947–68, June 2007.

[7] I. W. Selesnick and Mário A T Figueiredo. Signal restoration with overcomplete wavelet transforms: comparison of analysis and synthesis priors. In *Proc. SPIE 7446 Wavelets XIII*, page 74460D, 2009. Wavelets XIII.

[8] Y. Wang, J. Yang, W. Yin, and Y. Zhang. A new alternating minimization algorithm for total variation image reconstruction. *SIAM J. Imaging Sci.*, 1(3):248–72, 2008.

[9] M. V. Afonso, José M Bioucas-Dias, and Mário A T Figueiredo. Fast image recovery using variable splitting and constrained optimization. *IEEE Trans. Im. Proc.*, 19(9):2345–56, September 2010.

[10] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. & Trends in Machine Learning*, 3(1):1–122, 2010.

[11] J. Eckstein and D. P. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55(1-3):293–318, April 1992.

[12] J. Douglas and H. H. Rachford. On the numerical solution of heat conduction problems in two and three space variables. *tams*, 82(2):421–39, July 1956.

[13] T. Goldstein and S. Osher. The split Bregman method for L1-regularized problems. *SIAM J. Imaging Sci.*, 2(2):323–43, 2009.

[14] S. J. Reeves. Fast image restoration without boundary artifacts. *IEEE Trans. Im. Proc.*, 14(10):1448–53, October 2005.

[15] M. G. McGaffin, S. Ramani, and J. A. Fessler. Reduced memory augmented Lagrangian algorithm for 3D iterative X-ray CT image reconstruction. In *Proc. SPIE 8313 Medical Imaging 2012: Phys. Med. Im.*, page 831327, 2012.

[16] H. Erdoğan and J. A. Fessler. Ordered subsets algorithms for transmission tomography. *Phys. Med. Biol.*, 44(11):2835–51, November 1999.

[17] 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.

[18] Z. Q. Luo. On the convergence of the LMS algorithm with adaptive learning rate for linear feedforward networks. *Neural Computation*, 32(2):226–45, June 1991.

[19] Y. Nesterov. Smooth minimization of non-smooth functions. *Mathematical Programming*, 103(1):127–52, May 2005.

[20] 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.
[21] 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.
[22] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.*, 2(1):183–202, 2009.
[23] M. J. Allison, S. Ramani, and J. A. Fessler. Accelerated regularized estimation of MR coil sensitivities using augmented Lagrangian methods. *IEEE Trans. Med. Imag.*, 32(3):556–64, March 2013.
[24] M. J. Allison and J. A. Fessler. Accelerated computation of regularized field map estimates. In *Proc. Intl. Soc. Mag. Res. Med.*, page 0413, 2012.
[25] J. H. Cho, S. Ramani, and J. A. Fessler. Motion-compensated image reconstruction with alternating minimization. In *Proc. 2nd Intl. Mtg. on image formation in X-ray CT*, pages 330–3, 2012.
[26] J. H. Cho, S. Ramani, and J. A. Fessler. Alternating minimization approach for multi-frame image reconstruction. In *IEEE Workshop on Statistical Signal Processing*, pages 225–8, 2012.
[27] M. McGaffin and J. A. Fessler. Sparse shift-varying FIR preconditioners for fast volume denoising. In *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, pages 284–7, 2013.
[28] H. Nien and J. A. Fessler. Combining augmented Lagrangian method with ordered subsets for X-ray CT image reconstruction. In *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, pages 280–3, 2013.
[29] M. Le, S. Ramani, and J. A. Fessler. An efficient variable splitting based algorithm for regularized SENSE reconstruction with support constraint. In *Proc. Intl. Soc. Mag. Res. Med.*, page 2654, 2013. To appear.
[30] S. Ramani and J. A. Fessler. Accelerated non-Cartesian SENSE reconstruction using a majorize-minimize algorithm combining variable-splitting. In *Proc. IEEE Intl. Symp. Biomed. Imag.*, pages 700–3, 2013.
[31] S. Y. Chun, Y. K. Dewaraja, and J. A. Fessler. Alternating direction method of multiplier for emission tomography with non-local regularizers. In *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, pages 62–5, 2013.
[32] D. Weller, S. Ramani, and J. A. Fessler. Augmented Lagrangian with variable splitting for faster non-Cartesian L1-SPIRiT MR image reconstruction. *IEEE Trans. Med. Imag.*, 2013. Submitted.
[33] J. A. Fessler and W. L. Rogers. Spatial resolution properties of penalized-likelihood image reconstruction methods: Space-invariant tomographs. *IEEE Trans. Im. Proc.*, 5(9):1346–58, September 1996.