Testing the Nullspace Property using Semidefinite Programming

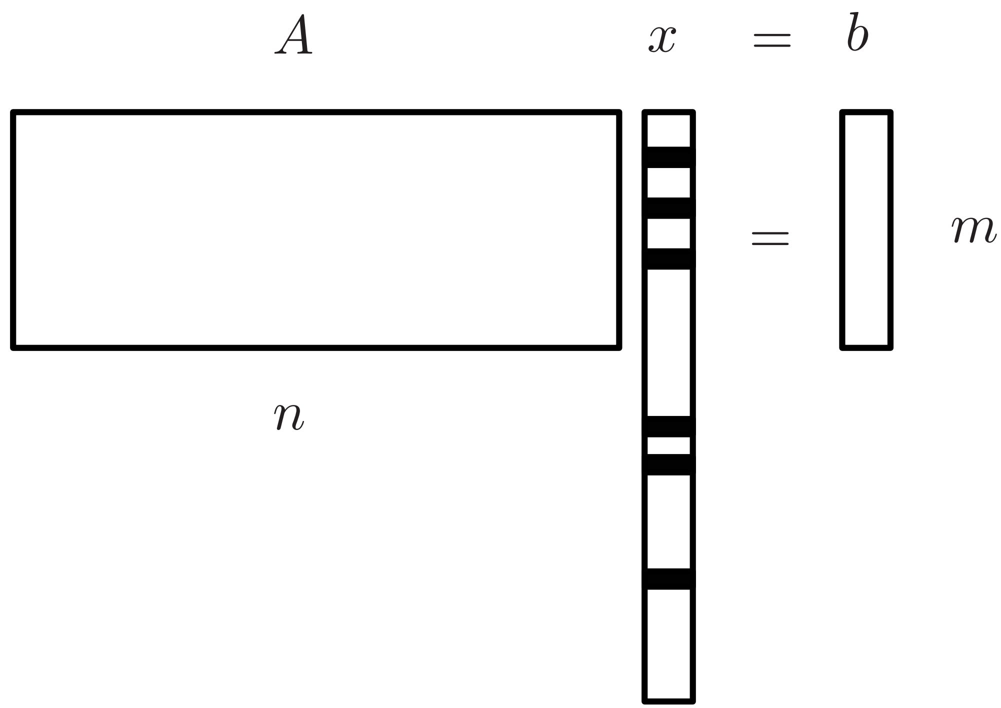
Alex d'Aspremont, Francis Bach, Laurent El Ghaoui

Princeton University, École Normale Supérieure, U.C. Berkeley.

Support from NSF, DHS and Google.

Introduction

Consider the following underdetermined linear system



where $A \in \mathbf{R}^{m \times n}$, with $n \gg m$.

Can we find the **sparsest** solution?

Introduction

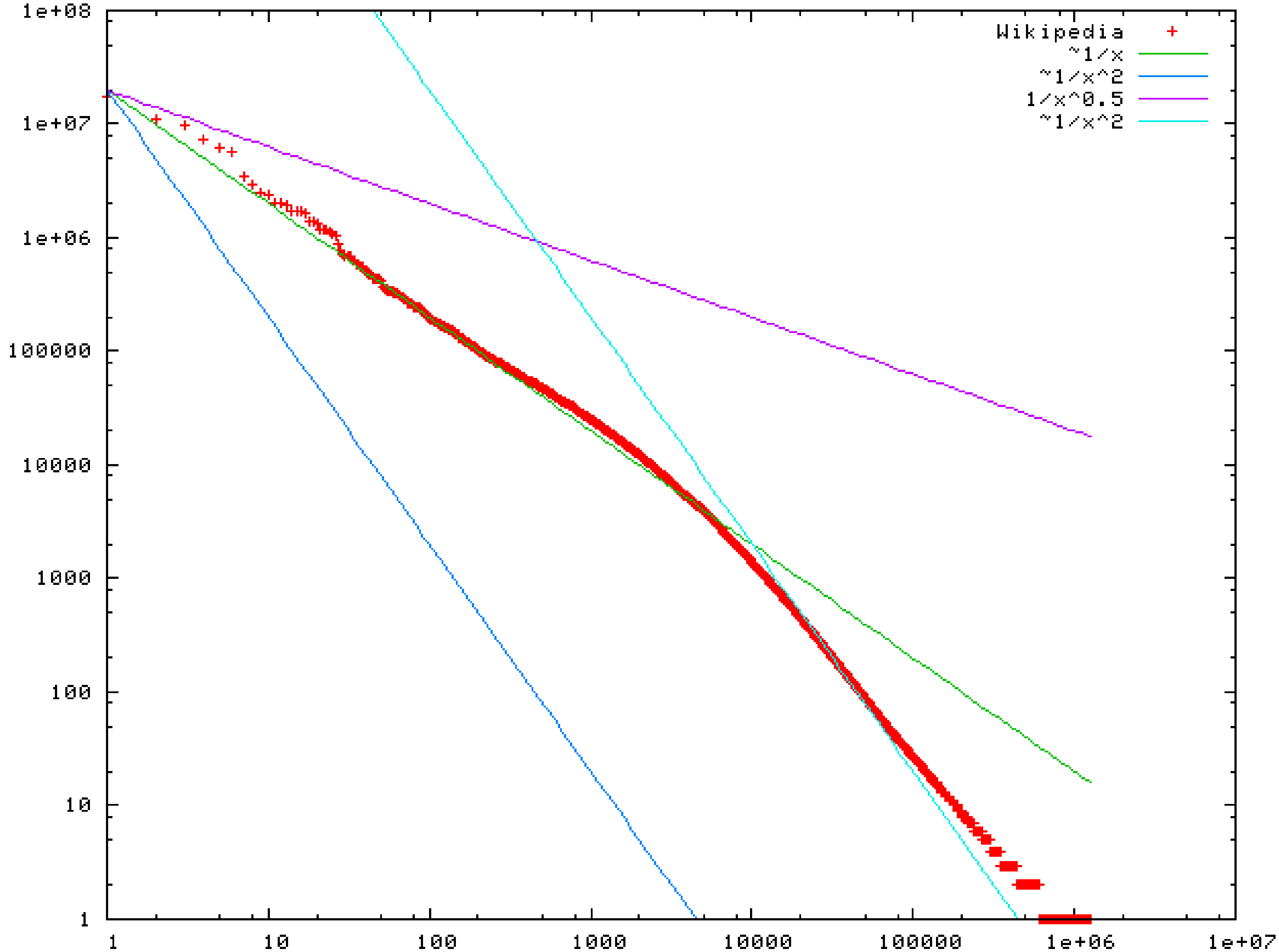
- **Signal processing:** We make a few measurements of a high dimensional signal, which admits a sparse representation in a well chosen basis (e.g. Fourier, wavelet). Can we reconstruct the signal exactly?
- **Coding:** Suppose we transmit a message which is corrupted by a few errors. How many errors does it take to start losing the signal?
- **Statistics:** Variable selection in regression (LASSO, etc).

