Tractable Upper Bounds on the Restricted Isometry Constant

Alex d'Aspremont, Francis Bach, Laurent El Ghaoui

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

Support from NSF, DHS and Google.

Introduction

Variable selection

- Error bounds in compressed sensing.
- Control model consistency and MSE in LASSO estimation of sparse models.

Sparse eigenvalues

- Extremal eigenvalues with a restriction on the cardinality of eigenvectors.
- Bounds using semidefinite relaxations.

Compressed Sensing

Following Candès & Tao (2005) and Donoho & Tanner (2005), recover a signal $f \in \mathbf{R}^n$ from corrupted measurements y:

$$y = Af + e,$$

where $A \in \mathbf{R}^{n \times m}$ is a coding matrix and $e \in \mathbf{R}^n$ is an unknown **sparse** vector of errors.

- Under certain conditions on $F \in \mathbf{R}^{p \times n}$ with $p < n$, such that $FA = 0$, this amounts to solving the following (combinatorial) problem:

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

- With stronger conditions on F, we only need to solve the **linear program**:

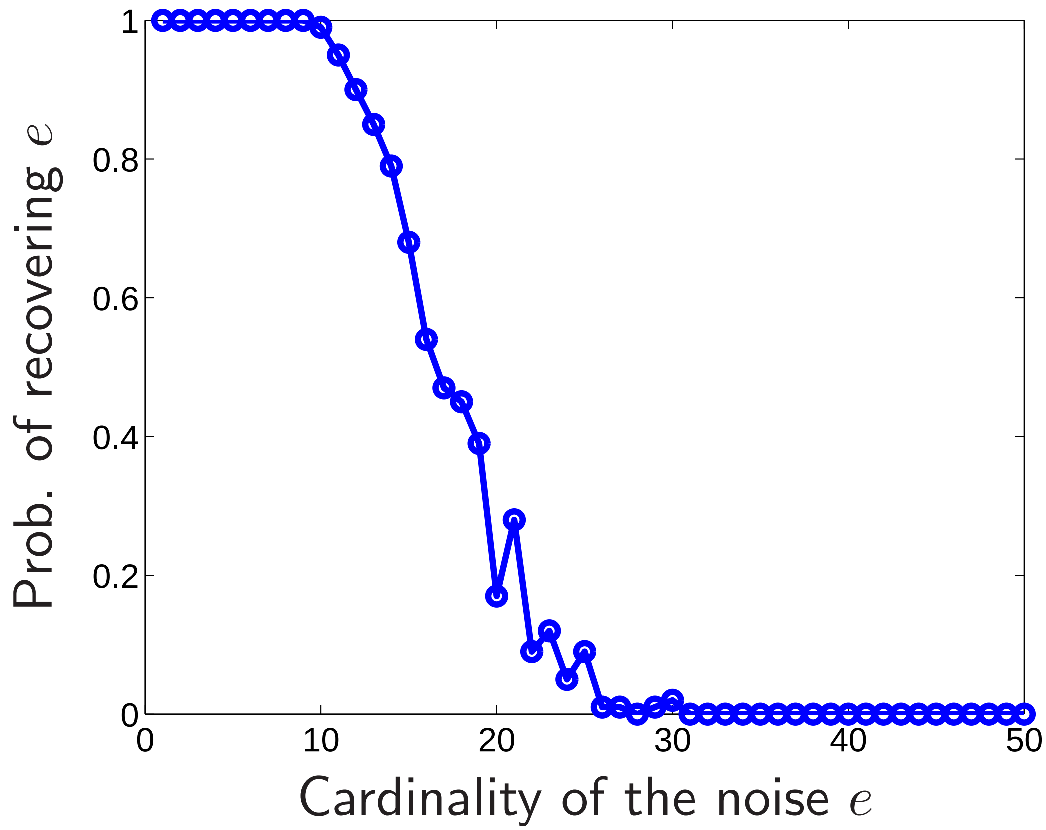
$$\begin{array}{ll} \text{minimize} & \|x\|_1 \\ \text{subject to} & Fx = Fy \end{array}$$

Compressed sensing: sparse recovery

Example: we plot probability of perfectly recovering e (hence f) by solving:

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

with $n = 50$ and $m = 30$.



Sparse eigenvalues

Given a matrix $C \in \mathbf{S}_n$.

