Incremental Optimization Transfer Algorithms: Application to Transmission Tomography

Sangtae Ahn, Jeffrey A. Fessler, Doron Blatt, and Alfred O. Hero

EECS Department
The University of Michigan

2004 IEEE NSS-MIC

October 21, 2004

Introduction

- Desirable properties of statistical image reconstruction algorithms
  - Convergence to a solution
  - Fast convergence rate

- Ordered subsets (OS) algorithms
  - Synonyms: incremental gradient, block iterative, row action, . . .
  - Fast initial convergence rates but not convergent

- Convergent OS type algorithms for emission tomography
  - Relaxed OS algorithms (Ahn and Fessler, 2003):
    inconvenient to determine relaxation parameters
  - Incremental EM (COSEM) (Hsiao *et al.*, 2002)

- For transmission tomography?

- Incremental optimization transfer: generalizes incremental EM
  - Transmission incremental optimization transfer (TRIOT)

Review: Optimization Transfer Methods

- Optimization transfer: general framework for designing iterative optimization algorithms to find $\hat{x} = \arg\max_{x \in D} \Phi(x)$

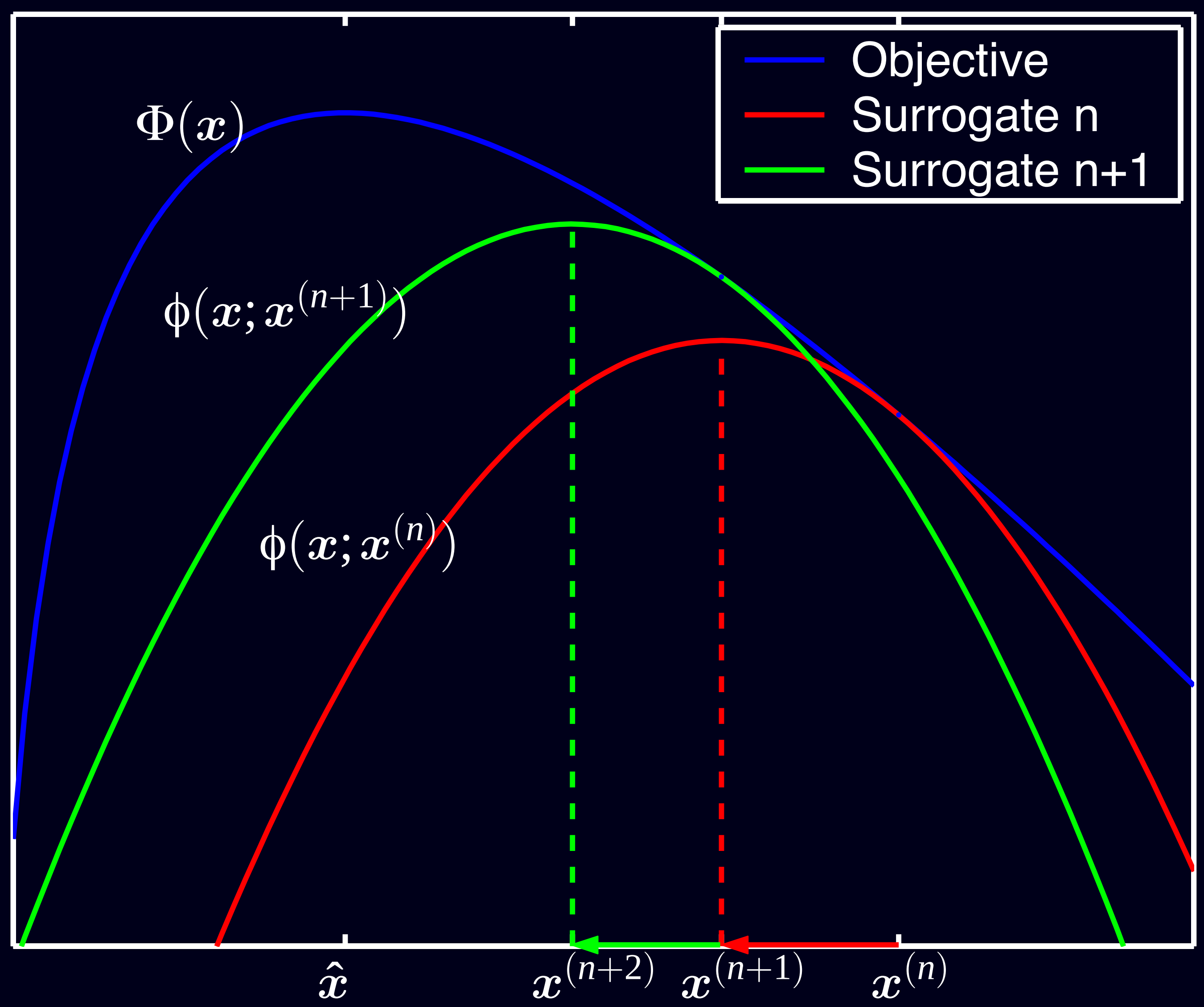
- For each iterate $x^{(n)}$,
  [S-step] choose a surrogate function $\phi(\cdot; x^{(n)})$
  [M-step] maximize the surrogate: $x^{(n+1)} = \arg\max_{x \in D} \phi(x; x^{(n)})$

