Basis Pursuit Denoising and the Dantzig Selector

West Coast Optimization Meeting

University of Washington

Seattle, WA, April 28–29, 2007

Michael Friedlander and Michael Saunders

Dept of Computer Science
University of British Columbia
Vancouver, BC V6K 2C6
mpf@cs.ubc.ca

Dept of Management Sci & Eng
Stanford University
Stanford, CA 94305-4026
saunders@stanford.edu

Abstract

Many imaging and compressed sensing applications seek **sparse solutions to under-determined least-squares problems**. The Lasso and Basis Pursuit Denoising (BPDN) approaches of bounding the 1-norm of the solution have led to several computational algorithms.

Abstract

Many imaging and compressed sensing applications seek **sparse solutions to under-determined least-squares problems**. The Lasso and Basis Pursuit Denoising (BPDN) approaches of bounding the 1-norm of the solution have led to several computational algorithms.

PDCO uses an interior method to handle general linear constraints and bounds. Homotopy, LARS, OMP, and STOMP are specialized active-set methods for handling the implicit bounds. l1_ls and GPSR are further recent entries in the $\ell 1$**-regularized least-squares competition**, both based on bound-constrained optimization.

Abstract

Many imaging and compressed sensing applications seek **sparse solutions to under-determined least-squares problems**. The Lasso and Basis Pursuit Denoising (BPDN) approaches of bounding the 1-norm of the solution have led to several computational algorithms.

PDCO uses an interior method to handle general linear constraints and bounds. Homotopy, LARS, OMP, and STOMP are specialized active-set methods for handling the implicit bounds. l1_ls and GPSR are further recent entries in the $\ell 1$**-regularized least-squares competition**, both based on bound-constrained optimization.

The Dantzig Selector of Candes and Tao is promising in its production of **sparse solutions using only linear programming**. Again, interior or active-set (simplex) methods may be used. We compare the BPDN and DS approaches via their dual problems and some numerical examples.

Sparse x

Lasso(t)   Tibshirani 1996

$$\min_x \ \tfrac{1}{2}\|b - Ax\|_2^2 \quad \text{s.t.} \quad \|x\|_1 \le t$$

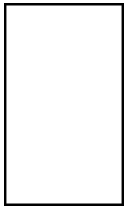
$A = \square$

explicit

Sparse x

Lasso(t)   Tibshirani 1996

$$\min_x \ \tfrac{1}{2}\|b - Ax\|_2^2 \quad \text{s.t.} \quad \|x\|_1 \le t$$

$A =$ 

explicit

Basis Pursuit   Chen, Donoho & S 2001

$$\min_x \ \|x\|_1 \quad \text{s.t.} \quad Ax = b$$

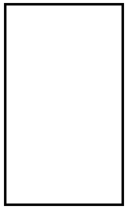
$A =$ 

fast operator

Sparse x

Lasso(t)   Tibshirani 1996

$$\min_x \tfrac{1}{2}\|b - Ax\|_2^2 \quad \text{s.t.} \quad \|x\|_1 \le t$$

$A =$ 

explicit

BPDN(λ)   Chen, Donoho & S 2001

$$\min_x \tfrac{1}{2}\|b - Ax\|_2^2 \ + \ \lambda\|x\|_1$$

$A =$ 

fast operator

BP and BPDN Algorithms

OMP	Davis, Mallat et al 1997	Greedy
BPDN-interior	Chen, Donoho & S, 1998	Interior, CG
PDSCO, PDCO	S 1997, 2002	Interior, LSQR
BCR	Sardy, Bruce & Tseng 2000	Orthogonal blocks
Homotopy	Osborne et al 2000	Active-set, all λ
LARS	Efron, Hastie et al 2004	Active-set, all λ
STOMP	Donoho, Tsaig et al 2006	Double greedy
l1_ls	Kim, Koh et al 2007	Primal barrier, PCG
GPSR	Figueiredo, Nowak & Wright 2007	Gradient-projection

Basis Pursuit Denoising (BPDN)