Maximum eigenvalue. We solve:

$$\lambda_{\max}(C) = \begin{array}{ll} \text{max.} & x^T C x \\ \text{subject to} & \|x\| = 1, \end{array}$$

in the variable $x \in \mathbf{R}^n$.

Sparse maximum eigenvalue. We solve instead:

$$\lambda^k_{\max}(C) = \begin{array}{ll} \text{max.} & x^T C x \\ \text{subject to} & \mathbf{Card}(x) = k \\ & \|x\| = 1, \end{array}$$

in the variable $x \in \mathbf{R}^n$, where $\mathbf{Card}(x)$ is the number of nonzero coefficients in the vector x and $k > 0$ is a parameter controlling **sparsity**.

Outline

- Introduction
- **Variable selection**
  - Compressed Sensing
  - LASSO
- Sparse eigenvalues
  - Semidefinite Relaxation
  - Computational challenges
- Numerical Experiments

Compressed Sensing: Restricted Isometry Property

The key quantity here is the **restricted isometry** constant δ_S of the matrix F:

- Given $0 < S \leq n$, the constant δ_S is the smallest number such that:

$$(1-\delta_S)\|z\|_2^2 \leq \|F_I z\|_2^2 \leq (1+\delta_S)\|z\|_2^2,$$

  for all $z \in \mathbf{R}^{|I|}$, for any index subset $I \subset [1,n]$ of cardinality at most S.

- The constant δ_S measures how far sparse subsets of the columns of F are from being an isometry.

Compressed sensing: perfect recovery

Following Candès & Tao (2005), Donoho & Tanner (2005) (see also Cohen, Dahmen & DeVore (2006) for a simple proof). Suppose the error has cardinality

$$\mathbf{Card}(e) = k.$$

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

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

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

- If $\delta_{2k} < 1/3$, we can recover the error e by solving:

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

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

Compressed sensing: restricted isometry

The restricted isometry constant δ_S can be computed by solving the following **sparse eigenvalue** problem:

$$(1+\delta_S^{\max}) = \begin{array}{ll} \text{max.} & x^T(F^TF)x \\ \text{s. t.} & \mathbf{Card}(x) \leq S \\ & \|x\| = 1, \end{array}$$

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

- Candès & Tao (2005), Donoho & Tanner (2005) obtain an **asymptotic** proof that some random matrices satisfy the restricted isometry condition with **overwhelming probability** (i.e. exponentially small probability of failure) at an optimal rate.
- Numerical upper bounds for sparse eigenvalues prove **deterministically** and with **polynomial complexity** that a finite dimensional matrix satisfies the restricted isometry property with constant δ_S.

LASSO

Assume that observations $(Y_1, \ldots, Y_n)$ follow a linear model:

$$Y = X\beta + \epsilon$$

where $\beta \in \mathbf{R}^p$ and $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$. We define the LASSO estimator of β as:

$$\hat{\beta} = \mathrm{argmin}_\beta \|Y - X\beta\|_2^2 + \lambda\|\beta\|_1.$$

Consistency.

- Suppose β is **sparse** with cardinality $s(n)$, Meinshausen & Yu (2007) show:

$$\lim_{n\to\infty} \mathbf{P}\left(\|\beta - \hat{\beta}\|_2^2 \leq M\sigma^2 \,\frac{s(n)\log p(n)}{n}\, \frac{e_n^2}{\lambda_{\min}^{s(n)e_n^2}(X^TX)} \right) = 1$$

  if

$$\liminf_{n\to\infty} e_n \lambda_{\min}^{e_n^2 s(n)}(X^TX) \geq 18\lambda_{\min}^{s(n)+\min\{n,p\}}(X^TX)$$

- Meinshausen & Yu (2007) also show sign consitency based on sparse eigenvalues. Similar non-asymptotic result by Candès & Tao (2007).

Outline

- Introduction
- Variable selection
  - Compressed Sensing
  - LASSO
- **Sparse eigenvalues**
  - Semidefinite Relaxation
  - Computational challenges
- Numerical Experiments

Sparse eigenvalues

Given a matrix $C \in \mathbf{S}_n$.

Maximum eigenvalue. We solve:

$$\lambda_{\max}(C) = \begin{array}{ll} \text{max.} & x^T C x \\ \text{subject to} & \|x\| = 1, \end{array}$$

in the variable $x \in \mathbf{R}^n$.

