Fast Alternating Direction Methods

Tom Goldstein
Brendan O'Donoghue
Simon Setzer

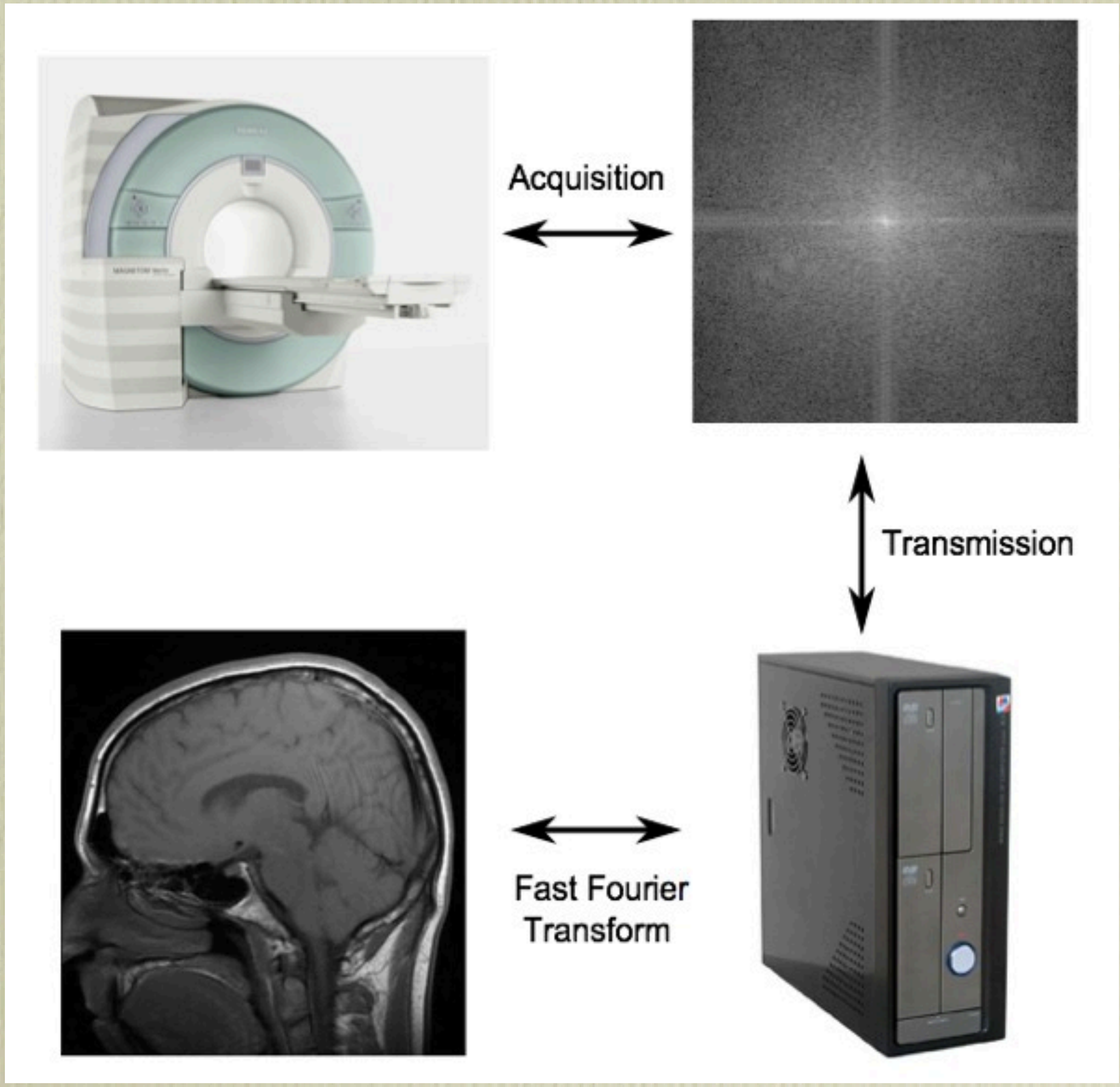
Non-Differentiable Problems

- ROF Denoising $\min |\nabla u| + \frac{\mu}{2}\|u - f\|^2$
- TV Deblurring $\min |\nabla u| + \frac{\mu}{2}\|Ku - f\|^2$
- General Problem:

$$\text{minimize} \quad H(u) + G(Av)$$

Compressed MRI

- Data is acquired in the Fourier Domain



Compressive MRI

$$\min |\nabla u| + \frac{\mu}{2}\|R\mathcal{F}u - f\|^2$$

- R is a 'Row Selector' Matrix

$$R = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Split Bregman Method

$$\min |\nabla u| + \frac{\mu}{2}\|Au - f\|^2$$

- Make change of Variables: $d \leftarrow \nabla u$
- 'Split Bregman' form:

$$\begin{array}{ll} \text{minimize} & |v| + \frac{\mu}{2}\|Au - f\|^2 \\ \text{subject to} & u - \nabla v = 0 \end{array}$$