- Desirable properties of surrogate function $\phi(\cdot; x^{(n)})$
  - minorization conditions

$$\begin{cases} \phi(x^{(n)}; x^{(n)}) = \Phi(x^{(n)}) & \text{“matched current value”} \\ \phi(x; x^{(n)}) \leq \Phi(x), \quad \forall x \in D & \text{“lie below”} \\ \nabla_x \phi(x; x^{(n)})\big|_{x=x^{(n)}} = \nabla \Phi(x)\big|_{x=x^{(n)}} & \text{“matched gradient”} \end{cases}$$

    $\Longrightarrow$ monotonicity $\Longrightarrow$ convergence (?)
  - easier to maximize (*e.g.* separable, . . .)
  - low curvature $\Longrightarrow$ fast convergence rate
  - example: EM surrogates (EM), quadratic surrogates (PS/SPS), . . .

1D Illustration



Review: Ordered Subsets Methods

- Goal: accelerate convergence rates

- What to need
  - partially-separable objective function (*e.g.*, independent observed data)

$$\Phi = \sum_{m=1}^{M} \Phi_m, \qquad \phi = \sum_{m=1}^{M} \phi_m, \qquad \begin{cases} \phi_m(\boldsymbol{x};\boldsymbol{x}) = \Phi_m(\boldsymbol{x}) \\ \phi_m(\boldsymbol{x};\boldsymbol{x}^{(n)}) \le \Phi_m(\boldsymbol{x}) \end{cases}$$

  - surrogate $\phi_m(\cdot;\boldsymbol{x}^{(n)})$ for each subobjective function Φ_m

- Idea: In M-step, use ϕ_k instead of $\phi = \sum_{m=1}^{M} \phi_m$ for some k

$$\boldsymbol{x}^{\text{new}} = \arg\max_{\boldsymbol{x}\in D} \phi_k(\boldsymbol{x};\boldsymbol{x}^{(n)})$$

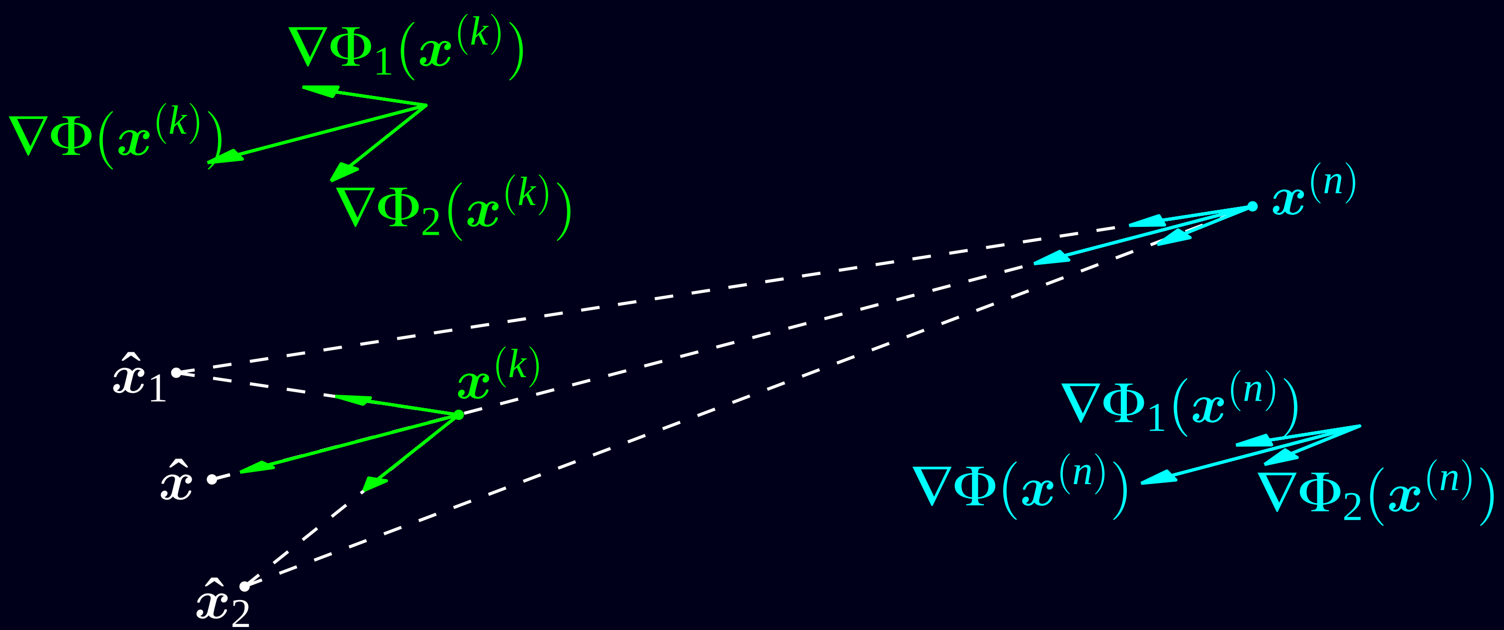
- In terms of "gradients": use partial gradient $\nabla\Phi_k$ instead of $\nabla\Phi$

