Axial block coordinate descent (ABCD) algorithm for X-ray CT image reconstruction

Jeffrey A. Fessler and Donghwan Kim

EECS Department
University of Michigan

Fully 3D Image Reconstruction Conference

July 13, 2011

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

Full disclosure

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

Goal:

Faster iterative (fully statistical) 3D CT reconstruction

Thin-slice FBP ASIR Statistical

Cost function

Penalized weighted least-squares (PWLS):

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \Psi(\boldsymbol{x}), \; \Psi(\boldsymbol{x}) = \sum_{i=1}^{M} \frac{w_i}{2}(y_i - [\boldsymbol{A}\boldsymbol{x}]_i)^2 + \mathsf{R}(\boldsymbol{x})$$

- unknown 3D image $\boldsymbol{x} = (x_1, \ldots, x_N)$ with N voxels
- $\boldsymbol{y} = (y_1, \ldots, y_M)$ CT (log) projection data with M rays
- w_i statistical weighting for ith ray, $i = 1, \ldots, M$
- $\boldsymbol{A}$: $M \times N$ system matrix
- $\mathsf{R}(\boldsymbol{x})$: edge-preserving regularizer
- forward projector : $[\boldsymbol{A}\boldsymbol{x}]_i = \sum_{j=1}^{N} a_{ij} x_j$.

The principles generalize readily to other statistical models.

Traditional iterative minimization algorithms

- Iterative coordinate descent (ICD)
  Sauer & Bouman, 1993; Thibault *et al.*, 2007
  + few iterations
  - challenging to parallelize because sequential
- Preconditioned conjugate gradient (PCG)
  + simultaneous update of all voxels using all views
  - more iterations
  - challenging to precondition effectively for 3D WLS
  - challenging to precondition effectively for nonquadratic $\mathsf{R}(\boldsymbol{x})$
  Fessler & Booth, 1999
- Ordered-subsets (OS) based on separable quadratic surrogates (SQS)
  Kamphuis & Beekman, 1998, Erdoğan & Fessler, 1999
  + update all pixels simultaneously using some views
  - regularizer gradient $\nabla \mathsf{R}(\boldsymbol{x})$ for every block of views
  - does not converge, worsening for large number of subsets
  - requires many more iterations to converge than ICD
  Deman *et al.*, 2005

Update each voxel sequentially or update all voxels simultaneously?

Block coordinate descent / Grouped coordinate descent

- Update a block of voxels simultaneously.
- Loop over all blocks.

Long history in general optimization
Bertsekas, 1999, *Nonlinear programming*

Global convergence for strictly convex cost functions

Long history in general statistical estimation problems
Hathaway and Bezdek, 1991; Jensen, 1991

Applications to tomographic image reconstruction
Sauer *et al.*, 1995; Fessler *et al.*, 1995; Fessler *et al.*, 1997; Benson *et al.*, 2010

Choice of order important for fastest possible convergence
Yu *et al.*, 2011

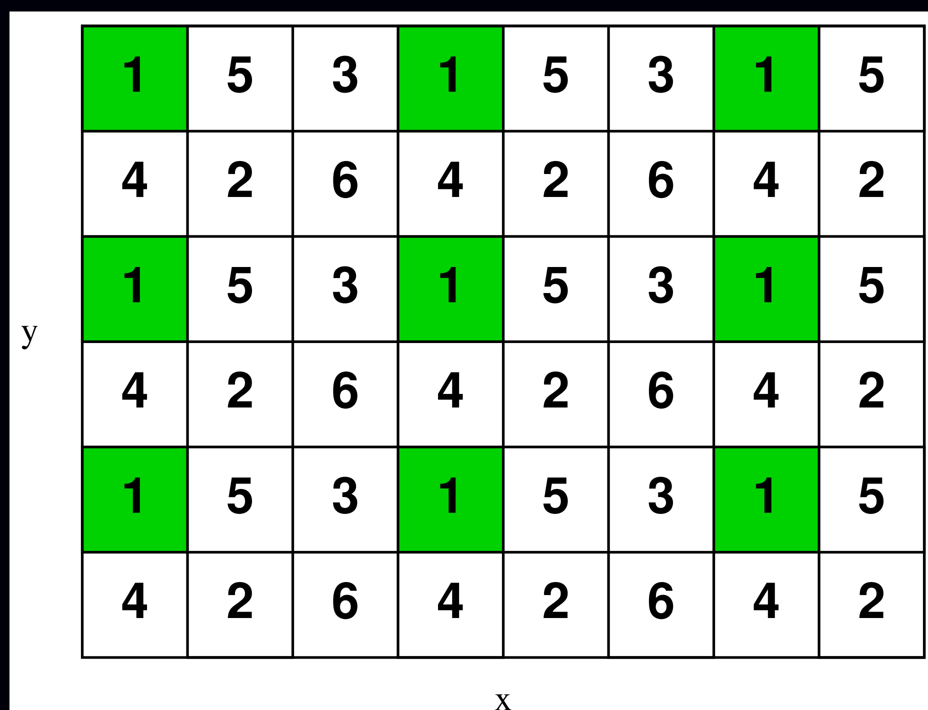
2D grouped coordinate descent

Fessler *et al.*, 1997

- Spatially separated grouped of pixels (in 2D)
- Pixels within group updated simultaneously using optimization transfer
- Moderately strong coupling of pixels within slice

