Computing approximate PageRank vectors by Basis Pursuit Denoising

Michael Saunders
Systems Optimization Laboratory, Stanford University

Joint work with
Holly Jin, LinkedIn Corp

SIAM Annual Meeting
San Diego, July 7–11, 2008

Outline

Abstract

The PageRank eigenvector problem involves a square system $Ax = b$ in which x is naturally nonnegative and somewhat sparse (depending on b). We seek an approximate x that is nonnegative and extremely sparse. We experiment with an active-set optimization method designed for the dual of the BPDN problem, and find that it tends to extract the important elements of x in a greedy fashion.

Supported by the Office of Naval Research

Introduction

Many applications (statistics, signal analysis, imaging) seek *sparse solutions to square or rectangular systems*

$$Ax \approx b \qquad x \text{ sparse}$$

Introduction

Many applications (statistics, signal analysis, imaging) seek *sparse solutions to square or rectangular systems*

$$Ax \approx b \qquad x \text{ sparse}$$

Basis Pursuit Denoising (BPDN) seeks sparsity by solving

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

Introduction

Many applications (statistics, signal analysis, imaging) seek *sparse solutions to square or rectangular systems*

$$Ax \approx b \qquad x \text{ sparse}$$

Basis Pursuit Denoising (BPDN) seeks sparsity by solving

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

Personalized PageRank eigenvectors involve square systems

$$(I - \alpha H^T)x = v \qquad H, v \text{ sparse}$$

where $\quad 0 < \alpha < 1 \qquad He \leq e \qquad x \geq 0$, sparse

Introduction

Many applications (statistics, signal analysis, imaging) seek *sparse solutions to square or rectangular systems*

$$Ax \approx b \qquad x \text{ sparse}$$

Basis Pursuit Denoising (BPDN) seeks sparsity by solving

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

Personalized PageRank eigenvectors involve square systems

$$(I - \alpha H^T)x = v \qquad H, v \text{ sparse}$$

where $\quad 0 < \alpha < 1 \qquad He \leq e \qquad x \geq 0$, sparse

Perhaps BPDN can find a very sparse approximate x

PageRank

PageRank

Langville and Meyer 2006

$$H = \begin{pmatrix} 1/3 & & 1/3 & 1/3 \\ & 1/2 & 1/2 & \\ 0 & 0 & 0 & 0 \\ & 1/2 & & 1/2 \end{pmatrix}$$ Hyperlink matrix $He \le e$

PageRank

Langville and Meyer 2006

$$H = \begin{pmatrix} 1/3 & & 1/3 & 1/3 \\ & 1/2 & 1/2 & \\ 0 & 0 & 0 & 0 \\ & 1/2 & & 1/2 \end{pmatrix}$$ Hyperlink matrix $He \leq e$

$v = \frac{1}{n}e$ or e_i Teleportation vector or personalization vector $e^T v = 1$

PageRank

Langville and Meyer 2006

$$H = \begin{pmatrix} 1/3 & & 1/3 & 1/3 \\ & 1/2 & 1/2 & \\ 0 & 0 & 0 & 0 \\ & 1/2 & & 1/2 \end{pmatrix}$$ Hyperlink matrix $He \leq e$

$v = \frac{1}{n}e$ or e_i Teleportation vector or personalization vector $e^T v = 1$

$S = H + av^T$ Stochastic matrix $Se = e$

PageRank

Langville and Meyer 2006

$H = \begin{pmatrix} 1/3 & & 1/3 & 1/3 \\ & 1/2 & 1/2 & \\ 0 & 0 & 0 & 0 \\ & 1/2 & & 1/2 \end{pmatrix}$	Hyperlink matrix	$He \leq e$
$v = \frac{1}{n}e$ or e_i	Teleportation vector or personalization vector	$e^T v = 1$
$S = H + av^T$	Stochastic matrix	$Se = e$
$G = \alpha S + (1-\alpha)ev^T$	Google matrix $\alpha = 0.85$ say	$Ge = e$

