The Zoom strategy for accelerating and warm-starting interior methods

INFORMS Annual Meeting

Pittsburgh, PA, November 5–8, 2006

Michael Saunders and Leo Tenenblat

Dept of Management Sci & Eng	IMPA
Stanford University	Rio de Janeiro
Stanford, CA 94305-4026	Brazil
saunders@stanford.edu	leo@async.com.br

First presented at ISMP 19, Rio de Janeiro, Brazil — August 2006

Abstract

Interior methods using iterative solvers for each search direction can require drastically increasing work per iteration as higher accuracy is sought.

The Zoom strategy solves first to low accuracy, and then solves for a correction to both primal and dual variables, again to low accuracy. We "zoom in" on the correction by scaling it up, thus permitting a cold start for the correction.

The same strategy applies to warm-starting in general.

Motivation

The problem that started it all

Image reconstruction

Nagy and Strakoš 2000 Byunggyoo Kim thesis, 2002

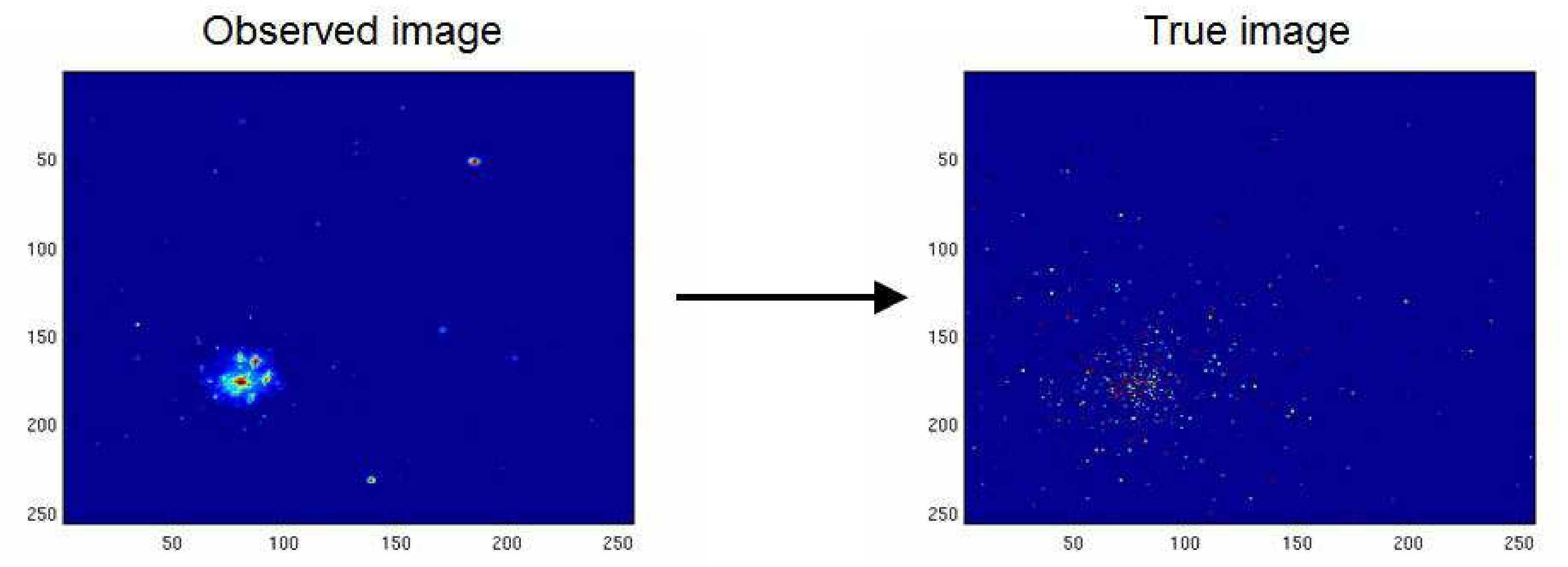


Image Reconstruction

$$\min \quad \lambda e^T x + \tfrac{1}{2}\|r\|^2$$
$$\text{st} \quad Ax + r = b, \quad x \geq 0$$

NNLS: Non-negative least squares $\quad \lambda = 10^{-4}$

A is an expensive operator 2-D DFT

65K $\times$ 65K

PDCO uses LSQR for each dual search direction Δy:

$$\min \left\| \begin{pmatrix} DA^T \\ I \end{pmatrix} \Delta y - \begin{pmatrix} Dw \\ r_1 \end{pmatrix} \right\|$$