  $\Longrightarrow$ undesirably high surrogate curvatures

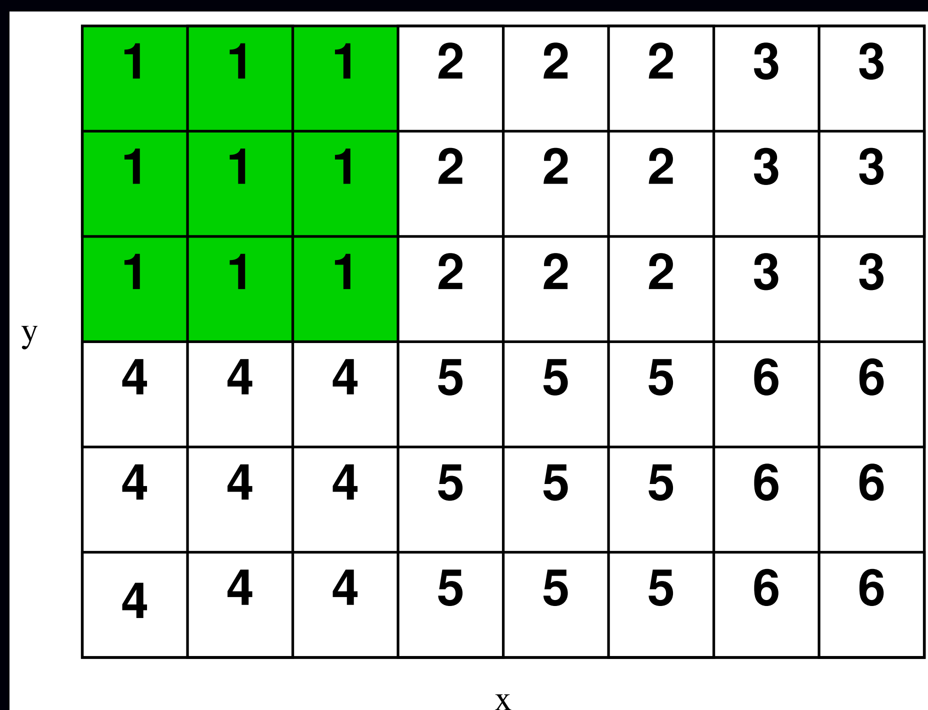
  $\Longrightarrow$ modest acceleration compared to all-voxel SPS



3D (transaxial) block coordinate descent

Benson *et al.*, 2010

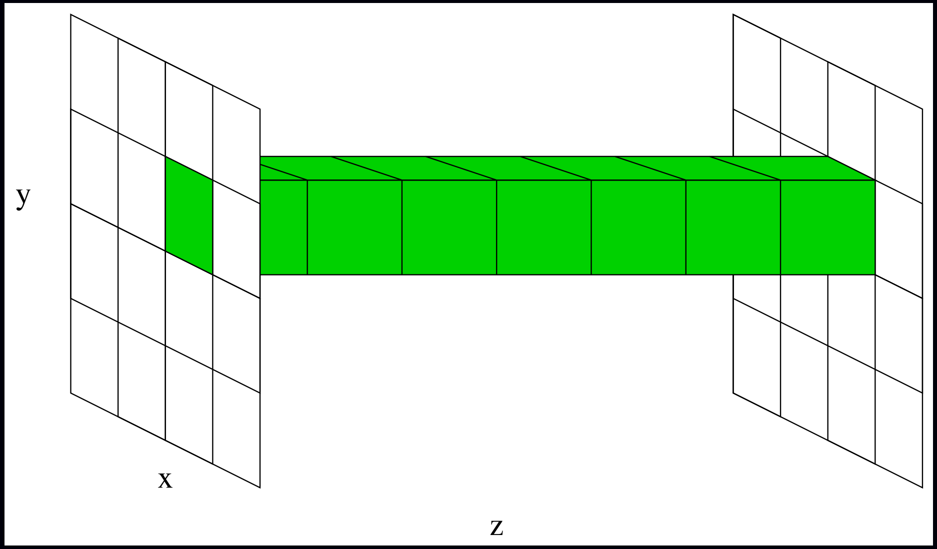
- Blocks of $k \times k$ neighboring pixels – strongly coupled
- Solved simultaneously by inverting a *dense* $k^2 \times k^2$ matrix
- Loop over z before proceeding to next transaxial block



3D axial block coordinate descent (ABCD)

Proposed approach:

- update a block of all N_z voxels along an axial line simultaneously
- loop over all x, y locations sequentially
  (possibly inhomogeneously, *cf.* Yu *et al.*, T-IP, 2011)



Axial block coordinate descent (ABCD) outline

`for` $k = 1, \ldots, K$: (K = # of x-y locations $\leq N_x N_y$)

$$\boldsymbol{x}_k^{(n+1)} = \arg\min_{\boldsymbol{x}_k \in \mathbb{R}^{N_z}} \Psi\left(\boldsymbol{x}_1^{(n+1)}, \ldots, \boldsymbol{x}_{k-1}^{(n+1)}, \boldsymbol{x}_k, \boldsymbol{x}_{k+1}^{(n)}, \ldots, \boldsymbol{x}_K^{(n)}\right).$$

