Generalized Row-Action Methods for Tomographic Imaging

Sparse Tomo Days, Technical University of Denmark, March 27, 2014

Martin S. Andersen

joint work with
Per Christian Hansen

Scientific Computing Section

Outline

1. Algebraic methods for X-ray computed tomography
2. Relaxed incremental proximal methods
3. Numerical experiments

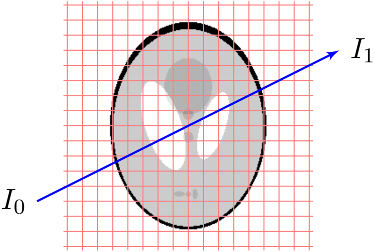
X-ray Computed Tomography

Measurement model

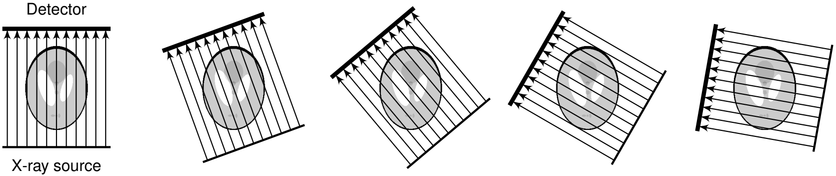
$$I_1 = I_0 \exp\left(- \int_l \mu(u_1, u_2)\, ds \right)$$

$$\log(I_0/I_1) \approx a_i^T x$$

$$b = Ax + e$$



Parallel beam measurement geometry



Algebraic reconstruction technique

Projection on hyperplane $\mathcal{H}_i = \{x \mid a_i^T x = b_i\}$

$$P_{\mathcal{H}_i}(x_k) = \underset{x \in \mathcal{H}_i}{\operatorname{argmin}} \|x - x_k\| = x_k - \frac{a_i(a_i^T x_k - b_i)}{\|a_i\|^2}$$

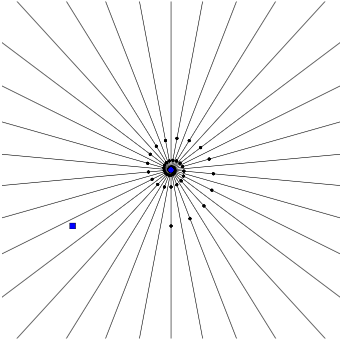
Kaczmarz's method / ART

$$x_{k+1} = \rho P_{\mathcal{H}_{i_k}}(x_k) + (1-\rho)x_k, \qquad i_k \in \{1, \ldots, m\}$$