Chen, Donoho and S 1998

Pure LS

$$\min_{x,r} \tfrac{1}{2}\|r\|_2^2 \qquad \text{s.t.} \qquad r = b - Ax$$

Basis Pursuit Denoising (BPDN)

Chen, Donoho and S 1998

Pure LS

$$\min_{x,r} \tfrac{1}{2}\|r\|_2^2 \qquad \text{s.t.} \qquad r = b - Ax$$

If x not unique, need regularization

Basis Pursuit Denoising (BPDN)

Chen, Donoho and S 1998

Pure LS

$$\min_{x,r} \tfrac{1}{2}\|r\|_2^2 \qquad \text{s.t.} \qquad r = b - Ax$$

If x not unique, need regularization

BPDN(λ)

$$\min_{x,r} \lambda\|x\|_1 \; + \; \tfrac{1}{2}\|r\|_2^2 \qquad \text{s.t.} \qquad r = b - Ax$$

Basis Pursuit Denoising (BPDN)

Chen, Donoho and S 1998

Pure LS

$$\min_{x,r} \tfrac{1}{2}\|r\|_2^2 \qquad \text{s.t.} \qquad r = b - Ax$$

If x not unique, need regularization

BPDN(λ)

$$\min_{x,r} \lambda\|x\|_1 \; + \; \tfrac{1}{2}\|r\|_2^2 \qquad \text{s.t.} \qquad r = b - Ax$$

smaller $\|x\|_1$ bigger $\|r\|_2$

The Dantzig Selector (DS)

Candès and Tao 2007

Pure LS

$$A^T r \quad = 0, \quad r = b - Ax$$

The Dantzig Selector (DS)

Candès and Tao 2007

Pure LS

$$A^T r \quad = 0, \quad r = b - Ax$$

Plausible regularization

$$\min_{x,r} \|x\|_1 \quad \text{s.t.} \quad A^T r \quad = 0, \quad r = b - Ax$$

The Dantzig Selector (DS)

Candès and Tao 2007

Pure LS

$$A^T r \quad = 0, \quad r = b - Ax$$

Plausible regularization

$$\min_{x,r} \|x\|_1 \quad \text{s.t.} \quad A^T r \quad = 0, \quad r = b - Ax$$

DS(λ)

$$\min_{x,r} \|x\|_1 \quad \text{s.t.} \quad \|A^T r\|_\infty \le \lambda, \quad r = b - Ax$$

The Dantzig Selector (DS)

Candès and Tao 2007

Pure LS

$$A^T r \quad = 0, \quad r = b - Ax$$

Plausible regularization

$$\min_{x,r} \|x\|_1 \quad \text{s.t.} \quad A^T r \quad = 0, \quad r = b - Ax$$

DS(λ)

$$\min_{x,r} \|x\|_1 \quad \text{s.t.} \quad \|A^T r\|_\infty \le \lambda, \quad r = b - Ax$$

smaller $\|x\|_1$ bigger $\|A^T r\|_\infty$

BP Denoising and the Dantzig Selector

Dual problems

BPDN(λ)

$$\min_{x,r} \lambda\|x\|_1 + \tfrac{1}{2}\|r\|_2^2 \qquad \text{s.t.} \qquad r = b - Ax \qquad \text{QP}$$

BP Denoising and the Dantzig Selector

Dual problems

BPDN(λ)

$$\min_{x,r} \lambda\|x\|_1 + \tfrac{1}{2}\|r\|_2^2 \qquad \text{s.t.} \qquad r = b - Ax \qquad \text{QP}$$

DS(λ)

$$\min_{x,r} \|x\|_1 \quad \text{s.t.} \quad \|A^T r\|_\infty \le \lambda, \quad r = b - Ax \qquad \text{LP}$$

BP Denoising and the Dantzig Selector

Dual problems

BPDN(λ)

$$\min_{x,r} \lambda\|x\|_1 + \tfrac{1}{2}\|r\|_2^2 \quad \text{s.t.} \quad r = b - Ax \qquad \text{QP}$$