- Key to success
  - To compute $\nabla\Phi_k$ is cheaper than $\nabla\Phi$.
  - Subset gradient balance: $\nabla\Phi_1 \approx \cdots \approx \nabla\Phi_M$, that is, $\nabla\Phi \approx M\nabla\Phi_m$

2D Illustration

- Two subset case:

$\Phi = \Phi_1 + \Phi_2$ with $\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} \Phi(\boldsymbol{x})$, $\hat{\boldsymbol{x}}_1 = \arg\max_{\boldsymbol{x}} \Phi_1(\boldsymbol{x})$, $\hat{\boldsymbol{x}}_2 = \arg\max_{\boldsymbol{x}} \Phi_2(\boldsymbol{x})$



- For $\boldsymbol{x}^{(n)}$ far from $\hat{\boldsymbol{x}}$ (early iterations), even partial gradients $\nabla\Phi_1$ or $\nabla\Phi_2$ point approximately at $\hat{\boldsymbol{x}}$.
- For $\boldsymbol{x}^{(k)}$ near $\hat{\boldsymbol{x}}$ (late iterations), $\nabla\Phi_1 \cong -\nabla\Phi_2$ since $\nabla\Phi \cong \mathbf{0}$
$\Rightarrow$ usually gets into a limit cycle $\Rightarrow$ convergence problem!

Incremental Optimization Transfer

- Goal: achieve convergence in OS approach

- Key idea
  - In M-step, use all most recent sub-surrogates ϕ_m (or $\nabla\Phi_m$).
  - Update only one of ϕ_m's (or compute only one of $\nabla\Phi_m$'s) at a time.

- Define augmented objective function

$$F(x; \bar{x}_1, \cdots, \bar{x}_M) = \sum_{m=1}^{M} \phi_m(x; \bar{x}_m).$$

- Fact:

$$\arg\max_{(x;\bar{x}_1,\cdots,\bar{x}_M)\in D^{M+1}} F(x; \bar{x}_1, \cdots, \bar{x}_M) = (\hat{x}, \cdots, \hat{x})$$
$$\hat{x} = \arg\max_{x\in D} \Phi(x)$$

  So, maximize F w.r.t $(x, \bar{x}_1, \cdots, \bar{x}_M) \Longleftrightarrow$ maximize Φ w.r.t. x

- Alternate between updating x and one of $\bar{x}_m$'s
  $\Longrightarrow$ incremental optimization transfer

- Special case: use of EM surrogates $\rightarrow$ incremental EM

Incremental Optimization Transfer Algorithms

Initialize $\boldsymbol{x}^{(0)}, \bar{\boldsymbol{x}}_1^{(0)}, \cdots, \bar{\boldsymbol{x}}_M^{(0)} \in D$

for $n = 0, \cdots, n_{\text{iter}} - 1$

for $m = 1, \cdots, M$

$$\boldsymbol{x}^{\text{new}} = \arg\max_{x \in D} F\left(\boldsymbol{x}; \bar{\boldsymbol{x}}_1^{(n+1)}, \cdots, \bar{\boldsymbol{x}}_{m-1}^{(n+1)}, \bar{\boldsymbol{x}}_m^{(n)}, \bar{\boldsymbol{x}}_{m+1}^{(n)}, \cdots \bar{\boldsymbol{x}}_M^{(n)}\right)$$

$$\bar{\boldsymbol{x}}_m^{(n+1)} = \arg\max_{\bar{x}_m \in D} F\left(\boldsymbol{x}^{\text{new}}; \bar{\boldsymbol{x}}_1^{(n+1)}, \cdots, \bar{\boldsymbol{x}}_{m-1}^{(n+1)}, \bar{\boldsymbol{x}}_m, \bar{\boldsymbol{x}}_{m+1}^{(n)}, \cdots \bar{\boldsymbol{x}}_M^{(n)}\right) = \boldsymbol{x}^{\text{new}}$$

end

$$\boldsymbol{x}^{(n+1)} = \bar{\boldsymbol{x}}_M^{(n+1)}$$

end

- Monotone in F (not necessarily in Φ)
- Convergence is ensured under mild conditions.

