“RIP” 学习笔记

黄冬勃

2015 年 4 月 24 日

Contents

Chapter 1

Math Basics

1.1 Background in Linear Algebra

1.1.1 Types of Matrices

1. Symmetric matrices: $A^T = A$.
2. Hermitian matrices: $A^H = A$.
3. Skew-symmetric matrices: $A^T = -A$.
4. Skew-Hermitian matrices: $A^H = -A$.
5. Normal matrices: $A^H A = AA^H$.
6. Nonnegative matrices: $a_{ij} \geq 0, i, j = 1, \ldots, n$ (similar definition for nonpositive, positive, and negative matrices).
7. Unitary matrices: $Q^H Q = I$.

1.2 Moore-Penrose pseudoinverse

1.2.1 Notation

- K will denote one of the fields of real or complex numbers, denoted $\mathbb{R}, \mathbb{C}$, respectively. The vector space of $m \times n$ matrices over K is denoted by $M(m, n; K)$.
- For $A \in M(m, n; K)$, A^T and A^* denote the transpose and Hermitian transpose (also called conjugate transpose) respectively. If $K = \mathbb{R}$, then $A^* = A^T$.
- For $A \in M(m, n; K)$, then $\mathrm{im}(A)$ denotes the range of A (the space spanned by the column vectors of A) and $\ker(A)$ denotes the kernel (null space) of A.
- Finally, for any positive integer $n, I_n \in M(n, n; K)$ denotes the $n \times n$ identity matrix.

1.2.2 Definition

For $A \in M(m, n; K)$, a pseudoinverse of A is defined as a matrix $A^+ \in M(n, m; K)$ satisfying all of the following four criteria:

1. $AA^+A = A$ (AA^+ need not be the general identity matrix, but it maps all column vectors of A to themselves);
2. $A^+AA^+ = A^+$ (A^+ is a weak inverse for the multiplicative semigroup);

3. $(AA^+)^* = AA^+$ (AA^+ is Hermitian); and
4. $(A^+A)^* = A^+A$ (A^+A is also Hermitian).

Matrix A^+ exists for any matrix A, but when the latter has full rank, A^+ can be expressed as a simple algebra formula.

In particular, when A has *full column rank* (and thus matrix A^*A is invertible), A^+ can be computed as:

$$A^+ = (A^*A)^{-1}A^*. \tag{1.2.1}$$

This particular pseudoinverse constitutes a *left inverse*, since, in this case, $A^+A = I$.

When A has *full row rank* (matrix AA^* is invertible), A^+ can be computed as:

$$A^+ = A^*(AA^*)^{-1}. \tag{1.2.2}$$

This is a *right inverse*, as $AA^+ = I$.

1.2.3 Properties

- If A has real entries, then so does A^+.
- If A is invertible, its pseudoinverse is its inverse. That is: $A^+ = A^{-1}$.
- The pseudoinverse of a zero matrix is its transpose.
- The pseudoinverse of the pseudoinverse is the original matrix: $(A^+)^+ = A$.
- Pseudoinverse commutes with transposition, conjugation, and taking the conjugate transpose:

$$(A^T)^+ = (A^+)^T, \overline{A}^+ = \overline{A^+}, (A^*)^+ = (A^+)^*. \tag{1.2.3}$$

- The pseudoinverse of a scalar multiple of A is the reciprocal multiple of A^+:

$$(\alpha A)^+ = \alpha^{-1}A^+ for \alpha \neq 0. \tag{1.2.4}$$

Identities

$$\begin{aligned}
A^+ &= A^+ \quad A^{+*} \quad A^* \\
A^+ &= A^* \quad A^{+*} \quad A^+ \\
A &= A^{+*} \quad A^* \quad A \\
A &= A \quad A^* \quad A^{+*} \\
A^* &= A^* \quad A \quad A^+ \\
A^* &= A^+ \quad A \quad A^*
\end{aligned} \tag{1.2.5}$$

1.2.4 Reduction to Hermitian case

The computation of the pseudoinverse is reducible to its construction in the Hermitian case. This is possible through the equivalences:

- $A^+ = (A^*A)^+A^*$
- $A^+ = A^*(AA^*)^+$

as A^*A and AA^* are obviously Hermitian.

1.2.5 Products

If $A \in M(m,n;K)$, $B \in M(n,p;K)$ and either,

- A has orthonormal columns (i.e. $A^*A = I_n$) or,
- B has orthonormal rows (i.e. $BB^* = I_n$) or,
- A has all columns linearly independent (full column rank) and B has all rows linearly independent (full row rank) or,
- $B = A^*$ (i.e. B is the conjugate transpose of A),

then $(AB)^+ = B^+A^+$.

The last property yields the equivalences:

$$(AA^*)^+ = A^{+*}A^+$$
$$(A^*A)^+ = A^+A^{+*}$$

1.3 Basic Algorithms

The algorithms are divided into three categories: *optimization methods, greedy methods, and thresholding-based methods*.

1.3.1 Optimization Methods

An *optimization problem* is a problem of the type

$$\underset{x \in \mathbb{R}^N}{\text{minimize}}\, F_0(x) \quad \text{subject to } F_i(x) \le b_i,\ i \in [n],$$

where the function $F_0 : \mathbb{R}^N \to \mathbb{R}$ is called an *objective function* and the functions $F_1, \ldots, F_n : \mathbb{R}^N \to \mathbb{R}$ are called *constrained functions*. This general framework also encompasses equality constraints of the type $G_i(x) = c_i$. Since the equality $G_i(x) = c_i$ is equivalent to the inequalities $G_i(x) \le c_i$ and $-G_i(x) \le -c_i$. If $F_0, F_1, \ldots, F_n$ are all convex functions, then the problem is called a *convex optimization problem*. If $F_0, F_1, \ldots, F_n$ are all linear functions, then the problem is called a *linear program*. The sparse recovery problem is in fact an optimization problem, since it translates into

$$\text{minimize}\, \|z\|_0 \quad \text{subject to } \mathbf{A}z = y. \tag{P_0}$$

This is a nonconvex problem and NP-hard in general. However, keeping in mind that $\|z\|_q^q$ approaches $\|z\|_0$ as $q > 0$ tends to zero, we can approximate eq. (P_0) by the problem

$$\text{minimize}\, \|z\|_q \quad \text{subject to } \mathbf{A}z = y. \tag{P_q}$$

and *basis pursuit or ℓ_1-minimization*:

$$\text{minimize}\, \|z\|_1 \quad \text{subject to } \mathbf{A}z = y. \tag{P_1}$$

Basis pursuit

Input: measurement matrix $\mathbf{A}$, measurement vector y.

Instruction:

$$x^\sharp = \text{argmin}\, \|z\|_1 \quad \text{subject to } \mathbf{A}z = y. \tag{BP}$$

Output: the vector $x^\sharp$.

Theorem 1.3.1. *Let $\mathbf{A} \in \mathbb{R}^{m\times N}$ be a measurement matrix with columns $a_1, \dots, a_N$. Assuming the uniqueness of a minimizer $x^\sharp$ of*

$$\underset{z\in\mathbb{R}^N}{\text{minimize}}\, \|z\|_1 \quad \textit{subject to } \mathbf{A}z = y,$$

the system $\{a_j, j \in \operatorname{supp}(x^\sharp)\}$ is linearly independent, and in particular

$$\|x^\sharp\|_0 = \operatorname{card}(\operatorname{supp}(x^\sharp)) \le m.$$

A general ℓ_1-minimization taking measurement error into account:

$$\text{minimize}\, \|z\|_1 \quad \text{subject to } \|\mathbf{A}z - y\|_2 \le \eta \tag{$P_{1,\eta}$}$$

Given a vector $z \in \mathbb{C}^N$, we introduce its real and imaginary parts $\mathbf{u}, \mathbf{v} \in \mathbb{R}^N$ and a vector $\mathbf{c} \in \mathbb{R}^N$ such that $c_j \ge |z_j| = \sqrt{u_j^2 + v_j^2}$ for all $j \in [N]$. The problem eq. ($P_{1,\eta}$) is then equivalent to the following problem with optimization variables $\mathbf{c}, \mathbf{u}, \mathbf{v} \in \mathbb{R}^N$:

$$\underset{\mathbf{c},\mathbf{u},\mathbf{v}\in\mathbb{R}^N}{\text{minimize}} \sum_{j=1}^N c_j \quad \text{subject to } \left\| \begin{bmatrix} Re(\mathbf{A}) & -Im(\mathbf{A}) \\ Im(\mathbf{A}) & Re(\mathbf{A}) \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} \right\|_2 \le \eta, \tag{$P'_{1,\eta}$}$$

$$\sqrt{u_1^2 + v_1^2} \le c_1,$$
$$\vdots$$
$$\sqrt{u_N^2 + v_N^2} \le c_N.$$

This is an instance of a *second-order cone problem*;

Quadratically constrained basis pursuit

Input: measurement matrix $\mathbf{A}$, measurement vector y, noise level η.

Instruction:

$$x^\sharp = \operatorname{argmin} \|z\|_1 \qquad \text{subject to } \|\mathbf{A}z - y\|_2 \le \eta. \tag{BP_η}$$

Output: the vector $x^\sharp$.

The solution $x^\sharp$ of

$$\underset{z\in\mathbb{C}^N}{\text{minimize}}\, \|z\|_1 \qquad \text{subject to } \|\mathbf{A}z - y\|_2 \le \eta \tag{1.3.1}$$

for some parameter $\lambda \ge 0$,

$$\underset{z\in\mathbb{C}^N}{\text{minimize}}\, \lambda\|z\|_1 + \|\mathbf{A}z - y\|_2^2. \tag{1.3.2}$$

The solution of eq. (1.3.1) is also related to the output of the *LASSO*, which consists in solving, for some parameter $\tau \ge 0$,

$$\underset{z\in\mathbb{C}^N}{\text{minimize}}\, \|\mathbf{A}z - y\|_2 \qquad \text{subject to } \|z\|_1 \le \tau. \tag{1.3.3}$$

Precisely, some links between the three approaches are given below.

Proposition 1.3.1. *(a) If x is a minimizer of the basis pursuit denoising eq.* (1.3.2) *with $\lambda > 0$, then there exists $\eta = \eta_x \ge 0$ such that x is a minimizer of the quadratically constrained basis pursuit eq.* (1.3.1)*.*

(b) If x is unique minimizer of the quadratically constrained basis pursuit eq. (1.3.1) *with $\eta \ge 0$, then there exists $\tau = \tau_x \ge 0$ such that x is a unique minimizer of the LASSO eq.* (1.3.3)*.*

(c) If x is a minimizer of the LASSO eq. (1.3.3) *with $\tau \ge 0$, then there exists $\lambda = \lambda_x \ge 0$ such that x is a minimizer of the basis pursuit denoising eq.* (1.3.2)*.*

Another type of ℓ_1-minimization problem is the *Dantzig selector*,

$$\underset{z\in\mathbb{C}^N}{\text{minimize}}\, \|z\|_1 \qquad \text{subject to } \|\mathbf{A}^*(\mathbf{A}z - y)\|_\infty \le \tau. \tag{1.3.4}$$

This is again a convex optimization problem. The intuition for the constraints is that the residual $r = \mathbf{A}z - y$ should have small correlation with all columns a_j of the matrix $\mathbf{A}$—indeed, $\|\mathbf{A}^*(\mathbf{A}z - y)\|_\infty = \max_{j\in[N]} |\langle r, a_j\rangle|$.

1.3.2 Greedy Methods

Orthogonal matching pursuit (OMP)

Input: measurement matrix $\mathbf{A}$, measurement vector y.

Initialization: $S^0 = \emptyset, x^0 = 0$.

Iteration: repeat until a stopping criterion is met at $n = \bar{n}$:

$$S^{n+1} = S^n \cup \{j_{n+1}\}, j_{n+1} := \underset{j \in [N]}{\operatorname{argmax}} \left\{ |(\mathbf{A}^*(y - \mathbf{A}x^n))_j| \right\}, \tag{OMP_1}$$

$$x^{n+1} = \underset{z \in \mathbb{C}^N}{\operatorname{argmin}} \left\{ \|y - \mathbf{A}z\|_2, \operatorname{supp}(z) \subset S^{n+1} \right\}. \tag{OMP_2}$$

Output: the $\bar{n}$-sparse vector $x^\sharp = x^{\bar{n}}$.

The projection step eq. (OMP_2) is the most costly part of the orthogonal matching pursuit algorithm, which can be accelerated by using the QR-decomposition of $\mathbf{A}_{S_n}$.

The choice of the index j_{n+1} is dictated by a greedy strategy where one aims to reduce the ℓ_2-norm of the residual $y - \mathbf{A}x^n$ as much as possible at each iteration.

Lemma 1.3.1. *Let $\mathbf{A} \in \mathbb{C}^{m \times N}$ be a matrix with ℓ_2-normalized columns. Given $S \subset [N]$, v supported on S, and $j \in [N]$, if*

$$w := \underset{z \in \mathbb{C}^N}{\operatorname{argmin}} \left\{ \|y - \mathbf{A}z\|_2. \operatorname{supp}(z) \subset S \cup \{j\} \right\},$$

then

$$\|y - \mathbf{A}w\|_2^2 \leq \|y - \mathbf{A}v\|_2^2 - |(\mathbf{A}^*(y - \mathbf{A}v))_j|^2.$$

Proof. Since any vector of the form $v + te_j$ with $t \in \mathbb{C}$ is supported on $S \cup \{j\}$, we have

$$\|y - \mathbf{A}w\|_2^2 \leq \min_{t \in \mathbb{C}} \|y - \mathbf{A}(v + te_j)\|_2^2.$$

Writing $t = \rho e^{i\theta}$ with $\rho \geq 0$ and $\theta \in [0, 2\pi)$, we compute

$$\begin{aligned}
\|y - \mathbf{A}(v + te_j)\|_2^2 &= \|y - \mathbf{A}v - t\mathbf{A}e_j\|_2^2 \\
&= \|y - \mathbf{A}v\|_2^2 + |t|^2 \|\mathbf{A}e_j\|_2^2 - 2\operatorname{Re}\left(\bar{t}\langle y - \mathbf{A}v, \mathbf{A}e_j\rangle\right) \\
&= \|y - \mathbf{A}v\|_2^2 + \rho^2 - 2\operatorname{Re}\left(\rho e^{-i\theta}(\mathbf{A}^*(y - \mathbf{A}v))_j\right) \\
&\geq \|y - \mathbf{A}v\|_2^2 + \rho^2 - 2\rho\,|(\mathbf{A}^*(y - \mathbf{A}v))_j|,
\end{aligned}$$

with equality for a property chosen θ. As a quadratic polynomial in ρ, the latter expression is minimized when $\rho = |(\mathbf{A}^*(y - \mathbf{A}u))_j|$. This shows that

$$\min_{t \in \mathbb{C}} \|y - \mathbf{A}(v + te_j)\|_2^2 = \|y - \mathbf{A}v\|_2^2 - |(\mathbf{A}^*(y - \mathbf{A}u))_j|^2,$$

which concludes the proof. $\square$

The step eq. (OMP_2) also reads as

$$x^{n+1}_{S^{n+1}} = \mathbf{A}^\dagger_{S^{n+1}} y,$$

where $x^{n+1}_{S^{n+1}}$ denotes the restriction of x^{n+1} to its support set S^{n+1} and where $\mathbf{A}^\dagger_{S^{n+1}}$ is the pseudo-inverse of $\mathbf{A}_{S^{n+1}}$. This simply says that $z = x^{n+1}_{S^{n+1}}$ is a solution of $\mathbf{A}^{n+1}_{S^{n+1}} \mathbf{A}_{S^{n+1}} z = \mathbf{A}^*_{S^{n+1}} y$.