Sparse maximum eigenvalue. We solve instead:

$$\lambda^k_{\max}(C) = \begin{array}{ll} \text{max.} & x^T C x \\ \text{subject to} & \mathbf{Card}(x) = k \\ & \|x\| = 1, \end{array}$$

in the variable $x \in \mathbf{R}^n$, where $\mathbf{Card}(x)$ is the number of nonzero coefficients in the vector x and $k > 0$ is a parameter controlling **sparsity**.

Related Work

This problem is usually called sparse PCA.

- Cadima & Jolliffe (1995): the loadings with small absolute value are thresholded to zero.
- Johnstone & Lu (2004) apply this to ECG data and show model consistency.
- Zou, Hastie & Tibshirani (2006), non-convex algo. (SPCA) based on a l_1 penalized representation of PCA as a regression problem.
- Non-convex methods (SCoTLASS) by Jolliffe, Trendafilov & Uddin (2003).
- A greedy search algorithm by Moghaddam, Weiss & Avidan (2006).

All these codes produce approximate solutions, hence lower bounds on sparse maximum eigenvalues. Here we seek **upper** bounds.

Semidefinite relaxation

In d'Aspremont, El Ghaoui, Jordan & Lanckriet (2007), we combine two classic relaxation techniques:

- The lifting procedure à la MAXCUT by Goemans & Williamson (1995).
- A ℓ_1 norm relaxation of the cardinality constraint. Used in basis pursuit by Chen, Donoho & Saunders (2001), LASSO by Tibshirani (1996), etc.

Start from:

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

write everything in terms of $X = xx^T$, or also $X \succeq 0,\ \ \mathbf{Rank}(X) = 1$:

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

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 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**): $\mathbf{1}^T|X|\mathbf{1} \leq k$.
- Simply drop the rank constraint.

Semidefinite Programming

Semidefinite relaxation:

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

- This is a **semidefinite program** in the variable $X \in \mathbf{S}^n$...
- The optimum value of this semidefinite relaxation is an **upper** bound on the sparse maximum eigenvalue. Any dual feasible point will also produce a valid bound.
- Another relaxation discussed in d'Aspremont, Bach & El Ghaoui (2007).
- Solve small problems (a few hundred variables) using IP solvers, etc.

Solution: use first order algorithms...