`end`

If the regularizer is quadratic, then the ABCD update is simply:

$$\boldsymbol{x}_k^{(n+1)} = \boldsymbol{x}_k^{(n)} - \left[\boldsymbol{H}_k^{(n)}\right]^{-1} \nabla_{\boldsymbol{x}_k} \Psi\left(\boldsymbol{x}_1^{(n+1)}, \ldots, \boldsymbol{x}_{k-1}^{(n+1)}, \boldsymbol{x}_k, \boldsymbol{x}_{k+1}^{(n)}, \ldots, \boldsymbol{x}_K^{(n)}\right)\Big|_{\boldsymbol{x}_k = \boldsymbol{x}_k^{(n)}}.$$

Requires inverting the $N_z \times N_z$ Hessian matrix

$$\begin{aligned} \boldsymbol{H}_k^{(n)} &= \nabla^2_{\boldsymbol{x}_k} \Psi\left(\boldsymbol{x}_1^{(n+1)}, \ldots, \boldsymbol{x}_{k-1}^{(n+1)}, \boldsymbol{x}_k, \boldsymbol{x}_{k+1}^{(n)}, \ldots, \boldsymbol{x}_K^{(n)}\right)\Big|_{\boldsymbol{x}_k = \boldsymbol{x}_k^{(n)}} \\ &= \boldsymbol{A}_k' \boldsymbol{W} \boldsymbol{A}_k + \nabla^2_{\boldsymbol{x}_k} \mathsf{R}(\boldsymbol{x}) \end{aligned}$$

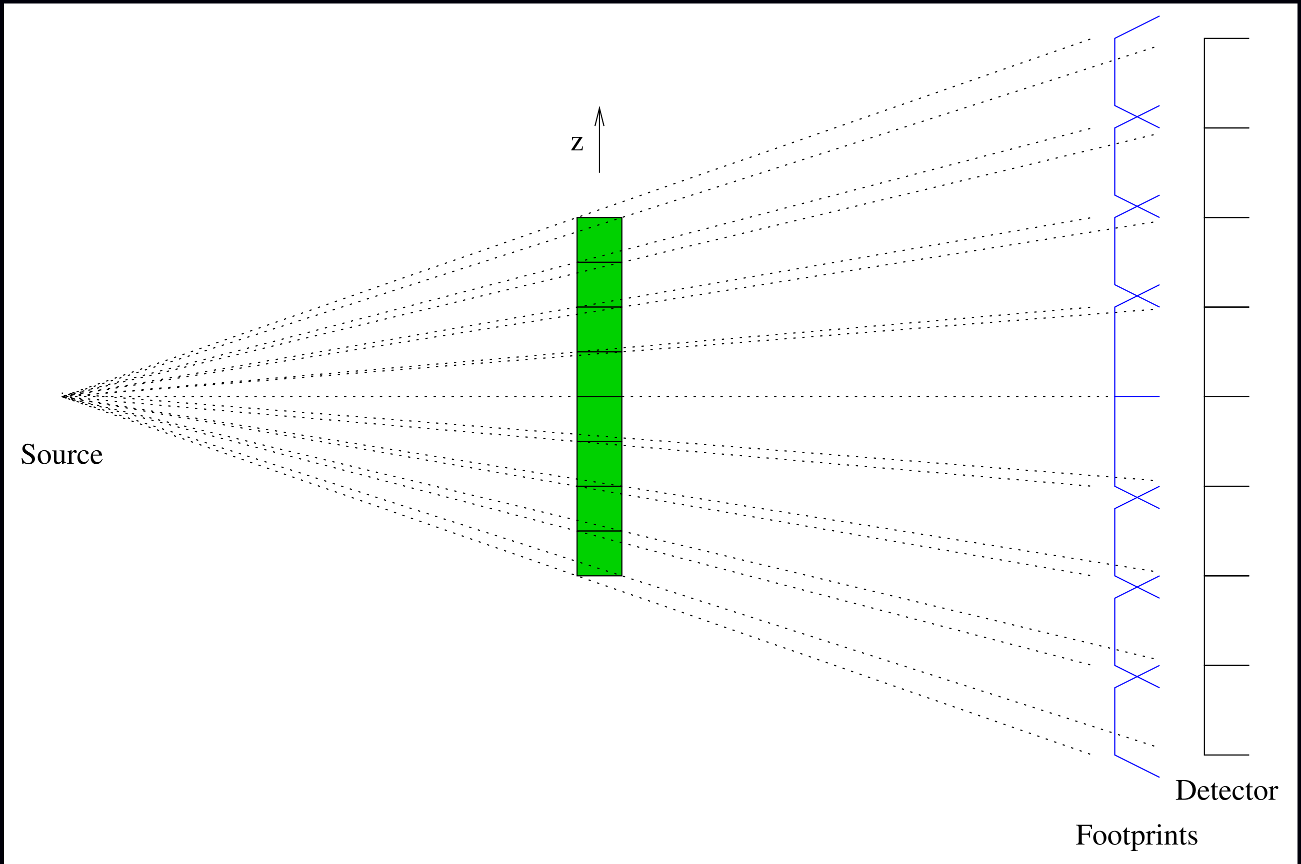
where $\boldsymbol{A}_k$ is the $M \times N_z$ submatrix of $\boldsymbol{A}$ with the columns that correspond to the voxels in the block being updated. $\boldsymbol{A} = [\boldsymbol{A}_1\ \boldsymbol{A}_2\ \ldots\ \boldsymbol{A}_K]$

(For edge-preserving case we use a *quadratic surrogate* for the regularizer.)

