Notes

on

Compressed Sensing

for

Math 394

Simon Foucart

Spring 2009

Foreword

I have started to put these notes together in December 2008. They are intended for a graduate course on Compressed Sensing in the Department of Mathematics at Vanderbilt University. I will update them during the Spring semester of 2009 to produce the draft of a coherent manuscript by May 2009. I expect to continue improving it afterwards. You should keep in mind that, because Compressed Sensing is a young and rapidly evolving field, these notes may become quickly obsolete. Besides, because Compressed Sensing is at the intersection of many scientific disciplines, several lines of approach for a course on the subject can be taken. The perspective adopted here is definitely a mathematical one, since I am a mathematician by training and by taste. This bias has obviously influenced the selection of topics I chose to cover, and so has my exposition to the area during the past few years. Other than reading research articles, this consisted of a Shanks workshop given by Joel Tropp at Vanderbilt University, a workshop on the ℓ_1-norm held at Texas A&M University, and a series of talks given by Rick Chartrand, Ron DeVore, Anna Gilbert, Justin Romberg, Roman Vershynin, and Mike Wakin at Vanderbilt University.

For a more exhaustive view on Compressed Sensing, the following online resources are recommended:

- IMA short course on *Compressive Sampling and Frontiers in Signal Processing*:
  http://www.ima.umn.edu/2006-2007/ND6.4-15.07/abstracts.html
- Rice Compressed Sensing web site:
  http://www.compressedsensing.com/
- Igor Carron's blog, aka Nuit Blanche:
  http://nuit-blanche.blogspot.com/

I hope that you will enjoy reading these notes. By the end of the course, you will know almost as much as I do on the subject, and you should be able to — enthusiastically — conduct research in the area. Any corrections and suggestions are welcome. E-mail me at simon.foucart@centraliens.net.

Nashville, TN

Contents

Overview

...

Chapter 1
Motivations and Applications

In Biomedical Imaging, for instance in Magnetic Resonance Imaging, it is not conceivable to collect a number of measurements equal to the number of unknown pixels. Likewise, in wideband radio frequency analysis, limitations in Analog-to-Digital converter technology prevents the acquisition of a full signal based on Nyquist–Shannon paradigm — see Section 1.1. However, there is a special feature of images/signals that one can exploit to reconstruct them from such incomplete sets of information: they are compressible, in the sense that they essentially depend on a number of degrees of freedom much smaller than the complete information level. Modern transform coders such as JPEG2000 already rely on the fact that images have a sparse — well, almost — representation in a fixed basis. It is however common to acquire a signal entirely, to compute the complete set of transform coefficients, to encode the largest ones, and finally to discard all the others. This process is quite wasteful. A digital camera, for instance, uses millions of sensors — the pixels — to finally encode a picture on a few kilobytes. Compressed Sensing offers a way to acquire just about what is needed, by sampling and compressing simultaneously and by providing efficient decompression algorithms. The ideas of Compressed Sensing are now used on the hardware side to produce new sensing devices, in particular the one-pixel camera is much talked about. They are also used in statistical estimation, in studies of cholesterol level and gene expression, to name but a few, and will probably interface with other fields soon.

1.1 New Sampling Paradigm

A traditional paradigm in Magnetic Imaging Resonance, Astrophysics, and other fields of science consists of retrieving a compactly supported function by measuring some of its frequency coefficients. This is based on the following theorem — up to duality between time and frequency.

Theorem 1.1 (Nyquist–Shannon)**.** If the Fourier transform $\hat{f}$ of a function f satisfies $\mathrm{supp}(\hat{f}) \subseteq [-\Omega, \Omega]$, then f is completely determined by its sample values at the points $n\pi/\Omega$, $n \in \mathbb{Z}$, via

$$f(t) = \sum_{n=-\infty}^{\infty} f\left(n\,\frac{\pi}{\Omega}\right) \mathrm{sinc}\left(\Omega\left(t - n\,\frac{\pi}{\Omega}\right)\right).$$

Now suppose that $\mathrm{supp}(\hat{f})$ is small, but centered away from the origin, so that Ω is large. The theorem becomes of little value in practice because the different time samples have to be very close. Compressed Sensing offers the following new paradigm to reconstruct f: sample at m random times $t_1, \ldots, t_m$, with m of the order of Ω, and reconstruct e.g. by minimizing the ℓ_1-norm of the Fourier coefficients subject to the sampling conditions. The aim of this course is to understand precisely why this works.

1.2 Sparsest Solutions of Underdetermined Linear Systems

In the setting just described, the signal f is expanded as

$$f = \sum_{j=1}^{N} x_j \psi_j,$$

for a certain orthonormal basis $(\psi_1, \ldots, \psi_N)$, and such a representation is sparse in the sense that the number s of nonzero x_i's is small compared to N. The signal is then sampled with a number m of linear measurements, again small compared to N, to obtain the values

$$y_i := \langle f, \varphi_i \rangle, \qquad i \in [1:m]. \tag{1.1}$$

In the previous setting, we took φ_i as the Dirac distribution δ_{t_i}. The big question is how to choose $\varphi_1, \ldots, \varphi_m$ so that the signal f can be reconstructed from the measurements $y_1, \ldots, y_m$. Note that knowledge of the signal f and knowledge of the coefficient vector $\mathbf{x}$ are equivalent. Thus, we will usually work with $\mathbf{x}$, abusively designating it the signal. The equalities (1.1) translate into the matricial equality $\mathbf{y} = A\mathbf{x}$, where $A_{i,j} = \langle \psi_j, \varphi_i \rangle$. We adopt once and for all the following formulation for our standard problem:

Can we find an $m \times N$ sensing matrix A with $m \ll N$ such that any s-sparse vector $\mathbf{x} \in \mathbb{R}^N$ can be recovered from the mere knowledge of the measurement vector $\mathbf{y} = A\mathbf{x} \in \mathbb{R}^m$?

Of course, since the recovery process only sees the measurements $\mathbf{y}$, two s-sparse vectors $\mathbf{x}$ and $\mathbf{x}'$ satisfying $A\mathbf{x} = A\mathbf{x}' = \mathbf{y}$ must be equal, so that any s-sparse vector $\mathbf{x}$ with $A\mathbf{x} = \mathbf{y}$

is necessarily the sparsest solution of the underdetermined linear system $A\mathbf{z} = \mathbf{y}$. For this reason, if $\|\mathbf{z}\|_0^0$ represents the number of nonzero components of the vector $\mathbf{z}$, we may also consider the related problem

$$(\mathrm{P}_0) \qquad \text{minimize } \|\mathbf{z}\|_0^0 \quad \text{subject to} \quad A\mathbf{z} = \mathbf{y}.$$

We can easily determine a necessary and sufficient condition on the matrix A and the sparsity level s for the standard problem to be solvable. Indeed, writing Σ_s for the set of s-sparse vectors in $\mathbb{R}^N$, and noticing that $\Sigma_s - \Sigma_s = \Sigma_{2s}$, the condition is

$$\forall\, \mathbf{x} \in \Sigma_s,\, \forall\, \mathbf{x}' \in \Sigma_s \setminus \{\mathbf{x}\},\ \ A\mathbf{x} \neq A\mathbf{x}', \qquad \text{that is} \qquad \forall\, \mathbf{u} \in \Sigma_{2s} \setminus \{0\},\ \ A\mathbf{u} \neq 0.$$

In other words, the necessary and sufficient condition is

$$\boxed{\Sigma_{2s} \cap \ker A = \{0\}}. \tag{1.2}$$

However, finding matrices A satisfying this condition is not the end of the story, mainly because we do not have a reconstruction procedure available at the moment. Several procedures will be introduced in these notes, most notably the ℓ_1-minimization recovery algorithm. To anticipate slightly, it consists of solving the optimization problem

$$(\mathrm{P}_1) \qquad \text{minimize } \|\mathbf{z}\|_1 \quad \text{subject to} \quad A\mathbf{z} = \mathbf{y}.$$

Here are a few MATLAB commands to illustrate the discussion. The ℓ_1-magic software, available online, is required.

```
>> N=512; m=128; s=25;
>> A=randn(m,N);
>> permN=randperm(N); supp=sort(permN(1:s));
>> x=zeros(N,1); x(supp)=rand(s,1);
>> y=A*x;
>> x1=A\y; xstar=l1eq_pd(x1,A,[],y,1e-3);
>> norm(x-xstar)
 ans =

   2.1218e-05
```

1.3 Error Correction

Suppose that we encode a plaintext $\mathbf{y} \in \mathbb{R}^m$ by a ciphertext $\mathbf{x} = B\mathbf{y} \in \mathbb{R}^N$, with $N > m$. We think of the $N \times m$ matrix B as representing a linear code. If B has full rank, then we can

decode $\mathbf{y}$ from $\mathbf{x}$ as

$$\mathbf{y} = [B^\top B]^{-1} B^\top \mathbf{x}. \tag{1.3}$$

Now suppose that the ciphertext is corrupted. Thus, we have knowledge of

$$\mathbf{x} = B\mathbf{y} + \mathbf{e} \qquad \text{for some error vector } \mathbf{e} \in \mathbb{R}^N.$$

It turns out that the plaintext $\mathbf{y}$ can still be exactly recovered if B is suitably chosen and if the number $s := |\{i \in [1:N] : e_i \neq 0\}|$ of corrupted entries is not too large. Indeed, take $N \geq 2m$ and consider an $m \times N$ matrix A satisfying (1.2) for $2s \leq m$. We then choose B as an $N \times m$ matrix that satisfies $AB = 0$ — pick the m columns of B as linearly independent vectors in $\ker A$, which is at least m-dimensional. Because we know the vector $A\mathbf{x} = A(B\mathbf{y} + \mathbf{e}) = A\mathbf{e}$, Condition (1.2) enable us to recover the error vector $\mathbf{e}$. Finally, the equality $B\mathbf{y} = \mathbf{x} - \mathbf{e}$ yields the decoding of the plaintext $\mathbf{y}$ as

$$\mathbf{y} = [B^\top B]^{-1} B^\top (\mathbf{x} - \mathbf{e}).$$

Exercises

Ex.1: Prove the identity (1.3).

Ex.2: Suppose that $\mathbf{x} \in \mathbb{R}^N$ is a piecewise constant vector with only a small number s of jumps, and suppose that we only know the measurement vector $\mathbf{y} = A\mathbf{x}$. It is possible to recover $\mathbf{x}$ by minimizing the ℓ_1-norm of a vector $\mathbf{z}'$ depending on a vector $\mathbf{z}$ subject to $A\mathbf{z} = \mathbf{y}$. What is this vector $\mathbf{z}'$?

Ex.3: Prove Nyquist–Shannon theorem

Ex.4: Observe that Condition (1.2) implies $m \geq 2s$.

Ex.5: Suppose that $2s > m$. For $2s \times m$ matrices B, show that plaintexts $\mathbf{y} \in \mathbb{R}^{\mathbf{m}}$ cannot be decoded from ciphertexts $\mathbf{x} = B\mathbf{y} + \mathbf{e}$ with s corrupted entries.

Ex.6: Imitate the MATLAB commands given in Section 1.2 to verify the error correction procedure described in Section 1.3.

Ex.7: Suppose that the sparsity is now measured by $\|\mathbf{z}\|_{0,w}^0 := \sum_{i:z_i \neq 0} w_i$ for some weights $w_1, \ldots, w_N > 0$. Find a necessary and sufficient condition in order for each $\mathbf{x} \in \mathbb{R}^N$ with $\|\mathbf{x}\|_{0,w}^0 \leq s$ to be the unique minimizer of $\|\mathbf{z}\|_{0,w}^0$ subject to $A\mathbf{z} = A\mathbf{x}$.

Chapter 2
Theoretical Limitations

As indicated in Chapter 1, one of the goals of Compressed Sensing is to recover 'high-dimensional' signals from the mere knowledge 'low-dimensional' measurements. To state such a problem in its full generality, we assume that the signals $\mathbf{x}$ live in a signal space X, and that they are sampled to produces measurements $\mathbf{y} = f(\mathbf{x})$ living in a measurement space Y. We call the map $f : X \to Y$ the measurement map — note that it can always be assumed to be surjective by reducing Y to $f(X)$. We wish to recover the signal $\mathbf{x} \in X$, i.e. we wish to find a reconstruction map $g : Y \to X$ such that $g(f(\mathbf{x})) = \mathbf{x}$. Typically, we take $X = \mathbb{R}^N$ and $Y = \mathbb{R}^m$ with $m < N$, or better $m \ll N$, and we choose f as a linear map. Since a linear map $f : \mathbb{R}^N \to \mathbb{R}^m$ cannot be injective, the reconstruction identity $g(f(\mathbf{x})) = \mathbf{x}$ cannot be valid for all signals $\mathbf{x} \in \mathbb{R}^N$. Instead, we impose the signals $\mathbf{x}$ to belong to a recoverable class Σ. Typically, the latter class is taken to be the set Σ_s of all s-sparse vectors, i.e. the set of all vectors with no more than s nonzero components. Note that Σ_s can be written as a union of linear subspaces, precisely $\Sigma_s = \cup_{|S| \le s} \Sigma_S$, where Σ_S is defined as $\Sigma_S := \{\mathbf{x} \in \mathbb{R}^N : \ x_i = 0, \, i \notin S\}$ for each index set $S \subseteq [1 : N]$.

2.1 Minimal Number of Measurements

Given a sparsity level s, we want to know how few measurements are necessary to recover s-sparse vectors. This depends on the signal and measurement spaces X and Y and on the possible restrictions imposed on the measurement and reconstruction maps f and g. In others words, if we translate these restrictions by $(f, g) \in R_{X,Y}$, we want to find

$$m_*(s; R_{X,Y}) := \inf \big\{ m \in \mathbb{N} : \text{ there exists } (f, g) \in R_{X,Y} \text{ such that } g(f(\mathbf{x})) = \mathbf{x}, \text{ all } \mathbf{x} \in \Sigma_s \big\}.$$

We shall distinguish several cases according to the underlying fields of the signal and measurement spaces. Namely, we take

$$X = \mathbb{F}^N, \quad Y = \mathbb{K}^m, \quad \text{with } \mathbb{F}, \mathbb{K} \in \big\{\mathbb{Q}, \mathbb{R}\big\}.$$

2.1.1 $\mathbb{F} = \mathbb{R}$, $\mathbb{K} = \mathbb{Q}$, no restriction on f and g

It can be argued that measurements are necessarily rational-valued. In this case, it is impossible to recover real-valued signals. We have

$$m_*(s; \mathbb{F} = \mathbb{R}, \mathbb{K} = \mathbb{Q}) = +\infty.$$

Indeed, if a suitable m existed, then the reconstruction identity would imply that the map $f_{|\Sigma_s} : \Sigma_s \to \mathbb{Q}^m$ is injective, and in turns that Σ_s is countable. But this is not so, since Σ_s contains all the real lines $\Sigma_{\{i\}}$, $i \in [1 : N]$.

We recall here that a set $\mathcal{S}$ is called countable if

(2.1) there exists an injection from $\mathcal{S}$ to $\mathbb{N}$,

or equivalently if

(2.2) there exists a surjection from $\mathbb{N}$ to $\mathcal{S}$.

2.1.2 $\mathbb{F} = \mathbb{Q}$, $\mathbb{K} = \mathbb{Q}$, no restriction on f and g

If both measurements and signals are rational-valued, then it is possible to recover any signal from one single measurement as long as we can freely select the measurement and reconstruction maps. In short,

$$m_*(s; \mathbb{F} = \mathbb{Q}, \mathbb{K} = \mathbb{Q}) = 1.$$

Indeed, because $\mathbb{Q}^N$ is countable, there exists a surjection $g : \mathbb{Q} \to \mathbb{Q}^N$. Thus, for all $\mathbf{x} \in \mathbb{Q}^N$, we can choose $\mathbf{y} =: f(\mathbf{x}) \in \mathbb{Q}$ such that $g(\mathbf{y}) = \mathbf{x}$. With such measurement and reconstruction maps f and g, we have $g(f(\mathbf{x})) = \mathbf{x}$ for all $\mathbf{x} \in \mathbb{Q}^N$, not just for all $\mathbf{x} \in \Sigma_s$.

2.1.3 $\mathbb{F} = \mathbb{R}$, $\mathbb{K} = \mathbb{R}$, no restriction on f and g

For real-valued signals and measurements, it also happens that one single measurement is enough to ensure recovery of all $\mathbf{x} \in \mathbb{Q}^N$, not just of all $\mathbf{x} \in \Sigma_s$, so long as we can freely

select the measurement and reconstruction maps. In short,

$$m_*(s; \mathbb{F} = \mathbb{R}, \mathbb{K} = \mathbb{R}) = 1.$$

Indeed, classical space-filling maps provide surjections $g : \mathbb{R} \to \mathbb{R}^N$, and we construct measurement maps $f : \mathbb{R}^N \to \mathbb{R}$ just as before.

2.1.4 $\mathbb{F} = \mathbb{R}$, $\mathbb{K} = \mathbb{R}$, f continuous and antipodal

For real-valued signals and measurements, we shall now impose the measurement map $f : \mathbb{R}^N \to \mathbb{R}^m$ to be continuous and to be antipodal, that is to satisfy $f(-\mathbf{x}) = -f(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^N$. For example, a linear map $\mathbb{R}^N \to \mathbb{R}^m$ meets these requirements. The minimal number of measurements necessary to recover sparse vectors is in this case twice the sparsity. In other words,

$$m_*(s; \mathbb{F} = \mathbb{R}, \mathbb{K} = \mathbb{R}, f \text{ continuous and antipodal}) \; = \; 2s.$$

For the first part of the proof — that is the inequality $m_* \geq 2s$ — we shall use Borsuk–Ulam theorem. A proof can be found in Appendix 2.

Theorem 2.1 (Borsuk–Ulam)**.** A continuous antipodal map $F : \mathbb{S}^n \to \mathbb{R}^n$ from the sphere $\mathbb{S}^n$ of $\mathbb{R}^{n+1}$ — relative to any norm — into $\mathbb{R}^n$ vanishes at least once, i.e. there is a point $\mathbf{x} \in \mathbb{S}^n$ for which $F(\mathbf{x}) = 0$.

Given $m < 2s$, let us now assume that it is possible to find a continuous antipodal map $f : \mathbb{R}^N \to \mathbb{R}^m$ which is injective on Σ_s. Then, taking $U := \Sigma_{[1:s]}$ and $V := \Sigma_{[s+1:2s]}$, we define the continuous antipodal map

$$F : (\mathbf{u}, \mathbf{v}) \in U \times V \mapsto f(\mathbf{u}) - f(\mathbf{v}) \in \mathbb{R}^m.$$

Since $\dim(U \times V) = 2s > m$, we can apply Borsuk–Ulam theorem to obtain $\mathbf{u} \in U$ and $\mathbf{v} \in V$ with $\|\mathbf{u}\|_1 + \|\mathbf{v}\|_1 = 1$ such that $F(\mathbf{u}, \mathbf{v}) = 0$. This means that $f(\mathbf{u}) = f(\mathbf{v})$. The injectivity of f on Σ_s then implies that $\mathbf{u} = \mathbf{v}$. But this yields $\mathbf{u} = \mathbf{v} = 0$, which contradicts $\|\mathbf{u}\|_1 + \|\mathbf{v}\|_1 = 1$. At this point, we have established the inequality

$$m_*(s; \mathbb{F} = \mathbb{R}, \mathbb{K} = \mathbb{R}, f \text{ continuous and antipodal}) \; \geq \; 2s.$$

The reverse inequality is established in the next section. As it turns out, we will consider linear measurement maps to recover s-sparse vectors from $2s$ measurements.

2.2 Totally Positive Matrices

Definition 2.2. A square matrix M is called totally positive, resp. totally nonnegative, if

$$\det(M_{I,J}) > 0, \qquad \text{resp.} \quad \det(M_{I,J}) \geq 0,$$

for all index sets I and J of same cardinality. Here $M_{I,J}$ represents the submatrix of M formed by keeping the rows indexed by I and the columns indexed by J.

Let us now suppose that $m = 2s$. We consider an $N \times N$ totally positive matrix M, from which we extract m rows indexed by a set I to form an $m \times N$ submatrix A. For each index set J of cardinality $m = 2s$, the submatrix $A_J := M_{I,J}$ is invertible. Therefore, for any nonzero $2s$-sparse vector $\mathbf{u} \in \mathbb{R}^N$, say with $\mathrm{supp}(\mathbf{u}) \subseteq J$, $|J| = 2s$, we have $A\mathbf{u} = A_J \mathbf{u}_J \neq 0$. This establishes Condition (1.2) that $\Sigma_{2s} \cap \ker A = \{0\}$. Thus, the linear — in particular, continuous and antipodal — measurement map defined by $f(\mathbf{x}) = A\mathbf{x}$ allows reconstruction of every s-sparse vector from $m = 2s$ measurements. This means that

$$m_*(s; \mathbb{F} = \mathbb{R}, \mathbb{K} = \mathbb{R}, f \text{ continuous and antipodal}) \ \leq \ 2s.$$

This completes our proof, so long as we can exhibit a totally positive matrix M. We take the classical example of a Vandermonde matrix.

Proposition 2.3. Given $x_n > \cdots > x_1 > x_0 > 0$, the Vandermonde matrix $V := \big[x_i^j\big]_{i,j=0}^n$ is totally positive.

Proof. We start by proving Descartes' rule of sign, namely for all $(a_0, \ldots, a_n) \in \mathbb{R}^{n+1} \setminus \{0\}$, one has

$$Z_{(0,\infty)}(p) \leq S^-(a_0, \ldots, a_n), \qquad p(x) := \sum_{k=0}^n a_k x^k,$$

where $Z_{(0,\infty)}(p)$ represents the number of zeros of the polynomial p in the interval $(0, \infty)$ and where $S^-(a_0, \ldots, a_n)$ represents the number $|\{i \in [1:n] : \ a_{i-1}a_i < 0\}|$ of strong sign changes for the sequence $(a_0, \ldots, a_n)$. We proceed by induction on $n \geq 0$. For $n = 0$, the required result is obvious — it reads $0 \leq 0$. Let us now assume that the required result holds up to an integer $n-1$, $n \geq 1$. We want to establish that, given $(a_0, \ldots, a_n) \in \mathbb{R}^{n+1} \setminus \{0\}$ and $p(x) := \sum_{k=0}^n a_k x^k$, we have $Z_{(0,\infty)}(p) \leq S^-(a_0, \ldots, a_n)$. Note that we may suppose $a_0 \neq 0$, otherwise the result would be clear from the induction hypothesis, in view of

$$S^-(a_0, \ldots, a_n) = S^-(a_1, \ldots, a_n) \geq Z_{(0,\infty)}\Big(\sum_{k=0}^{n-1} a_{k+1} x^k\Big) = Z_{(0,\infty)}\Big(\sum_{k=0}^{n-1} a_{k+1} x^{k+1}\Big) = Z_{(0,\infty)}(p).$$

Now let ℓ be the smallest index in $[1:n]$ such that $a_\ell \neq 0$ — if no such index exists, then the result is clear. Up to the change $p \leftrightarrow -p$, there are two cases to consider: $[a_0 > 0, a_\ell < 0]$ or $[a_0 > 0, a_\ell > 0]$.

1/ $[a_0 > 0, a_\ell < 0]$. Applying Rolle's theorem and the induction hypothesis, we obtain

$$S^-(a_0, \ldots, a_n) = S^-(a_1, \ldots, a_n) + 1 \geq Z_{(0,\infty)}(p') + 1 \geq Z_{(0,\infty)}(p),$$

which is the required result.

2/ $[a_0 > 0, a_\ell > 0]$. Let t be the smallest positive zero of p — again, if no such t exists, then the result is clear. Suppose that p' does not vanish on $(0,t)$. This implies that p' has a constant sign on $(0,t)$. Since $p'(x) = \sum_{k=\ell}^n a_k k x^{k-1}$, there holds $p'(x) > 0$ on a certain right neighborhood of 0. Thus we obtain $p'(x) > 0$ for all $x \in (0,t)$, and consequently $0 = p(t) > p(0) = a_0$, which is not the case. Therefore, there is a zero of p' in $(0,t)$. Taking into account the zeros of p' guaranteed by Rolle's theorem, and using the induction hypothesis, we obtain

$$S^-(a_0, \ldots, a_n) = S^-(a_1, \ldots, a_n) \geq Z_{(0,\infty)}(p') \geq Z_{(0,\infty)}(p),$$

which is the required result.

The inductive proof of Descartes' rule of sign is now complete. Next, we shall prove that, for all $0 < x_0 < x_1 < \cdots < x_n$, $1 \leq i_1 \leq \cdots \leq i_\ell \leq n$, $1 \leq j_1 \leq \cdots \leq j_\ell \leq n$, and $1 \leq \ell \leq n$, one has

$$\begin{vmatrix} x_{i_1}^{j_1} & \ldots & x_{i_1}^{j_{\ell-1}} & x_{i_1}^{j_\ell} \\ \vdots & \ldots & \vdots & \vdots \\ x_{i_{\ell-1}}^{j_1} & \ldots & x_{i_{\ell-1}}^{j_{\ell-1}} & x_{i_{\ell-1}}^{j_\ell} \\ x_{i_\ell}^{j_1} & \ldots & x_{i_\ell}^{j_{\ell-1}} & x_{i_\ell}^{j_\ell} \end{vmatrix} > 0.$$

We proceed by induction on $\ell \in [1:n]$. For $\ell = 1$, the required result is nothing else than the positivity of all the x_i's. Let us now assume that the required result holds up to an integer $\ell - 1$, $\ell \geq 2$. Suppose that the required result fails for ℓ, i.e. that

$$\begin{vmatrix} x_{i_1}^{j_1} & \ldots & x_{i_1}^{j_{\ell-1}} & x_{i_1}^{j_\ell} \\ \vdots & \ldots & \vdots & \vdots \\ x_{i_{\ell-1}}^{j_1} & \ldots & x_{i_{\ell-1}}^{j_{\ell-1}} & x_{i_{\ell-1}}^{j_\ell} \\ x_{i_\ell}^{j_1} & \ldots & x_{i_\ell}^{j_{\ell-1}} & x_{i_\ell}^{j_\ell} \end{vmatrix} \leq 0 \tag{2.3}$$

for some $0 < x_0 < x_1 < \cdots < x_n$, $1 \leq i_1 \leq \cdots \leq i_\ell \leq n$, $1 \leq j_1 \leq \cdots \leq j_\ell \leq n$, and $1 \leq \ell \leq n$.

Let us introduce the polynomial

$$p(x) := \begin{vmatrix} x_{i_1}^{j_1} & \dots & x_{i_1}^{j_{\ell-1}} & x_{i_1}^{j_\ell} \\ \vdots & \dots & \vdots & \vdots \\ x_{i_{\ell-1}}^{j_1} & \dots & x_{i_{\ell-1}}^{j_{\ell-1}} & x_{i_{\ell-1}}^{j_\ell} \\ x^{j_1} & \dots & x^{j_{\ell-1}} & x^{j_\ell} \end{vmatrix}.$$

Expanding with respect to the last row and invoking Descartes' rules of sign, we get $Z_{(0,\infty)}(p) \leq \ell - 1$. But the polynomial p vanishes at the positive points $x_{i_1}, \dots, x_{i_{\ell-1}}$, hence vanishes only at these points in $(0, \infty)$. In view of (2.3), we derive that $p(x) < 0$ for all $x > x_{i_{\ell-1}}$. But this is absurd because, using the induction hypothesis, we have

$$\lim_{x\to\infty} \frac{p(x)}{x^{j_\ell}} = \begin{vmatrix} x_{i_1}^{j_1} & \dots & x_{i_1}^{j_{\ell-1}} \\ \vdots & \dots & \vdots \\ x_{i_{\ell-1}}^{j_1} & \dots & x_{i_{\ell-1}}^{j_{\ell-1}} \end{vmatrix} > 0.$$

We deduce that the required result holds for ℓ. This concludes the inductive proof. □

There is an interesting characterization of totally nonnegative matrices that we mention here without justification. By a weighted planar network G of order n, we mean an acyclic planar directed graph where $2n$ boundary vertices are distinguished as n sources $s_1, \dots, s_n$ and n sinks $t_1, \dots, t_n$ and where each edge e is assigned a weight $w(e) > 0$. The path matrix W of the weighted planar network G is defined by

$$W_{i,j} := \sum_{p \text{ path from } s_i \text{ to } t_j} w(p) := \sum_{p \text{ path from } s_i \text{ to } t_j} \prod_{e \text{ edge in } p} w(e).$$

The next lemma provides a simple interpretation for the determinant of W.

Lemma 2.4 (Lindström)**.** The determinant of the path matrix W of a weighted planar network G equals the weighted number of families of nonintersecting paths from the sources to the sinks, i.e.

$$\det(W) = \sum_{p_1,\dots,p_n \text{ non intersecting paths, } p_i \text{ path } s_i \to t_i} w(p_1) \cdots w(p_n) \quad \geq 0.$$

One can verify from Figure 2.1 that the 3×3 Vandermonde determinant is given by

$$\det\left[x_i^j\right]_{i,j=1}^3 = (x_2 - x_1)(x_2 - x_0)(x_1 - x_0).$$

We can apply Lindström's lemma to any submatrix $G_{I,J}$ of the path matrix G. We would obtain the first part of the next theorem. It is quite interesting that the converse also holds.

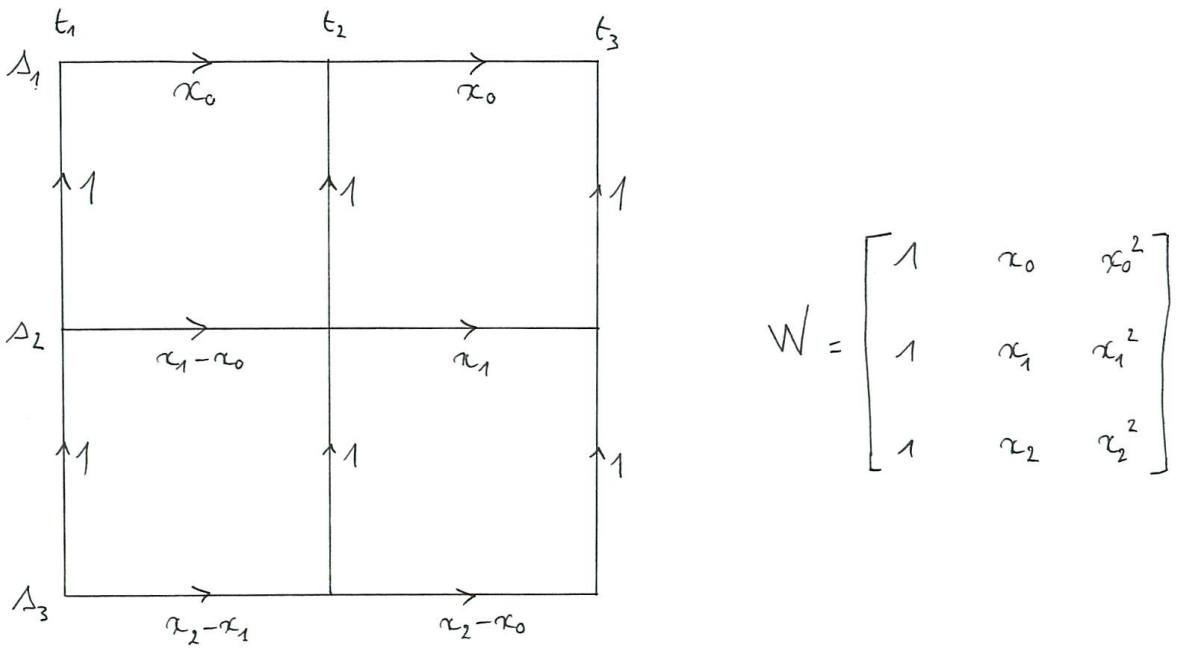


Figure 2.1: The weighted planar network for the 3×3 Vandermonde matrix

Theorem 2.5. The path matrix of any weighted planar network is totally nonnegative. Conversely, every totally nonnegative matrix is the path matrix of some weighted planar network.

Exercises

Ex.1: Check that (2.1) and (2.2) are indeed equivalent. Using (2.1) or (2.2), establish that $\mathbb{Q}^m$ is countable and that $\mathbb{R}$ is not countable.

Ex.2: Give an example of a plane filling-map $\mathbb{R} \to \mathbb{R}^2$.

Ex.3: Prove that $m_*(s; \mathbb{F} = \mathbb{R}, \mathbb{K} = \mathbb{R}, f \text{ continuous}) \geq s$. Can you do better?

Ex.4: Suppose that Borsuk–Ulam theorem holds relative to some particular norm. Show that it consequently holds relative to any norm.

Ex.5: Check that the stated formulation of Borsuk–Ulam theorem is equivalent to the following formulation:
Theorem. A continuous map $G : \mathbb{S}^n \to \mathbb{R}^n$ from the sphere $\mathbb{S}^n$ of $\mathbb{R}^{n+1}$ into $\mathbb{R}^n$ sends two antipodal points to the same point, i.e. there exists $\mathbf{x} \in \mathbb{S}^n$ for which $G(-\mathbf{x}) = G(\mathbf{x})$.

Ex.6: Prove that the product of two totally positive matrices is a totally positive matrix.

Ex.7: Let $0 < x_0 < \cdots < x_n < 1$. Use the total positivity of the Vandermonde matrix to establish that the collocation determinant

$$\det \left[B_i^n(x_j)\right]_{i,j=1}^n, \qquad B_i^n(x) := \binom{n}{i} x^i (1-x)^{n-i},$$

of the Bernstein polynomials $B_0^n, \ldots, B_n^n$ at the points $x_0, \ldots, x_n$ is totally positive.