PDCO Solver

MATLAB primal-dual interior method

http://www.stanford.edu/group/SOL/software.html

Nominal problem:

$$\text{NP} \qquad \underset{x}{\text{minimize}} \quad \phi(x) \qquad \text{subject to} \quad Ax = b, \quad \ell \le x \le u$$

$\phi(x)$ convex, separable

Regularized problem:

$$\text{NP}(\gamma, \delta) \qquad \underset{x,\, r}{\text{minimize}} \quad \phi(x) + \tfrac{1}{2}\|\gamma x\|^2 + \tfrac{1}{2}\|r\|^2 \qquad \text{subject to} \quad Ax + \delta r = b, \quad \ell \le x \le u$$

PDCO search directions

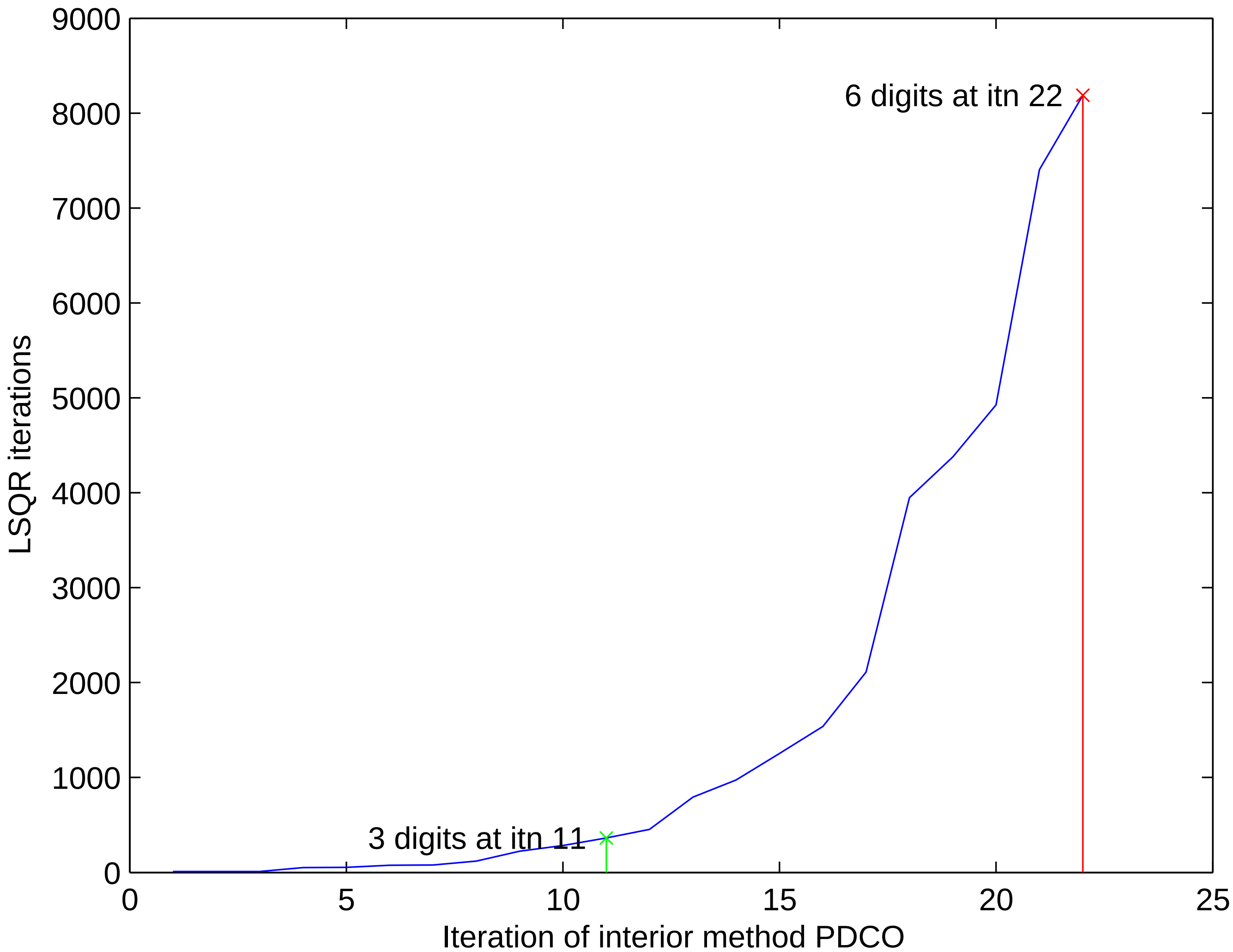
3 methods for computing Δy:

- Cholesky on $(AD^2A^T + \delta^2 I)\Delta y = AD^2 w + \delta r_1$
- Sparse QR on $\min \left\| \begin{pmatrix} DA^T \\ \delta I \end{pmatrix} \Delta y - \begin{pmatrix} Dw \\ r_1 \end{pmatrix} \right\|$
- LSQR on same LS problem (iterative solver)

Must use LSQR when A is an operator

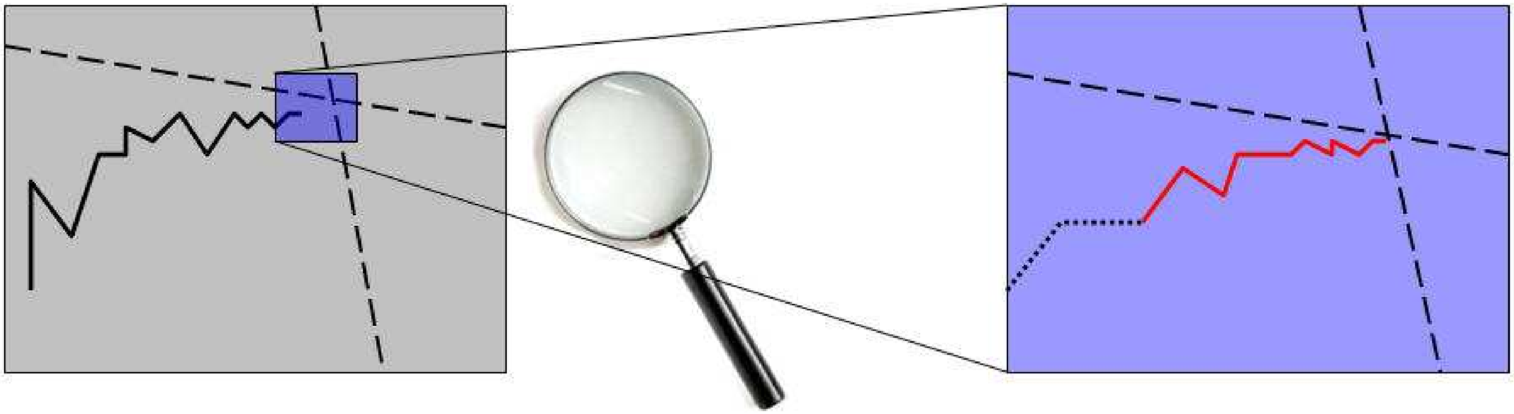
Motivation

LSQR iterations increase exponentially with requested accuracy



Zoom strategy: Accelerating IPMs

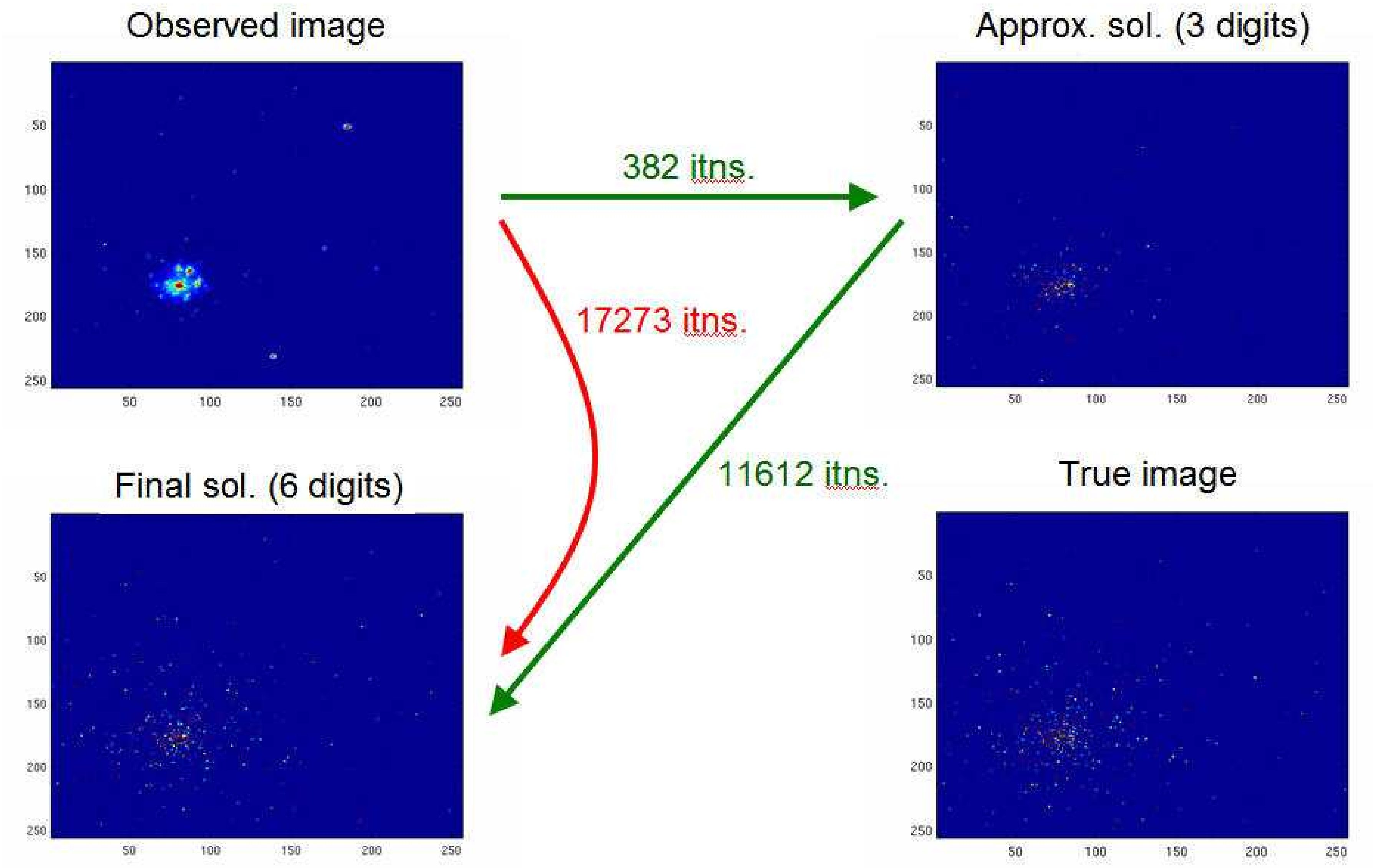
- Solve to 3 digits: cheap approximation to x, y, z
- Define new problem for correction dx, dy, dz
- Zoom in (scale up correction)
- Solve to 3 digits: cheap approximation to dx, dy, dz



