Convex Optimization M2

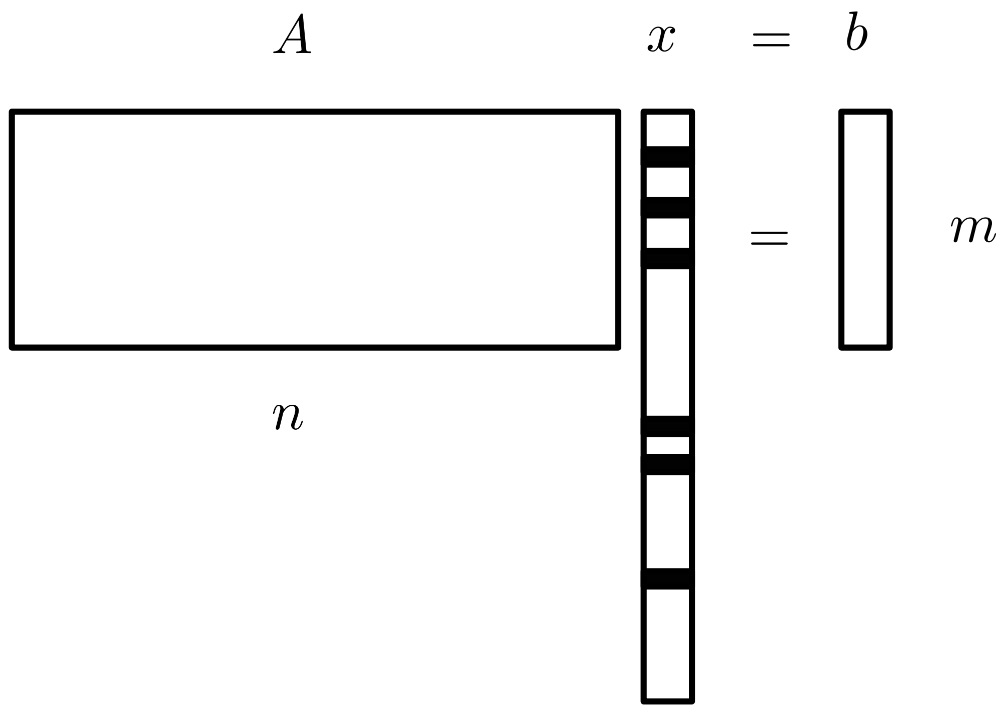
ℓ_1-recovery, compressed sensing

Today

- Sparsity, low complexity models.
- ℓ_1-recovery results: three approaches.
- Extensions: matrix completion, atomic norms.
- Algorithmic implications.

Low complexity models

Consider the following underdetermined linear system



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

Can we find the **sparsest** solution?

