Tractable performance bounds for compressed sensing.

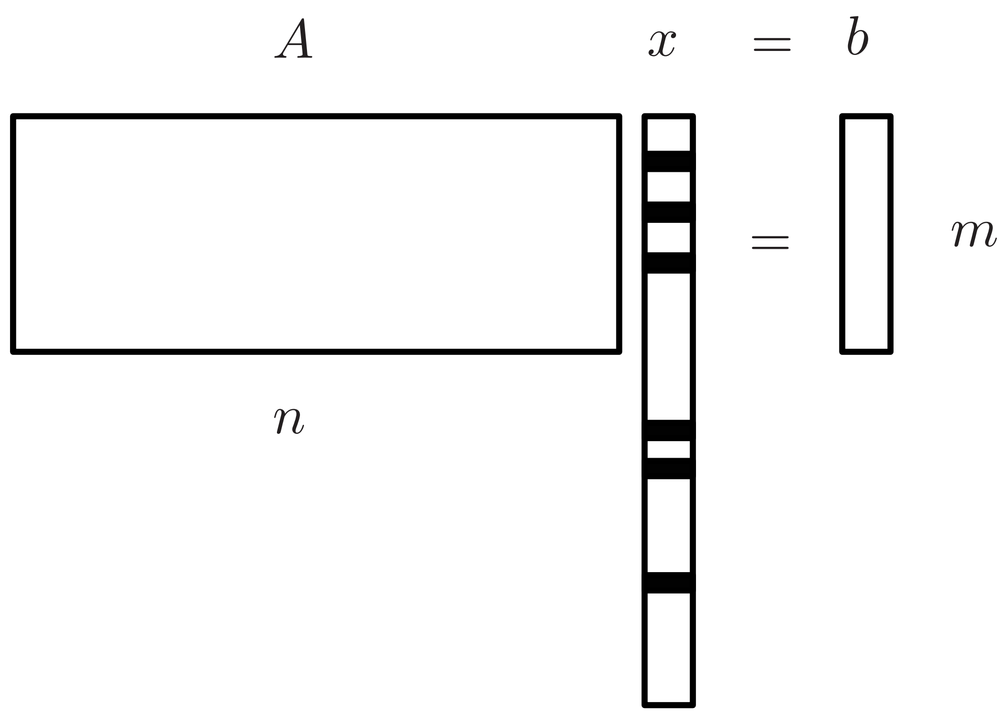
Alex d'Aspremont, Francis Bach, Laurent El Ghaoui

Princeton University, École Normale Supérieure/INRIA, 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 $n \geq m$.

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?
(Donoho, 2004; Donoho and Tanner, 2005; Donoho, 2006)

- **Coding:** Suppose we transmit a message which is corrupted by a few errors. How many errors does it take to start losing the signal?
(Candès and Tao, 2005, 2006)

- **Statistics:** Variable selection & regression (LASSO, . . .).
(Zhao and Yu, 2006; Meinshausen and Yu, 2008; Meinshausen et al., 2007; Candès and Tao, 2007; Bickel et al., 2007)

Introduction

Many variants. . .

- The observations could be **noisy**.
- **Approximate solutions** might be sufficient.
- We might have strict **computational limits** on the decoding side.