Introduction

Why **sparsity**?

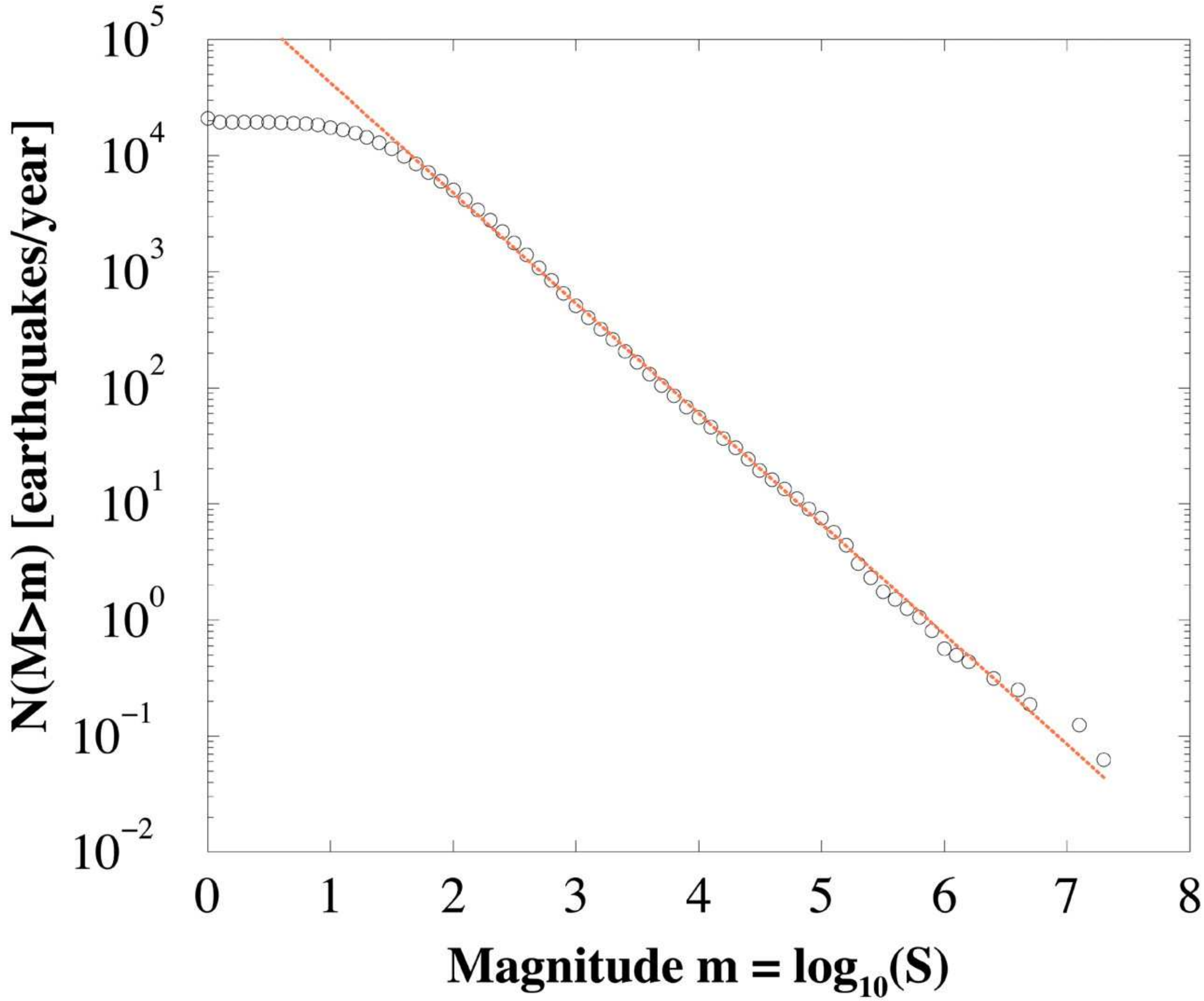
- Sparsity is a proxy for **power laws**. Most results stated here on sparse vectors apply to vectors with a power law decay in coefficient magnitude.
- Power laws appear everywhere. . .
  - Zipf law: word frequencies in natural language follow a power law.
  - Ranking: pagerank coefficients follow a power law.
  - Signal processing: $1/f$ signals
  - Social networks: node degrees follow a power law.
  - Earthquakes: Gutenberg-Richter power laws
  - River systems, cities, net worth, etc.