Introduction

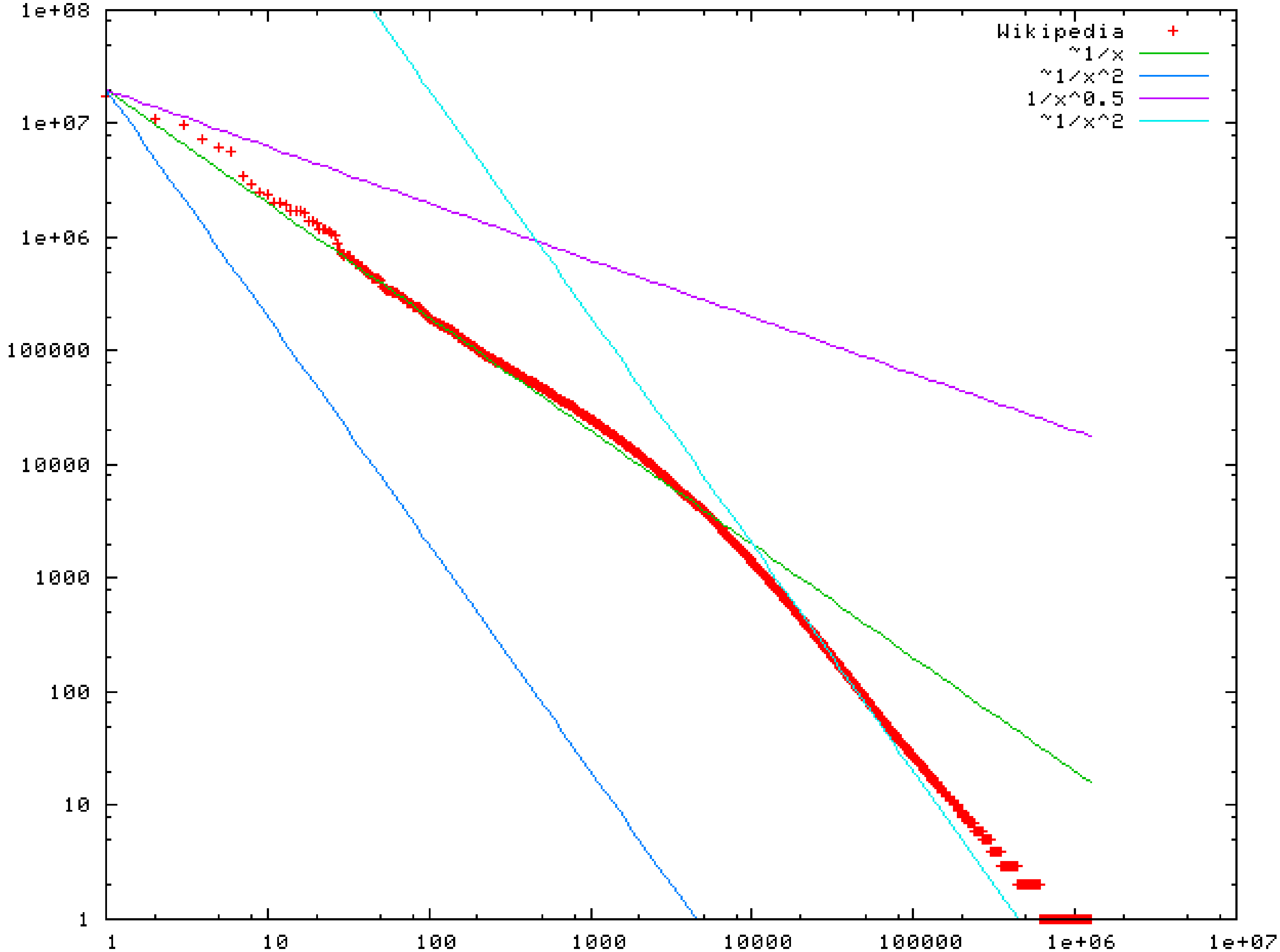
- **Signal processing:** We make a few measurements of a high dimensional signal, which admits a sparse representation in a well chosen basis (e.g. Fourier, wavelet). Can we reconstruct the signal exactly?
- **Coding:** Suppose we transmit a message which is corrupted by a few errors. How many errors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

Introduction

Why **sparsity**?

