SCIENCES
SORBONNE
UNIVERSITÉ

Structured measurements

Claire Boyer

Recall

What we have seen

- Subgaussian random matrices = optimal measurement matrices for CS
- But "completely random" matrices $\Rightarrow$ limited interest for practical purposes
- Difficult to store

What we would like

However, structure is important for several reasons:

- to model physical or other constraints;
- structure often allows to have fast matrix–vector multiplication algorithms
  - see `FFT`
  - crucial for speedups in recovery algorithms (including ℓ^1-minimization)
  - towards large-scale problems

Towards structured measurements?

Adding structure?

- May structured random matrices provide recovery guarantees similar to the ones for subgaussian random matrices?

Structured random matrix?

- Structured matrix that is generated by a random choice of parameters
- $\rightsquigarrow$ Randomly sampling functions whose expansion in a bounded orthonormal system (BOS) is sparse or compressible. ex:
  - sampling sparse trigonometric polynomials
  - taking random samples from the Fourier transform of a sparse vector

Different kind of results

- Analysis of such matrices becomes more involved than the analysis for subgaussian random matrices because the entries are not independent anymore.
- $\rightsquigarrow$ "non-uniform" results easier to derive.

- finite-dimensional setting
- orthonormal basis
- whose elements are uniformly bounded in the L^∞-norm
  ex: in a discrete setup, the resulting matrix is a random partial Fourier matrix, which was the first structured random matrix investigated in CS.

Summary

Definitions

Let $A_0 \in \mathbb{K}^{d \times d}$ be a full sensing orthogonal matrix (that can physically model the acquisition),

$$A_0 = \begin{pmatrix} a_1^* \\ \vdots \\ a_d^* \end{pmatrix}.$$

A_0 is supposed to be orthogonal, i.e.

$$A_0^* A_0 = \sum_{i=1}^{d} a_i a_i^* = \mathrm{Id}_d,$$

and then $(a_i)_{1 \le i \le d}$ forms a orthonormal basis of $\mathbb{K}^d$.

Definitions

Suppose that we are subsampling this orthonormal basis according to some probability distribution π, i.e. the probability to choose the j-th element of this basis is π_j

$$\mathbb{P}(J = j) = \pi_j, \qquad \text{for} \quad j = 1, \dots, d.$$

So by using the independent r.v. $(J_k)_{1 \leq k \leq m}$ with the same distribution than J, one can construct the following sensing matrix A:

Sensing matrix with BOS

$$A = \frac{1}{\sqrt{m}} \begin{pmatrix} \frac{a_{J_1}^*}{\sqrt{\pi_{J_1}}} \\ \vdots \\ \frac{a_{J_m}^*}{\sqrt{\pi_{J_m}}} \end{pmatrix} \in \mathbb{K}^{m \times d}. \tag{1}$$

About isotropy I

The normalization may appear obscure, but it ensures the isotropy condition:

- consider the random vector a such that

$$\mathbb{P}(a = a_k/\sqrt{\pi_k}) = \pi_k, \qquad 1 \leq k \leq d.$$

Then, for any vector $x \in \mathbb{K}^d$,

$$\begin{aligned}
\mathbb{E}|\langle a, x\rangle|^2 &= \mathbb{E}x^* a a^* x = x^* \mathbb{E}(aa^*)x \\
&= x^* \left(\sum_{k=1}^{d} \mathbb{P}(a = a_k/\sqrt{\pi_k}) \frac{a_k}{\sqrt{\pi_k}} \frac{a_k^*}{\sqrt{\pi_k}} \right) x \\
&= x^* \left(\sum_{k=1}^{d} \pi_k \frac{a_k}{\sqrt{\pi_k}} \frac{a_k^*}{\sqrt{\pi_k}} \right) x = x^* \left(\sum_{k=1}^{d} a_k a_k^* \right) x \\
&= x^* A_0^* A_0 x \\
&= x^* x = \|x\|_2^2.
\end{aligned}$$

About isotropy II

- Again, the renormalization in $1/\sqrt{m}$ ensures that A is at least an isometry in expectation: for any vector $x \in \mathbb{K}^d$,

$$\begin{aligned}\mathbb{E}\|Ax\|_2^2 = \mathbb{E}x^*A^*Ax = \mathbb{E}x^*\left(\frac{1}{m}\sum_k \frac{a_{J_k}a_{J_k}^*}{\pi_{J_k}}\right)x \\ = x^*\frac{m}{m}\mathbb{E}aa^*x \\ = \|x\|_2^2.\end{aligned}$$

Coherence I

One can define the coherence of such a system as follows.