Ex.8: Recall the necessary and sufficient condition for an $n \times n$ invertible matrix M to admit a Doolittle's factorization, i.e. an LU-factorization with ones on the diagonal of L. Observe that any totally positive matrix admits a Doolittle's factorization. Use Newton's form of the polynomial interpolant to exhibit the Doolittle's factorization of the transpose of the Vandermonde matrix.

Ex.9: Try to imitate the MATLAB commands of Section 1.2 when A is the matrix formed by m rows of an $N \times N$ Vandermonde matrix.

Chapter 3
Reed-Solomon Decoding

We have seen that, if the $m \times N$ matrix A is obtained from an $N \times N$ totally positive matrix by selecting m of its rows, then the measurement map defined by $f(\mathbf{x}) = A\mathbf{x}$, $\mathbf{x} \in \mathbb{R}^N$, allows to reconstruct every s-sparse vector with only $m = 2s$ measurements. In this case, the reconstruction map is given by

$$g(\mathbf{y}) \in \operatorname{argmin}\big\{ \|\mathbf{z}\|_0^0 : \ A\mathbf{z} = \mathbf{y} \big\}.$$

To find $g(\mathbf{y})$ in a straightforward way, it is required to perform a combinatorial search where all $\binom{N}{s}$ overdetermined linear systems $A_S \mathbf{z}_S = \mathbf{y}$, $|S| = s$, have to be solved. This is not feasible in practice. In this chapter, we shall introduce a practical reconstruction procedure that seems to do the job with only $m = 2s$ measurements. This procedure, however, has important faults that we shall expose.

3.1 The reconstruction procedure

Let $\mathbf{x}$ be an s-sparse vector. In fact, we consider $\mathbf{x}$ as a function x defined on $[0 : N-1]$ with $\operatorname{supp}(x) \subseteq S$, $|S| = s$. We shall measure only its first $2s$ discrete Fourier coefficients, namely

$$\hat{x}(j) := \sum_{k=0}^{N-1} x(k) e^{-i2\pi jk/N}, \qquad j \in [0 : 2s-1].$$

We then consider the trigonometric polynomial of degree s defined by

$$p(t) := \prod_{k \in S} (1 - e^{-i2\pi k/N} e^{i2\pi t/N}),$$

which vanishes exactly for $t \in S$. In view of $0 = p(t) \cdot x(t)$, $t \in [0, N-1]$, we obtain by discrete convolution

$$0 = (\hat{p} * \hat{x})(j) = \sum_{k=0}^{N-1} \hat{p}(k) \cdot \hat{x}(j-k), \qquad j \in [0 : N-1]. \tag{3.1}$$

We take into account that the coefficient $\hat{p}(k)$ of the trigonometric polynomial $p(t)$ on the monomial $e^{i2\pi kt/N}$ vanishes for $k > s$ and that its coefficient $\hat{p}(0)$ on the constant monomial 1 equals 1 to rewrite the equations of (3.1) corresponding to $j \in [s : 2s-1]$ as

$$\begin{array}{ccccccccc}
\hat{x}(s) & + & \hat{p}(1)\cdot\hat{x}(s-1) & + & \cdots & + & \hat{p}(s)\cdot\hat{x}(0) & = & 0, \\
\hat{x}(s+1) & + & \hat{p}(1)\cdot\hat{x}(s) & + & \cdots & + & \hat{p}(s)\cdot\hat{x}(1) & = & 0, \\
\vdots & & \vdots & & \cdots & & \vdots & & \vdots \\
\hat{x}(2s-1) & + & \hat{p}(1)\cdot\hat{x}(2s-2) & + & \cdots & + & \hat{p}(s)\cdot\hat{x}(s-1) & = & 0.
\end{array}$$

This translates into the Toeplitz system [1]

$$\begin{bmatrix}
\hat{x}(s-1) & \hat{x}(s-2) & \cdots & \hat{x}(0) \\
\hat{x}(s) & \hat{x}(s-1) & \cdots & \hat{x}(1) \\
\vdots & & \ddots & \vdots \\
\hat{x}(2s-2) & \hat{x}(2s-3) & \cdots & \hat{x}(s-1)
\end{bmatrix}
\begin{bmatrix} \hat{p}(1) \\ \hat{p}(2) \\ \vdots \\ \hat{p}(s) \end{bmatrix}
= -
\begin{bmatrix} \hat{x}(s) \\ \hat{x}(s+1) \\ \vdots \\ \hat{x}(2s-1) \end{bmatrix}.$$

Because $\hat{x}(0), \ldots, \hat{x}(2s-1)$ are known, we can solve for $\hat{p}(1), \ldots, \hat{p}(s)$. This determines $\hat{p}$ completely. In turns, the trigonometric polynomial p is completely determined by taking the inverse discrete Fourier transform. Then we *just need* to find the zeros of p to obtain the support of x. Once this is done, we can deduce x exactly by solving an overdetermined system of linear equations.

3.2 Implementation

We recall right away that the process of finding the roots of a polynomial — trigonometric or algebraic — is highly unstable. Therefore, instead of solving $p(t) = 0$ to find the support of x, we will simply select the indices j yielding the s smallest values for $|p(j)|$. Here are the MATLAB commands to test the Reed-Solomon decoding we have just described.

```
>> N=500; s=18;
>> supp=sort(randsample(N,s));
```

[1] the Toeplitz matrix is not always invertible: take e.g. $x = [1, 0, \ldots, 0]^\top$, so that $\hat{x} = [1, 1, \ldots, 1]^\top$,

```
>> x=zeros(N,1); x(supp)=randn(s,1);
>> xhat=fft(x); y=xhat(1:2*s);
>> phat=zeros(N,1); phat(1)=1;
>> A=toeplitz(y(s:2*s-1),y(s:-1:1));
>> phat(2:s+1)=-A\y(s+1:2*s);
>> p=ifft(phat);
>> [sorted_p,ind]=sort(abs(p)); rec_supp=sort(ind(1:s));
>> [supp';rec_supp']
 ans =
   17  43  45  48  73  90  91 141 154 253 255 307 321 344 439 456 486 492
   17  43  45  48  73  90  91 141 154 253 255 307 321 344 439 456 486 492
```

3.3 Non-robustness

In practice, we cannot measure the discrete Fourier coefficients with infinite precision, so our $2s$ measurement are in fact a slight perturbation of $\hat{x}(0), \ldots, \hat{x}(2s-1)$. It turns out that this small inaccuracy causes the procedure to badly misbehave. The following numerical experiment illustrates this point.

```
>> N=500; s=18;
>> supp=sort(randsample(N,s));
>> x=zeros(N,1); x(supp)=randn(s,1);
>> xhat=fft(x); noise=randn(2*s,1)/10^4; y=xhat(1:2*s)+noise;
>> phat=zeros(N,1); phat(1)=1;
>> A=toeplitz(y(s:2*s-1),y(s:-1:1));
>> phat(2:s+1)=-A\y(s+1:2*s);
>> p=ifft(phat);
>> [sorted_p,ind]=sort(abs(p)); rec_supp=sort(ind(1:s));
>> [supp';rec_supp']
 ans =
   8  23  91 167 177 212 214 220 248 266 284 338 354 410 424 433 489 491
   8   9  23  91 167 177 212 248 266 284 338 354 410 424 433 487 488 489
```

Exercises

Ex.1: A circulant matrix is a particular Toeplitz matrix of the form

$$\begin{bmatrix} c_0 & c_1 & c_2 & \cdots & c_{N-1} \\ c_{N-1} & c_0 & c_1 & \cdots & c_{N-2} \\ c_{N-2} & c_{N-1} & c_0 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & c_1 \\ c_1 & c_2 & \cdots & c_{N-1} & c_0 \end{bmatrix}.$$

Under which conditions is a circulant matrix invertible? Calculate the determinant of such a matrix using its eigenstructure.

Ex.2: Find the determinant of the Toeplitz matrix

$$\begin{bmatrix} a & b & b & \cdots & b \\ c & a & b & \cdots & b \\ c & c & a & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & b \\ c & c & \cdots & c & a \end{bmatrix}.$$

Ex.3: Prove that the discrete Fourier transform converts product into discrete convolution, and vice versa.

Ex.4: This classical example illustrates the instability of root finding. We consider the Wilkinson polynomial $p(x) = (x-1)(x-2)\cdots(x-20)$. Alter this polynomial slightly to form $p(x)+10^{-8}x^{19}$, and investigate numerically what happens to the largest roots.

Ex.5: This exercise illustrates the non-stability of Reed–Solomon decoding. Assume that $\mathbf{x}$ is not an s-sparse vector, but is close to one. Apply the Reed–Solomon procedure and determine if the result is close to the original vector $\mathbf{x}$.

Chapter 4
ℓ_q-Strategy: Null-Space Property

The sparsity $\|\mathbf{z}\|_0^0$ of a given vector $\mathbf{z} \in \mathbb{R}^N$ can be approximated by the q-th power of its ℓ_q-quasinorm when $q > 0$ is small. The observation that

$$\|\mathbf{z}\|_q^q := \sum_{j=1}^N |z_j|^q \underset{q \to 0}{\longrightarrow} \sum_{j=1}^N \mathbf{1}_{\{z_j \neq 0\}} = \|\mathbf{z}\|_0^0, \quad \mathbf{z} \in \mathbb{R}^N,$$

is the premise of this chapter. It suggests substituting the problem (P_0) by the problem

$$(\mathrm{P}_q) \qquad \text{minimize } \|\mathbf{z}\|_q^q \quad \text{subject to} \quad A\mathbf{z} = \mathbf{y}.$$

4.1 Null-Space Properties

We need to highlight the following fact as a prerequisite to our analysis.

Lemma 4.1. For $0 < q \leq 1$, the q-th power of the ℓ_q-quasinorm induces a metric on $\mathbb{R}^N$ defined by $d(\mathbf{u}, \mathbf{v}) := \|\mathbf{u} - \mathbf{v}\|_q^q$ for $\mathbf{u}, \mathbf{v} \in \mathbb{R}^N$

Proof. That $d(\mathbf{u}, \mathbf{v}) = d(\mathbf{v}, \mathbf{u})$ and that $[d(\mathbf{u}, \mathbf{v}) = 0] \iff [\mathbf{u} = \mathbf{v}]$ are clear. To establish the triangle inequality $d(\mathbf{u}, \mathbf{w}) \leq d(\mathbf{u}, \mathbf{v}) + d(\mathbf{v}, \mathbf{w})$ for $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}$, it is enough to show that $\|\mathbf{a} + \mathbf{b}\|_q^q \leq \|\mathbf{a}\|_q^q + \|\mathbf{b}\|_q^q$ for $\mathbf{a}, \mathbf{b} \in \mathbb{R}^N$. Working component by component, it suffices to prove that $(\alpha + \beta)^q \leq \alpha^q + \beta^q$ whenever $\alpha, \beta \geq 0$. If $\alpha = 0$, then this is obvious. If otherwise $\alpha > 0$, then we need to show that $(1 + \gamma)^q \leq 1 + \gamma^q$ for $\gamma := \beta/\alpha \geq 0$. Simply observe that the function h defined by $h(\gamma) := (1 + \gamma)^q - 1 - \gamma^q$ is negative on $(0, \infty)$, since $h(0) = 0$ and $h'(\gamma) = q[(1 + \gamma)^{q-1} - \gamma^{q-1}] < 0$ for $\gamma > 0$. □

Let us point out that the assumption $m \geq 2s$ is made throughout this chapter — and implicitly in the rest of these notes. Recall that it is in any case necessary for the exact

reconstruction of s-sparse vectors $\mathbf{x} \in \mathbb{R}^N$ from the measurements $\mathbf{y} = A\mathbf{x} \in \mathbb{R}^m$. Let us now suppose that, given any s-sparse vector $\mathbf{x} \in \mathbb{R}^N$, solving the problem ($\mathrm{P}_q$) with $\mathbf{y} = A\mathbf{x}$ returns the vector $\mathbf{x}$ as unique output. Then, for a vector $\mathbf{v} \in \ker A \setminus \{0\}$ and an index set $S \subseteq [1:N]$ with $|S| \le s$, we have $A(-\mathbf{v}_{\bar{S}}) = A\mathbf{v}_S$. Since the vector $\mathbf{v}_S$ is s-sparse and different from the vector $-\mathbf{v}_{\bar{S}}$, we must have $\|\mathbf{v}_S\|_q^q < \|\mathbf{v}_{\bar{S}}\|_q^q$. Conversely, let us suppose that $\|\mathbf{v}_S\|_q^q < \|\mathbf{v}_{\bar{S}}\|_q^q$ for all $\mathbf{v} \in \ker A \setminus \{0\}$ and all $S \subseteq [1:N]$ with $|S| \le s$. Then, given an s-sparse vector $\mathbf{x} \in \mathbb{R}^N$ and a different vector $\mathbf{z} \in \mathbb{R}^N$ satisfying $A\mathbf{z} = A\mathbf{x}$, we have $\mathbf{v} := \mathbf{x} - \mathbf{z} \in \ker A \setminus \{0\}$. Specifying $S = \mathrm{supp}(\mathbf{x})$, we get

$$\|\mathbf{x}\|_q^q \le \|\mathbf{x} - \mathbf{z}_S\|_q^q + \|\mathbf{z}_S\|_q^q = \|\mathbf{v}_S\|_q^q + \|\mathbf{z}_S\|_q^q < \|\mathbf{v}_{\bar{S}}\|_q^q + \|\mathbf{z}_S\|_q^q = \|-\mathbf{z}_{\bar{S}}\|_q^q + \|\mathbf{z}_S\|_q^q = \|\mathbf{z}\|_q^q.$$

Thus, the s-sparse vector $\mathbf{x}$ is the unique solution of (P_q) with $\mathbf{y} = A\mathbf{x}$. At this point, we have established a necessary and sufficient condition for exact recovery of all s-sparse vectors by ℓ_q-minimization. This condition on the matrix A and the sparsity s is called the Null-Space Property relative to ℓ_q. It reads

(NSP_q) $$\boxed{\forall\, \mathbf{v} \in \ker A \setminus \{0\}, \quad \forall\, |S| \le s, \qquad \|\mathbf{v}_S\|_q^q < \|\mathbf{v}_{\bar{S}}\|_q^q.}$$

By adding $\|\mathbf{v}_S\|_q^q$ to both sides of the inequality and rearranging the terms, we can also state the Null-Space Property in the form

$$\boxed{\forall\, \mathbf{v} \in \ker A \setminus \{0\}, \quad \forall\, |S| \le s, \qquad \|\mathbf{v}_S\|_q^q < \tfrac{1}{2}\|\mathbf{v}\|_q^q.}$$

Let us briefly mention two properties that a reasonable reconstruction scheme should posses: if we add some measurements, then the recovery should be preserved, and if we amplify some measurements, then the recovery should also be preserved. Mathematically speaking, this translates into replacing the $m \times N$ matrix A by an $m' \times N$ matrix $\widetilde{A} := \begin{bmatrix} A \\ \hline B \end{bmatrix}$ for an $(m'-m) \times N$ matrix B, or by an $m \times N$ matrix $\widehat{A} = DA$ for a nonsingular $m \times m$ diagonal matrix D. In these two cases, ℓ_q-recovery is preserved, because the corresponding Null-Space Properties remain fulfilled, in view of $\ker \widetilde{A} \subseteq \ker A$ and $\ker \widehat{A} = \ker A$.

4.2 Reconstruction exponents

It is natural to enquire about the success of ℓ_q-recovery as a function of the exponent q. The main result of this section is the justification of the intuitive belief that ℓ_q-recovery should imply ℓ_r-recovery for all $r < q$. Before establishing this, let us start with the simple observation that ℓ_q-recovery is impossible if $q > 1$.

Lemma 4.2. If $q > 1$, then for any $m \times N$ matrix A with $m < N$, there exists a 1-sparse vector which is not recovered by ℓ_q-minimization.

Proof. Let us consider an exponent $q > 1$. For $j \in [1 : N]$, let $\mathbf{e}_j \in \mathbb{R}^N$ be the 1-sparse vector whose j-th component equals one. Now suppose that, for all $\mathbf{z} \in \mathbb{R}^N$ satisfying $A\mathbf{z} = A\mathbf{e}_j$ and $\mathbf{z} \neq \mathbf{e}_j$, we have $\|\mathbf{z}\|_q^q > \|\mathbf{e}_j\|_q^q = 1$. Considering a vector $\mathbf{v} \in \ker A \setminus \{0\}$ and a real number $t \neq 0$ with $|t| < 1/\|\mathbf{v}\|_\infty$, we obtain

$$1 < \|\mathbf{e}_j + t\mathbf{v}\|_q^q = |1 + t\,v_j|^q + \sum_{k=1,\,k\neq j}^{N} |t\,v_k|^q = (1 + t\,v_j)^q + t^q \sum_{k=1,\,k\neq j}^{N} |v_k|^q \underset{t\to 0}{\sim} 1 + q\,t\,v_j.$$

For this to happen, we need $v_j = 0$. The fact that this should be true for all $j \in [1 : N]$ is clearly in contradiction with $\mathbf{v} \neq 0$. □

The next observation concerns the set $\mathcal{Q}_s(A)$ of reconstruction exponents, that is the set of all exponents $0 < q \leq 1$ for which every s-sparse vector $\mathbf{x} \in \mathbb{R}^N$ is recovered as the unique solution of (P_q) with $\mathbf{y} = A\mathbf{x}$. Although we will not use the following observation, let us notice that, according to the Null-Space Property relative to ℓ_q, the set of reconstruction exponents can be written as

$$\mathcal{Q}_s(A) := \big\{q \in (0,1] :\ \forall\, |S| \leq s,\ \|R_{A,S}\|_q < (1/2)^{1/q}\big\},$$

where for an index set $S \subseteq [1 : N]$, the notation $R_{A,S}$ denotes the restriction operator from $\ker A$ into $\mathbb{R}^N$ as defined by $R_{A,S}(\mathbf{v}) := \mathbf{v}_S$, $\mathbf{v} \in \ker A$, and where the expression $\|R_{A,S}\|_q := \sup\big\{\|R_{A,S}(\mathbf{v})\|_q,\ \|\mathbf{v}\|_q = 1\big\}$ represents the ℓ_q-quasinorm of the operator $R_{A,S}$

Proposition 4.3. The set $\mathcal{Q}_s(A)$ of reconstruction exponents is a — possibly empty — open interval $\big(0, q^*(A)\big)$. The right endpoint $q_s^*(A) \in [0,1]$ is called the critical reconstruction exponent of the matrix A with respect to the sparsity s.

Proof. Let us remark that to establish the Null-Space Property for a given $\mathbf{v} \in \ker A \setminus \{0\}$, it is enough to consider the index set S of the s largest absolute-value components of $\mathbf{v}$. Note then that the condition $\|\mathbf{v}_S\|_q^q < \|\mathbf{v}_{\bar{S}}\|_q^q$ can be rewritten as

$$\sum_{j\in S} \frac{|v_j|^q}{\sum_{k\in\bar{S}} |v_k|^q} \ < \ 1. \tag{4.1}$$

Now, given an index $j \in S$, the quantity

$$\frac{|v_j|^q}{\sum_{k\in\bar{S}} |v_k|^q} \ = \ \frac{1}{\sum_{k\in\bar{S}} \big(|v_k|/|v_j|\big)^q}$$

is an increasing function of $q \in (0,1]$, since $|v_k|/|v_j| \le 1$ for $k \in \bar{S}$. Thus, summing over $j \in S$, we see that the inequality (4.1) is fulfilled for all $r \in (0,q)$ as soon as it if fulfilled for a certain $q \in (0,1]$. This shows that $\mathcal{Q}_s(A)$ is an interval of the type $\big(0, q^*(A)\big)$ or of the type $\big(0, q^*(A)\big]$. Let us prove that it is of the former type. For this purpose, let us consider a sequence (q_n) of exponents in $\overline{\mathcal{Q}_s(A)}$ converging to $q := q_s^*(A) \in (0,1]$. For each integer n, there exists an index set $S_n \subseteq [1:N]$ with $|S_n| \le s$ and a vector $\mathbf{v}_n \in \ker A$ with $\|\mathbf{v}_n\|_{q_n}^{q_n} = 1$ and $\|\mathbf{v}_{n,S_n}\|_{q_n}^{q_n} \ge 1/2$. Note that we can extract a constant subsequence $(S_{n_k}) =: (S)$ out of the sequence (S_n), since there is only a finite number of subsets of cardinality s in $[1:N]$. Then, because the sequence $(\mathbf{v}_{n_k})$ has values in the unit ball of the finite-dimensional space $\ker A$ endowed with the ℓ_∞-norm, we can extract a subsequence that converges to some $\mathbf{v} \in \ker A$. The equality $\|\mathbf{v}_{n_k}\|_{q_{n_k}}^{q_{n_k}} = 1$ and the inequality $\|\mathbf{v}_{n_k,S}\|_{q_{n_k}}^{q_{n_k}} \ge 1/2$ pass to the limit to give $\|\mathbf{v}\|_q^q = 1$ and $\|\mathbf{v}_S\|_q^q \ge 1/2$. Thus, the Null-Space Property relative to ℓ_q is not satisfied. This proves that $q = q_s^*(A) \in \overline{\mathcal{Q}_s(A)}$, as required. □

4.3 Reconstruction and sign pattern of sparse vectors

Although we were varying the index set S in the previous sections, most of our analysis remains valid if the index set S is fixed. For such a context, we present in this section a result reminiscent of the Null-Space Property. It has an interesting corollary, which roughly states that ℓ_1-minimization succeeds in recovering vectors supported exactly on S only according to their sign patterns. But beware, recovering $\mathbf{x}$ by ℓ_1-minimization means here that $\mathbf{x}$ is *a* minimizer of $\|\mathbf{z}\|_1$ subject to $A\mathbf{z} = A\mathbf{x}$, not necessarily *the unique* minimizer.

Proposition 4.4. Given an index set $S \subseteq [1:N]$, and given a vector $\mathbf{x} \in \mathbb{R}^N$ whose support is exactly S, one has

$$[\forall\, \mathbf{z} \in \mathbb{R}^N \text{ with } A\mathbf{z} = A\mathbf{x},\ \|\mathbf{z}\|_1 \ge \|\mathbf{x}\|_1] \iff [\forall\, \mathbf{v} \in \ker A,\ \sum_{j \in S} \operatorname{sgn}(x_j) v_j \le \|\mathbf{v}_{\bar{S}}\|_1].$$

Proof. The left-hand side condition is equivalent to

$$\forall\, \mathbf{v} \in \ker A,\ \|\mathbf{x} - \mathbf{v}\|_1 \ge \|\mathbf{x}\|_1.$$

The desired result will follow from a characterization of best approximation in ℓ_1-norm. We refer to the Appendix for a characterization of best approximation in a general normed space. Using this result and the fact that

$$\operatorname{Ex}(B_{\ell_1^N}^*) \cong \operatorname{Ex}(B_{\ell_\infty^N}) = \big\{\varepsilon \in \mathbb{R}^N : \forall\, j \in [1:N], \varepsilon_j = \pm 1\big\},$$

the characterization takes the form

$$\forall\, \mathbf{v} \in \ker A,\ \exists\, \varepsilon_1, \ldots, \varepsilon_N = \pm 1 :\ \sum_{j=1}^N \varepsilon_j (x_j - v_j) \geq \sum_{j=1}^N \varepsilon_j x_j = \sum_{j=1}^N |x_j|.$$

The latter equality implies that $\varepsilon_j = \operatorname{sgn}(x_j)$ on $\operatorname{supp}(\mathbf{x}) = S$. Simplifying, we obtain the equivalent condition

$$\forall\, \mathbf{v} \in \ker A,\ \exists\, (\eta_j)_{j \in \bar{S}} \in \{-1, 1\}^{\bar{S}} :\ \sum_{j \in S} \operatorname{sgn}(x_j) v_j \leq \sum_{j \in \bar{S}} \eta_j v_j.$$

Finally, this turns out to be equivalent to the condition

$$\forall\, \mathbf{v} \in \ker A,\ \sum_{j \in S} \operatorname{sgn}(x_j) v_j \leq \|\mathbf{v}_{\bar{S}}\|_1,$$

as expected. □

4.4 Mixed-Norm Null-Space Properties

Typically, it is difficult to work directly with the ℓ_q-quasinorm, hence it is often preferable to obtain first estimates for the Euclidean ℓ_2-norm, and then to derive estimates for the ℓ_q-quasinorm. This step involves the following classical inequalities.

Lemma 4.5. Given $0 < q < p \leq \infty$, there holds

$$\|\mathbf{x}\|_p \ \leq\ \|\mathbf{x}\|_q \ \leq\ n^{1/q - 1/p} \|\mathbf{x}\|_p, \qquad \mathbf{x} \in \mathbb{R}^n,$$

and these inequalities are sharp.

Proof. Consider first of all the vectors $\mathbf{x} = [1, 0, \ldots, 0]^\top$ and $\mathbf{x} = [1, 1, \ldots, 1]^\top$ to observe that the constants 1 and $n^{1/q-1/p}$ in the inequalities $\|\mathbf{x}\|_p \leq 1 \cdot \|\mathbf{x}\|_q$ and $\|\mathbf{x}\|_q \leq n^{1/q-1/p} \cdot \|\mathbf{x}\|_p$ cannot be improved. Next, to establish the first inequality, we observe that it is sufficient to prove it for $\|\mathbf{x}\|_q = 1$, by homogeneity. In this case, we have $|x_i| \leq 1$ for all $i \in [1 : n]$, so that $|x_i|^p \leq |x_i|^q$. Summing over all i's yields $\|\mathbf{x}\|_p^p \leq 1$, i.e. $\|\mathbf{x}\|_p \leq 1$, which is what we wanted. Finally, to establish the second inequality, we simply use Hölder's inequality to write

$$\|\mathbf{x}\|_q^q = \sum_{i=1}^n 1 \cdot |x_i|^q \leq \Big[\sum_{i=1}^n 1^r\Big]^{1/r} \cdot \Big[\sum_{i=1}^n \big(|x_i|^q\big)^{p/q}\Big]^{q/p} = n^{1/r} \cdot \|\mathbf{x}\|_p^q,$$

where r satisfies $1/r + q/p = 1$, i.e. $1/r = 1 - q/p$. Taking the q-th root, we obtain the required inequality. □

Given $0 < q \leq 1$ and $p \geq q$, the previous lemma implies that $\|\mathbf{v}_S\|_q \leq s^{1/q-1/p} \cdot \|\mathbf{v}_{\bar{S}}\|_p$ for all s-sparse vectors $\mathbf{v} \in \mathbb{R}^N$. Therefore, the following conditions, which we call ℓ_p-Strong Null-Space Properties relative to ℓ_q, are enough to guarantee exact reconstruction of s-sparse vectors $\mathbf{x} \in \mathbb{R}^N$ from the measurements $\mathbf{y} = A\mathbf{x} \in \mathbb{R}^m$ by ℓ_q-minimization. They read

$$\text{(NSP}p,q\text{)} \qquad \forall\, \mathbf{v} \in \ker A, \quad \forall\, |S| \leq s, \qquad \|\mathbf{v}_S\|_q < \begin{cases} \dfrac{1}{s^{1/q-1/p}} \, \|\mathbf{v}_{\bar{S}}\|_q, \\ \\ \dfrac{(1/2)^{1/q}}{s^{1/q-1/p}} \, \|\mathbf{v}\|_q. \end{cases}$$

Exercises

Ex.1: A quasinorm $/\!/\!/ \cdot /\!/\!/$ on a vector space X is a function from X into $[0,\infty)$ for which there exists a constant $c > 0$ such that $/\!/\!/ \lambda \mathbf{x} /\!/\!/ = |\lambda| \, /\!/\!/ \, x /\!/\!/$, $/\!/\!/ \mathbf{x} + \mathbf{y} /\!/\!/ \leq c\big(\, /\!/\!/ \, \mathbf{x} \, /\!/\!/ + /\!/\!/ \, \mathbf{y} \, /\!/\!/ \, \big)$, and $[/\!/\!/ \mathbf{x} /\!/\!/ = 0] \Rightarrow [\mathbf{x} = 0]$. Prove that the expressions $\|\mathbf{x}\|_q = \sum_{i=1}^n |x_i|^q$, $\mathbf{x} \in \mathbb{R}^n$, and $\|T\|_q = \sup\big\{\|T\mathbf{x}\|_q, \, \|\mathbf{x}\|_q = 1\big\}$, $T \in \mathbb{R}^{n \times n}$, indeed define quasinorms. Note that a quasinorm is different from a seminorm $/\!/\!/ \cdot /\!/\!/ : X \to [0,\infty)$, also called pseudonorm, which satisfies $/\!/\!/ \lambda \mathbf{x} /\!/\!/ = |\lambda| /\!/\!/ x /\!/\!/$ and $/\!/\!/ \mathbf{x} + \mathbf{y} /\!/\!/ \leq /\!/\!/ \mathbf{x} /\!/\!/ + /\!/\!/ \mathbf{y} /\!/\!/$, but not $[/\!/\!/ \mathbf{x} /\!/\!/ = 0] \Rightarrow [\mathbf{x} = 0]$.

Ex.2: Make sure that the Null-Space Property implies the condition $\Sigma_{2s} \cap \ker A = \{0\}$.

Ex.3: Fixing a sparsity s, find a matrix whose critical reconstruction exponent is equal to a prescribed $q \in (0,1]$.

Ex.4: Consider the strengthened Null-Space Property

$$\exists\, c \in (0,1] : \quad \forall\, \mathbf{v} \in \ker A, \quad \forall\, |S| \leq s, \quad \|\mathbf{v}_S\|_q^q \leq \|\mathbf{v}_{\bar{S}}\|_q^q - c\|\mathbf{v}\|_q^q$$

and the strengthened Minimality Property

$$\exists\, C \in (0,1] : \quad \forall\, s\text{-sparse } \mathbf{x} \in \mathbb{R}^N, \quad \forall\, \mathbf{z} \in \mathbb{R}^N \text{ with } A\mathbf{z} = A\mathbf{x}, \quad \|\mathbf{x}\|_q^q \leq \|\mathbf{z}\|_q^q - C\|\mathbf{x}-\mathbf{z}\|_q^q.$$

Prove that the equivalence of the two properties. What is the relation between the constants c and C?

Ex.5: What is the critical reconstruction exponent $q_1^*(A)$ of the matrix

$$A = \begin{bmatrix} 2 & 1 & -2 & 1 \\ 1 & 1 & -1 & 1 \end{bmatrix}.$$

Chapter 5
ℓ_q-Strategy: Stability, Robustness

In the previous chapter, we have suggested recovering sparse vectors by ℓ_q-minimization, thus using the reconstruction map $g : \mathbb{R}^m \to \mathbb{R}^N$ defined by

$$g(\mathbf{y}) \in \operatorname{argmin}\big\{ \|\mathbf{z}\|_q^q : \ A\mathbf{z} = \mathbf{y} \big\}.$$

A nice feature of this reconstruction scheme is that it is nonadaptive, i.e. the measurement procedure is independent of the signal $\mathbf{x}$ we wish to acquire, that is, the measurements $\mathbf{y} = A\mathbf{x}$ are collected using one and only one sensing matrix A. The ℓ_q-reconstruction scheme possesses other favorable properties presented in this chapter, mainly stability and robustness. The continuity property presented next has less interest, since we will aim at obtaining $h = \mathrm{id}_{\Sigma_s}$, which is of course continuous.

5.1 Continuity

The ℓ_q-scheme turns out to be continuous as soon as it can be defined unambiguously. Precisely, we state the following proposition.

Proposition 5.1. Suppose that, for a fixed $q \in (0, 1]$, the minimizers of $\|\mathbf{z}\|_q^q$ subject to $A\mathbf{z} = A\mathbf{x}$ are unique whenever the vectors $\mathbf{x}$ belong to a certain subset Σ of $\mathbb{R}^N$. Then the map h defined by

$$h(\mathbf{x}) := \operatorname{argmin}\big\{ \|\mathbf{z}\|_q^q : \ A\mathbf{z} = A\mathbf{x} \big\}, \qquad \mathbf{x} \in \Sigma,$$

is a continuous map on Σ.