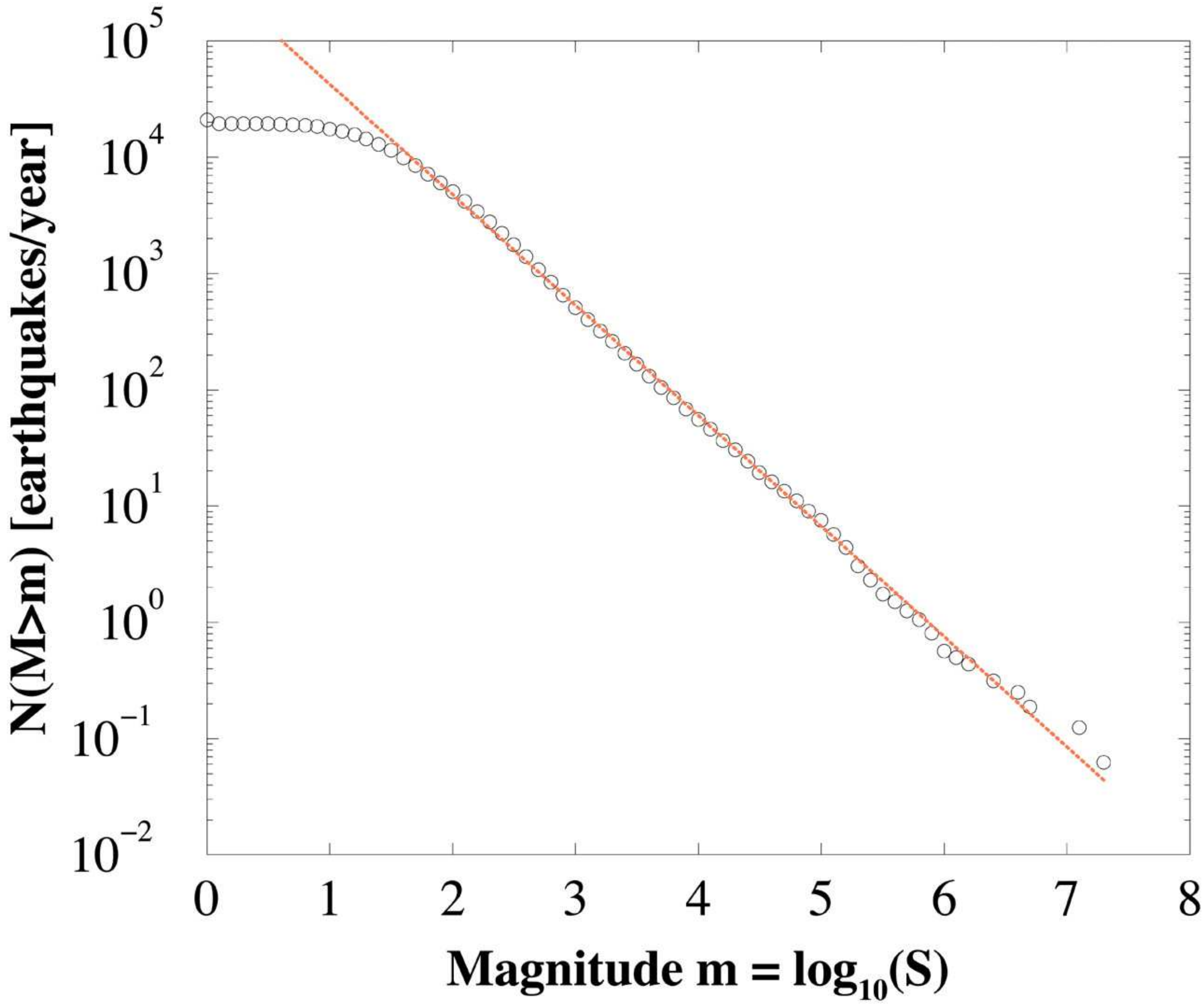
- Sparsity is a proxy for **power laws**. Most results stated here on sparse vectors apply to vectors with a power law decay in coefficient magnitude.
- Power laws appear everywhere. . .
  - Zipf law: word frequencies in natural language follow a power law.
  - Ranking: pagerank coefficients follow a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