Lemma 1.3.2. *Given an index set $S \subset [N]$, if*

$$v := \underset{z \in \mathbb{C}^N}{\operatorname{argmin}} \left\{ \|y - \mathbf{A}z\|_2, \operatorname{supp}(z) \subset S \right\},$$

then

$$(\mathbf{A}^*(y - \mathbf{A}v))_S = 0. \tag{1.3.5}$$

Proposition 1.3.2. *Given a matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$, every nonzero vector $x \in \mathbb{C}^N$ supported on a set S of size s is recovered from $y = \mathbf{A}x$ after at most s iterations of orthogonal matching pursuit if and only if the matrix $\mathbf{A}_S$ is injective and*

$$\max_{j \in S} |(\mathbf{A}^* r)_j| > \max_{\ell \in \overline{S}} |(\mathbf{A}^* r)_\ell| \tag{1.3.6}$$

for all nonzero $r \in \{\mathbf{A}z, \operatorname{supp}(z) \subset S\}$.

Remark 1. *A more concise way to formulate the necessary and sufficient conditions of proposition 1.3.2 is the exact recovery condition, which reads*

$$\|\mathbf{A}_S^\dagger \mathbf{A}_{\overline{S}}\|_{1 \to 1} < 1; \tag{1.3.7}$$

see A.1 for the definition of matrix norms. Implicitly, the existence of the pseudo-inverse $\mathbf{A}_S^\dagger = (\mathbf{A}_S^* \mathbf{A}_S)^{-1} \mathbf{A}_S^*$ is equivalent to the injectivity of $\mathbf{A}_S$.

Compressive sampling matching pursuit (CoSaMP)

Input: measurement matrix $\mathbf{A}$, measurement vector y, sparsity level s.

Initialization: s-sparse vector x^0, typically $x^0 = 0$.

Iteration: repeat until a stopping criterion is met at $n = \bar{n}$:

$$U^{n+1} = \operatorname{supp}(x^n) \cup L_{2s}(\mathbf{A}^*(y - \mathbf{A}x^n)), \tag{CoSaMP$_1$}$$

$$u^{n+1} = \operatorname*{argmin}_{z \in \mathbb{C}^N} \left\{ \|y - \mathbf{A}z\|_2, \operatorname{supp}(z) \subset U^{n+1} \right\}, \tag{CoSaMP$_2$}$$

$$x^{n+1} = H_s(u^{n+1}). \tag{CoSaMP$_3$}$$

Output: the s-sparse vector $x^\sharp = x^{\bar{n}}$.

1.3.3 Thresholding-Based Methods

Basic thresholding

Input: measurement matrix $\mathbf{A}$, measurement vector y, sparsity level s.

Instruction:

$$S^\sharp = L_s(\mathbf{A}^* y), \tag{BT$_1$}$$

$$x^\sharp = \operatorname*{argmin}_{z \in \mathbb{C}^N} \left\{ \|y = \mathbf{A}z\|_2, \operatorname{supp}(z) \subset S^\sharp \right\}. \tag{BT$_2$}$$

Output: the s-sparse vector $x^\sharp$.

Proposition 1.3.3. *A vector $x \in \mathbb{C}^N$ supported on a set S is recovered from $y = \mathbf{A}x$ via basic thresholding if and only if*

$$\min_{j \in S} |(\mathbf{A}^* y)_j| > \max_{\ell \in \overline{S}} |(\mathbf{A}^* y)_\ell|. \tag{1.3.8}$$

Proof. It is clear that the vector x is recovered if and only if the index set $S^\sharp$ defined in eq. (BT$_1$) coincides with the set S, that is to say, if and only if any entry of $\mathbf{A}^* y$ on S is greater than any entry of $\mathbf{A}^* y$ on $\overline{S}$. This is property eq. (1.3.8). □

The more elaborate *iterative hard thresholding algorithm* is an iterative algorithm to solve the rectangular system $\mathbf{A}z = y$, knowing that the solution is s-sparse. We shall solve the square system $\mathbf{A}^*\mathbf{A}z = \mathbf{A}^* y$ instead, which can be interpreted as the fixed-point equation $z = (\mathbf{Id} - \mathbf{A}^*\mathbf{A})z + \mathbf{A}^* y$. Classical iterative methods suggest the *fixed-point iteration* $x^{n+1} = (\mathbf{Id} - \mathbf{A}^*\mathbf{A})x^n + \mathbf{A}^* y$. Since we target s-sparse vectors, we only keep the *s largest absolute entries of $(\mathbf{Id} - \mathbf{A}^*\mathbf{A})x^n + \mathbf{A}^* y = x^n + \mathbf{A}^*(y - \mathbf{A}x^n)$ at each iteration*. The resulting algorithm:

Iterative hard thresholding (HIT)

Input: measurement matrix $\mathbf{A}$, measurement vector y, sparsity level s.

Initialization: s-sparse vector x^0, typically $x^0 = 0$.

Iteration: repeat until a stopping criterion is met at $n = \bar{n}$:

$$x^{n+1} = H_s(x^n + \mathbf{A}^*(y - \mathbf{A}x^n)). \tag{IHT}$$

Output: the s-sparse vector $x^\sharp = x^{\bar{n}}$.

The above algorithm does not require computation of any orthogonal projection. If the orthogonal projection needs to be paid attention, like in the greedy methods, *it makes sense to look at the vector with the same support as x^{n+1} that best fits the measurements*. This leads to the *hard thresholding pursuit algorithm*:

Hard thresholding pursuit (HTP)

Input: measurement matrix $\mathbf{A}$, measurement vector y, sparsity level s.

Initialization: s-sparse vector x^0, typically $x^0 = 0$.

Iteration: repeat until stopping criterion is met an $n = \bar{n}$:

$$S^{n+1} = L_s(x^n + \mathbf{A}^*(y = \mathbf{A}x^n)), \tag{HTP$_1$}$$

$$x^{n+1} = \operatorname*{argmin}_{z \in \mathbb{C}^N} \left\{ \|y - \mathbf{A}z\|_2, \operatorname{supp}(z) \subset S^{n+1} \right\}. \tag{HTP$_2$}$$

Output: the s-sparse vector $x^\sharp = x^{\bar{n}}$.

Subspace Pursuit

Input: measurement matrix $\mathbf{A}$, measurement vector y, sparsity level s.

Initialization: s-sparse vector x^0, typically $x^0 = 0, S^0 = \operatorname{supp}(x^0)$.

Iteration: repeat until a stopping criterion is met at $n = \bar{n}$:

$$U^{n+1} = S^n \cup L_s(\mathbf{A}^*(y - \mathbf{A}x^n)), \tag{SP$_1$}$$

$$u^{n+1} = \operatorname*{argmin}_{z \in \mathbb{C}^N} \left\{ \|y - \mathbf{A}z\|_2, \operatorname{supp}(z) \subset U^{n+1} \right\}, \tag{SP$_2$}$$

$$S^{n+1} = L_s(u^{n+1}), \tag{SP$_3$}$$

$$x^{n+1} = \operatorname*{argmin}_{z \in \mathbb{C}^N} \left\{ \|y - \mathbf{A}z\|_2, \operatorname{supp}(z) \subset S^{n+1} \right\}.. \tag{SP$_4$}$$

Output: the s-sparse vector $x^\sharp = x^{\bar{n}}$.

1.3.4 Notes

Which Algorithm Should One Choose?

- The minimal number m of measurements for a sparsity s and a signal length N may vary with each algorithm. Comparing the recovery rates and identifying the best algorithm is a matter of numerical tests. For this criterion, the recovery performance of *basic thresholding* is significantly worse than the one of other algorithms, although it is the fastest algorithm since it identifies the support in only one step.

- The speed of the algorithm is also a criterion, and it is also a matter of numerical tests. If the sparsity s is quite small, then *orthogonal matching pursuit* is extremely fast because the speed essentially depends on *the number of iterations, which typically equals s when the algorithm succeeds. Compressive sampling matching pursuit and hard thresholding pursuit* are fast for small s, because each step involves the computation of an orthogonal projection relative to $\mathbf{A}_S$ with small $S \subset [N]$. But if the sparsity s is not that small compared to N, then orthogonal matching pursuit may nonetheless require a significant time. The same applies to the homotopy method which builds the support of an ℓ_1-minimizer iteratively. The runtime of iterative hard thresholding is almost not influenced by the sparsity s at all.

Basis pursuit section 1.3.1, per se, is not an algorithm, so the runtime depends on the actual algorithm that is used for the minimization. For *Chambolle and Pock's primal dual algorithm*, which constructs a sequence converging to an ℓ_1-minimizer, the sparsity s has no serious influence on the speed. Hence, for mildly large s, it can be significantly faster than orthogonal matching pursuit. A point of view of regarding the orthogonal matching pursuit are always faster than ℓ_1-minimization — is only true for small sparsity. The iteratively reweighted least squares method may also be a good alternative for mildly large sparsity.

- Another Important feature of an algorithm is the *possibility to exploit fast matrix-vector multiplication routines* that are available for $\mathbf{A}$ and $\mathbf{A}^*$. In principle, any of the proposed methods can be sped up in this case, but the task is complicated if *orthogonal projection* steps are involved. Fast matrix-vector multiplications are easily integrated in the *iterative hard thresholding algorithm and in Chambolle and Pock's primal dual algorithm for ℓ_1-minimization.* The acceleration achieved in this context depends on the algorithm, and the fastest algorithm should again be determined by numerical tests in the precise situation.

1.4 Basis Pursuit

1.4.1 Null Space Property

The null space property is a necessary and sufficient condition for exact recovery of sparse vectors via basis pursuit.

Definition 1.4.1. *A matrix $\mathbf{A} \in \mathbb{K}^{m \times N}$ is said to satisfy the null space property relative to a set $S \subset [N]$ if*

$$\|v_S\|_1 < \|v_{\bar{S}}\|_1 \quad \textit{for all } v \in \ker \mathbf{A} \setminus \{0\}. \tag{1.4.1}$$

It is to satisfy the null space property of order s if it satisfies the null space property relative to any set $S \subset [N]$ with $\operatorname{card}(S) \leq s$.

Remark 2. *It is important to observe that, for a given $v \in \ker \mathbf{A} \setminus \{0\}$, the condition $\|v_S\|_1 < \|v_{\bar{S}}\|_1$ holds for any set $S \subset [N]$ with $card(S) \leq s$ as soon as it holds for an index set of s largest (in modulus) entries of v.*

Definition 1.4.2. *For $p > 0$, the ℓ_p-error of best s-term approximation to a vector $x \in \mathbb{C}^N$ is defined by*

$$\sigma_s(x)_p := \inf\left\{\|x - z\|_p, z \in \mathbb{C}^N \textit{ is } s\textit{-sparse}\right\}.$$

Remark 3. *There are two convenient reformulations of the null space property. The first one is obtained by adding $\|v_S\|_1$ to both sides of the inequality $\|v_S\|_1 < \|v_{\bar{S}}\|_1$. Thus, the null space property relative to S reads*

$$2\|v_S\|_1 < \|v\|_1 \quad \textit{for all } v \in \ker \mathbf{A} \setminus \{0\}. \tag{1.4.2}$$

The second one is obtained by choosing S as an index set of s largest (in modulus) entries of v and this time by adding $\|v_{\bar{S}}\|_1$ to both sides of the inequality. Thus, the null space property of order s reads

$$\|v\|_1 < 2\sigma_s(v)_1 \quad \textit{for all } v \in \ker \mathbf{A} \setminus \{0\}, \tag{1.4.3}$$

where we recall from definition 1.4.2 that, for $p > 0$, the ℓ_p-error of best s-term approximation *to $x \in \mathbb{K}^N$ is defined by*

$$\sigma_s(x)_p = \inf_{\|z\|_0 \leq s} \|x - z\|_p.$$

Theorem 1.4.1. *Given a matrix $\mathbf{A} \in \mathbb{K}^{m \times N}$, every vector $x \in \mathbb{K}^N$ supported on a set S is the unique solution of eq. (P_1) with $y = \mathbf{A}x$ if and only if $\mathbf{A}$ satisfies the null space property relative to S.*

Remark 4. *(a) This theorem theorem 3.2.2 shows that for every $y = \mathbf{A}x$ with s-sparse vector x the ℓ_1-minimization strategy eq. (P_1) actually solves the ℓ_0-minimization problem eq. (P_0) when the null space property of order s holds. Indeed, assume that every s-sparse vector x is recovered via ℓ_1-minimization from $y = \mathbf{A}x$. Let z be the minimizer of the ℓ_0-minimization problem eq. (P_0) with $y = \mathbf{A}x$ then $\|z\|_0 \leq \|x\|_0$ so that also z is s-sparse. But since every s-sparse vector is the unique ℓ_1-minimizer, it follows that $x = z$.*

(b) *It is desirable for any reconstruction scheme to preserve sparse recovery if some measurements are rescaled, reshuffled, or added. Basis pursuit actually features such properties. Indeed, mathematically speaking, these operations consist in replacing the original measurement matrix* $\mathbf{A}$ *by new measurement matrices* $\hat{\mathbf{A}}$ *and* $\hat{\mathbf{A}}$ *defined by*

$$\hat{\mathbf{A}} := \mathbf{G}\mathbf{A}, \quad \textit{where } \mathbf{G} \textit{ is some invertible } m \times m \textit{ matrix},$$

$$\hat{\mathbf{A}} := \begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix}, \quad \textit{where } \mathbf{B} \textit{ is some } m' \times N \textit{ matrix}.$$

One can observe that $\ker\hat{\mathbf{A}} = \ker\mathbf{A}$ *and* $\ker\hat{\mathbf{A}} \subset \ker\mathbf{A}$*, hence the null space property for the matrices* $\hat{\mathbf{A}}$ *and* $\hat{\mathbf{A}}$ *remains fulfilled if it is satisfied for the matrix* $\mathbf{A}$*. It is not true that the null space property remains valid if we multiply on the right by an invertible matrix.*

The *real null space relative to a set S*:

$$\sum_{j\in S} |v_j| < \sum_{\ell\in\bar{S}} |v_\ell| \quad \text{for all } v \in \ker_{\mathbb{R}}\mathbf{A}, v \neq 0, \tag{1.4.4}$$

and, on the other hand, to the *complex null space property relative to S*:

$$\sum_{j\in S} \sqrt{v_j^2 + w_j^2} < \sum_{\ell\in\bar{S}} \sqrt{v_\ell^2 + w_\ell^2} \quad \text{for all } v, w \in \ker_{\mathbb{R}}\mathbf{A}, (v, w) \neq (0, 0). \tag{1.4.5}$$

The real and complex version are in fact equivalent. These vectors can be interpreted as real or as complex vectors. The followed theorem explains why we usually work in the complex setting.