PageRank

Langville and Meyer 2006

$H = \begin{pmatrix} 1/3 & & 1/3 & 1/3 \\ & 1/2 & 1/2 & \\ 0 & 0 & 0 & 0 \\ & 1/2 & & 1/2 \end{pmatrix}$	Hyperlink matrix	$He \leq e$
$v = \frac{1}{n}e$ or e_i	Teleportation vector or personalization vector	$e^T v = 1$
$S = H + av^T$	Stochastic matrix	$Se = e$
$G = \alpha S + (1-\alpha)ev^T$	Google matrix $\alpha = 0.85$ say	$Ge = e$

eigenvector	$\equiv$	linear system
$G^T x = x$	$\equiv$	$(I - \alpha H^T)x = v$

PageRank

Langville and Meyer 2006

$H = \begin{pmatrix} 1/3 & & 1/3 & 1/3 \\ & 1/2 & 1/2 & \\ 0 & 0 & 0 & 0 \\ & 1/2 & & 1/2 \end{pmatrix}$	Hyperlink matrix	$He \le e$
$v = \frac{1}{n} e$ or e_i	Teleportation vector or personalization vector	$e^T v = 1$
$S = H + av^T$	Stochastic matrix	$Se = e$
$G = \alpha S + (1-\alpha) e v^T$	Google matrix $\alpha = 0.85$ say	$Ge = e$

eigenvector	$\equiv$	linear system
$G^T x = x$	$\equiv$	$(I - \alpha H^T) x = v$

In both cases, $x \leftarrow x / e^T x$ (and $x \ge 0$)

Web matrices H

Name	n (pages)	nnz (links)	i	$(v = e_i)$
csStanford	9914	36854	4	cs.stanford.edu
Stanford	275689	1623817	23036	cs.stanford.edu
			6753	calendus.stanford.edu
Stanford-Berkeley	683446	7583376	6753	unknown

All data collected around 2001

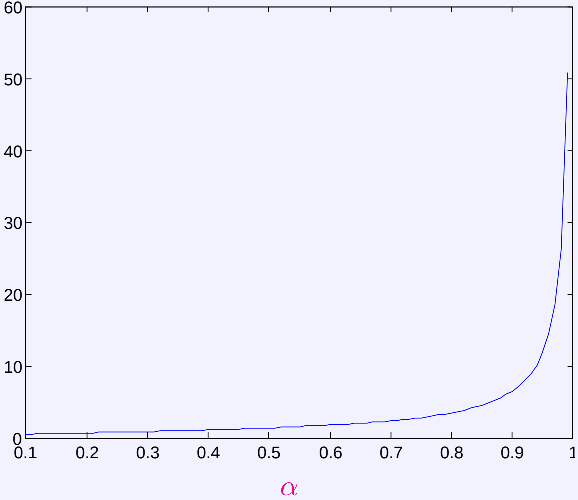
Cleve Moler
Sep Kamvar
David Gleich

$$(I - \alpha H^T)x = v$$

Power method on $G^T x = x$

Stanford, $n = 275000$, $v = e_{6753}$ (calendus.stanford.edu)

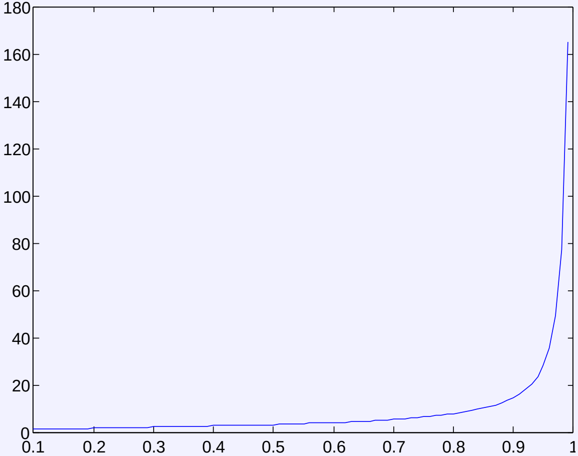
Time (secs) vs $\alpha = 0.1 : 0.01 : 0.99$



Power method

Stanford-Berkeley, $n = 683000$, $v = e_{6753}$

Time (secs) vs $\alpha = 0.1 : 0.01 : 0.99$

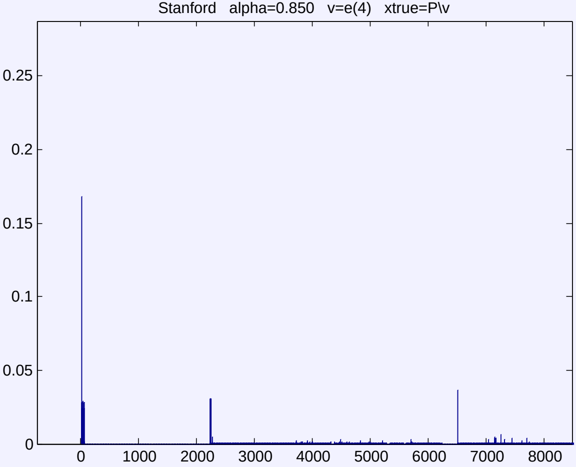


α

Exact $x(\alpha)$

csStanford, $n = 9914$, $v = e_4$ (cs.stanford.edu)