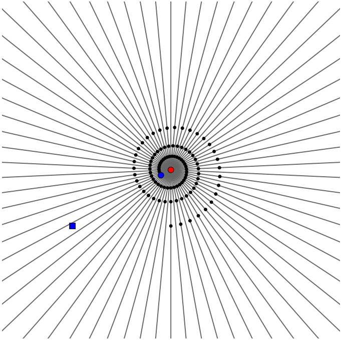
Relaxation parameter $\rho \in (0, 2)$

Example — consistent system

17 equations, $\rho = 1.0$

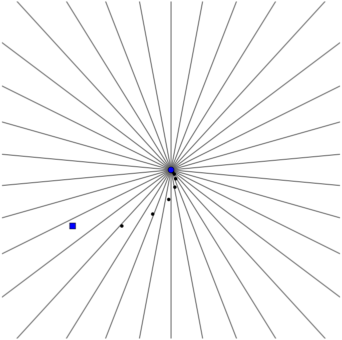


35 equations, $\rho = 1.0$

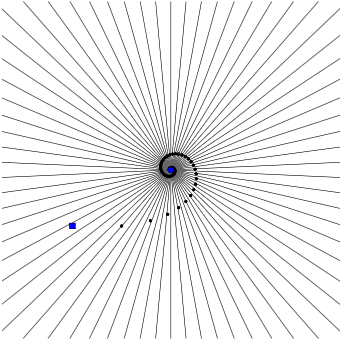


Example — consistent system

17 equations, $\rho = 0.5$

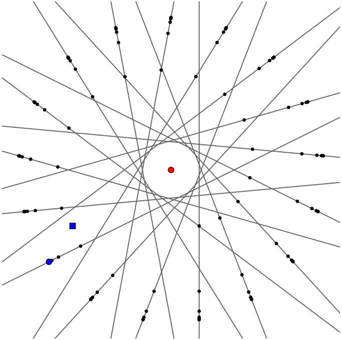


35 equations, $\rho = 0.5$

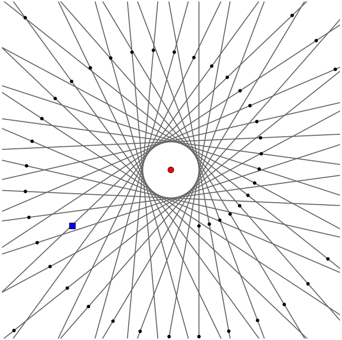


Example — inconsistent system

17 equations, $\rho = 1.0$

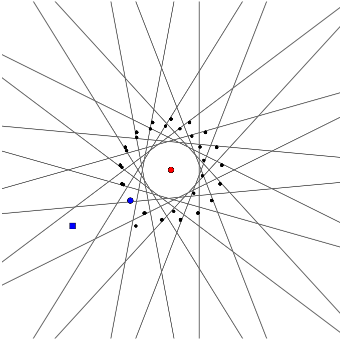


35 equations, $\rho = 1.0$

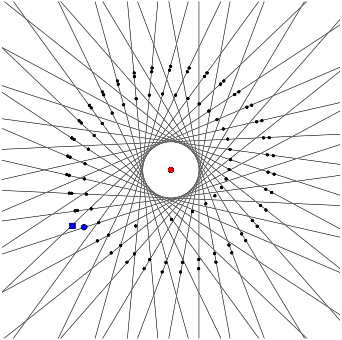


Example — inconsistent system

17 equations, $\rho = 0.5$

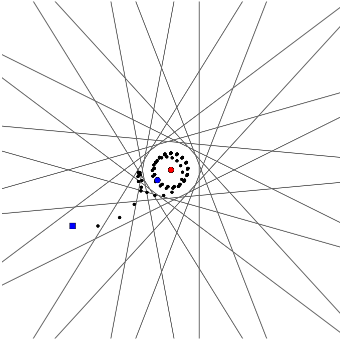


35 equations, $\rho = 0.5$

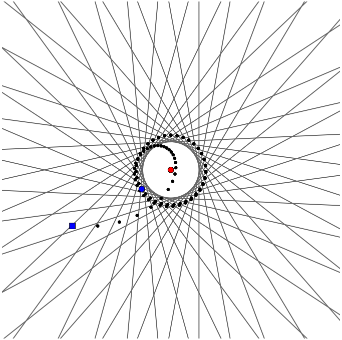


Example — inconsistent system

17 equations, $\rho = 0.2$

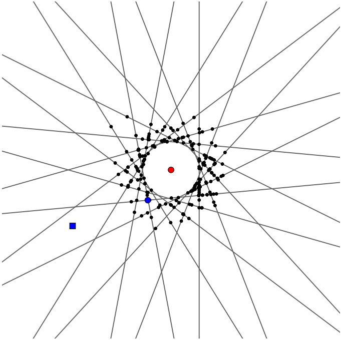


35 equations, $\rho = 0.2$

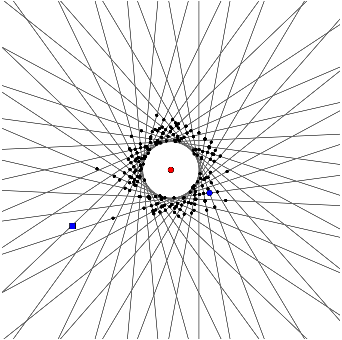