Theorem 1.4.2. *Given a matrix* $\mathbf{A} \in \mathbb{R}^{m\times N}$*, the real null space property eq.* (1.4.4) *relative to a set* S *is equivalent to the complex null space property eq.* (1.4.5) *relative to this set* S*.*

In particular, the real null space property of order s *is equivalent to the complex null space property of order* s*.*

Proof. eq. (1.4.4) immediately follows from eq. (1.4.5) by setting $w = 0$. So one can assume that eq. (1.4.4) holds. We consider $v, w \in \ker_{\mathbb{R}\mathbf{A}}$ with $(v, w) \neq (0, 0)$. If v and w are linearly dependent, then the inequality $\sum_{j\in S} \sqrt{v_j^2{}_+ w_j^2} < \sum_{\ell\in\bar{S}} \sqrt{v_\ell^2 + w_\ell^2}$ is clear, sow we suppose that they are linearly independent. Then $u := \cos\theta v + \sin\theta w \in \ker_{\mathbb{R}}\mathbf{A}$ is nonzero, and eq. (1.4.4) yields, for any $\theta \in \mathbb{R}$,

$$\sum_{j\in S} |\cos\theta v_j + \sin\theta w_j| < \sum_{\ell\in\bar{S}} |\cos\theta v_\ell + \sin\theta w_\ell| \,. \tag{1.4.6}$$

For each $k \in [N]$, we define $\theta_k \in [-\pi, \pi]$ by the equalities

$$v_k = \sqrt{v_k^2 + w_k^2}\cos\theta_k, \qquad w_k = \sqrt{v_k^2 + w_k^2}\sin\theta_k,$$

so that eq. (1.4.6) reads

$$\sum_{j\in S} \sqrt{v_j^2 + w_j^2}\,|\cos(\theta - \theta_j)| < \sum_{\ell\in\bar{S}} \sqrt{v_\ell^2 + w_\ell^2}\,|\cos(\theta - \theta_\ell)| \,.$$

We now integrate over $\theta \in [-\pi, \pi]$ to obtain

$$\sum_{j\in S} \sqrt{v_j^2 + w_j^2} \int_{-\pi}^{\pi} |\cos(\theta - \theta_j)|\, d\theta < \sum_{\ell\in\bar{S}} \sqrt{v_\ell^2 + w_\ell^2} \int_{-\pi}^{\pi} |\cos(\theta - \theta_\ell)|\, d\theta.$$

For the inequality $\sum_{j\in S} \sqrt{v_j^2 + w_j^2} < \sum_{\ell\in\bar{S}} \sqrt{v_\ell^2 + w_\ell^2}$, it remains to observe that

$$\int_{-\pi}^{\pi} |\cos(\theta - \theta')|\, d\theta$$

is a positive constant independent of $\theta' \in [-\pi, \pi]$—namely,4. □

Nonconvex Minimization

The number of nonzero entries of a vector $z \in \mathbb{C}^N$ is approximated by the qth power of its ℓ_q-quasinorm,

$$\sum_{j=1}^{N} |z_j|^q \xrightarrow[q \to 0]{} \sum_{j=1}^{N} 1_{\{z_j \neq 0\}} = \|z\|_0.$$

This observation suggests to replace the ℓ_0-minimization problem eq. (P_0) by the optimization problem

$$\underset{z \in \mathbb{C}^N}{\text{minimize}} \, \|z\|_q \quad \text{subject to } \mathbf{A}z = y. \tag{P$_q$}$$

The problem eq. (P_q) does not provide a worse approximation of the original problem eq. (P_0) when q gets smaller even though it becomes nonconvex and even NP-hard. An analog is needed to justify the null space property for $0 < q < 1$.

Theorem 1.4.3. *Given a matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$ and $0 < q \leq 1$, every s-sparse vector $x \in \mathbb{C}^N$ is the unique solution of eq. (P_q) with $y = \mathbf{A}x$ if and only if, for any set $S \subset [N]$ with card$(S) \leq s$,*

$$\|v_S\|_q < \|v_{\bar{S}}\|_q \quad \textit{for all } v \in \ker\mathbf{A} \backslash \{0\}.$$

Theorem 1.4.4. *Given a matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$ and $0 < p < q \leq 1$, if every s-sparse vector $x \in \mathbb{C}^N$ is the unique solution of eq. $(\mathrm{P}_q)(P_q)$ with $y = \mathbf{A}x$, then every s-sparse vector $x \in \mathbb{C}^N$ is also the unique solution of (P_p) with $y = \mathbf{A}x$.*

1.4.2 Stability

The vectors we aim to recover via basis pursuit–or other schemes, for that matter–are sparse only in idealized situations. In more realistic scenarios, we can only claim that they are close to sparse vectors. In such cases, we would like to recover a vector $x \in \mathbb{C}^N$ with an error controlled by its distance to s-sparse vectors. This property: *stability* of the reconstruction scheme with respect to sparsity defect.

Definition 1.4.3. *A matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$ is said to satisfy the stable null space property with constant $0 < \rho < 1$ relative to a set $S \subset [N]$ if*

$$\|v_S\|_1 \leq \rho \|v_{\bar{S}}\|_1 \quad \textit{for all } v \in \ker\mathbf{A}.$$

It is to satisfy the stable null space property of order s with constant $0 < \rho < 1$ if it satisfies the stable null space property with constant $0 < \rho < 1$ relative to any set $S \subset [N]$ with card$(S) \leq s$.

Theorem 1.4.5. *Suppose that a matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$ satisfies the stable null space property of order s with constant $0 < \rho < 1$. Then, for any $x \in \mathbb{C}^N$, a solution $x^\sharp$ of eq. (P_1) with $y = \mathbf{A}x$ approximates the vector x with ℓ_1-error*

$$\|x - x^\sharp\|_1 \leq \frac{2(1+\rho)}{(1-\rho)} \sigma_s(x)_1. \tag{1.4.7}$$

Remark 5. *In contrast to theorem 1.4.1 we cannot guarantee uniqueness of the ℓ_1-minimizer anymore—although nonuniqueness is rather pathological. In any case, even when the ℓ_1-minimizer is not unique, the theorem above states that* every solution $x^\sharp$ of eq. (P_1) with $y = \mathbf{A}x$ satisfies eq. (1.4.7).

Apart from improving theorem 1.4.5, the result also says that, under the stable null space property relative to S, the distance between a vector $x \in \mathbb{C}^N$ supported on S and a vector $z \in \mathbb{C}^N$ satisfying $\mathbf{A}z = \mathbf{A}x$ is controlled by the difference between their norms.

Theorem 1.4.6. *The matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$ satisfies the stable null space property with constant $0 < \rho < 1$ relative to S if and only if*

$$\|z - x\|_1 \leq \frac{1+\rho}{1-zr} \left(\|z\|_1 - \|x\|_1 + 2\|x_{\bar{S}}\|_1 \right) \tag{1.4.8}$$

for all vectors $x, z \in \mathbb{C}^N$ with $\mathbf{A}z = \mathbf{A}x$.

The error bound eq. (1.4.7) follows from theorem 1.4.6 as follows: Take S to be a set of s largest absolute coefficients of x, so that $\|x_{\bar{S}}\|_1 = \sigma_s(x)_1$. If $x^\sharp$ is a minimizer of eq. (P_1), then $\|x^\sharp\|_1 \leq \|x\|_1$ and $\mathbf{A}x^\sharp = \mathbf{A}x$. The right-hand side of inequality eq. (1.4.8) with $z = x^\sharp$ can therefore be estimated by the right hand of eq. (1.4.7).

Lemma 1.4.1. *Given a set $S \subset [N]$ and vectors $x, z \in \mathbb{C}^N$,*

$$\|(x-z)_{\overline{S}}\|_1 \leq \|z\|_1 - \|x\|_1 + \|(x-z)_S\|_1 + 2\|x_{\overline{S}}\|_1$$

We assume that the matrix $\mathbf{A}$ satisfies eq. (1.4.8) for all vectors $x, z \in \mathbb{C}^N$ with $\mathbf{A}z = \mathbf{A}x$. Given a vector $v \in \ker\mathbf{A}$, since $\mathbf{A}v_{\overline{S}} = \mathbf{A}_{(-v_S)}$, we can apply eq. (1.4.8) with $x = -v_S$ and $z = v_{\overline{S}}$. It yields

$$\|v\|_1 \leq \frac{1+\rho}{1-\rho}\left(\|v_{\overline{S}}\|_1 - \|v_S\|_1\right).$$

This can be written as

$$(1-\rho)(\|v_S\|_1 + \|v_{\overline{S}}\|_1) \leq (1+\rho)(\|v_{\overline{S}}\|_1 - \|v_S\|_1).$$

After rearranging:

$$\|v_S\|_1 \leq \rho\|v_{\overline{S}}\|_1,$$

and we recognize the stable null space property with constant $0 < \rho < 1$ relative to S.

Conversely, we now assume that the matrix $\mathbf{A}$ satisfies the stable null space property with constant $0 < \rho < 1$ relative to S. For $x, z \in \mathbb{C}^N$ with $\mathbf{A}z = \mathbf{A}x$, since $v := z - x \in \ker\mathbf{A}$, the stable null space property yields

$$\|v_S\|_1 \leq \rho\|v_{\overline{S}}\|_1 \tag{1.4.9}$$

Moreover, lemma 1.4.1 gives

$$\|v_{\overline{S}}\|_1 \leq \|z\|_1 - \|x\|_1 + \|v_S\|_1 + 2\|x_{\overline{S}}\|_1. \tag{1.4.10}$$

Substituting eq. (1.4.9) into eq. (1.4.10):

$$\|v_{\overline{S}}\|_1 \leq \|z\|_1 - \|x\|_1 + \rho\|v_{\overline{S}}\|_1 + 2\|x_{\overline{S}}\|_1.$$

Since $\rho < 1$, this can be rewritten as:

$$\|v_{\overline{S}}\|_1 \leq \frac{1}{1-\rho}\left(\|z\|_1 - \|x\|_1 + 2\|x_{\overline{S}}\|_1\right),$$

Using eq. (1.4.9) again:

$$\|v_1\| = \|v_{\overline{S}}\|_1 + \|v_S\|_1 \leq (1+\rho)\|v_{\overline{S}}\|_1 \leq \frac{1+\rho}{1-\rho}\left(\|z\|_1 - \|x\|_1 + 2\|x_{\overline{S}}\|_1\right),$$

which proves the theorem 1.4.6.

Remark 6. *Given the matrix $\mathbf{A} \in \mathbb{C}^{m\times N}$, we consider, for each index set $S \subset [N]$ with card$(S) \leq s$, the operator $\mathbf{R}_S$ defined on ker$\mathbf{A}$ by $\mathbf{R}_S(v) = v_S$. The formulation eq. (1.4.2) of the null space property says that*

$$\mu := \max\left\{\|\mathbf{R}_S\|_{1\to1} : S \subset [N], card(S) \leq s\right\} < 1/2.$$

It then follows that $\mathbf{A}$ satisfies the stable null space property with constant $\rho := \mu/(\bigcap \mu) < 1$. Thus, the stability of the basis pursuit comes for free if sparse vectors are exactly recovered. However, the constant $1(1+\rho)/(1-\rho)$ in eq. (1.4.7) may be very large if ρ is close to one.

1.4.3 Robustness

In realistic situations, it is also inconceivable to measure a signal $x \in \mathbb{C}^N$ with infinite precision. This means that the measurement vector $y \in \mathbb{C}^m$ is only an approximation of the vector $\mathbf{A}x \in \mathbb{C}^m$, with

$$\|\mathbf{A}x - y\| \leq \eta$$

for some $\eta \geq 0$ and for some norm $\|\cdot\|$ on $\mathbb{C}^m$—usually the ℓ_2-norm, but the ℓ_1-norm will also be considered. In this case, the reconstruction scheme should be required to output a vector $x^* \in \mathbb{C}^N$ whose distance to the original vector $x \in \mathbb{C}^N$ is controlled by the measurement error $\eta \geq 0$, which is referred to as the *robustness* of the reconstruction scheme with respect to measurement error. Then the eq. (P_1) is replaced by eq. $(P_{1,\eta})$.

Definition 1.4.4. *The matrix $\mathbf{A} \in \mathbb{C}^{m\times N}$ is said to satisfy the robust null space property (with respect to $\|\cdot\|$) with constants $0 < \rho < 1$ and $\tau > 0$ relative to a set $S \subset [N]$ if*

$$\|v_S\|_1 \leq \rho\|v_{\overline{S}}\|_1 + \tau\|\mathbf{A}v\| \quad \textit{for all } v \in \mathbb{C}^N. \tag{1.4.11}$$

Remark 7. *Observe that the above definition does not require that v is contained in ker$\mathbf{A}$. In fact, if $v \in$ ker$\mathbf{A}$, then the term $\|\mathbf{A}v\|$ in eq. (1.4.11) vanishes, and we see that the robust null space property implies the stable null space property in definition 1.4.3.*

The theorem followed constitutes the first main result of this section. It incorporates the conclusion of theorem 1.4.5 as the special case $\eta = 0$. The special case of an s-sparse vector $x \in \mathbb{C}^N$ is also worth a separate look.

Theorem 1.4.7. *Suppose that a matrix $\mathbf{A} \in \mathbb{C}^{m\times N}$ satisfies the robust null space property of order s with constants $0 < \rho < 1$ and $\tau > 0$. Then, for any $x \in \mathbb{C}^N$, a solution $x^\sharp$ of eq. $(P_{1,\eta})$ with $y = \mathbf{A}x + e$ and $\|e\| \leq \eta$ approximates the vector x with ℓ_1-error*

$$\|x - x^\sharp\|_1 \leq \frac{2(1+\rho)}{(1-\rho)}\sigma_s(x)_1 + \frac{4\tau}{1-\rho}\eta.$$

A stronger "if and only if" statement valid for any index set S is proved as follows:

Theorem 1.4.8. *The matrix $\mathbf{A} \in \mathbb{C}^{m\times N}$ satisfies the robust null space property with constants $0 < \rho < 1$ and $\tau > 0$ relative to S if and only if*

$$\|z - x\|_1 \leq \frac{1+\rho}{1-\rho}(\|z\|_1 - \|x\|_1 + 2\|x_{\overline{S}}\|_1) + \frac{2\tau}{1-\rho}\|\mathbf{A}(z-x)\| \tag{1.4.12}$$

for all vectors $x, z \in \mathbb{C}^N$.

The second main result enhances the previous robustness result by *replacing the ℓ_2-error estimate by and ℓ_p-error estimate for $p \geq 1$*. A final strengthening of the null space property is required. The corresponding property could be defined relative to any fixed set $S \subset [N]$, but it is not introduced as such because this will not be needed later.

Definition 1.4.5. *Given $q \geq 1$, the matrix $\mathbf{A} \in \mathbb{C}^{m\times N}$ is said to satisfy the* ℓ_q-robust null space property *of order s (with respect to $\|\cdot\|$) with constants $0 < \rho < 1$ and $\tau > 0$ if, for any set $S \subset [N]$ with card$(S) \leq s$,*

$$\|v_S\|_q \leq \frac{\rho}{s^{1-1/q}}\|V_{\bar{S}}\|_1 + \tau\|\mathbf{A}v\| \quad \textit{for all } v \in \mathbb{C}^N.$$

In view of the *inequality $\|v_S\|_p \leq s^{1/p-1/q}\|v_S\|_q$ for $1 \leq p \leq q$*, we observe that the ℓ_p-robust null space property with constants $0 < \rho < 1$ and $\tau > 0$ implies that, for any set $S \subset [N]$ with card$(S) \leq s$,

$$\|v_S\|_p \leq \frac{\rho}{s^{1-1/p}}\|v_{\overline{S}}\|_1 + \tau s^{1/p-1/q}\|\mathbf{A}v\| \quad \text{for all } v \in \mathbb{C}^N.$$

Thus, for $1 \leq p \leq q$, the ℓ_q-robust null space property implies the ℓ_p-robust null space property with identical constants, modulo the change of norms $\|\cdot\| \leftarrow s^{1/p-1/q}\|\cdot\|$. This justifies in particular that the ℓ_q-robust null space property is a strengthening of the previous robust null space property. The robustness of the quadratically constrained basis pursuit algorithm is then deduced according to the following theorem:

Theorem 1.4.9. *Suppose that the matrix $\mathbf{A} \in \mathbb{C}^{m\times N}$ satisfies the ℓ_2-robust null space property of order s with constants $0 < \rho < 1$ and $\tau > 0$. Then, for any $x \in \mathbb{C}^N$, a solution $x^\sharp$ of eq. $(P_{1,\eta})$ with $\|\cdot\| = \|\cdot\|_2, y = \mathbf{A}x + e$, and $\|e\|_2 \leq \eta$ approximates the vector x with ℓ_p-error*

$$\|x - x^\sharp\|_p \leq \frac{C}{s^{1-1/p}}\sigma_s(x)_1 + Ds^{1/p-1/2}\eta, \qquad 1 \leq q \leq 2, \tag{1.4.13}$$

for some constants $C, D > 0$ depending only on ρ and τ.

The estimates for the extremal values $p = 1$ and $p = 2$ are the most familiar. They read

$$\begin{aligned} \|x - x^\sharp\|_1 &\leq C\sigma_s(x)_1 + D\sqrt{s}\eta, \\ \|x - x^\sharp\|_2 &\leq \frac{C}{\sqrt{s}}\sigma_s(x)_1 + D\eta. \end{aligned} \tag{1.4.14}$$

The coefficient of $\sigma_s(x)_1$ is a constant for $p = 1$ and scales like $1/\sqrt{s}$ for $p = 2$, while the coefficient of η scales like $\sqrt{s}$ for $p = 1$ and is a constant for $p = 2$.