Introduction



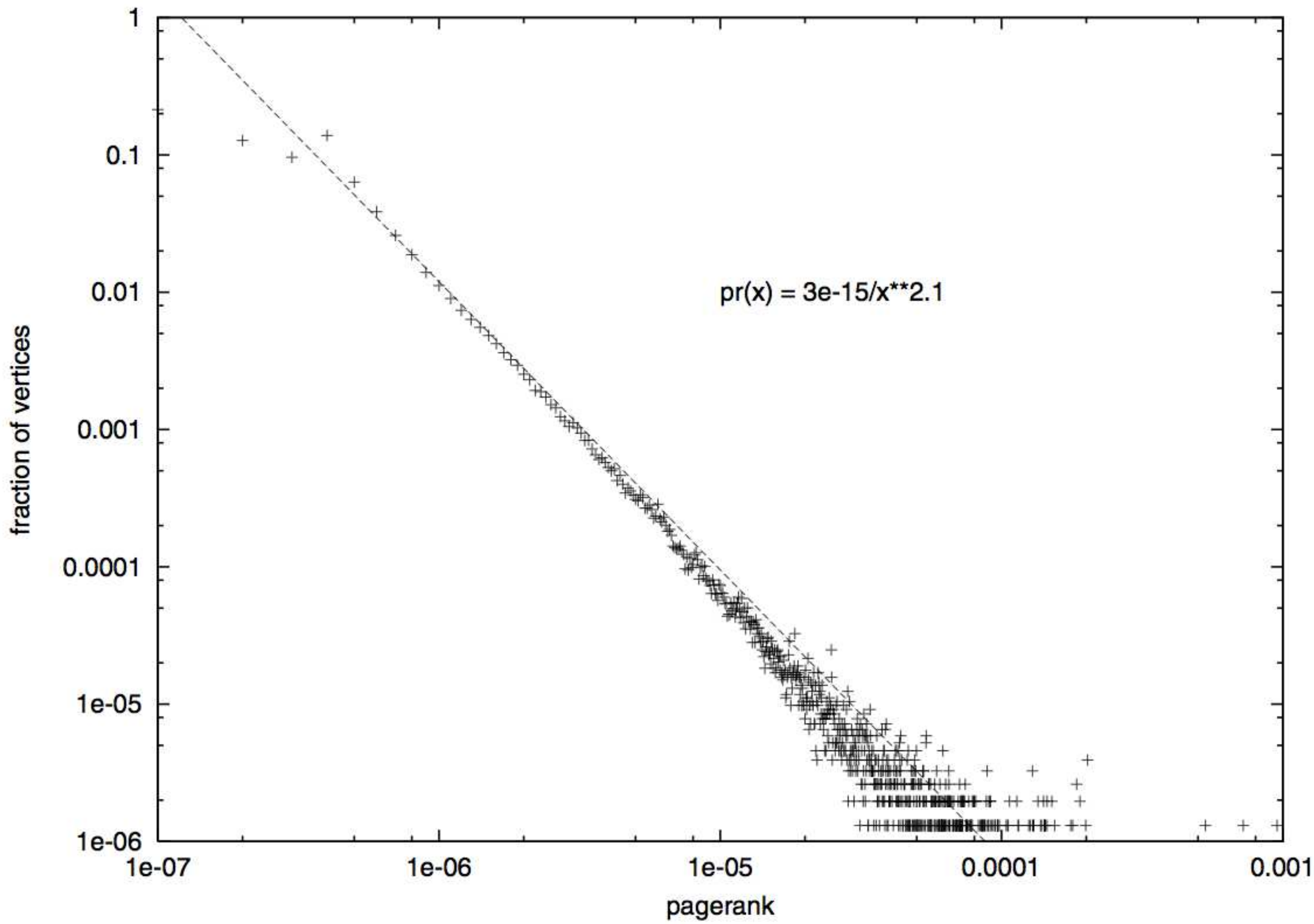
Frequency vs. word in Wikipedia (from Wikipedia).