Proof. Given a sequence $(\mathbf{x}_n)$ in Σ converging to some $\mathbf{x} \in \Sigma$, we need to prove that the sequence $h(\mathbf{x}_n)$ converges to $h(\mathbf{x})$. To start with, let us consider a subsequence $(h(\mathbf{x}_{n_k}))$ of the sequence $(h(\mathbf{x}_n))$ converging to a given limit $\mathbf{x}' \in \mathbb{R}^N$. Remark that the equalities

$Ah(\mathbf{x}_{n_k}) = A\mathbf{x}_{n_k}$ pass to the limit as $k \to \infty$ to give $A\mathbf{x}' = A\mathbf{x}$. Let us now consider $\mathbf{z} \in \mathbb{R}^N$ satisfying $A\mathbf{z} = A\mathbf{x}$. Observe that

$$\mathbf{z}_{n_k} := \mathbf{x}_{n_k} - \mathbf{x} + \mathbf{z} \quad \text{satisfies} \quad A\mathbf{z}_{n_k} = A\mathbf{x}_{n_k},$$

so that

$$\|h(\mathbf{x}_{n_k})\|_q^q \le \|\mathbf{z}_{n_k}\|_q^q.$$

Taking the limit as $k \to \infty$, we derive

$$\|\mathbf{x}'\|_q^q \le \|\mathbf{z}\|_q^q.$$

Thus, the vector $\mathbf{x}' \in \mathbb{R}^N$ is a minimizer of $\|\mathbf{z}\|_q^q$ subject to $A\mathbf{z} = A\mathbf{x}$. By assumption, the vector $h(\mathbf{x}) \in \mathbb{R}^N$ is the unique such minimizer. We deduce that $\mathbf{x}' = h(\mathbf{x})$. At this point, we have established that any convergent subsequence of the sequence $(h(\mathbf{x}_n))$ actually converges to $h(\mathbf{x})$. Let us now assume by contradiction that the whole sequence $(h(\mathbf{x}_n))$ does not converge to $h(\mathbf{x})$. Then there exists a number $\varepsilon > 0$ and a subsequence $(h(\mathbf{x}_{n_k}))$ of the sequence $(h(\mathbf{x}_n))$ such that

$$\|h(\mathbf{x}_{n_k}) - h(\mathbf{x})\|_\infty \ge \varepsilon. \tag{5.1}$$

Observe that the sequence $(h(\mathbf{x}_{n_k}))$ is bounded form above by some constant $C > 0$, since $\|h(\mathbf{x}_{n_k})\|_\infty \le \|h(\mathbf{x}_{n_k})\|_q \le \|\mathbf{x}_{n_k}\|_q$, the latter being bounded because convergent. Hence, the sequence $(h(\mathbf{x}_{n_k}))$ has values in the compact set

$$\big\{\mathbf{z} \in \mathbb{R}^N : \ \|\mathbf{z}\|_\infty \le C, \ \|\mathbf{z} - h(\mathbf{x})\|_\infty \ge \varepsilon\big\}.$$

We can therefore extract from the sequence $(h(\mathbf{x}_{n_k}))$ a subsequence that converges to some $\mathbf{x}' \in \mathbb{R}^N$. Our previous argument implies that $\mathbf{x}' = h(\mathbf{x})$, but (5.1) yields $\|\mathbf{x}' - h(\mathbf{x})\|_\infty \ge \varepsilon$. This contradiction shows that the sequence $(h(\mathbf{x}_n))$ does converge to $h(\mathbf{x})$, as expected. □

5.2 Stability

In practice, the signals to be recovered are almost sparse, but not exactly. We should ask our reconstruction procedure to perform well in this case, in the sense that the reconstruction error should be controlled by the distance to sparse signals. Precisely, given a vector $\mathbf{x} \in \mathbb{R}^N$, if

$$\sigma_s(\mathbf{x})_q := \inf\big\{\|\mathbf{x} - \mathbf{z}\|_q, \ \|\mathbf{z}\|_0^0 \le s\big\}$$

represents the ℓ_q-error of best approximation to $\mathbf{x}$ by s-sparse vectors, and if $\mathbf{x}^\star$ is the output of the reconstruction algorithm applied to $\mathbf{x}$, we wish to have an inequality

$$\|\mathbf{x} - \mathbf{x}^\star\|_q \le C\, \sigma_s(\mathbf{x})_q \qquad \text{for some constant } C > 0.$$

This property is called Instance Optimality of order s relative to ℓ_q (with constant C). As it happens, as soon as ℓ_q-minimization provides exact reconstruction of sparse vectors, it also provides Instance Optimality. Let us state the following proposition.

Proposition 5.2. Given $0 < q \le 1$, we assume that the Null-Space Property of order s relative to ℓ_q holds, i.e. that

$$\forall\, \mathbf{v} \in \ker A \setminus \{0\}, \quad \forall\, |S| \le s, \quad \|\mathbf{v}_S\|_q^q < \frac{1}{2}\|\mathbf{v}\|_q^q.$$

Then the Instance Optimality of order s relative to ℓ_q holds for the ℓ_q-reconstruction, i.e.

$$\forall\, \mathbf{x} \in \mathbb{R}^N, \quad \|\mathbf{x} - \mathbf{x}^\star\|_q \le C\, \sigma_s(\mathbf{x})_q, \qquad \text{where } \mathbf{x}^\star \in \operatorname{argmin}\{\|\mathbf{z}\|_q^q :\ A\mathbf{z} = A\mathbf{x}\}$$

The constant C depends on s, q, and $\ker A$.

Proof. For each index set S of cardinality s, we have $\|\mathbf{v}_S\|_q^q < 1/2$ whenever $\mathbf{v} \in \ker A \cap S_q^N$, where S_q^N is the unit sphere of $\mathbb{R}^N$ relative to the ℓ_q-quasinorm. Since there are only finitely many such index sets and since $\ker A \cap S_q^N$ is compact, we have

$$c \ := \ 2 \sup_{|S| \le s} \ \sup_{\mathbf{v} \in \ker A \cap S_q^N} \|\mathbf{v}_S\|_q^q \ < \ 1.$$

Note that the constant c depends on s, q, and $\ker A$. For $|S| \le s$ and $\mathbf{v} \in \ker A$, the inequality $\|\mathbf{v}_S\|_q^q \le \frac{c}{2}\|\mathbf{v}\|_q^q$ can also be written as $\|\mathbf{v}_S\|_q^q \le \frac{1}{2}\|\mathbf{v}\|_q^q - \frac{1-c}{2}\|\mathbf{v}\|_q^q$. Subtracting $\frac{1}{2}\|\mathbf{v}_S\|_q^q$, this is also equivalent to

$$\|\mathbf{v}_S\|_q^q \le \|\mathbf{v}_{\bar S}\|_q^q - (1-c)\|\mathbf{v}\|_q^q, \qquad \mathbf{v} \in \ker A, \quad |S| \le s. \tag{5.2}$$

Let us now consider $\mathbf{x} \in \mathbb{R}^N$. We specify $\mathbf{v} \in \ker A$ to be $\mathbf{x} - \mathbf{x}^\star$ and S to be an index set of s largest absolute-value components of $\mathbf{x}$. Note that the inequality

$$\|\mathbf{x}\|_q^q \ge \|\mathbf{x}^\star\|_q^q$$

implies that

$$\|\mathbf{x}_S\|_q^q + \|\mathbf{x}_{\bar S}\|_q^q \ge \|(\mathbf{x}-\mathbf{v})_S\|_q^q + \|(\mathbf{x}-\mathbf{v})_{\bar S}\|_q^q \ge \|\mathbf{x}_S\|_q^q - \|\mathbf{v}_S\|_q^q + \|\mathbf{v}_{\bar S}\|_q^q - \|\mathbf{x}_{\bar S}\|_q^q.$$

Rearranging the latter, we obtain

$$\|\mathbf{v}_{\bar{S}}\|_q^q \le \|\mathbf{v}_S\|_q^q + 2\|\mathbf{x}_{\bar{S}}\|_q^q = \|\mathbf{v}_S\|_q^q + 2\sigma_s(\mathbf{x})_q^q. \tag{5.3}$$

Finally, we use (5.2) and (5.3) to deduce

$$\|\mathbf{v}\|_q^q = \|\mathbf{v}_S\|_q^q + \|\mathbf{v}_{\bar{S}}\|_q^q \le (\|\mathbf{v}_{\bar{S}}\|_q^q - (1-c)\|\mathbf{v}\|_q^q) + (\|\mathbf{v}_S\|_q^q + 2\sigma_s(\mathbf{x})_q^q) = c\|\mathbf{v}\|_q^q + 2\sigma_s(\mathbf{x})_q^q.$$

In other words, we have

$$\|\mathbf{x} - \mathbf{x}^\star\|_q^q \le \frac{2}{1-c}\sigma_s(\mathbf{x})_q^q, \qquad \text{all } \mathbf{x} \in \mathbb{R}^N.$$

which is the required result with $C := \left(\frac{2}{1-c}\right)^{1/q}$. □

Remark. There is a clear converse to the previous proposition. Indeed, for any s-sparse vector $\mathbf{x} \in \mathbb{R}^N$, we have $\sigma_s(\mathbf{x})_q = 0$, so that any minimizer $\mathbf{x}^\star$ of $\|\mathbf{z}\|_q^q$ subject to $A\mathbf{z} = A\mathbf{x}$ is the vector $\mathbf{x}$ itself. This exact reconstruction of s-sparse vectors by ℓ_q-minimization is known to be equivalent to the Null-Space Property of order s relative to ℓ_q.

5.3 Robustness

In practice, it is also impossible to measure a signal $\mathbf{x} \in \mathbb{R}^N$ with infinite precision. This means that the measurement vector $\mathbf{y} \in \mathbb{R}^m$ is only approximates the vector $A\mathbf{x} \in \mathbb{R}^m$ with an error bounded by some small constant $\varepsilon > 0$. Precisely, for a norm that need not be specified, we suppose that, for all $\mathbf{x} \in \mathbb{R}^N$, we have

$$\|\mathbf{y} - A\mathbf{x}\| \le \varepsilon.$$

We should ask our reconstruction procedure to perform well in this case, too, in the sense that the reconstruction error should be controlled by the measurement error. Therefore, if $\mathbf{x}$ is an s-sparse vector and $\mathbf{x}^\star$ is the output of the reconstruction algorithm applied to $\mathbf{x}$, we wish to have an inequality

$$\|\mathbf{x} - \mathbf{x}^\star\|_q \le D\,\varepsilon \qquad \text{for some constant } D > 0.$$

The constant D will not depend on the vector $\mathbf{x}$, but it will typically depend on the sparsity, for instance as $D \propto s^{1/q-1/2}$ if the Euclidian norm is chosen to evaluate the measurement error. The robustness inequality will be achieved when reconstructing via the following modified version of the ℓ_q-minimization:

$$(\mathrm{P}_{q,\varepsilon}) \qquad \text{minimize } \|\mathbf{z}\|_q^q \quad \text{subject to} \quad \|A\mathbf{z} - \mathbf{y}\| \le \varepsilon.$$

The sufficient condition is just a slight strengthening of the Null-Space Property.

Proposition 5.3. Given $0 < q \leq 1$, we assume that

$$\forall\, \mathbf{v} \in \mathbb{R}^N, \quad \forall\, |S| \leq s, \quad \|\mathbf{v}_S\|_q^q \leq c\, \|\mathbf{v}_{\bar{S}}\|_q^q + d\, \|A\mathbf{v}\|^q \tag{5.4}$$

for some norm $\|\cdot\|$ on $\mathbb{R}^m$ and some constants $0 < c < 1$, $d > 0$. Then, for every s-sparse vector $\mathbf{x} \in \mathbb{R}^N$ and any vector $\mathbf{y} \in \mathbb{R}^m$ satisfying $\|A\mathbf{x} - \mathbf{y}\| \leq \varepsilon$, defining $\mathbf{x}^\star$ as

$$\mathbf{x}^\star \in \operatorname{argmin}\big\{\|\mathbf{z}\|_q^q : \|A\mathbf{z} - \mathbf{y}\| \leq \varepsilon\big\},$$

there is a constant $D > 0$ depending only on c, d, and q, such that

$$\|\mathbf{x} - \mathbf{x}^\star\|_q \leq D\, \varepsilon.$$

Proof. We set $\mathbf{v} := \mathbf{x} - \mathbf{x}^\star \in \mathbb{R}^N$. Note first of all that

$$\|A\mathbf{v}\| = \|A\mathbf{x} - A\mathbf{x}^\star\| \leq \|A\mathbf{x} - \mathbf{y}\| + \|\mathbf{y} - A\mathbf{x}^\star\| \leq \varepsilon + \varepsilon = 2\varepsilon. \tag{5.5}$$

Let S be the support of the s-sparse vector $\mathbf{x}$. Form the minimality property of $\mathbf{x}^\star$, we derive

$$\|\mathbf{x}\|_q^q \geq \|\mathbf{x}^\star\|_q^q = \|(\mathbf{x} - \mathbf{v})_S\|_q^q + \|(\mathbf{x} - \mathbf{v})_{\bar{S}}\|_q^q \geq \|\mathbf{x}\|_q^q - \|\mathbf{v}_S\|_q^q + \|\mathbf{v}_{\bar{S}}\|_q^q.$$

Thus, we obtain

$$\|\mathbf{v}_{\bar{S}}\|_q^q \leq \|\mathbf{v}_S\|_q^q. \tag{5.6}$$

In conjunction with (5.4), this implies

$$\|\mathbf{v}_S\|_q^q \leq c\, \|\mathbf{v}_S\|_q^q + d\, \|A\mathbf{v}\|^q.$$

Rearranging the latter, and in view of (5.5), we deduce

$$\|\mathbf{v}_S\|_q^q \leq \frac{d}{1-c}(2\varepsilon)^q = \frac{2^q d}{1-c}\varepsilon^q.$$

Using (5.6) once more, we conclude that

$$\|\mathbf{v}\|_q^q = \|\mathbf{v}_S\|_q^q + \|\mathbf{v}_{\bar{S}}\|_q^q \leq 2\|\mathbf{v}_S\|_q^q \leq \frac{2^{q+1} d}{1-c}\varepsilon^q,$$

which is the required result with $D = \dfrac{2^{1+1/q} d^{1/q}}{(1-c)^{1/q}}$. □

Exercises

Ex.1: Let (X,d) be a metric space and let K be a compact subset of X. Suppose that every $x \in X$ has a unique best approximation from K, i.e. a unique $p_K(x) \in K$ such that $d(x,p_K(x)) \le d(x,k)$ for all $k \in K$. Prove that the best approximation map $x \in X \mapsto p_K(x) \in K$ is continuous.

Ex.2: Does the best approximation – assuming its uniqueness — to a vector $\mathbf{x}$ of $\mathbb{R}^N$ by s-sparse vectors of $\mathbb{R}^N$ depend on the ℓ_q-(quasi)norm when q runs in $(0,\infty]$?

Ex.3: Prove that the Null-Space Property with constant $\gamma < 1$ relative to ℓ_1 may be stated as

$$\forall\, \mathbf{v} \in \ker A, \quad \|\mathbf{v}\|_1 \le (1+\gamma)\sigma_s(\mathbf{v})_1.$$

We say that an $m \times N$ sensing matrix A exhibits Instance Optimality of order s with constant C relative to ℓ_1 if there exists a reconstruction map $g : \mathbb{R}^m \to \mathbb{R}^N$, not necessarily given by ℓ_1-minimization, such that

$$\forall\, \mathbf{x} \in \mathbb{R}^N, \quad \|\mathbf{x} - g(A\mathbf{x})\|_1 \le C\, \sigma_s(\mathbf{x})_1.$$

Prove that Instance Optimality of order s with constant C relative to ℓ_1 implies the Null-Space Property of order $2s$ with constant $C-1$ relative to ℓ_1, which itself implies Instance Optimality of order s with constant $2C$ relative to ℓ_1.

Ex.4: Suppose that an $m \times N$ sensing matrix A exhibits Instance Optimality of order s relative to ℓ_2. Prove that we necessarily have $m \ge c\,N$ for some constant $c > 0$.

Ex.5: This question aims at combining stability and robustness. Given $0 < q \le 1$, we assume that

$$\forall\, \mathbf{v} \in \mathbb{R}^N, \quad \forall\, |S| \le s, \quad \|\mathbf{v}_S\|_q^q \le c\, \|\mathbf{v}_{\bar{S}}\|_q^q + d\, \|A\mathbf{v}\|^q$$

for some norm $\|\cdot\|$ on $\mathbb{R}^m$ and some constants $0 < c < 1$, $d > 0$. Prove that, for every vector $\mathbf{x} \in \mathbb{R}^N$ and any vector $\mathbf{y} \in \mathbb{R}^m$ satisfying $\|A\mathbf{x} - \mathbf{y}\| \le \varepsilon$, defining $\mathbf{x}^\star$ as

$$\mathbf{x}^\star \in \operatorname{argmin}\big\{ \|\mathbf{z}\|_q^q : \|A\mathbf{z} - \mathbf{y}\| \le \varepsilon \big\},$$

one has

$$\|\mathbf{x} - \mathbf{x}^\star\|_q \;\le\; C\, \sigma_s(\mathbf{x})_q \;+\; D\, \varepsilon,$$

for some constant $C, D > 0$. What do these constants depend on?

Ex.6: Combine the inequalities $\|\mathbf{v}_{\bar{S}}\|_q^q \leq \|\mathbf{v}_S\|_q^q + 2\sigma_s(\mathbf{x})_q^q$ and $\|\mathbf{v}_S\|_q^q \leq c\|\mathbf{v}_{\bar{S}}\|_q^q$, $0 < c < 1$, in a simple way to obtain the bound $\|\mathbf{v}\|_q^q \leq \dfrac{2(1+c)}{(1-c)}\sigma_s(\mathbf{x})_q^q$, as stated in the middle column of next page table. Combine the inequalities in a more elaborate way to obtain the improved bound $\|\mathbf{v}\|_q^q \leq \dfrac{2}{(1-c)}\sigma_s(\mathbf{x})_q^q$, as was done in the proof of Proposition 5.2.

Pictorial Summary of Chapters 4 and 5

Exact Recovery $\mathbf{x}$ is s-sparse	**Stable Recovery** $\mathbf{x}$ is arbitrary	**Robust Recovery** $\mathbf{x}$ is s-sparse
Measure with a suitable sensing matrix A: (NSP$_q$) $\forall\, \mathbf{v} \in \ker A \setminus \{0\}, \quad \forall\, \lvert S\rvert \le s,$ $\lVert \mathbf{v}_S\rVert_q^q < \lVert \mathbf{v}_{\bar{S}}\rVert_q^q$	Measure with a suitable sensing matrix A: (NSP$_q(0<c<1)$) $\forall\, \mathbf{v} \in \ker A, \quad \forall\, \lvert S\rvert \le s,$ $\lVert \mathbf{v}_S\rVert_q^q \le c\lVert \mathbf{v}_{\bar{S}}\rVert_q^q$	Measure with a suitable sensing matrix A: (NSP'$_q(0<c<1, 0<d)$) $\forall\, \mathbf{v} \in \mathbb{R}^N, \quad \forall\, \lvert S\rvert \le s,$ $\lVert \mathbf{v}_S\rVert_q^q \le c\lVert \mathbf{v}_{\bar{S}}\rVert_q^q + d\lVert A\mathbf{v}\rVert^q$
Reconstruct by ℓ_q-minimization: $\mathbf{y} = A\mathbf{x},\ \mathbf{x}^\star \in \operatorname{argmin}\{\lVert \mathbf{z}\rVert_q^q : A\mathbf{z} = \mathbf{y}\}$ $\mathbf{v} = \mathbf{x} - \mathbf{x}^\star, \quad S = \operatorname{supp}(\mathbf{x})$ (MP$_q$) $\lVert \mathbf{v}_{\bar{S}}\rVert_q^q \le \lVert \mathbf{v}_S\rVert_q^q$	Reconstruct by ℓ_q-minimization: $\mathbf{y} = A\mathbf{x},\ \mathbf{x}^\star \in \operatorname{argmin}\{\lVert \mathbf{z}\rVert_q^q : A\mathbf{z} = \mathbf{y}\}$ $\mathbf{v} = \mathbf{x} - \mathbf{x}^\star, \quad S : s \text{ largest } \lvert x_j\rvert$ (MP'$_q$) $\lVert \mathbf{v}_{\bar{S}}\rVert_q^q \le \lVert \mathbf{v}_S\rVert_q^q + 2\sigma_s(\mathbf{x})_q^q$	Reconstruct by ℓ_q-minimization: $\lVert A\mathbf{x} - \mathbf{y}\rVert \le \varepsilon,\ \mathbf{x}^\star \in \operatorname{argmin}\{\lVert \mathbf{z}\rVert_q^q : \lVert A\mathbf{z} - \mathbf{y}\rVert \le \varepsilon\}$ $\mathbf{v} = \mathbf{x} - \mathbf{x}^\star, \quad S = \operatorname{supp}(\mathbf{x})$ (MP$_q$) $\lVert \mathbf{v}_{\bar{S}}\rVert_q^q \le \lVert \mathbf{v}_S\rVert_q^q$
(NSP$_q$) + (MP$_q$) $\Downarrow$ $\lVert \mathbf{x} - \mathbf{x}^\star\rVert_q^q = 0$	(NSP$_q(0<c<1)$) + (MP'$_q$) $\Downarrow$ $\lVert \mathbf{x} - \mathbf{x}^\star\rVert_q^q \le \dfrac{2(1+c)}{(1-c)}\sigma_s(\mathbf{x})_q^q$	(NSP'$_q(0<c<1, 0<d)$) + (MP$_q$) $\Downarrow$ $\lVert \mathbf{x} - \mathbf{x}^\star\rVert_q^q \le \dfrac{2^{q+1}d}{1-c}\varepsilon^q$

Chapter 6
A Primer on Convex Optimization

We have seen that it is possible to recover sparse vectors $\mathbf{x} \in \mathbb{R}^N$ by solving the problem

$$(\mathrm{P}_q) \qquad \text{minimize } \|\mathbf{z}\|_q^q \quad \text{subject to} \quad A\mathbf{z} = A\mathbf{x},$$

provided the sensing matrix A satisfies the Null-Space Property relative to ℓ_q. However, we did not touch upon the practicality of such minimization problems. In fact, for $q < 1$, the minimization problem is not a convex one. There is no truly reliable algorithm available in this case. The ideal situation occurs when the critical reconstruction exponent is as large as possible, i.e. when $q_s^*(A) = 1$, in which case the minimization problem (P_1) can be solved efficiently, since it is a convex optimization problem — in fact, it can be reformulated as a linear optimization problem.

6.1 Convex optimization

Let us start with the common terminology. The minimization problem

$$\text{minimize } F_0(\mathbf{z}) \quad \text{subject to } \begin{cases} F_1(\mathbf{z}) \le 0, & \dots, \quad F_k(\mathbf{z}) \le 0, \\ G_1(\mathbf{z}) = 0, & \dots, \quad G_\ell(\mathbf{z}) = 0, \end{cases} \tag{6.1}$$

is said to be a convex minimization problem if the objective, or cost, function $F_0 : \mathbb{R}^n \to \mathbb{R}$ and the inequality constraint functions $F_1, \dots, F_k : \mathbb{R}^n \to \mathbb{R}$ are convex functions, and if the equality constraints $G_1, \dots, G_\ell : \mathbb{R}^n \to \mathbb{R}$ are linear functions. Note that the problem may be unconstrained, which means that there are no inequality nor equality constraints. Of course, we should require the convexity of the domain $\mathcal{D}$ of the convex optimization problem (6.1), defined by

$$\mathcal{D} := \Big[\bigcap_{i=0}^{k} \mathrm{dom}(F_i)\Big] \bigcap \Big[\bigcap_{i=1}^{\ell} \mathrm{dom}(G_i)\Big].$$

It is readily seen that the feasible, or constraint, set

$$\mathcal{C} := \left\{ \mathbf{z} \in \mathcal{D} : \begin{array}{cccc} F_1(\mathbf{z}) \le 0, & \ldots, & F_k(\mathbf{z}) \le 0, \\ G_1(\mathbf{z}) = 0, & \ldots, & G_\ell(\mathbf{z}) = 0 \end{array} \right\}$$

is a convex subset of $\mathcal{D}$. It might be empty, in which case we say that the minimization problem is infeasible. Otherwise, we notice that the problem (6.1) translates into the minimization of a convex function over a nonempty convex set.

We can already notice that the minimization problems

$$\text{(P}_1\text{)} \qquad \text{minimize } \|\mathbf{z}\|_1 \quad \text{subject to} \quad A\mathbf{z} = \mathbf{y},$$

$$\text{(P}_{1,\varepsilon}\text{)} \qquad \text{minimize } \|\mathbf{z}\|_1 \quad \text{subject to} \quad \|A\mathbf{z} - \mathbf{y}\| \le \varepsilon,$$

are convex problems. Indeed, the objective function $\|\cdot\|_1$ is a convex function, there are no inequality constraints but only linear equality constraints in the first problem, and there are no equality constraints but only convex inequality constraints in the second problem.

The essential feature of convex optimization is that local minimizers are automatically global minimizers, as established in the next lemma. This means that algorithms designed to find local minimizers are reliable in this context.

Lemma 6.1. Given a convex set $\mathcal{C}$ and a convex function $F_0 : \mathcal{C} \to \mathbb{R}$, one has

$$[\forall\, \mathbf{z} \in \mathcal{C},\ F_0(\mathbf{z}^*) \le F_0(\mathbf{z})] \iff [\exists\, \varepsilon > 0 :\ \forall\, \mathbf{z} \in \mathcal{C} \text{ with } \|\mathbf{z} - \mathbf{z}^*\|_2 \le \varepsilon,\ F_0(\mathbf{z}^*) \le F_0(\mathbf{z})].$$

Proof. The direct implication is obvious. We now focus on the reverse implication. Let us consider $\mathbf{z} \in \mathcal{C}$. For $t \in (0,1)$, we define

$$\mathbf{z}' := (1-t)\mathbf{z}^* + t\mathbf{z} \in \mathcal{C}.$$

Because $\|\mathbf{z}' - \mathbf{z}^*\|_2 = t\|\mathbf{z} - \mathbf{z}^*\|_2 < \varepsilon$ as soon as $t < \varepsilon / \|\mathbf{z} - \mathbf{z}^*\|_2$, we have $F_0(\mathbf{z}^*) \le F_0(\mathbf{z}')$. By the convexity of F_0, this yields

$$F_0(\mathbf{z}^*) \le F_0((1-t)\mathbf{z}^* + t\mathbf{z}) \le (1-t)F_0(\mathbf{z}^*) + tF_0(\mathbf{z}).$$

This clearly implies $F_0(\mathbf{z}^*) \le F_0(\mathbf{z})$, as required. □

6.2 Linear optimization

A linear optimization problem is an optimization problem in which the objective function, the inequality constraint functions, and the equality constraint functions are all linear functions. In these favorable conditions, we have at our disposal some algorithms that perform even better than convex optimization algorithms.

It is crucial to realize that the problem (P_1) can be reformulated as a linear optimization problem by introducing N slack variables $t_1, \ldots, t_N$. Indeed, the problems

$$\text{(P_1)} \qquad \text{minimize } \sum_{j=1}^{N} |z_j| \quad \text{subject to} \quad A\mathbf{z} = \mathbf{y}$$

and

$$\text{(P'_1)} \qquad \text{minimize } \sum_{j=1}^{N} t_j \quad \text{subject to} \quad \mathbf{z} - \mathbf{t} \leq 0,\ -\mathbf{z} - \mathbf{t} \leq 0,\ A\mathbf{z} = \mathbf{y}$$

are equivalent. This means that if $\mathbf{z}^\star$ is a minimizer of (P_1), then $(\mathbf{z}^\star, |\mathbf{z}^\star|)$ is a minimizer of (P'_1), and conversely that if $(\mathbf{z}^\star, \mathbf{t}^\star)$ is a minimizer of (P'_1), then $\mathbf{z}^\star$ is a minimizer of (P_1).

We shall now make two classical observations in linear programming. The first one is at the basis of the so-called simplex method. It says that one can find a minimizer — or a maximizer — of a linear optimization problem among the vertices of the feasible set, which happens to be a convex polygon.

Proposition 6.2. For a compact and convex set $\mathcal{K}$ and a continuous and convex function $F_0 : \mathcal{K} \to \mathbb{R}$, one has

$$\sup_{\mathbf{z} \in \mathcal{K}} F_0(\mathbf{z}) = \max_{\mathbf{z} \in \mathrm{Ex}(\mathcal{K})} F_0(\mathbf{z}).$$

Proof. Because F_0 is continuous and $\mathcal{K}$ is compact, the supremum $M := \sup_{\mathbf{z} \in \mathcal{K}} F_0(\mathbf{z})$ is in fact a maximum. Thus, the set

$$\mathcal{E} := \big\{ \mathbf{z} \in \mathcal{K} : F_0(\mathbf{z}) = M \big\}$$

is nonempty. Observe that it is also compact, as a closed subset of a compact set. By Krein–Mil'man theorem, the set $\mathrm{Ex}(\mathcal{E})$ of extreme points of $\mathcal{E}$ is nonempty. Let us then pick $\mathbf{z} \in \mathrm{Ex}(\mathcal{E})$. The result will be established once we show that $\mathbf{z} \in \mathrm{Ex}(\mathcal{K})$. Suppose that this is not the case, i.e. that

$$\mathbf{z} = (1-t)\mathbf{z}_1 + t\mathbf{z}_2, \qquad \text{for some } \mathbf{z}_1, \mathbf{z}_2 \in \mathcal{K},\ \mathbf{z}_1 \neq \mathbf{z}_2, \text{ and } t \in (0,1).$$

Since $\mathbf{z} \in \mathcal{E}$ and since F_0 is convex, we get

$$M = F_0(\mathbf{z}) = F_0((1-t)\mathbf{z}_1 + t\mathbf{z}_2) \leq (1-t)F_0(\mathbf{z}_1) + tF_0(\mathbf{z}_2) \leq (1-t)M + tM = M.$$

Thus, equality must hold all the way through. In particular, we have $F_0(\mathbf{z}_1) = F_0(\mathbf{z}_2) = M$. This means that $\mathbf{z}_1, \mathbf{z}_2 \in \mathcal{E}$. Therefore, $\mathbf{z}$ appears as a strict convex combination of two distinct elements of $\mathcal{E}$, so that $\mathbf{z} \notin \mathrm{Ex}(\mathcal{E})$. This is the required contradiction. □

The second observation is the duality theorem of linear programming.

Theorem 6.3. Given an $n \times k$ matrix A and given vectors $\mathbf{b} \in \mathbb{R}^n$ and $\mathbf{c} \in \mathbb{R}^k$, the problems

$$\text{maximize } \mathbf{c}^\top \mathbf{x} \quad \text{subject to } A\mathbf{x} \leq \mathbf{b},\ \mathbf{x} \geq 0,$$

$$\text{minimize } \mathbf{b}^\top \mathbf{y} \quad \text{subject to } A^\top \mathbf{y} \geq \mathbf{c},\ \mathbf{y} \geq 0,$$

are dual, in the sense that if they are both feasible, then extremizers $\mathbf{x}^\star$ and $\mathbf{y}^\star$ exist, and that one has

$$\mathbf{c}^\top \mathbf{x}^\star = \mathbf{b}^\top \mathbf{y}^\star.$$

The same conclusion holds if only one of the extremizers $\mathbf{x}^\star$ or $\mathbf{y}^\star$ is known to exist.

Proof. For the first statement, let us first observe that

$$\begin{bmatrix} -A & 0 & \mathbf{b} \\ 0 & A^\top & -\mathbf{c} \\ I_k & 0 & 0 \\ 0 & I_n & 0 \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \\ 1 \end{bmatrix} \geq 0 \iff \mathbf{x} \text{ and } \mathbf{y} \text{ are feasible}$$

$$\implies \begin{bmatrix} -\mathbf{c} \\ \mathbf{b} \\ 0 \end{bmatrix}^\top \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \\ 1 \end{bmatrix} = -\mathbf{c}^\top \mathbf{x} + \mathbf{b}^\top \mathbf{y} \geq (-A^\top \mathbf{y})^\top \mathbf{x} + (A\mathbf{x})^\top \mathbf{y} = 0.$$

Then, according to Farkas lemma — see Appendix — we get

$$\begin{bmatrix} -\mathbf{c} \\ \mathbf{b} \\ 0 \end{bmatrix} = \begin{bmatrix} -A^\top & 0 & I_k & 0 \\ 0 & A & 0 & I_n \\ \mathbf{b}^\top & -\mathbf{c}^\top & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{y}^\star \\ \mathbf{x}^\star \\ \mathbf{x}' \\ \mathbf{y}' \end{bmatrix}, \qquad \text{for some } \mathbf{y}^\star, \mathbf{x}^\star, \mathbf{x}', \mathbf{y}' \geq 0.$$

The first and second block of rows say that $\mathbf{y}^\star$ and $\mathbf{x}^\star$ are feasible, while the third block of rows say that $\mathbf{b}^\top \mathbf{y}^\star = \mathbf{c}^\top \mathbf{x}^\star$. Thus, it remains to show that $\mathbf{x}^\star$ and $\mathbf{y}^\star$ provide extrema. This is true because, for instance, given a feasible $\mathbf{x}$, we get

$$\mathbf{c}^\top \mathbf{x}^\star = \mathbf{b}^\top \mathbf{y}^\star \geq (A\mathbf{x})^\top \mathbf{y}^\star = \mathbf{x}^\top (A^\top \mathbf{y}^\star) = \mathbf{x}^\top (\mathbf{c} + \mathbf{x}') \geq \mathbf{x}^\top \mathbf{c} = \mathbf{c}^\top \mathbf{x}.$$

As for the second statement, assuming for instance that a maximizer $\mathbf{x}^\star$ exists, we have

$$\begin{bmatrix} -A & 0 & \mathbf{b} \\ 0 & -A & \mathbf{b} \\ I_k & 0 & 0 \\ 0 & I_k & 0 \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{x}^\star \\ 1 \end{bmatrix} \geq 0 \iff \mathbf{x} \text{ is feasible} \implies \begin{bmatrix} -\mathbf{c} \\ \mathbf{c} \\ 0 \end{bmatrix}^\top \begin{bmatrix} \mathbf{x} \\ \mathbf{x}^\star \\ 1 \end{bmatrix} = -\mathbf{c}^\top \mathbf{x} + \mathbf{c}^\top \mathbf{x}^\star \geq 0.$$

Therefore, Farkas lemma implies

$$\begin{bmatrix} -\mathbf{c} \\ \mathbf{c} \\ 0 \end{bmatrix} = \begin{bmatrix} -A^\top & 0 & I_k & 0 \\ 0 & -A^\top & 0 & I_k \\ \mathbf{b}^\top & \mathbf{b}^\top & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{y}' \\ \mathbf{y}'' \\ \mathbf{x}' \\ \mathbf{x}'' \end{bmatrix}, \qquad \text{for some } \mathbf{y}', \mathbf{y}'', \mathbf{x}', \mathbf{x}'' \geq 0.$$

The first block of rows imply that $\mathbf{y}'$ is feasible. We can then make use of the first part. □

6.3 How does ℓ_1-magic work?

The ℓ_1-magic package is a MATLAB code designed specifically to solve the problems (P_1) and ($\mathrm{P}_{1,\varepsilon}$), among others. Simple techniques such as simplex methods or descent methods do not perform well in this case — try using MATLAB's own optimization toolbox! Instead, the implementation of ℓ_1-magic relies on interior points methods: problems reformulated as linear programs, such as (P_1), use a generic path-following primal-dual algorithm, and problems reformulated as second-order cone programs, such as ($\mathrm{P}_{1,\varepsilon}$), use a generic log-barrier algorithm. More details can be found in the ℓ_1-magic user's guide [2] and in Chapter 11 of the book [1].