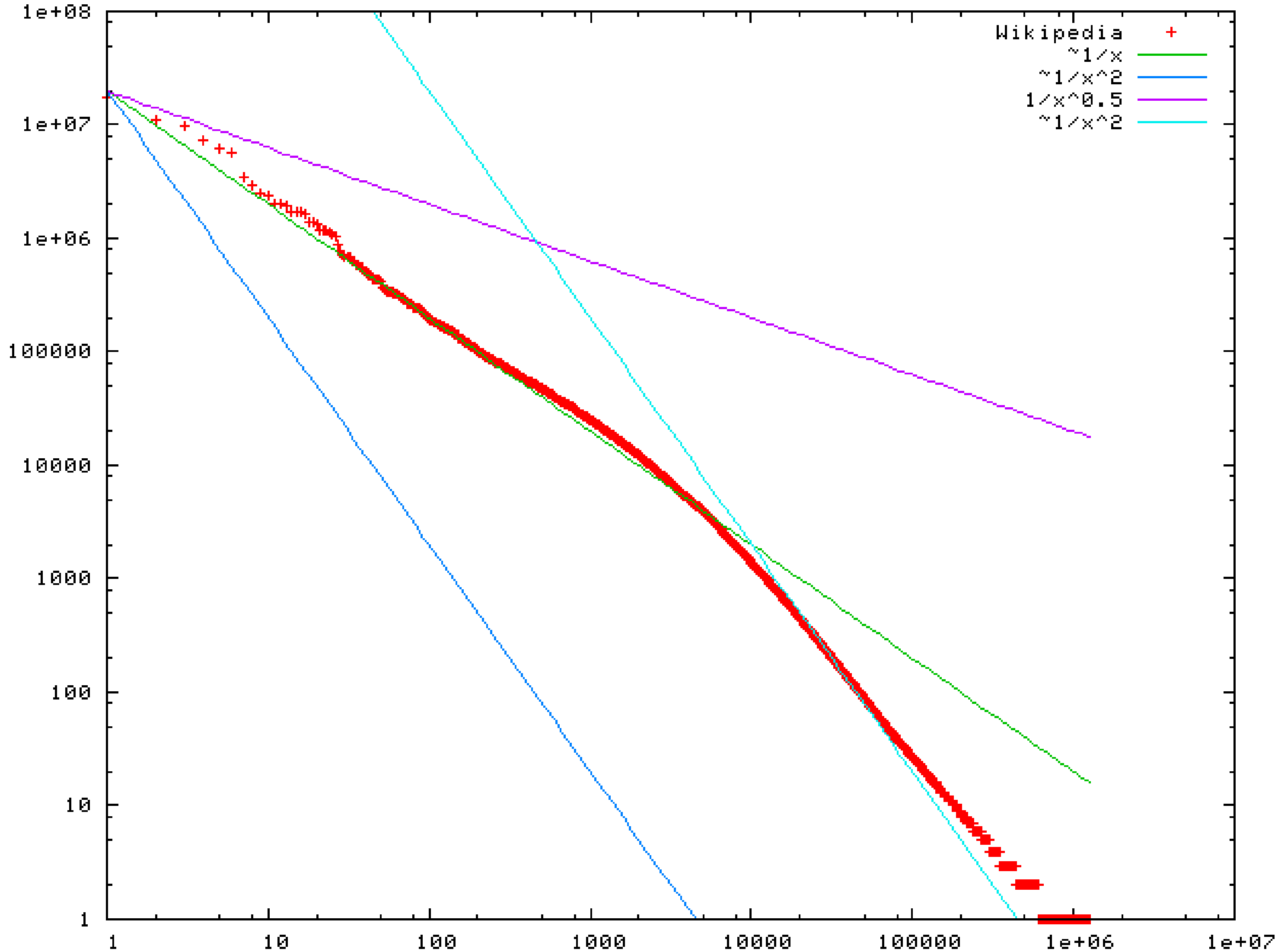
In this talk: use simplest formulation possible, focus on the **complexity** of recovery conditions.

Introduction

Why **sparsity**?

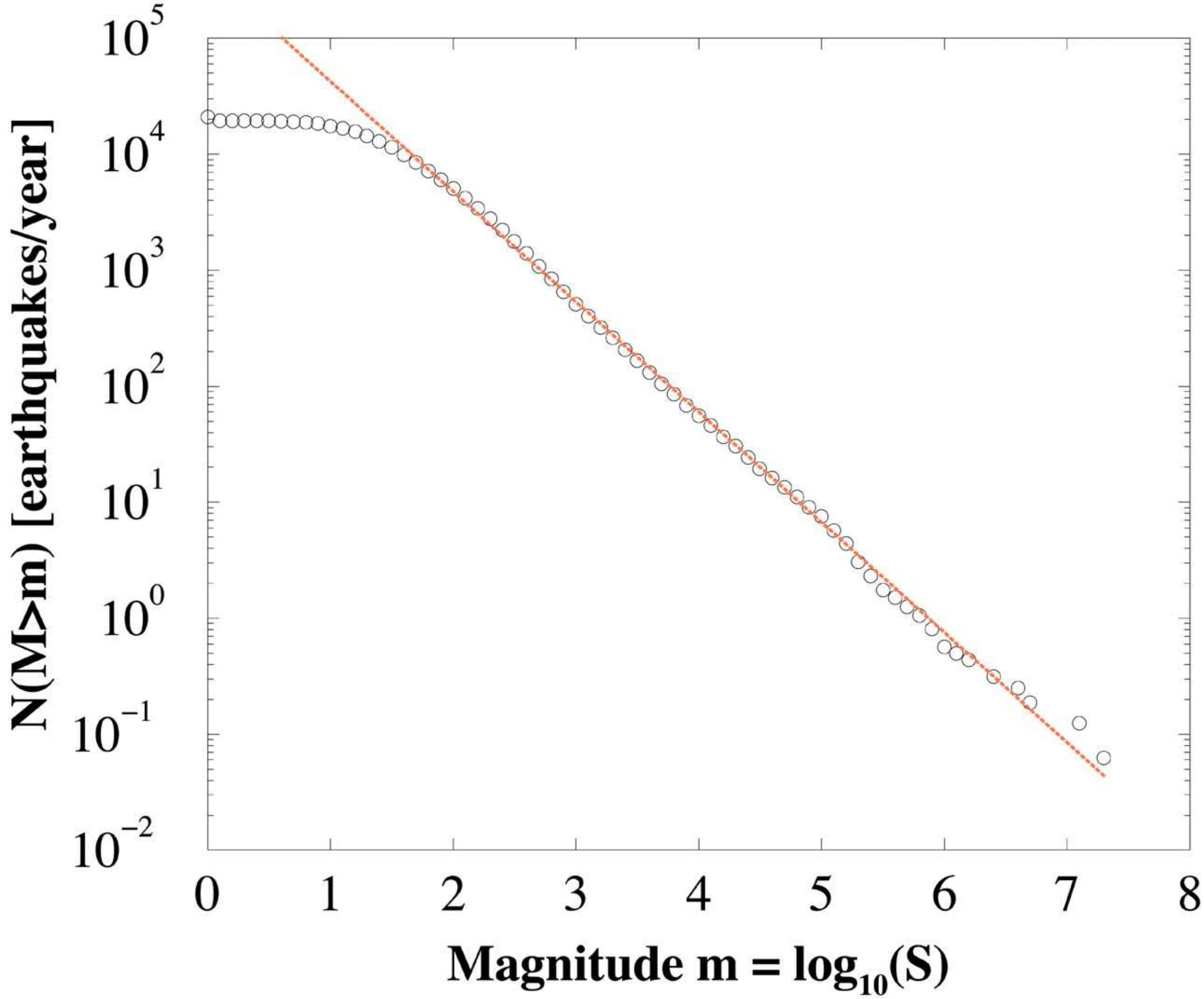
- Sparsity is a proxy for **power laws**. Most results stated here on sparse vectors apply to vectors with a power law decay in coefficient magnitude.
- Power laws appear everywhere. . .
  - **Text:** word frequencies in natural language follow a Zipf power law.
  - **Ranking:** pagerank coefficients 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