Introduction



Frequency vs. magnitude for earthquakes worldwide. Christensen, Danon, Scanlon & Bak (2002)

Introduction



Pages vs. Pagerank on web sample. Pandurangan, Raghavan & Upfal (2006)

Introduction

- Getting the sparsest solution means solving:

$$\begin{array}{ll}\text{minimize} & \mathbf{Card}(x) \\ \text{subject to} & Ax = b\end{array}$$

which is a (hard) **combinatorial** problem in $x \in \mathbf{R}^n$.

- A classic heuristic is to solve instead:

$$\begin{array}{ll}\text{minimize} & \|x\|_1 \\ \text{subject to} & Ax = b\end{array}$$

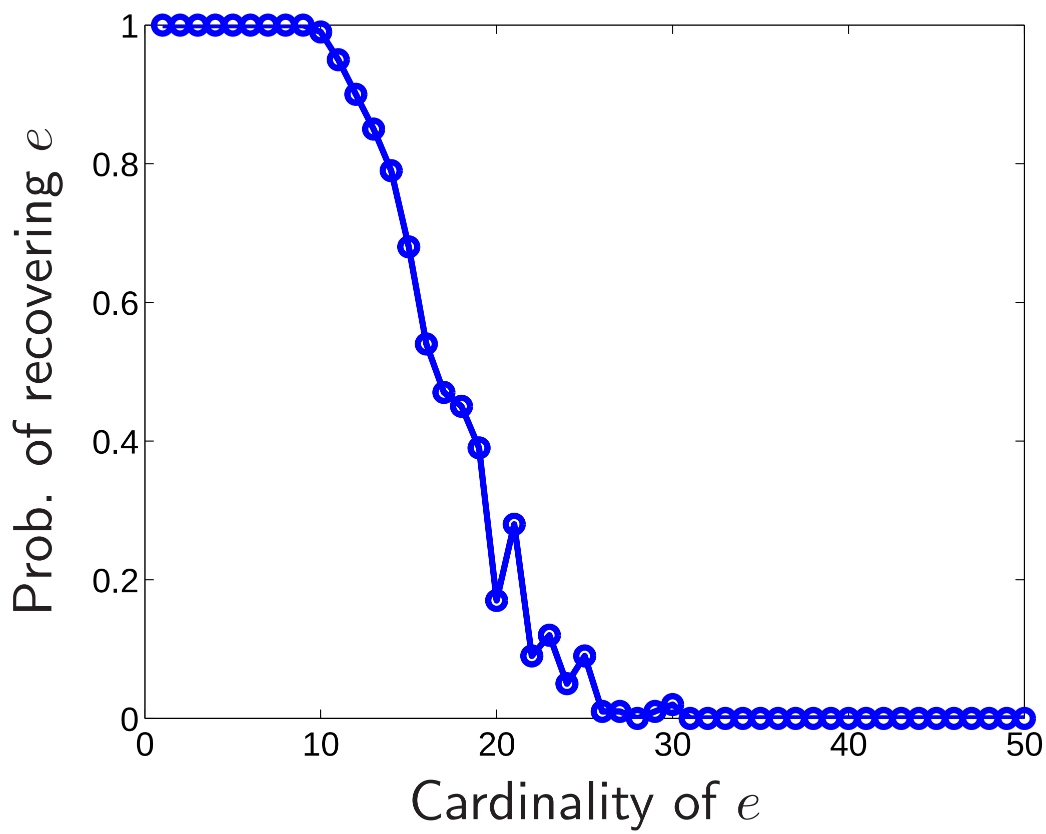
which is equivalent to an (easy) **linear program**.

Introduction

Example: we fix A, we draw many **sparse** signals e and plot the probability of perfectly recovering e by solving

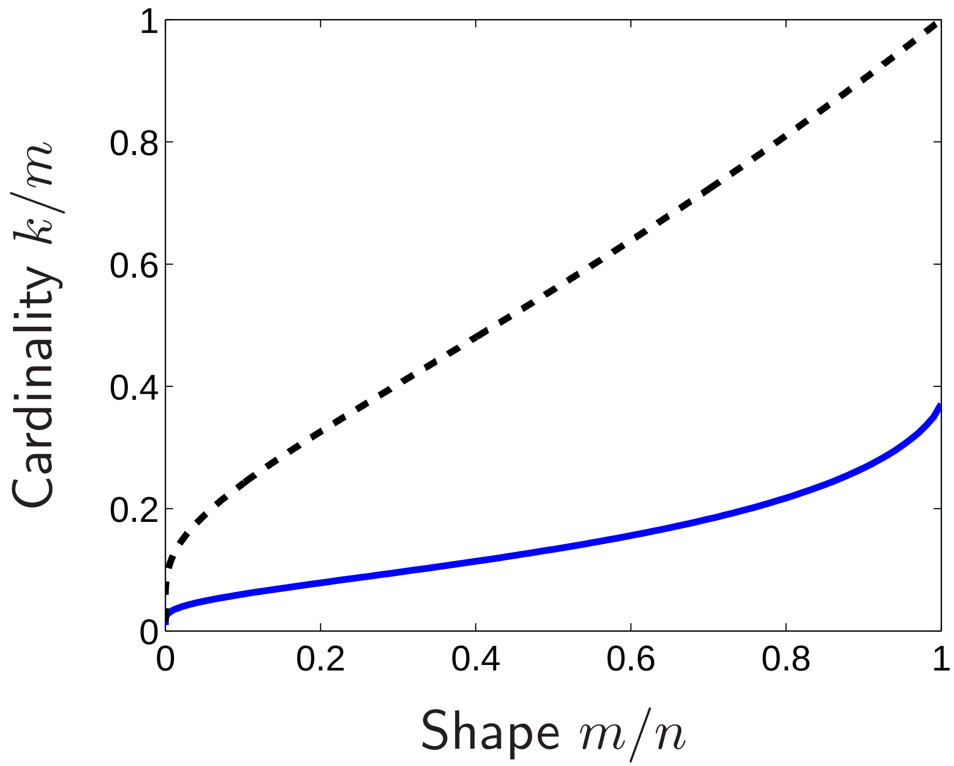
$$\begin{array}{ll} \text{minimize} & \|x\|_1 \\ \text{subject to} & Ax = Ae \end{array}$$