Axial block coordinate descent (ABCD) properties

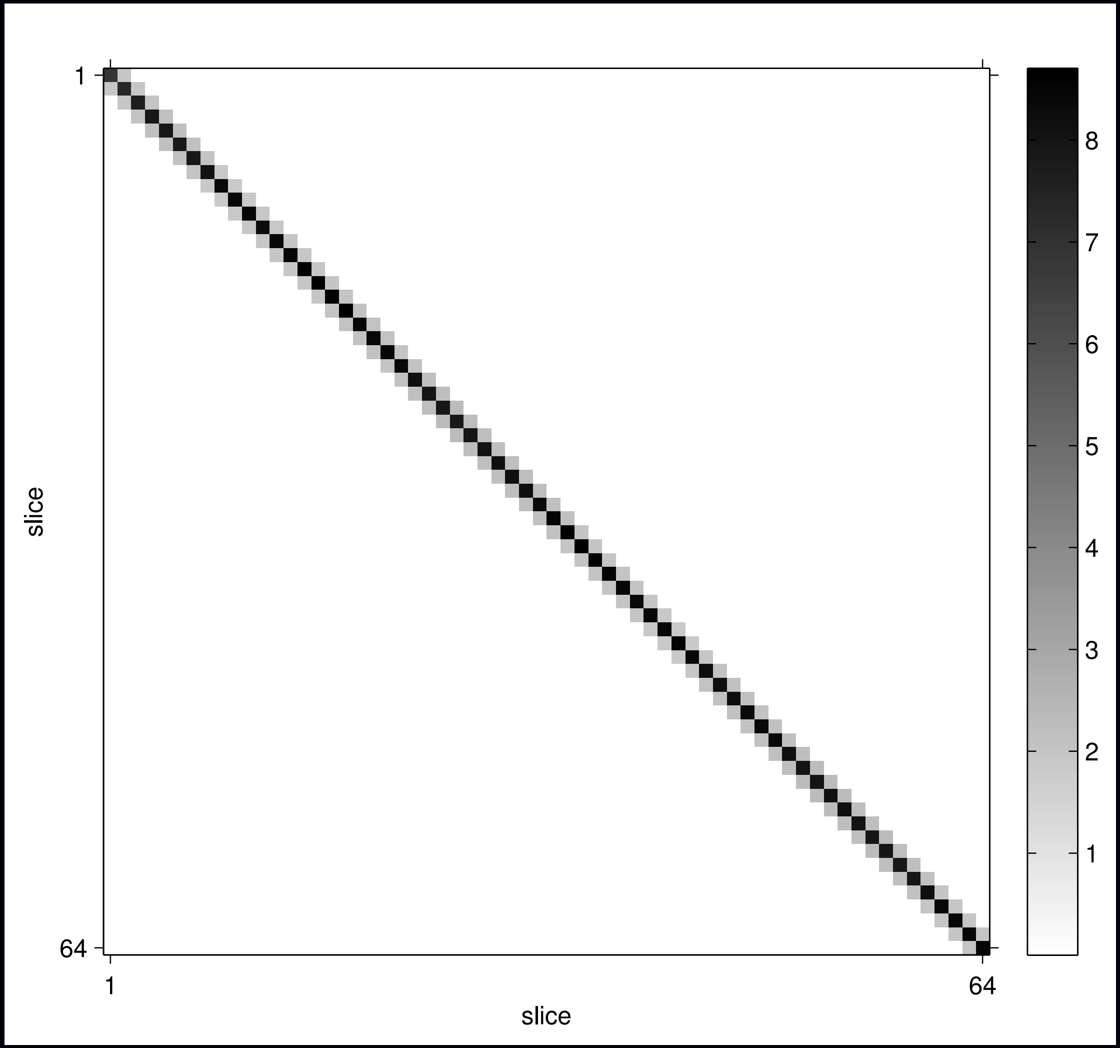
- N_z-times more parallelism opportunities than ICD
  (*e.g.*, $N_z = 64$ for axial study; $N_z = 700$ for helical scan)
- Weak coupling among voxels axially $\Longrightarrow$ reasonably fast convergence
- $N_z \times N_z$ Hessian matrix is banded; typically tri-diagonal or penta-diagonal.
  Invertible in $O(N_z)$ operations, not $O(N_z^2)$
- Particularly well suited to separable footprint (SF) projector
  Long *et al.*, 2010.
  Assumes alignment of rotation axis with detector axis (no C-arms?)
- Converges much faster than conventional optimization transfer methods based on separable quadratic surrogates [5, 16].