Let us mention, without justification, that the central-path for the convex problem (6.1) is a curve $(\mathbf{z}^\star(\tau))_{\tau>0}$, where $\mathbf{z}^\star(\tau)$ is a minimizer of

$$\text{minimize } \tau F_0(\mathbf{z}) + \Phi(\mathbf{z}) \quad \text{subject to } G\mathbf{z} = b, \tag{6.2}$$

where the barrier function Φ is the convex function defined by

$$\Phi(\mathbf{z}) := -\sum_{i=1}^{k} \log(-F_i(\mathbf{z})).$$

Each minimization problem (6.2) is solved via the Karush–Kuhn–Tucker conditions, aka KKT conditions, just as many convex optimization algorithms operate. The KKT conditions generalize the Lagrange multipliers method to inequality constraints. They are necessary

conditions for $\mathbf{x}^\star \in \mathbb{R}^n$ to be a local minimizer of a — not necessarily convex — problem of the type (6.1), provided that some regularity conditions are fulfilled and that the objective function F_0, the inequality constraint functions $F_1, \ldots, F_k$ and the inequality constraint functions $G_1, \ldots, G_\ell$ are all differentiable. Furthermore, if the optimization problem is convex, then the conditions are also sufficient for the vector $\mathbf{x}^\star \in \mathbb{R}^n$ to be a local — hence global — minimizer. The KKT conditions on $\mathbf{x}^\star \in \mathbb{R}^n$ state that there exist $\lambda^\star \in \mathbb{R}^k$ and $\nu^\star \in \mathbb{R}^\ell$ such that

$$\begin{aligned}
\text{primal feasibility:} \quad & F_i(\mathbf{x}^\star) \leq 0,\ G_j(\mathbf{x}^\star) = 0, \quad i \in [1:k], j \in [1:\ell], \\
\text{dual feasibility:} \quad & \lambda_i^\star \geq 0, \quad i \in [1:k], \\
\text{complementary slackness:} \quad & \lambda_i^\star F_i(\mathbf{x}^\star) = 0, \quad i \in [1:k], \\
\text{stationary:} \quad & \nabla F_0(\mathbf{x}^\star) + \sum_{i=1}^{k} \lambda_i^\star \nabla F_i(\mathbf{x}^\star) + \sum_{j=1}^{\ell} \nu_j^\star \nabla G_j(\mathbf{x}^\star) = 0.
\end{aligned}$$

Exercises

Ex.1: Prove that equality constraints can always be eliminated in a convex optimization problem.

Ex.2: Show that the problem of best approximation to an element $\mathbf{x} \in \mathbb{R}^n$ by elements of a linear subspace $\mathcal{V}$ of $\mathbb{R}^n$ relative to the max-norm, that is the minimization problem

$$\underset{\mathbf{v} \in \mathcal{V}}{\text{minimize}} \ \|\mathbf{x} - \mathbf{v}\|_\infty$$

can be reformulated as a linear optimization problem.

Ex.3: Verify carefully the equivalence between the problems (P_1) and ($\mathrm{P'}_1$).

Ex.4: Can the continuity assumption be dropped in Proposition 6.2?

Ex.5: Given an $m \times N$ sensing matrix with complex entries and a complex measurement vector $\mathbf{y} \in \mathbb{C}^m$, reformulate the problem

$$\text{minimize } \|\mathbf{z}\|_1 \quad \text{subject to} \quad A\mathbf{z} = \mathbf{y}$$

as a second-order cone programming problem.

Chapter 7
Coherence and Recovery by ℓ_1-minimization

Now that we have noticed that ℓ_1-minimization offers a practical way to reconstruct sparse vectors whenever the Null-Space Property is fulfilled, we must supply matrices satisfying the Null-Space Property. This is the aim of the next few chapters. We shall derive here the Null-Space Property from the notion of coherence. This will enable us to underline a deterministic process to reconstruct s-sparse vectors from $m \asymp s^2$ measurements. This is not the optimal order $m \asymp s$ yet.

7.1 Definitions and Estimates

The term coherence can either apply to an ℓ_2-normalized system of vectors — called a dictionary if it spans the whole space — or to a matrix whose columns are ℓ_2-normalized. We will also see in Section 7.3 the related notion of mutual coherence that applies to a pair of orthonormal bases.

Definition 7.1. The coherence of an ℓ_2-normalized system of vectors $\mathcal{A} = (\mathbf{a}_1, \dots, \mathbf{a}_N)$ in the space $\mathbb{C}^m$ is defined as

$$\mu(\mathcal{A}) := \max_{1 \le i \ne j \le N} |\langle \mathbf{a}_i, \mathbf{a}_j \rangle|.$$

For an $m \times N$ matrix A whose columns are ℓ_2-normalized, The coherence of A is the coherence of the systems composed by the columns of A.

We may already observe that the coherence of an orthonormal basis equals zero. In general, the theoretical guarantees for ℓ_1-minimization or for greedy algorithms improve when the

coherence becomes smaller. The following result, known as Welch bound, tells us how small the coherence can become.

Proposition 7.2. For any $m \times N$ matrix A whose columns are ℓ_2-normalized, one has

$$\sqrt{\frac{N-m}{m(N-1)}} \leq \mu(A) \leq 1.$$

Proof. Let $\mathbf{a}_1, \ldots, \mathbf{a}_N \in \mathbb{C}^m$ be the ℓ_2-renormalized columns of the matrix A. The upper bound is clear, in view of

$$|\langle \mathbf{a}_i, \mathbf{a}_j \rangle| \leq \|\mathbf{a}_i\|_2 \cdot \|\mathbf{a}_j\|_2 = 1 \cdot 1 = 1, \qquad \text{for all } i, j \in [1:N].$$

Let us now establish the upper bound. We introduce the $N \times N$ Gram matrix G associated to the system $(\mathbf{a}_1, \ldots, \mathbf{a}_N)$, as defined by

$$G_{i,j} := \langle \mathbf{a}_i, \mathbf{a}_j \rangle, \qquad i, j \in [1:N].$$

Let us notice that $G = A^\top A$. Let us also introduce the $m \times m$ matrix $\widetilde{G} := AA^\top$. On the one hand, because the system $(\mathbf{a}_1, \ldots, \mathbf{a}_k)$ is ℓ_2-normalized, we have

$$\operatorname{tr}(G) = \sum_{i=1}^{N} \|\mathbf{a}_i\|_2^2 = N. \tag{7.1}$$

On the other hand, remembering that the inner product

$$\langle\langle U, V \rangle\rangle := \operatorname{tr}(U^\top V) = \sum_{i,j=1}^{n} u_{i,j} \overline{v_{i,j}} \tag{7.2}$$

induces the so-called Froebenius norm $|||\cdot|||$ on the space of $n \times n$ matrices, we have

$$\operatorname{tr}(G) = \operatorname{tr}(\widetilde{G}) = \langle\langle I_m, \widetilde{G} \rangle\rangle \leq |||I_m||| \cdot |||\widetilde{G}||| = \sqrt{\operatorname{tr}(I_m)} \cdot \sqrt{\operatorname{tr}(\widetilde{G}^\top \widetilde{G})} = \sqrt{m} \cdot \sqrt{\operatorname{tr}(\widetilde{G}^\top \widetilde{G})}. \tag{7.3}$$

But now observe that

$$\begin{aligned} \operatorname{tr}(\widetilde{G}^\top \widetilde{G}) &= \operatorname{tr}(AA^\top AA^\top) = \operatorname{tr}(A^\top AA^\top A) = \operatorname{tr}(G^\top G) \\ &= \sum_{1 \leq i,j \leq N} |\langle \mathbf{a}_i, \mathbf{a}_j \rangle|^2 = \sum_{1 \leq i \leq N} \|\mathbf{a}_i\|_2^2 + \sum_{1 \leq i \neq j \leq N} |\langle \mathbf{a}_i, \mathbf{a}_j \rangle| \leq N + (N^2 - N) \cdot \mu(A)^2. \end{aligned} \tag{7.4}$$

Combining (7.1), (7.3), and (7.4), we obtain

$$N \leq \sqrt{m} \cdot \sqrt{N + (N^2 - N) \cdot \mu(A)^2}.$$

The required result is just a rearrangement of this inequality. □

Observe that the Welch bound behaves like $1/\sqrt{m}$ if N is large. It is easy to construct an $m \times (2m)$ matrix whose coherence equals $1/\sqrt{m}$ by concatenating the identity and Fourier matrices of size m. We now wish to construct an $m \times N$ matrix whose coherence equals $1/\sqrt{m}$ with a much larger N. We shall achieve this with $N = m^2$. Note that the Welch bound equals $1/\sqrt{m+1}$ in this case.

Proposition 7.3. If m is a prime number not equal to 2 nor 3, then there exists an $m \times m^2$ matrix A with $\mu(A) = \dfrac{1}{\sqrt{m}}$.

Proof. We identify $[1:m]$ with $\mathbb{Z}_m := \mathbb{Z}/mZ$. For $k, \ell \in \mathbb{Z}_m$, we define the translation and modulation operators $T_k, M_\ell : \mathbb{C}^{\mathbb{Z}_m} \to \mathbb{C}^{\mathbb{Z}_m}$ by setting, for $\mathbf{z} \in \mathbb{C}^{\mathbb{Z}_m}$,

$$\begin{aligned} (T_k \mathbf{z})_j &= z_{j-k}, \\ (M_\ell \mathbf{z})_j &= e^{i2\pi\ell j/m} z_j. \end{aligned}$$

We then define a vector $\mathbf{x} \in \mathbb{C}^{\mathbb{Z}_m}$ and an $m \times m^2$ matrix A by

$$x_j := \frac{1}{\sqrt{m}} e^{i2\pi j^3/m}, \qquad A = \Big[M_1 T_1 \mathbf{x} \,\Big|\, \cdots \,\Big|\, M_1 T_m \mathbf{x} \,\Big|\, \cdots \,\Big|\, M_m T_1 \mathbf{x} \,\Big|\, \cdots \,\Big|\, M_m T_m \mathbf{x} \Big].$$

Observe first that the columns $M_\ell T_k \mathbf{x}$ of A are ℓ_2-normalized, because the translation and modulation operators are isometries of $\ell_2(\mathbb{Z}_m)$ and because the vector $\mathbf{x}$ is ℓ_2-normalized. Then, for $(k, \ell) \neq (k', \ell')$, we have

$$\begin{aligned} \langle M_\ell T_k \mathbf{x}, M_{\ell'} T_{k'} \mathbf{x} \rangle &= \sum_{j \in \mathbb{Z}_m} (M_\ell T_k \mathbf{x})_j \cdot \overline{(M_{\ell'} T_{k'} \mathbf{x})_j} \\ &= \sum_{j \in \mathbb{Z}_m} e^{i2\pi\ell j/m} \, x_{j-k} \cdot \overline{e^{i2\pi\ell' j/m} \, x_{j-k'}} \\ &= \sum_{j \in \mathbb{Z}_m} e^{i2\pi(\ell-\ell')j/m} \cdot \frac{1}{\sqrt{m}} e^{i2\pi(j-k)^3/m} \cdot \frac{1}{\sqrt{m}} e^{-i2\pi(j-k')^3/m} \\ &= \frac{1}{m} \sum_{j \in \mathbb{Z}_m} e^{i2\pi(\ell-\ell')j/m} \cdot e^{i2\pi[(j-k)^3 - (j-k')^3]/m}. \end{aligned}$$

Let us set $a := \ell - \ell'$ and $b := k - k'$. We make the change of summation index $h = j - k'$ in the first sum to get

$$\begin{aligned} |\langle M_\ell T_k \mathbf{x}, M_{\ell'} T_{k'} \mathbf{x} \rangle| &= \frac{1}{m} \Big| \sum_{h \in \mathbb{Z}_m} e^{i2\pi a(h+k')/m} \cdot e^{i2\pi[(h-b)^3 - h^3]/m} \Big| \\ &= \frac{1}{m} \Big| \sum_{h \in \mathbb{Z}_m} e^{i2\pi a h/m} \cdot e^{i2\pi[3bh^2 - 3b^2 h - b^3]/m} \Big| \\ &= \frac{1}{m} \Big| \sum_{h \in \mathbb{Z}_m} e^{i2\pi[3bh^2 + (a-3b^2)h]/m} \Big|. \end{aligned}$$

We now set $c := 3b$ and $d := a - 3b^2$. Instead of concentrating on the previous modulus, we will in fact look at its square. We obtain

$$\begin{aligned}
|\langle M_\ell T_k \mathbf{x}, M_{\ell'} T_{k'} \mathbf{x} \rangle|^2 &= \frac{1}{m^2} \Big(\sum_{h \in \mathbb{Z}_m} e^{i2\pi[ch^2+dh]/m} \Big) \cdot \Big(\overline{\sum_{h' \in \mathbb{Z}_m} e^{i2\pi[ch'^2+dh']/m}} \Big) \\
&= \frac{1}{m^2} \sum_{h,h' \in \mathbb{Z}_m} e^{i2\pi[ch^2-ch'^2+dh-dh']/m} \\
&= \frac{1}{m^2} \sum_{h,h' \in \mathbb{Z}_m} e^{i2\pi(h-h')[c(h+h')+d]/m} \\
&\underset{j=h-h'}{=} \frac{1}{m^2} \sum_{h,j \in \mathbb{Z}_m} e^{i2\pi j[c(2h-j)+d]/m} \\
&= \frac{1}{m^2} \sum_{j \in \mathbb{Z}_m} e^{i2\pi j[-cj+d]/m} \Big(\sum_{h \in \mathbb{Z}_m} e^{i2\pi\, 2jch/m} \Big).
\end{aligned}$$

We now notice that, for each $j \in \mathbb{Z}_m$, we have

$$\sum_{h \in \mathbb{Z}_m} e^{i2\pi\, 2jch/m} = \begin{cases} m & \text{if } 2jc = 0 \mod m, \\ 0 & \text{if } 2jc \neq 0 \mod m. \end{cases}$$

At this point, we separate two cases:
1/ $c = 0 \mod m$. Because 3 is nonzero in the field $\mathbb{Z}_m$, this means that $b = 0$, i.e. that $k = k'$. This implies that $\ell \neq \ell'$, i.e. that $a \neq 0$, and therefore that $d \neq 0$. Thus, we derive

$$|\langle M_\ell T_k \mathbf{x}, M_{\ell'} T_{k'} \mathbf{x} \rangle|^2 = \frac{1}{m^2} \sum_{j \in \mathbb{Z}_m} e^{i2\pi j[-cj+d]/m} \cdot m = \frac{1}{m} \sum_{j \in \mathbb{Z}_m} e^{i2\pi jd/m} = 0.$$

2/ $c \neq 0 \mod m$. Because 2 is nonzero in the field $\mathbb{Z}_m$, the only possibility to have $2jc = 0$ is $j = 0$. We then derive

$$|\langle M_\ell T_k \mathbf{x}, M_{\ell'} T_{k'} \mathbf{x} \rangle|^2 = \frac{1}{m^2} \cdot 1 \cdot m = \frac{1}{m}.$$

The conclusion $\mu(A) \leq \dfrac{1}{\sqrt{m}}$ follows, since we have established that, for all $(k, \ell) \neq (k', \ell')$, there holds $|\langle M_\ell T_k \mathbf{x}, M_{\ell'} T_{k'} \mathbf{x} \rangle| \leq \dfrac{1}{\sqrt{m}}$. □

7.2 Small Coherence Implies ℓ_1-Recovery

We shall use the equivalence between sparse recovery and Null-Space Property to establish that s-sparse vectors $\mathbf{x} \in \mathbb{R}^N$ can be reconstructed from the measurements $\mathbf{y} = A\mathbf{x} \in \mathbb{R}^m$ provided that the coherence of the sensing matrix A is small enough, namely provided that $\mu(A) < 1/(2s-1)$.

Theorem 7.4. Suppose that the $m \times N$ sensing matrix A has a coherence obeying

$$\mu(A) < \frac{1}{2s-1}.$$

Then the matrix A satisfy the Null-Space Property of order s relative to ℓ_1.

Proof. Let us consider a vector $\mathbf{v} \in \ker A$ and an index set S with $|S| \le s$. Denoting by $\mathbf{a}_1, \ldots, \mathbf{a}_N$ the ℓ_2-normalized columns of the matrix A, the condition $\mathbf{v} \in \ker A$ translates into

$$\sum_{\ell=1}^{N} v_\ell \mathbf{a}_\ell = 0.$$

Thus, for any $j \in [1:N]$, we have

$$v_j \mathbf{a}_j = - \sum_{\ell=1, \ell \neq j}^{N} v_\ell \mathbf{a}_\ell.$$

Taking the inner product with $\mathbf{a}_j$, we obtain

$$v_j = - \sum_{\ell=1, \ell \neq j}^{N} v_\ell \langle \mathbf{a}_j, \mathbf{a}_\ell \rangle.$$

It then follows that

$$|v_j| \le \sum_{\ell=1, \ell \neq j}^{N} |v_\ell| \mu(A) = \mu(A)(\|\mathbf{v}\|_1 - |v_j|).$$

Rearranging and summing over $j \in S$, we obtain

$$\|\mathbf{v}_S\|_1 \le s \, \frac{\mu(A)}{1+\mu(A)} \|\mathbf{v}\|_1.$$

Therefore, the Null-Space Property is fulfilled as soon as

$$s \, \frac{\mu(A)}{1+\mu(A)} < \frac{1}{2}, \quad \text{i.e. } \; 2s\,\mu(A) < 1 + \mu(A), \quad \text{or } \; (2s-1)\mu(A) < 1.$$

This is the required sufficient condition. □

Corollary 7.5. The $m \times m^2$ matrix of Proposition 7.3 allows reconstruction of s-sparse vectors by ℓ_1-minimization as soon as

$$s < \frac{\sqrt{m}+1}{2} \asymp \sqrt{m}.$$

Proof. The coherence of this matrix A is given by $\mu(A) = 1/\sqrt{m}$. Thus, the sufficient condition of Theorem 7.4 is equivalent to $1/\sqrt{m} < 1/(2s-1)$, that is $2s - 1 < \sqrt{m}$, or $s < (\sqrt{m}+1)/2$. □

7.3 Mutual Coherence

We now introduce the notion of mutual coherence. Theorem 7.8, which will be stated but not proved, claims that if the sensing basis and the representation basis are incoherent — i.e. have a small mutual coherence — then s-sparse recovery is achievable with an almost optimal number of measurements.

Definition 7.6. The mutual coherence between two orthonormal bases $\Phi = (\phi_1, \dots, \phi_m)$ and $\Psi = (\psi_1, \dots, \psi_m)$ of $\mathbb{C}^m$ is given by

$$\mu(\Phi, \Psi) := \sqrt{m} \max_{1 \le i \ne j \le N} |\langle \phi_i, \psi_j \rangle|.$$

The mutual coherence between the orthonormal bases Φ and Ψ is of course closely related to the coherence of the system obtained by concatenation of the bases Φ and Ψ, precisely

$$\mu(\Phi, \Psi) = \sqrt{m} \cdot \mu((\phi_1, \dots, \phi_m, \psi_1, \dots, \psi_m)) = \sqrt{m} \cdot \mu(A),$$
$$\text{where} \quad A := \left[\begin{array}{c|c|c|c|c} \phi_1 & \cdots & \phi_m & \psi_1 & \cdots & \psi_m \end{array} \right].$$

Applying Proposition 7.2 to such a situation, in which case $N = 2m$, we get

$$\sqrt{\frac{m}{2m-1}} \le \mu(\Phi, \Psi) \le \sqrt{m}.$$

This is not quite optimal. In fact, we have the following stronger result.

Proposition 7.7. The mutual coherence between two orthonormal bases Φ and Ψ of $\mathbb{C}^m$ satisfies

$$1 \le \mu(\Phi, \Psi) \le \sqrt{m}.$$

These inequalities are sharp, since

$$\begin{aligned} \mu(\Phi, \Phi) &= \sqrt{m} \quad \text{for any othonormal basis } \Phi, \\ \mu(\Phi, \Psi) &= 1 \quad \text{for the canonical basis } \Phi \text{ and the Fourier basis } \Psi. \end{aligned}$$

Proof. We only need to consider the lower estimate. Given two orthonormal bases Φ and Ψ of $\mathbb{C}^m$, we have, for any $k \in [1 : m]$,

$$1 = \|\phi_k\|_2^2 = \sum_{j=1}^m |\langle \phi_k, \psi_j \rangle|^2 \le \sum_{j=1}^m \left(\frac{\mu(\Phi, \Psi)}{\sqrt{m}} \right)^2 = \mu(\Phi, \Psi)^2,$$

as was required. To prove that this estimate is sharp, we consider the orthonormal bases $\Phi = (\phi_1, \dots, \phi_m)$ and $\Psi = (\psi_1, \dots, \psi_m)$ of $\mathbb{C}^m$ defined by

$$\begin{aligned} \phi_k &= \big[0, \dots, 0, \overbrace{1}^{\text{pos. } k}, 0, \dots, 0\big]^\top, \\ \psi_j &= \frac{1}{\sqrt{m}} \big[1, e^{i2\pi j/m}, \dots, e^{i2\pi j(m-1)/m}\big]^\top. \end{aligned}$$

We have, for any $k, j \in [1 : m]$,

$$|\langle \phi_k, \psi_j \rangle| = \Big| \frac{1}{\sqrt{m}} e^{i2\pi jk/m} \Big| = \frac{1}{\sqrt{m}},$$

which immediately implies that $\mu(\Phi, \Psi) = 1$, as announced. □

We shall now state the announced theorem. As an informal corollary, we notice that if the Fourier basis is used as the sensing basis Φ and the canonical basis as the representation basis Ψ, in which case $\mu(\Phi, \Psi) = 1$, then exact reconstruction of s-sparse vectors from m random Fourier samples by ℓ_1-minimization occurs with probability $\geq 1 - \delta$, provided that

$$m \geq \mathrm{cst} \cdot s \cdot \log(N/\delta).$$

Theorem 7.8. Let the sensing basis Φ and the representation basis Ψ be two orthonormal bases of $\mathbb{C}^N$. Let $S \subseteq [1 : N]$ be a fixed index set. We choose a set $M \subseteq [1 : N]$ of m measurements and a sign sequence σ on S uniformly at random. There exists absolute constants C_1 and C_2 such that, for all $\delta > 0$, if

$$m \geq \max\big[C_1 \cdot \mu(\Phi, \Psi)^2 \cdot s \cdot \log(N/\delta),\, C_2 \cdot \log^2(N/\delta)\big],$$

then

$$\mathbb{P}\Big(\forall\, \mathbf{x} \in \Sigma_S :\ \|\mathbf{x}\|_1 \leq \|\mathbf{z}\|_1 \text{ whenever } \langle \textstyle\sum_j z_j \psi_j, \phi_i \rangle = \langle \sum_j x_j \psi_j, \phi_i \rangle, \text{ all } i \in M\Big) \geq 1 - \delta.$$

Exercises

Ex.1: Find the systems of three vectors in $\mathbb{R}^2$ with the smallest coherence.

Ex.2: Establish that the Gram matrix associated to a system $(\mathbf{a}_1, \dots, \mathbf{a}_k)$ is a symmetric positive-semidefinite matrix, and that it is positive-definite whenever the system $(\mathbf{a}_1, \dots, \mathbf{a}_k)$ is linearly independent.

Ex.3: Verify that the expression of $\langle\langle U, V \rangle\rangle$ given in (7.2) indeed defines an inner product on the algebra $\mathcal{M}_n(\mathbb{C})$ of complex $n \times n$ matrices. Show that, for the induced Froebenius norm, one has $|||U||| = |||U^\top||| = |||U^\top U|||^{1/2}$ for all $U \in \mathcal{M}_n(\mathbb{C})$. Prove at last that the Froebenius norm is a matrix norm, in the sense that $|||UV||| \leq |||U||| \cdot |||V|||$ for all $U, V \in \mathcal{M}_n(\mathbb{C})$.

Ex.4: Verify that the Fourier basis $(\psi_1, \ldots, \psi_m)$ introduced in Proposition 7.7 is an orthonormal basis of $\mathbb{C}^m$.

Ex.5: Establish that the reconstruction of s-sparse vectors by ℓ_1-minimization is stable whenever $(2s-1)\mu(A) \leq \gamma$ for some $0 < \gamma < 1$.

Ex.6: Follow the steps involved in the proof of Theorem 7.4 to derive an analogous result for ℓ_q-minimization. Does the condition guaranteeing recovery becomes weaker when the exponent q decreases?

Ex.7: Imitate the MATLAB commands of Section 1.2 to illustrate Theorem 7.8. Select m rows of the Fourier matrix at random first, then try making some deterministic choices.

Chapter 8
Restricted Isometry Property and Recovery by ℓ_1-Minimization

In the previous chapter, we have isolated a condition on the coherence of the measurement matrix that guarantees sparse recovery by ℓ_1-minimization. Here, we present a condition on the so-called restricted isometry constants of the measurement matrix that guarantees sparse recovery by ℓ_1-minimization, too. The proof of Theorem 8.2 is a high point of the course. Note that it is not optimal, though, since a stronger statement will be established in Chapter 10, but that it is the simplest and most natural proof.

8.1 Restricted Isometry Property

Given the $m \times N$ measurement matrix A, suppose that we can recover s-sparse vectors $\mathbf{x} \in \mathbb{R}^N$ from the knowledge of the measurement vector $\mathbf{y} = A\mathbf{x} \in \mathbb{R}^m$ by solving the minimization problem

$$\text{(P}_1\text{)} \qquad \text{minimize } \|\mathbf{z}\|_1 \quad \text{subject to} \quad A\mathbf{z} = \mathbf{y}.$$

Because there is a reconstruction map — given by ℓ_1-minimization — associated to the measurement matrix A, we know that

$$\Sigma_{2s} \cap \ker A = \{0\}.$$

This condition is equivalent to

$$\forall\, \mathbf{v} \in \Sigma_{2s}, \qquad A\mathbf{v} \neq 0.$$

Fixing $p, r \in (0, \infty]$, compactness arguments show that this is also equivalent to

$$\text{there exists a constant } \alpha > 0 : \quad \forall\, \mathbf{v} \in \Sigma_{2s}, \qquad \|A\mathbf{v}\|_r^r \geq \alpha \, \|\mathbf{v}\|_p^p.$$

We denote by $\alpha_{2s}^{[p,r]}(A)$ the largest such constant. On the other hand, it is clear that we can define a constant $\beta_{2s}^{[p,r]}(A)$ as the smallest constant β such that

$$\forall\, \mathbf{v} \in \Sigma_{2s}, \qquad \|A\mathbf{v}\|_r^r \leq \beta\, \|\mathbf{v}\|_p^p.$$

Thus, assuming that s-sparse reconstruction is achievable via ℓ_1-minimization, we can define a finite quantity as the ration of these two constants. Conversely, it will turn out that the possibility of s-sparse reconstruction via ℓ_1-minimization is dictated by how small this ratio is, i.e. by how close the lower and upper constants α and β are.

Definition 8.1. For $p, r \in (0, \infty]$, the k-th order Restricted Isometry Ratio of the measurement matrix A as an operator from ℓ_p^N into ℓ_r^m is defined by

$$\gamma_k^{[p,r]}(A) \;:=\; \frac{\alpha_k^{[p,r]}(A)}{\beta_k^{[p,r]}(A)} \quad \in [1, \infty],$$

where $\alpha_k^{[p,r]}(A)$ and $\beta_k^{[p,r]}(A)$ are the largest and smallest positive constants α and β such that

$$\forall\, \mathbf{v} \in \Sigma_k, \qquad \alpha\, \|\mathbf{v}\|_p^p \;\leq \|A\mathbf{v}\|_r^r \;\leq\; \beta\, \|\mathbf{v}\|_p^p.$$

When the superscript is omitted, we implicitly understand

$$\gamma_k(A) \;=\; \gamma_k^{[p,r]}(A).$$

What is traditionally called the k-th order Restricted Isometry Property with constant $\delta \in (0, 1)$ for the matrix A is the fact that

$$\forall\, \mathbf{v} \in \Sigma_k, \qquad (1 - \delta)\, \|\mathbf{v}\|_2^2 \;\leq\; \|A\mathbf{v}\|_2^2 \;\leq\; (1 + \delta)\, \|\mathbf{v}\|_2^2.$$

The smallest such constant δ is called the Restricted Isometry Constant of the matrix A, and is denoted by $\delta_k(A)$. Following the previous model, we can also define a k-th order Restricted Isometry Constant $\delta_k^{[p,r]}(A)$ for the matrix A as an operator from ℓ_p^N into ℓ_r^m. The relations with the Restricted Isometry Ratio are

$$\gamma_k^{[p,r]}(A) \;:=\; \frac{1 + \delta_k^{[p,r]}(A)}{1 - \delta_k^{[p,r]}(A)}, \qquad \delta_k^{[p,r]}(A) \;:=\; \frac{\gamma_k^{[p,r]}(A) - 1}{\gamma_k^{[p,r]}(A) + 1}.$$

In general, we prefer to deal with the Restricted Isometry Ratio rather than the Restricted Isometry Constant, because the latter is not homogeneous in the matrix A, while the former is homogeneous, that is to say

$$\gamma_k^{[p,r]}(cA) \;=\; \gamma_k^{[p,r]}(A), \qquad \text{for all } c \in \mathbb{R},\ c \neq 0.$$

8.2 Recovery by ℓ_1-minimization

According to the introductory remark in Section 8.1, we shall try, as much as possible, to link s-sparse recovery with $2s$-th order Restricted Isometry Ratio when establishing that a small Restricted Isometry Ratio implies ℓ_1-recovery. We recall ℓ_1-recovery is guaranteed as soon as the Null-Space Property is fulfilled, and in fact as soon as the stronger form of the Null-Space Property introduced at the end of Chapter 4 is fulfilled. Let us mention that the proof below appears only valid for real-valued signals and measurements, but that the result also holds in the complex case.

Theorem 8.2. Under the condition that

$$\gamma_{2s}(A) \ < \ 2, \tag{8.1}$$

the measurement matrix A satisfies the ℓ_2-Strong Null-Space Property relative to ℓ_1, i.e.

$$\forall\, \mathbf{v} \in \ker A, \quad \forall\, |S| \le s, \quad \|\mathbf{v}_S\|_2 \le \frac{\eta}{\sqrt{s}}\, \|\mathbf{v}\|_1, \qquad \text{with } \eta := \frac{\gamma_{2s}-1}{2} < \frac{1}{2},$$

so that every s-sparse vector $\mathbf{x} \in \mathbb{R}^N$ is the unique solution of (P_1) with $\mathbf{y} = A\mathbf{x}$.

Proof. Given $\mathbf{v} \in \ker A$, it is enough to prove the result for the index set $S = S_0$ corresponding to the s largest absolute-value components of $\mathbf{v}$. We also partition the complement of S_0 as $\bar{S_0} = S_1 \cup S_2 \cup \cdots$, where

$$\begin{aligned} S_1 &:= \{\text{indices of the next } s \text{ largest absolute-value components of } \mathbf{v} \text{ in } \overline{S}\},\\ S_2 &:= \{\text{indices of the next } s \text{ largest absolute-value components of } \mathbf{v} \text{ in } \overline{S}\},\\ &\ \ \vdots \end{aligned}$$

Because of the fact that $\mathbf{v} \in \ker A$, we have $A\mathbf{v}_S = A(-\mathbf{v}_{S_1} - \mathbf{v}_{S_2} - \cdots)$, thus

$$\|\mathbf{v}_S\|_2^2 \le \frac{1}{\alpha_{2s}} \|A\mathbf{v}_S\|_2^2 = \frac{1}{\alpha_{2s}} \langle A\mathbf{v}_S, -A\mathbf{v}_{S_1} - A\mathbf{v}_{S_2} - \cdots \rangle = \frac{1}{\alpha_{2s}} \sum_{k \ge 1} \langle A\mathbf{v}_S, -A\mathbf{v}_{S_k} \rangle. \tag{8.2}$$

For $k \ge 1$, we may write $-\mathbf{v}_{S_k} = \|\mathbf{v}_{S_k}\|_2\, \mathbf{u}_{S_k}$ for some ℓ_2-normalized vector $\mathbf{u}_{S_k}$ which is supported on S_k. Likewise, we may write $\mathbf{v}_S = \|\mathbf{v}_S\|_2\, \mathbf{u}_S$ for some ℓ_2-normalized vector $\mathbf{u}_S$ which is supported on S. We derive

$$\begin{aligned} \langle A\mathbf{v}_S, -A\mathbf{v}_{S_k} \rangle &= \langle A\mathbf{u}_S, A\mathbf{u}_{S_k} \rangle\, \|\mathbf{v}_S\|_2 \|\mathbf{v}_{S_k}\|_2 = \frac{1}{4}\Big[\|A(\mathbf{u}_S - \mathbf{u}_{S_k})\|_2^2 - \|A(\mathbf{u}_S + \mathbf{u}_{S_k})\|_2^2 \Big]\, \|\mathbf{v}_S\|_2 \|\mathbf{v}_{S_k}\|_2 \\ &\le \frac{1}{4}\Big[\beta_{2s}\, \|\mathbf{u}_S - \mathbf{u}_{S_k}\|_2^2 - \alpha_{2s}\, \|\mathbf{u}_S + \mathbf{u}_{S_k}\|_2^2 \Big]\, \|\mathbf{v}_S\|_2 \|\mathbf{v}_{S_k}\|_2 \\ &= \frac{1}{4}\big[\beta_{2s} \cdot 2 - \alpha_{2s} \cdot 2 \big]\, \|\mathbf{v}_S\|_2 \|\mathbf{v}_{S_k}\|_2. \end{aligned}$$

Substituting into (8.2) and simplifying by $\|\mathbf{v}_S\|_2$, we obtain

$$\|\mathbf{v}_S\|_2 \leq \frac{1}{\alpha_{2s}} \sum_{k\geq 1} \frac{1}{2}\big[\beta_{2s} - \alpha_{2s}\big] \, \|\mathbf{v}_{S_k}\|_2 = \frac{\gamma_{2s}-1}{2} \sum_{k\geq 1} \|\mathbf{v}_{S_k}\|_2 =: \eta \sum_{k\geq 1} \|\mathbf{v}_{S_k}\|_2. \tag{8.3}$$

Observe that, for $k \geq 1$, the inequality

$$|v_i| \leq |v_j|, \qquad i \in S_k, \quad j \in S_{k-1},$$

yields, by averaging over j, the inequality

$$|v_i| \leq \frac{1}{s} \, \|\mathbf{v}_{S_{k-1}}\|_1, \qquad i \in S_k.$$

Then, by squaring and summing over i, we obtain

$$\|\mathbf{v}_{S_k}\|_2^2 \leq \frac{1}{s} \, \|\mathbf{v}_{S_{k-1}}\|_1^2, \qquad \text{i.e.} \qquad \|\mathbf{v}_{S_k}\|_2 \leq \frac{1}{\sqrt{s}} \, \|\mathbf{v}_{S_{k-1}}\|_1.$$

In view of (8.3), we now deduce that

$$\|\mathbf{v}_S\|_2 \leq \frac{\eta}{\sqrt{s}} \sum_{k\geq 1} \|\mathbf{v}_{S_{k-1}}\|_1 \leq \frac{\eta}{\sqrt{s}} \, \|\mathbf{v}\|_1,$$

which is the required inequality. We finally observe that the condition $\eta < 1/2$ is fulfilled as soon as $\gamma_{2s} < 2$, which is exactly Condition (8.1). □

Let us remark that, in terms of Restricted Isometry Constant, Condition (8.1) translates into the condition

$$\delta_{2s}(A) \, < \, \frac{1}{3}. \tag{8.1'}$$

Exercises

Ex.1: What is the complex equivalent of the polarization formula

$$\langle \mathbf{u}, \mathbf{v} \rangle = \frac{1}{4}\big[\|\mathbf{u}+\mathbf{v}\|_2^2 - \|\mathbf{u}-\mathbf{v}\|_2^2\big].$$

Ex.2: Find the 2×3 real matrices with smallest second order Restricted Isometry Ratio.