Definition

We call the coherence of a bounded orthonormal system (BOS) the following quantity:

$$\mu = \max_{1 \leq k \leq d} \frac{\|a_k\|_\infty^2}{\pi_k}.$$

Coherence II

Coherence measures the ability to catch information by averaging

Since rows of A are picked in a BOS, their ℓ^2-norm is 1. The ℓ_∞-norm is then able to measure if the mass in the row vectors is localized or not: consider a vector u is such that $\|u\|_2^2 = 1$,

- if $\|u\|_\infty$ is large, it means that a few coefficients in u carries the total energy of the vector, i.e. a few coefficients in u are non-zero;
- if $\|u\|_\infty$ is small, it means that the mass in u is spread accross all its components, i.e. all the components in u are non-zero, but they have to be small such that the sum of their square gives 1.

Using the definition of the sensing matrix in (1), the data vector y can then be written as follows

$$y = Ax \in \mathbb{K}^m.$$

How to use this structured randomness in practice?
Even if randomness is still part of the definition, and might be hard to be found in applications, the sensing matrix A considered here can be at least easily handled on a computer.

Example with Fourier measurements

- Let A_0 be the Fourier matrix (with entries $(e^{2i\pi k\ell/d}/\sqrt{d})_{k\ell}$.
- Choose π to be the uniform distribution $\Rightarrow$ uniform subsampling in the frequency domain..
- Draw m frequencies and store the location in a mask Ω which coefficients are equal to 1 if the corresponding frequency is sampled, and 0 otherwise.
- **No need** of forming the sensing matrix A: use fast transforms (`fft`) and then perform pointwise product between the result and the mask to keep the selected Fourier coefficients.

Summary

BOS & Coherence

Here are some BOS that could be used for random subsampling:

- **Fourier measurements**:
  - the sparsity is then in the canonical basis,
  - for such an acquisition system, one gets for all k,

$$\|a_k\|_\infty^2 = 1/d.$$

  - The Fourier basis and the canonical basis are thus maximally incoherent.
  - Thus if we uniformly subsample the Fourier coefficients, meaning that $\pi_k = 1/d$, one gets

$$\mu = 1.$$

BOS & Coherence

- **Hadamard measurements**:
  - The Hadamard transform is maximally incoherent with the canonical basis.
  - The Hadamard transform is $H = H_p \in \mathbb{R}^{2^p \times 2^p}$ is defined recursively by

$$H_p = \frac{1}{\sqrt{2}} \begin{pmatrix} H_{p-1} & H_{p-1} \\ H_{p-1} & -H_{p-1} \end{pmatrix} \qquad \text{and} \qquad H_0 = 1$$

  - It can be computed in $O(d' log(d'))$ time and is useful in modelling systems where there are "on/off" measurements, such as the single-pixel camera, or in fluorescence microscopy.

BOS & Coherence

- **Fourier-wavelets case**:
  - the measurements are performed in the Fourier basis,
  - and the sparsity is supposed to be in the wavelet domain.
  - In that case, A_0 can be viewed as

$$A_0 = \mathcal{F}H$$

    where $\mathcal{F} \in \mathbb{C}^{d \times d}$ is the Fourier matrix and $H \in \mathbb{R}^{d \times d}$ is the Haar wavelets transform.
  - Both are orthogonal transforms, and so is A_0.

- What about the sampling distribution?
- What about the coherence?

Define the function $j : [|d|] \to \{0, \dots, J\}$ where J is the maximum level in the wavelet decomposition. The function j maps the coefficient index to its corresponding level, one gets

$$|(A_0)_{k\ell}|^2 \lesssim 2^{-j(k)} 2^{-|j(k)-j(\ell)|}, \quad 1 \leq k, \ell \leq d.$$

$$\mu = \max_{1 \leq k \leq d} \frac{\|a_k\|_\infty^2}{\pi_k}$$

How to choose π_k?

One should choose $\pi_k \propto \|a_k\|_\infty^2$ i.e. choosing with larger probability the more coherent rows, which leads to

$$\mu = \sum_{k=1}^{d} \|a_k\|_\infty^2 \rightsquigarrow \log(d).$$

Summary

RIP for BOS I

For such sensing matrices, one can derive uniform results, i.e. recovery guarantees for *all* s-sparse vectors.