in $x \in \mathbf{R}^n$, with $n = 50$ and $m = 30$.



Introduction

- Donoho & Tanner (2005) and Candès & Tao (2005) show that for certain classes of matrices, when the solution e is sparse enough, the solution of the ℓ_1-**minimization** problem is also the **sparsest** solution to $Ax = Ae$.
- Let $k = \mathbf{Card}(e)$, this happens even when $\mathbf{k = O(m)}$ asymptotically, which is provably optimal.
- Also obtain bounds on reconstruction error outside of this range.



Introduction

Similar results exist for **rank minimization**.

- The ℓ_1 norm is replaced by the trace norm on matrices.
- Exact recovery results are detailed in Recht, Fazel & Parrilo (2007), Candes & Recht (2008), ...

Introduction

Explicit conditions for perfect recovery:

- **Restricted Isometry Property** (RIP) from Candès & Tao (2005).
- **Nullspace Property** (NSP) from Donoho & Huo (2001), Cohen, Dahmen & DeVore (2006), . . .

Candès & Tao (2005) and Cohen et al. (2006) show that these conditions are satisfied by certain classes of random matrices.

One small problem. . .

Testing these conditions on general matrices is **harder** than finding the sparsest solution to an underdetermined linear system for example.

Outline

- Introduction
- **Testing the RIP**
- Testing the NSP
- Limits of performance

Testing the RIP

- Candès & Tao (2005) define the **restricted isometry constant** $\delta_k(A)$ of the matrix A as follows: given $0 < k \leq n$, the constant $\delta_k(A)$ is the smallest number such that

$$(1-\delta_k)\|z\|_2^2 \leq \|A_I z\|_2^2 \leq (1+\delta_k)\|z\|_2^2,$$

  for all $z \in \mathbf{R}^{|I|}$, for any index subset $I \subset [1, n]$ of cardinality at most k, where A_I is the submatrix formed by extracting the columns of A indexed by I.

- The constant $\delta_k(A)$ measures how far sparse subsets of the columns of A are from being an isometry.

Testing the RIP

Following Candès & Tao (2005), suppose the solution has cardinality k.

- If $\delta_{2k}(A) < 1$, we can recover the error e by solving:

$$\begin{array}{ll} \text{minimize} & \mathbf{Card}(x) \\ \text{subject to} & Ax = Ae \end{array}$$

  in the variable $x \in \mathbf{R}^n$, which is a **combinatorial** problem.

- If $\delta_{2k}(A) < \sqrt{2} - 1$, we can recover the error e by solving:

$$\begin{array}{ll} \text{minimize} & \|x\|_1 \\ \text{subject to} & Ax = Ae \end{array}$$

  in the variable $x \in \mathbf{R}^n$, which is a **linear program**.

Testing the RIP

- The restricted isometry constant $\delta_k(A)$ can be computed by solving the following **sparse eigenvalue** problem