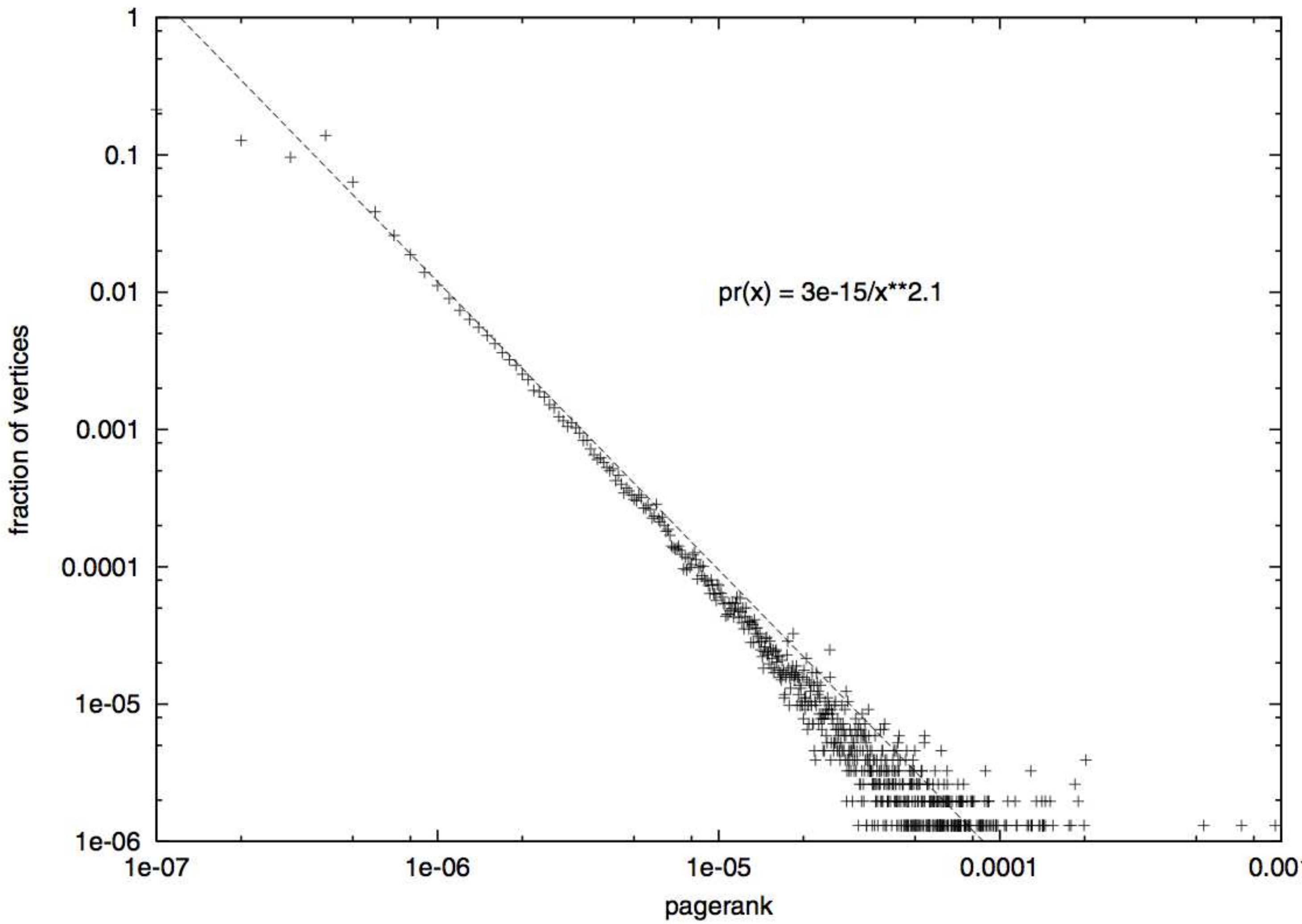
Frequency vs. word in Wikipedia (from Wikipedia).

Introduction



Frequency vs. magnitude for earthquakes worldwide. Christensen et al. [2002]

Introduction



Pages vs. Pagerank on web sample. Pandurangan et al. [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 $x \in \mathbb{R}^n$.

- 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

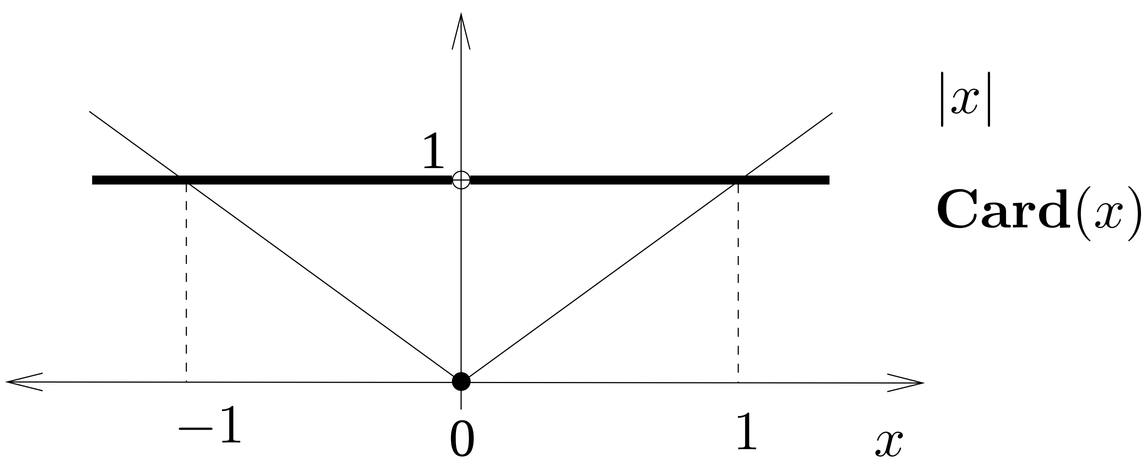
Assuming $|x| \leq 1$, we can replace:

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

with

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

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



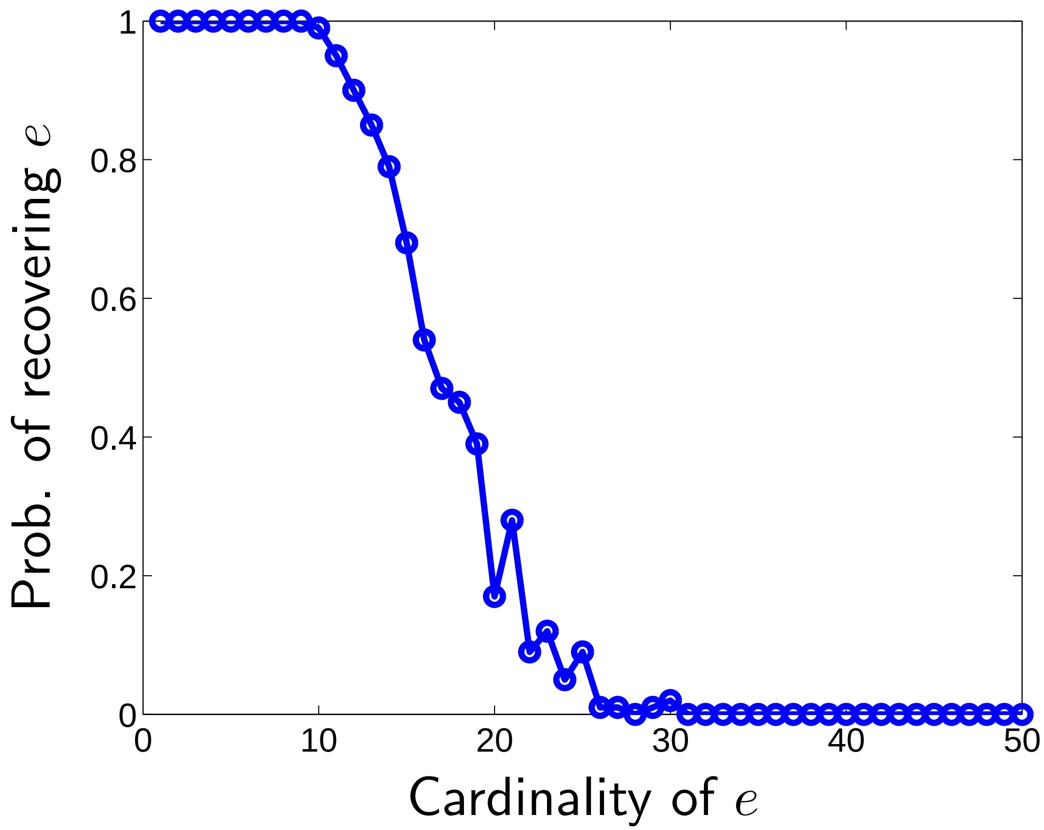
The l_1 norm is the **largest convex lower bound** on $\mathbf{Card}(x)$ in $[-1, 1]$.