Statistical Model for Monoenergetic Transmission Scan

$$\boxed{Y_i \sim \text{Poisson}\left\{b_i e^{-[\boldsymbol{A}\boldsymbol{x}^{\text{true}}]_i} + r_i\right\}, \quad i = 1, \cdots, N}$$

- $\boldsymbol{Y} = [Y_1, \cdots, Y_N]'$: measurement data
- $\boldsymbol{A} = \{a_{ij}\}$: system matrix $(N \times p)$
- $\boldsymbol{x}^{\text{true}} = [x_1^{\text{true}}, \cdots, x_p^{\text{true}}]'$: unknown attenuation coefficient
- $\boldsymbol{b} = [b_1, \cdots, b_N]'$: known blank scan counts
- $\boldsymbol{r} = [r_1, \cdots, r_N]'$: known background contributions

Goal: estimate $\boldsymbol{x}^{\text{true}}$ from observed data $\boldsymbol{Y}$

Penalized-Likelihood Reconstruction

$$\boxed{\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x} \in D} \Phi(\boldsymbol{x}), \qquad \Phi(\boldsymbol{x}) = L(\boldsymbol{x}) - R(\boldsymbol{x})}$$

- Log-likelihood (data-fit term):

$$L(\boldsymbol{x}) = \sum_{i=1}^{N} h_i([\boldsymbol{A}\boldsymbol{x}]_i), \qquad h_i(l) = y_i \log(b_i e^{-l} + r_i) - (b_i e^{-l} + r_i)$$

- Roughness penalty (regularization term):

$$R(\boldsymbol{x}) = \frac{\beta}{2} \sum_{j=1}^{p} \sum_{k \in N_j} w_{jk} \psi(x_j - x_k)$$

β: regularization parameter, ψ: potential function
w_{jk}: weight, N_j: neighborhood of pixel j

- Box constraint: $D = \{\boldsymbol{x} \in \mathbb{R}^p : 0 \leq x_j \leq U, \;\; \forall j\}$ for some bound U
- If $r_i > 0$ and $y_i > r_i$ for some i, then Φ can be nonconcave, complicating maximization.

Application to Transmission Tomography: TRIOT

Use separable paraboloidal surrogates (SPS) (Erdoğan and Fessler, 1999)

$$\bar{\boldsymbol{x}}_m := \left[\left[\sum_{k=1}^{M} \breve{C}_k(\bar{\boldsymbol{x}}_k)\right]^{-1} \sum_{k=1}^{M} \left[\breve{C}_k(\bar{\boldsymbol{x}}_k)\bar{\boldsymbol{x}}_k + \nabla\Phi_k(\bar{\boldsymbol{x}}_k)\right]\right]^{+}, \quad m = 1, \cdots, M$$

where

$$\left[[\boldsymbol{x}]^{+}\right]_j = \mathrm{median}\{0, x_j, U\}$$

$$\breve{C}_k(\boldsymbol{x}) = \mathrm{diag}_j\{\breve{c}_{kj}(\boldsymbol{x})\}$$

$$\breve{c}_{mj}(\boldsymbol{x}) = \max\{\sum_{i \in S_m} a_{ij} a_i c_i([\boldsymbol{A}\boldsymbol{x}]_i) + \frac{2\beta}{M} \sum_{k \in N_j} w_{jk}\omega_\psi(x_j - x_k),\, \varepsilon\}$$

for some small $\varepsilon > 0$.

- Choices for curvature $c_i(\cdot)$
  - Maximum curvature (MC): convergent, slow
  - Optimum curvature (OC): convergent, fast, but extra backprojection
  - Precomputed curvature (PC): faster, practically convergent
- Note: built-in weighted averaging in TRIOT

Methods

- 2D attenuation map reconstruction
- Real PET data (Siemens/CIT ECAT EXACT 921 PET scanner)
- 128×128 image, 160×192 sinogram
- Test algorithms:
  SPS-MC/PC, OS-SPS, TRIOT-MC/PC
  (the second part denotes curvature type)
- Regularization parameter, $\beta = 2^{18.5}$, chosen by visual inspection
- Edge-preserving nonquadratic penalty