Remark 8. *Let us comment on the fact that, regardless of the ℓ_p-sparse in which the error is estimated, the best s-term approximation error $\sigma_s(x)_1$ with respect to the ℓ_1-norm always appears on the right-hand side. One may wonder why* the error estimate in ℓ_2 does not involve $\sigma_s(x)_1/\sqrt{s}$. *In fact, we will see in theorem 1.4.10 that such an estimate is impossible in parameter regimes of (m, N) that are interesting for Compressive sensing. Besides, we have seen that unit ℓ_p-balls with $q < 1$ provide good models for compressible vectors by virtue of theorem 1.4.11 and its refinement theorem 1.4.12. Indeed, if $\|x\|_q \leq 1$ for $q < 1$, then, for $p \geq 1$,*

$$\sigma_s(x)_p \leq s^{1/p-1/q}.$$

Theorem 1.4.10. *If a pair of measurement matrix and reconstruction map is ℓ_2-instance optimal of order $s \geq 1$ with constant C, then*

$$m \geq cN \tag{1.4.15}$$

for some constant c depending only on C.

Theorem 1.4.11. *For any $q > p > 0$ and any $x \in \mathbb{C}^N$,*

$$\sigma_s(x)_q \leq \frac{1}{s^{1/p-1/q}}\|x\|_p.$$

Theorem 1.4.12. *For any $q > p > 0$ and any $x \in \mathbb{C}^N$, the inequality*

$$\sigma_s(x)_q \leq \frac{c_{p,q}}{s^{1/p-1/q}}\|x\|_p$$

holds with

$$c_{p,q} := \left[\left(\frac{p}{q}\right)^{p/q}\left(1 - \frac{p}{q}\right)^{1-p/q}\right]^{1/p} \leq 1.$$

Assuming perfect measurements (that is, η), the error bound eq. (1.4.13) yields

$$\|x - x^\sharp\|_p \leq \frac{C}{s^{1-1/p}}\sigma_s(x)_1 \leq Cs^{1/p-1/q}, \quad 1 \leq p \leq 2.$$

Remark 9. *The ℓ_q-robust null space property may seem mysterious at first sight, but it is necessary—save for the condition $\rho < 1$—to obtain estimates of the type*

$$\|x - x^\sharp\|_q \leq \frac{C}{s^{1-1/q}}\sigma_s(x)_1 + D\eta, \tag{1.4.16}$$

where $x^\sharp$ is a minimizer of eq. $(P_{1,\eta})$ with $y = \mathbf{A}x + e \leq \eta$. Indeed, given $v \in \mathbb{C}^N$ and $S \subset [N]$ with $\operatorname{card}(S) \leq s$, we apply eq. (1.4.16) with $x = v, e = -\mathbf{A}v$ and $\eta = \|\mathbf{A}v\|$, so that $x^\sharp = 0$, to obtain

$$\|v\|_q \leq \frac{C}{s^{1-1/q}}\|v_{\overline{S}}\|_1 + D\|\mathbf{A}v\|,$$

and in particular

$$\|v_S\|_q \leq \frac{C}{s^{1-1/q}}\|v_{\overline{S}}\|_1 + D\|\mathbf{A}v\|.$$

Theorem 1.4.13. *Given $1 \leq p \leq q$, suppose that the matrix $\mathbf{A} \in \mathbb{C}^{m\times N}$ satisfies the ℓ_q-robust null space property of order s with constants $0 < \rho < 1$ and $\tau > 0$. Then, for any $x, z \in \mathbb{C}^N$,*

$$\|z - x\|_p \leq \frac{C}{s^{1-1/p}}\left(\|z\|_1 - \|x\|_1 + 2\sigma_s(x)_1\right) + Ds^{1/p-1/q}\|\mathbf{A}(z - x)\|,$$

where $C := (1 + \rho)^2/(1 - \rho)$ and $D := (3 + \rho)\tau/(1 - \rho)$.

Proof. Let us first remark that the ℓ_p-robust null space properties imply the ℓ_1-robust and ℓ_p-robust null space property ($p \leq q$) in the forms

$$\|v_S\|_1 \leq \rho \|v_{\overline{S}}\|_1 + \tau s^{1-1/q} \|\mathbf{A}v\|, \tag{1.4.17}$$

$$\|v_S\|_p \leq \frac{\rho}{s^{1-1/p}} \|v_{\overline{S}}\|_1 + \tau s^{1/p-1/q} \|\mathbf{A}v\|, \tag{1.4.18}$$

for all $v \in \mathbb{C}^N$ and all $S \subset [N]$ with $\mathrm{card}(S) \leq s$. Thus, in view of eq. (1.4.17), applying theorem 1.4.8 with S chosen as an index set of s largest (in modulus) entries of x leads to

$$\|z - x\|_1 \leq \frac{1+\rho}{1-\rho}(\|z\|_1 - \|x\|_1 + 2\sigma_s(x)_1) + \frac{2\tau`}{1-\rho} s^{1-1/q} \|\mathbf{A}(z-x)\|. \tag{1.4.19}$$

Then, choosing S as an index of s largest (in modulus) entries of $z - x$, we use theorem 1.4.12 to notice that

$$\|z - x\|_p \leq \|(z-x)_{\overline{S}}\|_p + \|(z-x)_S\|_p \leq \frac{1}{s^{1-1/p}} \|z - x\|_1 + \|(z-x)_S\|_p.$$

In view of eq. (1.4.18), we derive

$$\begin{aligned} \|z - x\|_p &\leq \frac{1}{s^{1-1/p}} \|z - x\|_1 + \frac{\rho}{s^{1-1/p}} \|(z-x)_{\overline{S}}\|_1 + \tau s^{1/p-1/q} \|\mathbf{A}(z-x)\| \\ &\leq \frac{1+\rho}{s^{1-1/p}} \|z - x\|_1 + \tau s^{1/p-1/q} \|\mathbf{A}(z-x)\|. \end{aligned} \tag{1.4.20}$$

□

1.4.4 Recovery of Individual Vectors

In some case, specific sparse vectors need to be dealt with rather than with all vectors supported on a given set or all vectors with a given sparsity. Therefore, some recovery conditions finer than the null space property are required. Such conditions are provided as follows, with a subtle difference between the real and complex settings, due to the fact the *sign* of a number z, defined as

$$\mathrm{sng}(z) := \begin{cases} \frac{z}{|z|} & if\ z \neq 0, \\ 0 & if\ z = 0, \end{cases}$$

is a discrete quantity when z is real, but it is not when z is complex. For a vector $x \in \mathbb{C}^N$, we denote by $\mathrm{sgn}(x) \in \mathbb{C}^N$ the vector with components $\mathrm{sgn}(x_j), j \in [N]$. The following contents start with the complex version of a recovery condition valid for sparse vectors.

Theorem 1.4.14. *Given a matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$, a vector $x \in \mathbb{C}^N$ with support S is the unique minimizer of $\|z\|_1$ subject to $\mathbf{A}z = \mathbf{A}x$ if one of the following equivalent conditions holds:*

(a) $\left| \sum_{j \in S} \overline{\mathrm{sgn}(x_j)} v_j \right| < \|v_{\overline{S}}\|_1$ *for all* $v \in \ker \mathbf{A} \backslash \{0\}$,

(b) $\mathbf{A}_S$ *is injective, and there exists a vector* $h \in \mathbb{C}^m$ *such that*

$$(\mathbf{A}^* h)_j = \mathrm{sgn}(x_j), \quad j \in S, \qquad |(\mathbf{A}^* h)_\ell| < 1, \quad \ell \in \overline{S}.$$

Proof. Let us start by proving that (a) implies that x is the unique minimizer of $\|z\|_1$ subject to $\mathbf{A}z = \mathbf{A}x$. For a vector $z \neq x$ such that $\mathbf{A}z = \mathbf{A}x$, we just have to write, with $v := x - z \in \ker\mathbf{A}\backslash\{0\}$,

$$\begin{aligned} \|z\|_1 &= \|z_S\|_1 + \|z_{\overline{S}}\|_1 = \|(x-v)_S\|_1 + \|v_{\overline{S}}\|_1 \\ &> |\langle x - v, \mathrm{sgn}(x)_S \rangle| + |\langle v, \mathrm{sgn}(x)_S \rangle| \geq |\langle x, \mathrm{sgn}(x)_S \rangle| = \|x\|_1. \end{aligned}$$

The implication $(b) \Rightarrow (a)$ is also simple. Indeed, observing that $\mathbf{A}v_S = -\mathbf{A}v_{\overline{S}}$ for $v \in \ker\mathbf{A}\backslash\{0\}$, we write

$$\begin{aligned} \left| \sum_{j \in S} \overline{\mathrm{sgn}(x_j)} v_j \right| &= |\langle v_S, \mathbf{A}^* h \rangle| = |\langle \mathbf{A}v_S, h \rangle| = |\langle \mathbf{A}v_{\overline{S}}, h \rangle| \\ &= \left| \left\langle v_{\overline{S}, \mathbf{A}^* h} \right\rangle \right| \leq \max_{\ell \in \bar{S}} |(\mathbf{A}^* h)_\ell| \, \|v_{\overline{S}}\|_1 < \|v_{\overline{S}}\|_1. \end{aligned}$$

The strict inequality holds since $\|v_{\overline{S}}\|_1 > 0$; otherwise, the nonzero vector $v \in \ker\mathbf{A}$ would be supported on S, contradicting the injectivity of $\mathbf{A}_S$.

The remaining implication $(a) \Rightarrow (b)$ requires more work. We start by noticing that (a) implies $\|v_{\overline{S}}\| > 0$ for all $v \in \ker\mathbf{A}\backslash\{0\}$. It follows that the matrix $\mathbf{A}_S$ is injective. Indeed, assume $\mathbf{A}_S v_S = 0$ for some $v_S \neq 0$ and complete v_S to a vector $v \in \mathbb{C}^N$ by setting $v_{\overline{S}} = 0$. Then v is contained in $\ker\mathbf{A}\backslash\{0\}$, which is in contradiction with $\|v_{\overline{S}} > 0\|$ for all $v \in \ker\mathbf{A}\backslash\{0\}$. Next, since the continuous function $v \mapsto |\langle v, \mathrm{sgn}(x)_S\rangle| / \|v_{\overline{S}}\|_1$ takes values less than one on the unit sphere of $\ker\mathbf{A}$, which is compact, its maximum μ satisfies $\mu < 1$. By homogeneity, we deduce

$$|\langle v, \mathrm{sgn}(x)_S\rangle| \leq \mu \|v_{\overline{S}}\|_1 \quad \text{for all } v \in \ker\mathbf{A}.$$

We then define, for $\mu < \nu < 1$, the convex set $\mathcal{C}$ and the affine set $\mathcal{D}$ by

$$\begin{aligned} \mathcal{C} &:= \left\{z \in \mathbb{C}^N : \|z_S\|_1 + \nu\|z_{\overline{S}}\|_1 \leq \|x\|_1\right\}, \\ \mathcal{D} &:= \left\{z \in \mathbb{C}^N : \mathbf{A}z = \mathbf{A}x\right\}. \end{aligned}$$

The intersection $\mathcal{C} \cup \mathcal{D}$ reduces to $\{x\}$. Indeed, we observe that $x \in \mathcal{C} \cup \mathcal{D}$, and if $z \neq x$ belongs to $\mathcal{C} \cup \mathcal{D}$, setting $v := x - z \in \ker\mathbf{A}\backslash\{0\}$, we obtain a contradiction from

$$\begin{aligned} \|x\|_1 &\geq \|z_S\|_1 + \nu\|z_{\overline{S}}\|_1 = \|(x-v)_S\|_1 + \nu\|v_{\overline{S}}\|_1 \\ &> \|(x-v)_S\|_1 + \mu\|v_{\overline{S}}\|_1 \geq |\langle x - v, \mathrm{sgn}(x)_S\rangle| + |\langle v, \mathrm{sgn}(x)_S\rangle| \\ &\geq |\langle x, \mathrm{sgn}(x)_S\rangle| = \|x\|_1. \end{aligned}$$

Thus, by the separation of convex sets via hyperplanes there exists a vector $w \in \mathbb{C}^N$ such that

$$\mathcal{C} \subset \left\{z \in \mathbb{C}^N : \mathrm{Re}\,\langle z, w\rangle \leq \|x\|_1\right\}, \tag{1.4.21}$$

$$\mathcal{D} \subset \left\{z \in \mathbb{C}^N : \mathrm{Re}\,\langle z, w\rangle = \|x\|_1\right\}. \tag{1.4.22}$$

In view of section 1.4.4, we have

$$\begin{aligned} \|x\|_1 &\geq \max_{\|z_S + \nu z_{\overline{S}}\|_1 \leq \|x\|_1} \mathrm{Re}\,\langle z, w\rangle \\ &= \max_{\|z_S + \nu z_{\overline{S}}\|_1 \leq \|x\|_1} \mathrm{Re}\left(\sum_{j\in S} z_j\overline{w_j} + \sum_{j\in\overline{S}} \nu z_j \overline{w_j}/\nu\right) \\ &= \max_{\|z_S + \nu z_{\overline{S}}\|_1 \leq \|x\|_1} \mathrm{Re}\,\langle z_S + \nu z_{\overline{S}}, w_s + (1/\nu)w_{\overline{S}}\rangle \\ &= \|x\|_1 \|w_S + (1/\nu)w_{\overline{S}}\|_\infty = \|x\|_1 \max\{\|w_S\|_\infty, (1/\nu)\|w_{\overline{S}}\|_\infty\}. \end{aligned}$$

Setting aside the case $x \neq 0$ (where the choice $h = 0$ would do), we obtain $\|w_S\|_\infty \leq$ ‘ and $\|w_{\overline{S}}\|_\infty \leq \nu <$ ‘. From section 1.4.4, we derive $\mathrm{Re} < x, w >= \|x\|_1$, i.e., $w_j = \mathrm{sgn}(x_j)$ for all $j \in S$, and also $\mathrm{Re}< v, w >= 0$ for all $v \in \ker\mathbf{A}$, i.e., $w \in (\ker\mathbf{A})^\perp$. Since $(\ker\mathbf{A})^\perp = \mathrm{ran}\mathbf{A}^*$, we write $w = \mathbf{A}^* h$ for some $h \in \mathbb{C}^m$. This establishes (b).

□

Remark 10. *(a) If a vector $x \in \mathbb{C}^N$ with support S satisfies condition (a) of the previous theorem, then all vectors $x' \in \mathbb{C}^N$ with support $S' \subset S$ and $sgn(x')_{S'}$ are also recovered via basis pursuit. Indeed, for $v \in ker\mathbf{A}\backslash\{0\}$,*

$$\begin{aligned} \left|\sum_{j\in S'} sgn(x'_j)v_j\right| &= \left|\sum_{j\in S} sgn(x_j)v_j - \sum_{j\in S\backslash S'} sgn(x_j)v_j\right| \\ &\leq \left|\sum_{j\in S} sgn(x_j)v_j\right| + \sum_{j\in S\backslash S'} |v_j| < \|v_{\overline{S}}\|_1 + \|v_{S\backslash S'}\|_1 = \|v_{\overline{S'}}\|_1. \end{aligned}$$

(b) theorem 1.4.14 can be made stable under noise on the measurements and under passing to compressible vectors. However, the resulting error bounds are slightly weaker than the ones of theorem 1.4.13 under the ℓ_2-robust null space property.

Corollary 1.4.1. *let $a_1, \ldots, a_N$ be the columns of $\mathbf{A} \in \mathbb{C}^{c \times N}$. For $x \in \mathbb{C}^N$ with support S, if the matrix $\mathbf{A}_S$ is injective and if*

$$\left|\left\langle \mathbf{A}_S^{\dagger} a_\ell, sgn(x_S)\right\rangle\right| < 1 \quad \textit{for all } \ell \in \overline{S}, \tag{1.4.23}$$

then the vector x is the unique solution of eq. (P_1) with $y = \mathbf{A}x$.