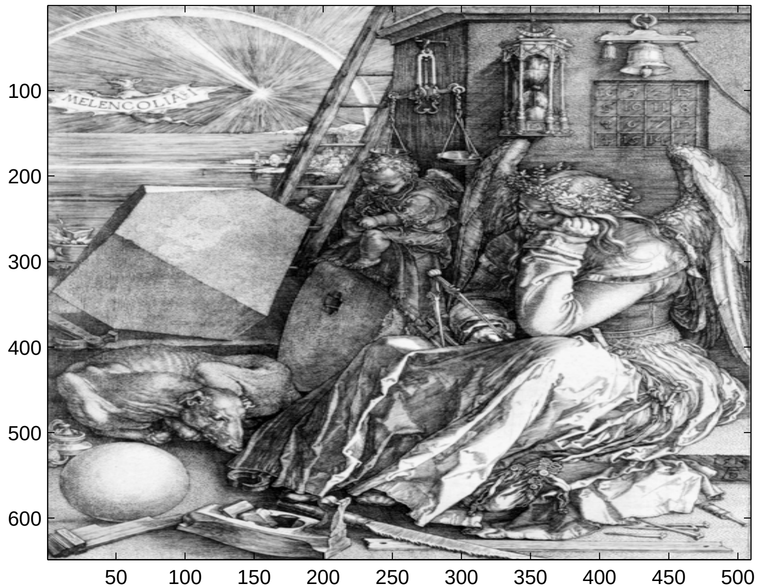
Frequency vs. word in Wikipedia (from Wikipedia).

Introduction



Frequency vs. magnitude for earthquakes worldwide. (Christensen et al., 2002)

Introduction



Original image

9% wavelet coefs.

Left: Original image.

Right: Same image reconstructed from 9% largest wavelet coefficients.

Introduction

- Getting the **sparsest** solution means solving

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

- Given an a priori bound on the solution, this can be formulated as a Mixed Integer Linear Program:

$$\begin{array}{ll} \text{minimize} & \mathbf{1}^T u \\ \text{subject to} & Ax = b \\ & |x| \preceq Bu \\ & u \in \{0,1\}^n. \end{array}$$