Theorem (RIP for BOS)

Let $A \in \mathbb{K}^{m \times d}$ *be the sensing matrix constructed as in* (1). *If*

$$m \geq c_0 \frac{\mu s}{\delta^2} \ln^4(d),$$

then with probability $1 - d^{-\ln^3(d)}$, *the restricted isometry constant* δ_s *of* A *satsifies* $\delta_s \leq \delta$.

RIP for BOS

- The price to pay for adding structure in the acquisition matrix (compared to Gaussian matrices) leads to extra log factors in the required number of measurements.
  Note that $\ln^4(d)$ is not so small... for $d = 1000$, $\ln(d) \simeq 6.9$ but $\ln^4(d) > 2d$.
- The concentration is also slower than Gaussian matrices: look at the probability of satisfying RIP for BOS, one gets $1 - d^{-\ln^3(d)}$ versus $1 - 2\exp(-c_0 m)$ for Gaussian matrices.
- The ϵ-net techniques with union bound used for Gaussian matrices leads to suboptimal results in this case: indeed, the bound for m scales quadratically in s, while the desired estimate obeys a linear scaling up to logarithmic factors ($\rightsquigarrow$ chaining techniques or Dudley's inequality).

Exercise (Showing net-techniques are insufficient for BOS - exam 2017)

The goal of this exercise is to show that the net-techniques lead to suboptimal result. Suppose that we are given an orthonormal basis $(v_1, \dots, v_d)$ *of* $\mathbb{R}^d$*, such that*

$$\|v_i\|_\infty^2 \leq \mu \quad a.s.$$

The sensing matrix $A \in \mathbb{R}^{m \times d}$ *is formed by concatenating independent random vectors* $a_1, \dots, a_m$ *uniformly drawn in* $(v_1, \dots, v_d)$*, meaning that the* (a_i)*'s are independent copies of the random vector* a*, such that for all* $i = 1, \dots, n$*,* $\mathbb{P}(a = v_i) = 1/d$*, and*

$$A = \begin{pmatrix} a_1^* \\ \vdots \\ a_m^* \end{pmatrix}.$$

Exercise (Showing net-techniques are insufficient for BOS - exam 2017)

1. *Using Bernstein's concentration result for matrices, show that for all $t > 0$*

$$\mathbb{P}\left(\|A_S^* A_S - I_S\|_{2\to 2} \geq 1/8\right) \leq 2s \exp\left(-c\frac{m}{s\mu + 1}\right).$$

2. *Using the previous concentration result, show that such matrix A satisfies RIP with a RIP constant $1/2$ for a certain bound on m.*
3. *Conclude on the tightness of the obtained bound on m.*

Helper: Matrix Bernstein inequality

Proposition (Matrix Bernstein inequality)

Consider a finite sequence (M_k) of independent, random matrices with dimension $d_1 \times d_2$. Assume that

$$\mathbb{E} M_k = 0 \qquad \|M_k\|_{2\to 2} \leq R \quad a.s.$$

Introduce the random matrix

$$Z = \sum_k M_k,$$

compute the variance parameter

$$\sigma^2 = \sigma^2(Z) = \max\left(\|\mathbb{E}(ZZ^*)\|_{2\to 2}, \|\mathbb{E}(Z^* Z)\|_{2\to 2}\right).$$

One has for $t \geq 0$

$$\mathbb{P}\left(\|Z\|_{2\to 2} \geq t\right) \leq (d_1 + d_2) \exp\left(-\frac{t^2/2}{\sigma^2 + Rt/3}\right)$$

Some ideas of proof for RIP with BOS

- Symmetrization
- Moments of supremum of Rademacher process
- Linking moments to tails.

Blackboard time.

Lemma (Symmetrization)

Consider $(\xi_j)_{1 \le j \le m}$ a sequence of independent random vectors in $\mathbb{C}^d$ equipped with a norm $\|\cdot\|$, with expectation $\mathbb{E}\xi_j = x_j$. Then for $1 \le p < \infty$,

$$\left(\mathbb{E} \left\| \sum_{j=1}^{m} (\xi_j - x_j) \right\|^p \right)^{1/p} \le 2 \left(\mathbb{E} \left\| \sum_{j=1}^{m} \epsilon_j \xi_j \right\|^p \right)^{1/p}$$

where $(\epsilon_j)_j$ is a Rademacher sequence independent of $(\xi_j)_j$.