$x_j(\alpha), \;\; j = 1:n$

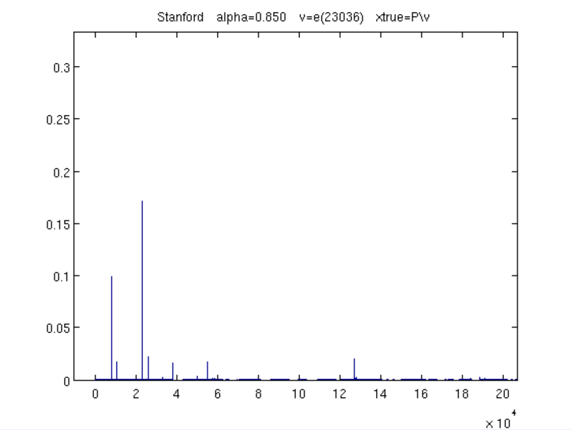


$\alpha = 0.85$

Exact $x(\alpha)$

Stanford, $n = 275000$, $v = e_{23036}$ (cs.stanford.edu)

$x_j(\alpha), \ j = 1:n$

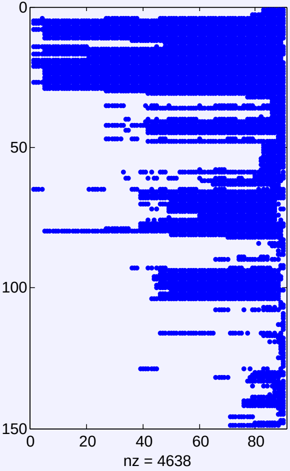


$\alpha = 0.85$

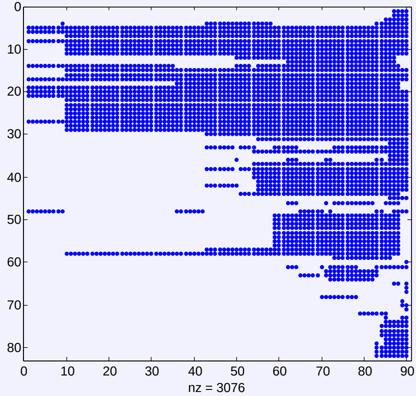
Sparsity of $x(\alpha)$

Stanford, $n = 275000$, $v = e_{23036}$ (cs.stanford.edu)

$x > 250/n$



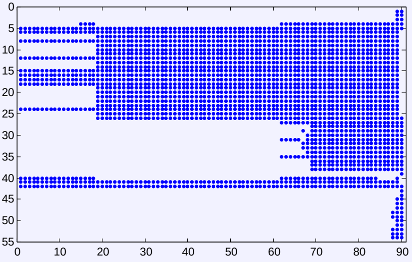
$x > 500/n$



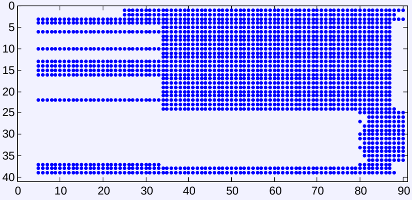
$\alpha = 0.1 : 0.01 : 0.99$ (90 values)

$n = 275000$

$x > 1000/n$



$x > 2000/n$



$\alpha = 0.1 : 0.01 : 0.99$ (90 values)

Basis Pursuit Denoising

min L1 $\Rightarrow$ sparse x

Lasso(ν) Tibshirani 1996

$$\min \ \tfrac{1}{2}\|Ax - b\|^2 \quad \text{st} \quad \|x\|_1 \leq \nu \qquad\qquad A = \boxed{}$$

min L1 $\Rightarrow$ sparse x

Lasso(ν) Tibshirani 1996

$$\min \ \tfrac{1}{2}\|Ax-b\|^2 \quad \text{st} \quad \|x\|_1 \le \nu$$

$A =$ [tall rectangle]

Basis Pursuit Chen, Donoho & S 1998

$$\min \ \|x\|_1 \quad \text{st} \quad Ax = b$$

$A =$ fast operator

(Ax, A^Ty are efficient)

min L1 $\Rightarrow$ sparse x

Lasso(ν) Tibshirani 1996

$$\min \tfrac{1}{2}\|Ax - b\|^2 \quad \text{st} \quad \|x\|_1 \le \nu \qquad\qquad A = \boxed{}$$

Basis Pursuit Chen, Donoho & S 1998

$$\min \|x\|_1 \quad \text{st} \quad Ax = b \qquad\qquad A = \boxed{\text{fast operator}}$$