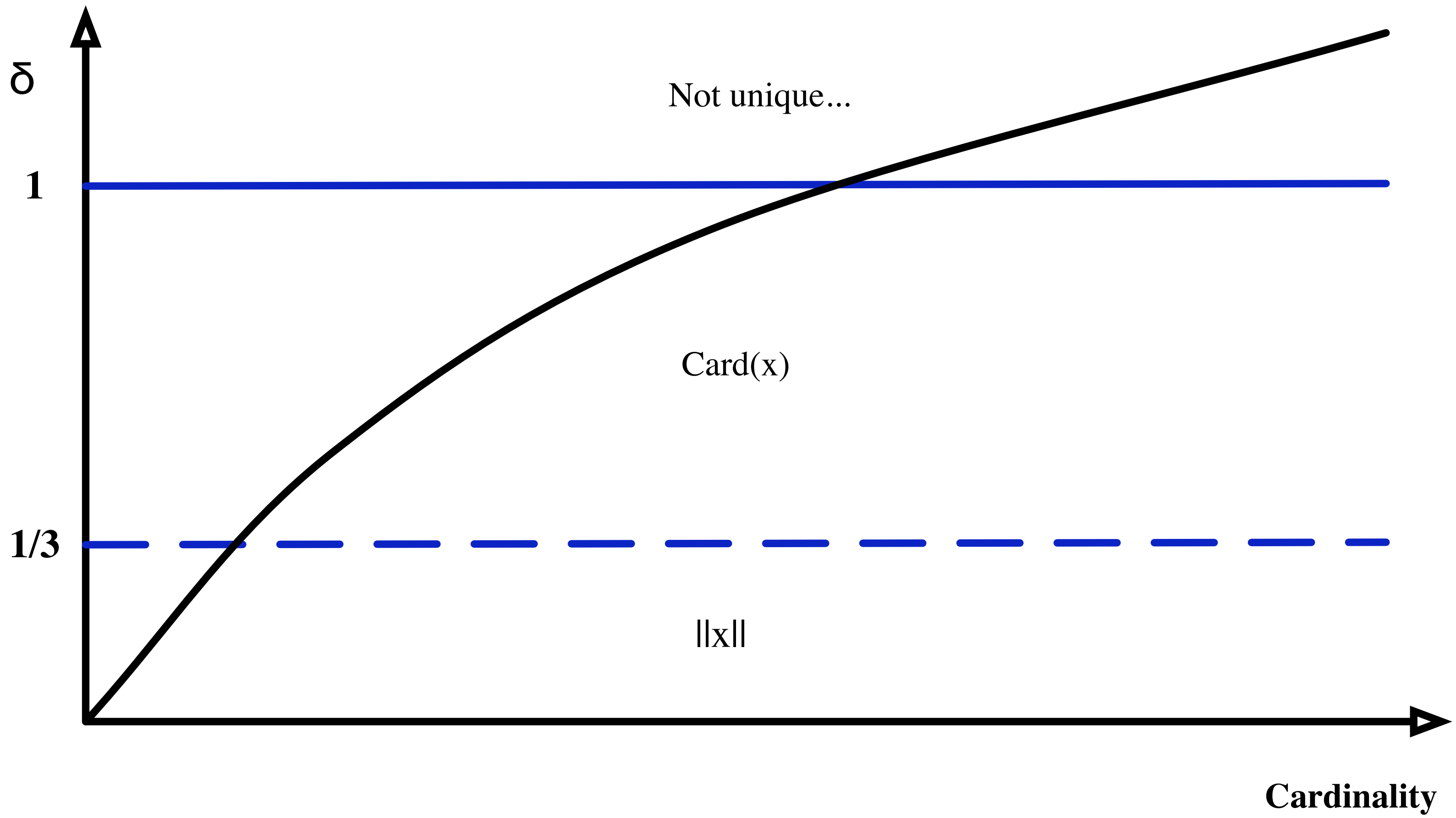
Outline

- Introduction
- Variable selection
  - Compressed Sensing
  - LASSO
- Sparse eigenvalues
  - Semidefinite Relaxation
  - Computational challenges
- **Numerical Experiments**