- Now, apply splitting methods (e.g. ADMM)

Splitting Methods

$$\begin{array}{ll} \text{minimize} & H(u) + G(v) \\ \text{subject to} & Au + Bv = b \end{array}$$

- Idea: Remove Constraint
  - Add Lagrange Multiplier
  - Add Penalty Term

$$\max_{\lambda} \min_{u,v} H(u) + G(v) + \langle \lambda, b - Au - Bv \rangle + \|b - Au - Bv\|^2$$

Alternating Direction Method of Multipliers

Algorithm 1 ADMM

Require: $v_0 \in R^{N_v}$, $\lambda_0 \in R_b^N$, $\tau > 0$

for $k = 0, 1, \dots$ **do**

$\quad u_{k+1} = \operatorname{argmin}_u H(u) + \langle \lambda_k, -Au \rangle + \frac{\tau}{2}\|b - Au - Bv_k\|^2$

$\quad v_{k+1} = \operatorname{argmin}_v G(v) + \langle \lambda_k, -Bv \rangle + \frac{\tau}{2}\|b - Au_{k+1} - Bv\|^2$

$\quad \lambda_{k+1} = \lambda_k + \tau(b - Au_{k+1} - Bv_{k+1})$

end for

Alternating Minimization Algorithm

Algorithm 1 AMA

Require: $\lambda_0 \in R^{N_b}$, $\tau > 0$

for $k = 0, 1, \ldots$ **do**

$\quad u_{k+1} = \operatorname{argmin}_u H(u) + \langle \lambda_k, -Au \rangle$

$\quad v_{k+1} = \operatorname{argmin}_v G(v) + \langle \lambda_k, -Bv \rangle + \frac{\tau}{2} \| b - Au_{k+1} - Bv \|^2$

$\quad \lambda_{k+1} = \lambda_k + \tau (b - Au_{k+1} - Bv_{k+1})$

end for

Why is Split Bregman Good?

$$\min |\nabla u| + \frac{\mu}{2}\|Au - f\|^2$$

$$\begin{array}{ll} \text{minimize} & |v| + \frac{\mu}{2}\|Au - f\|^2 \\ \text{subject to} & u - \nabla v = 0 \end{array}$$

Algorithm 1 ADMM for TV

Require: $v_0 \in R^{N_v}$, $\lambda_0 \in R_b^N$, $\tau > 0$

for $k = 0, 1, \ldots$ **do**

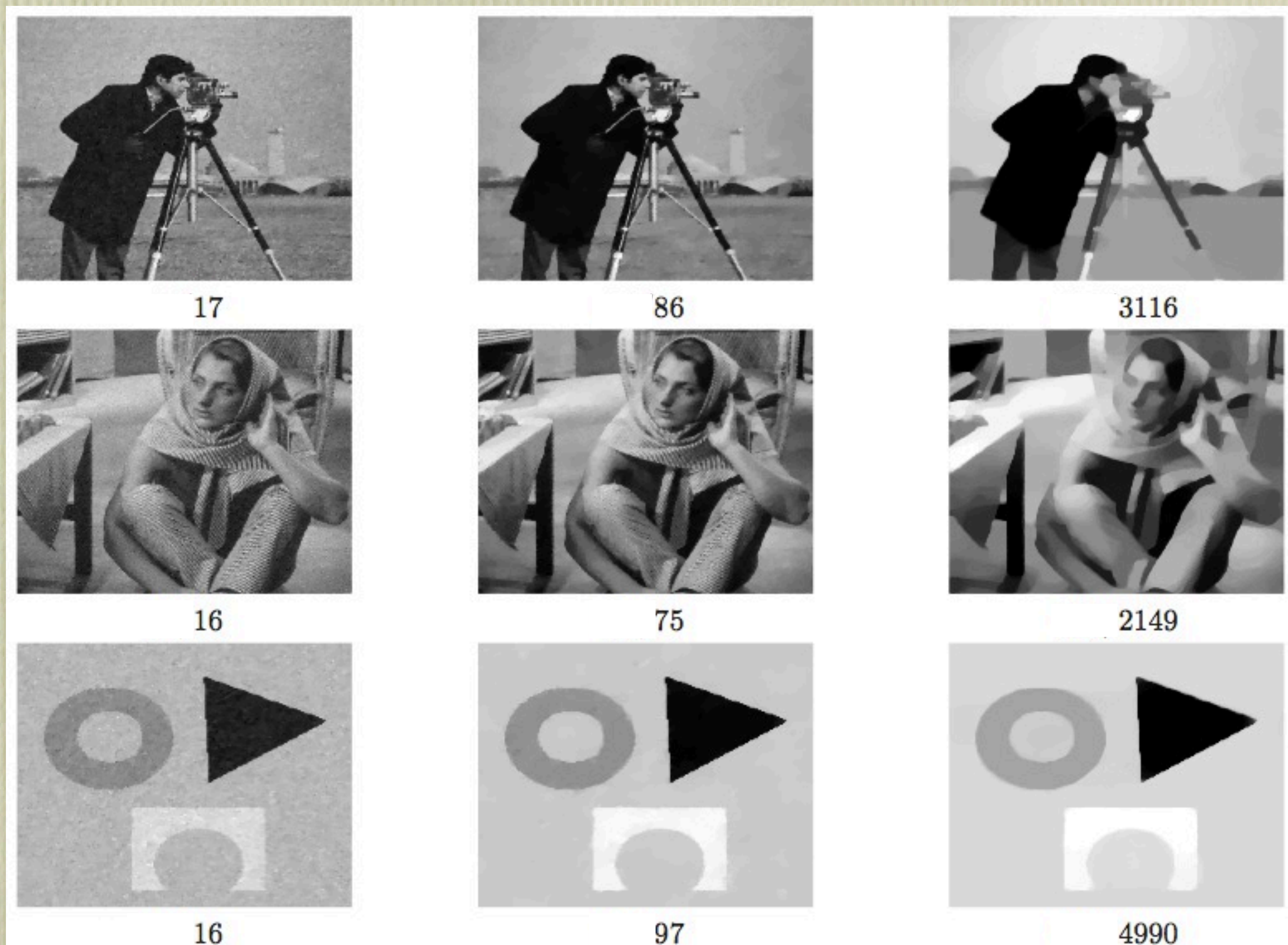
  $u_{k+1} = \text{argmin}_u \|Au - f\| + \frac{\tau}{2}\|v_k - \nabla u - \lambda_k\|^2$