(Ax, $A^T y$ are efficient)

BPDN(λ) Same paper

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

fast operator, any shape

BP and BPDN algorithms

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

OMP	Davis, Mallat et al 1997	Greedy
BPDN-interior	Chen, Donoho & S, 1998, 2001	Interior, CG
PDSCO, PDCO	Saunders 1997, 2002	Interior, LSQR
BCR	Sardy, Bruce & Tseng 2000	Orthogonal blocks
Homotopy	Osborne et al 2000	Active-set, all λ
LARS	Efron, Hastie, Tibshirani 2004	Active-set, all λ
STOMP	Donoho, Tsaig, et al 2006	Double greedy
l1_ls	Kim, Koh, Lustig, Boyd et al 2007	Primal barrier, PCG
GPSR	Figueiredo, Nowak & Wright 2007	Gradient Projection
BCLS	Friedlander 2006	Projection, LSQR
SPGL1	van den Berg & Friedlander 2007	Spectral GP, all λ
BPdual	Friedlander & S 2007	Active-set on dual
LPdual	Friedlander & S 2007	For $x \geq 0$

$\|x\|_1$ when $x \geq 0$

Suggests regularized LP problems:

$$\text{LPprimal}(\lambda) \quad \min_{x,\,y} \quad e^T x + \tfrac{1}{2}\lambda\|y\|^2 \quad \text{st} \quad Ax + \lambda y = b,\ x \geq 0$$

$\|x\|_1$ when $x \geq 0$

Suggests regularized LP problems:

$$\text{LPprimal}(\lambda) \quad \min_{x,\,y} \quad e^T x + \tfrac{1}{2}\lambda\|y\|^2 \quad \text{st} \quad Ax + \lambda y = b,\ x \geq 0$$

$$\text{LPdual}(\lambda) \quad \min_{y} \ -b^T y + \tfrac{1}{2}\lambda\|y\|^2 \quad \text{st} \quad A^T y \leq e$$

$\|x\|_1$ when $x \geq 0$

Suggests regularized LP problems:

$$\text{LPprimal}(\lambda) \quad \min_{x,\,y} \quad e^T x + \tfrac{1}{2}\lambda\|y\|^2 \quad \text{st} \quad Ax + \lambda y = b,\ x \geq 0$$

$$\text{LPdual}(\lambda) \quad \min_{y} \ -b^T y + \tfrac{1}{2}\lambda\|y\|^2 \quad \text{st} \quad A^T y \leq e$$

Friedlander and S 2007: New active-set MATLAB codes

LPdual solver for $Ax \approx b$, $x \geq 0$

$$\min_{y} \; -b^T y + \tfrac{1}{2}\lambda\|y\|^2 \quad \text{st} \quad A^T y \leq e$$

LPdual solver for $Ax \approx b$, $x \geq 0$

$$\min_y \; -b^T y + \tfrac{1}{2}\lambda \|y\|^2 \quad \text{st} \quad A^T y \leq e$$

A few active constraints: $B^T y = e$

Initially B is empty, $y = 0$

Select columns of B in almost *greedy* manner

LPdual solver for $Ax \approx b$, $x \geq 0$

$$\min_y \; -b^T y + \tfrac{1}{2}\lambda \|y\|^2 \quad \text{st} \quad A^T y \leq e$$

A few active constraints: $B^T y = e$
Initially B is empty, $y = 0$
Select columns of B in almost *greedy* manner

Main work per itn:
Solve $\min \|Bx - g\|$
Form $dy = (g - Bx)/\lambda$
Form $dz = A^T dy$

LPdual solver for $Ax \approx b$, $x \geq 0$

$$\min_y \; -b^T y + \tfrac{1}{2}\lambda \|y\|^2 \quad \text{st} \quad A^T y \leq e$$

A few active constraints: $B^T y = e$

Initially B is empty, $y = 0$

Select columns of B in almost *greedy* manner

Main work per itn:

Solve $\min \|Bx - g\|$ — Update $QB = \binom{R}{0}$ without Q

Form $dy = (g - Bx)/\lambda$

Form $dz = A^T dy$

Sparse PageRank with Neumann Series

Neumann series

$$(I - \alpha H^T)x = v \qquad (*)$$

$$\begin{aligned} x &= (I - \alpha H^T)^{-1} v \\ &= (I + \alpha H^T + (\alpha H^T)^2 + \cdots)v \end{aligned}$$

A sum of nonnegative terms

Neumann series

$$(I - \alpha H^T)x = v \qquad (*)$$

$$\begin{aligned} x &= (I - \alpha H^T)^{-1} v \\ &= (I + \alpha H^T + (\alpha H^T)^2 + \cdots) v \end{aligned}$$

A sum of nonnegative terms

- Equivalent to Jacobi's method on (*)   Thanks David!