DS(λ)

$$\min_{x,r} \|x\|_1 \quad \text{s.t.} \quad \|A^T r\|_\infty \le \lambda, \quad r = b - Ax \qquad \text{LP}$$

BPdual(λ)

$$\min_{r} -b^T r + \tfrac{1}{2}\|r\|_2^2 \quad \text{s.t.} \quad \|A^T r\|_\infty \le \lambda$$

BP Denoising and the Dantzig Selector

Dual problems

BPDN(λ)

$$\min_{x,r} \ \lambda\|x\|_1 \ + \ \tfrac{1}{2}\|r\|_2^2 \qquad \text{s.t.} \qquad r = b - Ax \qquad \text{QP}$$

DS(λ)

$$\min_{x,r} \ \|x\|_1 \quad \text{s.t.} \quad \|A^T r\|_\infty \leq \lambda, \quad r = b - Ax \qquad \text{LP}$$

BPdual(λ)

$$\min_{r} \ -b^T r + \tfrac{1}{2}\|r\|_2^2 \quad \text{s.t.} \quad \|A^T r\|_\infty \leq \lambda$$

DSdual(λ)

$$\min_{r,z} \ -b^T r + \lambda\|z\|_1 \quad \text{s.t.} \quad \|A^T r\|_\infty \leq \lambda, \quad r = Az$$

BPDN(λ) implementation

Chen, Donoho & S 1998

$$\min_{v,w,r} \quad \lambda \mathbf{1}^T(v+w) + \tfrac{1}{2} r^T r$$

$$\text{s.t.} \quad \begin{bmatrix} A & -A \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix} + r = b, \quad v, w \geq 0$$

2007: Apply PDCO (MATLAB primal-dual interior method)

BPDN(λ) implementation

Chen, Donoho & S 1998

$$\min_{v,w,r} \quad \lambda \mathbf{1}^T(v+w) + \tfrac{1}{2} r^T r$$

$$\text{s.t.} \quad \begin{bmatrix} A & -A \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix} + r = b, \quad v, w \ge 0$$

2007: Apply PDCO (MATLAB primal-dual interior method)

Dense A in test problems $\Rightarrow$ Dense Cholesky

Double-handling of A

$\begin{bmatrix} A & -A \end{bmatrix} \begin{bmatrix} D_1 & \\ & D_2 \end{bmatrix} \begin{bmatrix} A^T \\ -A^T \end{bmatrix}$ could be coded as $A(D_1 + D_2)A^T$

DS(λ) implementation

Candès and Tao 2007

$$\begin{aligned} \min_{x,u} \quad & \mathbf{1}^T u \\ \text{s.t.} \quad & -u \le x \le u, \\ & -\lambda \mathbf{1} \le A^T(b - Ax) \le \lambda \mathbf{1}, \end{aligned}$$

Apply l1dantzig_pd (MATLAB primal-dual interior method)
Romberg 2005 Part of ℓ_1-magic Candes 2006

DS(λ) implementation

Candès and Tao 2007

$$\min_{x,u} \quad \mathbf{1}^T u$$

s.t. $\quad -u \le x \le u,$

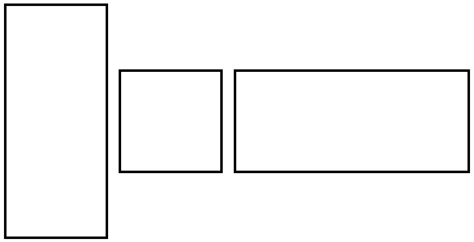
$$-\lambda \mathbf{1} \le \quad A^T(b - Ax) \quad \le \lambda \mathbf{1},$$

Apply l1dantzig_pd (MATLAB primal-dual interior method)

Romberg 2005 Part of ℓ_1-magic Candes 2006

Dense A in test problems

Dense Cholesky on $A^T(AD_1A^T)A + D_2$ (much bigger)