Easy proof.

Towards deterministic compressed sensing

- Using the concept of coherence, it is easy (see e.g. [Foucart & Rauhut, Chapter 5]) to deterministically produce matrices that allow for stable and robust s-sparse recovery with a number m of rows of the order of s^2.
- Reducing this number of rows to $m \sim s^\beta$ with $\beta < 2$ would already be a breakthrough.
- In fact, such a breakthrough was achieved in [Bourgain, Dilworth, Ford, Konyagin, Kutzarova]: an explicit choice of $m = s^{2-\epsilon}$ rows from the discrete Fourier matrix was shown to yield a matrix with the RIP of order s, but $\epsilon > 0$ was ever so small that this theoretical feat is of no practical interest.

Summary

Non-uniform vs. uniform

In this section we present non-uniform results. Until now, we have only quantified the probability of recovering *any* s-sparse signal:

Uniform results

$$\mathbb{P}(\forall x \in \Sigma_s, \text{(BP) uniquely recovers } x) \geq 1 - \eta.$$

Now we are going to derive results such that

Non-uniform results

$$\forall x \in \Sigma_s, \quad \mathbb{P}(\text{(BP) uniquely recovers } x) \geq 1 - \eta.$$

Statement

Theorem (Non-uniform recovery for BOS)

Let $x \in \mathbb{K}^d$ be an s-sparse vector. Let A be the sampling matrix defined as in (1). If

$$m \geq c_0 \cdot \mu s \cdot \ln(d) \ln(\varepsilon^{-1}),$$

then x is the unique minimizer of (BP) with probability higher than $1 - \varepsilon$.

Remark

- *Note that the coherence is a criterion easier to verify for a sampling matrix than the RIP property.*
- *The price to pay is that the recovery success is guaranteed for a fixed vector x to be reconstructed.*

Starting point

Given an s-sparse signal $x \in \mathbb{R}^d$ such that $S = \mathrm{supp}(x)$ and $y = Ax$, we want to recover x using (BP).
Using Lagrangian methods or Fermat's rule, one should end with the following optimality conditions:

- x is a solution to (BP),
- if $\exists h \in \mathbb{R}^m$

$$0 \in \partial_{\|\cdot\|_1}(x) + A^* h, \tag{2}$$

$$y = Ax. \tag{3}$$

Optimality conditions for BP

Given the value of $\partial_{\|\cdot\|}$, one could reformulate as follows: x is a solution of (BP) if $\exists h \in \mathbb{R}^m$, such that

$$(A^*h)_S = \text{sign}(x_S), \tag{4}$$

$$|(A^*h)_i| \leq 1, \quad \text{for} \quad i \in S^c. \tag{5}$$

In addition, if A is injective on the set $\mathcal{S}$ of all vectors supported on S and we can find an h such that the second condition is

$$|(A^*h)_i| < 1, \quad \text{for} \quad i \in S^c,$$

then x is the unique solution.

The vector $v = A^*h$ is called a *dual certificate*.

Exercise

Show the following statement:

Let $x \in \mathbb{C}^d$ be an s-sparse signal of support S. If there exists $v = A^* h$ such that

$$v_S = \operatorname{sign}(x_S)$$
$$\|v_{S^c}\|_\infty < 1,$$

then x is the unique solution of

$$\min \|z\|_1 \qquad s.t. \quad (y =) Ax = Az.$$

Hint: Use that $|\langle z, v\rangle| < \|v\|_1 \|z\|_\infty$ if v is non-zero and the entries of z have not the same moduli.

Constructing a dual certificate

Let A_S be the matrix extracted from A containing the rows in S. Recovering x via (BP) is equivalent to being able to construct a dual certificate $v(= A^*h)$. A possible approach is to study an ansatz, which is the solution v to:

$$\min \|v\|_2 \qquad \text{s.t.} \quad v \in \mathrm{Im}(A^*) \quad \text{and} \quad P_S v = \operatorname{sign}(x).$$

where P_S is the projection onto the linear span $\mathcal{S}$ of vectors with the same support S as x.