Remark 11. *In general, there is no converse to theorem 1.4.14. Let us consider, for instance,*

$$\mathbf{A} := \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix}, x = \begin{bmatrix} e^{-\pi i/3} \\ e^{\pi i/3} \\ 0 \end{bmatrix}.$$

We can verify that x is the unique minimizer of $\|z\|_1$ subject to $\mathbf{A}z = \mathbf{A}x$; However, (a) in theorem 1.4.14 fails. Indeed, for a vector $v = [\zeta, \zeta, \zeta] \in \ker \mathbf{A} \backslash \{0\}$, we have $|\overline{\text{sgn}(x_1)}v_1 + \overline{\text{sgn}(x_2)}v_2| = |(e^{\pi i/3} + e^{-\pi i/3})\zeta| = |\zeta|$, while $\|v_{\{3\}}\|_1 = |\zeta|$. A converse to theorem 1.4.14 holds in the real setting.

Theorem 1.4.15. *Given a matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$, a vector $x \in \mathbb{R}^N$ with support S is the unique minimizer of $\|z\|_1$ subject to $\mathbf{A}z = \mathbf{A}x$ if and only if one of the following equivalent conditions holds:*

(a) $\left|\sum_{j \in X} \text{sgn}(x_j) v_j\right| < \|v_{\overline{S}}\|_1$ *for all* $v \in \ker \mathbf{A} \backslash \{0\}$.

(b) $\mathbf{A}_S$ *is injective, and there exists a vector $h \in \mathbb{R}^m$ such that*

$$(\mathbf{A}^T h)_j = \text{sgn}(x_j), \quad j \in S, \qquad \left|(\mathbf{A}^T h)_\ell\right| < 1, \quad \ell \in \overline{S}.$$

Remark 12. *theorem 1.4.15 shows that in the real setting the recovery of a given vector via basis pursuit depends only on its sign pattern, but not on the magnitude of its entries. Moreover, by remark 10(a), if a vector $x \in \mathbb{R}^N$ with support $S' \subset S$ and $sgn(x')_{S'} = \text{sgn}(x)_{S'}$ are slso exactly recovered via basis pursuit.*

The construction of the "dual vector" h described in property (b) of theorem 1.4.14 and theorem 1.4.15 is not always straightforward. The following condition involving an "inexact dual vector" is somtimes easier to verify.

Theorem 1.4.16. *Let $a_1, \ldots, a_N$ be the columns of $\mathbf{A} \in \mathbb{C}^{m \times N}$ and let $x \in \mathbb{C}^N$ with support S. For $\alpha, \beta, \gamma, \theta \geq 0$, assume that*

$$\|(\mathbf{A}_S^* \mathbf{A}_S)^{-1}\|_{2\to 2} \leq \alpha, \qquad \max_{\ell \in \overline{S}} \|\mathbf{A}_S^* a_\ell\|_2 \leq \beta, \tag{1.4.24}$$

and that there exists a vector $u = \mathbf{A}^* h \in \mathbb{C}^N$ with $h \in \mathbb{C}^m$ such that

$$\|u_S - \text{sgn}(x_S)\|_2 \leq \gamma \quad \textit{and} \quad \|u_{\overline{S}}\|_\infty \leq \theta. \tag{1.4.25}$$

If $\theta + \alpha\beta\gamma < 1$, then x is the unique minimizer of $\|z\|_1$ subject to $\mathbf{A}z = \mathbf{A}x$.

Theorem 1.4.17. *Let $a_1, \ldots, a_N$ be the columns of $\mathbf{A} \in \mathbb{C}^{m \times N}$, let $x \in \mathbb{C}^N$ with s largest absolute entries supported on S, and let $y = \mathbf{A}x + e$ with $\|e\|_2 \leq \eta$. For $\delta, \beta, \gamma, \tau \geq 0$ with $\delta < 1$, assume that*

$$\|\mathbf{A}_S^* \mathbf{A}_S - \mathbf{Id}\|_{2\to 2} \leq \delta, \qquad \|\mathbf{A}_S^* a_\ell\|_2 \leq \beta, \tag{1.4.26}$$

and that there exists a vector $u = \mathbf{A}^* h \in \mathbb{C}^N$ with $h \in \mathbb{C}^m$ such that

$$\|u_S - \text{sgn}(x_S)\|_2 \leq \gamma, \quad \|u_{\overline{S}}\|_\infty \leq \theta, \quad \textit{and} \quad \|h\|_2 \leq \tau\sqrt{s}. \tag{1.4.27}$$

if $\rho := \theta + \beta\gamma/(1-\delta) < 1$, then a minimizer $x^\sharp$ of $\|z\|_1$ subject to $\|\mathbf{A}z - y\|_2 \leq \eta$ satisfies

$$\|x - x^\sharp\|_2 \leq C_1 \sigma_s(x)_1 + (C_2 + C_3\sqrt{s})\eta$$

for some constant $C_1, C_2, C_3 > 0$ depending only on $\delta, \beta, \gamma, \theta, \tau$.

Remark 13. *The proof reveals explicit values of the constants, namely,*

$$C_1 = \frac{2}{1-\rho}\left(1 + \frac{\beta}{1-\delta}\right), \qquad C_2 = \frac{2\sqrt{1+\delta}}{1-\delta}\mu\left(\frac{\gamma}{1-\rho}\left(1+\frac{\beta}{1-\delta}\right)+1\right),$$

$$C_3 = \frac{2\tau}{1-\rho}\left(1+\frac{\beta}{1-\delta}\right).$$

For instance, the specific choice $\delta = \beta = \gamma = 1/2, \theta = 1/4$, and $\tau = 2$, for which $\rho = 3/4$, results in $C_1 \approx 16, C_2 = 10\sqrt{6} \approx 24.49$, and $C_3 \approx 32$.

Proof. Observe that x is feasible for the quadratically constrained ℓ_1-minimization problem due to the assumed ℓ_2-bound of the perturbation e. Setting $v := x^\sharp - x$, the minimality of $\|x^\sharp\|_1$ implies

$$\begin{aligned} \|x\|_1 &\geq \|x^\sharp\|_1 = \|x+v\|_1 = \|(x+v)_S\|_1 + \|(x+v)_{\overline{S}}\|_1 \\ &\geq \operatorname{Re}\langle (x+v)_S, \operatorname{sgn}(x_S)\rangle + \|v_{\overline{S}}\|_1 - \|x_{\overline{S}}\|_1 \\ &= \|x_S\|_1 + \operatorname{Re}\langle v_S, \operatorname{sgn}(x_S)\rangle + \|v_{\overline{S}}\|_1 - \|x_{\overline{S}}\|_1. \end{aligned}$$

Rearranging and using the fact that $\|x\|_1 = \|x_S\|_1 + \|x_{\overline{S}}\|_1$ yields

$$\|v_{\overline{S}}\|_1 \leq 2\|x_{\overline{S}}\|_1 + |\langle v_S\rangle, \operatorname{sgn}(x_S)|. \tag{1.4.28}$$

In view of eq. (1.4.27), we have

$$\begin{aligned} |\langle v_S, \operatorname{sgn}(x_S)\rangle| &\leq |\langle v_S, \operatorname{sgn}(x_S) - u_S\rangle| + |\langle v_S, u_S\rangle| \\ &\leq \gamma\|v_S\|_2 + |\langle v, u\rangle| + |\langle v_{\overline{S}}, u_{\overline{S}}\rangle|. \end{aligned} \tag{1.4.29}$$

The first inequality of eq. (1.4.26) guarantees that $\|(\mathbf{A}_S^*\mathbf{A}_S)^{-1}\|_{2\to 2} \leq 1/(1-\delta)$ and $\|\mathbf{A}_S^*\|_{2\to 2} \leq \sqrt{1+\delta}$. Hence,

$$\|v_S\|_2 \leq \frac{1}{1-\delta}\|\mathbf{A}_S^*\mathbf{A}_S v_S\|_2 \leq \frac{1}{1-\delta}\|\mathbf{A}_S^*\mathbf{A}_{\overline{S}} v_{\overline{S}}\|_2 + \frac{1}{1-\delta}\|\mathbf{A}_S^*\mathbf{A}v\|_2 \tag{1.4.30}$$

$$\leq \frac{1}{1-\delta}\sum_{\ell\in\overline{S}} |v_\ell|\, \|\mathbf{A}_S^* a_\ell\|_2 + \frac{\sqrt{1+\delta}}{1-\delta}\|\mathbf{A}v\|_2 \tag{1.4.31}$$

$$\leq \frac{\beta}{1-\delta}\|v_{\overline{S}}\|_1 + \frac{2\sqrt{1+\delta}}{1-\delta}\eta. \tag{1.4.32}$$

The last step involved the inequality $\|\mathbf{A}v\|_2 \leq 2\eta$, which follows from the optimization constraints as

$$\|\mathbf{A}v\|_2 = \|\mathbf{A}(x^\sharp - x)\|_2 \leq \|\mathbf{A}x^\sharp - y\|_2 + \|y - \mathbf{A}x\|_2 \leq 2\eta.$$

The latter inequalituy combined with $\|h\|_2 \leq \tau\sqrt{s}$ also gives

$$|\langle v, u\rangle| = |\langle v, \mathbf{A}^* h\rangle| = |\langle \mathbf{A}v, h\rangle| \leq \|\mathbf{A}v\|_2\|h\|_2 \leq 2\tau\eta\sqrt{s},$$

while $\|u_{\overline{S}}\|_\infty \leq \theta$ implies $|\langle v_{\overline{S}}, u_{\overline{S}}\rangle| \leq \theta\|v_{\overline{S}}\|_1$. Substituting these estimates in section 1.4.4 and in turn in eq. (1.4.28) yields

$$\|v_{\overline{S}}\|_1 \leq 2\|\overline{S}\|_1 + \left(\theta + \frac{\beta\gamma}{1-\delta}\right)\|v_{\overline{S}}\|_1 + \left(2\gamma\frac{\sqrt{1+\delta}}{1-\delta} + 2\tau\sqrt{s}\right)\eta.$$

Since $\rho = \theta + \beta\gamma/(1-\delta) < 1$, this can be rearranged as

$$\|v_{\overline{S}}\|_1 \leq \frac{2}{1-\rho}\|x_{\overline{S}}\|_1 + \frac{2(\mu\gamma + \tau\sqrt{s})}{1-\rho}\eta, \tag{1.4.33}$$

where $\mu := \sqrt{1+\delta}/(1-\delta)$. Using section 1.4.4 once again, we derive

$$\|v_S\|_2 \leq \frac{2\beta}{(1-\rho)(1-\delta)}\|x_{\overline{S}}\|_1 + \left(\frac{2\beta(\mu\gamma+\tau\sqrt{s})}{(1-\rho)(1-\delta)} + 2\mu\right)\eta. \tag{1.4.34}$$

Finally, combining eq. (1.4.33) and eq. (1.4.34), we obtain

$$\begin{aligned} \|v\|_2 &\leq \|v_{\overline{S}}\|_2 + \|v_S\|_2 \leq \|v_{\overline{S}}\|_1 + \|v_S\|_2 \\ &\leq \frac{2}{1-\rho}\left(1+\frac{\beta}{1-\delta}\right)\|x_{\overline{S}}\|_1 + \left(\frac{2(\mu\gamma+\tau\sqrt{s})}{1-\rho}\left(1+\frac{\beta}{1-\delta}\right) + 2\mu\right)\eta. \end{aligned}$$

Taking $\|x_{\overline{S}}\|_1 = \sigma_s(x)_1$ into account, we arrive at the desired result. $\square$

The next characterization of exact recovery via ℓ_1-minimization involves *tangent cones* to the ℓ_1-ball. For a vector $x \in \mathbb{R}^N$, we introduce the convex cone

$$T(x) = \operatorname{cone}\left\{z - x : z \in \mathbb{R}^N, \|z\|_1 \leq \|x\|_1\right\}, \tag{1.4.35}$$

where the notation *cone represents the conic hll;*

> The *norm cone* associated with the norm $\|\cdot\|$ is the set
>
> $$\mathcal{C} = \{(x,t) \mid \|x\| \leq t\} \subseteq \mathbf{R}^{n+1}.$$
>
> It is (as the name suggests) a **convex cone**. **Example 2.3** *The second-order cone* is the norm cone for the Euclidean norm, i.e.,
>
> $$\begin{aligned} \mathcal{C} &= \left\{(x,t) \in \mathbf{R}^{n+1} \mid \|x\|_2 \leq t\right\} \\ &= \left\{ \begin{bmatrix} x \\ t \end{bmatrix} \,\middle|\, \begin{bmatrix} x \\ t \end{bmatrix}^T \begin{bmatrix} I & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ t \end{bmatrix} \leq 0, t \geq 0 \right\}. \end{aligned}$$

Theorem 1.4.18. *For $\mathbf{A} \in \mathbb{R}^{m\times N}$, a vector $x \in \mathbb{R}^N$ is the unique minimizer of $\|z\|_1$ subject to $\mathbf{A}z = \mathbf{A}x$ if and only if $ker\mathbf{A} \cup T(x) = \{0\}$.*

Theorem 1.4.19. *For $\mathbf{A} \in \mathbb{R}^{m\times N}$, let $x \in \mathbb{R}^N$ and $y = \mathbf{A}x + e \in \mathbb{R}^m$ with $\|e\|_2 \leq \eta$. If*

$$\inf_{v\in T(x), \|v\|_2=1} \|\mathbf{A}v\|_2 \geq \tau$$

for some $\tau > 0$, then a minimizer $x^\sharp$ of $\|z\|_1$ subject to $\|\mathbf{A}z - y\|_2 \leq \eta$ satisfies

$$\|x - x^\sharp\|_2 \leq \frac{2\eta}{\tau} \tag{1.4.36}$$

Proof. The inequality $\|x^\sharp\|_1 \leq \|x\|_1$ yields $v := (x^\sharp - x)/\|x^\sharp - x\|_2 \in T(x)$—note that $x^\sharp - x \neq 0$ can be safely assumed. Since $\|v\|_2 = 1$, the assumption implies $\|\mathbf{A}v\|_2 \geq \tau$, i.e., $\|\mathbf{A}(x^\sharp - x)\|_2 \geq \tau\|x^\sharp - x\|_2$. It remains to remark that

$$\|\mathbf{A}(x^\sharp - x)\|_2 \leq \|\mathbf{A}x^\sharp - y\|_2 + \|\mathbf{A}x - y\|_2 \leq 2\eta$$

in order to obtain the desired result. □

<++>

1.4.5 The Projected Cross-Polytope

1.4.6 Low-Rank Matrix Recovery

1.4.7 Notes

Throughout the section, we have insisted on sparse vectors to be unique solutions of eq. (P_1). If the uniqueness requirement is dropped, then a necessary and sufficient condition for every s-sparse vector to be a solution of eq. (P_1) would be a *weak null space property where the strict inequality is replaced by a weak inequality sign*.

The null space property is somewhat folklore in the compressive sensing literature. It appears implicitly in works in [1], in [2], and in [3]. In [4] the notion was also isolated. The name was first used in [5], albeit for a property slightly more general then eq. (1.4.3), namely, $\|v\|_1 \leq C\sigma_s(v)_1$ for all $v \in \ker\mathbf{A}$, where $C \geq 1$ is an unspecified constant.

The equivalence between the real and complex null space properties was established in [6]. using a *different argument* than the one of theorem 1.4.2. The result was generalized in [7]. The proof of theorem 1.4.2 follows their argument.

The term *instance optimality* is sometimes also used for what we called stability. The stability and robustness of sparse reconstruction via basis pursuit, as stated after theorem 1.4.9, were established in [8] under a restricted isometry property—condition on the measurement matrix.

The fact that sparse recovery via ℓ_p-minimization implies sparse recovery via ℓ_p-minimization whenever $0 < p < q \leq 1$ was proved in [9].