Axial footprint overlap



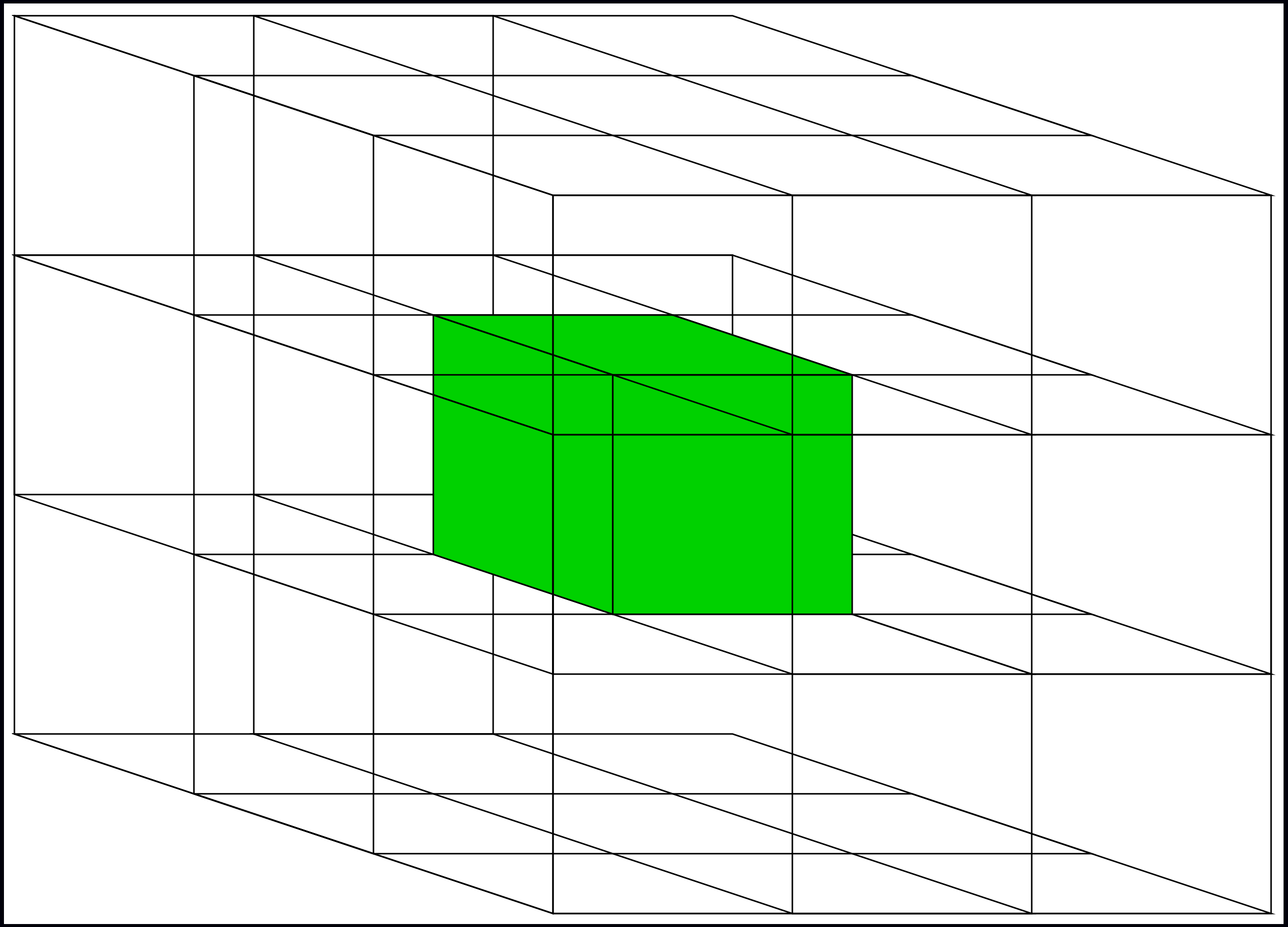
Typically the axial footprints of 2-3 voxels overlap on any given detector cell.
Amount of overlap depends on magnification factor.
The $N_z \times N_z$ Hessian matrix is banded; typically penta-diagonal.
(In contrast, for transaxial blocks the Hessian is dense.)

Banded Hessian matrix for axial block



Example for axial scan with $N_z = 64$ slices.
In contrast, for any transaxial block the Hessian is *dense*.

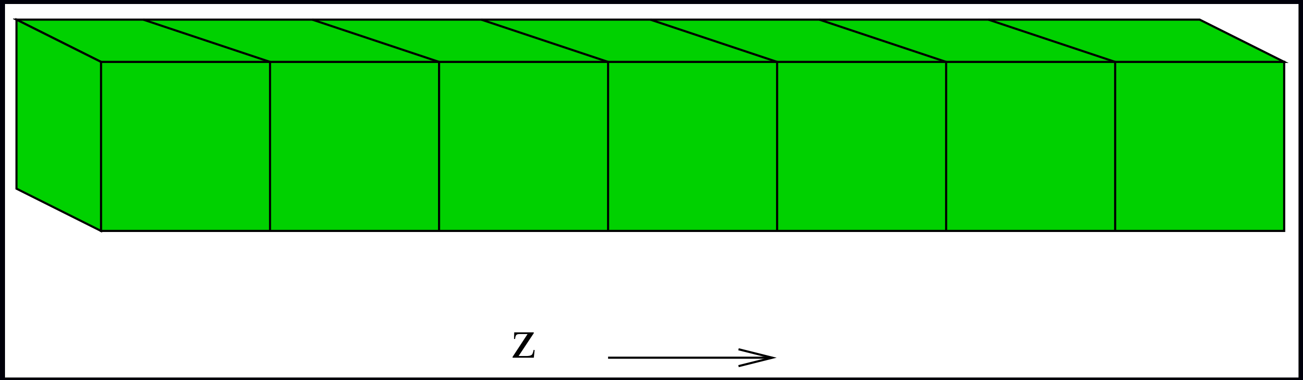
3D Regularizer



3D edge-preserving regularizer couples each voxel to 26 nearest neighbors:

$$R = \sum_{x,y,z} \sum_{j,k,l \in \{-1,1\}} \psi(f[x+j, y+k, z+l] - f[x,y,z]) \,.$$

3D Regularizer for Axial Block



3D regularizer couples each voxel in an axial block to two adjacent voxels.
(One in the slice above, one in the slice below.)
$\therefore$ The $N_z \times N_z$ Hessian of the regularizer for each axial block is tri-diagonal.