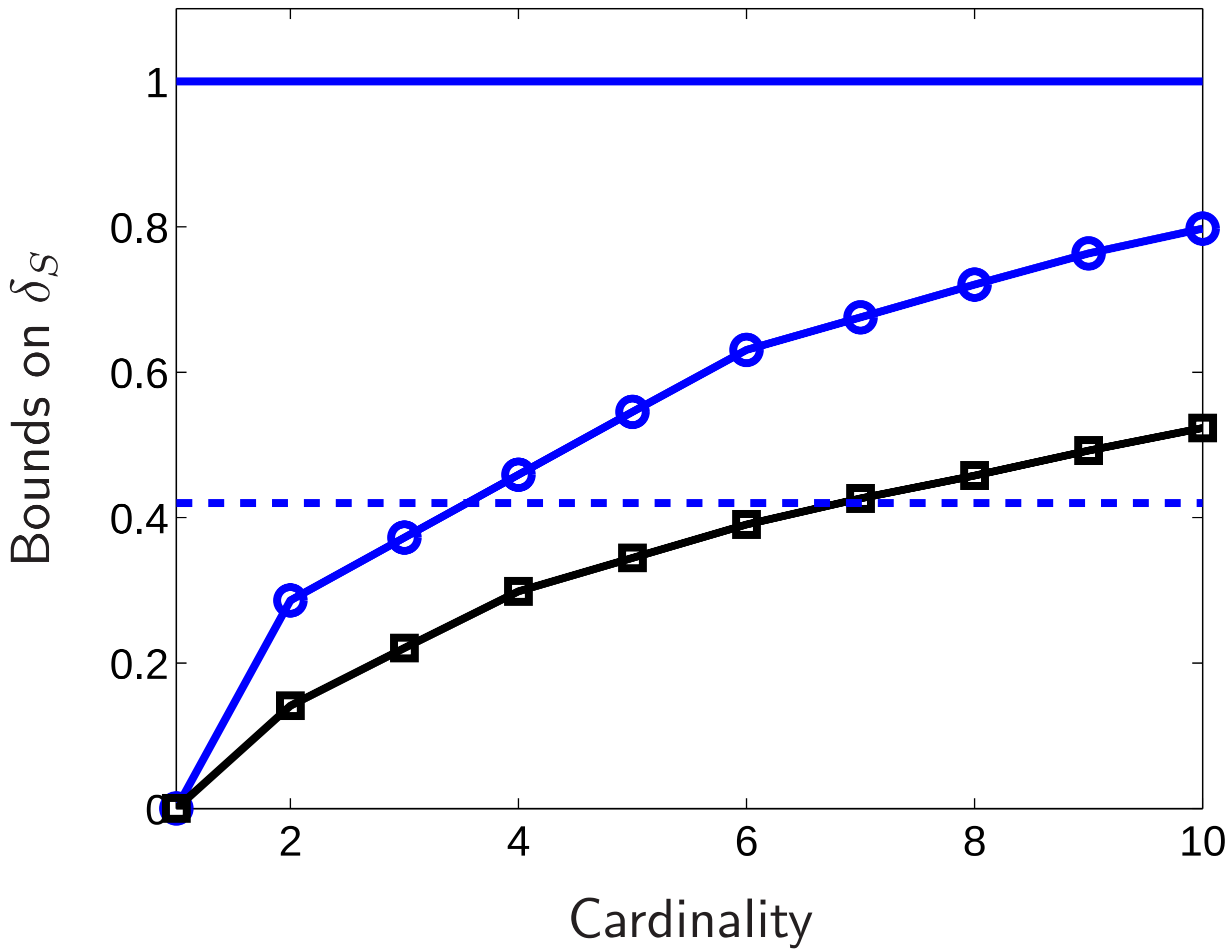
$$(1+\delta_k^{\max}) = \begin{array}{ll} \text{max.} & x^T(A^TA)x \\ \text{s. t.} & \mathbf{Card}(x) \leq k \\ & \|x\| = 1, \end{array}$$

in $x \in \mathbf{R}^m$ (a similar problem gives $\delta_k^{\min}$ and $\delta_k(A) = \max\{\delta_k^{\min}, \delta_k^{\max}\}$).

- SDP relaxation in d'Aspremont, El Ghaoui, Jordan & Lanckriet (2007):

$$\begin{array}{ll} \text{maximize} & x^TA^TAx \\ \text{subject to} & \|x\|_2 = 1 \\ & \mathbf{Card}(x) \leq k, \end{array} \quad \textbf{is bounded by} \quad \begin{array}{ll} \text{maximize} & \mathbf{Tr}(A^TAX) \\ \text{subject to} & \mathbf{Tr}(X) = 1 \\ & \mathbf{1}^T|X|\mathbf{1} \leq k \\ & X \succeq 0, \end{array}$$

Testing the RIP



Upper bound on δ_S using approximate sparse eigenvectors, for a Bernoulli matrix of dimension $n = 1000$, $p = 750$ (blue cicles).
Lower bound on δ_S using approximate sparse eigenvectors (black squares).

Outline

- Introduction
- Testing the RIP
- **Testing the NSP**
- Limits of performance

Testing the NSP

Given $A \in \mathbf{R}^{m \times n}$ and $k > 0$, Donoho & Huo (2001) or Cohen et al. (2006) among others, define the **Nullspace Property** of the matrix A as

$$\|x_T\|_1 \leq \alpha_k \|x\|_1$$

for all vectors $x \in \mathbf{R}^n$ with $Ax = 0$ and index subsets $T \subset [1, n]$ with cardinality k, for some $\alpha_k \in [0, 1)$.

Once again, two thresholds:

- $\alpha_{2k} < 1$ means recovery is guaranteed by solving a ℓ_0 minimization problem.
- $\alpha_k < 1/2$ means recovery is guaranteed by solving a ℓ_1 minimization problem.

Cohen et al. (2006) show that $RIP(2k, \delta)$ implies NSP with $\alpha = (1 + 5\delta)/(2 + 2\delta)$, so the NSP is a **weaker** condition for sparse recovery.