  $v_{k+1} = \text{argmin}_v |v| + \frac{\tau}{2}\|v - \nabla u_{k+1} - \lambda_k\|^2$

  $\lambda_{k+1} = \lambda_k + \tau(\nabla u_{k+1} - v_{k+1})$

end for

Why is Split Bregman Bad?



How Can We Fix This Problem?

- Problem: Poor conditioning
- Solution: Acceleration convergence
- Inspired by Nesterov's method

Gradient Descent

$$\min_x F(x)$$

- Simplest approach:

Algorithm 1 Gradient Descent

1: **for** $k = 0, 1, \cdots$ **do**
2: $\quad x_{k+1} = x_k - \tau \nabla F(x_k)$
3: **end for**

- Slow: $F(x_k) - F(x^*) < O(1/k)$

Nesterov's Optimal Descent

Algorithm 1 Nesterov's Optimal Gradient Descent

Require: $\alpha_0 = 1$, $x_0 = y_1 \in R^N$, $\tau < 1/L(\nabla F)$

for $k = 1, 2, 3 \ldots$ **do**

$\quad x_k = y_k - \tau \nabla F(y_k)$

$\quad \alpha_{k+1} = (1 + \sqrt{4\alpha_k^2 + 1})/2$

$\quad y_{k+1} = x_k + (\alpha_k - 1)(x_k - x_{k-1})/\alpha_{k+1}$

end for

- Converges Fast: $F(x_k) - F(x^*) < O(1/k^2)$
- Optimal Rate (Nemirovski and Yundin '83)

Complexity of Constrained Problems

$$\begin{array}{ll} \text{minimize} & H(u) + G(v) \\ \text{subject to} & Au + Bv = b \end{array}$$

- How do we measure convergence rate?
- Saddle Point Formulation (Lagrange Mult's)

$$\max_{\lambda} \min_{u,v} H(u) + G(v) + \langle \lambda, b - Au - Bv \rangle$$

- Convert into Dual functional:

$$\text{maximize} \quad \left(D(\lambda) = -H^*(A^T\lambda) + \langle \lambda, b \rangle - G^*(B^T\lambda) \right)$$

Fast ADMM

Require: $v_{-1} = \hat{v}_0 \in R^{N_v}, \lambda_{-1} = \hat{\lambda}_0 \in R^{N_b}, \tau > 0$

for $k = 1, 2, 3 \ldots$ **do**

$u_k = \operatorname{argmin} H(u) + \langle \hat{\lambda}_k, -Au \rangle + \frac{\tau}{2} \|b - Au - B\hat{v}_k\|^2$

$v_k = \operatorname{argmin} G(v) + \langle \hat{\lambda}_k, -Bv \rangle + \frac{\tau}{2} \|b - Au_k - Bv\|^2$

$\lambda_k = \hat{\lambda}_k + \tau(b - Au_k - Bv_k)$

$\alpha_{k+1} = \frac{1+\sqrt{1+4\alpha_k^2}}{2}$

$\hat{v}_{k+1} = v_k + \frac{\alpha_k - 1}{\alpha_{k+1}}(v_k - v_{k-1})$

$\hat{\lambda}_{k+1} = \lambda_k + \frac{\alpha_k - 1}{\alpha_{k+1}}(\lambda_k - \lambda_{k-1})$

end for

Fast AMA

Require: $\alpha_0 = 1,\ \lambda_{-1} = \hat{\lambda}_0 \ \in R^{N_b}, \tau < \sigma_H / \rho(A^T A)$

for $k = 0, 1, 2 \ldots$ **do**

$\quad u_k = \operatorname{argmin} H(u) + \langle \hat{\lambda}_k, -Au \rangle$

$\quad v_k = \operatorname{argmin} G(v) + \langle \hat{\lambda}_k, -Bv \rangle + \frac{\tau}{2} \|b - Au_k - Bv\|^2$