 $+ D_2$

Two other DS LP implementations

Introduce $s = -A^T r$

DS1

Interior

$$\min_{v,w,s} \quad \mathbf{1}^T(v+w)$$

$$\text{s.t.} \quad \begin{bmatrix} A^TA & -A^TA & I \end{bmatrix} \begin{bmatrix} v \\ w \\ s \end{bmatrix} = A^Tb, \quad v, w \geq 0, \quad \|s\|_\infty \leq \lambda$$

Two other DS LP implementations

Introduce $s = -A^T r$

DS1

Interior

$$\min_{v,w,s} \quad \mathbf{1}^T(v+w)$$

$$\text{s.t.} \quad \begin{bmatrix} A^TA & -A^TA & I \end{bmatrix} \begin{bmatrix} v \\ w \\ s \end{bmatrix} = A^Tb, \quad v, w \geq 0, \quad \|s\|_\infty \leq \lambda$$

DS2

Interior, Simplex

$$\min_{v,w,r,s} \quad \mathbf{1}^T(v+w)$$

$$\text{s.t.} \quad \begin{bmatrix} A & -A & I & \\ & & A^T & I \end{bmatrix} \begin{bmatrix} v \\ w \\ r \\ s \end{bmatrix} = \begin{bmatrix} b \\ 0 \end{bmatrix}, \quad v, w \geq 0, \quad \|s\|_\infty \leq \lambda$$

Test data

A, b depend on dimensions m, n, T

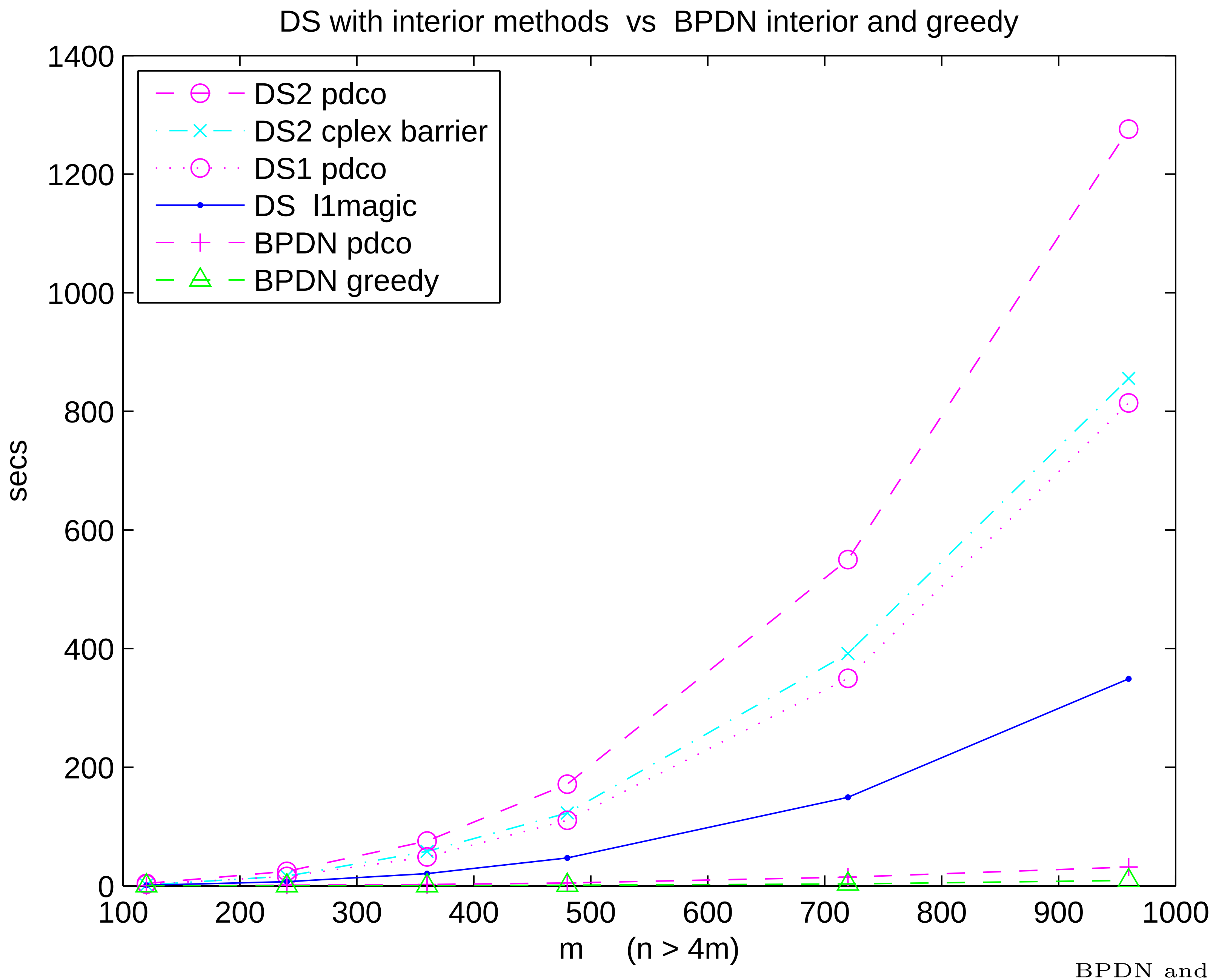
```
rand ('state',0);                  % initialize generators
randn('state',0);
x     = zeros(n,1);                % random +/-1 signal
q     = randperm(m);
x(q(1:T)) = sign(randn(T,1));
[A,R] = qr(randn(n,m),0);
A     = A';                        % m x n measurement mtx
sigma = 0.005;
b     = A*x + sigma*randn(m,1);    % noisy observations
```