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

l_1 relaxation

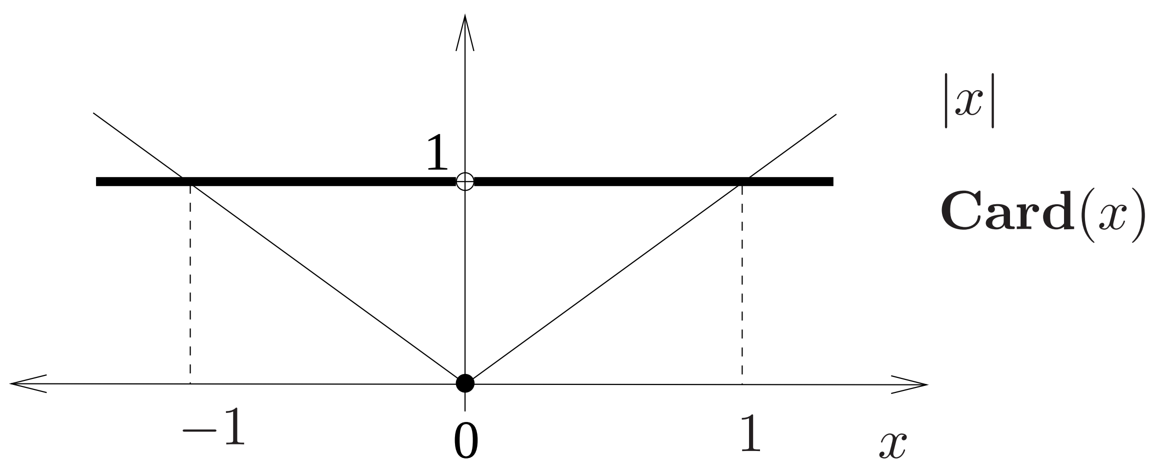
Assuming $|x| \leq 1$, we can replace:

$$\mathbf{Card}(x) = \sum_{i=1}^{n} 1_{\{x_i \neq 0\}}$$

with

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|$$

Graphically, assuming $x \in [-1, 1]$ this is:



The l_1 norm is the **largest convex lower bound** on $\mathbf{Card}(x)$ in $[-1, 1]$.

l_1 relaxation

$$\begin{array}{ll} \text{minimize} & \mathbf{Card}(x) \\ \text{subject to} & Ax = b \end{array} \quad \textbf{becomes} \quad \begin{array}{ll} \text{minimize} & \|x\|_1 \\ \text{subject to} & Ax = b \end{array}$$

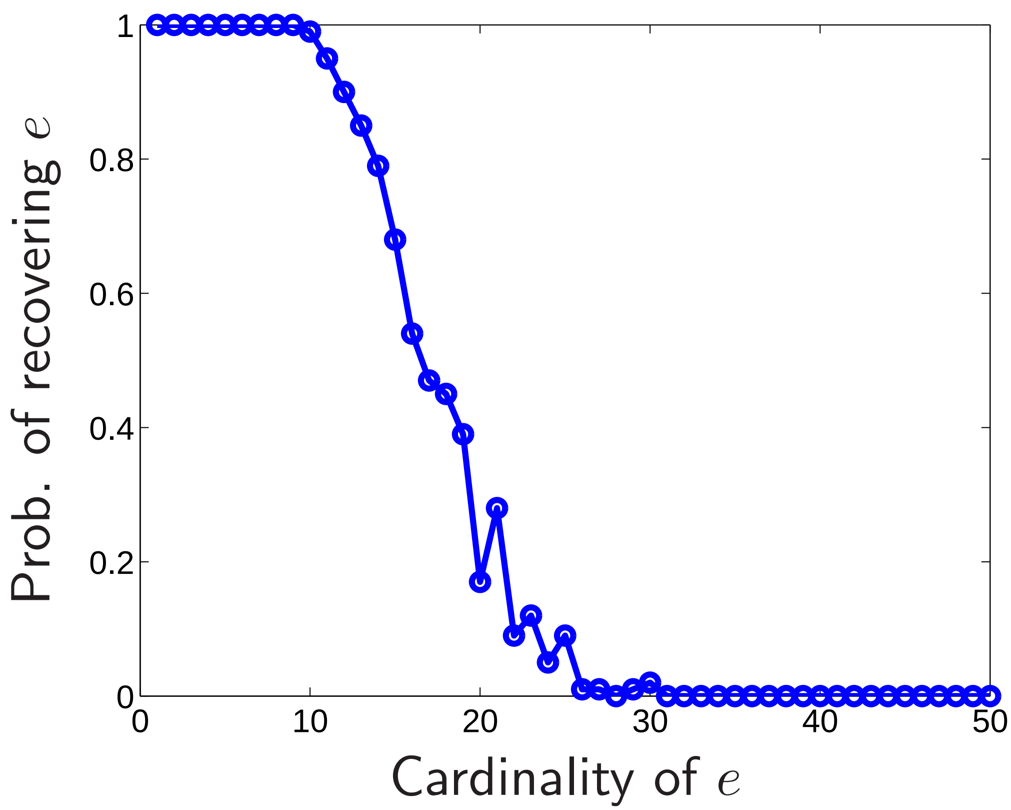
- Relax the constraint $u \in \{0, 1\}^n$ as $u \in [0, 1]^n$ in the MILP formulation.
- Same result if we relax a nonconvex quadratic program with $u \in \{0, 1\}$ replaced by $u(1 - u) = 0$ (see Lemaréchal and Oustry (1999) for a general discussion).
- Same trick can be generalized: **minimum rank** semidefinite program by Fazel et al. (2001).