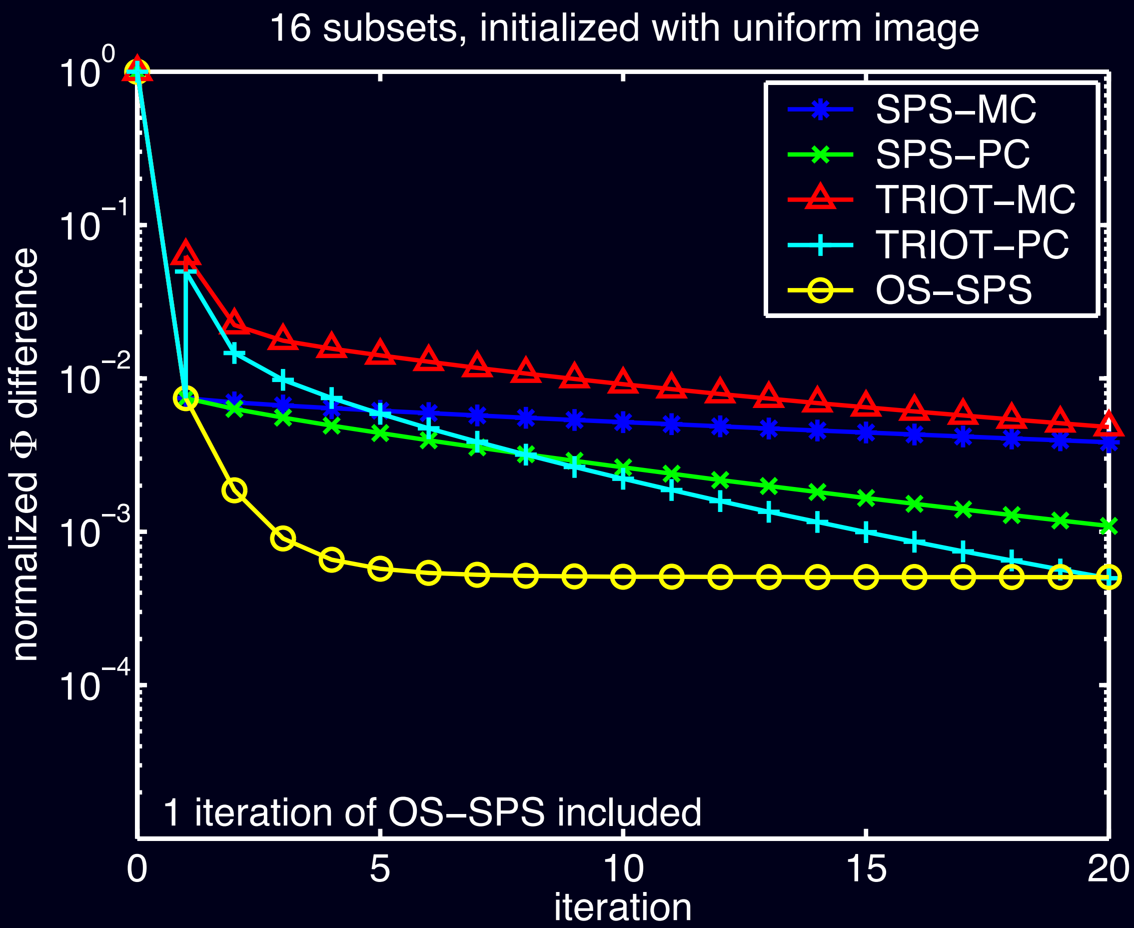
$$\psi(t) = \delta^2[|t/\delta| - \log(1 + |t/\delta|)]$$

  where $\delta = 4 \times 10^{-4}\mathrm{mm}^{-1}$, chosen by visual inspection

- Initialized with a uniform image

Results (16 subsets)