Neumann series

$$(I - \alpha H^T)x = v \qquad (*)$$

$$\begin{aligned} x &= (I - \alpha H^T)^{-1} v \\ &= (I + \alpha H^T + (\alpha H^T)^2 + \cdots) v \end{aligned}$$

A sum of nonnegative terms

- Equivalent to Jacobi's method on (*) Thanks David!
- $\|x - x_k\|$ decreases monotonically

Neumann series

$$(I - \alpha H^T)x = v \qquad (*)$$

$$\begin{aligned} x &= (I - \alpha H^T)^{-1} v \\ &= (I + \alpha H^T + (\alpha H^T)^2 + \cdots) v \end{aligned}$$

A sum of nonnegative terms

- Equivalent to Jacobi's method on (*)   Thanks David!
- $\|x - x_k\|$ decreases monotonically
- $\|r_k\| \le \|x - x_k\| \le \frac{1+\alpha}{1-\alpha}\|r_k\|$ Error bounded by residual

Neumann series

$$(I - \alpha H^T)x = v \qquad (*)$$

$$\begin{aligned} x &= (I - \alpha H^T)^{-1} v \\ &= (I + \alpha H^T + (\alpha H^T)^2 + \cdots) v \end{aligned}$$

A sum of nonnegative terms

- Equivalent to Jacobi's method on (*) Thanks David!
- $\|x - x_k\|$ decreases monotonically
- $\|r_k\| \le \|x - x_k\| \le \frac{1+\alpha}{1-\alpha}\|r_k\|$ Error bounded by residual
- Eventually small $(x_k)_j \leftarrow 0$, update remainder

Neumann series

$$(I - \alpha H^T)x = v \qquad (*)$$

$$\begin{aligned} x &= (I - \alpha H^T)^{-1} v \\ &= (I + \alpha H^T + (\alpha H^T)^2 + \cdots)v \end{aligned}$$

A sum of nonnegative terms

- Equivalent to Jacobi's method on (*) Thanks David!
- $\|x - x_k\|$ decreases monotonically
- $\|r_k\| \le \|x - x_k\| \le \frac{1+\alpha}{1-\alpha}\|r_k\|$ Error bounded by residual
- Eventually small $(x_k)_j \leftarrow 0$, update remainder

See Gleich and Polito (2006) for sparse Power Method

Sparse PageRank with BPDN (LPdual)

Sparse PageRank

$$(I - \alpha H^T)x = v \qquad v \text{ sparse}$$

Regard as

$$Ax \approx v \qquad v = e_i \text{ say}$$

Sparse PageRank

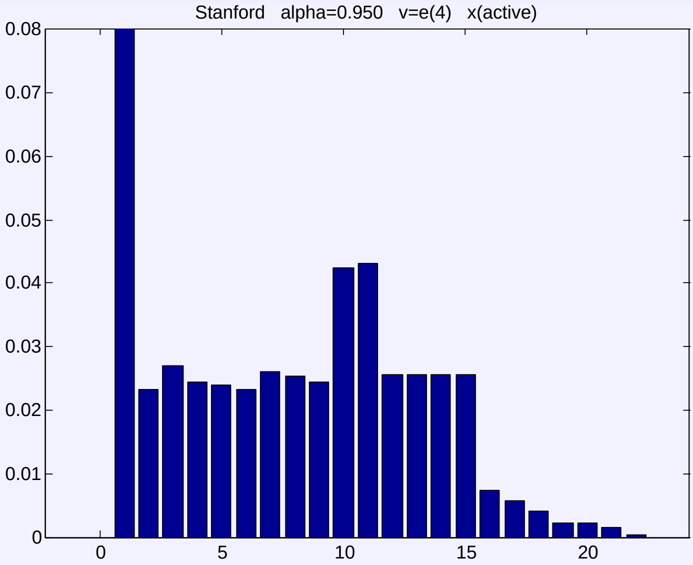
$$(I - \alpha H^T)x = v \qquad v \text{ sparse}$$

Regard as

$$Ax \approx v \qquad v = e_i \text{ say}$$

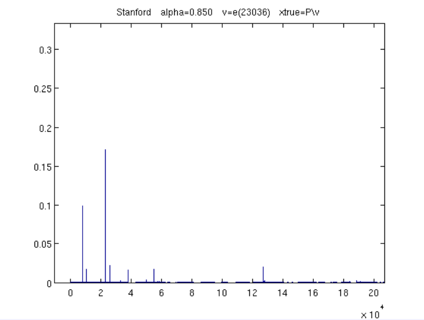
Apply active-set solver LPdual to dual of