$\quad \lambda_k = \hat{\lambda}_k + \tau(b - Au_k - Bv_k)$

$\quad \alpha_{k+1} = (1 + \sqrt{1 + 4\alpha_k^2})/2$

$\quad \hat{\lambda}_{k+1} = \lambda_k + \frac{\alpha_k - 1}{\alpha_{k+1}} (\lambda_k - \lambda_{k-1})$

end for

Convergence Results

- Convergence of Fast AMA is straight Forward:
  - Regularity assumption the same
  - Stepsize restriction is a little tighter

Theorem: If H is strongly convex and $\tau < \sigma_H/\rho(A^T A)$, then fast AMA converges in the dual objective with rate $O(1/k^2)$. More precisely, if D denotes the dual objective, then

$$D(\lambda^\star) - D(\lambda_k) \leq O(1/k^2).$$

Convergence Results

- To prove formal convergence bounds, we need some assumptions:
  - Strong convexity of the objective
  - Stepsize restriction

Theorem: Suppose that H and G are strongly convex, and that

$$\tau^3 \leq \frac{\sigma_H \sigma_G^2}{\rho(A^T A)\rho(B^T B)^2}.$$

Then we have the convergence bound

$$D(\lambda^\star) - D(\lambda_k) \leq \frac{2\tau \|\hat{\lambda}_1 - \lambda^\star\|^2}{(k+2)^2}$$

Primal and Dual Residuals

$$\begin{array}{ll} \text{minimize} & |v| + \frac{\mu}{2}\|Au - f\|^2 \\ \text{subject to} & u - \nabla v = 0 \end{array}$$

- Optimality/KKT conditions:

$$\begin{aligned} 0 &\in \partial H(u^\star) - A^T\lambda^\star \\ 0 &\in \partial G(v^\star) - B^T\lambda^\star \end{aligned}$$

- Residuals measure how close iterates are to solution

$$\begin{aligned} r_k &= b - Au_k - Bv_k \\ d_k &= \tau A^T B(v_k - v_{k-1}) \end{aligned}$$

- Restart Rule: Reset acceleration parameters any time $\max\{r_k, d_k\}$ increases

Nestorov's Method is Not Optimal!

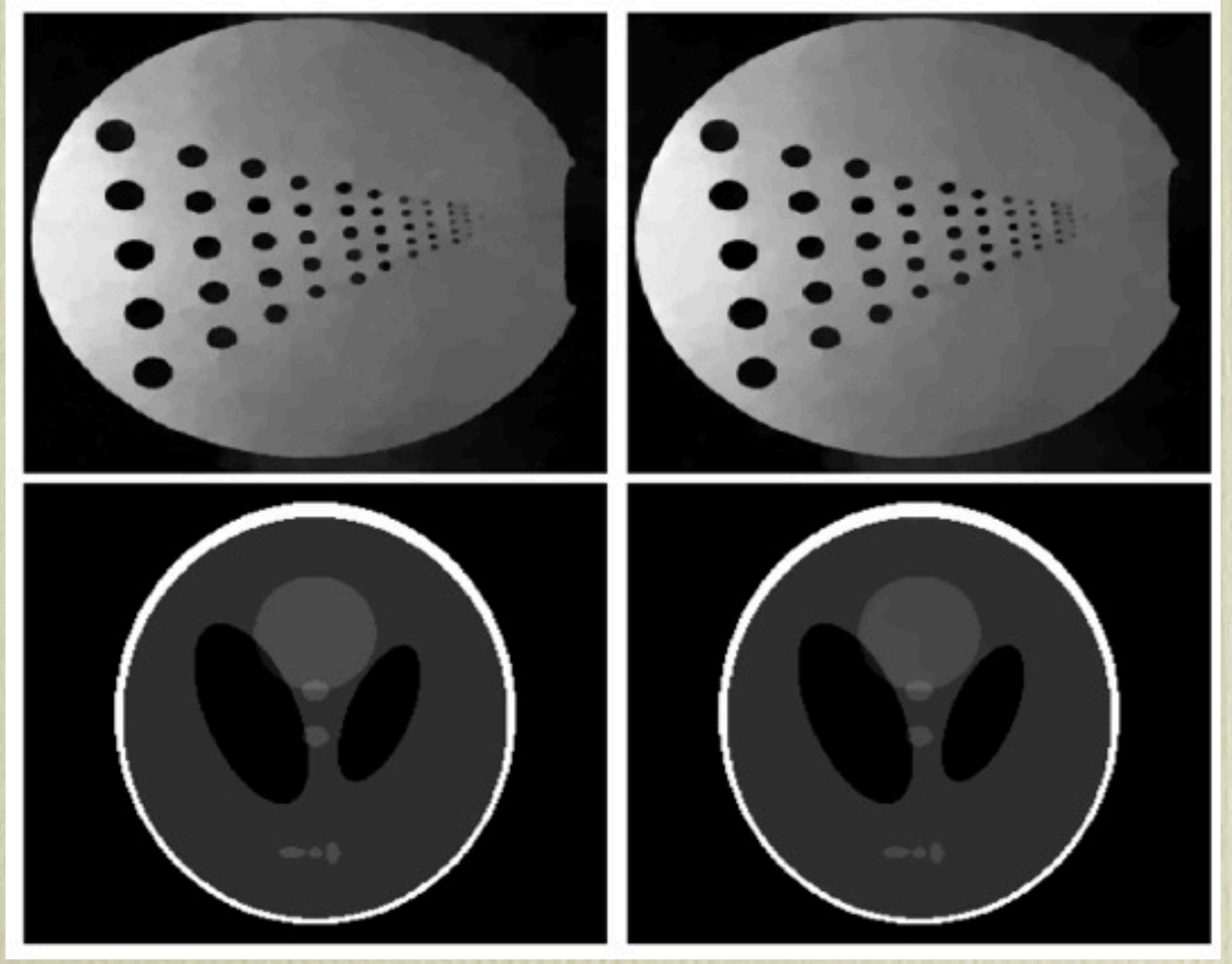
- Nesterov's method achieved 'Optimal' complexity over the set of ALL smooth minimization problems
- In practice, we can do better using 'restart' rules (O'Donoghue & Candes '12)