The motivation for this ansatz is twofold:

1. it is known in closed form and can be expressed as

   $$v = A^* A_S (A_S^* A_S)^{-1} \operatorname{sign}(x_S).$$

   ($A_S^* A_S$ is invertible if and only if $\mathcal{S} \cap \ker(A) = \{0\}$).
2. it is the solution to a least-squares problem and that by minimizing its Euclidean norm we hope to make its dual norm small as well.

Using the random sensing matrix?

At this point it is important to recall the random sampling model in which the rows of A are *i.i.d.* samples from some distribution so that

$$A^*A = \frac{1}{m}\sum_{i=1}^{m} \tilde{a}_i \tilde{a}_i^* \qquad \rightsquigarrow \text{empirical covariance matrix}$$

with the $(\tilde{a}_i)$'s isotropic random vectors (such that $\tilde{a}_i = \frac{a_{J_i}}{\sqrt{\pi_i}}$). Under the isotropy assumption, $\mathbb{E}A^*A = \mathrm{Id}$, $\mathbb{E}A_S^*A_S = \mathrm{Id}_{\mathcal{S}}$.
Of course, A^*A cannot be close to the identity since it has rank $m \ll n$ but we can nevertheless ask whether its restriction to S is close to the identity on S.

(i) One asks

$$\frac{1}{2}\mathrm{Id}_{\mathcal{S}} \preceq A_S^*A_S \preceq \frac{3}{2}\mathrm{Id}_{\mathcal{S}}, \tag{6}$$

meaning that $A_S^*A_S$ should be close to its expectation.

Using the random sensing matrix?

The idea is to develop bounds on moments of the difference between the sampled covariance matrix and its expectation,

$$H_S = \mathrm{Id}_S - A_S^* A_S,$$

- controlling the growth of $\mathbb{E}(H_S^{2k})$ for large powers give control of $\|H_S\|_{2\to 2}$
- lack of independency: it is not possible to invoke standard moment calculation methods
- $\rightsquigarrow$ delicate combinatorial issues

Next?

(ii) To show that the ansatz is indeed a dual certificate, one can expand the inverse of $A_S^*A_S$ as a Neumann series

Neumann series for matrix

If $\|B\|_{2\to 2} < 1$ then $(I-B)^{-1}$ exists and

$$(I-B)^{-1} = \sum_{k\geq 0} B^k.$$

$$v = A^*A_S(A_S^*A_S)^{-1}\,\mathrm{sign}(x_S) = A^*A_S\left[\sum_{j\geq 0}(\mathrm{Id} - A_S^*A_S)^j\right]\mathrm{sign}(x_S)$$

$$= \sum_{j\geq 0} v_j, \qquad v_j := A^*A_S H_S^j\,\mathrm{sign}(x).$$

Problem: each term is not independent.

And then?

(iii) In the ℓ^1 problem, we would need to show that

$$\|v_{S^c}\|_\infty < 1,$$

In [1], this is achieved by a combinatorial method bounding the size of each term v_j by controlling an appropriately large moment $\mathbb{E}|v_j|^{2k}$. This strategy yields to $20s\log(d)$ measurements.

[1] Emmanuel J Candès, Justin Romberg, and Terence Tao.
Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information.
IEEE Transactions on information theory, 52(2):489–509, 2006.

$\rightsquigarrow$ There is an easier way to show that the restricted sampled covariance matrix is close to its mean.

Proposition (Matrix Bernstein [1])

Consider a finite sequence (M_k) of independent, random matrices with dimension $d_1 \times d_2$. Assume that

$$\mathbb{E} M_k = 0 \qquad \|M_k\|_{2\to 2} \leq R \quad a.s.$$

Introduce the random matrix and its variance parameter

$$Z = \sum_k M_k,$$

$$\sigma^2 = \max\left(\|\mathbb{E}(ZZ^*)\|_{2\to 2}, \|\mathbb{E}(Z^*Z)\|_{2\to 2}\right).$$

One has

$$\mathbb{P}\left(\|Z\|_{2\to 2} \geq t\right) \leq (d_1 + d_2) \exp\left(-\frac{t^2/2}{\sigma^2 + Rt/3}\right)$$

[1] Joel A Tropp.
User-friendly tail bounds for sums of random matrices.
Foundations of computational mathematics, 12(4):389–434, 2012.

- Matrix-valued analog of the classical Bernstein inequality for sums of independent random variables
- Gives tail bounds on the spectral norm of a sum of mean-zero independent random matrices.
- Both $\mathrm{Id} - A^*A$ and its restriction to $\mathcal{S}$ are of this form! $\rightsquigarrow$ Show it!
- One downside is that these general tools are unfortunately not as precise as combinatorial methods.