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

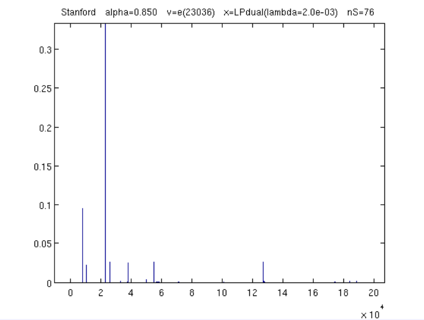

Stanford alpha=0.950 v=e(4) x(active)
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
0
5
10
15
20

Stanford web, $n = 275000$
$\alpha = 0.85$
$v = e_{23036}$ (cs.stanford.edu)

Direct solve:
`xtrue = A\v`

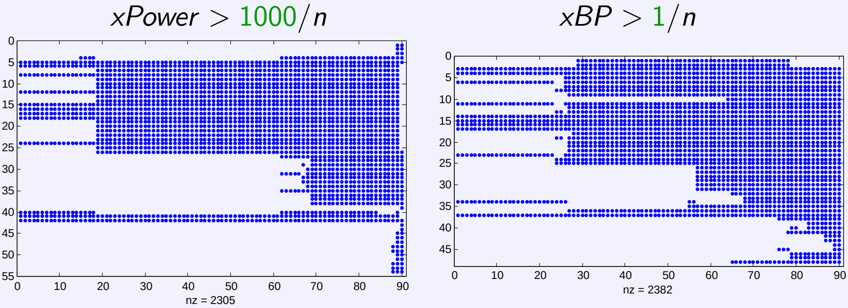


BPDN
$\lambda = 2\text{e-}3$
76 nonzero x_j



Sparsity of $xBP(\alpha)$

Stanford, $n = 275000$, $v = e_{23036}$ (cs.stanford.edu)

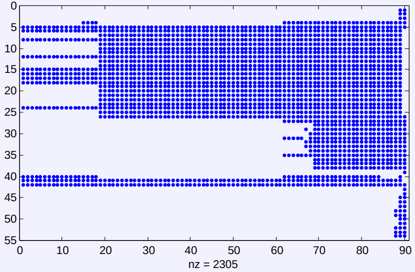


$\alpha = 0.1 : 0.01 : 0.99$ (90 values)

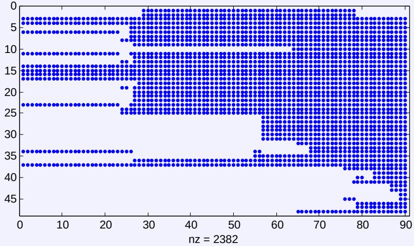
Sparsity of $xBP(\alpha)$

Stanford, $n = 275000$, $v = e_{23036}$ (cs.stanford.edu)

$xPower > 1000/n$



$xBP > 1/n$



$\alpha = 0.1 : 0.01 : 0.99$ (90 values)

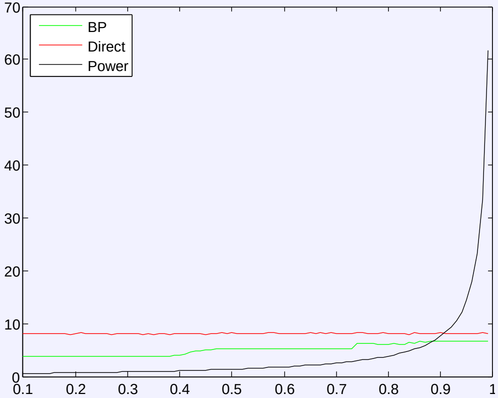
$xPower > 1/n$ has
6000 rows, 155000 nonzeros

$\lambda = 4e\text{-}3$

Power method vs BP

Stanford, $n = 275000$, $v = e_{23036}$ (cs.stanford.edu)

Time (secs) for each α



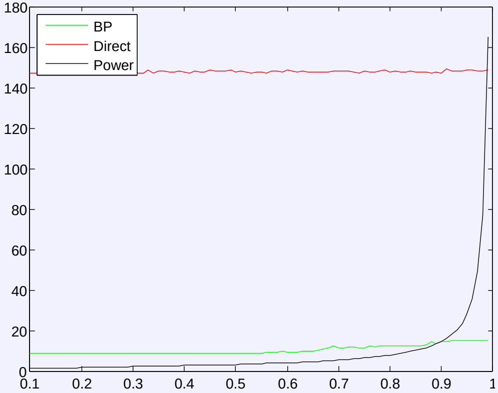
$\alpha = 0.1 : 0.01 : 0.99$ (90 values)