Testing the NSP

- By homogeneity, we have

$$\alpha_k = \max_{\{Ax=0,\ \|x\|_1=1\}} \ \max_{\{\|y\|_\infty=1,\ \|y\|_1\leq k\}} y^T x$$

- An upper bound can be computed by solving

$$\begin{array}{ll} \mbox{maximize} & \mathbf{Tr}(Z) \\ \mbox{subject to} & AXA^T = 0,\ \|X\|_1 \leq 1, \\ & \|Y\|_\infty \leq 1,\ \|Y\|_1 \leq k^2,\ \|Z\|_1 \leq k, \\ & \left(\begin{array}{cc} X & Z^T \\ Z & Y \end{array}\right) \succeq 0, \end{array}$$

  which is a **semidefinite program** in $X, Y \in \mathbf{S}_n,\ Z \in \mathbf{R}^{n\times n}$.

- This is a standard semidefinite relaxation, except for the redundant constraint $\|Z\|_1 \leq k$ which significantly improves performance. Extra column-wise redundant constraints further tighten it.
- Another LP-based relaxation was derived in Juditsky & Nemirovski (2008).

Testing the NSP

- Use an **elimination result** for LMIs in Boyd, El Ghaoui, Feron & Balakrishnan (1994, §2.6.2) to reduce the size of the problem and express it in terms of a matrix P where $AP = 0$ with $P^T P = \mathbf{I}$.

- Compute the dual and using **binary search** to certify $\alpha_k \leq 1/2$, we solve

$$\begin{array}{ll} \text{maximize} & \lambda_{\min}\begin{pmatrix} P^T U_1 P & -\frac{1}{2}P^T(\mathbf{I}+U_4) \\ -\frac{1}{2}(\mathbf{I}+U_4^T)P & U_2+U_3 \end{pmatrix} \\ \text{subject to} & \|U_1\|_\infty + k^2\|U_2\|_\infty + \|U_3\|_1 + k\|U_4\|_\infty \leq 1/2 \end{array}$$

in the variables $U_1, U_2, U_3 \in \mathbf{S}_n$ and $U_4 \in \mathbf{R}^{n\times n}$.

Testing the NSP

- The complexity of computing the Euclidean projection $(x_0, y_0, z_0, w_0) \in \mathbf{R}^{3n}$ on

$$\|x\|_\infty + k^2\|y\|_\infty + \|z\|_1 + k\|w\|_\infty \leq \alpha$$

is bounded by $O(n \log n \log_2(1/\epsilon))$, where ϵ is the target precision in projecting.

- Using smooth optimization techniques as in Nesterov (2007), we get the following complexity bound:

$$O\left(\frac{n^4\sqrt{\log n}}{\epsilon}\right)$$

- In practice, this is still **slow**. Much slower than the LP relaxation in Juditsky & Nemirovski (2008). Slower also than a similar algorithm in d'Aspremont et al. (2007) to bound the RI constant.

Testing the NSP

- We can use **randomization** to generate certificates that $\alpha_k > 1/2$.
- **Concentration result**: let $X \in \mathbf{S}_n$, $x \sim \mathcal{N}(0, X)$ and $\delta > 0$, we have

$$\mathbf{P}\left(\frac{\|x\|_1}{\left(\sqrt{2/\pi}+\sqrt{2\log\delta}\right)\sum_{i=1}^n (X_{ii})^{1/2}} \geq 1\right) \leq \frac{1}{\delta}$$

- Highlights the importance of the redundant constraint on Z:

$$\|Z\|_1 \leq \left(\sum_{i=1}^n (X_{ii})^{1/2}\right)\left(\sum_{i=1}^n (Y_{ii})^{1/2}\right)$$

with equality when the SDP solution has rank one.

Testing the NSP

- **Tightness:** writing SDP_k the optimal value of the relaxation, we have

$$\frac{SDP_k - \epsilon}{g(X,\delta)h(Y,n,k,\delta)} \leq \alpha_k \leq SDP_k$$

where

$$g(X,\delta) = (\sqrt{2/\pi} + \sqrt{2\log\delta}) \textstyle\sum_{i=1}^n (X_{ii})^{1/2}$$

and

$$\begin{aligned} h(Y,n,k,\delta) \quad = \quad & \max\{(\sqrt{2\log 2n} + \sqrt{2\log\delta}) \max_{i=1,\ldots,n} (Y_{ii})^{1/2}, \\ & \frac{(\sqrt{2/\pi} + \sqrt{2\log\delta}) \sum_{i=1}^n (Y_{ii})^{1/2}}{k}\} \end{aligned}$$

- Because $\sum_{i=1}^n (X_{ii})^{1/2} \leq \sqrt{n}$ here, this is roughly

$$\frac{SDP_k - \epsilon}{\max\left\{\sqrt{2\log 2n}, \sqrt{\frac{m}{k}}\sqrt{\frac{n}{m}}\sqrt{\frac{1}{k}}\right\} C\sqrt{n}} \leq \alpha_k \leq SDP_k$$