Example — inconsistent system, randomization

17 equations, $\rho = 1.0$

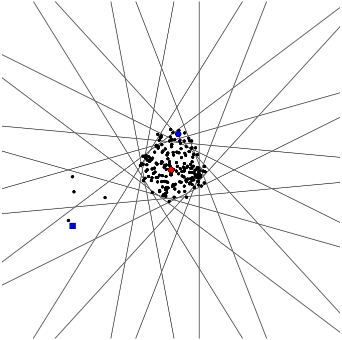


35 equations, $\rho = 1.0$

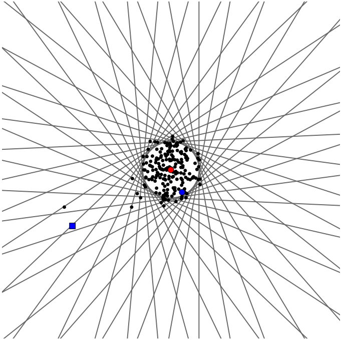


Example — inconsistent system, randomization

17 equations, $\rho = 0.5$

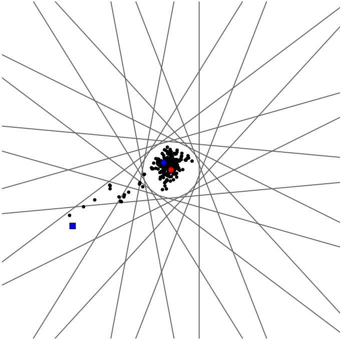


35 equations, $\rho = 0.5$

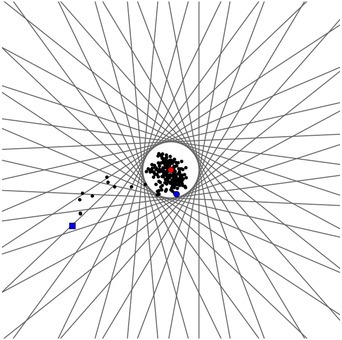


Example — inconsistent system, randomization

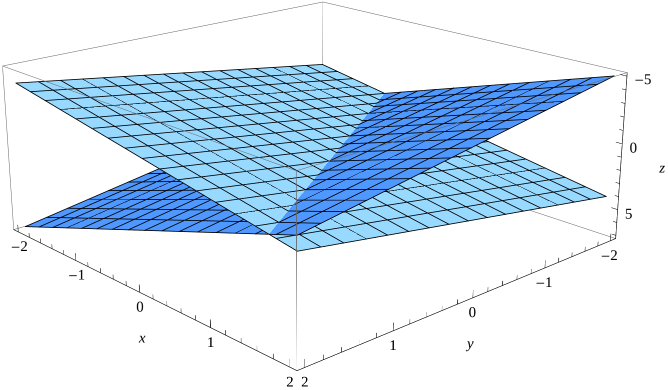
17 equations, $\rho = 0.2$



35 equations, $\rho = 0.2$

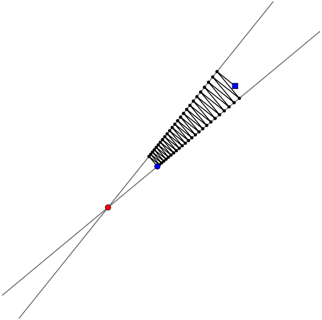


Example — underdetermined system



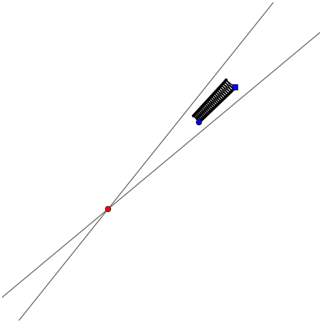
Example — consistent system

$\rho = 1.0$



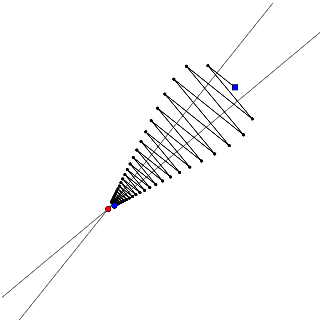
Example — consistent system

$\rho = 0.5$



Example — consistent system

$$\rho = 1.5$$



Tomographic image reconstruction

Ill-posed inverse problem — we need regularization!