Introduction

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

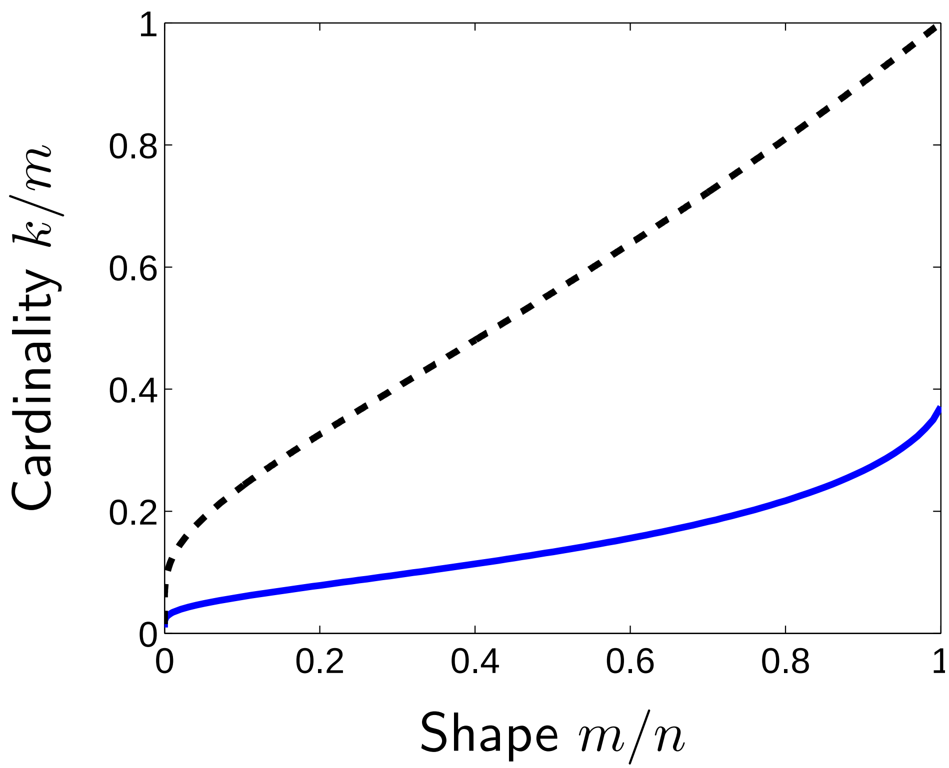
$$\begin{array}{ll}\text{minimize} & \|x\|_1 \\ \text{subject to} & Ax = Ae\end{array}$$

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



Introduction

- Donoho and Tanner [2005] and Candès and 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.



ℓ_1 recovery

Diameter

Kashin and Temlyakov [2007]: Simple relationship between the **diameter** of a section of the ℓ_1 ball and the size of signals recovered by ℓ_1-minimization.

Proposition

Diameter & Recovery threshold. *Given a coding matrix $A \in \mathbb{R}^{m \times n}$, suppose that there is some $k > 0$ such that*

$$\sup_{\substack{Ax=0\\ \|x\|_1 \leq 1}} \|x\|_2 \leq \frac{1}{\sqrt{k}} \tag{1}$$

then sparse recovery $x^{\mathrm{LP}} = u$ is guaranteed if $\mathbf{Card}(u) \leq k/4$, and

$$\|u - x^{\mathrm{LP}}\|_1 \leq \; 4 \min_{\{\mathbf{Card}(y) \leq k/16\}} \|u - y\|_1$$

where x^{LP} solves the ℓ_1-minimization LP and u is the true signal.

Diameter

Proof. Kashin and Temlyakov [2007]. Suppose

$$\sup_{\substack{Ax=0\\ \|x\|_1\leq 1}} \|x\|_2 \leq k^{-1/2}$$

Let u be the true signal, with $\mathbf{Card}(u) \leq k/4$. If x satisfies $Ax = 0$, for any support set Λ with $|\Lambda| \leq k/4$,

$$\sum_{i\in\Lambda} x_i \leq \sqrt{|\Lambda|}\ \|x\|_2 \leq \sqrt{|\Lambda|/k}\ \|x\|_1 \leq\ \|x\|_1/2,$$

Now let $\Lambda = \mathrm{supp}(u)$ and let $v \neq u$ such that $x = v - u$ satisfies $Ax = 0$, then

$$\|v\|_1 = \sum_{i\in\Lambda} |u_i+x_i| + \sum_{i\notin\Lambda} |x_i| \geq \sum_{i\in\Lambda} |u_i| - \sum_{i\in\Lambda} |x_i| + \sum_{i\notin\Lambda} |x_i| = \|u\|_1 + \|x\|_1 - 2\sum_{i\in\Lambda} |x_i|$$

and

$$\|x\|_1 - 2\sum_{i\in\Lambda} |x_i| > 0$$

means that $\|v\|_1 > \|u\|_1$, so $x^{\mathrm{LP}} = u$. The error bound follows from similar arg.