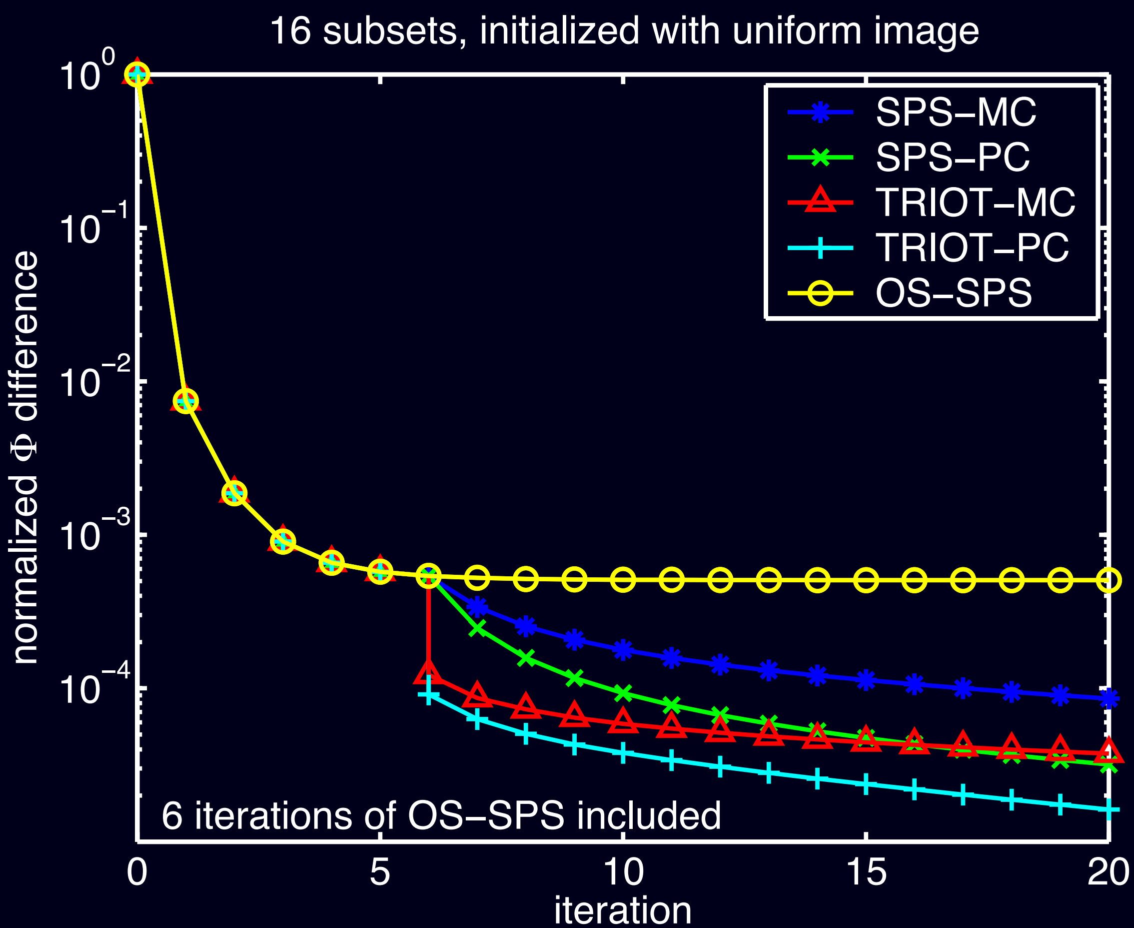
- 1 iteration of OS-SPS + SPS, TRIOT



- Initial convergence rates of TRIOT are slow. But . . .

Results (16 subsets)

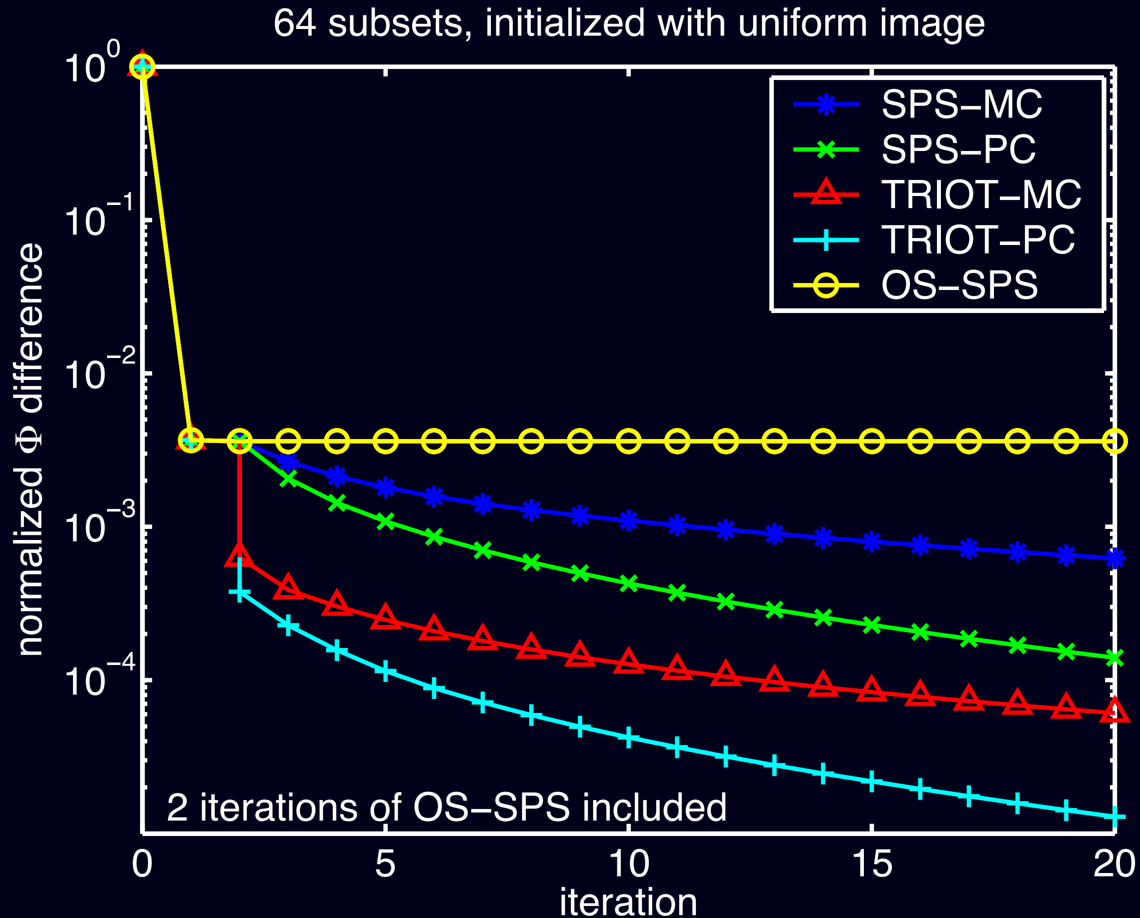
- If TRIOT is applied when OS-SPS reaches a limit cycle, ...
- 6 iterations of OS-SPS + SPS, TRIOT