Incorporate a priori knowledge in reconstruction problem

- spatial information (smoothness, piecewise constant/affine, ...)
- bounds (nonnegativity, box constraints, ...)
- sparsity
- ...

Large-scale optimization problem

- gradient computation is expensive
- may not be differentiable

Superiorization

Kaczmarz's method is pertubation resilient (Herman et al., 2009)

$$x_{k+1} = \mathcal{P}\big(x_k + t_k v_k\big), \qquad \mathcal{P} = P_{\mathcal{H}_m} \circ \cdots \circ P_{\mathcal{H}_2} \circ P_{\mathcal{H}_1}$$

Converges if $Ax = b$ is consistent and $t_k \to 0$ for $k \to \infty$

Perturbed iteration yields a "superior" solution in some sense

Incremental methods (I)

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{m} f_i(x) \\ \text{subject to} & x \in \mathcal{C} \end{array}$$

Incremental (sub)gradient iteration:

$$x_{k+1} = P_{\mathcal{C}}\big(x_k - t_k \widetilde{\nabla} f_{i_k}(x_k)\big), \qquad \widetilde{\nabla} f_{i_k}(x_k) \in \partial f_{i_k}(x_k)$$

- sublinear rate of convergence — *initial convergence* often very fast
- diminishing stepsize or “oscillation” that depends on stepsize
- goes back to Kibardin (1980), Litvakov (1966)
- Bertsekas (1996): incremental least-squares and extended Kalman filter

Incremental methods (II)

Incremental proximal iteration:

$$x_{k+1} = \underset{x \in \mathcal{C}}{\operatorname{argmin}} \Big\{ f_{i_k}(x) + \frac{1}{2t_k} \|x - x_k\|^2 \Big\}$$

equivalently

$$x_{k+1} = x_k - t_k g_{k+1}, \quad g_{k+1} \in \partial f_{i_k}(x_{k+1}) + N_{\mathcal{C}}(x_{k+1})$$

Linearized proximal iteration: let $g_k \in \partial f_{i_k}(x_k)$

$$\begin{aligned} x_{k+1} &= P_{\mathcal{C}}\Big(\underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \Big\{ f_{i_k}(x_k) + g_k^T x + \frac{1}{2t_k} \|x - x_k\|^2 \Big\} \Big) \\ &= P_{\mathcal{C}}\big(x_k - t_k g_k \big) \end{aligned}$$

Incremental proximal gradient methods (I)

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{m}\big(g_i(x) + h_i(x)\big) \\ \text{subject to} & x \in \mathcal{C} \end{array}$$

Algorithm 1 (Bertsekas, 2011)

$$z_k = \underset{u\in\mathbb{R}^n}{\operatorname{argmin}}\Big\{g_{i_k}(u) + \frac{1}{2t_k}\|u - x_k\|^2\Big\}$$

$$x_{k+1} = P_{\mathcal{C}}\big(z_k - t_k\widetilde{\nabla} h_{i_k}(z_k)\big)$$

Interpretation

$$x_{k+1} = P_{\mathcal{C}}\big(x_k - t_k\widetilde{\nabla} g_{i_k}(z_k) - t_k\widetilde{\nabla} h_{i_k}(z_k)\big)$$

Incremental proximal gradient methods (II)

Algorithm 2 (Bertsekas, 2011)

$$z_k = x_k - t_k \widetilde{\nabla} h_{i_k}(x_k)$$
$$x_{k+1} = \operatorname*{argmin}_{u \in \mathcal{C}} \Big\{ g_{i_k}(u) + \frac{1}{2t_k} \|u - z_k\|^2 \Big\}$$