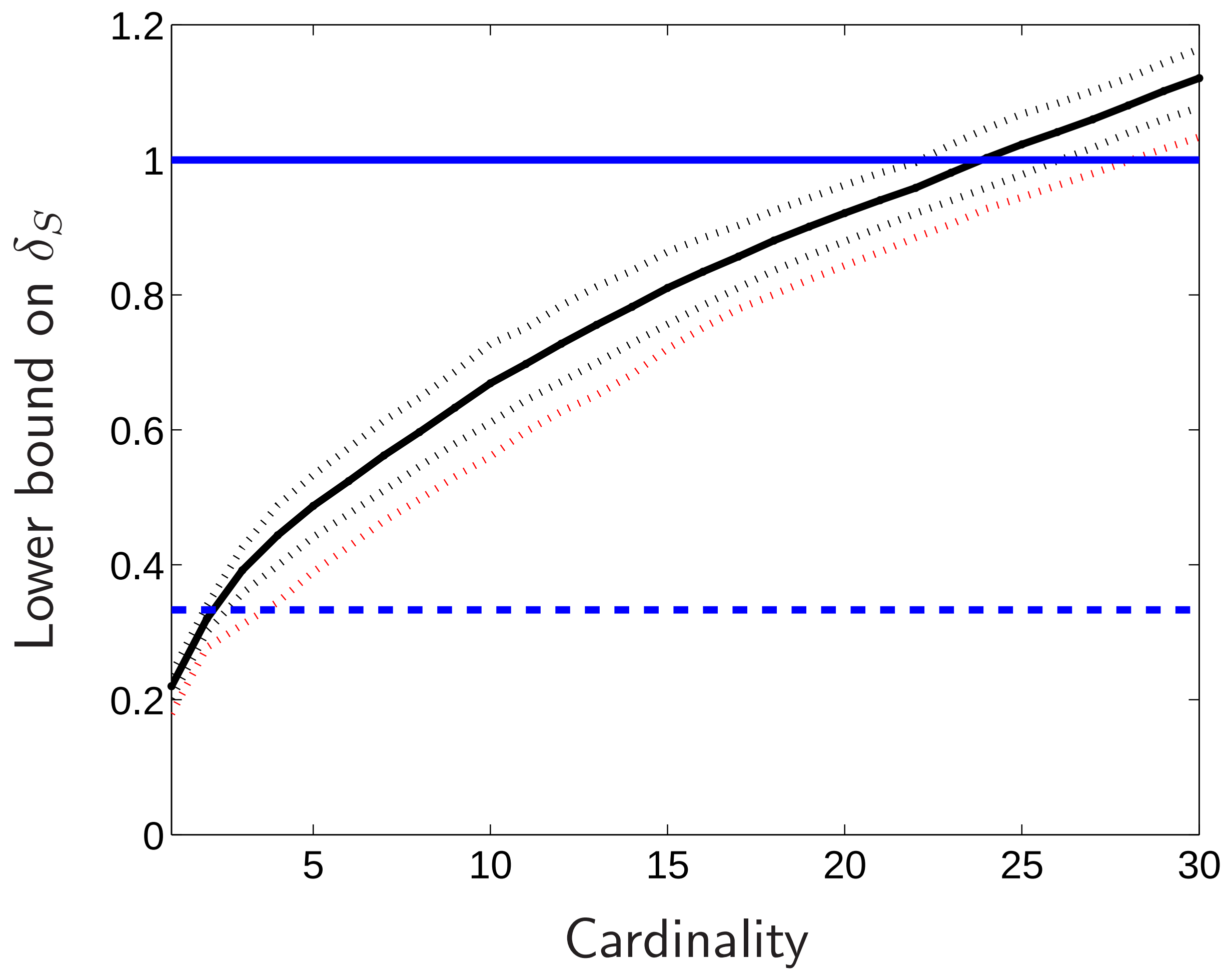
Compressed sensing: restricted isometry

- Generate random Gaussian or Bernoulli matrices X as in Candès & Tao (2005).
- Compute lower bounds on $\delta_S(X^TX)$ by finding approximate sparse eigenvectors.
- Compute **upper bounds** on $\delta_S(X^TX)$ by solving the semidefinite relaxation.
- Compare these bounds with asymptotic ones.