Chapter 2

Coherence

2.1 Definitions and Basic Properties

Definition 2.1.1. *Let* $\mathbf{A} \in \mathbb{C}^{m\times N}$ *be a matrix with* ℓ_2*-normalized columns* $a_1, \dots, a_N$*, i.e.,* $\|a_i\|_2 = 1$ *for all* $i \in [N]$*. The* coherence $\mu = \mu(A)$ *of the matrix* $\mathbf{A}$ *is defined as*

$$\mu := \max_{1\leq i\neq j\leq N} |\langle a_i, a_j\rangle| . \tag{2.1.1}$$

Next the more general concept of ℓ_1-coherence function is introduced, which incorporates the usual coherence as the particular value $s = 1$ of its argument.

Definition 2.1.2. *Let* $\mathbf{A} \in \mathbb{C}^{m\times N}$ *be a matrix with* ℓ_2*-normalized columns* $a_1, \dots, a_N$*. The* ℓ_1-coherence function μ_1 *of the matrix* $\mathbf{A}$ *is defined for* $s \in [N-1]$ *by*

$$\mu_1(s) := \max_{i\in[N]} \max \left\{ \sum_{j\in S} |\langle a_i, a_j\rangle| , S \subset [N], card(S) = s, i \notin S \right\} .$$

It is straightforward to observe that, for $1 \leq s \leq N-1$,

$$\mu \leq \mu_1(s) \leq s\mu, \tag{2.1.2}$$

and more generally that, for $1 \leq s, t \leq N-1$ with $s + t \leq N-1$,

$$\max\{\mu_1(s), \mu_1(t)\} \leq \mu_1(s+t) \leq \mu_1(s) + \mu_1(t). \tag{2.1.3}$$

The coherence, more generally the ℓ_1-coherence function, is invariant under multiplication on the left by a *unitary matrix* $\mathbf{U}$, for the columns of $\mathbf{UA}$ are the *ℓ_2-normalized vectors* $\mathbf{U}a_1, \dots, \mathbf{U}a_N$ and they satisfy $\langle \mathbf{U}a_i, \mathbf{U}a_j\rangle = \langle a_i, a_j\rangle$. Moreover, because of the *Cauchy-Schwarz* inequality $|\langle a_i, a_j\rangle| \leq \|a_i\|_2\|a_j\|_2$, it is clear that the coherence of a matrix is bounded above by one, i.e.,

$$\mu \leq 1$$

.

We observe that $\mu = 0$ if and only if the columns of $\mathbf{A}$ from an orthonormal system. In particular, in the case of a square matrix, we have $\mu = 0$ if and only if $\mathbf{A}$ is a *unitary matrix*. But from the *compressive sensing* point of view, we only consider matrices $\mathbf{A} \in \mathbb{C}^{m\times N}$ with $m < N$. $\mathbf{A}_S$ denotes the matrix formed by the columns of $\mathbf{A} \in \mathbb{C}^{m\times N}$ indexed by *a subset* S *of* $[N]$.

Theorem 2.1.1. *Let* $\mathbf{A} \in \mathbb{C}^{m\times N}$ *be a matrix with* ℓ_2*-normalized columns and let* $s \in [N]$*. For all* s*-sparse vectors* $x \in \mathbb{C}^N$*,*

$$(1 - \mu_1(s-1))\|x\|_2^2 \leq \|\mathbf{A}x\|_2^2 \leq (1 + \mu_1(s-1))\|x\|_2^2,\text{[1]}$$

[1] It looks like RIP

or equivalently, for each set $S \subset [N]$ with card$(S) \leq s$, the eigenvalues of the matrix $\mathbf{A}_S^*\mathbf{A}_S$ lie in the interval $[1-\mu_1(s-1), 1+\mu_1(s-1)]$. In particular, if $\mu_1(s-1) < 1$, then $\mathbf{A}_S^*\mathbf{A}_S$ is invertible.

Proof. For a set $S \subset [N]$ with $card(S) \leq s$, since the matrix $\mathbf{A}_S^*\mathbf{A}_S$ is positive semidefinite, it has an orthonormal basis of eigenvectors associated with real, positive eigenvalues. The minimal eigenvalue is denoted by λ_{min} and the maximal eigenvalue by λ_{max}. Then, since $\mathbf{A}x = \mathbf{A}_S x_S$ for any $x \in \mathbb{C}^N$ supported on S, it is easy to see that the maximum of

$$\|\mathbf{A}x\|_2^2 = \langle \mathbf{A}_S x_S, \mathbf{A}_S x_S\rangle = \langle \mathbf{A}_s^*\mathbf{A}_S x_S, x_S\rangle$$

over the set $\{x \in \mathbb{C}^N, suppx \subset S, \|x\|_2 = 1\}$ is λ_{max} and that its minimum is λ_{min}. Due to the normalizations $\|a_j\|_2 = 1$ for all $j \in [N]$, the diagonal entries of $\mathbf{A}_S^*\mathbf{A}_S$ all equal one. By *Gershgorin's disk theorem theorem 2.1.2*, the eigenvalues of $\mathbf{A}_S^*\mathbf{A}_S$ are contained in the union of the disks centered at 1 with radii

$$r_j := \sum_{\ell_S, \ell \neq j} |(\mathbf{A}_S^*\mathbf{A}_S)_{j,\ell}| = \sum_{\ell \in S, \ell \neq j} |\langle a_\ell, a_j\rangle| \leq \mu_1(s-1), \qquad j \in S.$$

Since these eigenvalues are real, they must lie in $[1-\mu_1(s-1), 1+\mu_1(s-1)]$, as announced.

Theorem 2.1.2. *Let λ be an eigenvalue of a square matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$. There exists an index $j \in [n]$ such that*

$$|\lambda - A_{j,j}| \leq \sum_{\ell \in [n]\setminus\{j\}} |A_{j,\ell}|.$$

□

Corollary 2.1.1. *Given a matrix $\mathbf{A} \in \mathbb{C}^{m\times N}$ with ℓ_2-normalized columns and an integer $s \geq 1$, if*

$$\mu_1(s) + \mu_1(s-1) < 1,$$

then, for each set $S \subset [N]$ with card$(S) \leq 2s$, the matrix $\mathbf{A}_S^*\mathbf{A}_S$ is invertible and the matrix $\mathbf{A}_S$ injective. In particular, the conclusion holds if

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

2.2 Matrices with Small Coherence

The lower bounds is given for the coherence and for the ℓ_1-coherence function of a matrix $\mathbf{A} \in \mathbb{C}^{m\times N}$ with $m < M$. An example of a matrix is given with an almost minimal coherence. The analysis is carried out for matrices $\mathbf{A} \in \mathbb{K}^{m\times N}$, where the field $\mathbb{K}$ can either be $\mathbb{R}$ or $\mathbb{C}$, because the matrices *achieving the lower bounds have different features in the real and complex settings*. In both cases, however, *their columns are equiangular tight frames, which are defined below.*

Definition 2.2.1. *A system of ℓ_2-normalized vectors $(a_1, \ldots, a_N)$ in $\mathbb{K}^m$ is called equiangular if there is a constant $c \geq 0$ such that*

$$|\langle a_i, a_j\rangle| = c \qquad \text{for all } i, j \in [N], i \neq j.$$

Tight frames can be defined by several conditions.

Definition 2.2.2. *A system of vectors $(a_1, \ldots, a_N)$ in $\mathbb{K}^m$ is called a tight frame if there exists a constant $\lambda > 0$ such that one of the following equivalent conditions holds:*

(a) $\|x\|_2^2 = \lambda \sum_{j=1}^{N} |\langle x, a_j\rangle|^2$ *for all* $x \in \mathbb{K}^m$,

(b) $x = \lambda \sum_{j=1}^{N} \langle x, a_j\rangle a_j$ *for all* $x \in \mathbb{K}^m$,

(c) $\mathbf{A}\mathbf{A}^* = \frac{1}{\lambda}\mathbf{Id}_m$, *where* $\mathbf{A}$ *is the matrix with columns* $a_1, \ldots, a_N$.

A system of ℓ_2-normalized vectors is called an ***equiangular tight frame*** if it is both an equiangular system and a tight frame. Such systems are the ones achieving the lower bound given below and known as the *Welch bound*.

Theorem 2.2.1. *The coherence of a matrix* $\mathbf{A}$ *in*$\mathbb{K}^{m\times N}$ *with* ℓ_2*-normalized columns satisfies*

$$\mu \geq \sqrt{\frac{N-m}{m(N-1)}}. \tag{2.2.1}$$

Equality holds if and only if the columns $a_1, \ldots, a_N$ *of the matrix* $\mathbf{A}$ *form an equiangular tight frame.*

Proof. The *Gramm matrix* $\mathbf{G} := \mathbf{A}^*\mathbf{A} \in \mathbb{K}^{N\times N}$ of the system $(a_1, \ldots, a_N)$ is introduced, which has entries

$$G_{i,j} = \overline{\langle a_i, a_j\rangle} = \langle a_j, a_i\rangle, \qquad i, j \in [N],$$

and the matrix $\mathbf{H} := \mathbf{A}\mathbf{A}^* \in \mathbb{K}^{m\times m}$. On the other hand, since the system $(a_1, \ldots, a_N)$ is ℓ_2-normalized, we have

$$tr(\mathbf{G}) = \sum_{i=1}^{N} \|a_i\|_2^2 = N. \tag{2.2.2}$$

On the other hand, since the inner product

$$\langle \mathbf{U}, \mathbf{V}\rangle_F := tr(\mathbf{U}\mathbf{V}^*) = \sum_{i,j=1}^{n} U_{i,j}\overline{V_{i,j}}$$

induces the so-called *Froebenius norm* $\|\cdot\|_F$ on $\mathbb{K}^{n\times n}$, the *Cauchy-Schwarz inequality* yields

$$tr(\mathbf{H}) = \langle \mathbf{H}, \mathbf{Id}_m\rangle_F \leq \|\mathbf{H}\|_F \|\mathbf{Id}_m\|_F = \sqrt{m}\sqrt{tr(\mathbf{H}\mathbf{H}^*)}. \tag{2.2.3}$$

Now it is observed that

$$\begin{aligned} tr(\mathbf{H}\mathbf{H}^*) &= tr(\mathbf{A}\mathbf{A}^*\mathbf{A}\mathbf{A}^*) = tr(\mathbf{A}^*\mathbf{A}\mathbf{A}^*\mathbf{A}) = tr(\mathbf{G}\mathbf{G}^*) = \sum_{i,j=1}^{N} |\langle a_i, a_j\rangle|^2 && (2.2.4)\\ &= \sum_{i=1}^{N} \|a_i\|_2^2 + \sum_{i,j=1, i\neq j}^{N} |\langle a_i, a_j\rangle|^2 = N + \sum_{i,j=1, i\neq j}^{N} |\langle a_i, a_j\rangle|^2. && (2.2.5) \end{aligned}$$

In view of $tr(G) = tr(H)$, combining eqs. (2.2.2) and (2.2.3) and section 2.2 yields

$$N^2 \leq m\left(N + \sum_{i,j\neq 1, i\neq j}^{N} |\langle a_i, a_j\rangle|^2\right). \tag{2.2.6}$$

Taking into account that

$$|\langle a_i, a_j\rangle| \leq \mu \qquad \text{for all } i, j \in [N], i \neq i, \tag{2.2.7}$$

we obtain

$$N^2 \leq m(N + (N^2 - N)\mu^2),$$

$\square$

The *Welch bound* can be extended to the ℓ_1-coherence function for small values of its argument.

Theorem 2.2.2. *The* ℓ_1 *coherence function of a matrix* $\mathbf{A} \in \mathbb{K}^{m\times N}$ *with* ℓ_2*-normalized columns satisfies*

$$\mu_1(s) \geq s\sqrt{\frac{N-m}{m(N-1)}} \qquad \textit{whenever} \qquad s < \sqrt{N-1}. \tag{2.2.8}$$

Equality holds if and only if the columns $a_1, \ldots, a_N$ *of the matrix* $\mathbf{A}$ *form an equiangular tight frame.*

The proof is based on the following lemma.

Lemma 2.2.1. *For $k < \sqrt{n}$, if the finite sequence $(\alpha_1, \alpha_2, \ldots, \alpha_n)$ satisfies*

$$\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n \geq 0 \qquad \text{and} \qquad \alpha_1^2 + \alpha_2^2 + \cdots + \alpha_n^2 \geq \frac{n}{k^2},$$

then

$$\alpha_1 + \alpha_2 + \cdots + \alpha_k \geq 1,$$

with equality if and only if $\alpha_1 = \alpha_2 = \cdots = \alpha_n = 1/k$.

Proof. The equivalent statement

$$\left.\begin{array}{r} \alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n \geq 0 \\ \alpha_1 + \alpha_2 + \cdots + \alpha_k \leq 1 \end{array}\right\} \Longrightarrow \alpha_1^2 + \alpha_2^2 + \cdots + \alpha_n^2 \leq \frac{n}{k^2},$$

with equality if and only if $\alpha_1 = \alpha_2 = \cdots = \alpha_n = 1/k$. This is the problem of maximizing the convex function

$$f(\alpha_1, \alpha_2, \ldots, \alpha_n) := \alpha_1^2 + \alpha_2^2 + \cdots + \alpha_n^2$$

over the convex polygon

$$\mathcal{C} := \{(\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n : \alpha_1 \geq \cdots \geq \alpha_n \geq 0 \quad \text{and} \quad \alpha_1 + \cdots + \alpha_k \leq 1\}.$$

Because any point in $\mathcal{C}$ is a convex combination of its vertices (so that the extreme points of $\mathcal{C}$ are vertices) and because the function of is convex, the maximum is attained at a vertex of $\mathcal{C}$ by theorem 2.2.3. The vertices of $\mathcal{C}$ are obtained as intersections of n hyperplanes arising by turning n of the $(n+1)$ inequality constraints into equalities. Thus, we have the following possibilities:

1. If $\alpha_1 = \alpha_2 = \cdots = `a_n = 0$, then $f(\alpha_1, \alpha_2, \ldots, \alpha_n) = 0$.
2. If $\alpha_1 + \cdots + \alpha_k = 1$ and $\alpha_1 = \cdots = \alpha_\ell > \alpha_{\ell+1} = \cdots = \alpha_n$ for $1 \leq \ell \leq k$, then $\alpha_1 = \cdots = \alpha_\ell = 1/\ell$, and consequently $f(\alpha_1, \alpha_2, \ldots, \alpha_n) = 1/\ell$.
3. If $\alpha_1 + \cdots + \alpha_k = 1$ and $\alpha_1 = \cdots = \alpha_\ell > \alpha_{\ell+1} = \cdots = \alpha_n = 0$ for $k < \ell < n$, then $\alpha_1 = \cdots = \alpha_\ell = 1/k$, and consequently $f(\alpha_1, \alpha_2, \ldots, \alpha_n) = \ell/k^2$.

Taking $k < \sqrt{n}$ into account, it follows that

$$\max_{(\alpha_1,\ldots,\alpha_n)\in\mathcal{C}} f(\alpha_1, \ldots, \alpha_n) = \max\left\{\max_{1\leq\ell\leq k} \frac{1}{\ell}, \max_{k<\ell<n} \frac{\ell}{k^2}\right\} \max\left\{1, \frac{n}{k^2}\right\} = \frac{n}{k^2},$$

with equality only in the case $\ell = n$ where $\alpha_1 = \alpha_2 = \cdots = \alpha_n = 1/k$.

Theorem 2.2.3. *Let $K \subset \mathbb{R}^N$ be a compact convex set and let $F := K \to \mathbb{R}$ be a convex function. Then F attains its maximum at an extreme point of K.*

□

The case of equality in lemma 2.2.1 implies that $|\langle a_{i^*}, a_j\rangle| = \sqrt{(N-m)/(m(N-1))}$ for all $j \in [N], j \neq i^*$. Since the index i^* can be arbitrarily chosen in $[N]$, the system $(a_1, \ldots, a_N)$ is also equiangular. Conversely, the proof that equiangular tight frames yields equality in eq. (2.2.8) follows easily from theorem 2.2.1 and eq. (2.1.2).