Results: ROF

$$\min |\nabla u| + \frac{\mu}{2}\|u - f\|^2$$

Results: CS

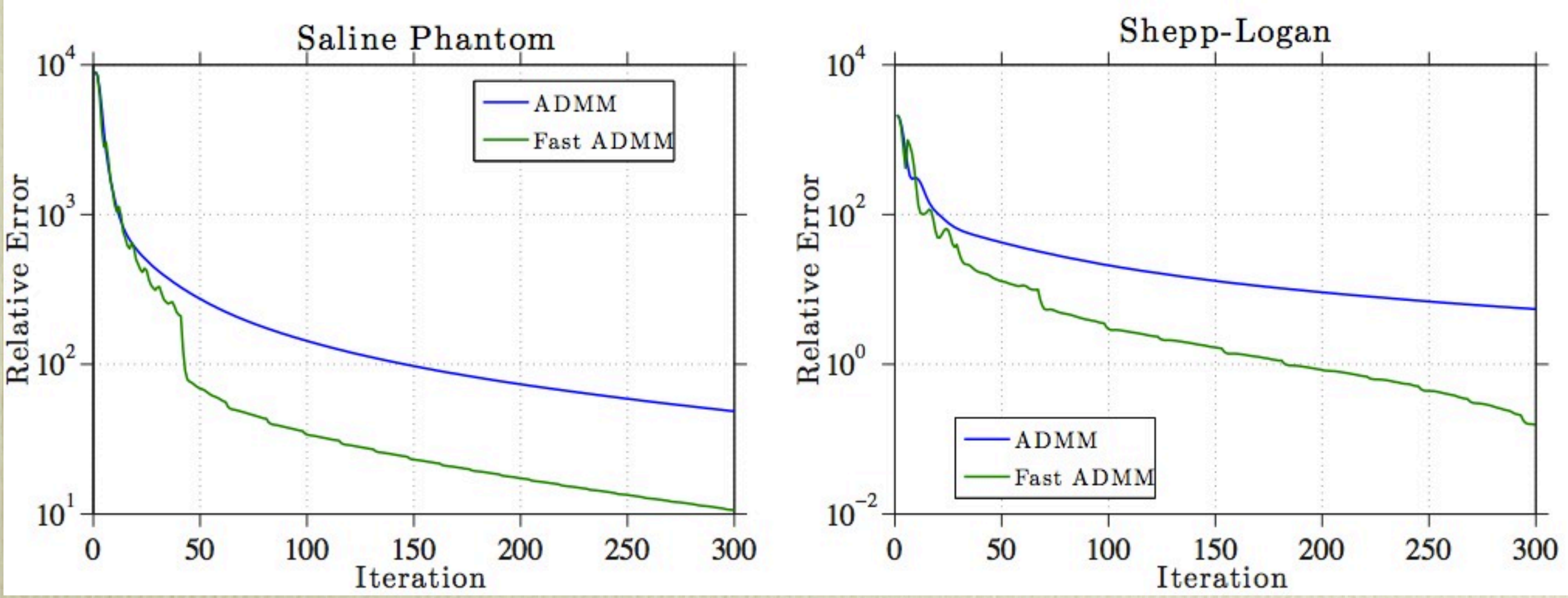
$$\min |\nabla u| + \frac{\mu}{2}\|R\mathcal{F}u - f\|^2$$