Ex.3: Verify that adding measurements does not increase the Restricted Isometry Ratio.

Ex.4: Write a program to evaluate the Restricted Isometry Ratio of a matrix A. Does it work well for large dimensions?

Ex.5: Repeat the proof of Theorem 8.2 by replacing the ℓ_1-norm with an ℓ_q-quasinorm. You will need to partition $\bar{S}_0$ into index sets of size t with $t \geq s$. Does the sufficient condition become weaker when the exponents q decreases?

Ex.6: Can you improve the sufficient condition (8.1)?

Ex.7: Given a measurement matrix A whose columns are ℓ_2-normalized, prove that A obeys the k-th order Restricted Isometry Property with constant $(k-1)\mu(A)$ for all $k < 1 + 1/\mu(A)$.

Chapter 9
Restricted Isometry Property for Random Matrices

In this chapter, we prove that, with 'high probability', an $m \times N$ 'random matrices' satisfy the k-th order Restricted Isometry Property with a prescribed constant $\delta \in (0,1)$, so long as $m \geq \mathrm{cst}(\delta) \cdot k \cdot \ln(eN/k)$. Thus, if we fix $\delta = 1/4$, say, Theorem 8.2 implies that it is possible to reconstruct s-sparse vectors by ℓ_1-minimization using certain random matrices as measurement matrices. For this purpose, the number m of measurements has to be of the order of $s \cdot \ln(N/s)$, which is close to the lower bound $m \geq 2s$. In Section 9.1, we prove the that Restricted Isometry Property follows from a certain Concentration Inequality. In Section 9.2, we introduce strictly subgaussian random matrices, and in Section 9.3, we establish that the latter obey the required Concentration Inequality.

9.1 Concentration Inequality Implies Restricted Isometry Property

This section is devoted to the proof of the following theorem.

Theorem 9.1. Given integers m, N, suppose that the matrix A is drawn according to a probability distribution satisfying, for each $\mathbf{x} \in \mathbb{R}^N$, the concentration inequality

$$\mathbb{P}\Big(\big|\|A\mathbf{x}\|_2^2 - \|\mathbf{x}\|_2\big| > \varepsilon \|\mathbf{x}\|_2^2\Big) \;\leq\; 2\,\exp(-c(\varepsilon)\,m), \qquad \varepsilon \in (0,1), \tag{9.1}$$

where $c(\varepsilon)$ is a constant depending only on ε. Then, for each $\delta \in (0,1)$, there exist constants $c_0(\delta), c_1(\delta) > 0$ depending on δ and on the probability distribution such that

$$\mathbb{P}\Big(\big|\|A\mathbf{x}\|_2^2 - \|\mathbf{x}\|_2\big| > \delta \|\mathbf{x}\|_2^2, \quad \text{for some } \mathbf{x} \in \Sigma_k\Big) \;\leq\; 2\,\exp(-c_0(\delta)\,m),$$

provided that

$$m \geq c_1(\delta) \cdot k \cdot \ln(eN/k).$$

The following lemma will be needed in the upcoming arguments.

Lemma 9.2. If $\mathcal{S}$ is the unit sphere of $\mathbb{R}^n$ relative to an arbitrary norm $\|\cdot\|$, then there exists a set $\mathcal{U} \in \mathcal{S}$ with

$$\forall\, \mathbf{z} \in \mathcal{S}, \quad \min_{\mathbf{u} \in \mathcal{U}} \|\mathbf{z} - \mathbf{u}\| \leq \delta, \qquad \text{and} \qquad |\mathcal{U}| \leq \Big(1 + \frac{2}{\delta}\Big)^n.$$

Proof. Let $(\mathbf{u}_1, \ldots, \mathbf{u}_h)$ be a set of h points on the sphere $\mathcal{S}$ such that $\|\mathbf{u}_i - \mathbf{u}_j\| > \delta$ for all $i \neq j$. We choose k as large as possible. Thus, it is clear that

$$\forall\, \mathbf{z} \in \mathcal{S}, \quad \min_{i \in [1:h]} \|\mathbf{z} - \mathbf{u}_i\| \leq \delta$$

Let $\mathcal{B}$ be the unit ball of $\mathbb{R}^n$ endowed with the norm $\|\cdot\|$. We have that

$$\mathbf{u}_1 + \frac{\delta}{2}\mathcal{B}, \ldots, \mathbf{u}_h + \frac{\delta}{2}\mathcal{B} \text{ are disjoints},$$

because, if $\mathbf{z}$ would belong to $\big[\mathbf{u}_i + \frac{\delta}{2}\mathcal{B}\big] \cap \big[\mathbf{u}_j + \frac{\delta}{2}\mathcal{B}\big]$, then we would have

$$\|\mathbf{u}_i - \mathbf{u}_j\| \leq \|\mathbf{u}_i - \mathbf{z}\| + \|\mathbf{u}_j - \mathbf{z}\| \leq \frac{\delta}{2} + \frac{\delta}{2} = \delta.$$

Besides, we also have that

$$\mathbf{u}_1 + \frac{\delta}{2}\mathcal{B}, \ldots, \mathbf{u}_h + \frac{\delta}{2}\mathcal{B} \text{ are included in } \Big(1 + \frac{\delta}{2}\Big)\mathcal{B},$$

because, if $\mathbf{z} \in \mathcal{B}$, then we have

$$\|\mathbf{u}_i + \frac{\delta}{2}\mathbf{z}\| \leq \|\mathbf{u}_i\| + \frac{\delta}{2}\|\mathbf{z}\| \leq 1 + \frac{\delta}{2}.$$

By comparison of volumes, we get

$$h\mathrm{Vol}\Big(\frac{\delta}{2}\mathcal{B}\Big) = \sum_{i=1}^{k} \mathrm{Vol}\Big(\mathbf{u}_i + \frac{\delta}{2}\mathcal{B}\Big) \leq \mathrm{Vol}\Big(\Big(1 + \frac{\delta}{2}\Big)\mathcal{B}\Big),$$

and then, by n-homogeneity of the volume,

$$h\Big(\frac{\delta}{2}\Big)^n \mathrm{Vol}(\mathcal{B}) \leq \Big(1 + \frac{\delta}{2}\Big)^n \mathrm{Vol}(\mathcal{B}),$$

which implies

$$h \leq \Big(1 + \frac{2}{\delta}\Big)^n.$$

□

Proof of Theorem 9.1. We suppose here that the concentration inequality (9.1) is fulfilled. Let for the moment K be a fixed index set of cardinality $|K| = k$. According to Lemma 9.2, we can find a subset $\mathcal{U}$ of the unit sphere $\mathcal{S}_{\Sigma_K}$ of Σ_K relative to the ℓ_2-norm such that

$$\forall\, \mathbf{z} \in \mathcal{S}_{\Sigma_K}, \quad \min_{\mathbf{u} \in \mathcal{U}} \|\mathbf{z} - \mathbf{u}\|_2 \le \frac{\delta}{4+\delta}, \qquad \text{and} \qquad |\mathcal{U}| \le \Big(3 + \frac{8}{\delta}\Big)^k.$$

Applying the concentration inequality with $\varepsilon = \delta/4$ to each element of $\mathcal{U}$ yields

$$\begin{aligned}&\mathbb{P}\Big(\big|\|A\mathbf{u}\|_2^2 - \|\mathbf{u}\|_2\big| > \frac{\delta}{4}\|\mathbf{u}\|_2^2, \quad \text{for some } \mathbf{u} \in \mathcal{U}\Big)\\ &\quad \le \sum_{\mathbf{u} \in \mathcal{U}} \mathbb{P}\Big(\big|\|A\mathbf{u}\|_2^2 - \|\mathbf{u}\|_2\big| > \frac{\delta}{4}\|\mathbf{u}\|_2^2\Big) \le 2|\mathcal{U}| \exp(-c(\delta/4)\, m) \le 2\Big(3 + \frac{8}{\delta}\Big)^k \exp(-c(\delta/4)\, m).\end{aligned}$$

Now suppose that the draw of the matrix A gives

$$\big|\|A\mathbf{u}\|_2^2 - \|\mathbf{u}\|_2\big| \le \frac{\delta}{4}\|\mathbf{u}\|_2^2, \qquad \text{all } \mathbf{u} \in \mathcal{U}. \tag{9.2}$$

In other words, we have

$$\Big(1 - \frac{\delta}{4}\Big)\|\mathbf{u}\|_2^2 \;\le\; \|A\mathbf{u}\|_2^2 \;\le\; \Big(1 + \frac{\delta}{4}\Big)\|\mathbf{u}\|_2^2, \qquad \text{all } \mathbf{u} \in \mathcal{U}.$$

We consider δ' to be the smallest number such that

$$\|A\mathbf{x}\|_2^2 \le (1+\delta')\|\mathbf{x}\|_2^2, \qquad \text{all } \mathbf{x} \in \Sigma_K.$$

Given $\mathbf{x} \in \Sigma_K$ with $\|\mathbf{x}\|_2 = 1$, we pick $\mathbf{u} \in \mathcal{U}$ such that $\|\mathbf{x} - \mathbf{u}\|_2 \le \delta/(4+\delta)$. We then derive

$$\|A\mathbf{x}\|_2 \le \|A\mathbf{u}\|_2 + \|A(\mathbf{x}-\mathbf{u})\|_2 \le \|A\mathbf{u}\|_2 + \sqrt{1+\delta'}\,\|\mathbf{x}-\mathbf{u}\|_2 \le \sqrt{1 + \frac{\delta}{4}} + \sqrt{1+\delta'} \cdot \frac{\delta}{4+\delta}.$$

Since δ' is the smallest number such that $\|A\mathbf{x}\|_2 \le \sqrt{1+\delta'}$ for every $\mathbf{x} \in \Sigma_K$ with $\|\mathbf{x}\|_2 = 1$, we obtain

$$\sqrt{1+\delta'} \le \sqrt{1 + \frac{\delta}{4}} + \sqrt{1+\delta'} \cdot \frac{\delta}{4+\delta}.$$

It follows that

$$1 + \frac{\delta}{4} \ge (1+\delta')\Big(1 - \frac{\delta}{4+\delta}\Big)^2 = (1+\delta')\Big(\frac{4}{4+\delta}\Big)^2 = (1+\delta')\Big(\frac{1}{1+\delta/4}\Big)^2.$$

Taking into account that $(1+t)^3 \le (1+4t)$ for $t \in (0, 1/4)$, we deduce

$$1 + \delta' \le \Big(1 + \frac{\delta}{4}\Big)^3 \le 1 + \delta.$$

The inequality $\delta' \le \delta$ that we have just obtained means that

$$\|A\mathbf{x}\|_2^2 \le (1+\delta)\|\mathbf{x}\|_2^2, \qquad \text{all } \mathbf{x} \in \Sigma_K.$$

On the other hand, given $\mathbf{x} \in \Sigma_K$ with $\|\mathbf{x}\|_2 = 1$, we still pick $\mathbf{u} \in \mathcal{U}$ such that $\|\mathbf{x} - \mathbf{u}\|_2 \le \delta/(4+\delta)$ to derive

$$\begin{aligned}
\|A\mathbf{x}\|_2^2 &\ge \big(\|A\mathbf{u}\|_2 - \|A(\mathbf{x}-\mathbf{u})\|_2\big)^2 \ge \|A\mathbf{u}\|_2^2 - 2\|A\mathbf{u}\|_2\|A(\mathbf{x}-\mathbf{u})\|_2 \\
&\ge 1 - \frac{\delta}{4} - 2\cdot\sqrt{1+\frac{\delta}{4}}\cdot\sqrt{1+\delta}\cdot\frac{\delta}{4+\delta} \ge 1 - \frac{\delta}{4} - 2\cdot\sqrt{1+\frac{\delta}{4}}\cdot\Big(1+\frac{\delta}{2}\Big)\cdot\frac{\delta}{4(1+\delta/4)} \\
&= 1 - \frac{\delta}{4} - \frac{\delta}{2}\cdot\frac{1+\delta/2}{\sqrt{1+\delta/4}} \ge 1 - \frac{\delta}{4} - \frac{\delta}{2}\Big(1+\frac{\delta}{2}\Big) \ge 1 - \frac{\delta}{4} - \frac{\delta}{2} - \frac{\delta}{4} = 1-\delta.
\end{aligned}$$

Thus, we have obtained

$$(1-\delta)\|\mathbf{x}\|_2^2 \;\le\; \|A\mathbf{x}\|_2^2 \;\le\; (1+\delta)\|\mathbf{x}\|_2^2, \qquad \text{all } \mathbf{x} \in \Sigma_K,$$

as soon as the inequality (9.2) holds. We therefore have

$$\begin{aligned}
&\mathbb{P}\Big(\big|\|A\mathbf{x}\|_2^2 - \|\mathbf{x}\|_2^2\big| > \delta\|\mathbf{x}\|_2^2, \quad \text{for some } \mathbf{x} \in \Sigma_K\Big) \\
&\qquad \le \mathbb{P}\Big(\big|\|A\mathbf{u}\|_2^2 - \|\mathbf{u}\|_2^2\big| > \frac{\delta}{4}\|\mathbf{u}\|_2^2, \quad \text{for some } \mathbf{u} \in \mathcal{U}\Big) \le 2\Big(3+\frac{8}{\delta}\Big)^k \exp(-c(\delta/4)\,m).
\end{aligned}$$

We now make the index set K vary, taking into account that Σ_k is the union of $\binom{N}{k}$ spaces Σ_K to derive

$$\begin{aligned}
\mathbb{P}\Big(\big|\|A\mathbf{x}\|_2^2 - \|\mathbf{x}\|_2^2\big| > \delta\|\mathbf{x}\|_2^2, \quad \text{for some } \mathbf{x} \in \Sigma_k\Big) &\le \binom{N}{k}\cdot 2\Big(3+\frac{8}{\delta}\Big)^k \exp(-c(\delta/4)\,m) \\
&\le 2\Big(\frac{eN}{k}\Big)^k\Big(3+\frac{8}{\delta}\Big)^k \exp(-c(\delta/4)\,m) \le 2\,\exp\Big(k\big[\ln(eN/k) + \ln(3+8/\delta)\big] - c(\delta/4)\,m\Big) \\
&\le 2\,\exp\Big(\big[1+\ln(3+8/\delta)\big]k\ln(eN/k) - c(\delta/4)\,m\Big).
\end{aligned}$$

By taking $[1+\ln(3+8/\delta)]k\ln(eN/k) \le \dfrac{1}{2}c(\delta/4)\,m$, that is

$$m \;\ge\; c_1(\delta)\cdot k\cdot\ln(eN/k), \qquad\qquad c_1(\delta) := \frac{2[1+\ln(3+8/\delta)]}{c(\delta/4)},$$

we finally obtain

$$\mathbb{P}\Big(\big|\|A\mathbf{x}\|_2^2 - \|\mathbf{x}\|_2^2\big| > \delta\|\mathbf{x}\|_2^2, \quad \text{for some } \mathbf{x} \in \Sigma_k\Big) \;\le\; 2\,\exp(-c_0(\delta)\,m), \qquad\qquad c_0(\delta) := \frac{c(\delta/4)}{2},$$

as announced. □

9.2 Subgaussian and strictly subgaussian random variables

In the next section, we will establish that the Concentration Inequality (9.1) holds with constant $c(\varepsilon) = \varepsilon^2/4 - \varepsilon^3/6$ if the entries of the matrix A are (independent identically) distributed strictly subgaussian random variables. Before doing so, we must recall in this section the definitions and basic properties of subgaussian and strictly subgaussian random variables.

Definition 9.3. A random variable ξ is called subgaussian if there is a number $a > 0$ such that

$$\mathbb{E}\left(\exp(\lambda\xi)\right) \leq \exp\left(\frac{a^2\lambda^2}{2}\right), \qquad \text{all } \lambda \in \mathbb{R}.$$

The subgaussian standard $\tau := \tau(\xi)$ is the smallest such number a.

Lemma 9.4. If ξ is a subgaussian random variable, then

$$\mathbb{E}(\xi) = 0, \qquad \mathbb{E}(\xi^2) \leq \tau^2(\xi).$$

Proof. We expand both terms of the previous inequality up to degree 2 in λ to get

$$1 + \mathbb{E}(\xi)\lambda + \frac{\mathbb{E}(\xi^2)}{2}\lambda^2 \leq 1 + \frac{a^2}{2}\lambda^2 + \mathcal{O}(\lambda^3).$$

For small values of λ, this implies that $\mathbb{E}(\xi) = 0$ and that $\mathbb{E}(\xi^2) \leq a^2$. We obtain the required result by taking the minimum over a. □

Definition 9.5. A subgaussian random variable is called strictly subgaussian if

$$\mathbb{E}(\xi^2) = \tau^2(\xi),$$

i.e., setting $\sigma^2 := \mathbb{E}(\xi^2)$, if and only if

$$\mathbb{E}\left((\exp(\lambda\xi)\right) \leq \exp\left(\frac{\sigma^2\lambda^2}{2}\right), \qquad \text{all } \lambda \in \mathbb{R}.$$

Lemma 9.6. If ξ is a strictly subgaussian random variable, then

$$\mathbb{E}(\xi^3) = 0, \qquad \mathbb{E}(\xi^4) \leq 3\,\mathbb{E}(\xi^2)^2.$$

Proof. With $\sigma^2 := \mathbb{E}(\xi^2)$, we expand up to degree 4 in λ to obtain

$$1 + \frac{\sigma^2}{2}\lambda^2 + \frac{\mathbb{E}(\xi^3)}{6}\lambda^3 + \frac{\mathbb{E}(\xi^4)}{24}\lambda^4 \leq 1 + \frac{\sigma^2}{2}\lambda^2 + \frac{\sigma^4}{8}\lambda^4 + \mathcal{O}(\lambda^5).$$

For small values of λ, this implies $\mathbb{E}(\xi^3) = 0$ and $\mathbb{E}(\xi^4)/24 \leq \sigma^4/8$, that is $\mathbb{E}(\xi^4) \leq 3\,\sigma^4$. □

Lemma 9.7. If $\xi_1, \ldots, \xi_n$ are independent subgaussian random variables, then the sum $\xi_1 + \cdots + \xi_n$ is also a subgaussian random variable, and

$$\tau^2(\xi_1 + \cdots + \xi_n) \leq \tau^2(\xi_1) + \cdots + \tau^2(\xi_n).$$

Proof. Using the independence of $\xi_1, \ldots, \xi_n$, we have, for all $\lambda \in \mathbb{R}$,

$$\mathbb{E}\left(\exp(\lambda\xi_1) \cdots \exp(\lambda\xi_n)\right) = \mathbb{E}(\exp(\lambda\xi_1)) \cdots \mathbb{E}(\exp(\lambda\xi_n)).$$

Since the random variables $\xi_1, \ldots, \xi_n$ are subgaussian, this yields, for all $\lambda \in \mathbb{R}$,

$$\begin{aligned} \mathbb{E}\left(\exp(\lambda(\xi_1 + \cdots + \xi_n))\right) &\leq \exp\left(\frac{\tau^2(\xi_1)\lambda^2}{2}\right) \cdots \exp\left(\frac{\tau^2(\xi_n)\lambda^2}{2}\right) \\ &= \exp\left(\frac{[\tau^2(\xi_1) + \cdots + \tau^2(\xi_n)]\lambda^2}{2}\right). \end{aligned}$$

This immediately translates into the required result. □

Lemma 9.8. If $\xi_1, \ldots, \xi_n$ are independent strictly subgaussian random variables, then the sum $\xi_1 + \cdots + \xi_n$ is also a strictly subgaussian random variable.

Proof. We need to show that $\sigma^2\left(\xi_1 + \cdots + \xi_n\right) = \tau^2\left(\xi_1 + \cdots + \xi_n\right)$. The inequality

$$\sigma^2\left(\xi_1 + \cdots + \xi_n\right) \leq \tau^2\left(\xi_1 + \cdots + \xi_n\right)$$

is acquired from Lemmas 9.8 and 9.4. As for the reverse inequality, Lemma 9.8 and the fact that $\xi_1, \ldots, \xi_n$ are strictly subgaussian imply

$$\tau^2(\xi_1 + \cdots + \xi_n) \leq \tau^2(\xi_1) + \cdots + \tau^2(\xi_n) = \sigma^2(\xi_1) + \cdots + \sigma^2(\xi_n).$$

We conclude by remarking that, due to the independence of $\xi_1, \ldots, \xi_n$, we have

$$\sigma^2(\xi_1) + \cdots + \sigma^2(\xi_n) = \sigma^2(\xi_1 + \cdots + \xi_n).$$

□

Lemma 9.9. If ξ is a subgaussian random variable and if $\tau := \tau(\xi) > 0$, then

$$\mathbb{E}\left(\exp\left(\frac{t\xi^2}{2\tau^2}\right)\right) \leq \frac{1}{\sqrt{1-t}}, \qquad 0 \leq t < 1.$$

Proof. We recall that

$$\int_{-\infty}^{\infty} \exp\left(-\pi x^2\right) dx = 1,$$

which we are going to use in the form

$$\int_{-\infty}^{\infty} \exp\left(-ax^2\right) dx = \sqrt{\frac{\pi}{a}}, \qquad a > 0.$$

If F denotes the distribution function of the random variable ξ, that ξ is subgaussian reads

$$\int_{-\infty}^{\infty} \exp(\lambda x) dF(x) \leq \exp\left(\frac{\lambda^2 \tau^2}{2}\right), \qquad \text{all } \lambda \in \mathbb{R}.$$

Since the case $t = 0$ is clear, we can multiply by $\exp(-\lambda^2\tau^2/2t)$ and integrate with respect to λ to get

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left(\lambda x - \frac{\lambda^2\tau^2}{2t}\right) dF(x)\, d\lambda \leq \int_{-\infty}^{\infty} \exp\left(-\frac{\lambda^2\tau^2(1-t)}{2t}\right) d\lambda.$$

It follows that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left(-\left[\frac{\lambda\tau}{\sqrt{2t}} - \sqrt{\frac{t}{2}}\frac{x}{\tau}\right]^2 + \frac{t}{2}\frac{x^2}{\tau^2}\right) d\lambda\, dF(x) \leq \sqrt{\frac{2\pi t}{\tau^2(1-t)}},$$

that is to say

$$\int_{-\infty}^{\infty} \exp\left(\frac{t}{2}\frac{x^2}{\tau^2}\right) \cdot \sqrt{\frac{2\pi t}{\tau^2}} \cdot dF(x) \leq \sqrt{\frac{2\pi t}{\tau^2(1-t)}}.$$

After simplification, this simply reads

$$\mathbb{E}\left(\left(\frac{t\xi^2}{2\tau^2}\right)\right) \leq \sqrt{\frac{1}{1-s}}.$$

□

Proposition 9.10. Zero-mean Gaussian variables are strictly subgaussian.

Proof. Suppose that a random variable ξ follows a normal distribution with variance σ. By symmetry, it is clear that we have

$$\mathbb{E}(\xi^{2k+1}) = 0, \qquad k \text{ nonnegative integer}.$$

On the other hand, we have

$$\begin{aligned}\mathbb{E}(\xi^{2k}) &= \int_{-\infty}^{\infty} x^{2k} \cdot \frac{1}{\sigma\sqrt{2\pi}} \exp\Big(\frac{-x^2}{2\sigma^2}\Big) dx = \frac{2}{\sqrt{\pi}} \int_0^{\infty} x^{2k} \exp\Big(\frac{-x^2}{2\sigma^2}\Big) d\Big(\frac{x}{\sqrt{2}\sigma}\Big) \\ &\underset{t=x^2/(2\sigma^2)}{=} \frac{(2\sigma^2)^k}{\sqrt{\pi}} \int_0^{\infty} t^{k-1/2} \exp(-t) dt = \frac{(2\sigma^2)^k}{\sqrt{\pi}} \Gamma\Big(\frac{2k+1}{2}\Big).\end{aligned}$$

We recall here that the Γ function, defined by

$$\Gamma(z) := \int_0^\infty t^{z-1} \exp(-t) dt, \qquad \Re(z) > 0,$$

has a value at the half-integer $(2k+1)/2$ given by

$$\Gamma\left(\frac{2k+1}{2}\right) = \frac{\sqrt{\pi}}{2^k} (2k-1)!! := \frac{\sqrt{\pi}}{2^k} (2k-1)(2k-3)\cdots 1 = \frac{\sqrt{\pi}}{2^{2k}} \frac{(2k)!}{k!}. \tag{9.3}$$

Thus, we obtain the moment condition

$$\mathbb{E}(\xi^{2k}) = \frac{(2k)!}{2^k\, k!}\, \sigma^{2k}, \qquad k \text{ nonnegative integer}.$$

It finally follows that

$$\mathbb{E}\left(\exp(\lambda\xi)\right) = \sum_{k=0}^\infty \frac{1}{(2k)!} \lambda^{2k} \mathbb{E}(\xi^{2k}) = \sum_{k=0}^\infty \frac{1}{k!} \left(\frac{\lambda^2\sigma^2}{2}\right)^k = \exp\left(\frac{\lambda^2\sigma^2}{2}\right),$$

where the required inequality appears to be an equality in this case. □

Proposition 9.11. Random variables uniformly distributed on $[-1, 1]$ are strictly subgaussian.

Proof. Let ξ be a random variable uniformly distributed on $[-1, 1]$. We first observe that

$$\sigma^2 := \mathbb{E}(\xi^2) = \frac{1}{2} \int_{-1}^1 x^2\, dx = \int_0^1 x^2\, dx = \frac{1}{3},$$

and that

$$\mathbb{E}(\exp(\lambda\xi)) = \frac{1}{2} \int_{-1}^1 \exp(\lambda x) dx = \frac{1}{2\lambda} \left[\exp(\lambda) - \exp(-\lambda)\right] = 1 + \sum_{k=1}^\infty \frac{1}{(2k+1)!} \lambda^{2k}.$$

To prove that the latter is bounded by

$$\exp\left(\frac{\lambda^2}{6}\right) = 1 + \sum_{k=1}^\infty \frac{1}{k!\, 6^k} \lambda^{2k},$$

it is enough to show that $(2k+1)! \geq k!\, 6^k$, $k \geq 1$. This is seen, for $k \geq 1$, from

$$(2k+1)! = \left[(2k+1)\cdot(2k-1)\cdots 3\right] \cdot \left[2k\cdot(2k-2)\cdots 2\right] \geq \left[3^k\right] \cdot \left[2^k\, k!\right] = k!\, 6^k.$$

□

Proposition 9.12. Random variables with the distribution

$$\mathbb{P}(\xi = -1) = \mathbb{P}(\xi = 1) = \frac{1-\mu}{2}, \qquad \mathbb{P}(\xi = 0) = \mu,$$

are strictly subgaussian for and only for $\mu \in [0, 2/3] \cup \{1\}$.

Proof. We observe first that

$$\sigma^2 := \mathbb{E}(\xi^2) = \mu \cdot 0 + \frac{1-\mu}{2} \cdot (-1)^2 + \frac{1-\mu}{2} \cdot (1)^2 = 1 - \mu$$

and that

$$\mathbb{E}(\exp(\lambda\xi)) = \mu \cdot 1 + \frac{1-\mu}{2} \cdot \exp(\lambda) + \frac{1-\mu}{2} \cdot \exp(-\lambda) = \lambda + \sum_{k=0}^{\infty} \frac{\lambda^{2k}}{(2k)!} = 1 + \sum_{k=1}^{\infty} \frac{\lambda^{2k}}{(2k)!}.$$

To prove that the latter is bounded by

$$\exp\left(\frac{(1-\mu)\lambda^2}{2}\right) = 1 + \sum_{k=1}^{\infty} \frac{1}{k!} \frac{(1-\mu)^k}{2^k} \lambda^{2k},$$

it is enough to show — when $\mu \neq 1$ — that $(2k)! \geq k!\, 2^k/(1-\mu)^{k-1}$, $k \geq 1$. As before, this is seen, for $k \geq 1$, from

$$(2k)! = [2k \cdot (2k-2) \cdots 2] \cdot [(2k-1) \cdot (2k-3) \cdots 3] \geq [2^k\, k!] \cdot [3^{k-1}] \geq k!\, 2^k/(1-\mu)^{k-1}$$

as soon as $3 \geq 1/(1-\mu)$, i.e. $\mu \leq 2/3$. If on the other hand we have $\mu \in (2/3, 1)$, the random variable ξ is not strictly subgaussian, because the inequality $\mathbb{E}(\xi^4) \leq 3\sigma^4$ is not satisfied, in view of $\mathbb{E}(\xi^4) = 1 - \mu$. □

9.3 Concentration Inequality for Strictly Subgaussian Random Matrices

We suppose now that the entries $a_{i,j}$ of the $m \times N$ matrix A are independent realizations of subgaussian random variables with standard deviation σ and subgaussian standard τ — not necessarily the same distribution for all entries. Only later will these random variables be assumed to be strictly subgaussian. For $\mathbf{x} \in \mathbb{R}^N$ and $\varepsilon \in (0, 1)$, we shall prove (9.1) in s slightly different form — equivalent after renormalization of A — namely

$$\mathbb{P}\Big(\big|\|A\mathbf{x}\|_2^2 - m\sigma^2\|\mathbf{x}\|_2^2\big| > \varepsilon m\sigma^2\|\mathbf{x}\|_2^2\Big) \leq 2\, \exp\Big(\Big(-\frac{\varepsilon^2}{4} + \frac{\varepsilon^3}{6}\Big)\, m\Big). \tag{9.4}$$

Note that we have

$$\|A\mathbf{x}\|_2^2 = \sum_{i=1}^m X_i^2, \qquad \text{where } X_i := \sum_{j=1}^N a_{i,j}x_j.$$

As a sum of independent subgaussian random variables, the random variable X_i itself is subgaussian. Moreover, by the independence of the $a_{i,j}$'s, we have

$$\mathbb{E}(X_i^2) = \sigma^2\left(\sum_{j=1}^N a_{i,j}x_j\right) = \sum_{j=1}^N x_j^2\sigma^2(a_{i,j}) = \sigma^2\|\mathbf{x}\|_2^2, \qquad \text{so that } \mathbb{E}(\|A\mathbf{x}\|_2^2) = m\sigma^2\|\mathbf{x}\|_2^2.$$

We now set

$$\xi_i := X_i^2 - \mathbb{E}(X_i^2), \qquad \text{so that } \sum_{i=1}^m \xi_i = \|A\mathbf{x}\|_2^2 - m\sigma^2\|\mathbf{x}\|_2^2.$$

We will bound

$$\begin{aligned}
&\mathbb{P}\Big(\big|\|A\mathbf{x}\|_2^2 - m\sigma^2\|\mathbf{x}\|_2^2\big| > \varepsilon m\sigma^2\|\mathbf{x}\|_2^2\Big)\\
&\qquad = \mathbb{P}\Big(\|A\mathbf{x}\|_2^2 - m\sigma^2\|\mathbf{x}\|_2^2 > \varepsilon m\sigma^2\|\mathbf{x}\|_2^2\Big) + \mathbb{P}\Big(\|A\mathbf{x}\|_2^2 - m\sigma^2\|\mathbf{x}\|_2^2 < -\varepsilon m\sigma^2\|\mathbf{x}\|_2^2\Big)\\
&\qquad = \mathbb{P}\left(\sum_{i=1}^m \xi_i > \varepsilon m\sigma^2\|\mathbf{x}\|_2^2\right) + \mathbb{P}\left(\sum_{i=1}^m \xi_i < -\varepsilon m\sigma^2\|\mathbf{x}\|_2^2\right)
\end{aligned}$$

in two steps, one for each of the terms in this sum.