Compressed sensing: restricted isometry

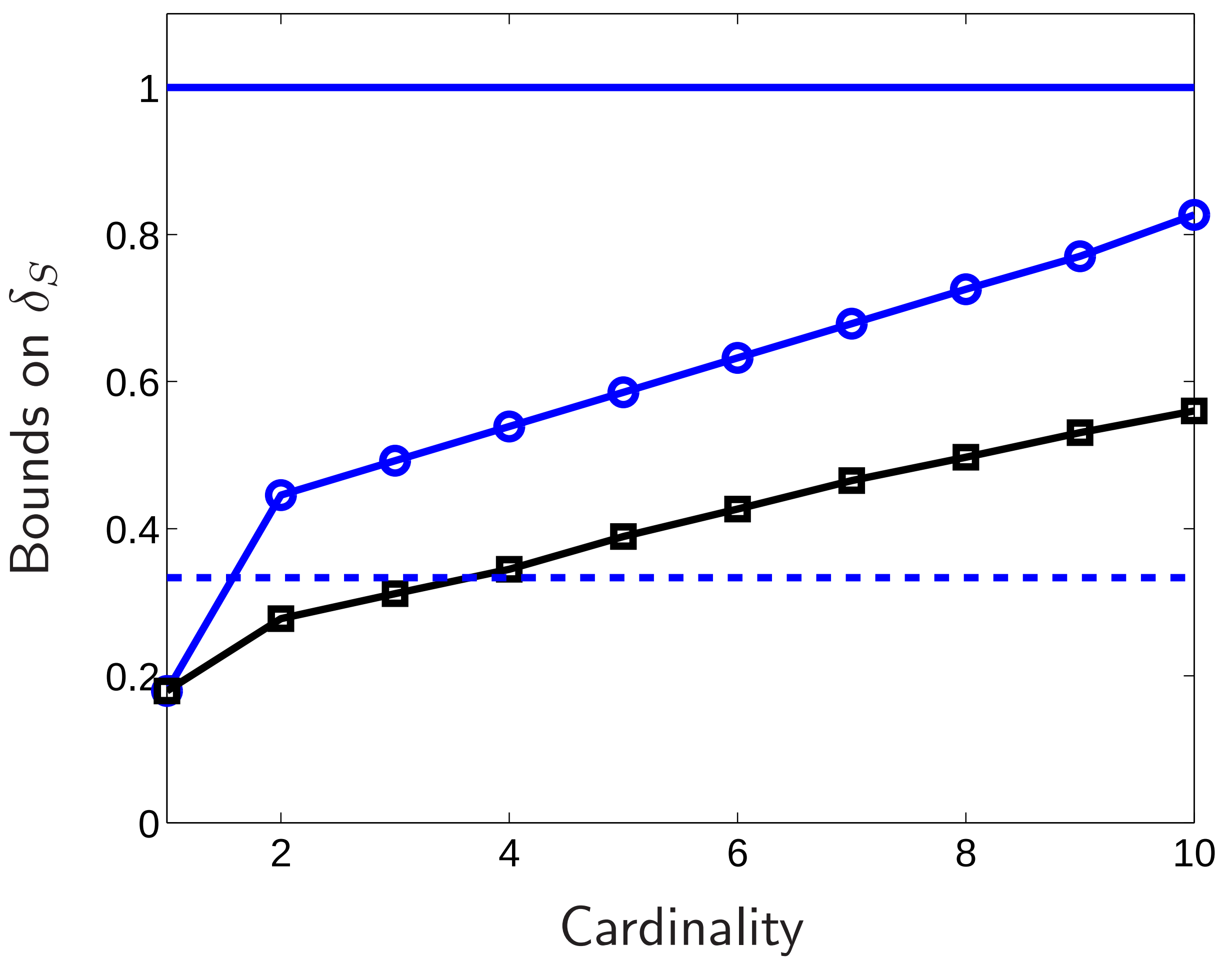


Compressed sensing: restricted isometry



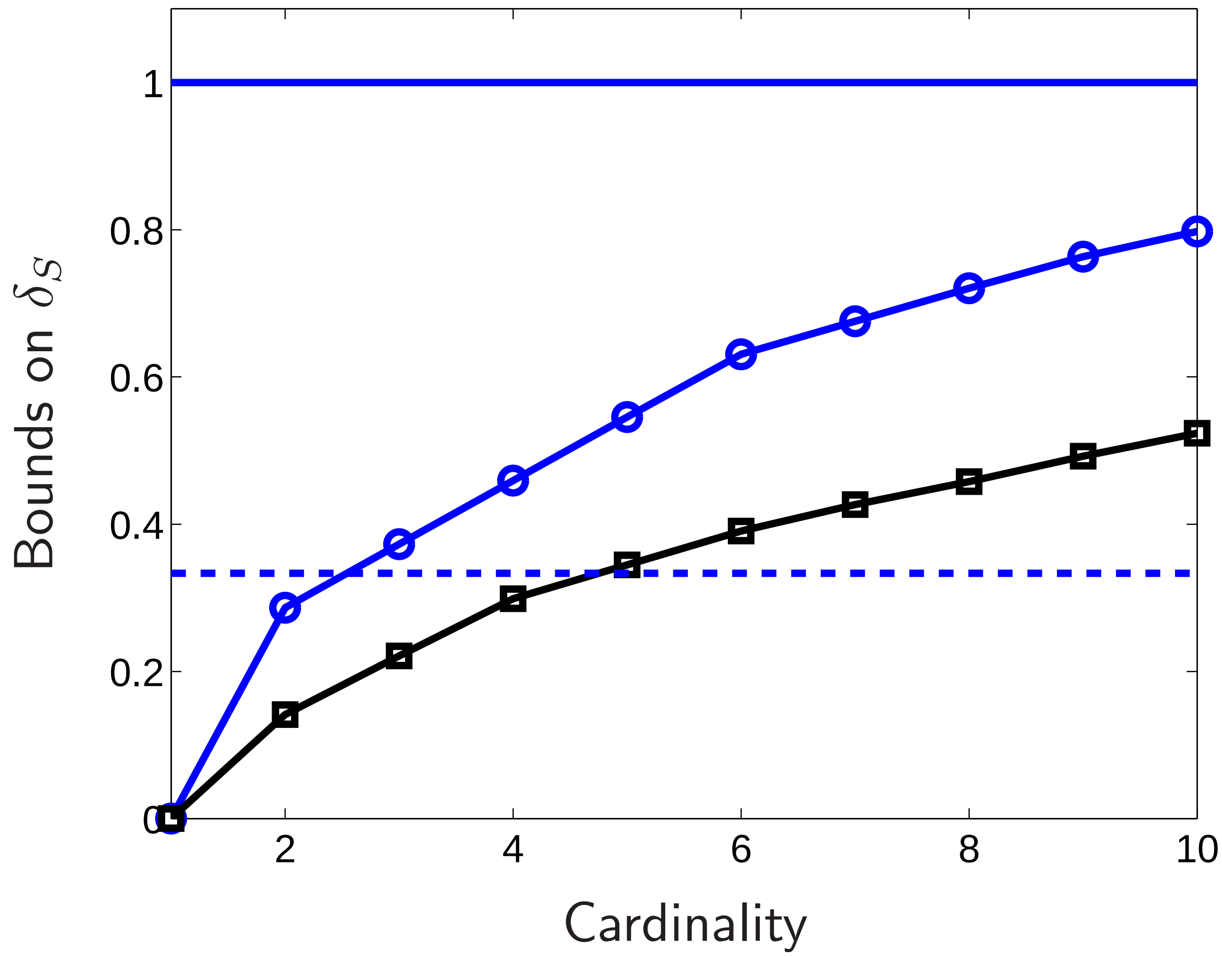
Lower bound on δ_S using approximate sparse eigenvectors, for Gaussian matrices of dimension $n = 1000$, $p = 750$, over 100 samples (solid line). Black dotted lines at plus and minus three stdev. Red dotted line at the pointwise sample minimum.