Theorem 2.2.4. *The cardinality N of an equiangular system $(a_1, \ldots, a_N)$ of ℓ_2-normalized vectors in $\mathbb{K}^m$ satisfies*

$$N \leq \frac{m(m+1)}{2} \qquad \text{when } \mathbb{K} = \mathbb{R},$$
$$N \leq m^2 \qquad \text{when } \mathbb{K} = \mathbb{C}.$$

If equality is achieved, then the system $(a_1, \ldots, a_N)$ is also tight frame.

Theorem 2.2.5. *For $m \geq 3$, if there is an equiangular system of $m(m+1)/2$ vectors in $\mathbb{R}^m$, then $m+2$ is necessarily the square of an odd integer.*

Proposition 2.2.1. *For each prime number $m \geq 5$, there is an explicit $m \times m^2$ complex matrix with coherence $\mu = 1/\sqrt{m}$.*

2.2.1 Analysis of Orthogonal Matching Pursuit

Theorem 2.2.6. *Let* $\mathbf{A} \in \mathbb{C}^{m\times N}$ *be a matrix with* ℓ_2*-normalized columns. If*

$$\mu_1(s) + \mu_1(s-1) < 1 \tag{2.2.9}$$

then every s*-sparse vector* $x \in \mathbb{C}^N$ *is exactly recoved from the measurement vector* $y = \mathbf{A}x$ *after at most* s *iterations of orthogonal matching pursuit.*

2.2.2 Analysis of Basis Pursuit

As a matter of fact, any condition guaranteeing the success of the recovery of all vecotrs supported on a set S via card(S) iterations of orthogonal matching pursuit also guarantees the success of the recovery of all vectors supported on S via basis pursuit. Indeed, given $\mathbf{v} \in ker\ \mathbf{A}\backslash\{0\}$, we have $\mathbf{A}_S\mathbf{v}_S = -\mathbf{A}_{\overline{S}}\mathbf{v}_{\overline{S}}$, and section 1.2

$$\|\mathbf{v}\|_1 = \|\mathbf{A}_S^\dagger\mathbf{A}_S\mathbf{v}_S\|_1 = \|\mathbf{A}_S^\dagger\mathbf{A}_{\overline{S}}\mathbf{v}_{\overline{S}}\|_1 \le \|\mathbf{A}_S^\dagger\mathbf{A}_{\overline{S}}\|_{1\to1}\|\mathbf{v}_{\overline{S}}\|_1 < \|\mathbf{v}_{\overline{S}}\|_1$$

Theorem 2.2.7. *Let* $\mathbf{A} \in \mathbb{C}^{m\times N}$ *be a matrix with* ℓ_2*-normalized columns. If*

$$\mu_1(s) + \mu_1(s-1) < 1, \tag{2.2.10}$$

then every s*-sparse vector* $x \in \mathbb{C}^N$ *is exactly recovered from the measurement vecotr* $y = \mathbf{A}x$ *via basis pursuit.*

2.2.3 Analysis of Thresholding Algorithms

Theorem 2.2.8. *Let* $\mathbf{A} \in \mathbb{C}^{m\times N}$ *be a matrix with* ℓ_2*-normalized columns and let* $x \in \mathbb{C}^N$ *be a vector supported on a set* S *of size* s*. If*

$$\mu_1(s) + \mu_1(s-1) < \frac{\min_{i\in S}|x_i|}{\max_{i\in S}|x_i|}, \tag{2.2.11}$$

then the vecotr $x \in \mathbb{C}^N$ *is exactly recovered from the measurement vecotr* $y = \mathbf{A}x$ *via basic thresholding.*

Theorem 2.2.9. *Let* $\mathbf{A} \in \mathbb{C}^{m\times N}$ *be a matrix with* ℓ_2*-normalized columns. If*

$$2\mu_1(s) + \mu_1(s-1) < 1,$$

then every s*-sparse vector* $x \in \mathbb{C}^N$ *is exactly recoved from the measurement vecotr* $y = \mathbf{A}x$ *after at most* s *iterations of hard thresholding prusuit.*

Note 1. *The conclusion of theorem 2.2.9 can be achieved under the sufficient condition* $\mu < 1/(3s-1)$*. Similarly, the conclusion of theorem 2.2.7 can be achieved under the sufficient condition* $\mu < 1/(2s-1)$*.*

theorem 2.2.6 and theorem 2.2.7 in their present form were established by Tropp. What we call ℓ_1*-coherence function here is called* cumulative coherence function *there. This concept also appears under the name* Babel function*. A straightforward extension to any* $p > 0$ *would be the* ℓ_p*-coherence function of a matrix* $\mathbf{A}$ *with* ℓ_2*-normalized columns* $a_1, \dots, a_N$ *defined by*

$$\mu_p(s) := \max_{i\in[N]} \max \left\{ \left(\sum_{j\in S} |\langle a_i, a_j\rangle|^p \right)^{1/p}, S \subseteq [N], card(S) = s, i \notin S \right\}.$$

Chapter 3

Restricted isometry constants and restricted orthogonality constants

3.1 Definitions and Basic Properties

Unlike the *coherence, which only takes paires of columns of a matrix into account*, the restricted isometry constant of *order s involves all s-tuples of columns and is therefore more suited to assess the quality of the matrix*.

Definition 3.1.1. *The sth* restricted isometry constant $\delta_s = \delta_s(\mathbf{A})$ *of a matrix* $\mathbf{A} \in \mathbb{C}^{\gtrdot \times \mathbb{N}}$ *is the smallest* $\delta \geq 0$ *such that*

$$(1-\delta)\|x\|_2^2 \leq \|\mathbf{A}x\|_2^2 \leq (1+\delta)\|x\|_2^2 \tag{3.1.1}$$

for all s-sparse vectors $x \in \mathbb{C}^N$. Equivalently, it is given by

$$\delta_s = \max_{S \subset [N], card(S) \leq s} \|\mathbf{A}_S^* \mathbf{A}_S - \mathbf{Id}\|_{2 \to 2}. \tag{3.1.2}$$

We say that $\mathbf{A}$ satisfies the *restricted isometry property* if *δ_s is small for reasonably large s—the meaning of small δ_s and large s will be made precise later*.

Indeed, eq. (3.1.2) syas that each column submatrix $\mathbf{A}_S, S \subset [N]$ with $card(S) \leq s$, has all its singular values in the interval $[1-\delta_s, 1+\delta_s]$ and is therefore injective when $\delta_s < 1$.

To prove eq. (3.1.2) and for the equivalence of eq. (3.1.1) and eq. (3.1.2) in the complex setting, we start by noticing that eq. (3.1.1) is equivalent to

$$\left| \|\mathbf{A}_S x\|_2^2 - \|x\|_2^2 \right| \leq \delta \|x\|_2^2 \quad \text{for all } S \subset [N], card(S) \leq s, \text{and all } x \in \mathbb{C}^S.$$

One then observes that, for $x \in \mathbb{C}^S$,

$$\|\mathbf{A}_S x\|_2^2 - \|x\|_2^2 = \langle \mathbf{A}_S x, \mathbf{A}_S x \rangle - \langle x, x \rangle = \langle (\mathbf{A}_S^* \mathbf{A}_S - \mathbf{Id})\, x, x \rangle .$$

Since the matrix $\mathbf{A}_S^* \mathbf{A}_S - \mathbf{Id}$ is **Hermitian**, we have

$$\max_{x \in \mathbb{C}^S \setminus \{0\}} \frac{\langle (\mathbf{A}_S^* \mathbf{A}_S - \mathbf{Id}) x, x \rangle}{\|x\|_2^2} = \|\mathbf{A}_S^* \mathbf{A}_S - \mathbf{Id}\|_{2 \to 2},$$

so that eq. (3.1.1) is equivalent to

$$\max_{S \subset [N], card(S) \leq s} \|\mathbf{A}_S^* \mathbf{A}_S - \mathbf{Id}\|_{2 \to 2} \leq \delta.$$

This proves the identity eq. (3.1.2), as δ_s is the smallest such δ.

Proposition 3.1.1. *If the matrix A has ℓ_2-normalized columns $a_1, \dots, a_N$, i.e., $\|a_j\|_2 = 1$ for all $j \in [N]$, then*

$$\delta_1 = 0, \qquad \delta_2 = \mu, \qquad \delta_s \leq \mu_1(s-1) \leq (s-1)\mu, \quad s \geq 2.$$

Proof. The ℓ_2-normalization of the columns means that $\|\mathbf{A}e_j\|_2^2 = \|e_j\|_2^2$ for all $j \in [N]$, that is to say $\delta_1 = 0$. Next, with $a_1, \dots, a_N$ denoting the columns of the matrix $\mathbf{A}$, we have

$$\delta_2 = \max_{1 \leq i \neq j \leq N} \|\mathbf{A}_{i,j}^* \mathbf{A}_{i,j} - \mathbf{Id}\|_{2\to 2}, \qquad \mathbf{A}_{i,j}^* \mathbf{A}_{i,j} = \begin{bmatrix} 1 & \langle a_j, a_i \rangle \\ \langle a_i, a_j \rangle & 1 \end{bmatrix}.$$

The eigenvalues of the matrix $\mathbf{A}_{i,j}^* \mathbf{A}_{i,j} - \mathbf{Id}$ are $|\langle a_i, a_j \rangle|$ and $-|\langle a_i, a_j \rangle|$, so its operator norm is $|\langle a_i, a_j \rangle|$. Taking the maximum over $1 \leq i \neq i \leq N$ yields the equality $\delta_2 = \mu$. The inequality $\delta_s \leq \mu_1(s-1) \leq (s-1)\mu$ folows from theorem 2.1.1. $\square$

In view of the existence of $m \times m^2$ matrices with coherence μ equal to $1/\sqrt{m}$ (see chapter 2), this already shows the existence of $m \times m^2$ matrices with restricted isometry constant $\delta_s < 1$ for $s \leq \sqrt{m}$. We will establish that, given $\delta < 1$, there exist $m \times N$ matrices with restricted isometry constant $\delta_s \leq \delta$ for $s \leq cm/\ln(eN/m)$, where c is a constant depending only on δ. *Matrices with a small restricted isometry constant of this optimal order are informally said to satisfy the* ***restricted isometry property*** *and* ***uniform uncertainty principle***.

A simple but essential observation will be made, which *motivates the related notion of restricted orthogonality constant*.

Proposition 3.1.2. *Let $\mathbf{u}, \mathbf{v} \in \mathbb{C}^N$ be vectors with $\|\mathbf{u}\|_0 \leq s$ and $\|\mathbf{v}\|_0 \leq t$. If supp$(\mathbf{u}) \cap$ supp$(\mathbf{v}) = \emptyset$, then*

$$|\langle \mathbf{Au}, \mathbf{Av} \rangle| \leq \delta_{s+t} \|\mathbf{u}\|_2 \|\mathbf{v}\|_2. \tag{3.1.3}$$

Definition 3.1.2. *The (s,t)-restricted orthogonality constant $\theta_{s,t} = \theta_{s,t}(\mathbf{A})$ of a matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$ is the smallest $\theta \geq 0$ such that*

$$|\langle \mathbf{Au}, \mathbf{Av} \rangle| \leq \theta \|\mathbf{u}\|_2 \|\mathbf{v}\|_2 \tag{3.1.4}$$

for all disjointly supported s-sparse and t-sparse vectors $\mathbf{u}, \mathbf{v} \in \mathbb{C}^N$. Equivalently, it is given by

$$\theta_{s,t} = \max\left\{\|\mathbf{a}_T^* \mathbf{A}_s\|_{2\to 2}, S \cap T = \emptyset, card(S) \leq s, card(T) \leq t\right\}. \tag{3.1.5}$$

Proposition 3.1.3. *Restricted isometry constants and restricted orthogonality constants are related by*

$$\theta_{s,t} \leq \delta_{s+t} \leq \frac{1}{s+t}(s\delta_s + t\delta_t + 2\sqrt{st}\theta_{s,t}).$$

The special case $t = s$ gives the inequalities

$$\theta_{s,s} \leq \delta_{2s} \qquad \textit{and} \qquad \delta_{2s} \leq \delta_s + \theta_{s,s}.$$

Restricted isometry constants and restricted orthogonality constants of high order can be controlled by those of lower order.

Proposition 3.1.4. *For integers $r, s, t \geq 1$ with $t \geq s$,*

$$\begin{aligned} \theta_{t,r} &\leq \sqrt{\frac{t}{s}}\theta_{s,r}, \\ \theta_t &\leq \frac{t-d}{s}\delta_{2s} + \frac{d}{s}\delta_s \qquad d := gcd(s,t). \end{aligned}$$

The special case $t = cs$ gives

$$\delta_{cs} \leq c\delta_{2s}.$$

Remark 14. *There are other relations enabling to control constants of higher order by constants of lower order*

Theorem 3.1.1. *For $\mathbf{A} \in \mathbb{C}^{m\times N}$ and $1 \le s \le N$, one has*

$$m \ge c\frac{s}{\delta_s^2} \tag{3.1.6}$$

provided $N \ge Cm$ and $\delta_s \le \delta_*$, where the constants c, C and δ_* depend only on each other. For instance, the choices $c = 1/162, C = 30$ and $\delta_* = 2/3$ are valid.

Proof. We set $t := s/2 \ge 1$, and decompose the matrix $\mathbf{A}$ in blocks of size $m \times t$—except possibly the last one which may have less columns—as

$$\mathbf{A} = [\mathbf{A}_1|\mathbf{A}_2|\dots|\mathbf{A}_n], N \le nt.$$

From eq. (3.1.2) and eq. (3.1.5), we recall that, for all $i, j \in [n], i \ne i$

$$\|\mathbf{A}_i^*\mathbf{A}_i - \mathbf{Id}\|_{2\to2} \le \delta_t \le \delta_s, \qquad \|\mathbf{A}_i^*\mathbf{A}_j\|_{2\to2} \le \theta_{t,t} \le \delta_{2t} \le \delta_s,$$

so that the eigenvalues of $\mathbf{A}_i^*\mathbf{A}_i$ and the singular values of $\mathbf{A}_i^*\mathbf{A}_j$ satisfy

$$1 - \delta_s \le \lambda_k(\mathbf{A}_i^*\mathbf{A}_i) \le 1 + \delta_s, \qquad \sigma_k(\mathbf{A}_i^*) \le \delta_s$$

Let us introduce the matrices

$$\mathbf{H} := \mathbf{A}\mathbf{A}^* \in \mathbb{C}^{m\times m}, \qquad \mathbf{G} := \mathbf{A}^*\mathbf{A} = [\mathbf{A}_i^*\mathbf{A}_i]_{1\le i,j\le n} \in \mathbb{C}^{N\times N}.$$

On the one hand, we have the lower bound

$$tr(\mathbf{H}) = tr(\mathbf{G}) = \sum_{i=1}^{n} tr(\mathbf{A}_i^*\mathbf{A}_i) = \sum_{i=1}^{n}\sum_{k=1}^{t} \lambda_k(\mathbf{A}_i^*\mathbf{A}_i) \ge nt(1-\delta_s). \tag{3.1.7}$$

On the other hand, writing $\langle \mathbf{M}_1, \mathbf{M}_2\rangle_F = tr(\mathbf{M}_2^*\mathbf{M}_1)$ for the Frobenius inner product of two matrices $\mathbf{M}_1$ and $\mathbf{M}_2$, we have

$$tr(\mathbf{H})^2 = \langle \mathbf{Id}_m, \mathbf{H}\rangle_F^2 \le \|\mathbf{Id}_m\|_F^2\|\mathbf{H}\|_F^2 = mtr(\mathbf{H}^*\mathbf{H}).$$