Cold start for both solves

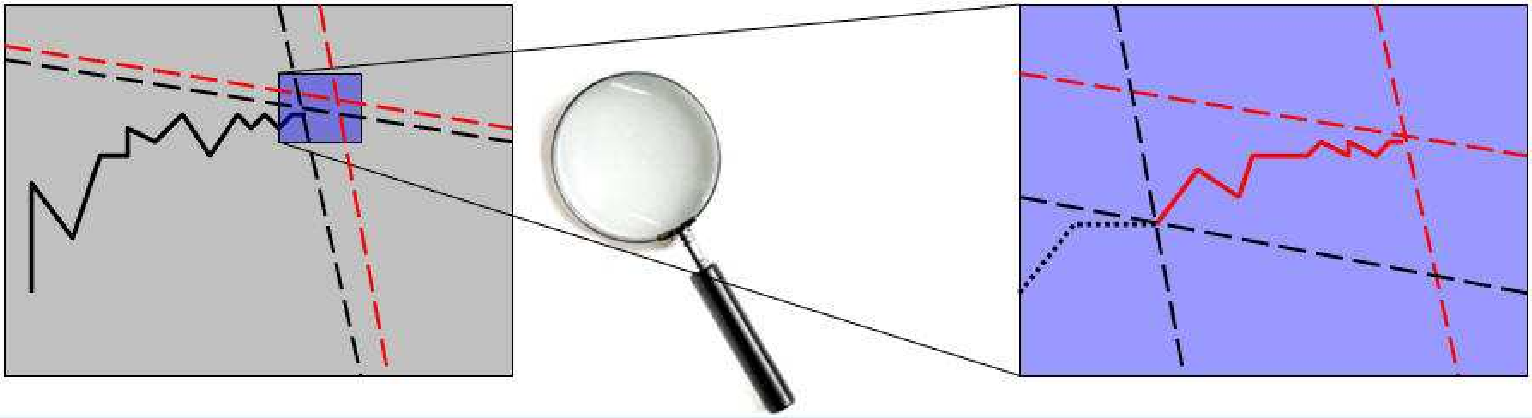
Results: Accelerating IPMs

LSQR iterations inside PDCO



Zoom strategy: Warm-starting IPMs

- Set solution to original LP as current approximation
- Define new problem for correction dx, dy, dz
- Zoom in (scale up correction)