Interpretation

$$\begin{aligned} x_{k+1} &= \operatorname*{argmin}_{u \in \mathcal{C}} \Big\{ g_{i_k}(u) + \widetilde{\nabla} h_{i_k}(x_k)^T u + \frac{1}{2t_k} \|u - x_k\|^2 \Big\} \\ &= P_{\mathcal{C}}\Big(x_k - t_k \widetilde{\nabla} g_{i_k}(x_{k+1}) - t_k \widetilde{\nabla} h_{i_k}(x_k) \Big) \end{aligned}$$

Relaxed incremental proximal gradient methods

Algorithm 1

$$
\begin{aligned}
w_k &= \operatorname*{argmin}_{u \in \mathbb{R}^n} \Big\{ g_{i_k}(u) + \frac{1}{2t_k} \|u - x_k\|^2 \Big\} \\
z_k &= w_k - t_k \widetilde{\nabla} h_{i_k}(w_k) \\
x_{k+1} &= P_{\mathcal{C}}\big(\rho\, z_k + (1-\rho) x_k\big)
\end{aligned}
$$

Algorithm 2

$$
\begin{aligned}
w_k &= w_k - t_k \widetilde{\nabla} h_{i_k}(w_k) \\
z_k &= \operatorname*{argmin}_{u \in \mathbb{R}^n} \Big\{ g_{i_k}(u) + \frac{1}{2t_k} \|u - w_k\|^2 \Big\} \\
x_{k+1} &= P_{\mathcal{C}}\big(\rho\, z_k + (1-\rho) x_k\big)
\end{aligned}
$$

Convergence results

Cyclic order or random order?

- cyclic order yields worst-case performance bounds
- random order yields expected performance bounds

Constant stepsize or diminishing stepize?

- constant stepsize yields convergence within an error bound
- diminishing stepsize yields exact convergence

Randomized cyclic order works well in practice

R-IPG vs. ART

$$\begin{array}{ll} \text{minimize} & (1/2)\sum_{i=1}^{m}(a_i^T x - b_i)^2 \\ \text{subject to} & x \in \mathcal{C} \end{array}$$

Let $g_i(x) = (1/2)(a_i^T x - b_i)^2$ and $h_i(x) = 0$:

$$x_{k+1} = P_{\mathcal{C}}\left(x_k - \rho \frac{a_{i_k}(a_{i_k}^T x_k - b_{i_k})}{t_k^{-1} + \|a_{i_k}\|^2}\right)$$

Interpretation: ART with damped step size

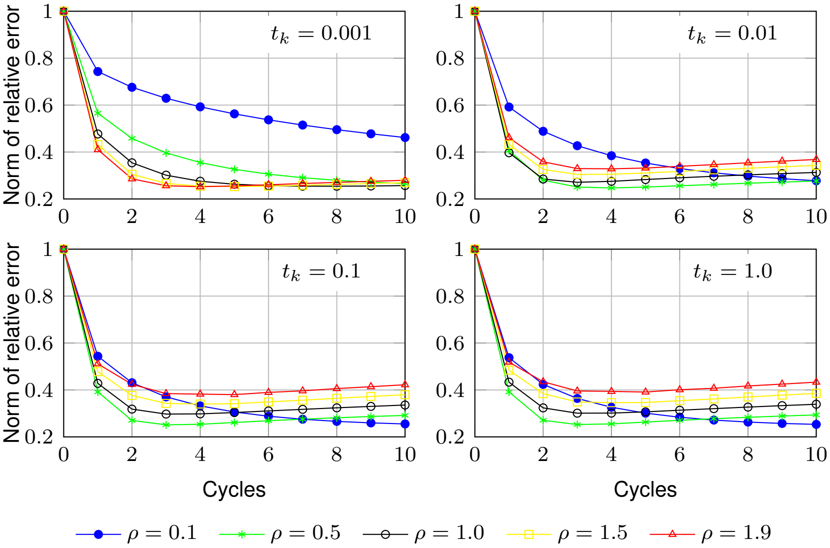
R-IPG vs. ART — numerical example

2D tomography problem

- 256×256 Shepp–Logan phantom
- $n = 256^2 = 65536$ variables
- $p = 120$ uniformly spaced angles
- $r = 362$ parallel rays
- $m = pr = 43440$ measurements
- Gaussian noise: $e_i \sim \mathcal{N}(0, 0.02 \cdot \|b\|_\infty)$