- Recall the built-in averaging in TRIOT.

Results (64 subsets)

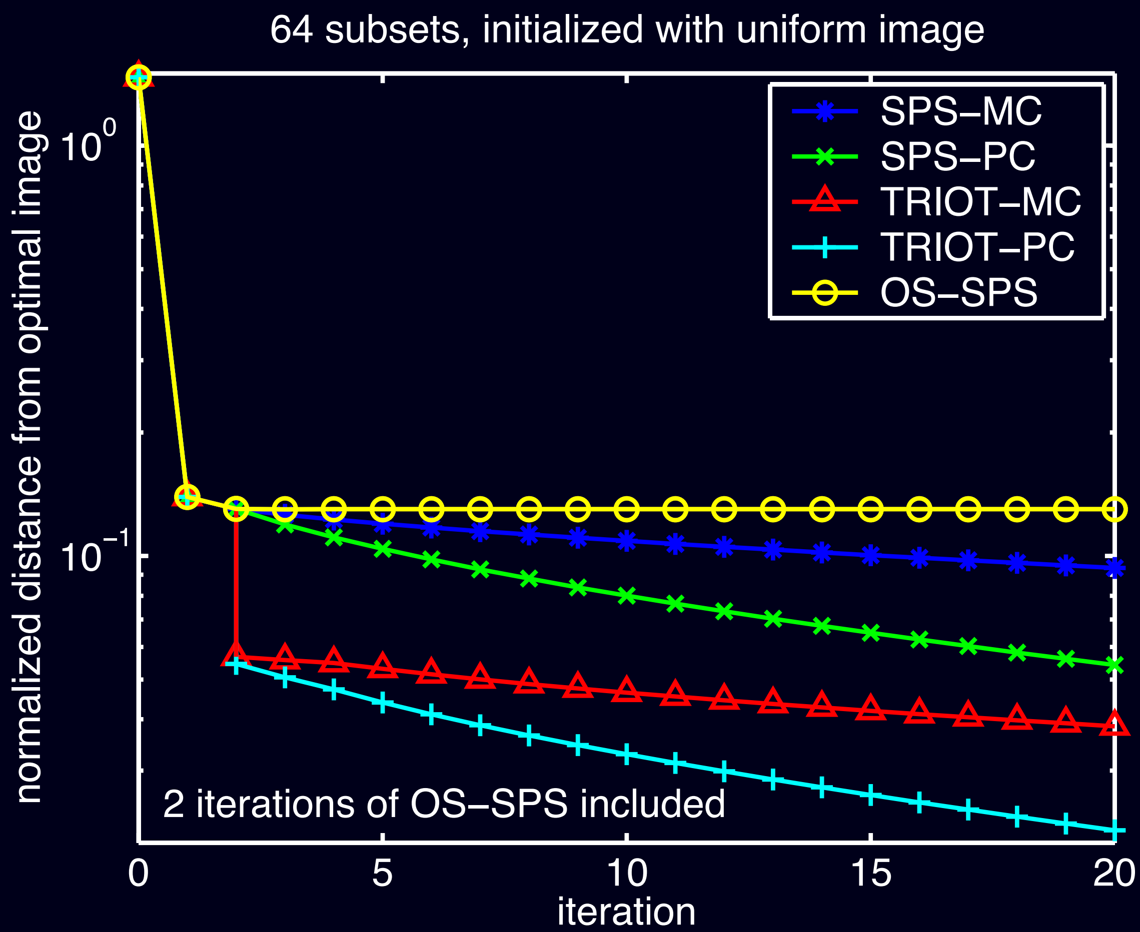
- 2 iterations of OS-SPS $+$ SPS, TRIOT



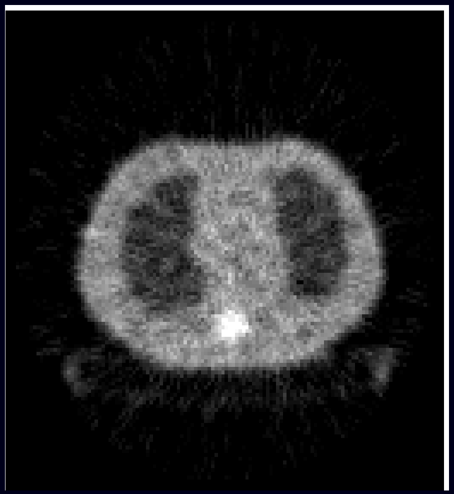
- TRIOT is quite effective.

Results (64 subsets)

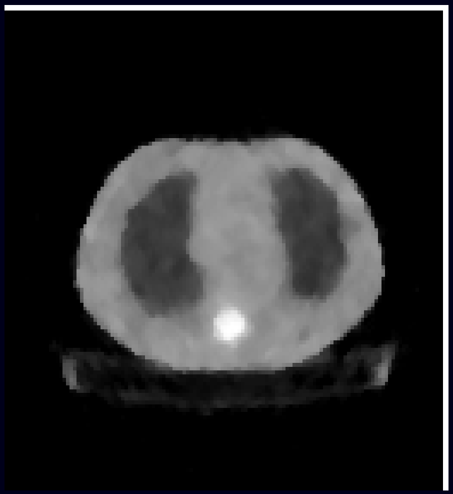
- The same as the previous one, but in terms of distance from the optimal image



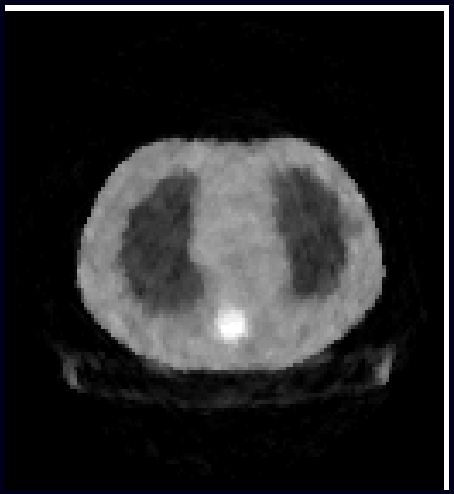
Reconstructed Images



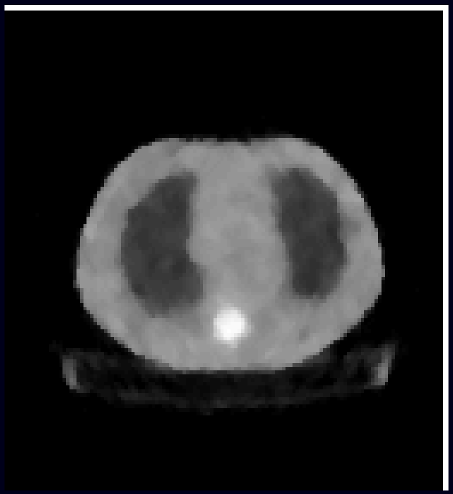
FBP



PL optimal image



OS-SPS
(64 subsets, 20 iterations)
one point of the limit cycle



TRIOT-PC
(64 subsets, 20 iterations
incl 2 iterations of OS-SPS)

Summary

- A broad family of convergent incremental optimization transfer algorithms that do not require relaxation parameters

- A particular algorithm: TRIOT for transmission tomography
  - not limited to transmission tomography

- Very effective to switch from OS-SPS to TRIOT when OS-SPS gets to a limit cycle (a couple of iterations of OS-SPS was enough for 64 subsets).

- Computational cost per iteration for TRIOT-PC/MC
  - 1 projection + 1 backprojection + "overhead" mainly due to penalty
  - Overhead depends on the number of subsets, system size, . . .

- TRIOT-OC and TRIOT-MC are provably convergent. TRIOT-PC is convergent?

- Ahn *et al.* (2004) submitted to TMI:
  `http://www.eecs.umich.edu/~fessler`

Acknowledgments

This research was supported in part by NIH Grants CA-60711 and CA-87634, by DOE Grant DE-FG02-87ER60561, and by a Rackham Predoctoral Fellowship.