- Solve loosely: cheap approximation to dx, dy, dz



Cold start for loose solve

Warm-starting IPMs

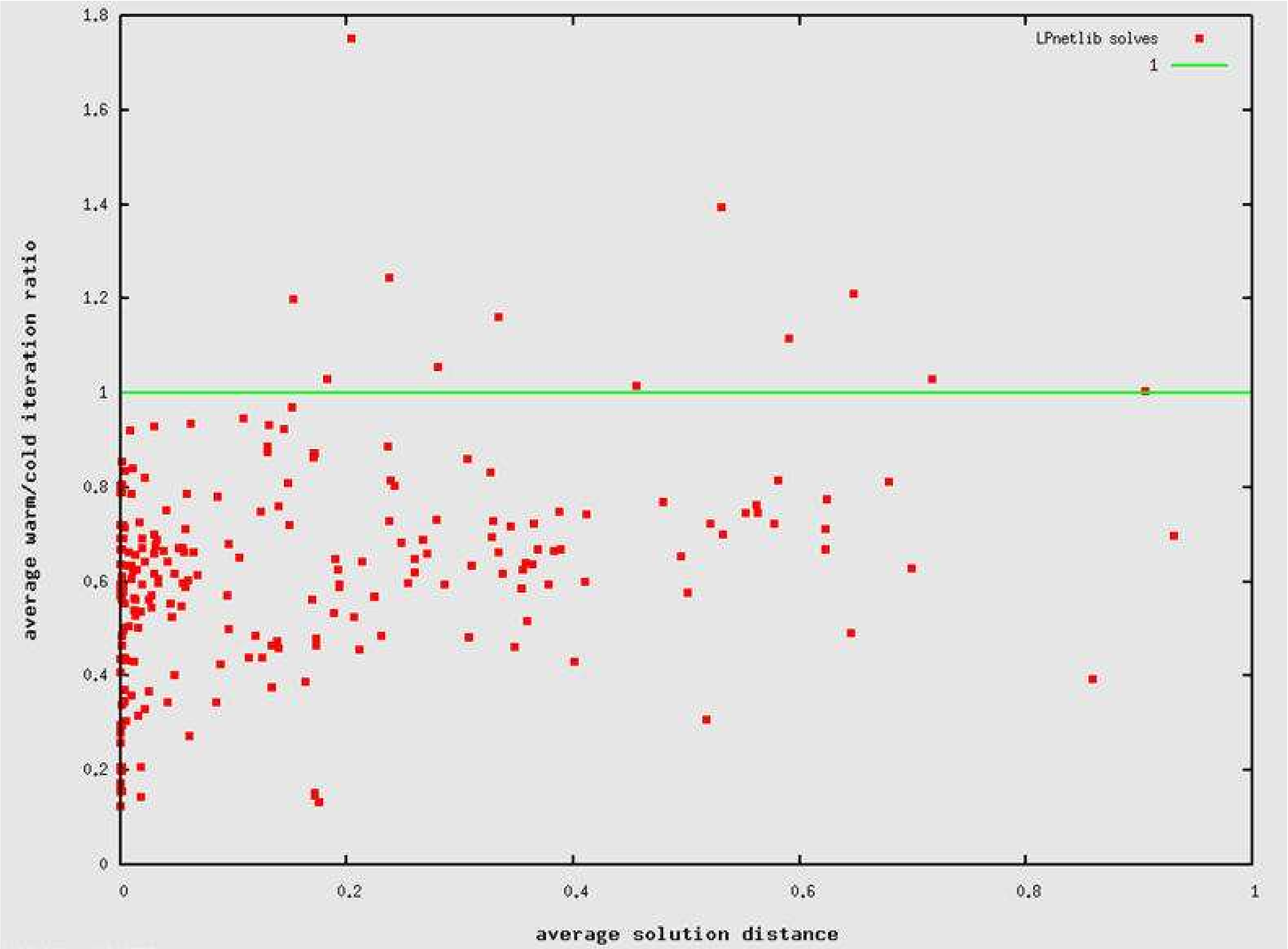
$$\text{LP}(\gamma,\delta) \qquad \underset{x,\,r}{\text{minimize}} \quad c^T x + \tfrac{1}{2}\|\gamma x\|^2 + \tfrac{1}{2}\|r\|^2$$
$$\text{subject to} \quad Ax + \delta r = b, \qquad \ell \le x \le u$$