Algorithm: R-IPG with constant step size

R-IPG vs. ART — numerical example



Regularized data fitting

$$\begin{array}{ll} \text{minimize} & f(x) \equiv (1/2)\|Ax - b\|^2 + \lambda\, h(x) \\ \text{subject to} & x \in \mathcal{C} \end{array}$$

Example: express $f(x)$ as

$$f(x) = \sum_{i=1}^{m} \underbrace{(1/2)(a_i^T x - b_i)^2}_{g_i(x)} + \sum_{i=1}^{m} \underbrace{\frac{\lambda}{m} h(x)}_{h_i(x)}$$

Total-variation regularization

$$\begin{array}{ll} \text{minimize} & (1/2)\|Ax-b\|_2^2 + \lambda \sum_{i=1}^n \|D_i x\|_2 \\ \text{subject to} & x \in \mathcal{C} \end{array}$$

Let $g_i(x) = \|A_i x - b_i\|_2^2$ and $h_i(x) = (\lambda/p) \sum_{i=1}^n \|D_i x\|_2$

$$\begin{aligned} z_k &= \operatorname*{argmin}_{u \in \mathbb{R}^n} \Big\{ g_{i_k}(u) + \frac{1}{2t_k} \|u - x_k\|^2 \Big\} \\ &= x_k - A_{i_k}^T (A_{i_k} A_{i_k}^T + t_k^{-1} I)^{-1} (A_{i_k} x_k - b_{i_k}) \\ x_{k+1} &= P_{\mathcal{C}}\big(\rho(z_k - t_k \widetilde{\nabla} h_{i_k}(z_k)) + (1-\rho) x_k\big) \end{aligned}$$

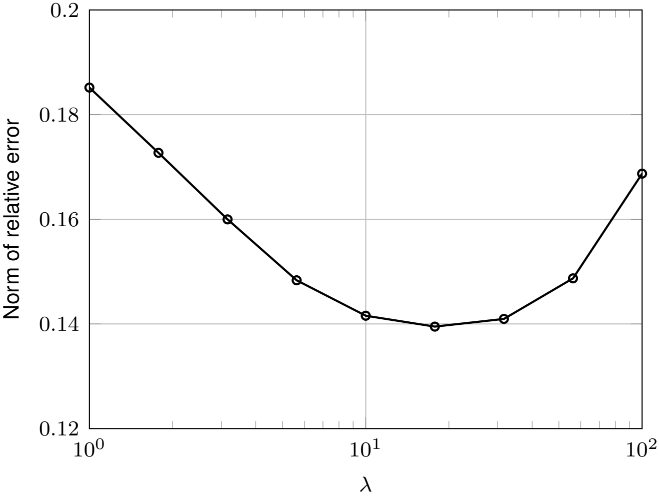
Total-variation regularization — numerical example

2D tomography problem

- 512×512 Shepp–Logan phantom
- $n = 512^2 = 262144$ variables
- $p = 60$ uniformly spaced angles
- $r = 724$ parallel rays
- $m = pr = 43440$ measurements
- Gaussian noise: $e_i \sim \mathcal{N}(0, 0.01 \cdot \|b\|_\infty)$

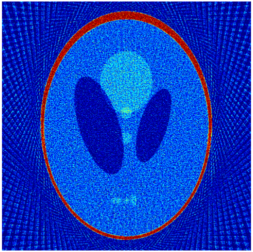
Algorithm: R-IPG1 with diminishing step-size sequence