Inverting $N_z \times N_z$ penta-diagonal + tri-diagonal matrix is easy.
Easily fits in cache.

Alternatives

- Use separable quadratic surrogate (diagonal Hessian) for the axial block.
  Less work per iteration but probably more iterations.
- Use quasi-separable surrogate with tri-diagonal Hessian.
  Compromise between work per iteration and convergence rate?

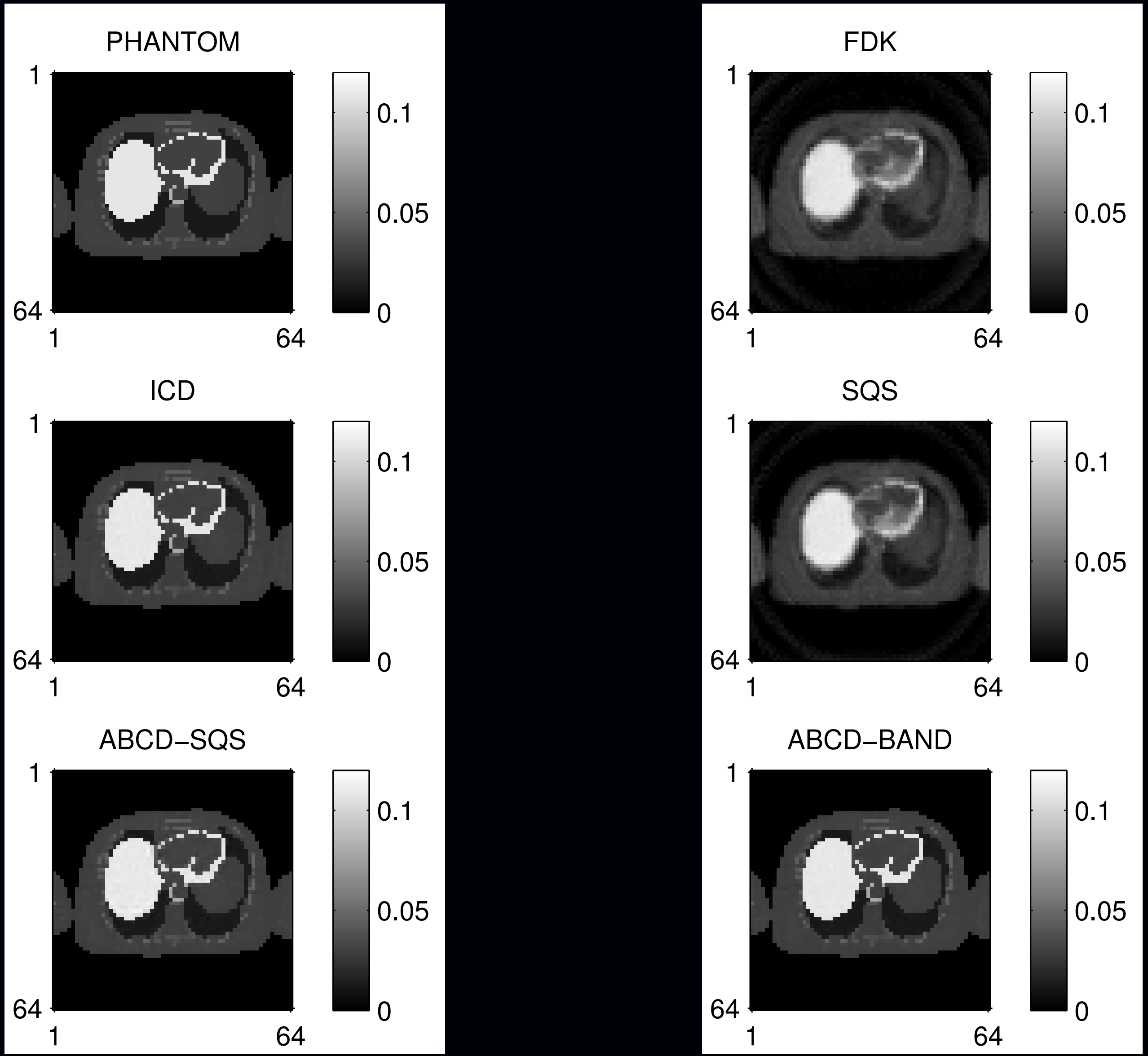
Algorithm comparison

- ICD: “blocks” with just one voxel
- ABCD-BAND: axial blocks with banded Hessian
- ABCD-SQS: axial blocks with separable quadratic surrogate (small diagonal Hessian)
- SQS: entire 3D image is one “block” with separable quadratic surrogate (large diagonal Hessian)

Expected wall time per iteration for well-parallelized implementations:

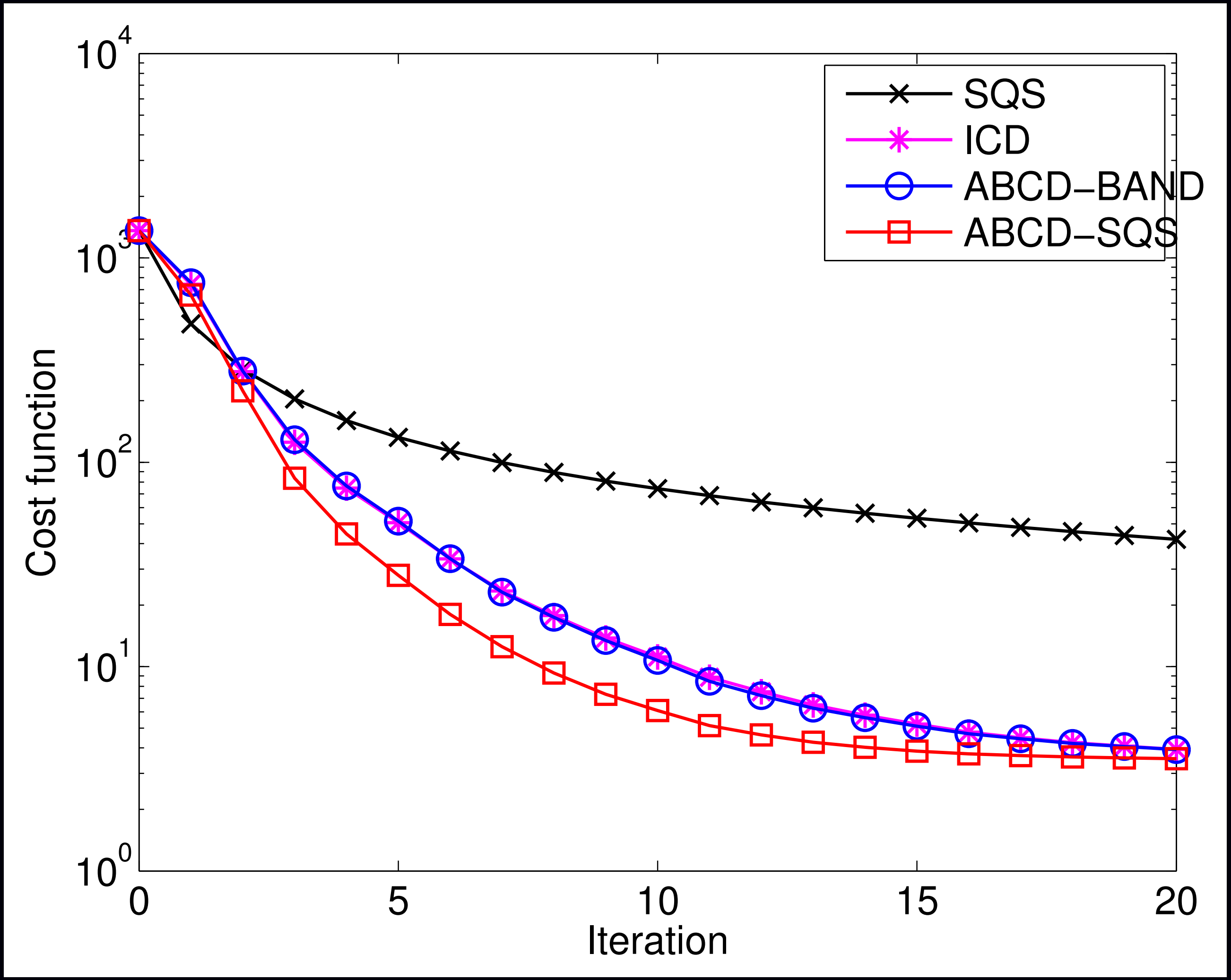
$$\text{SQS} < \text{ABCD-SQS} < \text{ABCD-BAND} < \text{ICD}$$

Matlab simulation example



Reconstructed images after 15 iterations for a small 3D problem.