Introduction

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

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

in $x \in \mathbf{R}^n$ over 100 samples, with $n = 50$ and $m = 30$.

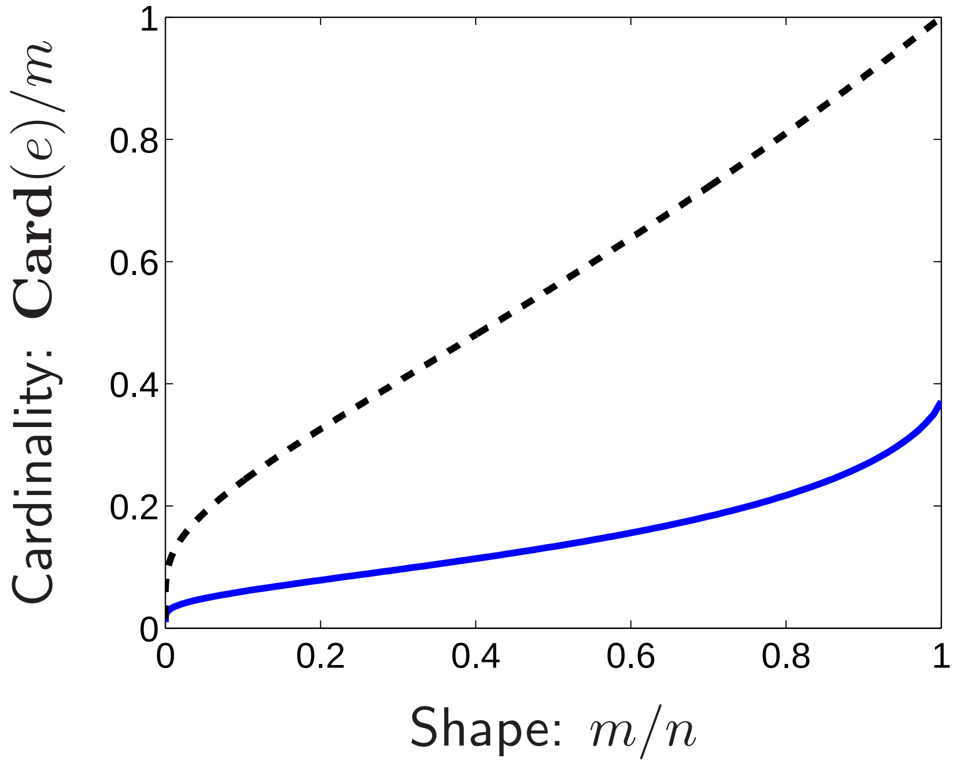


Introduction

- **Donoho and Tanner (2005), Candès and Tao (2005):**

  *For certain matrices A, when the solution e is sparse enough, the solution of the ℓ_1-**minimization** problem is also the **sparsest** solution to $Ax = Ae$.*

- This happens even when $\mathbf{Card(e) = O(m)}$ asymptotically in n when $m = \rho n$, which is provably optimal.



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 et al. (2007), Candes and Recht (2008), . . .

Introduction

Explicit conditions on the matrix A for perfect recovery of all sparse signals e.

- **Nullspace Property** (NSP) from Donoho and Huo (2001), Cohen et al. (2009), ...
- **Restricted Isometry Property** (RIP) from Candès and Tao (2005).

Candès and Tao (2005) and Baraniuk et al. (2007) show that these conditions are satisfied by certain classes of **random matrices**: Gaussian, Bernoulli, etc. (Donoho and Tanner (2005) use a geometric argument to obtain similar results)

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

- Given $0 < k \leq n$, Candès and Tao (2005) define the **restricted isometry constant** $\delta_k(A)$ as smallest number δ such that

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

  for all $z \in \mathbf{R}^{|I|}$ and 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.

- Candès and Tao (2005): $\delta_k(A)$ controls **sparse recovery** using ℓ_1-minimization.

Testing the RIP

Following Candès and 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 constant $\delta_{2k}(A) < 1$ also **controls reconstruction error** when exact recovery does not occur, with

$$\|x^* - e\|_1 \leq 2\frac{1 + (\sqrt{2} - 1)\delta_{2k}(A)}{1 - \delta_{2k}(A)/(\sqrt{2} - 1)}\sigma_k(e)$$

where x^* is the solution to the ℓ_1 minimization problem and e is the original signal, with

$$\sigma_k(x) = \min_{\mathbf{Card}(u) \leq k} \|u - e\|_1$$

denoting the **best possible approximation error**.

See Cohen et al. (2009) or Candes (2008) for simple proofs.