Inexact duality I

A bit later, David Gross provided an elegant construction of an *inexact dual certificate* he called the *golfing scheme*.
To begin with, it is not hard to see that if

$$\frac{1}{2}\mathrm{Id}_{\mathcal{S}} \preceq A_S^* A_S \preceq \frac{3}{2}\mathrm{Id}_{\mathcal{S}},$$

holds, then the existence of a vector $v \perp \ker(A)$ (i.e. $v \in \mathrm{Im}(A^*)$), obeying

$$\underbrace{\|v_S - \mathrm{sign}(x_S)\|_2 \leq \delta}_{\text{relaxation}}, \qquad \underbrace{\|v_{S^c}\|_\infty \leq 1/2}_{\textit{restriction}} \tag{7}$$

for some small δ, certifies that x is the unique solution.

Inexact duality II

So we relax the condition $v_S = \text{sign}(x_S)$ so that it only holds approximately.
To see why this is true, take v as in (7) and consider the perturbation $v' = v - A^* A_S (A_S^* A_S)^{-1}(\text{sign}(x) - v)$. Then $v' \perp \ker(A)$, $v'_S = \text{sign}(x_S)$ and

$$\|v'_{S^c}\|_\infty \leq 1/4 + \|A^* A_T (A_S^* A_S)^{-1}(\text{sign}(x) - v)\|_\infty.$$

Because the columns of A have Euclidean norm at most μ, then the "nonuniform RIP", together with Cauchy-Schwarz give that the second term in the right-hand side is bounded by $\delta\sqrt{2}\mu$, which is less than $1/2$ if δ is sufficiently small.

Inexact duality

Proposition

Let $x \in \mathbb{R}^d$ with support S of size s. Let $A \in \mathbb{K}^{m\times d}$ be the sensing matrix whose the columns are denoted $(a_{:\ell})_{1\leq \ell \leq d}$. Assume that

$$\|(A_S^* A_S)^{-1}\|_{2\to 2} \leq 2 \qquad \max_{\ell \in S^c} \|A_S^* a_{:\ell}\|_1 \leq 1, \tag{8}$$

and that there exists a vector $v = A^* w \in \mathbb{K}^d$ with $w \in \mathbb{C}^m$ such that

$$\|v_S - \operatorname{sign}(x_S)\|_2 \leq 1/4 \qquad \|v_{S^c}\|_\infty \leq 1/4. \tag{ID}$$

Then x is the unique ℓ^1-minimizer such that $y = Ax$.

Proof of inexact duality

Exercise

We're going to show the previous proposition. To do so, let us suppose that there exists $\hat{x}$ another solution to the ℓ^1-min problem.

1. *Show that for some $h \in \ker(A)$,*

$$\|\hat{x}\|_1 \geq \|x\|_1 - |\langle \operatorname{sign}(x_S), h_S\rangle| + \|h_{S^c}\|_1.$$

2. *Using Conditions* (ID)*, show that*

$$|\langle \operatorname{sign} x_S, h_S\rangle| \leq \frac{1}{4}(\|h_S\|_2 + \|h_{S^c}\|_1).$$

3. *By noticing that* $\|h_S\|_2 = \left\|(A_S^* A_S)^{-1}(A_S^* A_S)\, h_S\right\|_2$*, show that*

$$\|h_S\|_2 \leq 2\,\|h_{S^c}\|_1\,.$$

4. *Finish the proof.*

The golfing scheme I

The golfing scheme is an iterative process to construct an inexact dual certificate.

Now partition A into row blocks so that from now on, $A^{(1)}$ are the first m_1 rows of the matrix A, $A^{(2)}$ the next m_2 rows, and so on. The L matrices $(A^{(\ell)})_{1 \le \ell \le L}$ are independently distributed, and we have $m_1 + m_2 + \ldots + m_L = m$. The golfing scheme then starts with $v_0 = 0$, inductively defines

Iterative process

$$v_\ell = \frac{m}{m_\ell}(A^{(\ell)})^* A^{(\ell)} P_S(\operatorname{sign}(x_S) - v_{\ell-1}) + v_{\ell-1},$$

for $\ell = 1, \ldots, L$, and set $v = v_L$.

Clearly, v is in the row space of A, and thus perpendicular to the null space.