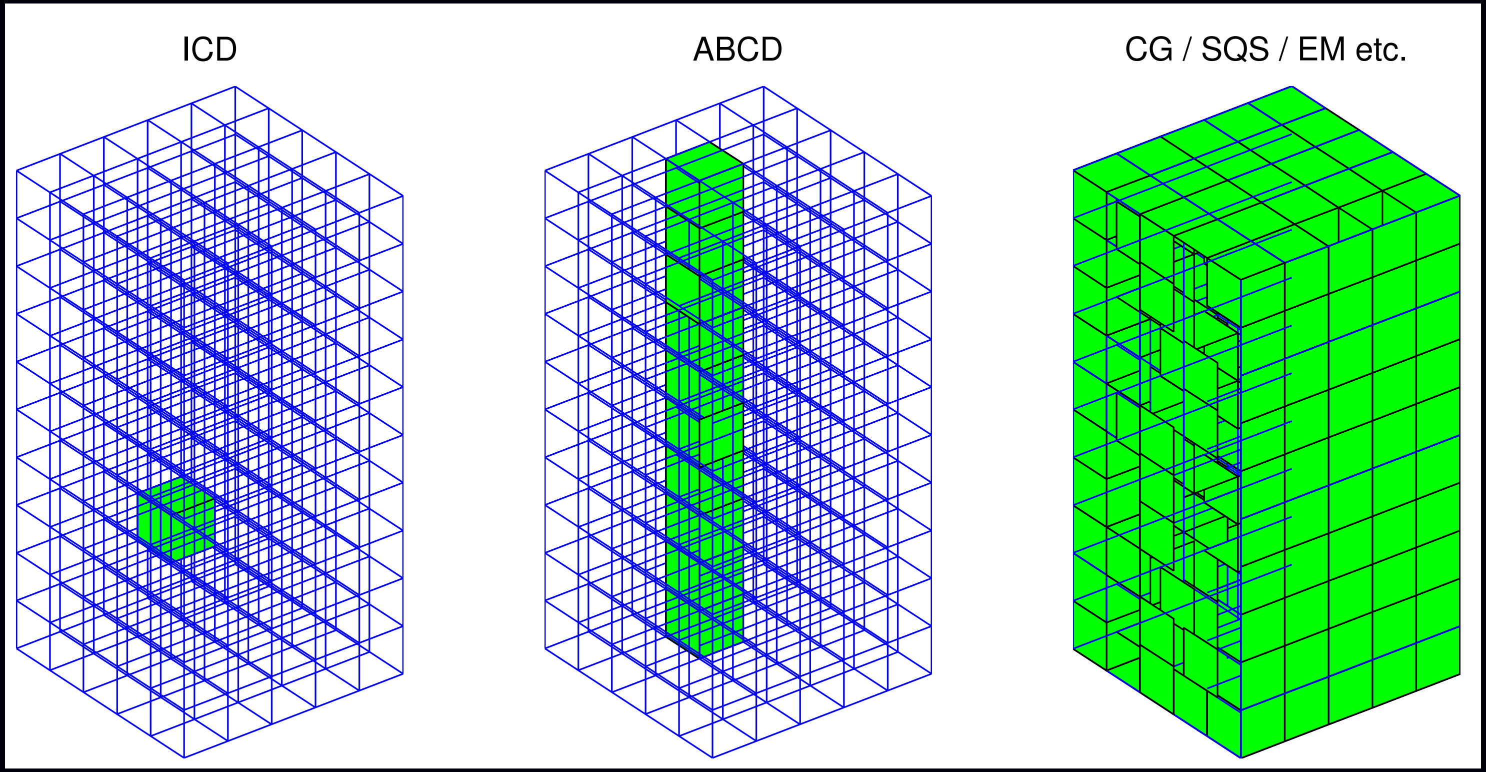
Convergence rate comparison



Cost function $\Psi(\boldsymbol{x}^{(n)})$ versus iteration n for four algorithms.

Summary

- ICD: small number of iterations but hard to parallelize
- ABCD: small number of iterations but more amenable to parallelization
- SQS: most amenable to parallelization but slowest convergence rate



Bibliography

[1] K. Sauer and C. Bouman. A local update strategy for iterative reconstruction from projections. *IEEE Trans. Sig. Proc.*, 41(2):534–48, February 1993.

[2] J-B. Thibault, K. Sauer, C. Bouman, and J. Hsieh. A three-dimensional statistical approach to improved image quality for multi-slice helical CT. *Med. Phys.*, 34(11):4526–44, November 2007.

[3] J. A. Fessler and S. D. Booth. Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction. *IEEE Trans. Im. Proc.*, 8(5):688–99, May 1999.

[4] C. Kamphuis and F. J. Beekman. Accelerated iterative transmission CT reconstruction using an ordered subsets convex algorithm. *IEEE Trans. Med. Imag.*, 17(6):1001–5, December 1998.

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

[6] B. De Man, S. Basu, J-B. Thibault, J. Hsieh, J. A. Fessler, C. Bouman, and K. Sauer. A study of different minimization approaches for iterative reconstruction in X-ray CT. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 5, pages 2708–10, 2005.

[7] R. J. Hathaway and J. C. Bezdek. Grouped coordinate minimization using Newton's method for inexact minimization in one vector coordinate. *J. Optim. Theory Appl.*, 71(3):503–16, December 1991.

[8] S. T. Jensen, S. Johansen, and S. L. Lauritzen. Globally convergent algorithms for maximizing a likelihood function. *Biometrika*, 78(4):867–77, December 1991.

[9] K. D. Sauer, S. Borman, and C. A. Bouman. Parallel computation of sequential pixel updates in statistical tomographic reconstruction. In *Proc. IEEE Intl. Conf. on Image Processing*, volume 3, pages 93–6, 1995.

[10] J. A. Fessler, E. P. Ficaro, N. H. Clinthorne, and K. Lange. Fast parallelizable algorithms for transmission image reconstruction. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 3, pages 1346–50, 1995.

[11] J. A. Fessler, E. P. Ficaro, N. H. Clinthorne, and K. Lange. Grouped-coordinate ascent algorithms for penalized-likelihood transmission image reconstruction. *IEEE Trans. Med. Imag.*, 16(2):166–75, April 1997.

[12] T. M. Benson, B. K. B. D. Man, L. Fu, and J-B. Thibault. Block-based iterative coordinate descent. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, 2010.

[13] Z. Yu, J-B. Thibault, C. A. Bouman, K. D. Sauer, and J. Hsieh. Fast model-based X-ray CT reconstruction using spatially non-homogeneous ICD optimization. *IEEE Trans. Im. Proc.*, 20(1):161–75, January 2011.

[14] G. H. Golub and C. F. Van Loan. *Matrix computations.* Johns Hopkins Univ. Press, 2 edition, 1989.

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

[16] H. Erdoğan and J. A. Fessler. Monotonic algorithms for transmission tomography. *IEEE Trans. Med. Imag.*, 18(9):801–14, September 1999.