Testing the RIP

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

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

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 et al. (2007):

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

Semidefinite relaxation

(Lovász and Schrijver, 1991; Goemans and Williamson, 1995) Start from

$$\begin{array}{ll} \text{maximize} & x^T A x \\ \text{subject to} & \|x\|_2 = 1 \\ & \mathbf{Card}(x) \leq k, \end{array}$$

where $x \in \mathbf{R}^n$. Let $X = xx^T$ and write everything in terms of the matrix X

$$\begin{array}{ll} \text{maximize} & \mathbf{Tr}(AX) \\ \text{subject to} & \mathbf{Tr}(X) = 1 \\ & \mathbf{Card}(X) \leq k^2 \\ & X = xx^T, \end{array}$$

Replace $X = xx^T$ by the equivalent $X \succeq 0,\ \mathbf{Rank}(X) = 1$

$$\begin{array}{ll} \text{maximize} & \mathbf{Tr}(AX) \\ \text{subject to} & \mathbf{Tr}(X) = 1 \\ & \mathbf{Card}(X) \leq k^2 \\ & X \succeq 0,\ \mathbf{Rank}(X) = 1, \end{array}$$

again, this is the **same problem**.

Semidefinite relaxation

We have made **some progress**:

- The objective $\mathbf{Tr}(AX)$ is now **linear** in X
- The (non-convex) constraint $\|x\|_2 = 1$ became a **linear** constraint $\mathbf{Tr}(X) = 1$.

But this is still a hard problem:

- The $\mathbf{Card}(X) \leq k^2$ is still non-convex.
- So is the constraint $\mathbf{Rank}(X) = 1$.

We still need to relax the two non-convex constraints above:

- If $u \in \mathbf{R}^p$, $\mathbf{Card}(u) = q$ implies $\|u\|_1 \leq \sqrt{q}\|u\|_2$. So we can replace $\mathbf{Card}(X) \leq k^2$ by the weaker (but **convex**): $\mathbf{1}^T|X|\mathbf{1} \leq k$.
- We simply drop the rank constraint

Semidefinite Programming

Semidefinite relaxation:

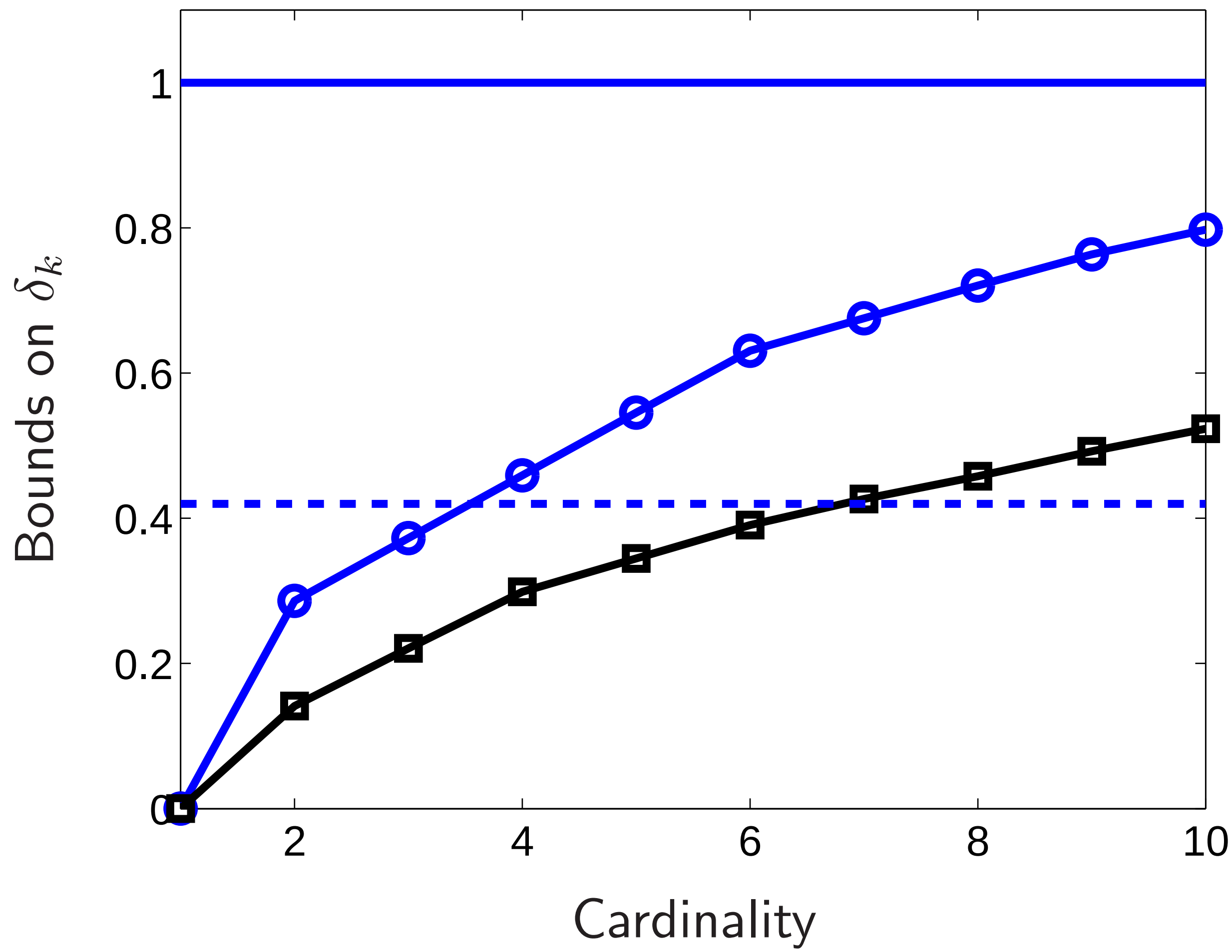
$$\begin{array}{ll} \text{max.} & x^T A x \\ \text{s.t.} & \|x\|_2 = 1 \\ & \textbf{Card}(x) \leq k, \end{array}$$