- Test images: (left) original (right) reconstructed from 30% samples

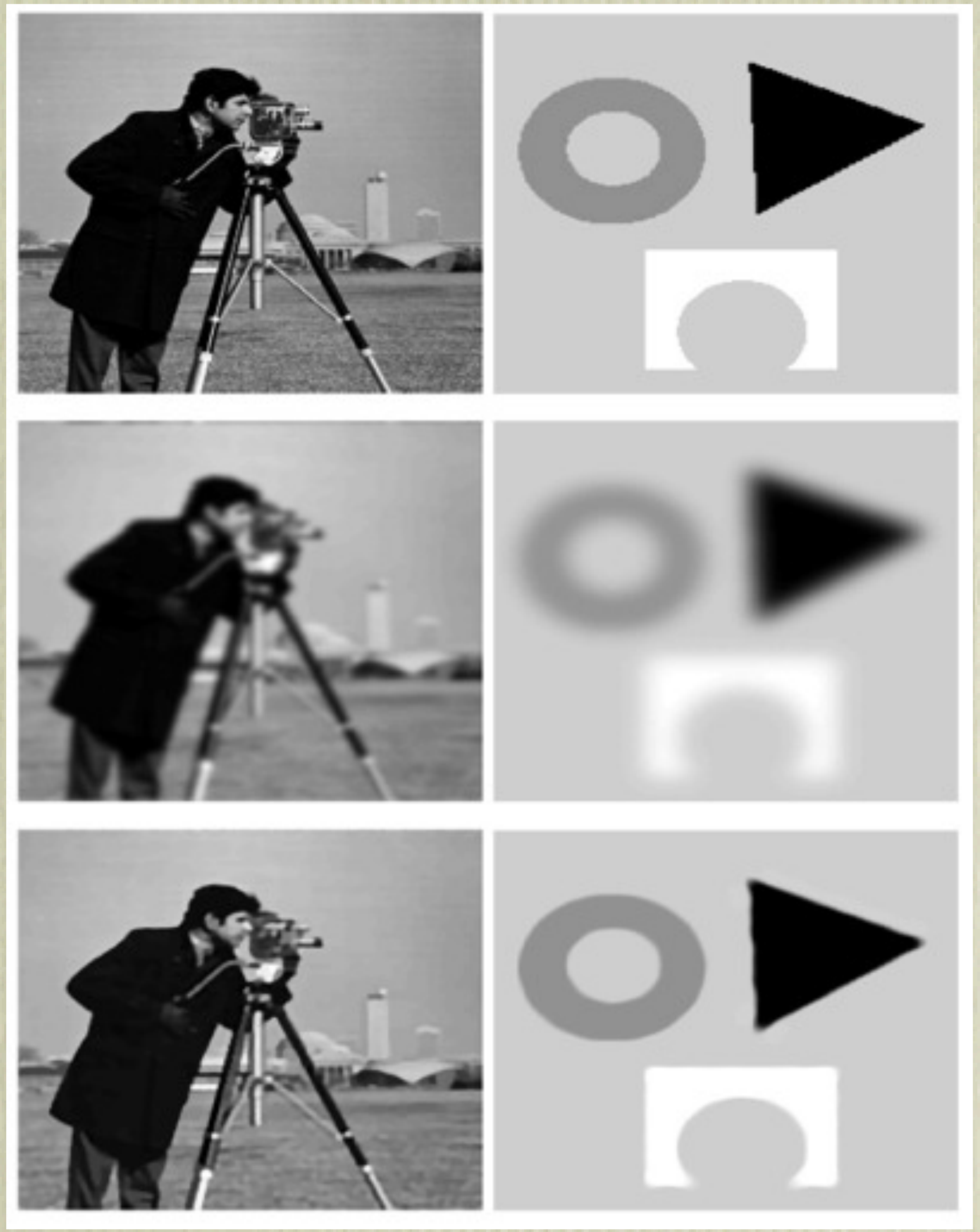
Results: CS

- Convergence curves: objective error



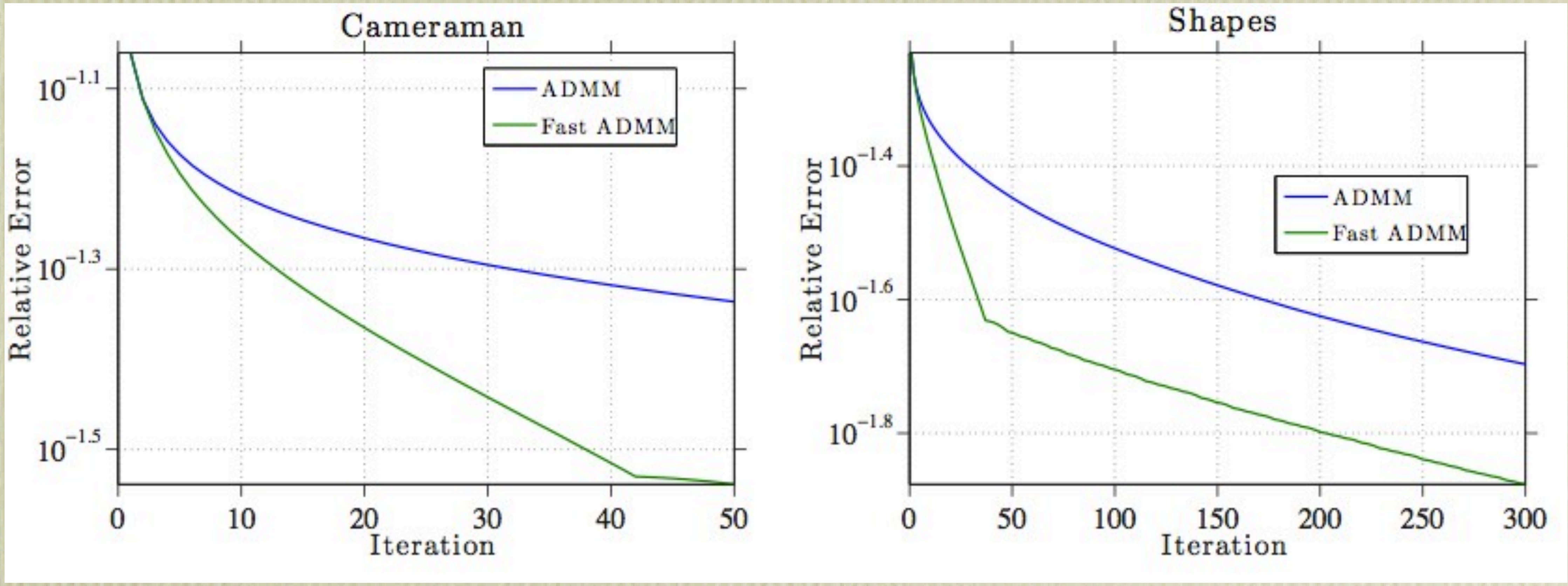
Results: Deblurring

$$\min |\nabla u| + \frac{\mu}{2}\|Ku - f\|^2$$