Diameter, low M^* estimate

Theorem

Low $\mathbf{M}^*$ estimate. *Let $E \subset \mathbb{R}^n$ be a subspace of codimension k chosen uniformly at random w.r.t. to the Haar measure on $\mathcal{G}_{n,n-k}$, then*

$$\mathbf{diam}(K \cap E) \leq c\sqrt{\frac{n}{k}} M(K^*) = c\sqrt{\frac{n}{k}} \int_{\mathbb{S}^{n-1}} \|x\|_{K^*} d\sigma(x)$$

with probability $1 - e^{-k}$, where c is an absolute constant.

Proof. See [Pajor and Tomczak-Jaegermann, 1986] for example.

We have $M(B_\infty^n) \sim \sqrt{\log n / n}$ asymptotically. This means that random sections of the ℓ_1 ball with dimension $n - k$ have diameter bounded by

$$\mathbf{diam}(B_1^n \cap E) \leq c\sqrt{\frac{\log n}{k}}$$

with high probability, where c is an absolute constant (a more precise analysis allows the $\log$ term to be replaced by $\log(n/k)$).

Sections of the ℓ_1 ball

The bound $\mathbf{diam}(B_1^n \cap E) \leq c\sqrt{\frac{\log n}{k}}$ means recovery of all signals with at most

$$O\left(\frac{k}{\log n}\right) \text{ coefficients, using k linear observations } Ae.$$

Results guaranteeing near-optimal bounds on the diameter can be traced back to Kashin and Dvoretzky's theorem.

- **Kashin decomposition** [Kashin, 1977]. Given $n = 2m$, there exists two orthogonal m-dimensional subspaces $E_1, E_2 \subset \mathbb{R}^n$ such that

$$\frac{1}{8}\|x\|_2 \leq \frac{1}{\sqrt{n}}\|x\|_1 \leq \|x\|_2, \quad \text{for all } x \in E_1 \cup E_2$$

- In fact, **most $\mathbf{m}$-dimensional subspaces** satisfy this relationship.

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], . . .

*

References

J. Bourgain, J. Lindenstrauss, and V. Milman. Minkowski sums and symmetrizations. *Geometric aspects of functional analysis*, pages 44–66, 1988.

E. Candes and B. Recht. Simple bounds for low-complexity model reconstruction. *arXiv preprint arXiv:1106.1474*, 2011.

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

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

V. Chandrasekaran, B. Rechtw, P.A. Parrilo, and A.S. Willskym. The convex geometry of linear inverse problems. *Arxiv preprint arXiv:1012.0621*, 2010.

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

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.

J.J. Fuchs. On sparse representations in arbitrary redundant bases. *Information Theory, IEEE Transactions on*, 50(6):1341–1344, 2004.

A. Giannopoulos, V.D. Milman, and A. Tsolomitis. Asymptotic formulas for the diameter of sections of symmetric convex bodies. *Journal of Functional Analysis*, 223(1):86–108, 2005.

A. A. Giannopoulos and V. D. Milman. On the diameter of proportional sections of a symmetric convex body. *International Math. Research Notices, No. 1 (1997) 5–19.*, (1):5–19, 1997.

Y. Gordon. On Milman's inequality and random subspaces which escape through a mesh in R^n. *Geometric Aspects of Functional Analysis*, pages 84–106, 1988.

B. Kashin. The widths of certain finite dimensional sets and classes of smooth functions. *Izv. Akad. Nauk SSSR Ser. Mat*, 41(2):334–351, 1977.

B.S. Kashin and V.N. Temlyakov. A remark on compressed sensing. *Mathematical notes*, 82(5):748–755, 2007.

A. Pajor and N. Tomczak-Jaegermann. Subspaces of small codimension of finite-dimensional banach spaces. *Proceedings of the American Mathematical Society*, 97(4):637–642, 1986.

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

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.

R. Vershynin. *Lectures in Geometric Functional Analysis*. In preparation, 2011. URL `http://www-personal.umich.edu/~romanv/papers/GFA-book/GFA-book.pdf`.