is bounded by

$$\begin{array}{ll} \text{max.} & \textbf{Tr}(AX) \\ \text{s.t.} & \textbf{Tr}(X) = 1 \\ & \mathbf{1}^T |X| \mathbf{1} \leq k \\ & X \succeq 0, \end{array}$$

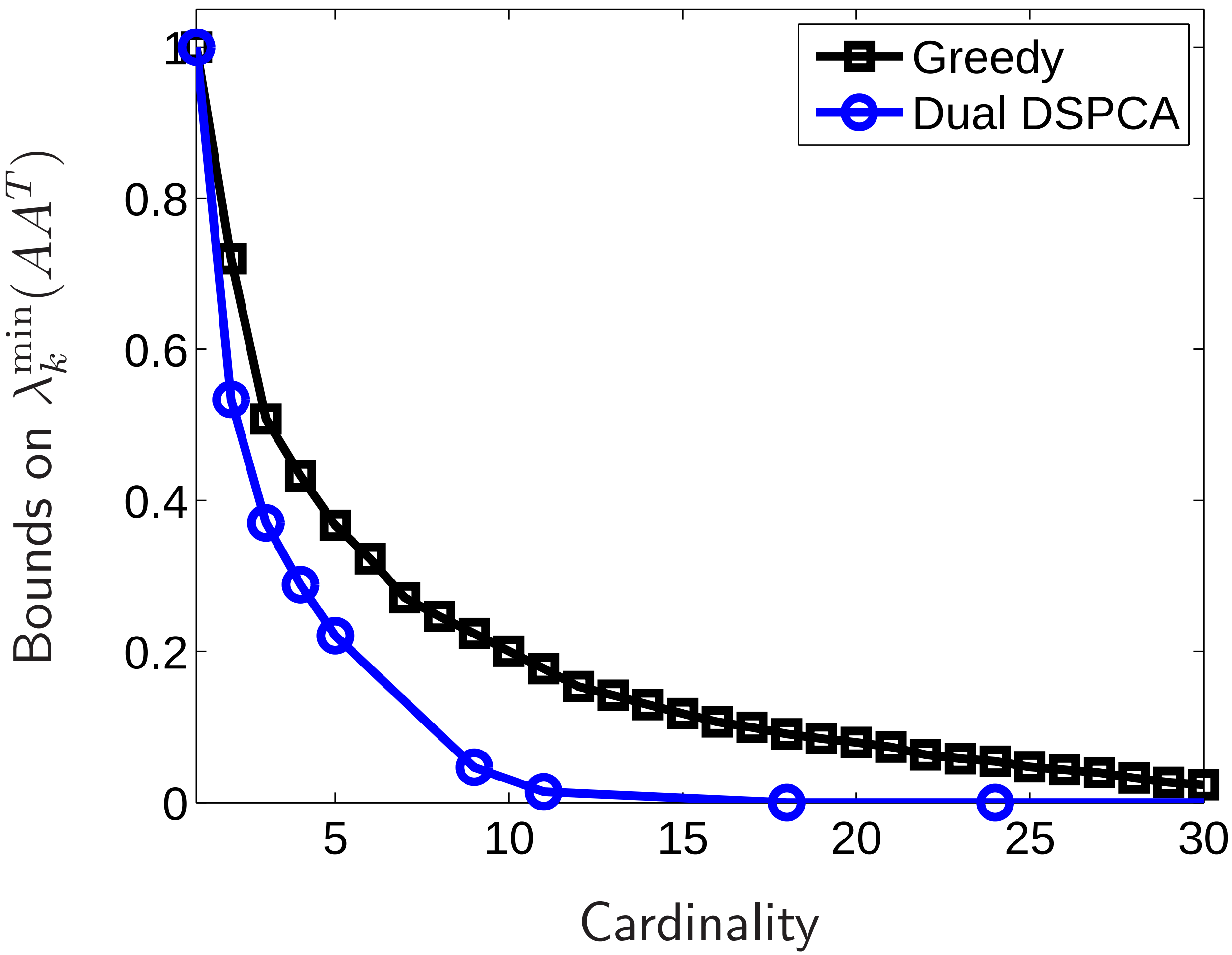
This is a (convex) **semidefinite program** in the variable $X \in \mathbf{S}^n$ and can be solved efficiently (roughly $O(n^4)$ in this case).