9.3.1 The bound $\mathbb{P}\left(\sum \xi_i > \varepsilon m\sigma^2\|\mathbf{x}\|_2^2\right) \le \exp(-\varepsilon^2/4 + \varepsilon^3/6)$

For any $t > 0$, using Markov inequality, we have

$$\begin{aligned}
\mathbb{P}\left(\sum_{i=1}^m \xi_i > \varepsilon m\sigma^2\|\mathbf{x}\|_2^2\right) &= \mathbb{P}\left(\exp\big(t\sum_{i=1}^m \xi_i\big) > \exp\big(t\,\varepsilon m\sigma^2\|\mathbf{x}\|_2^2\big)\right)\\
&\le \frac{\mathbb{E}\left(\exp\left(t\,\sum_{i=1}^m \xi_i\right)\right)}{\exp\left(t\,\varepsilon m\sigma^2\|\mathbf{x}\|_2^2\right)} = \frac{\mathbb{E}\left(\prod_{i=1}^m \exp(t\,\xi_i)\right)}{\prod_{i=1}^m \exp(t\,\varepsilon\sigma^2\|\mathbf{x}\|_2^2)}.
\end{aligned}$$

By the independence of the random variables $\xi_1, \ldots, \xi_m$, we obtain

$$\begin{aligned}
\mathbb{P}\left(\sum_{i=1}^m \xi_i > \varepsilon m\sigma^2\|\mathbf{x}\|_2^2\right) &\le \prod_{i=1}^m \frac{\mathbb{E}\big(\exp(t\,\xi_i)\big)}{\exp(t\,\varepsilon\sigma^2\|\mathbf{x}\|_2^2)} = \prod_{i=1}^m \frac{\mathbb{E}\big(\exp(t\,X_i^2)\big)\cdot\exp(-t\,\mathbb{E}(X_i^2))}{\exp(t\,\varepsilon\sigma^2\|\mathbf{x}\|_2^2)}\\
&= \prod_{i=1}^m \big\{\mathbb{E}\big(\exp(t\,X_i^2)\big)\cdot\exp(-(1+\varepsilon)t\,\sigma^2\|\mathbf{x}\|_2^2)\big\}\\
&= \prod_{i=1}^m \big\{\mathbb{E}\big(\exp(u\,X_i^2/\sigma^2\|\mathbf{x}\|_2^2)\big)\cdot\exp(-(1+\varepsilon)u)\big\},
\end{aligned}$$

where we have set $u := t\sigma^2\|\mathbf{x}\|_2^2$. We are going to show that, for an appropriate choice of $u > 0$, each term in this product can be bounded by $\exp(-\varepsilon^2/4 + \varepsilon^3/6)$. Hence it will follow that

$$\mathbb{P}\Big(\|A\mathbf{x}\|_2^2 - m\sigma^2\|\mathbf{x}\|_2^2 > \varepsilon m\sigma^2\|\mathbf{x}\|_2^2\Big) \le \exp\left(\Big(-\frac{\varepsilon^2}{4} + \frac{\varepsilon^3}{6}\Big)\, m\right).$$

Thus, we need to establish that

$$\big\{\mathbb{E}\big(\exp(u^* X_i^2/\sigma^2\|\mathbf{x}\|_2^2)\big) \cdot \exp(-(1+\varepsilon)u^*)\big\} \le \exp\Big(-\frac{\varepsilon^2}{4} + \frac{\varepsilon^3}{6}\Big) \qquad \text{for some } u^* > 0. \tag{9.5}$$

Because the subgaussian standard of the random variable X_i satisfies

$$\tau^2(X_i) = \tau^2\left(\sum_{j=1}^N a_{i,j}x_j\right) \le \sum_{j=1}^N x_j^2\tau^2(a_{i,j}) = \tau^2\|\mathbf{x}\|_2^2,$$

we derive, using Lemma 9.9, that

$$\mathbb{E}\big(\exp(u\,X_i^2/\sigma^2\|\mathbf{x}\|_2^2)\big) = \mathbb{E}\left(\left(\frac{2u\tau^2(X_i)}{\sigma^2\|\mathbf{x}\|_2^2}\,\frac{X_i^2}{2\tau^2(X_i)}\right)\right) \le \frac{1}{\sqrt{1 - 2u\tau^2(X_i)/\sigma^2\|\mathbf{x}\|_2^2}} \le \frac{1}{\sqrt{1-2u\tau^2/\sigma^2}},$$

provided that we can actually write this square root. It follows that

$$\big\{\mathbb{E}\big(\exp(u\,X_i^2/\sigma^2\|\mathbf{x}\|_2^2)\big) \cdot \exp(-(1+\varepsilon)u)\big\} \le \frac{\exp(-(1+\varepsilon)u)}{\sqrt{1-2u\tau^2/\sigma^2}}.$$

As a function of $u \in (0,\infty)$, the latter is minimized for

$$u^* = \frac{(\sigma^2/\tau^2 - 1) + (\sigma^2/\tau^2)\varepsilon}{2(1+\varepsilon)}.$$

Making the choice $u = u^*$, we obtain

$$\begin{aligned}
\big\{\mathbb{E}\big(\exp(u^* X_i^2/\sigma^2\|\mathbf{x}\|_2^2)\big) \cdot \exp(-(1+\varepsilon)u^*)\big\} &\le \frac{\exp\big(-\big((\sigma^2/\tau^2-1) + (\sigma^2/\tau^2)\varepsilon\big)/2\big)}{\sqrt{1 - \dfrac{(1-\tau^2/\sigma^2)+\varepsilon}{1+\varepsilon}}} \\
&= \frac{\sigma}{\tau}\sqrt{1+\varepsilon}\,\exp\big(-\big((\sigma^2/\tau^2-1) + (\sigma^2/\tau^2)\varepsilon\big)/2\big) \\
&= \left[\frac{\sigma^2}{\tau^2}(1+\varepsilon)\exp\big(1 - (\sigma^2/\tau^2) - (\sigma^2/\tau^2)\varepsilon\big)\right]^{1/2}.
\end{aligned}$$

We assume at present that the entries of the matrix A are strictly subgaussian, so that $\sigma = \tau$, and the previous inequality becomes

$$\big\{\mathbb{E}\big(\exp(u^* X_i^2/\sigma^2\|\mathbf{x}\|_2^2)\big) \cdot \exp(-(1+\varepsilon)u^*)\big\} \le \left[(1+\varepsilon)\exp(-\varepsilon)\right]^{1/2}.$$

It is easy to verify that

$$(1+\varepsilon)\exp(-\varepsilon) \le 1 - \frac{\varepsilon^2}{2} + \frac{\varepsilon^3}{3}, \qquad \varepsilon \in (0,1).$$

Therefore, we conclude that

$$\begin{aligned}\big\{\mathbb{E}\big(\exp(u^* X_i^2/\sigma^2\|\mathbf{x}\|_2^2)\big)\cdot\exp(-(1+\varepsilon)u^*)\big\} &\le \Big[1-\frac{\varepsilon^2}{2}+\frac{\varepsilon^3}{3}\Big]^{1/2} \le \Big[\exp\Big(-\frac{\varepsilon^2}{2}+\frac{\varepsilon^3}{3}\Big)\Big]^{1/2} \\ &= \exp\Big(-\frac{\varepsilon^2}{4}+\frac{\varepsilon^3}{6}\Big),\end{aligned}$$

as required by (9.5).

9.3.2 The bound $\mathbb{P}\left(\sum \xi_i < -\varepsilon m\sigma^2\|\mathbf{x}\|_2^2\right) \le \exp(-\varepsilon^2/4+\varepsilon^3/6)$

Once again, we call upon Markov inequality to derive, for $t>0$,

$$\begin{aligned}\mathbb{P}\left(\sum_{i=1}^m \xi_i < -\varepsilon m\sigma^2\|\mathbf{x}\|_2^2\right) &= \mathbb{P}\left(\exp\big(-t\sum_{i=1}^m \xi_i\big) > \exp\big(t\,\varepsilon m\sigma^2\|\mathbf{x}\|_2^2\big)\right) \\ &\le \frac{\mathbb{E}\left(\exp\big(-t\sum_{i=1}^m \xi_i\big)\right)}{\exp\big(t\,\varepsilon m\sigma^2\|\mathbf{x}\|_2^2\big)} = \frac{\mathbb{E}\left(\prod_{i=1}^m \exp(-t\,\xi_i)\right)}{\prod_{i=1}^m \exp(t\,\varepsilon\sigma^2\|\mathbf{x}\|_2^2)}.\end{aligned}$$

By the independence of the random variables $\xi_1,\ldots,\xi_m$, we obtain

$$\begin{aligned}\mathbb{P}\left(\sum_{i=1}^m \xi_i < -\varepsilon m\sigma^2\|\mathbf{x}\|_2^2\right) &\le \prod_{i=1}^m \frac{\mathbb{E}\big(\exp(-t\,\xi_i)\big)}{\exp(t\,\varepsilon\sigma^2\|\mathbf{x}\|_2^2)} = \prod_{i=1}^m \frac{\mathbb{E}\big(\exp(-t\,X_i^2)\big)\cdot\exp(t\,\mathbb{E}(X_i^2))}{\exp(t\,\varepsilon\sigma^2\|\mathbf{x}\|_2^2)} \\ &= \prod_{i=1}^m \big\{\mathbb{E}\big(\exp(-t\,X_i^2)\big)\cdot\exp((1-\varepsilon)t\,\sigma^2\|\mathbf{x}\|_2^2)\big\} \\ &= \prod_{i=1}^m \big\{\mathbb{E}\big(\exp(-u\,X_i^2/\sigma^2\|\mathbf{x}\|_2^2)\big)\cdot\exp((1-\varepsilon)u)\big\},\end{aligned}$$

where we have set $u := t\sigma^2\|\mathbf{x}\|_2^2$. Once again, we are going to establish that each term in this product can be bounded by $\exp(-\varepsilon^2/4+\varepsilon^3/6)$ for an appropriate choice of $u>0$. It will follow that

$$\mathbb{P}\Big(\|A\mathbf{x}\|_2^2 - m\sigma^2\|\mathbf{x}\|_2^2 < -\varepsilon m\sigma^2\|\mathbf{x}\|_2^2\Big) \le \exp\Big(\Big(-\frac{\varepsilon^2}{4}+\frac{\varepsilon^3}{6}\Big)\,m\Big).$$

Thus, we need to establish that

$$\text{(9.6)} \qquad \mathbb{E}\big(\exp(-u\,X_i^2/\sigma^2\|\mathbf{x}\|_2^2)\big) \le \exp\Big(-(1-\varepsilon)u - \frac{\varepsilon^2}{4}+\frac{\varepsilon^3}{6}\Big) \qquad \text{for some } u>0.$$

Using Lemma 9.6, we simply write

$$\begin{aligned}\mathbb{E}\big(\exp(-u\,X_i^2/\sigma^2\|\mathbf{x}\|_2^2)\big) &\leq \mathbb{E}\left(1-\frac{u\,X_i^2}{\sigma^2\|\mathbf{x}\|_2^2}+\frac{1}{2}\,\frac{u^2\,X_i^4}{\sigma^4\|\mathbf{x}\|_2^4}\right)=1-\frac{u\,\mathbb{E}(X_i^2)}{\sigma^2\|\mathbf{x}\|_2^2}+\frac{1}{2}\,\frac{u^2\,\mathbb{E}(X_i^4)}{\sigma^4\|\mathbf{x}\|_2^4}\\ &\leq 1-u+\frac{3}{2}\,u^2.\end{aligned}$$

We make the choice $u=\varepsilon/2$, so that it is enough to prove that

$$1-\frac{\varepsilon}{2}+\frac{3\varepsilon^2}{8}\leq\exp\left(-\frac{\varepsilon}{2}+\frac{\varepsilon^2}{4}+\frac{\varepsilon^6}{6}\right),$$

or, setting $\eta:=\varepsilon/2$ to simplify the calculations, that

$$1-\eta+\frac{3}{2}\eta^2\leq\exp\left(-\eta+\eta^2+\frac{4}{3}\eta^3\right),\qquad 0<\eta<\frac{1}{2}.$$

Taking logarithms instead, our objective is to show that the function h defined by

$$h(\eta):=\ln\left(1-\eta+\frac{3}{2}\eta^2\right)+\eta-\eta^2-\frac{4}{3}\eta^3$$

is negative on $(0,1/2)$. In view of $h(0)=0$, it suffices to show that

$$h'(\eta)=\frac{-1+3\eta}{1-\eta+\frac{3}{2}\eta^2}+1-2\eta-4\eta^2\;<\;0,\qquad 0<\eta<\frac{1}{2}.$$

This is equivalent to the inequality

$$1-3\eta>\left(1-\eta+\frac{3}{2}\eta^2\right)(1-2\eta-4\eta^2),\qquad 0<\eta<\frac{1}{2},$$

that is

$$1-3\eta>1-3\eta-\frac{1}{2}\eta^2+\eta^3-6\eta^4,\qquad 0<\eta<\frac{1}{2},$$

which is clearly satisfied.

Exercises

Ex.1: Verify that $\displaystyle\binom{N}{k}\leq\left(\frac{eN}{k}\right)^k$.

Ex.2: Prove an analog of Lemma 9.2 for the ℓ_q-quasinorm.

Ex.3: Make sure that, if $X_1,\ldots,X_n$ are independent $\mathcal{N}(0,1)$ random variables, then the linear combination $\sum_{j=1}^n c_jX_j$ is an $\mathcal{N}(0,\|\mathbf{c}\|_2)$ random variable.

Ex.4: Recall the proof of Markov inequality, which states that for any positive random variable X and any $a>0$, one has

$$\mathbb{P}(X>a)\leq\frac{\mathbb{E}(X)}{a}.$$

Ex.5: Prove the identity (9.3) giving the value of the Γ function at half-integers.

Chapter 10
Stable and Robust Recovery with Mixed Norms

See the following handwritten notes.

STABLE AND ROBUST RECOVERY WITH MIXED NORMS

I/

Consider the problem of recovering a vector $x \in \mathbb{R}^N$ — not necessarily sparse — from the flawed measurement $y \in \mathbb{R}^m$ which approximates the ideal measurement vector $Ax \in \mathbb{R}^m$.

Let θ represent a bound for the relative error between accurate and inaccurate measurements:

$$\|Ax - y\|_2 \le \sqrt{\beta_{2s}} \cdot \theta$$

[Note the invariance under the change $A \leftarrow cA$, $y \leftarrow cy$].

Recall that $\alpha_{2s}, \beta_{2s} > 0$ are the best constants in the inequality

$$\alpha_{2s} \|x\|_2^2 \le \|Ax\|_2^2 \le \beta_{2s} \|x\|_2^2 \qquad \text{for all } 2s\text{-sparse } x \in \mathbb{R}^N$$

We shall try to recover the original vector x by solving

$$(P_{1,\theta}) \quad \underset{z \in \mathbb{R}^N}{\text{minimize}} \ \|z\|_1 \quad \text{subject to} \quad \|Az - y\|_2 \le \sqrt{\beta_{2s}} \cdot \theta.$$

We have the following theorem:

Theorem If one has

$$\gamma_{2s} := \frac{\beta_{2s}}{\alpha_{2s}} < 4\sqrt{2} - 3 \approx 2.6569,$$

then, for any $p \in [1, 2]$, there holds

$$\|x - x^*\|_p \le \frac{C_p}{s^{1-\frac{1}{p}}} \sigma_s(x)_1 + D_p \, s^{\frac{1}{p}-\frac{1}{2}} \theta,$$

with x^* = solution of $(P_{1,\theta})$

II/

Proof: For $p=1$, x s-sparse, and $\theta=0$, this is a result of exact s-sparse recovery by ℓ_1-minimization. Thus, the Null-Space Property should be satisfied, and we are going to prove a little bit more

step 1: consequence of the assumption on γ_{2s}

Let $v \in \mathbb{R}^N$ and $S = S_0$ with $\mathrm{card}(S) \le s$ be arbitrary.

Partition $\overline{S} = [1:N] \setminus S$ as $\overline{S} = S_1 \cup S_2 \cup S_3 \cup \cdots$, with

$S_1 := \{$ indices of s-largest absolute-value entries of $v_{\overline{S}}$ $\}$

$S_2 := \{$ — next ..... $\}$

$\vdots$

* Observe that:

$$\|v_{S_0}\|_2^2 + \|v_{S_1}\|_2^2 = \|v_{S_0} + v_{S_1}\|_2^2 \le \frac{1}{\alpha_{2s}} \|A(v_{S_0} + v_{S_1})\|_2^2$$

$$= \frac{1}{\alpha_{2s}} \langle A(v - v_{S_2} - v_{S_3} - \cdots),\ A(v_{S_0} + v_{S_1}) \rangle$$

$$= \frac{1}{\alpha_{2s}} \langle Av,\ A(v_{S_0} + v_{S_1}) \rangle + \frac{1}{\alpha_{2s}} \sum_{k=1,2} \big[\langle A(-v_{S_k}),\ A(v_{S_0}) \rangle + \langle A(-v_{S_k}),\ A(v_{S_1}) \rangle \big]$$

Renormalize $-v_{S_k}$ and v_{S_0}: $u_k := \dfrac{-v_{S_k}}{\|v_{S_k}\|_2}$, $u_0 := \dfrac{v_{S_0}}{\|v_{S_0}\|_2}$; then:

$$\frac{\langle A(-v_{S_k}),\ A(v_{S_0}) \rangle}{\|v_{S_k}\|_2 \, \|v_{S_0}\|_2} = \langle A u_k,\ A u_0 \rangle = \frac{1}{4} \big[\|A(u_k + u_0)\|_2^2 - \|A(u_k - u_0)\|_2^2 \big]$$

$$\le \frac{1}{4} \big[\beta_{2s} \underbrace{\|u_k + u_0\|_2^2}_{=2} - \alpha_{2s} \underbrace{\|u_k - u_0\|_2^2}_{=2} \big] = \frac{1}{2} \big[\beta_{2s} - \alpha_{2s} \big]$$

The same can be done for S_1 in place of S_0, and we get

$$\langle A(-v_{S_k}),\ A(v_{S_0}) \rangle + \langle A(-v_{S_k}),\ A(v_{S_1}) \rangle \le \frac{\beta_{2s} - \alpha_{2s}}{2} \|v_{S_k}\|_2 \big(\|v_{S_0}\|_2 + \|v_{S_1}\|_2 \big)$$

III/

Besides : $\langle A r, A(r_{S_0}+r_{S_1})\rangle \le \|Ar\|_2 \times (\|Ar_{S_0}\|_2 + \|Ar_{S_1}\|_2)$

$\le \|Ar\|_2 \times \sqrt{\beta_{2\Delta}}\,(\|r_{S_0}\|_2 + \|r_{S_1}\|_2)$

We derive :

$$\|r_{S_0}\|_2^2 + \|r_{S_1}\|_2^2 \le \Big\{ \underbrace{\frac{\gamma_{2\Delta}}{\sqrt{\beta_{2\Delta}}}}_{=:c} \|Ar\|_2 + \underbrace{\frac{\gamma_{2\Delta}-1}{2}}_{=:d} \underbrace{\sum_{k\ge 2}}_{=:\Sigma} \|r_{S_k}\|_2 \Big\} \Big(\|r_{S_0}\|_2 + \|r_{S_1}\|_2 \Big)$$

Complete the square to get :

$$\Big[\|r_{S_0}\|_2 - \frac{c+d\Sigma}{2} \Big]^2 + \Big[\|r_{S_1}\|_2 - \frac{c+d\Sigma}{2} \Big]^2 \le \frac{(c+d\Sigma)^2}{2}.$$

This yields : $\dfrac{\|r_{S_0}\|_2}{\|r_{S_1}\|_2} \le \dfrac{c+d\Sigma}{2} + \dfrac{c+d\Sigma}{\sqrt{2}} = \dfrac{1+\sqrt{2}}{2}(c+d\Sigma)$

* Bound for Σ : for $k \ge 2$, $i \in S_k$, $j \in S_{k-1}$

$|r_i| \le |r_j|$ $\xrightarrow[\text{on } j]{\text{average}}$ $|r_i| \le \frac{1}{\Delta}\|r\|_{S_{k-1}}$ $\xrightarrow[\text{summation on } i]{\text{square and}}$ $\|r_{S_k}\|_2^2 \le \frac{1}{\Delta}\|r_{S_{k-1}}\|_2^2$

Thus : $\Sigma \le \sum_{k\ge 2} \frac{1}{\sqrt{\Delta}} \|r_{S_{k-1}}\|_2$, i.e. $\Sigma \le \frac{1}{\sqrt{\Delta}} \|r_{\bar S}\|_1$

* We use $\|r_S\|_1 \le \sqrt{\Delta}\,\|r_{S_0}\|_2$ to get

$$\|r_S\|_1 \le \frac{\lambda}{2\sqrt{\beta_{2\Delta}}} \sqrt{\Delta}\,\|Ar\|_2 + \mu \|r_{\bar S}\|_1$$

$$\lambda := (1+\sqrt{2})\gamma_{2\Delta} \quad , \quad \mu := \frac{1}{4}(1+\sqrt{2})(\gamma_{2\Delta}-1)$$

Note that $\mu < 1$, since

$\mu < 1 \iff \gamma_{2\Delta} - 1 < \dfrac{4}{\sqrt{2}+1} = 4(\sqrt{2}-1) \iff \gamma_{2\Delta} < 4\sqrt{2}-3$.

step 2: consequence of the ℓ_1-minimization

We now specify: $S = \{\text{indices of } s \text{ largest absolute value entries of } x\}$

$$v = x - x^*$$

We have already seen the following arguments:

$$\|x\|_1 \geq \|x^*\|_1 \quad \text{, i.e.} \quad \|x_S\|_1 + \|x_{\bar{S}}\|_1 \geq \|x^*_S\|_1 + \|x^*_{\bar{S}}\|_1$$

$$\geq \|x_S\|_1 - \|v_S\|_1 + \|v_{\bar{S}}\|_1 - \|x^*_{\bar{S}}\|_1$$

which yields: $\|v_{\bar{S}}\|_1 \leq 2\|x_{\bar{S}}\|_1 + \|v_S\|_1$

$$\underline{\|v_{\bar{S}}\|_1 \leq 2\sigma_s(x)_1 + \|v_S\|_1}$$

step 3: error estimate for $\rho = 1$

Observe that: $\|Av\|_2 \leq \|Ax - y\|_2 + \|Ax^* - y\|_2 \leq \sqrt{\beta_{2s}}\cdot\theta + \sqrt{\beta_{2s}}\cdot\theta = 2\sqrt{\beta_{2s}}\cdot\theta$

The results of steps 1 & 2 read:

$$\|v_S\|_1 \leq \lambda\sqrt{s}\,\theta + \mu\|v_{\bar{S}}\|_1$$

$$\|v_{\bar{S}}\|_1 \leq 2\sigma_s(x)_1 + \|v_S\|_1$$

The following arguments have already been seen:

$$\|v_{\bar{S}}\|_1 \leq 2\sigma_s(x)_1 + \lambda\sqrt{s}\,\theta + \mu\|v_{\bar{S}}\|_1$$

$$\xrightarrow[(\mu<1)]{} \quad \underline{\|v_{\bar{S}}\|_1 \leq \frac{2}{1-\mu}\sigma_s(x)_1 + \frac{\lambda}{1-\mu}\sqrt{s}\,\theta}$$

$$\Rightarrow \quad \|v_S\|_1 \leq \lambda\sqrt{s}\,\theta + \frac{2\mu}{1-\mu}\sigma_s(x)_1 + \frac{\mu\lambda}{1-\mu}\sqrt{s}\,\theta$$

$$= \frac{2\mu}{1-\mu}\sigma_s(x)_1 + \frac{\lambda}{1-\mu}\sqrt{s}\,\theta$$

$$\Rightarrow \underline{\|v\|_1 = \|v_S\|_1 + \|v_{\bar{S}}\|_1 \leq \frac{2(1+\mu)}{1-\mu}\sigma_s(x)_1 + \frac{(2+\mu)\lambda}{1-\mu}\sqrt{s}\,\theta}$$

V/

Step 4 : error estimate for any $p \in [1,2]$

We write : $\|v\|_p = \left[\sum_{k\geq 0} \|v_{S_k}\|_p^p\right]^{1/p} \leq \left[\sum_{k\geq 0} \left(\Delta^{\frac{1}{p}-\frac{1}{2}} \|v_{S_k}\|_2\right)^p\right]^{1/p}$

$$\leq \sum_{k\geq 0} \Delta^{\frac{1}{p}-\frac{1}{2}} \|v_{S_k}\|_2$$

so that : $\Delta^{\frac{1}{2}-\frac{1}{p}} \|v\|_p \leq \|v_{S_0}\|_2 + \|v_{S_1}\|_2 + \sum_{k\geq 2} \|v_{S_k}\|_2$

$$\leq 2\left[\lambda\theta + \frac{1}{\sqrt{\Delta}}\|v_S\|_2\right] + \sum \leq 2\lambda\theta + \frac{1}{\sqrt{\Delta}}\|v_{\bar{S}}\|_1$$

i.e. $\Delta^{\frac{1}{2}-\frac{1}{p}} \|v\|_p \leq \frac{3\lambda}{1-\mu}\theta + \frac{2\mu+1}{1-\mu}\Delta^{-\frac{1}{2}}\sigma_\Delta(a)_1$ $\quad 2\lambda\theta + \frac{2\mu+1}{\sqrt{\Delta}}\left(\frac{2}{1-\mu}\sigma_\Delta(a)_1 + \frac{1}{1-\mu}\sqrt{\Delta}\,\varepsilon\right)$

$$\boxed{\|v\|_p \leq \frac{2\mu+1}{1-\mu}\,\frac{1}{\Delta^{1-\frac{1}{p}}}\,\sigma_\Delta(a)_1 + \frac{3\lambda}{1-\mu}\Delta^{\frac{1}{p}-\frac{1}{2}}\theta}$$

□

Remark In terms of the Restricted Isometry Constant $\delta_{2\Delta} = \frac{\gamma_{2\Delta}-1}{\gamma_{2\Delta}+1}$,

the sufficient condition of the theorem reads

$$\delta_{2\Delta} < \frac{2(3-\sqrt{2})}{7} \approx 0.4531.$$

Chapter 11
Widths

11.1 Definitions and Basic Properties

Definition 11.1. Let X be a normed space and let C be a subset of X.
The Kolmogorov n-width of C in X is defined by

$$d_n(C,X) := \inf\big\{ \sup_{\mathbf{x}\in C} \inf_{\mathbf{y}\in X_n} \|\mathbf{x}-\mathbf{y}\|,\ X_n \text{ is a subspace of } X \text{ with } \dim X_n \le n\big\}.$$

The Gel'fand n-width of C in X is defined by

$$d^n(C,X) := \inf\big\{ \sup_{\mathbf{x}\in C\cap L^n} \|\mathbf{x}\|,\ L^n \text{ is a subspace of } X \text{ with } \operatorname{codim} L^n \le n\big\}.$$

Theorem 11.2 (Duality)**.** If U and V two finite-dimensional normed spaces with $U \subseteq V$, then

$$d_n(B_U, V) = d^n(B_{V^*}, U^*).$$

The proof is based on the following lemma.

Lemma 11.3. Let Y be a finite-dimensional subspace of a normed space X. For $\mathbf{x} \in X$ and $\mathbf{y}^\star \in Y$, one has

$$[\mathbf{y}^\star \text{ is a best approximation to } \mathbf{x} \text{ from } Y] \iff [\exists\, \lambda \in B_{X^*} :\ \lambda_{|Y} = 0 \text{ and } \lambda(\mathbf{x}) = \|\mathbf{x}-\mathbf{y}^\star\|].$$

In particular, one obtains

$$\inf_{\mathbf{y}\in Y} \|\mathbf{x}-\mathbf{y}\| = \sup_{\lambda\in B_{X^*},\, \lambda_{|Y}=0} \lambda(\mathbf{x}).$$

Proof. On the one hand, let us assume that $\|\mathbf{x}-\mathbf{y}^\star\| = \lambda(\mathbf{x})$ for some $\lambda \in B_{X^*}$ satisfying $\lambda_{|Y} = 0$. We have, for any $\mathbf{y} \in Y$,

$$\|\mathbf{x}-\mathbf{y}\| \ge \lambda(\mathbf{x}-\mathbf{y}) = \lambda(\mathbf{x}) = \|\mathbf{x}-\mathbf{y}^\star\|.$$

This proves that $\mathbf{y}^\star$ is a best approximation to $\mathbf{x}$ from Y.
On the other hand, let us assume that $\mathbf{y}^\star$ is a best approximation to $\mathbf{x}$ from Y. We define a linear functional $\widetilde{\lambda}$ on $[Y \oplus \mathrm{span}(\mathbf{x})]$ by

$$\widetilde{\lambda}(\mathbf{y} + t\mathbf{x}) = t \, \|\mathbf{x} - \mathbf{y}^\star\| \qquad \text{for all } \mathbf{y} \in Y \text{ and } t \in \mathbb{R}.$$

It is readily checked that $\widetilde{\lambda}_{|Y} = 0$, and that $\|\widetilde{\lambda}\| \leq 1$, since

$$|\widetilde{\lambda}(\mathbf{y} + t\mathbf{x})| = |t| \, \|\mathbf{x} - \mathbf{y}^\star\| \leq |t| \, \|\mathbf{x} - (-1/t)\mathbf{y}\| = \|\mathbf{y} + t\mathbf{x}\|.$$

Using the Hahn–Banach theorem, we finally extend the linear functional $\widetilde{\lambda}$ to the whole space X while preserving its norm. This gives rise to the required linear functional λ. □

Proof of Theorem 11.2. Left as an exercise. □

11.2 Relation to Compressed Sensing

Definition 11.4. Let X be a normed space and let C be a subset of X. We define

$$E_m(C, X) = \inf \left\{ \sup_{\mathbf{x} \in C} \|\mathbf{x} - g(f(x))\|, \ f \text{ linear map from } X \text{ to } \mathbb{R}^m, \, g \text{ map from } \mathbb{R}^m \text{ to } X \right\}.$$

Theorem 11.5. If a subset C of normed space X satisfies $-C = C$ and $C + C \subseteq \mathrm{cst} \cdot C$, then one has

$$d^m(C, X) \leq E_m(C, X) \leq \mathrm{cst} \cdot d^m(C, X).$$

Proof. Consider first a linear map $f : X \to \mathbb{R}^m$ and a map $g : \mathbb{R}^m \to X$. We introduce the subspace $L^m := \ker f$ of X of codimension $\leq m$. By definition of the Gel'fand width, we have

$$d^m(C, X) \leq \sup_{\mathbf{v} \in C \cap \ker f} \|\mathbf{v}\|.$$

For any $\mathbf{v} \in C \cap \ker f$, we have

$$\begin{aligned} \|\mathbf{v}\| &\leq \frac{1}{2}\|\mathbf{v} - g(0)\| + \frac{1}{2}\|-\mathbf{v} - g(0)\| \\ &\leq \frac{1}{2} \sup_{\mathbf{x} \in C} \|\mathbf{x} - g(f(\mathbf{x}))\| + \frac{1}{2} \sup_{\mathbf{x} \in C} \|\mathbf{x} - g(f(\mathbf{x}))\| = \sup_{\mathbf{x} \in C} \|\mathbf{x} - g(f(\mathbf{x}))\|. \end{aligned}$$

We derive that

$$d^m(C, X) \leq \sup_{\mathbf{x} \in C} \|\mathbf{x} - g(f(\mathbf{x}))\|.$$

The inequality

$$d^m(C,X) \le E_m(C,X)$$

now follows by taking the infimum over f and g.
For the other inequality, we consider a subspace L^m of X of codimension $\le m$. We choose a linear map $f : X \to \mathbb{R}^m$ such that $\ker f = L^m$. We then define the map

$$g : \mathbf{y} \in \mathbb{R}^m \mapsto \begin{cases} \text{any } \mathbf{z} \in C \text{ such that } f(\mathbf{z}) = \mathbf{y} & \text{if } \mathbf{y} \in f(C), \\ \text{anything} & \text{if } \mathbf{y} \notin f(C). \end{cases}$$

We derive

$$\begin{aligned} E_m(C,X) &\le \sup_{\mathbf{x}\in C} \|\mathbf{x} - g(f(\mathbf{x}))\| \le \sup_{\mathbf{x}\in C} \sup_{\mathbf{z}\in C,\, f(\mathbf{z})=f(\mathbf{x})} \|\mathbf{x}-\mathbf{z}\| \le \sup_{\mathbf{v}\in L^m,\, \mathbf{v}\in C-C} \|\mathbf{v}\| \\ &\le \mathrm{cst} \cdot \sup_{\mathbf{w}\in C\cap L^m} \|\mathbf{w}\|. \end{aligned}$$

The inequality

$$E_m(C,X) \le \mathrm{cst} \cdot d^m(C,X)$$

now follows by taking the infimum over L^m. □

11.3 Upper Estimate for $d^m(B_1^N, \ell_p^N)$

Using Compressed Sensing tools, it is possible to establish, in a simple way, the upper bound for the Gel'fand width of the ℓ_1-ball in ℓ_p^N, $p \in [1,2]$. Here is the main result.