Regularization curve



Reference

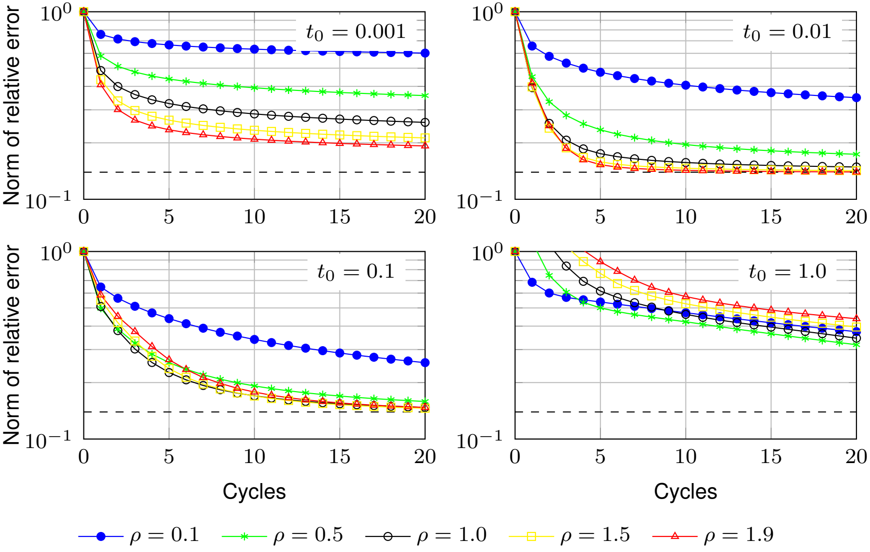
Original



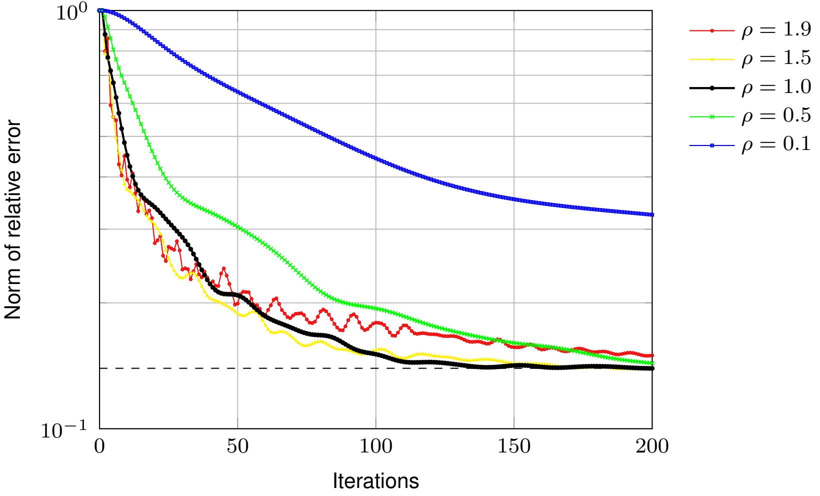
Filtered backprojection

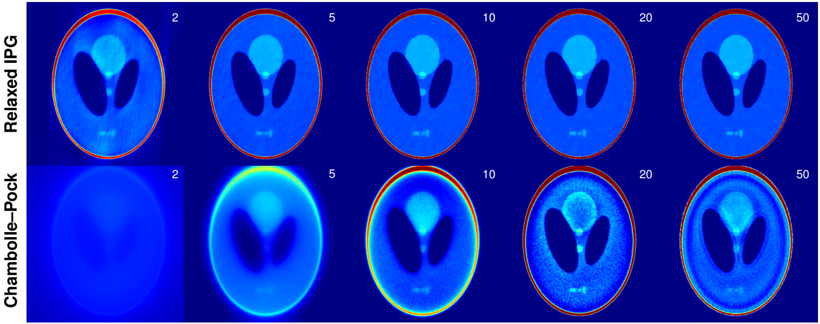
Total variation regularized LS

Relaxed incremental proximal subgradient method



Primal–dual first-order method (Chambolle & Pock)




Relaxed IPG
Chambolle–Pock
2
5
10
20
50
2
5
10
20
50

Summary

- relaxed incremental proximal gradient methods
- slow global rate of convergence
- often fast initial rate rate of convergence
- hybrid methods that transition from incremental to non-incremental method

Thank you for listening!

mskan@dtu.dk

http://compute.dtu.dk/~mskan