Testing the RIP



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

Testing the RIP



Lower bound on $\lambda_k^{\min}(AA^T)$ using the semidefinite relaxation, for a Bernoulli matrix of dimension $n = 100$, $p = 75$ (blue circles).
Upper bound 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 and Huo (2001) or Cohen et al. (2009) 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. (2009) 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

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

  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 and Nemirovski (2008).

Testing the NSP

- Use an **elimination result** for LMIs in (Boyd et al., 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

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

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

- Shows that the relaxation is **rotation invariant**.

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 and 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 $\alpha_k > 1/2$ and show that sparse recovery fails.

- **Concentration result**: let $X \in \mathbf{S}_n$, $x \sim \mathcal{N}(0, X)$ and $\delta > 0$, we have

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

- 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

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

- 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 and 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.
- Following Juditsky and 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: for any matrix A, feasible point of the SDP where $k = \sqrt{k^*}$ with objective greater than 1/2 in testing the NSP*).
- The LP relaxation in Juditsky and 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

- **Good news:** Tractable convex relaxations of sparse recovery conditions prove recovery at cardinality $k = O(\sqrt{k^*})$ for **any matrix** satisfying NSP at the optimal rate $k = O(k^*)$.

- **Bad news:** Testing recovery conditions on deterministic matrices at the optimal rate $O(m)$ 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

R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A Simple Proof of the Restricted Isometry Property for Random Matrices. *To appear*, 2007.

P. Bickel, Y. Ritov, and A. Tsybakov. Simultaneous analysis of lasso and dantzig selector. *Preprint Submitted to the Annals of Statistics*, 2007.

Stephen Boyd, Laurent El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*. SIAM, 1994.

E. Candès and T. Tao. The Dantzig selector: statistical estimation when p is much larger than n. *To appear in Annals of Statistics*, 2007.

E. J. Candès and T. Tao. Decoding by linear programming. *Information Theory, IEEE Transactions on*, 51(12):4203–4215, 2005.

E.J. Candes. The Restricted Isometry Property and Its Implications for Compressed Sensing. *CRAS*, 2008.

E.J. Candes and B. Recht. Exact matrix completion via convex optimization. *preprint*, 2008.

E.J. Candès and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Transactions on Information Theory*, 52(12):5406–5425, 2006.

K. Christensen, L. Danon, T. Scanlon, and P. Bak. Unified scaling law for earthquakes, 2002.

A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best k-term approximation. *Journal of the AMS*, 22(1):211–231, 2009.

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

D. L. Donoho. Neighborly polytopes and sparse solution of underdetermined linear equations. *Stanford dept. of statistics working paper*, 2004.

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

D.L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

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

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

M. Fazel, H. Hindi, and S. Boyd. A rank minimization heuristic with application to minimum order system approximation. *Proceedings American Control Conference*, 6:4734–4739, 2001.

M.X. Goemans and D.P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. ACM*, 42:1115–1145, 1995.

A. Juditsky and A.S. Nemirovski. On verifiable sufficient conditions for sparse signal recovery via ℓ_1 minimization. *ArXiv:0809.2650*, 2008.

C. Lemaréchal and F. Oustry. Semidefinite relaxations and Lagrangian duality with application to combinatorial optimization. *INRIA, Rapport de recherche*, 3710, 1999.

L. Lovász and A. Schrijver. Cones of matrices and set-functions and 0-1 optimization. *SIAM Journal on Optimization*, 1(2):166–190, 1991.

N. Meinshausen and B. Yu. Lasso-type recovery of sparse representations for high-dimensional data. *Annals of Statistics*, 37(1):246–270, 2008.

N. Meinshausen, G. Rocha, and B. Yu. A tale of three cousins: Lasso, l2boosting, and danzig. *Annals of Statistics*, 35(6):2373–2384, 2007.

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

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

P. Zhao and B. Yu. On model selection consistency of lasso. *Journal of Machine Learning Research*, 7:2541–2563, 2006.