Theorem 11.6. Given $N > m$ and $1 \le p \le 2$, one has

$$d^m(B_1^N, \ell_p^N) \le \min\left(1, \left[\mathrm{cst}\frac{\log(\mathrm{cst}'\, N/m)}{m}\right]^{1-1/p}\right).$$

Proof. Let us first of all remark that the inequality $d^m(B_1^N, \ell_p^N) \le 1$ is clear. Indeed, for any $\mathbf{x} \in B_1^N$ — a fortiori for any $\mathbf{x} \in B_1^N \cap L^m$ where L^m is a subspace of ℓ_p^N of codimension at most m — we have $\|\mathbf{x}\|_p \le \|\mathbf{x}\|_1 = 1$. We shall now give two justifications of the upper bound containing the log factor: the first one is less involved, but invokes more Compressed Sensing results, while the second one only uses the Restricted Isometry Property for random matrices.
First justification: in the previous chapter, we have seen that there exists a constant c such that, for any $\mathbf{x} \in \mathbb{R}^N$, if $\mathbf{x}^\star$ represents a minimizer of $\|\mathbf{z}\|_1$ subject to $A\mathbf{z} = A\mathbf{x}$, we have

$$\|\mathbf{x} - \mathbf{x}^\star\|_p \le \frac{c}{s^{1-1/p}}\, \sigma_s(\mathbf{x})_1, \qquad p \in [1,2], \tag{11.1}$$

provided that $\gamma_{2s} < 4\sqrt{2} - 3$. We have also seen that this is satisfied for random matrices as soon as $m \geq \mathrm{cst}\, s\, \ln(\mathrm{cst}'\, N/s)$, or equivalently as soon as $m \geq \mathrm{cst}\, s\, \ln(\mathrm{cst}'\, N/m)$. In view of $\sigma_s(\mathbf{x})_1 \leq \|\mathbf{x}\|_1$, Inequality (11.1) yields

$$E_m(B_1^N, \ell_p^N) \leq \frac{\mathrm{cst}}{s^{1-1/p}}, \qquad p \in [1,2],$$

as soon as $s \leq \mathrm{cst}\, m/\ln(\mathrm{cst}'\, N/m)$. We also know, according to Theorem 11.5, that

$$d^m(B_1^N, \ell_p^N) \leq E_m(B_1^N, \ell_p^N).$$

We can therefore conclude that

$$d^m(B_1^N, \ell_p^N) \leq \left[\mathrm{cst}\frac{\log(\mathrm{cst}'\, N/m)}{m}\right]^{1-1/p}.$$

Second justification: Fix a number $\gamma > 1$ and pick an $m \times N$ random matrix for which $\gamma_s(A) \leq \gamma$. This can be done for $s \leq \mathrm{cst}\, m/\log(\mathrm{cst}'\, N/m)$. Next, consider the subspace $L^m := \ker A$ of ℓ_p^N of codimension at most m. It is enough to establish that

$$\|\mathbf{x}\|_p \leq \left[\mathrm{cst}\frac{\log(\mathrm{cst}'\, N/m)}{m}\right]^{1-1/p} \qquad \text{for all } \mathbf{x} \in B_1^N \cap L^m.$$

So let us consider $\mathbf{x} \in \mathbb{R}^N$ with $\|\mathbf{x}\|_1 \leq 1$ and $A\mathbf{x} = 0$. Partitioning $[1:N]$ as $S_0 \cup S_1 \cup S_2 \cup \cdots$ with $|x_i| \geq |x_j|$ for all $i \in S_{k-1}$, $j \in S_k$, and $k \geq 1$, it has become usual to derive the inequality $\|\mathbf{x}_{S_k}\|_2 \leq \|\mathbf{x}_{S_{k-1}}\|_1/\sqrt{s}$. We then write

$$\begin{aligned}
\|\mathbf{x}\|_p &\leq \|\mathbf{x}_{S_0}\|_p + \|\mathbf{x}_{S_1}\|_p + \|\mathbf{x}_{S_2}\|_p + \cdots \leq s^{1/p-1/2}\big[\|\mathbf{x}_{S_0}\|_2 + \|\mathbf{x}_{S_1}\|_2 + \|\mathbf{x}_{S_2}\|_2 + \cdots\big] \\
&\leq \frac{s^{1/p-1/2}}{\sqrt{\alpha_s}}\big[\|A\mathbf{x}_{S_0}\|_2 + \|A\mathbf{x}_{S_1}\|_2 + \|A\mathbf{x}_{S_2}\|_2 + \cdots\big] \\
&= \frac{s^{1/p-1/2}}{\sqrt{\alpha_s}}\big[\|A(-\mathbf{x}_{S_1} - \mathbf{x}_{S_2} - \cdots)\|_2 + \|A\mathbf{x}_{S_1}\|_2 + \|A\mathbf{x}_{S_2}\|_2 + \cdots\big] \\
&\leq \frac{2\, s^{1/p-1/2}}{\sqrt{\alpha_s}} \sum_{k\geq 1} \|A\mathbf{x}_{S_k}\|_2 \leq \frac{2\sqrt{\beta_s}\, s^{1/p-1/2}}{\sqrt{\alpha_s}} \sum_{k\geq 1} \|\mathbf{x}_{S_k}\|_2 \leq \frac{2\sqrt{\gamma_s}}{s^{1-1/p}} \sum_{k\geq 1} \|\mathbf{x}_{S_{k-1}}\|_1 \\
&\leq \frac{2\sqrt{\gamma}}{s^{1-1/p}}\|\mathbf{x}\|_1 \leq \left[\mathrm{cst}\, \frac{\log(\mathrm{cst}'\, N/m)}{m}\right]^{1-1/p},
\end{aligned}$$

which is the required result. □

11.4 Lower Estimate for $d^m(B_1^N, \ell_p^N)$

We establish in this section a lower bound for $d^m(B_1^N, \ell_p^N)$, or equivalently for $d_m(B_{p^*}^N, \ell_\infty^N)$, whose order matches the order of the upper bound presented in Section 11.3. The corollary that follows is of utmost importance for Compressed Sensing.

Theorem 11.7. Given $N > m$ and $2 \le p \le \infty$, one has

$$d_m(B_p^N, \ell_\infty^N) \ge \frac{1}{4} \min\left(1, \left[\mathrm{cst}\frac{\ln(\mathrm{cst}' N/m)}{m}\right]^{1/p}\right).$$

Corollary 11.8. Suppose that there exist a linear measurement map $f : \mathbb{R}^N \to \mathbb{R}^m$ and a reconstruction map $\mathbb{R}^m \to \mathbb{R}^N$ such that, for some integer $s \ge 1$ and some $1 < p \le 2$, there holds

$$\|\mathbf{x} - g(f(x))\| \le \frac{\mathrm{cst}}{s^{1-1/p}} \sigma_s(\mathbf{x})_1 \qquad \text{for all } \mathbf{x} \in \mathbb{R}^N.$$

Then we necessarily have

$$m \ge \mathrm{cst}\, s\, \ln(\mathrm{cst}' N/s).$$

Proof. In view of $\sigma_s(\mathbf{x})_1 \le \|\mathbf{x}\|_1$, the assumption implies that

$$E_m(B_1^N, \ell_p^N) \le \frac{\mathrm{cst}}{s^{1-1/p}}.$$

We also know that

$$E_m(B_1^N, \ell_p^N) \ge d^m(B_1^N, \ell_p^N) = d_m(B_{p^*}^N, \ell_\infty^N).$$

Therefore, according to Theorem 11.7, we obtain

$$\frac{\mathrm{cst}}{s^{1-1/p}} \ge \left[\mathrm{cst}\frac{\ln(\mathrm{cst}' N/m)}{m}\right]^{1-1/p},$$

so that

$$m \ge \mathrm{cst}\, s\, \ln(\mathrm{cst}' N/m).$$

We now take into account that $t \ln t \ge -1/e$ for all $t \in [0,1]$, so that

$$m \ge \mathrm{cst}\, s\, \ln(\mathrm{cst}' N/s) + \mathrm{cst}\, m\, (s/m)\, \ln(\mathrm{cst}' s/m) \ge \mathrm{cst}\, s\, \ln(\mathrm{cst}' N/s) - (\mathrm{cst}/e)\, m.$$

We can therefore conclude that

$$m \ge \frac{\mathrm{cst}}{1 + \mathrm{cst}/e}\, s\, \ln(\mathrm{cst}' N/s).$$

□

For the proof of the Theorem 11.7, we need the following lemmas.

Lemma 11.9. If U and V are two finite-dimensional subspaces of a normed space X with $\dim V > \dim U$, then there exists a vector $\mathbf{v} \in V \setminus \{0\}$ for which the zero vector is a best approximation to $\mathbf{v}$ from U, i.e. for which

$$\|\mathbf{v}\| \le \|\mathbf{v} - \mathbf{u}\| \qquad \text{for all } \mathbf{u} \in U.$$

Proof. We can equip the finite-dimensional space $U \oplus V$ with a Euclidean norm $|\cdot|$. Then, for any $n \geq 1$, the norm $\|\cdot\|_n := \|\cdot\| + |\cdot|/n$ is strictly convex. This allows to define, for each $\mathbf{v} \in V$, a unique best approximation $P_U^n(\mathbf{v})$ to $\mathbf{v}$ from U with respect to the norm $\|\cdot\|_n$. The map $P_U^n : \mathcal{S}_V \to U$ is continuous [unique best approximations vary continuously] and antipodal. Furthermore, we have $\dim V > \dim U$. Borsuk–Ulam theorem then implies the existence of $\mathbf{v}_n \in V$ with $\|\mathbf{v}_n\| = 1$ such that $P_U^n(\mathbf{v}_n) = 0$, which means

$$\|\mathbf{v}_n\|_n \leq \|\mathbf{v}_n - \mathbf{u}\|_n \qquad \text{for all } \mathbf{u} \in U.$$

Note that we can extract from the sequence $(\mathbf{v}_n)$ a subsequence $(\mathbf{v}_{n_k})$ converging to some $\mathbf{v} \in V$. The previous inequality, written for $n = n_k$, finally passes to the limit as $k \to \infty$ to give

$$\|\mathbf{v}\| \leq \|\mathbf{v} - \mathbf{u}\| \qquad \text{for all } \mathbf{u} \in U,$$

as required. □

Lemma 11.10. For $1 \leq p \leq \infty$ and $m < N$, we have

$$d_m(B_p^N, \ell_\infty^N) \geq \frac{1}{(m+1)^{1/p}}.$$

Proof. We need to prove that for any subspace X_m of ℓ_∞^N with $\dim X_m \leq m$, we have

$$\sup_{\mathbf{x} \in B_p^N} \inf_{\mathbf{y} \in X_m} \|\mathbf{x} - \mathbf{y}\|_\infty \geq \frac{1}{(m+1)^{1/p}}.$$

Because X_m and ℓ_∞^{m+1} are finite-dimensional subspaces of ℓ_∞^N with $\dim \ell_\infty^{m+1} > \dim X_m$, Lemma 11.9 implies the existence of $\mathbf{v} \in \ell_\infty^{m+1} \setminus \{0\}$ such that

$$\inf_{\mathbf{y} \in X_m} \|\mathbf{v} - \mathbf{y}\|_\infty = \|\mathbf{v}\|_\infty.$$

Because $\|\mathbf{v}\|_p \leq (m+1)^{1/p} \|\mathbf{v}\|_\infty$, we obtain

$$\sup_{\mathbf{x} \in B_p^N} \inf_{\mathbf{y} \in X_m} \|\mathbf{x} - \mathbf{y}\|_\infty \geq \inf_{\mathbf{y} \in X_m} \left\| \frac{\mathbf{v}}{\|\mathbf{v}\|_p} - \mathbf{y} \right\|_\infty = \left\| \frac{\mathbf{v}}{\|\mathbf{v}\|_p} \right\|_\infty \geq \frac{1}{(m+1)^{1/p}},$$

as required. □

Lemma 11.11. Let C be a subset of a normed space X. For all $\varepsilon > d_m(C, X)$ and all $t > 0$, the ε-covering number of the set $C \cap tB_X$ in X satisfies

$$N(\varepsilon, C \cap tB_X, X) \leq \left(1 + 2\frac{t + d_m(C, X)}{\varepsilon - d_m(C, X)}\right)^m.$$

Proof. Let X_m be a subspace of X with $\dim X_m \leq m$ and with

$$\sup_{\mathbf{x}\in C} \inf_{\mathbf{y}\in X_m} \|\mathbf{x}-\mathbf{y}\| =: \alpha < \varepsilon.$$

Let $\mathcal{U}$ be a minimal $(\varepsilon-\alpha)$-net of $X_m \cap (t+\alpha)B_X$. We know that

$$\mathrm{card}(\mathcal{U}) \leq \left(1+\frac{2(t+\alpha)}{\varepsilon-\alpha}\right)^m.$$

We claim that $\mathcal{U}$ is also an ε-net of $C\cap tB_X$. Let indeed $\mathbf{x}\in C\cap tB_X$. We can find $\mathbf{y}\in X_m$ with $\|\mathbf{x}-\mathbf{y}\|\leq\alpha$. Therefore, we have $\mathbf{y}\in X_m\cap(t+\alpha)B_X$, since $\|\mathbf{y}\|\leq\|\mathbf{x}\|+\|\mathbf{x}-\mathbf{y}\|\leq t+\alpha$. Thus, we can find $\mathbf{z}\in\mathcal{U}$ with $\|\mathbf{y}-\mathbf{z}\|\leq\varepsilon-\alpha$. It follows that

$$\|\mathbf{x}-\mathbf{z}\|\leq\|\mathbf{x}-\mathbf{y}\|+\|\mathbf{y}-\mathbf{z}\|\leq\alpha+(\varepsilon-\alpha)=\varepsilon.$$

This shows that $\mathcal{U}$ is am ε-net for $C\cap tB_X$. We obtain

$$N(\varepsilon, C\cap tB_X, X)\leq\left(1+2\frac{t+\alpha}{\varepsilon-\alpha}\right)^m.$$

The result follows by letting α go to $d_m(C,X)$. □

Lemma 11.12. Given $1\leq p\leq\infty$, $1/N^{1/p}<2\varepsilon<1$, and $t\geq 2^{1+1/p}\varepsilon$, we have

$$N(\varepsilon, B_p^N\cap tB_\infty^N, \ell_\infty^N)\geq\left(2^{p+1}\varepsilon^p N\right)^{1/(2^{p+1}\varepsilon^p)}.$$

Proof. Note that it is enough to find a subset $\mathcal{S}$ of $B_p^N\cap tB_\infty^N$ of cardinality $\mathrm{card}(\mathcal{S})\geq\left(2^{p+1}\varepsilon^pN\right)^{1/(2^{p+1}\varepsilon^p)}$ in which every two points are separated by a distance larger than 2ε. Indeed, if $\{\mathbf{u}_1,\ldots,\mathbf{u}_n\}$ is an ε-net of $B_p^N\cap tB_\infty^N$, then each ball $B(\mathbf{u}_i,\varepsilon)$ contains at most one element of $\mathcal{S}$, so that $n\geq\mathrm{card}(\mathcal{S})$. Since $2\varepsilon<1$, we consider the largest integer $k\geq 1$ smaller than $1/(2\varepsilon)^p$. Because the integer $2k$ is larger than k, we can deduce

$$k<\frac{1}{(2\varepsilon)^p}\qquad\text{and}\qquad k\geq\frac{1}{2(2\varepsilon)^p}.$$

Note also that $k\leq N$, so we can consider the set

$$\mathcal{S}:=\left\{\mathbf{x}\in\mathbb{R}^N : \forall\, j\in[1:N], x_j\in\{-1/k^{1/p},0,1/k^{1/p}\},\ \|\mathbf{x}\|_0^0=k\right\}.$$

This is a suitable set, because for distinct $\mathbf{x},\mathbf{x}'\in\mathcal{S}$, we have

$$\|\mathbf{x}-\mathbf{x}'\|_\infty\geq 1/k^{1/p}>2\varepsilon,$$

because

$$\mathcal{S}\subseteq B_p^N\cap(1/k^{1/p})B_\infty^N\subseteq B_p^N\cap 2^{1+1/p}\varepsilon\subseteq B_p^N\cap tB_\infty^N,$$

and because

$$\mathrm{card}(\mathcal{S}) = 2^k \binom{N}{k} \geq 2^k \left(\frac{N}{k}\right)^k = \left(\frac{2N}{k}\right)^k > \left(\underbrace{2^{p+1}\varepsilon^p N}_{\geq 1}\right)^k > \left(2^{p+1}\varepsilon^p N\right)^{1/(2^{p+1}\varepsilon^p)}.$$

□

Proof of Theorem 11.7. We choose $\varepsilon = 2\, d_m(B_p^N, \ell_\infty^N)$ and $t = 2^{1+1/p}\varepsilon$. Thus, provided that $d_m(B_p^N, \ell_\infty^N) < 1/4$, the conditions of Lemmas 11.11 and 11.12 are fulfilled. Therefore, we obtain

$$\left(2^{p+1}\varepsilon^p N\right)^{1/(2^{p+1}\varepsilon^p)} \leq N(\varepsilon, B_p^N \cap tB_\infty^N, \ell_\infty^N) \leq \left(1 + 2\frac{2^{1+1/p}\varepsilon + \varepsilon/2}{\varepsilon/2}\right)^m \leq 19^m.$$

Taking the logarithm yields, in view of the definition of ε,

$$\frac{1}{2^{2p+1}\, d_m(B_p^N, \ell_\infty^N)^p}\, \ln\left(2^{2p+1}\, d_m(B_p^N, \ell_\infty^N)^p\, N\right) \leq m\, \ln(19).$$

Now, using Lemma 11.10, we derive

$$\frac{1}{2^{2p+1}\, d_m(B_p^N, \ell_\infty^N)^p}\, \ln\left(2^{2p+1}\, \frac{N}{m+1}\right) \leq m\, \ln(19),$$

and finally

$$d_m(B_p^N, \ell_\infty^N) \geq \frac{1}{2^{2+1/p}\ln(19)^{1/p}} \left[\frac{\ln\left(2^{2p+1}\, N/(m+1)\right)}{m}\right]^{1/p}$$

The restriction $p \geq 2$ now easily implies the required form for the lower bound. □

Exercises

Ex.1: Determine the Gel'fand 1-width $d^1(B_1^2, \ell_2^2)$ of the unit ℓ_1-ball of $\mathbb{R}^2$ when considered as a subspace of $\mathbb{R}^2$ endowed with the ℓ_2-norm.

Ex.2: Given an integer $n \geq 0$, given a real number α, and given subsets B, C of a normed linear space X with $B \subseteq C$, prove that

$$\begin{aligned} d_n(\alpha\, C) &= |\alpha|\, d_n(C), & \qquad d^n(\alpha\, C) &= |\alpha|\, d^n(C), \\ d_n(B) &\leq d_n(C), & \qquad d^n(B) &\leq d^n(C), \\ d_{n+1}(C) &\leq d_n(C), & \qquad d^{n+1}(C) &\leq d^n(C). \end{aligned}$$

Ex.3: Given a normed space X of dimension greater than n, prove that

$$d_k(S_X, X) = 1, \quad k \in [0 : n].$$

Ex.4: Given an $(n+1)$-dimensional subspace X_{n+1} of a normed space X, prove that

$$d_k(S_{X_{n+1}}, X) = 1, \quad k \in [0 : n].$$

Ex.5: Check the equivalence, for $k \le m$, between

$$[m \ge \mathrm{cst}\, k\, \ln(\mathrm{cst}'\, N/k)] \qquad \text{and} \qquad [m \ge \mathrm{cst}\, k\, \ln(\mathrm{cst}'\, N/m)].$$

Ex.6: Fill in the details of the proof of Lemma 11.9.

Chapter 12
Using Expander Graphs

See the following handwritten notes.

THE USE OF EXPANDER GRAPHS

I/ Definitions, preparatory lemmas

A graph with n vertices is called a <u>c-expander</u> if for every set S of $\leq \frac{n}{2}$ vertices, there are at least $c|S|$ edges between S and $\overline{S}$.

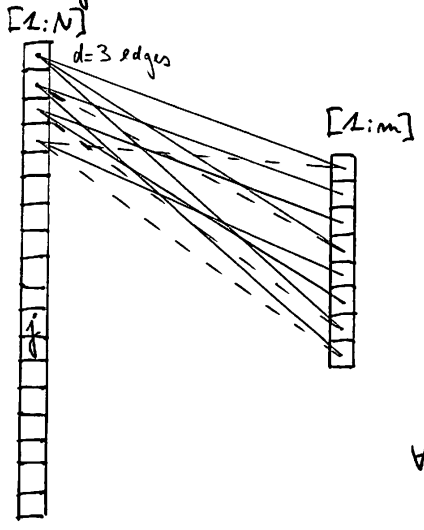
We will in fact consider bipartite expanders, where the edges join "vertices on the left" to "vertices on the right". More precisely, we consider <u>(k, ε)-unbalanced expanders with left degree d</u> for which

- exactly d edges emanate from each left vertex
- for each subset K of left vertices of cardinality $|K| \leq k$, the number of right vertices joined to K is $\geq (1-\varepsilon)\, d\, |K|$

[almost the same as if the right neighbors of K were all distinct]

Rather surprisingly, this is possible with more left vertices than right vertices

We identify the left vertices with $[1:N]$ and the right vertices with $[1:m]$



Notations:

E: set of all edges

for j left vertex, $R(j)$ set of right vertices joined to j, i.e.

$$R(j) = \{ i \in [1:m] : (j \to i) \in E \}$$

$\mathrm{Card}(R(j)) = d$, all $j \in [1:N]$

for $J \subseteq [1:N]$, $R(J) = \bigcup_{j \in J} R(j)$

$\forall |J| \leq k$, $|R(J)| \geq (1-\varepsilon)\, d\, |J|$

$E(J) = \{ (j \to i) \in E \text{ with } j \in J \}$

For $i \in [1:m]$, let $(\ell(i) \to i) \in E$ be the "first" edge coming to i, 2/

that is $\ell(i) = \min \{ j \in [1:N] : (j \to i) \in E \}$

We decompose $E(J)$ in two as

$$E'(J) = \{ (\ell(i) \to i),\ i \in R(K) \} \quad , \quad E''(J) = E(J) \setminus E'(J) \ ,$$

with the idea that, since the right neighbors of J are almost all distinct, $E''(J)$ should be small

Lemma 1: If $|K| \le k$, then $\mathrm{Card}(E''(K)) \le \varepsilon d |K|$

Pf We have $d|K| = \mathrm{Card}(E(K)) = \mathrm{Card}(E'(K)) + \mathrm{Card}(E''(K))$

Note that $\mathrm{Card}(E'(K)) = \mathrm{Card}\, R(K) \ge (1-\varepsilon)\, d|K|$.

Hence: $\mathrm{Card}(E''(K)) \le d|K| - (1-\varepsilon) d|K| = \varepsilon d |K|$ □

We note also the following:

Lemma 2: If K is of the form $[1:k]$, then $E(K) \cap E''(S) = E''(K)$ and S of the form $[1:s]$ with $s \ge k$

Pf $e \in E(K) \cap E''(S) \iff e = (j \to i)$ with $j \le k$ & $f(i) < j \le s$

$\iff e = (j \to i)$ with $f(i) < j \le k \iff e \in E''(K)$

2/ Null-Space Property for the Adjacency Matrix

The adjacency matrix A of the expander graph is the $m \times N$ matrix whose entries are given by

$$A_{ij} = \begin{cases} 1 & \text{if } (j \to i) \in E, \\ 0 & \text{otherwise.} \end{cases}$$

Theorem The adjacency matrix of a $(2s,\varepsilon)$-unbalanced expander with left degree d satisfies the s-th order Null-Space Property provided that $\varepsilon < 1/6$. Precisely,

$$\forall\, x \in \ker A,\ \forall\, |S| = s,\quad \|x_S\|_1 \le \frac{2\varepsilon}{1-2\varepsilon}\|x\|_1 \qquad (*)$$

Pf By reordering the left vertices, we may assume that

$$|x_1| \ge |x_2| \ge |x_3| \ge \cdots \ge |x_N|.$$

Thus, it is enough to establish $(*)$ for $S = S_1$, where

$$\underbrace{1, 2, \ldots, s}_{S_1};\ \underbrace{s+1, s+2, \ldots, 2s}_{S_2};\ \underbrace{2s+1, 2s+2, \ldots, 3s}_{S_3};\ \ldots$$

We have:

$$d\|x_S\|_1 = d\sum_{j\in S}|x_j| = \sum_{(j\to i)\in E(S)}|x_j| = \sum_{(j\to i)\in E'(S)}|x_j| + \sum_{(j\to i)\in E''(S)}|x_j|$$

$$= \sum_{i\in R(S)}|x_{\ell(i)}| + \sum_{(j\to i)\in E''(S)}|x_j| \qquad (**)$$

Observe that, for a fixed $i \in R(S)$, we have

$$0 = (Ax)_i = \sum_{j=1}^{N} A_{ij}x_j = \sum_{j:(j\to i)\in E} x_j = x_{\ell(i)} + \sum_{j\neq\ell(i):(j\to i)\in E} x_j$$

$$= x_{\ell(i)} + \sum_{j\neq\ell(i)\in S:(j\to i)\in E} x_j + \sum_{k\ge 2}\sum_{j\in S_k:(j\to i)\in E} x_j$$

Therefore, we obtain:

$$|x_{\ell(i)}| \le \sum_{j:(j\to i)\in E''(S)}|x_j| + \sum_{k\ge 2}\sum_{j\in S_k:(j\to i)\in E}|x_j|$$

Summing over all $i \in R(S)$, we get

$$\sum_{i \in R(S)} |\sigma_{\ell(i)}| \leq \sum_{(j \to i) \in E''(S)} |\sigma_j| + \sum_{k \geq 2} \sum_{\substack{j \in R(S),\ j \in S_k \\ (j \to i) \in E}} |\sigma_j|$$

Substituting in (**), we have

$$d \|\sigma_S\|_1 \leq 2 \sum_{(j \to i) \in E''(S)} |\sigma_j| + \sum_{k \geq 2} \sum_{\substack{i \in R(S),\ j \in S_k \\ (j \to i) \in E}} |\sigma_j| \qquad (***)$$

Bounding the first term in (***) :

$$\sum_{(j \to i) \in E''(S)} |\sigma_j| = \sum_{j=1}^{\hat{n}} \underbrace{\mathrm{Card}\{\ell \in E(j) \cap E''(S)\}}_{=: c_j} \times |\sigma_j|$$

note that the sets $\{\ell \in E(j) \cap E''(S)\}$, $j \in [\![1, \hat{n}]\!]$, are disjoints.

Thus, $c_1 + \dots + c_j = \mathrm{Card} \Big\{ \bigcup_{k \in [\![1,j]\!]} \{\ell \in E(k) \cap E''(S)\} \Big\}$

$$= \mathrm{Card}\{\ell \in E([\![1,j]\!]) \cap E''(S)\} = \mathrm{Card}\big(E([\![1,j]\!]) \cap E''(S)\big)$$

$$= \underset{\text{Lemma 2}}{\mathrm{Card}\, E''([\![1,j]\!])} \leq \underset{\text{Lemma 1}}{\varepsilon d j}$$

Now, by summation by parts, with $C_0 = 0$, $C_j = c_1 + \dots + c_j$, $j \geq 1$, we get

$$\sum_{j=1}^{\hat{n}} c_j |\sigma_j| = \sum_{j=1}^{\hat{n}-1} C_j \big(\underbrace{|\sigma_j| - |\sigma_{j+1}|}_{\geq 0}\big) + C_{\hat{n}} |\sigma_{\hat{n}}|$$

$$\leq \sum_{j=1}^{\hat{n}-1} \varepsilon d j \big(|\sigma_j| - |\sigma_{j+1}|\big) + \varepsilon d \hat{n} |\sigma_{\hat{n}}|$$

$$\left(\begin{array}{c}\text{invert the} \\ \text{summation} \\ \text{by parts}\end{array}\right) = \sum_{j=1}^{\hat{n}} \varepsilon d |\sigma_j| = \varepsilon d \|\sigma_S\|_1$$

Therefore : $\underline{\sum_{(j \to i) \in E''(S)} |\sigma_j| \leq \varepsilon d \|\sigma_S\|_1}$

Bounding the second term in ⋆⋆⋆ :

$k \geq 2$ Note that, for $j \in S_k$, we have $|\pi_j| \leq \frac{\|\pi_{S_{k-2}}\|_1}{\Delta}$. Thus,

$$\sum_{\substack{i \in R(S),\, j \in S_k \\ (j \to i) \in E}} |\pi_j| \leq \text{Card}\left\{ (j \to i) \in E, \text{ with } \begin{matrix} j \in S_k \\ i \in R(S) \end{matrix} \right\} \times \frac{\|\pi_{S_{k-2}}\|_1}{\Delta}$$

Note that, if $(j \to i) \in E$ with $j \in S_k$ & $i \in R(S)$, there holds

$$\ell(i) \leq \Delta < j \quad , \text{ therefore } \quad (j \to i) \in E''(S \cup S_k)$$

It follows that (Lemma 1)

$$\sum_{\substack{i \in R(S),\, j \in S_k \\ (j \to i) \in E}} |\pi_j| \leq 2\varepsilon d\Delta \times \frac{\|\pi_{S_{k-2}}\|_1}{\Delta} = 2\varepsilon d \|\pi_{S_{k-2}}\|_1$$

Then, $\sum_{k \geq 2} \sum_{\substack{i \in R(S),\, j \in S_k \\ (j \to i) \in E}} |\pi_j| \leq 2\varepsilon d \sum_{k \geq 2} \|\pi_{S_{k-1}}\|_1 \leq 2\varepsilon d \|\pi\|_1$

Finally, substituting the two bounds in ⋆⋆⋆, we obtain

$$d\|\pi_S\|_1 \leq 2\varepsilon d \|\pi_S\|_1 + 2\varepsilon d \|\pi\|_1 ,$$

that is : $\|\pi_S\|_1 \leq \frac{2\varepsilon}{1-2\varepsilon} \|\pi\|_1$, as expected.