Testing the NSP

Relaxation	ρ	α_1	α_2	α_3	α_4	α_5	Strong k	Weak k
LP	0.5	**0.27**	**0.49**	0.67	0.83	0.97	2	11
SDP	0.5	**0.27**	**0.49**	0.65	0.81	0.94	2	11
SDP low.	0.5	0.27	0.31	0.33	0.32	0.35	2	11
LP	0.6	**0.22**	**0.41**	0.57	0.72	0.84	2	12
SDP	0.6	**0.22**	**0.41**	0.56	0.70	0.82	2	12
SDP low.	0.6	0.22	0.29	0.31	0.32	0.36	2	12
LP	0.7	**0.20**	**0.34**	**0.47**	0.60	0.71	3	14
SDP	0.7	**0.20**	**0.34**	**0.46**	0.59	0.70	3	14
SDP low.	0.7	0.20	0.27	0.31	0.35	0.38	3	14
LP	0.8	**0.15**	**0.26**	**0.37**	**0.48**	0.58	3	16
SDP	0.8	**0.15**	**0.26**	**0.37**	**0.48**	0.58	3	16
SDP low.	0.8	0.15	0.23	0.28	0.33	0.38	3	16

Given ten sample *Gaussian* matrices of leading dimension $n = 40$, we list median upper bounds on the values of α_k for various cardinalities k and matrix shape ratios ρ. We also list the asymptotic upper bound on both strong and weak recovery computed in Donoho & Tanner (2008) and the lower bound on α_k obtained by randomization using the SDP solution (SDP low.).

Outline

- Introduction
- Testing the RIP
- Testing the NSP
- **Limits of performance**

Limits of performance

- The SDP relaxation is **tight** for α_1.
- Based on results in Juditsky & Nemirovski (2008), this also means that it can prove perfect recovery at cardinality $k = O(\sqrt{k^*})$ when A satisfies RIP at the optimal rate $k = O(k^*)$. It cannot do better than $k = O(\sqrt{k^*})$.
(*Counter-example by A. Nemirovski: feasible point of the SDP where $k = \sqrt{k^*}$ with objective greater than 1/2 in testing the NSP*).
- The LP relaxation in Juditsky & Nemirovski (2008) guarantees the same $k = O(\sqrt{k^*})$ when A satisfies RIP at $k = O(k^*)$. It also cannot do better than this rate.
- The same kind of argument shows that the DSCPA relaxation in d'Aspremont et al. (2007) cannot do better than $k = O(\sqrt{k^*})$.

This means that all current convex relaxations for testing sparse recovery conditions achieve a **maximum rate of $\mathbf{O(\sqrt{m})}$**. . .

Conclusion

- Convex relaxations of sparse recovery conditions prove recovery at cardinality $k = O(\sqrt{n})$ for **any matrix** satisfying NSP at the optimal rate $k = O(n)$.
- Testing recovery conditions on deterministic matrices at the optimal rate $O(n)$ remains an open problem.

What next?

- Improved relaxations.
- Test weak recovery instead.
- Prove hardness of testing NSP and RIP beyond $O(\sqrt{m})$: optimization would do worst than sampling a few Gaussian variables?

References

Boyd, S., El Ghaoui, L., Feron, E. & Balakrishnan, V. (1994), *Linear Matrix Inequalities in System and Control Theory*, SIAM.

Candès, E. J. & Tao, T. (2005), 'Decoding by linear programming', *Information Theory, IEEE Transactions on* **51**(12), 4203–4215.

Candes, E. & Recht, B. (2008), 'Exact matrix completion via convex optimization', *preprint* .

Christensen, K., Danon, L., Scanlon, T. & Bak, P. (2002), 'Unified scaling law for earthquakes'.

Cohen, A., Dahmen, W. & DeVore, R. (2006), 'Compressed sensing and best k-term approximation', *Submitted for publication* .

d'Aspremont, A., El Ghaoui, L., Jordan, M. & Lanckriet, G. R. G. (2007), 'A direct formulation for sparse PCA using semidefinite programming', *SIAM Review* **49**(3), 434–448.

Donoho, D. & Huo, X. (2001), 'Uncertainty principles and ideal atomic decomposition', *IEEE Transactions on Information Theory* **47**(7), 2845–2862.

Donoho, D. L. & Tanner, J. (2005), 'Sparse nonnegative solutions of underdetermined linear equations by linear programming', *Proc. of the National Academy of Sciences* **102**(27), 9446–9451.

Donoho, D. & Tanner, J. (2008), 'Counting the Faces of Randomly-Projected Hypercubes and Orthants, with Applications', *Arxiv preprint arXiv:0807.3590* .

Juditsky, A. & Nemirovski, A. (2008), 'On verifiable sufficient conditions for sparse signal recovery via ℓ_1 minimization', *ArXiv:0809.2650* .

Nesterov, Y. (2007), 'Smoothing technique and its applications in semidefinite optimization', *Mathematical Programming* **110**(2), 245–259.

Pandurangan, G., Raghavan, P. & Upfal, E. (2006), 'Using pagerank to characterize web structure', *Internet Mathematics* **3**(1), 1–20.

Recht, B., Fazel, M. & Parrilo, P. (2007), 'Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization', *Arxiv preprint arXiv:0706.4138* .