The golfing scheme II

To understand the scheme,

Iterative process

$$v_\ell = \frac{m}{m_\ell}(A^{(\ell)})^* A^{(\ell)} P_S(\operatorname{sign}(x_S) - v_{\ell-1}) + v_{\ell-1},$$

for $\ell = 1, \ldots, L$, and set $v = v_L$.

we can examine the first step

$$v_1 = \frac{m}{m_1}(A^{(1)})^* A^{(1)} P_S \operatorname{sign}(x_S),$$

and observe that it is perfect on the average since

$$\mathbb{E} v_1 = \operatorname{sign}(x_S).$$

The golfing scheme III

With finite sampling, we will not find ourselves at $\text{sign}(x)$ and, therefore, the next step should approximate $P_S(\text{sign}(x) - v_1)$, and read

$$v_2 = v_1 + \frac{m}{m_2}(A^{(2)})^* A^{(2)} P_S(\text{sign}(x_S) - v_1).$$

Set $q_\ell = P_S(\text{sign}(x) - v_\ell)$ to be the "error" of duality, and observe the recurrence relation

$$q_\ell = \left(\mathrm{Id}_{\mathcal{S}} - \frac{m}{m_\ell}(A^{(\ell)})_S)^* A_S^{(\ell)} \right) q_{\ell-1}.$$

If the block sizes are large enough so that

$$\left\| \mathrm{Id}_{\mathcal{S}} - \frac{m}{m_\ell}(A^{(\ell)})_S)^* A_S^{(\ell)} \right\|_{2\to 2} \leq 1/2,$$

(this is again the property that the empirical covariance matrix does not deviate too much from the identity), then we see that the size of the error decays exponentially to zero since it is at least halved at each iteration:

$$\|q_\ell\|_2 \leq \frac{1}{2}\|q_{\ell-1}\|_2.$$

Remark

Actually we do not require that

$$\left\| \mathrm{Id}_{\mathcal{S}} - \frac{m}{m_\ell}(A^{(\ell)})_S)^* A_S^{(\ell)} \right\|_{2\to 2} \leq 1/2,$$

we only require that <u>*for a fixed vector z*</u>*,*

$$\left\| (\mathrm{Id}_{\mathcal{S}} - \frac{m}{m_\ell}(A^{(\ell)})_S)^* A_S^{(\ell)})z \right\|_2 \leq \frac{1}{2}\|z\|_2,$$

with high probability. Since H_ℓ and $q_{\ell-1}$ are independent, this fact allows for smaller block sizes.

Now we examine the size of v on S^c, that is, outside of the support of x, and compute

$$v = \sum_{\ell=1}^{L} \frac{m}{m_\ell} (A^{(\ell)})^* A^{(\ell)} q_{\ell-1}.$$

The key point is that by construction, $(A^{(\ell)})^* A^{(\ell)}$ and $q_{\ell-1}$ are stochastically independent. In a nutshell, conditioned on $q_{\ell-1}$, $(A^{(\ell)})^* A^{(\ell)} q_\ell$ is just a random sum of the form $\sum_k a_k \langle a_k, q_{\ell-1} \rangle$ and one can use standard large deviation inequalities to bound the size of each term as follows:

$$\left\| \frac{m}{m_\ell} (A_S^{(\ell)})^* A^{(\ell)} q_{\ell-1} \right\|_\infty \leq t_\ell \|q_{\ell-1}\|_2,$$

for some scalar $t_\ell > 0$, with inequality holding with high probability.

An inexact dual certificate for recovery

The previous construction of an inexact dual certificate can be used leading to exact recovery by the following lemma.

Lemma

Let $x \in \mathbb{K}^d$ be an s-sparse vector supported on S. If

$$\|A_S^* A_S - \mathrm{Id}\|_{2\to 2} \leq 1/2 \qquad \max_{\ell \in S^c} \|A_S^* a[\ell]\|_2 \leq 2$$

where $a[\ell]$ is the ℓ-th column of A, and if there exists $v = A^* h$ such that

$$\|v_S - \mathrm{sign}(x_S)\|_2 \leq 1/4 \qquad \|v_{S^c}\|_\infty \leq 1/4$$

then x is the unique solution of (BP).

Such a general strategy along with many other estimates and ideas that we cannot possibly detail in this course note. Gross' method is very general and useful, although it is generally not as precise as the combinatorial approach.