Note that $\frac{2\varepsilon}{1-2\varepsilon} < \frac{1}{2}$ for $\varepsilon < \frac{1}{6}$. □

THE USE OF EXPANDER GRAPHS, CTD

III Existence

To prove: for all integer $k \geq 1$, for all $\varepsilon > 0$,
the existence of a (k, ε)-unbalanced expander (with left-degree to be determined)

We need: $\forall J \subseteq [1:N]$ with $\mathrm{Card}(J) \leq k$, $\mathrm{Card}(R(J)) \geq (1-\varepsilon)\, d\, \mathrm{Card}(J)$

We start by fixing $J \subseteq [1:N]$ with $\mathrm{Card}(J) \leq k$.
"k"

Let $i \in [1:m]$, observe that

$$P(i \notin R(J)) = \prod_{j \in J} P(i \notin R(j))$$

now for $j \in J$, $P(i \in R(j)) = \dfrac{\#\ d\text{-subsets of } [1:m] \text{ containing } i}{\#\ d\text{-subsets of } [1:m]}$

$R(j)$ is a d-subset of $[1:m]$ chosen uniformly at random

$$= \frac{\binom{d-1}{m-1}}{\binom{d}{m}} = \frac{d}{m}$$

so: $P(i \in R(j)) = \dfrac{d}{m}$, $P(i \notin R(j)) = 1 - \dfrac{d}{m}$

$$P(i \notin R(J)) = \left(1 - \frac{d}{m}\right)^k$$

For each $i \in [1:m]$, introduce the random variable

$$X_i = \begin{cases} 0 & \text{if } i \notin R(J), \\ 1 & \text{if } i \in R(J), \end{cases}$$

so that $\mathrm{Card}(R(J)) = \sum_{i=1}^{m} X_i$

Note that $\mathbb{E}(X_i) = 1 - \left(1 - \frac{d}{m}\right)^k$

(rk: therefore $\mathbb{E}(\text{Card } R(j)) = m\left(1 - \left(1 - \frac{d}{m}\right)^k\right) \leq dk$, as expected)

Chernoff bound: $X_1, X_2, \dots, X_m$ independent random variables taking the values 0 or 1. With $\mu := \mathbb{E}\left(\sum_{i=1}^m X_i\right)$, one has

$$P\left(\sum_{i=1}^m x_i < (1-\delta)\mu\right) \leq \exp\left(-\frac{\mu\delta^2}{2}\right), \text{ all } \delta > 0$$

Let $\mu_i := P(X_i = 1)$, so that $\mathbb{E}(X_i) = \mu_i$ and $\mathbb{E}\left(\sum X_i\right) = \sum_{i=1}^m \mu_i = \mu$

$$P(\Sigma X_i < (1-\delta)\mu) \underset{\text{all } t>0}{=} P\left(\exp(-t\Sigma X_i) > \exp(-t(1-\delta)\mu)\right)$$

$$\underset{\text{(Markov)}}{\leq} \frac{\mathbb{E}\left(\exp(-t\Sigma X_i)\right)}{\exp(-t(1-\delta)\mu)} = \frac{\mathbb{E}\left(\prod_{i=1}^m \exp(-tX_i)\right)}{\prod_{i=1}^m \exp(-t(1-\delta)\mu_i)} \underset{\text{independence}}{=} \prod_{i=1}^m \frac{\mathbb{E}(\exp(-tX_i))}{\exp(-t(1-\delta)\mu_i)}$$

Now we have:

$$\frac{\mathbb{E}(\exp(-tX_i))}{\exp(-t(1-\delta)\mu_i)} = \frac{(1-\mu_i) + \mu_i \exp(-t)}{\exp(-t(1-\delta)\mu_i)} = \frac{1 + \mu_i(\exp(-t) - 1)}{\exp(-t(1-\delta)\mu_i)} \leq \frac{1 + \mu_i\left(1 - t + \frac{t^2}{2} - 1\right)}{\exp(-t(1-\delta)\mu_i)}$$

$$\leq \frac{\exp\left(\mu_i\left(-t + \frac{t^2}{2}\right)\right)}{\exp(-t(1-\delta)\mu_i)} = \exp\left(\mu_i\left(\frac{t^2}{2} - t\delta\right)\right)$$

choose $t = \delta$: $\frac{\mathbb{E}(\exp(-\delta X_i))}{\exp(-\delta(1-\delta)\mu_i)} \leq \exp\left(-\frac{\mu_i\delta^2}{2}\right)$

Finally: $P(\Sigma X_i < (1-\delta)\mu) \leq \prod_{i=1}^m \exp\left(-\frac{\mu_i\delta^2}{2}\right) = \exp\left(-\frac{(\sum\mu_i)\delta^2}{2}\right)$

$$= \exp\left(-\frac{\mu\delta^2}{2}\right)$$

Exercise prove also that $P(\sum X_i > (1+\delta)\mu) \le \left(\frac{\exp(\delta)}{(1+\delta)^{1+\delta}}\right)^{\mu}$, all $\delta > 0$. 3/

Now back to our specific setting

$$P(\text{Card}(R(J)) < (1-\varepsilon)dk) = P\Big(\sum X_i < \underbrace{m\left[1-\left(1-\tfrac{1}{m}\right)^k\right]}_{\mu} - \underbrace{\left(m\left[1-\left(1-\tfrac{1}{m}\right)^k\right] - (1-\varepsilon)dk\right)}_{\delta\mu}\Big)$$

$$\le \exp\left(-\frac{\delta\mu)^2}{2\mu}\right) = \exp\left(-\frac{\left(m\left[1-\left(1-\frac{1}{m}\right)^k\right] - (1-\varepsilon)dk\right)^2}{2m\left[1-\left(1-\frac{1}{m}\right)^k\right]}\right) =: \exp\left(-\frac{\text{Num}}{\text{Den}}\right)$$

note that: $\left(1-\frac{1}{m}\right)^k \ge 1-\frac{k}{m}$, $\le 1-\frac{dk}{m}+2\frac{d^2k^2}{m^2}$; $\left[1-\left(1-\frac{1}{m}\right)^k\right] \le \frac{dk}{m}$, $\ge \frac{dk}{m}-2\frac{d^2k^2}{m^2}$

so: $\text{Den} \le 2dk$

$$\text{Num} \ge \left[\frac{dk}{m} - 2\frac{d^2k^2}{m} - dk + \varepsilon dk\right]^2 = \left[(dk)\left(\varepsilon - \frac{2dk}{m}\right)\right]^2, \text{ as long as } \left\{\varepsilon \ge \frac{2dk}{m}\right\}$$

Hence, we have

$$P(\text{Card}(R(J)) < (1-\varepsilon)dk) \le \exp\left(-\frac{dk}{2}\left(\varepsilon - 2\frac{dk}{m}\right)^2\right)$$

$$\le \exp\left(-\frac{dk}{2}\left(\varepsilon - 2\frac{dk}{m}\right)^2\right), \text{ as long as } \left\{\varepsilon \ge \frac{2dk}{m}\right\}$$

We choose d as the largest integer $< \frac{\varepsilon m}{6k}$ (this requires $\varepsilon m \ge 6k$, i.e. $m \ge \frac{6k}{\varepsilon}$ ⊛)

so that $\frac{\varepsilon m}{12k} \le d < \frac{\varepsilon m}{6k}$

then: $\varepsilon - \frac{2dk}{m} \ge \varepsilon - \frac{\varepsilon}{6} = \frac{5}{6}\varepsilon$

It follows: $P\left(\mathrm{Card}\left(R(J)\right) < (1-\varepsilon)\,dk\right) \le \exp\left(-\frac{\varepsilon m}{24} \times \left(\frac{\varepsilon}{6}\right)^2\right)$

$= \exp\left(-c\varepsilon^3 m\right)$

Finally, it remains to "unfix" the subset J of $[1:N]$

$P\left(\mathrm{Card}\left(R(J)\right) < (1-\varepsilon)\,dk \quad \text{for some } J \subseteq [1:N] \text{ with } \mathrm{card}(J) \le k\right)$

$\le \sum_{k=0}^{k} \binom{N}{k} \times \exp\left(-c\varepsilon^3 m\right) \le k\binom{N}{k} \exp\left(-c\varepsilon^3 m\right)$

$\binom{N}{k} \le \binom{N}{k}$ since $k \le N/2$ [assumed]

$\le \underbrace{k}_{\le 2^k} \left(\frac{eN}{k}\right)^k \exp\left(-c\varepsilon^3 m\right) \le \exp\left(k \ln\left(\frac{2eN}{k}\right) - c\varepsilon^3 m\right)$

We impose $k \ln\left(\frac{2eN}{k}\right) \le \frac{c}{2}\varepsilon^3 m$,

that is $m \ge c' \frac{k}{\varepsilon^3} \ln\left(\frac{2eN}{k}\right)$,

which then implies:

$P\left(\mathrm{Card}\left(R(J)\right) < (1-\varepsilon)\,dk \quad \text{for some } J \subseteq [1:N] \text{ with } \mathrm{Card}\, J \le k\right)$

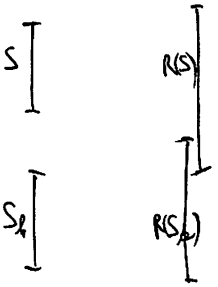
$\le \exp\left(-\frac{c}{2}\varepsilon^3 m\right)$

Corrigendum

Here is the correct argument to prove that (see Bounding the second term in [illegible])

$$\mathrm{Card}\left\{(j \to i) \in E \text{ , with } j \in S_\ell \text{ , } i \in R(S)\right\} \leq 2\varepsilon d s$$

Partition $E(S \cup S_\ell)$ as follows (disjoint union)



$$E(S \cup S_\ell) = E(S)$$
$$\cup \left\{(j \to i) \in E(S_\ell) \text{ , } i \notin R(S)\right\}$$
$$\cup \left\{(j \to i) \in E(S_\ell) \text{ , } i \in R(S)\right\}$$

so:

$$2sd = sd + \mathrm{Card}\left\{(j \to i) \in E(S_\ell) \text{ , } i \notin R(S)\right\} + \mathrm{Card}\left\{(j \to i) \in E(S_\ell) \text{ , } i \in R(S)\right\}$$
$$\geq sd + \underbrace{\mathrm{Card}\left(R(S_\ell) \setminus \left(R(S_\ell) \cap R(S)\right)\right)}_{= \mathrm{Card}(R(S \cup S_\ell)) - \mathrm{Card}(R(S))} + \mathrm{Card}\left\{(j \to i) \in E(S_\ell) \text{ , } i \in R(S)\right\}$$
$$= sd + 2sd(1-\varepsilon) - sd + \mathrm{Card}\left\{(j \to i) \in E(S_\ell) \text{ , } i \in R(S)\right\}$$

Thus: $\mathrm{Card}\left\{(j \to i) \in E(S_\ell) \text{ , } i \in R(S)\right\} \leq 2sd\varepsilon$, as expected.

Chapter 13
Orthogonal Matching Pursuit

See the following handwritten notes.

Orthogonal Matching Pursuit

Let $A = \left[A_1 | A_2 | \cdots | A_N \right]$ be the $m \times N$ measurement matrix.

Let $y = Ax \in \mathbb{R}^m$ be the measurement vector of an s-sparse vector $x \in \mathbb{R}^N$ supported on S

We construct sequences $(S_n)_{n=0}^{s}$, $(z_n)_{n=0}^{s}$, and $(r_n)_{n=0}^{s}$, where

S_n is a subset of $[1:N]$
z_n is a vector in $\mathbb{R}^N$ supported on S_n that approximates x
r_n is a vector in $\mathbb{R}^m$ representing the residual $y - Az_n$,

according to the following algorithm:

$S_0 = \emptyset$, $z_0 = 0$, $r_0 = y$

$S_n = S_{n-1} \cup \{j_n\}$, where $j_n := \underset{j \in [1:N]}{\operatorname{argmax}} |\langle r_{n-1}, A_j \rangle|$

$z_n = \underset{z \in \Sigma_{S_n}}{\operatorname{argmin}} \| y - Az \|_2$ ∉ orthogonal projection of y onto $\operatorname{Im}(A_{S_n})$

$r_n = y - Az_n$ ‖ see the MATLAB implementation

Of course, if $r_n = 0$ for some $n \leq s$, we stop the iteration, and we set $r_{n+1} = 0$, etc...

Observation 1: $r_n \perp \operatorname{span}(A_{j_1}, \ldots, A_{j_n}) \; (= \operatorname{Im} A_{S_n})$

Observation 2: if $r_{n-1} \neq 0$, then $j_n \notin S_{n-1}$

indeed, we have $\langle r_{n-1}, A_j \rangle = 0$ for all $j \in S_{n-1}$
and $\langle r_{n-1}, A_j \rangle \neq 0$ for some $j \in [1:N]$

Theorem Suppose that, for all $r \in \mathrm{Im}\, A_S \setminus \{0\}$,

$$\max_{j \in S} |\langle r, A_j \rangle| > \max_{\ell \in \bar{S}} |\langle r, A_\ell \rangle| . \quad (*)$$

Then $x_S = 0$.

Proof ← make the assumption that $x_0 \neq 0, x_1 \neq 0, \dots, x_{s-1} \neq 0$

By induction on n, we get $S_n \subseteq S$ for all $0 \leq n \leq s$.

indeed, if $S_{n-1} \subseteq S$, we can write $r_{n-1} = A(\underbrace{x}_{\in \Sigma_S} - \underbrace{z_{n-1}}_{\in \Sigma_S}) \in \mathrm{Im}\, A_S \setminus \{0\}$

so that $\max_{j \in S} |\langle r_{n-1}, A_j \rangle| > \max_{\ell \in \bar{S}} |\langle r_{n-1}, A_\ell \rangle|$ implies $j_n \in S$.

We have also seen that $\mathrm{Card}(S_n) = n$ for all $0 \leq n \leq s$.

So we can conclude that $S_s = S$.

Now $\underbrace{\max_{j \in S} |\langle r_s, A_j \rangle|}_{=0} > \max_{\ell \in \bar{S}} |\langle r_s, A_\ell \rangle|$ if $r_s \neq 0$ forces $r_s = 0$. □

Corollary Given an $m \times N$ matrix A whose columns are ℓ_2-normalized, if its coherence satisfies $\mu(A) < \frac{1}{2s-1}$, then every s-sparse vector $x \in \mathbb{R}^N$ is recovered from $y = Ax \in \mathbb{R}^m$ via OMP.

Rk We have seen that $\mu(A) < \frac{1}{2s-1}$ also guarantees recovery by ℓ_1-min.

Proof Let $r =: \sum_{i \in S} x_i A_i \in \mathrm{Im}\, A_S \setminus \{0\}$.

For $\ell \in \bar{S}$, we have $|\langle r, A_\ell \rangle| = \left|\sum_{i \in S} x_i \langle A_i, A_\ell \rangle\right| \leq s \mu(A) \|x\|_\infty$,

and for $j \in S$ with $|x_j| = \|x\|_\infty$, we have

$$|\langle r, a_j \rangle| = \Big|\sum_{i \in S} x_i \langle a_i, a_j \rangle\Big| \geq |x_j| - \sum_{\substack{i \in S \\ i \neq j}} |x_i| |\langle a_i, a_j \rangle| \geq \|x\|_\infty - (s-1)\mu(A)\|x\|_\infty .$$

Thus, the condition (*) is fulfilled as soon as

$$\|x\|_\infty \big(1 - (s-1)\mu(A)\big) > \|x\|_\infty \, s\mu(A) ,$$

or equivalently (since $\|x\|_\infty \neq 0$),

$$(2s-1)\mu(A) < 1 .$$

This implies that $x_S = 0$, i.e. $A(x - x_S) = 0$. To prove that $x - x_S = 0$, observe that, if $(c_1, \dots, c_N)$ are the components of $x - x_S$ ($c_\ell = 0$ if $\ell \in S$), then $0 = \sum_{j \in S} c_j a_j$, and then, with $j \in [1:N]$: $|c_j| = \|c\|_\infty$,

$$|c_j| |\langle a_j, a_j \rangle| = \Big|\sum_{\substack{i \in S \\ i \neq j}} c_i \langle a_i, a_j \rangle\Big| \longrightarrow \|c\|_\infty \leq (s-1)\mu(A)\|c\|_\infty < \frac{s-1}{2s-1}\|c\|_\infty$$

absurd if $\|c\|_\infty > 0$. ∎

We have exhibited a $m \times m^2$ measurement matrix A with $\mu(A) = \frac{1}{\sqrt{m}-1}$, so using this matrix we are guaranteed reconstruction of every s-sparse vector via OMP as soon as $\underline{m > (2s-1)^2 \asymp s^2}$.

This can — and must — be improved. The following theorem was proved by Gilbert and Tropp.

Theorem Let $x \in \mathbb{R}^N$ be an s-sparse vector.
Fix $0 < \delta < 1$ and choose $m \geq Cst \cdot s \log(N/\delta)$ [Cst is an absolute constant]
Let A be an $m \times N$ Gaussian random matrix.
OMP recovers the vector x from the knowledge of $y = Ax$ with probability $\geq 1-\delta$.

Some weaknesses of this result:

- The number of measurements $\asymp s \log(\frac{N}{\delta})$ not as good as $\asymp s \log(N)$
- The result is non uniform, it says
  $\forall x$, $P(x$ recovd$) \geq 1-\delta$
  which is different from
  $$P(\forall x, x \text{ recovd}) \geq 1-\delta$$
- The result does not apply to partial Fourier matrices.

Chapter 14
ROMP and CoSaMP

Appendix 1: Some Theorems and Their Proofs

Proof of Borsuk–Ulam Theorem: to come

Proof of Krein–Mil'man Theorem: to come

Proof of Farkas' Lemma: to come

Proof of Karush–Kuhn–Tucker: to come

It is a rather common problem to minimize a function defined by a maximum, e.g. when we look at a best approximation. The following theorem generalizes some characterizations of best approximations or of minimal projections. Roughly speaking, it allows the reduction of the domain of maximization.

Theorem 14.1. Let f be a function defined on $C \times K$ where C is a convex set and K is a compact set. We assume the convexity of $f(\bullet, y)$ for any $y \in K$, the continuity of $f(x, \bullet)$ for any $x \in C$ and the equicontinuity of $(f(\bullet, y))_{y \in K}$ at some point $x^* \in C$. The following propositions are equivalent:

(i) $\exists\, x \in C : \max\limits_{y \in K} f(x, y) < \max\limits_{y \in K} f(x^*, y)$,

(ii) $\exists\, x \in C : \forall\, z \in K$ satisfying $f(x^*, z) = \max\limits_{y \in K} f(x^*, y)$, one has $f(x, z) < f(x^*, z)$.

If in addition the set K is convex and the function $f(x^*, \bullet)$ is convex, the propositions (i)–(ii)

are also equivalent to

(iii) $\exists\, x \in C : \forall\, z \in \mathrm{Ex}(K)$ satisfying $f(x^*,z) = \max\limits_{y\in K} f(x^*,y)$, one has $f(x,z) < f(x^*,z)$.

Proof. The implications (i)⇒(ii) and (ii)⇒(iii) are straightforward.
For $x \in C$, we consider the non-empty compact subset D_x of K defined by

$$D_x := \big\{z \in K : f(x,z) = \max_{y\in K} f(x,y)\big\}.$$

Let us assume that (ii) holds, i.e. that there exists $x \in K$ for which

$$m := \max_{z\in D_{x^*}} f(x,z) < \max_{y\in K} f(x^*,y) =: M.$$

We consider the open neighborhood of D_{x^*} defined by $\mathcal{O} := \big\{y \in K :\ f(x,y) < (m+M)/2\big\}$.
For $y \in K \setminus \mathcal{O} \subseteq K \setminus D_{x^*}$, we have $f(x^*,y) < M$, and we set

$$m' := \max_{y\in K\setminus\mathcal{O}} f(x^*,y) < M.$$

Let $t > 0$ be small enough for $|f(x^*+t(x-x^*),y) - f(x^*,y)| < M - m'$ to hold for any $y \in K$.
We then get

$$\begin{aligned} y \in K\setminus\mathcal{O} \ &\Rightarrow\ f(x^*+t(x-x^*),y) < M - m' + f(x^*,y) \le M,\\ y \in \mathcal{O} \ &\Rightarrow\ f((1-t)x^*+tx,y) \le (1-t)f(x^*,y) + tf(x,y) \le (1-t)M + t(m+M)/2 < M. \end{aligned}$$

We have therefore obtained (i) in the form $\max\limits_{y\in K} f((1-t)x^*+tx,y) < M$.
Let us now assume that K is a convex set and that $f(x^*,\bullet)$ is a convex function. It follows that D_{x^*} is an extreme set of K, hence that $\mathrm{Ex}(D_{x^*}) = \mathrm{Ex}(K) \cap D_{x^*}$. The property (iii), assumed to hold, now reads, for some $x \in C$,

$$\forall\, z \in \mathrm{Ex}(D_{x^*}), \quad f(x,z) < M.$$

Thus, for any $y \in D_{x^*} \subseteq \overline{\mathrm{co}}[D_{x^*}] = \overline{\mathrm{co}}[\mathrm{Ex}(D_{x^*})]$, we have $f(x,y) \le M$. We aim at proving property (ii) as the statement that the set $S := \big\{y \in D_{x^*} : f(x^*,y) = M\big\}$ is empty. If it was not, due to the compactness of S, we would have $\mathrm{Ex}(S) \neq \varnothing$. But since S is an extreme set of D_{x^*}, we have $\mathrm{Ex}(S) = \mathrm{Ex}(D_{x^*}) \cap S$, which is empty. □

Corollary 14.2. Let V be a subspace of a normed space X. For $x \in X$ and $v^* \in V$, one has

$$\big[\, \|x-v^*\| = \inf_{v\in V} \|x-v\| \,\big] \iff \big[\, \forall\, v \in V,\ \exists\, \lambda \in \mathrm{Ex}(B_{X^*}) : \lambda(x-v) \ge \lambda(x-v^*) = \|x-v^*\| \,\big].$$

Appendix 2: Hints to the Exercises

Chapter 1. Ex.1: need to prove that $B^\top B$ is invertible; observe that $B^\top Bx = 0$ implies that $\|Bx\|_2^2 = \langle B^\top Bx, x\rangle = 0$ and in turn that $\mathbf{x} = 0$ because B is injective by the rank theorem. Ex.2: write $\mathbf{x} = U\mathbf{x}'$, where $\mathbf{x}'$ is s-sparse and $U_{i,j} = 1$ for $i \leq j$, 0 otherwise; recover $\mathbf{x}'$ by minimizing $\|U^{-1}z\|_1$ subject to $A\mathbf{z} = \mathbf{y}$; calculate $U^{-1}\mathbf{z}$. Ex.3: reduce the problem to the case $\Omega = \pi$; define a 2π-periodic function g by $g_{|[-\pi,\pi]} = \hat{f}_{|[-\pi,\pi]}$ and calculate its Fourier coefficients; use the inversion formula to express f in terms of g. Ex.4: dimension considerations. Ex.5: for $\mathbf{y} \neq \mathbf{y}' \in \mathbb{R}^m$, consider the errors $\mathbf{e} := (B\mathbf{y}' - B\mathbf{y})_{[\![1,s]\!]}$ and $\mathbf{e}' = (B\mathbf{y} - B\mathbf{y}')_{[\![s+1,2s]\!]}$. Ex.6: `N=512;m=128;s=20; y=rand(m,1); A=randn(m,N); [V,D]=eig(A'*A);B=V*[randn(N-m,m);zeros(m,m)]; permN=randperm(N);supp =sort(permN(1:s)); e=zeros(N,1);e(supp)=rand(s,1); estar=l1eq_pd(x,A,[] ,A*x); ystar=B\(x-estar); norm(y-ystar)` Ex.7: $\forall\, \mathbf{u} \in \ker A \setminus \{0\},\ |\mathrm{supp}(\mathbf{u})|_w > 2\max_{|S|\leq s} |\mathrm{supp}(\mathbf{u}) \cap S|_w$.

Chapter 2. Ex.1: if $f : \mathcal{S} \to \mathbb{N}$ is an injection, then extend $\left[f^{|f(\mathcal{S})}\right]^{-1}$ to obtain a surjection $g : \mathbb{N} \to \mathcal{S}$; if $g : \mathbb{N} \to \mathcal{S}$ is a surjection, then choose $f(x) \in g^{-1}(\{x\})$ for all $x \in \mathcal{S}$ to define an injection $f : \mathcal{S} \to \mathbb{N}$. Ex.2: ... Ex.3: $F : \mathbf{x} \in \Sigma_{[1:s]} \mapsto f(\mathbf{x}) - f(-\mathbf{x}) \in \mathbb{R}^m$ is continuous and antipodal; if $s > m$, then Borsuk–Ulam theorem yields a contradiction. Ex.4: if $\mathbb{S}^n_{(1)}$ and $\mathbb{S}^n_{(2)}$ are the unit spheres in $\mathbb{R}^{n+1}$ relative to two norms $\|\cdot\|_{(1)}$ and $\|\cdot\|_{(2)}$, then compose a continuous antipodal map from $\mathbb{S}^n_{(2)}$ to $\mathbb{R}^n$ with the map $\mathbf{x} \in \mathbb{S}^n_{(1)} \to \frac{\mathbf{x}}{\|\mathbf{x}\|_{(2)}} \in \mathbb{S}^n_{(2)}$ to obtain a continuous and antipodal map from $\mathbb{S}^n_{(1)}$ to $\mathbb{R}^n$. Ex.5: starting form G, define F by $F(\mathbf{x}) := G(\mathbf{x}) - G(-\mathbf{x})$ and apply the first formulation; starting form F, apply the second formulation and use antipodality. Ex.6: the concatenation of two weighted planar networks is a weighted planar network. Ex.7: start by factoring out the term $(1 - x_j)^n$ for the j-th column and $\binom{n}{i}$ for the i-th row. Ex.8: the condition necessary and sufficient is $\det M_{[1:k]} \neq 0$, all $k \in [1 : n-1]$, which is satisfied by totally positive matrices; Newton's interpolation yields $p(x) = \sum_{k=0}^n [x_0, \ldots, x_k]p \cdot (x - x_0) \cdots (x - x_k)$ for $p \in \mathcal{P}_n$, where the divided difference $[x_0, \ldots, x_k]p$ is the coefficient on x^k in the polynomial of degree $\leq k$

interpolating p at $x_0, \dots, x_k$; specify $p(x) = x^j$ and $x = x_i$. Ex.9: `N=512; m=128; s=20; R=sort(rand(1,N)); for i=1:m, A(i,:)=R.^(i-1); end; permN=randperm(N); supp=sort(permN(1:s)); x=zeros(N,1); x(supp)=rand(s,1); y=A*x; x1=A\y; xstar=l1eq_pd(x1,A,[],y,1e-3);`

Chapter 3. Ex.1: the eigenvectors are the $[1, e^{i2\pi j/N}, \dots, e^{i2\pi j(N-1)/N}]^\top$, $j \in [0 : N-1]$, and the eigenvalues are the discrete Fourier coefficients of $[c_0, \dots, c_{N-1}]^\top$. Ex.2: subtract a variable x to every entry, then the determinant is a linear function to be evaluated at two particular values for x. Ex.3: write down the definitions of $\widehat{u * v}(j)$ and of $(\hat{u} * \hat{v})(j)$, manipulate the sums to obtain $\hat{u}(j) \cdot \hat{v}(j)$ and $N\widehat{u \cdot v}(j)$. Ex.4: `P=[1]; for k=1:20, P=conv(P,[1,-k]); end; P(2)=P(2)+10^(-8); roots(P)';` Ex.5: `N=500; s=18; supp=sort(randsample(N,s));x=randn(N,1)/10^4;x(supp)=randn(s,1); xhat=fft(x);y=xhat(1:2*s);A=toeplitz(y(s:2*s-1),y(s:-1:1));phat=zeros(N,1); phat(1)=1; phat(2:s+1)=-A\y(s+1:2*s); p=ifft(phat); [sorted_p,ind]=sort (abs(p)); rec_supp=sort(ind(1:s)); [supp';rec_supp']`

Chapter 4. Ex.1: use the triangle inequality $\|\mathbf{x}+\mathbf{y}\|_q^q \le \|\mathbf{x}\|_q^q + \|\mathbf{y}\|_q^q$ and Hölder's inequality $\|\mathbf{x}\|_q^q + \|\mathbf{y}\|_q^q \le [1+1]^{1-q}[\|\mathbf{x}\|_q + \|\mathbf{y}\|_q]^q$ to derive $\|\mathbf{x}+\mathbf{y}\|_q \le 2^{(1-q)/q}[\|\mathbf{x}\|_q + \|\mathbf{y}\|_q]$; for $\|\mathbf{x}\|_q = 1$, write $\|(T+U)\mathbf{x}\|_q \le 2^{(1-q)/q}[\|T\mathbf{x}\|_q + \|U\mathbf{x}\|_q] \le 2^{(1-q)/q}[\|T\|_q + \|U\|_q]$ and take the supremum. Ex.2: if $\mathbf{v} \in \Sigma_{2s} \cap \ker A$, then take S to be an index set of s largest absolute-value components of $\mathbf{v}$ to get $\|\mathbf{v}_S\|_q^q \ge \|\mathbf{v}_{\bar{S}}\|_q^q$. Ex.3: take a $(2s) \times (2s+1)$ matrix whose kernel is spanned by $[\underbrace{a, \dots, a}_{s}, \underbrace{1, \dots, 1}_{s+1}]^\top$ with $a := (1+1/s)^{1/q}$. Ex.4: to prove the strengthened Minimality Property, apply the strengthened Null-Space Property with $\mathbf{v} = \mathbf{x} - \mathbf{z}$, to prove the strengthened Null-Space Property, apply the strengthened Minimality Property with $\mathbf{x} = \mathbf{v}_S$ and $\mathbf{z} = -\mathbf{v}_{\bar{S}}$, observe that $c = C$.

Chapter 5. Ex.1: imitate the proof of Proposition 5.1. Ex.2: if S is the index set of s largest absolute-value components of $\mathbf{x}$, then the best s-term approximation to $\mathbf{x}$ is provided by $\mathbf{x}_S$ independently of q. Ex.3: to establish Instance Optimality from the Null-Space Property, define the reconstruction map by $g(\mathbf{y}) \in \operatorname{argmin}\{\sigma_s(\mathbf{z})_1 : A\mathbf{z} = \mathbf{y}\}$, keep in mind the inequality $\sigma_s(\mathbf{a}+\mathbf{b})_1 \le \sigma_s(\mathbf{a})_1 + \sigma_1(\mathbf{b})$, $\mathbf{a}, \mathbf{b} \in \mathbb{R}^N$, conversely, to establish the Null-Space Property from Instance Optimality, given $\mathbf{v} \in \ker A$, choose an index set S so that $\|\mathbf{v}_{\bar{S}}\|_1 = \sigma_{2s}(\mathbf{v})_1$, and split $\mathbf{v}_S$ as $\mathbf{v}_S = \mathbf{v}_1 + \mathbf{v}_2$ with $\mathbf{v}_1, \mathbf{v}_2 \in \Sigma_s$, then justify that $\|\mathbf{v}\|_1 = \|\mathbf{v}_2 + \mathbf{v}_{\bar{S}} - g(A(\mathbf{v}_2 + \mathbf{v}_{\bar{S}}))\|_1 \le C\sigma_s(\mathbf{v}_2 + \mathbf{v}_{\bar{S}})_1 = C\|\mathbf{v}_{\bar{S}}\|_1 = C\sigma_{2s}(\mathbf{v})_1$. Ex.4: adapt Ex.3 to observe that, if A exhibits Instance Optimality of order s, then there is a constant $c < 1$ such that $\|\mathbf{v}_S\|_2 \le c\|\mathbf{v}\|_2$ for all $\mathbf{v} \in \ker A$ and $|S| \le s$, given the canonical basis $(\mathbf{e}_1, \dots, \mathbf{e}_N)$ of $\mathbb{R}^N$ and given an orthonormal basis $(\mathbf{v}_1, \dots, \mathbf{v}_{N-m})$ of $\ker A$, we get $\sum_{i=1}^{N-m} \langle \mathbf{e}_j, \mathbf{v}_i \rangle^2 =$

$\langle \sum_{i=1}^{N-m} \langle \mathbf{e}_j, \mathbf{v}_i \rangle \mathbf{v}_i, \mathbf{e}_j \rangle \leq c \| \sum_{i=1}^{N-m} \langle \mathbf{e}_j, \mathbf{v}_i \rangle \mathbf{v}_i \|_2 \leq c \|\mathbf{e}_j\|_2 = c$, sum over $j \in [1:N]$ and invert the summations to obtain $N - m \leq cN$.

Chapter 6. Ex.2: minimize t subject to $-t \leq x_i - v_i \leq t$ and $\langle \mathbf{v}, \mathbf{u}_1 \rangle = 0, \ldots, \langle \mathbf{v}, \mathbf{u}_k \rangle = 0$, where $(\mathbf{u}_1, \ldots, \mathbf{u}_k)$ denotes a basis of the orthogonal complement $\mathcal{V}^\perp$ of $\mathcal{V}$ in $\mathbb{R}^n$. Ex.5: minimize $\sum t_j$ subject to $z_{\mathrm{re},j}^2 + z_{\mathrm{im},j}^2 \leq t_j^2$, $A_{\mathrm{re}}\mathbf{z}_{\mathrm{re}} - A_{\mathrm{im}}\mathbf{z}_{\mathrm{im}} = \mathbf{y}_{\mathrm{re}}$, $A_{\mathrm{re}}\mathbf{z}_{\mathrm{im}} + A_{\mathrm{im}}\mathbf{z}_{\mathrm{re}} = \mathbf{y}_{\mathrm{im}}$.

Bibliography

[1] S. Boyd, L. Vandenberghe. *Convex Optimization*. Cambridge University Press.

[2] E. Candès, J. Romberg. ℓ_1-magic: Recovery of Sparse Signals via Convex Programing. Available online.

[3] R. Baraniuk, M. Davenport, R. DeVore, M. Wakin. A Simple Proof of the Restricted Isometry Property for Random Matrices, Constructive Approximation, ...

[4] Emmanuel Candès. Compressive Sampling, Proceeding of the INternational Congress of Mathematicians, Madrid, Spain, 2006.

[5] Emmanuel Candès. The Restricted Isometry Property and Its Implication for Compressed Sensing, Comptes Rendus de l'Académie des Sciences, Paris, Série I, 346, 589–592, 2008.

[6] Emmanuel Candès, Justin Romberg, Terence Tao. Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information, ...

[7] Emmanuel Candès, Justin Romberg, Terence Tao. Stabl Signal Recovery from Incomplete and Inaccurate Measurements, ...

[8] Emmanuel Candès, Terence Tao. Decoding by Linear Programing, ...

[9] Emmanuel Candès, Terence Tao. Near Optimal Signal Recovery from Random Projections: Universal Encoding Strategies?, ...

[10] E. Candès, M. Wakin, S. Boyd. Enhancing Sparsity by Reweighted ℓ_1 Minimization, ...

[11] Venkat Chandar. A Negative Result Concerning Explicit Matrices with the Restricted Isometry Property, ...

[12] Rick Chartrand, Valentina Stavena. Restricted Isometry Properties and Nonconvex Compressive Sensing, ...

[13] Albert Cohen, Wolfgang Dahmen, Ronald DeVore. Compressed Sensing and best k-term approximation, ...

[14] Albert Cohen, Wolfgang Dahmen, Ronald DeVore. Instance Optimal Decoding by Thresholding in Compressed Sensing, ...

[15] I. Daubechies, R. DeVore, M. Fornasier, C. Sinan Gúntürk. Iteratively Re-Weighted Least Squares Minimization for Sparse Recovery, ...

[16] Michael Davies, Rémi Gribonval. Restricted Isometry Constants where ℓ_p sparse recovery can fail for $0 < p \leq 1$, Preprint submitted to ... .

[17] Ronald DeVore, Deterministic Constructions of Compressed Sensing Matrices, ...

[18] Ronald DeVore, Guergana Petrova, Przemyslaw Wojtaszczyk. Instance-Optimality in Probability with an ℓ_1-Minimization Decoder, ...

[19] David Donoho. Compressed Sensing, ...

[20] David Donoho, For Most Large Underdetermined Systems of Linear Equations the Minimal ℓ^1-norm Solution is also the Sparsest Solution, ...

[21] David Donoho, For Most Large Underdetermined Systems of Equations, the Minimal ℓ^1-norm Near-Solution Approximates the Sparsest Near-Solution, ...

[22] David Donoho, Jared Tanner. Counting Faces of Randomly-Projected Polytopes when the Projection Radically Lowers Dimension, ...

[23] David Donoho, Jared Tanner. Counting the Faces of Randomly-Projected Hypercubes and Orthants, with Applications, ...

[24] S. Foucart, M.-J. Lai. Sparsest solutions of underdetermined linear systems via ℓ_q-minimization for $0 < q \leq 1$, Applied and Computational Harmonic Analysis, in press.

[25] Jean-Jacques Fuchs, On Sparse Representations in Arbitrary Redundant Bases, ...

[26] R. Berinde, A. Gilbert, P. Indyk, H. Karloff, J. Strauss. Combining Geometry and Combinatorics: a Unified Approach to Sparse Signal Recovery, ...

[27] R. Gribonval, M. Nielsen. Highly Sparse Representations from Dictionaries are Unique and Independent of the Sparsness Measure, ...

[28] B. S. Kashin, V. N. Temlyakov. A Remark on Compressed Sensing

[29] Basarab Matei, Yves Meyer. A Variant on the Compressed Sensing of Emmanuel Candès, ...

[30] D. Needell, J. Tropp. CoSamp: Iteravtive Signal Recovery from Incomplete and Inaccurate Samples, ...

[31] D. Needell, R. Vershynin. Uniform Uncertainty Principle and Signal Recovery via Regularized Orthogonal Matching Pursuit, ...

[32] D. Needell, R. Vershynin. Signal Recovery from Incomplete and Inaccurate Measurements vis Regularized Orthogonal Matching Pursuit, ...

[33] Yaakov Tsaig, David Donoho. Extensions of Compressed Sensing, ...

[34] P. Wojtaszczyk. Stability and Instance Optimality for Gaussian Measurements in Compressed Sensing, ...