$\lambda = 4\text{e-}3$

Power method vs BP

Stanford-Berkeley, $n = 683000$, $v = e_{6753}$ (unknown)

Time (secs) for each α



$\alpha = 0.1 : 0.01 : 0.99$ (90 values)

$\lambda = 5e\text{-}3$

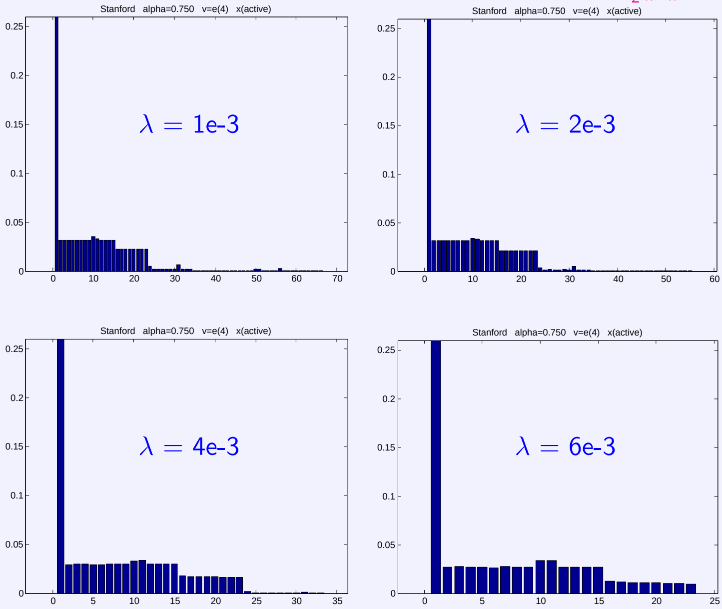
Varying α and λ

$$\min_{x,r} \lambda\|x\|_1 + \tfrac{1}{2}\|r\|^2 \quad \text{st} \quad (I - \alpha H^T)x + r = v$$

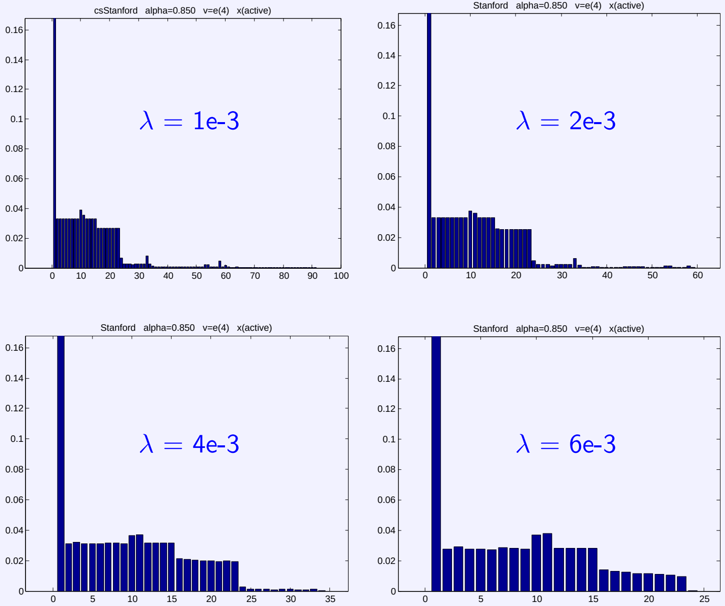
$$\begin{aligned} H &= 9914 \times 9914 \text{ csStanford web} \\ v &= e_4 \text{ (cs.stanford.edu)} \\ \alpha &= 0.75,\ 0.85,\ 0.95 \\ \lambda &= \text{1e-3, 2e-3, 4e-3, 6e-3} \end{aligned}$$