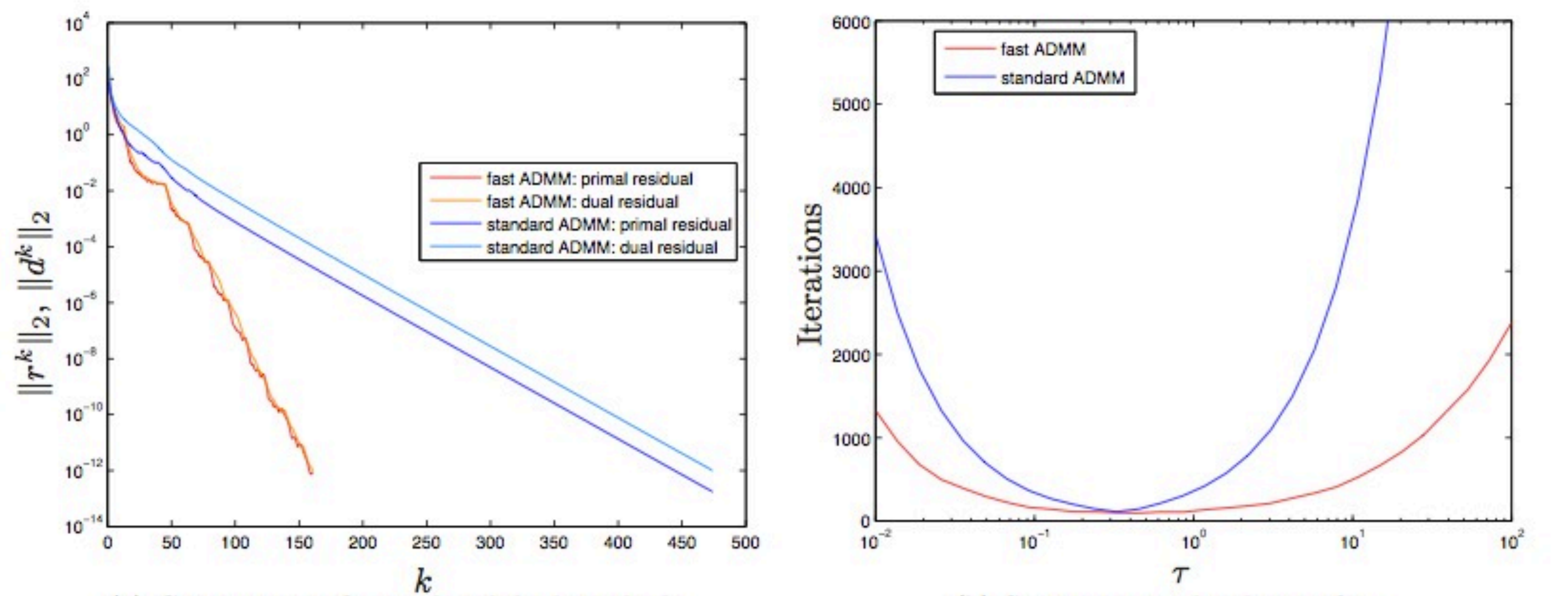
Results: Deblurring

$$\min |\nabla u| + \frac{\mu}{2}\|Ku - f\|^2$$



Results: QP

$$\begin{array}{ll} \text{minimize} & (1/2)u^TQu + q^Tu \\ \text{subject to} & Au \leq b, \end{array}$$