A dense $\quad AA^T = I \quad$ T components $x_j \approx \pm 1$

For example $\quad m = 500 \quad n = 2000 \quad T = 80$

$\lambda =$ 3e-3

DS vs BPDN


DS with interior methods vs BPDN interior and greedy
DS2 pdco
DS2 cplex barrier
DS1 pdco
DS l1magic
BPDN pdco
BPDN greedy
secs
m (n > 4m)
0
200
400
600
800
1000
1200
1400
100
200
300
400
500
600
700
800
900
1000

CPLEX dual simplex on DS2

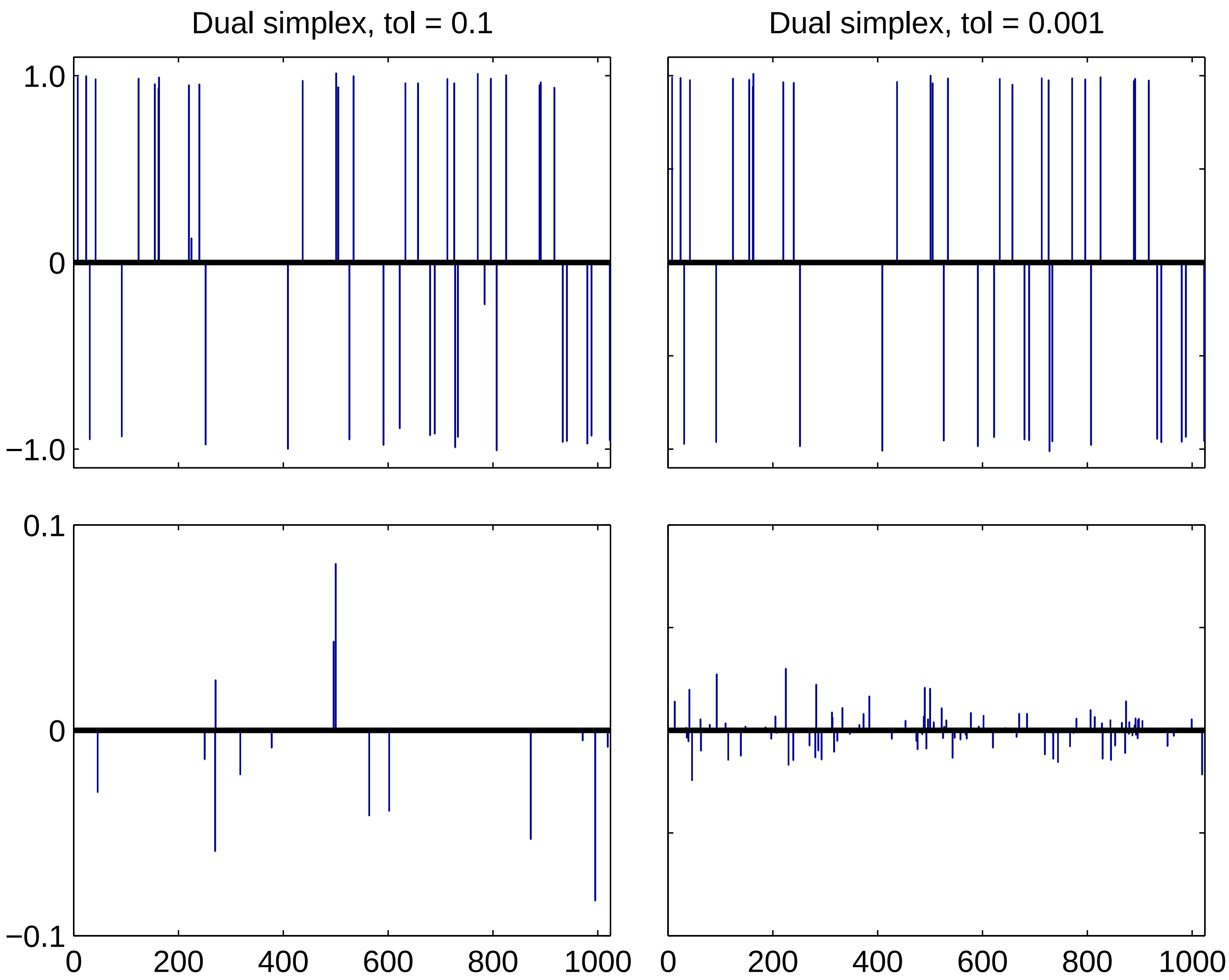
with loose and tight tols

sizes			tol $= 0.1$			tol $= 0.001$		
m	n	T	itns	time	$\|S\|$	itns	time	$\|S\|$
120	512	20	20	0.1	20	86	0.2	63
240	1024	40	58	0.4	56	405	2.3	150
360	1536	60	187	2.3	134	1231	15.1	215
480	2048	80	163	3.4	122	1277	26.7	275
720	3072	120	356	15.3	223	3006	146.6	420
960	4096	160	965	80.2	414	9229	891.6	567

Too many simplex iterations, too many $x_j \neq 0$

CPLEX dual simplex on DS2

Large and small $|x_j|$



More small values $\Rightarrow$ more simplex iterations and more time per iteration

References

- E. Candès, ℓ_1-*magic*, http://www.l1-magic.org/, 2006.
- E. Candès and T. Tao, *The Dantzig selector: Statistical estimation when* $p \gg n$, Annals of Statistics, to appear (2007).
- S. S. Chen, D. L. Donoho, and M. A. Saunders, *Atomic decomposition by basis pursuit*, SIAM Review, 43 (2001).
- B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, *Least angle regression*, Ann. Statist., 32 (2004).
- M. R. Osborne, B. Presnell, and B. A. Turlach, *The Lasso and its dual*, J. of Computational and Graphical Statistics, 9 (2000).
- M. R. Osborne, B. Presnell, and B. A. Turlach, *A new approach to variable selection in least squares problems*, IMA J. of Numerical Analysis, 20 (2000).
- M. A. Saunders, *PDCO*. Matlab software for convex optimization, http://www.stanford.edu/group/SOL/software/pdco.html, 2005.
- R. Tibshirani, *Regression shrinkage and selection via the lasso*, J. Roy. Statist. Soc. Ser. B, 58 (1996).
- Y. Tsaig, *Sparse Solution of Underdetermined Linear Systems: Algorithms and Applications*, PhD thesis, Stanford University, 2007.

Many thanks

Michael

Many thanks

Michael

and

Jim, Terry, Paul