Plot nonzero x_j in the order they are chosen

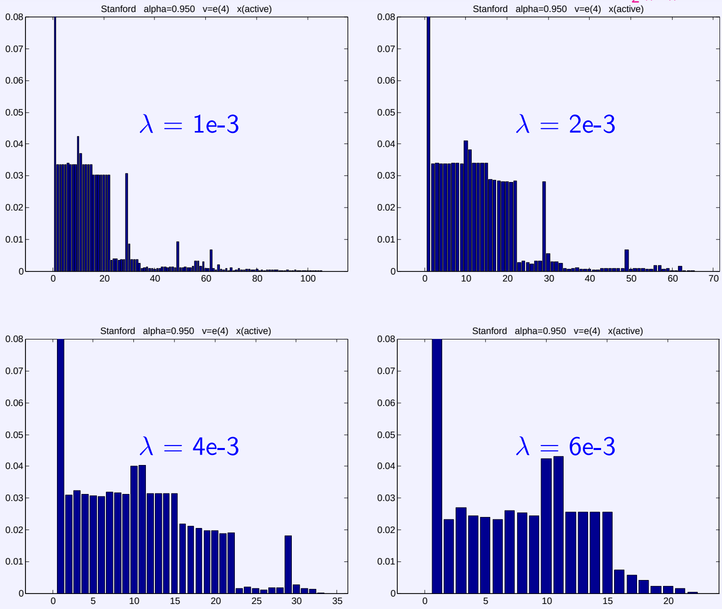
csStanford $v = e_4$ $\alpha = 0.75$ $\min \lambda e^T x + \frac{1}{2}\|r\|^2$



csStanford $v = e_4$ $\alpha = 0.85$ $\min \lambda e^T x + \frac{1}{2}\|r\|^2$



csStanford $v = e_4$ $\alpha = 0.95$ $\min \lambda e^T x + \frac{1}{2}\|r\|^2$



SPAI: Sparse Approximate Inverse

SPAI Preconditioning

A plausible future application

- Find sparse X such that $AX \approx I$

SPAI Preconditioning

A plausible future application

- Find sparse X such that $AX \approx I$
- Find sparse x_j such that $Ax_j \approx e_j$ for $j = 1 : n$

SPAI Preconditioning

A plausible future application

- Find sparse X such that $AX \approx I$
- Find sparse x_j such that $Ax_j \approx e_j$ for $j = 1 : n$
- Embarrassingly parallel use of sparse BP solutions

Summary

$$(I - \alpha H^T)x \approx v$$

Comments

- Sparse v *sometimes* gives sparse x

$$(I - \alpha H^T)x \approx v$$

Comments

- Sparse v *sometimes* gives sparse x
- LPprimal and LPdual work in greedy manner

$$(I - \alpha H^T)x \approx v$$

Comments

- Sparse v *sometimes* gives sparse x
- LPprimal and LPdual work in greedy manner
- Greedy solvers guarantee sparse x (unlike power method)

$$(I - \alpha H^T)x \approx v$$

Comments

- Sparse v *sometimes* gives sparse x
- LPprimal and LPdual work in greedy manner
- Greedy solvers guarantee sparse x (unlike power method)
- Not sensitive to α

$$(I - \alpha H^T)x \approx v$$

Comments

- Sparse v *sometimes* gives sparse x
- LPprimal and LPdual work in greedy manner
- Greedy solvers guarantee sparse x (unlike power method)
- Not sensitive to α
- Sensitive to λ but we can limit iterations

$$(I - \alpha H^T)x \approx v$$

Comments

- Sparse v *sometimes* gives sparse x
- LPprimal and LPdual work in greedy manner
- Greedy solvers guarantee sparse x (unlike power method)
- Not sensitive to α
- Sensitive to λ but we can limit iterations

Questions

$$(I - \alpha H^T)x \approx v$$

Comments

- Sparse v *sometimes* gives sparse x
- LPprimal and LPdual work in greedy manner
- Greedy solvers guarantee sparse x (unlike power method)
- Not sensitive to α
- Sensitive to λ but we can limit iterations

Questions

- Run much bigger examples

$$(I - \alpha H^T)x \approx v$$

Comments

- Sparse v *sometimes* gives sparse x
- LPprimal and LPdual work in greedy manner
- Greedy solvers guarantee sparse x (unlike power method)
- Not sensitive to α
- Sensitive to λ but we can limit iterations

Questions

- Run much bigger examples
- Which properties of $I - \alpha H^T$ lead to success of "greedy"

Thanks

- Shaobing Chen and David Donoho
- Michael Friedlander (BP solvers)

Thanks

- Shaobing Chen and David Donoho
- Michael Friedlander (BP solvers)

- Sou-Cheng Choi and Lek-Heng Lim (ICIAM 07)

Thanks

- Shaobing Chen and David Donoho
- Michael Friedlander (BP solvers)

- Sou-Cheng Choi and Lek-Heng Lim (ICIAM 07)

- Amy Langville and Carl Meyer (splendid book)
- David Gleich

Thanks

- Shaobing Chen and David Donoho
- Michael Friedlander (BP solvers)

- Sou-Cheng Choi and Lek-Heng Lim (ICIAM 07)

- Amy Langville and Carl Meyer (splendid book)
- David Gleich

- Holly Jin