Regularized LP $\quad \gamma = \delta = 10^{-3}$

PDCO with Cholesky on $AD^2A^T + \delta^2 I$

- LPnetlib problems with 5 random perturbations to A, b, or c
  (cf. Benson and Shanno 2005)
- Smaller problems ($<$ 100KB): 45 runs for each problem
- Compare Zoom to single solve

Results: Warm-starting IPMs



PDCO iterations (warm/cold) vs. perturbation to x, y

Zoom theory

Regularized LP:

$$\begin{array}{llll} \text{RLP} & \underset{x,r,x_1,x_2}{\text{minimize}} & c^T x + \frac{1}{2}\|\gamma x\|^2 + d^T r + \frac{1}{2}\|r\|^2 + c_1^T x_1 + c_2^T x_2 & \\ & \text{subject to} & Ax + \delta r = b & : y \\ & & x - x_1 = \ell & : z_1 \\ & & -x - x_2 = -u & : z_2 \\ & & x_1,\ x_2 \geq 0 & \end{array}$$

Suppose $(\widetilde{x}, \widetilde{y}, \widetilde{z}_1, \widetilde{z}_2, \widetilde{x}_1, \widetilde{x}_2, \widetilde{r})$ is an approximate solution

Redefine problem with

$$\begin{aligned} x &= \widetilde{x} + dx \\ r &= \widetilde{r} + dr \end{aligned}$$

Zoom theory

RLP′

$$
\begin{array}{lll}
\underset{dx,dr,x_1,x_2}{\text{minimize}} & c^T dx + \frac{1}{2}\|\gamma dx\|^2 + \cdots & \\
\text{subject to} & A dx + \delta dr = \widetilde{b} & : y \\
 & dx - x_1 = \tilde{\ell} & : z_1 \\
 & -dx - x_2 = -\tilde{u} & : z_2 \\
 & x_1,\ x_2 \geq 0 &
\end{array}
$$

where

$$
\begin{aligned}
\widetilde{b} &= b - A\widetilde{x} - \delta\widetilde{r} \\
\tilde{\ell} &= \ell - \widetilde{x} \\
\tilde{u} &= u - \widetilde{x}
\end{aligned}
$$

Zoom theory

Add Lagrangian terms

$$\widetilde{y}^T(\widetilde{b} - A dx - \delta dr) \qquad \widetilde{z}_1^T(\tilde{\ell} - dx + x_1) \qquad \widetilde{z}_2^T(-\tilde{u} + dx + x_2)$$

to objective:

RLP″

$$\begin{aligned} \underset{dx,dr,x_1,x_2}{\text{minimize}} \quad & \widetilde{c}^T dx + \tfrac{1}{2}\|\gamma dx\|^2 + \widetilde{d}^T dr + \tfrac{1}{2}\|dr\|^2 + \widetilde{c}_1^T dx_1 + \widetilde{c}_2^T dx_2 \\ \text{subject to} \quad & A dx + \delta dr = \widetilde{b} \qquad && : dy \\ & dx - x_1 = \tilde{\ell} \qquad && : dz_1 \\ & -dx - x_2 = -\tilde{u} \qquad && : dz_2 \\ & x_1,\ x_2 \geq 0 \end{aligned}$$

Same form as original RLP

Primal **and** dual variables are **small**

Next steps

- Multiple Zooms?
- How much is attributable to Zoom, to scaling?
- Explain outliers
  (e.g. Check size of residuals to decide Zoom scaling)

Conclusions

- Minor changes to existing primal-dual algorithms
- Zoom time reduced 40–60%