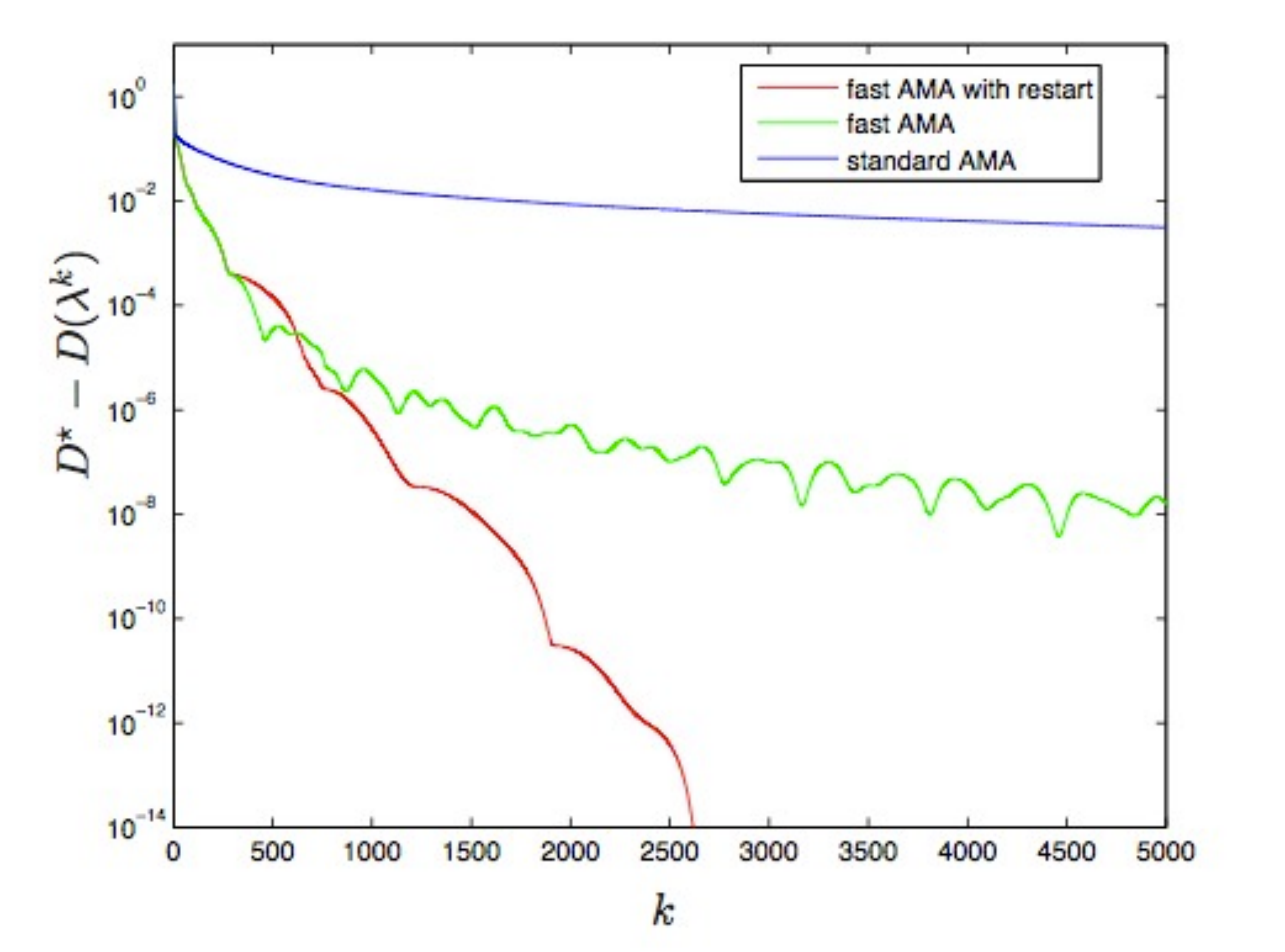


(a) Convergence of primal and dual residuals.

(b) Sensitivity to the choice of τ.

Results: QP

$$\begin{array}{ll} \text{minimize} & (1/2)u^T Q u + q^T u \\ \text{subject to} & Au \leq b, \end{array}$$



Thanks!

Acknowledgments:

This work was possible through the generous support of the National Science Foundation (grant number 08-582).