Compressed sensing: restricted isometry



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

Compressed sensing: restricted isometry



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).

Compressed sensing: restricted isometry

Semidefinite relaxation vs. probabilistic bound on δ_S for Gaussian matrices.

- For $n = 1000$, $p = 750$ and a confidence of 99%:

S	1	2	3	4	5	6	7	8	9	10
Prob.	0.67	0.87	1.04	1.21	1.37	1.55	1.78	...	...	...
SDP	0.18	0.45	0.49	0.54	0.59	0.63	0.68	0.73	0.77	0.83

- Equivalent n for $p/n = .75$ and a confidence of 99%:

S	1	2	3	4	5	6	7	8	9	10
Prob.	14200	3700	4200	4500	4700	4700	4700	4700	4700	4500
SDP	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Conclusion

- Relatively tight bounds on small matrices.
- Still computationally challenging.

- Slides online.
- Source code, binaries available at:

`www.princeton.edu/~aspremon/DSPCA.htm`

- More recent results in d'Aspremont, Bach & El Ghaoui (2007).

References

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

Cadima, J. & Jolliffe, I. T. (1995), 'Loadings and correlations in the interpretation of principal components', *Journal of Applied Statistics* **22**, 203–214.

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

Candès, E. & Tao, T. (2007), 'The Dantzig selector: statistical estimation when$ p$ is much larger than$ n$', *To appear in Annals of Statistics* .

Chen, S., Donoho, D. & Saunders, M. (2001), 'Atomic decomposition by basis pursuit.', *SIAM Review* **43**(1), 129–159.

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

d'Aspremont, A., Bach, F. R. & El Ghaoui, L. (2007), Full regularization path for sparse principal component analysis, *in* 'Proceedings of the 24th international conference on Machine learning', pp. 177–184.

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. 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.

Goemans, M. & Williamson, D. (1995), 'Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming', *J. ACM* **42**, 1115–1145.

Johnstone, I. & Lu, A. Y. (2004), 'Sparse principal components analysis', *Working Paper, Stanford department of statistics* .

Jolliffe, I. T., Trendafilov, N. & Uddin, M. (2003), 'A modified principal component technique based on the LASSO', *Journal of Computational and Graphical Statistics* **12**, 531–547.

Meinshausen, N. & Yu, B. (2007), Lasso-type recovery of sparse representations for highdimensional data, Technical report, To appear in Annals of Statistics.

Moghaddam, B., Weiss, Y. & Avidan, S. (2006), 'Spectral bounds for sparse PCA: Exact and greedy algorithms', *Advances in Neural Information Processing Systems* **18**.

Tibshirani, R. (1996), 'Regression shrinkage and selection via the LASSO', *Journal of the Royal statistical society, series B* **58**(1), 267–288.

Zou, H., Hastie, T. & Tibshirani, R. (2006), 'Sparse Principal Component Analysis', *Journal of Computational & Graphical Statistics* **15**(2), 265–286.