Summary

Stability and robustness for non-uniform approaches

Once an (exact or inexact) dual certificate has been constructed, stability and robustness result can be easily derived for non-uniform approaches. This is the purpose of the following result.

Theorem

Let $x \in \mathbb{K}^d$ be an s-sparse vector. Let A be the sampling matrix defined as in (1). Let $x^\sharp$ be a solution of

$$\min_z \|z\|_1 \qquad s.t. \qquad \|Az - y\|_2 \leq \eta.$$

If $m \geq c_0 \mu s \ln(d) \ln(\varepsilon^{-1})$, then the reconstruction error satisfies with probability higher than $1 - \varepsilon$,

$$\|x^\sharp - x\|_2 \leq c_1 \sigma_s(x)_1 + c_2 \sqrt{s}\eta,$$

for c_0, c_1 and c_2 some universal constants.

Stability and robustness for non-uniform approaches

The previous theorem can be shown by constructing an (inexact) dual certificate as in the following lemma.

Lemma

If

$$\|A_S^* A_S - \mathrm{Id}\|_{2\to 2} \le 1/2 \qquad \max_{\ell \in S^c} \|A_S^* a[\ell]\|_2 \le 1$$

where $a[\ell]$ *is the* ℓ*-th column of* A*, and if there exists* $v = A^* h$ *such that*

$$\|v_S - \mathrm{sign}(x_S)\|_2 \le 1/4 \qquad \|v_{S^c}\|_\infty \le 1/4 \qquad \|h\|_2 \le \tau\sqrt{s}$$

for some constant τ*, then a minimizer* $x^\sharp$ *of*

$$\min_z \|z\|_1 \qquad s.t. \qquad \|Az - y\|_2 \le \eta$$

satisfies

$$\|x^\sharp - x\|_2 \le c_1 \sigma_s(x)_1 + c_2\sqrt{s}\eta,$$

Stability and robustness for uniform approaches

Theorem

Suppose that

$$m \geq c \cdot \mu s \ln^4(d)$$

for a universal constant $c > 0$. *With probability* $1 - n^{-\ln^3(d)}$, *every* $x \in \mathbb{K}^d$ *is approximately recovered from inaccurate samples* $y = Ax + \omega$ *where* $\|\omega\|_2 \leq \eta$, *using the quadratically-constrained basis-pursuit*

$$\min_z \|z\|_1 \qquad s.t. \qquad \|Az - y\|_2 \leq \eta,$$

in the sense that

$$\|x^\sharp - x\|_2 \leq c_1 \frac{\sigma_s(x)_1}{\sqrt{s}} + c_2 \eta,$$

for c_1 *and* c_2 *some universal constants.*

Non-uniform vs. uniform

One can note that uniform approaches outperform non-uniform approaches by a $\sqrt{s}$-factor in the error bound.

Here is a better bound for robustness in a non-uniform setting.

Theorem

Let $x \in \mathbb{R}^d$, let A be an $m \times d$ measurement matrix, and let $y = Ax + \omega$ be a vector of measurements in $\mathbb{R}^m$ assuming that $\|\omega\|_2 \leq \eta$. Let $x^\sharp$ be a solution of

$$\min_z \|z\|_1 \qquad s.t. \qquad \|Az - y\|_2 \leq \eta,$$

Then,

$$\|x^\sharp - x\|_2 \leq \frac{2\eta}{\sigma_{\min}\left(A, \mathcal{D}(\|\cdot\|_1, x^\sharp)\right)},$$

where $\sigma_{\min}\left(A, \mathcal{D}(\|\cdot\|_1, x^\sharp)\right)$ is the minimal singular value of A restricted to the descent cone $\mathcal{D}(\|\cdot\|_1, x^\sharp)$ of the ℓ^1-norm at $x^\sharp$, i.e.

$$\sigma_{\min}\left(A, \mathcal{D}\right) := \inf\left\{\|Au\|_2, u \in \mathcal{D} \cap \mathbb{S}^{n-1}\right\}.$$

But...

Again, $\sigma_{\min}\left(A, \mathcal{D}(\|\cdot\|_1, x^\sharp)\right)$ can be controlled in the case of Gaussian measurements by using Gaussian width. But this is still a challenge to control such a quantity for structured measurements.

So far only Gaussian measurements bridge the gap of robustness between uniform and non-uniform approaches.

Notebook