Then, by cyclicity of the trace,

$$\begin{aligned}
tr(\mathbf{H}^*\mathbf{H}) &= tr(\mathbf{A}\mathbf{A}^*\mathbf{A}\mathbf{A}^*) = tr(\mathbf{A}^*\mathbf{A}\mathbf{A}^*\mathbf{A}) = tr(\mathbf{G}\mathbf{G}^*) \\
&= \sum_{i=1}^{n} tr(\sum_{j=1}^{*} \mathbf{A}_i^*\mathbf{A}_j\mathbf{A}_j^*\mathbf{A}_i) \\
&= \sum_{1\le i\ne j\le n}\sum_{k=1}^{t} \sigma_k(\mathbf{A}_i^*\mathbf{A}_j)^2 + \sum_{i=1}^{n}\sum_{k=1}^{t} \lambda_k(\mathbf{A}_i^*\mathbf{A}_i)^2 \\
&\le n(n-1)t\delta_s^2 + nt(1+\delta_s)^2.
\end{aligned}$$

The upper bound is derived

$$tr(\mathbf{H})^2 \le mnt\left((n-1)\delta_s^2 + (1+\delta_s)^2\right). \tag{3.1.8}$$

Combining the bounds eq. (3.1.7) and eq. (3.1.8) yields

$$m \ge \frac{nt(1-\delta_s)^2}{(n-1)\delta_s^2 + (1+\delta_s)^2}.$$

If $(n-1)\delta_s^2 < (1+\delta_s)^2/5$, we would obtain, using $\delta \le 2/3$,

$$m > \frac{nt(1-\delta_s)^2}{6(1+\delta_2)^2/5} \ge \frac{5(1-\delta_s)^2}{6(1+\delta_s)^2}N \ge \frac{1}{30}N,$$

which contradicts the assumption. We therefore have $(n-1)\delta_s^2 \ge (1+\delta_s)^2/5$, which yields, using $\delta_s \le 2/3$ again and $s \le 3t$,

$$m \ge \frac{nt(1-\delta-s)^2}{6(n-1)\delta_s^2} \ge \frac{1}{54}\frac{t}{\delta_s^2} \ge \frac{1}{162}\frac{s}{\delta_s^2}.$$

This is the desired result. $\square$

Let us compare the previous lower bound on restricted isometry constants, namely

$$\delta_s \geq \sqrt{cs/m}, \tag{3.1.9}$$

with upper bounds available so far. Precisely, choosing a matrix $\mathbf{A} \in \mathbb{C}^{m\times n}$ with a coherence of optimal order $\mu \leq c/\sqrt{m}$, proposition 3.1.1 implies that

$$\delta_s \leq (s-1)\mu \leq cs/\sqrt{m}. \tag{3.1.10}$$

There's a significant gap between eq. (3.1.9) and eq. (3.1.10). In particular, eq. (3.1.10) with the quadratic scaling

$$m \geq c's^2 \tag{3.1.11}$$

allows δ_s to be small, while this requires under eq. (3.1.9) that $m \geq c's$. However, whether such a condition can be sufficient is unknown at this point. Certain matrices $\mathbf{A} \in \mathbb{R}^{m\times N}$ satisfy $\delta_s \leq \delta$ with high probability for some $\delta > 0$ provided

$$m \geq C\delta^{-2} s \ln(eN/s). \tag{3.1.12}$$

$\delta_s \leq \delta$ requires $m \geq C_\delta s \ln(eN/s)$. Therefore, the lower bound $m \geq c'$ is optimal up to logarighmic factors, and eq. (3.1.6) is optimal regarding the scaling $C_\delta = C\delta^{-2}$.

Difficulty of Deterministic Constructions of matrices with RIP. As mensioned, random matrices will be used to obatin the restricted isometry property $\delta_s \leq \delta$ (abbreviated RIP) in the optimal regime eq. (3.1.12) for the number m of mearsurements in terms of the sparsity s and the vector length N. Finding deterministic matrices satisfying $\delta_s \leq \delta$ in this regime is a major open problem. Essentially all available estimations of δ_s for deterministic matrices combine a coherence estimation and proposition 3.1.1 in one form or another. This leads to bounds of the type eq. (3.1.10) and in turn to the quadratic bottleneck eq. (3.1.11). Thus, the lower bound of theorem 2.2.1 in principle prevents such a proof technique to generate improved results. *The intrinsic difficulty in bounding the restricted isometry constants of explicit matrices* $\mathbf{A}$ *lies in the basic tool for estimating the eigenvalues of* $\mathbf{A}_S^*\mathbf{A}_S - \mathbf{Id}$, *namely, Gershgorin's disk theorem theorem 2.1.2.* Assuming ℓ_2-normalization of the columns of $\mathbf{A}$ and taking the supremum over all $S \subset [N]$ with $\mathrm{card}(S) = s$ leads then to the ℓ_1-coherence function $\mu_1(s-1)$—this is how we showed the bound $\delta_s \leq \mu_1(s-1)$ of proposition 3.1.1; see also theorem 2.1.1.

3.2 Analysis of Basis Pursuit

Theorem 3.2.1. *Suppose that the* $2s$*th restricted isometry constant of the matrix* $\mathbf{A} \in \mathbb{C}^{m\times N}$ *satisfies*

$$\delta_{2s} < \frac{1}{3}. \tag{3.2.1}$$

Then every s*-sparse vector* $x \in \mathbb{C}^N$ *is the unique solution of*

$$\min_{z\in\mathbb{C}^N} \|z\|_1 \qquad \text{subject to } \mathbf{Az} = \mathbf{Ax}.$$

Lemma 3.2.1. *Given* $q > p > 0$, *if* $\mathbf{u} \in \mathbb{C}^t$ *satisfy*

$$\max_{i\in[s]} |u_i| \leq \min_{j\in[t]} |v_j|, \tag{3.2.2}$$

then

$$\|\mathbf{u}\|_q \leq \frac{s^{1/q}}{t^{1/p}} \|\mathbf{v}\|_p.$$

The special case $p = 1, q = 2$, *and* $t = s$ *gives*

$$\|\mathbf{u}\|_2 \leq \frac{1}{\sqrt{s}} \|\mathbf{v}\|_1.$$

Proof. For the first statement, we only need to notice that

$$\frac{\|\mathbf{u}\|_q}{s^{1/q}} = \left[\frac{1}{s}\sum_{i=1}^{s} |u_i|^q\right]^{1/q} \leq \max_{i\in[s]} |u_i|,$$

$$\frac{\|\mathbf{v}\|_p}{t^{1/p}} = \left[\frac{1}{t}\sum_{j=1}^{t} |v_j|^p\right]^{1/p} \geq \min_{j\in[t]} |v_j| \,,$$

and to use eq. (3.2.2). The second statement is an immediate consequence. □

Proof. (theorem 3.2.1) According to theorem 3.2.2, it is enough to establish the null space property of order s in the form

$$\|\mathbf{v}_S\|_1 < \frac{1}{2}\|\mathbf{v}\|_1 \quad \text{for all } \mathbf{v} \in \mathit{ker}\ \mathbf{A}\backslash\{0\} \text{ and all } S \subset [N] \text{ with } \mathrm{card}(S) = s.$$

This will follow from the stronger statement

$$\|\mathbf{v}_S\|_s \leq \frac{\rho}{2\sqrt{s}}\|\mathbf{v}\|_1 \quad \text{for all } \mathbf{v} \in \mathit{ker}\ \mathbf{A} \text{ and all } S \subset [N] \text{ with } \mathrm{card}(S) = s.$$

where

$$\rho := \frac{2\delta_{2s}}{1-\delta_{2s}}$$

satisfies $\rho < 1$ whenever $\delta_{2s} < 1/3$. Given $\mathbf{v} \in \mathit{ker}\mathbf{A}$, we notice that it is enough to consider an index set $S =: S_0$ of s large absolute entries of the vector $\mathbf{v}$. We partition the complement $\overline{S_0}$ of S_0 in $[N]$ as $\overline{S_0} = S_1 \cup S_2 \cup \ldots$, where

$$S_1 : \text{index set of } s \text{ largest absolute entries of } \mathbf{v} \text{ in } \overline{S_0},$$
$$S_0 : \text{index set of } s \text{ largest absolute entries of } \mathbf{v} \text{ in } \overline{S_0 \cup S_1},$$

etc. In view of $\mathbf{v} \in \mathit{ker}\mathbf{A}$, we have $\mathbf{A}(\mathbf{v}_{S_0}) = \mathbf{A}(-\mathbf{v}_{S_1} - \mathbf{v}_{S_2} - \ldots)$, so that

$$\begin{aligned} \|\mathbf{v}_{S_0}\|_2^2 &\leq \frac{1}{1-\delta_{2s}}\|\mathbf{A}(\mathbf{v}_{S_0})\|_2^2 = \frac{1}{1-\delta_{2s}}\left\langle \mathbf{A}(-\mathbf{v}_{S_1}), \mathbf{A}(-\mathbf{v}_{S_0}), \mathbf{A}(-\mathbf{v}_{S_1}) + \mathbf{A}(-\mathbf{v}_{S_2}) + \ldots \right\rangle \\ &= \frac{1}{1-\delta_{2s}}\sum_{k\geq 1}\left\langle \mathbf{A}(\mathbf{v}_{S_0}), \mathbf{A}(-\mathbf{v}_{S_k})\right\rangle. \end{aligned} \tag{3.2.3}$$

According to proposition 3.1.2, we also have

$$\left\langle \mathbf{A}(\mathbf{v}_{S_0}), \mathbf{A}(-\mathbf{v}_{S_k})\right\rangle \leq \delta_{2s}\|\mathbf{v}_{S_0}\|_2\|\mathbf{v}_{S_k}\|_2 \tag{3.2.4}$$

Substituting eq. (3.2.4) into section 3.2 and dividing by $\|\mathbf{v}_{S_0}\|_2 > 0$, we obtain

$$\|\mathbf{v}_{S_0}\|_2 \leq \frac{\delta_{2s}}{1-\delta_{2s}}\sum_{k\geq 1}\|\mathbf{v}_{S_k}\|_2 = \frac{\rho}{2}\sum_{k\geq 1}\|\mathbf{v}_{S_k}\|_2.$$

For $k \geq 1$, tht s absolute entries of $\mathbf{v}_{S_k}$ do not exceed the s absolute entries of $\mathbf{v}_{S_{k-1}}$, so that lemma 3.2.1 yields

$$\|\mathbf{v}_{S_k}\|_2 \leq \frac{1}{\sqrt{s}}\|\mathbf{v}_{S_{k-1}}\|_1.$$

We then derive

$$\|\mathbf{v}_{S_0}\|_2 \leq \frac{\rho}{2\sqrt{s}}\sum_{k\geq 1}\|\mathbf{v}_{S_{k-1}}\|_1 \leq \frac{\rho}{2\sqrt{s}}\|\mathbf{v}\|_1.$$

This is the desired inequality. □

Theorem 3.2.2. *Given a matrix $\mathbf{A} \in \mathbb{K}^{m\times N}$, every s-sparse vector $x \in \mathbb{K}^N$ is the unique solution of*

$$\min_{z\in\mathbb{C}^N} \|z\|_1 \qquad \textit{subject to } \mathbf{A}z = y.$$

with $y = \mathbf{A}x$ if and only if $\mathbf{A}$ satisfies the null space property of order s.

Theorem 3.2.3. *Suppose that the $2s$th restricted isometry constant of the matrix $\mathbf{A} \in \mathbb{C}^{m\times N}$ satisfies*

$$\delta_{2s} < \frac{4}{\sqrt{41}} \approx 0.6246. \tag{3.2.5}$$

Then, for any $x \in \mathbb{C}^N$ *and* $y \in \mathbb{C}^m$ *with* $\|\mathbf{A}x - y\|_2 \leq \eta$*, a solution* $x^\sharp$ *of*

$$\min_{z \in \mathbb{C}^N} \|z\|_1 \qquad \textit{subject to } \|\mathbf{A}z - y\|_2 \leq \eta$$

approximates the vector x *with errors*

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

where the constants $C, D > 0$ *depend only on* δ_{2s}*.*

Theorem 3.2.4. *If the* $2s$*th restricted isometry constant of* $\mathbf{A} \in \mathbb{C}^{m \times N}$ *obeys eq.* (3.2.5)*, then the matrix* $\mathbf{A}$ *satisfies the* ℓ_2*-robust null space property of order* s *with constants* $0 < p < 1$ *and* $\tau > 0$ *depending only on* δ_{2s}*.*

Lemma 3.2.2. *For* $a_1 \geq a_2 \geq \cdots \geq a_s \geq 0$*.*

$$\sqrt{a_1^2 + \cdots + a_s^2} \leq \frac{a_1 + \cdots + a_s}{\sqrt{s}} + \frac{\sqrt{s}}{4}a_s \leq 1.$$

Chapter 4

Math basic: John-Lindenstrauss lemma

The **Johnson-Lindenstrauss lemma** concerns low-distortion embeddings of points from high-dimensional into low-dimensional into low-dimensional *Euclidean space*. The lemma states that a small set of points in a high-dimensional space can be embedded into a space of much lower dimension in such a way that distances between the points are nearly preserved. The map used for the embedding is at least Lipschitz, and can even be taken to be an orthogonal projection.

The lemma is tight up to a factor $\log(1/\epsilon)$, i.e. there exists a set of points of size m that needs dimension

$$\Omega\left(\frac{\log(m)}{\epsilon^2 \log(1/\epsilon)}\right)$$

in order to preserve the distances between all pair of points.

4.1 Lemma

Definition 4.1.1. *Given $0 < \epsilon < 1$, a set X of m points in $\mathbf{R}^N$, and a number $n > 8\ln(m)/\epsilon^2$, there is a linear map $f : \mathbf{R}^N \to \mathbf{R}^n$ such that*

$$(1-\epsilon)\|u-v\|^2 \leq \|f(u)-f(v)\|^2 \leq (1+\epsilon)\|u-v\|^2$$

for all $u, v \in X$.

In discrete CS problem, we have

$$y = \Phi x,$$

where Φ is an $n \times N$ matrix and $y \in \mathbf{R}^n$. The matrix Φ maps $\mathbf{R}^N$, where N is generally large, into $\mathbf{R}^n$, where n is typically much smaller than N.chapter 5

<++>

Chapter 5

A simple proof of the restricted isometry property for random matrics

In *Compressed Sensing (CS)* [**?**, 10]<++>

Chapter 6

The restricted isometry property and its implications for compressed sensing

$$y = \Phi x \tag{6.0.1}$$

The solusion $x*$ to

$$\min_{\tilde{x} \in \mathbb{R}^n} \|\tilde{x}\|_{\ell_1} \qquad \text{subject to} \qquad \Phi \tilde{x} = y \tag{6.0.2}$$

recovers x exactly provided that

1) x is sufficiently sparse and

2) the matrix Φ obeys a condition known as the *restricted isometry property.*

Definition 6.0.2. *For each integer $s = 1, 2, \dots,$ define the isometry constant δ_s of a matrix Φ as the smallest number such that*

$$(1 - \delta_s)\|x\|_{\ell_2}^2 \leqslant \|\Phi x\|_{\ell_2}^2 \leqslant (1 + \delta_s)\|x\|_{\ell_2}^2 \tag{6.0.3}$$

holds for all s-sparse vectors. A vector is said to be s-sparse if it has at most s nonzero entries.

By comparing the reconstruction x^* with the *best sparse approximation* one could obtain if one knew exactly the locations and amplitudes of the s-largest entries of x; here and below, we denote this approximation by x_s, i.e. the vector x with all but the s-largest entries set to zero.

Theorem 6.0.1 (Noiseless recovery)**.** *Assume that $\delta_{2s} < \sqrt{2} - 1$. Then the solusion x^* to eq.* (6.0.2) *obeys*

$$\|x^* - x\|_{\ell_1} \leqslant C_0 \|x - x_s\|_{\ell_1} \tag{6.0.4}$$

and

$$\|x^* - x\|_{\ell_2} \leqslant C_0 s^{-1/2} \|x - x_s\|_{\ell_1} \tag{6.0.5}$$

for some constant C_0 given explicitly below. In particular, if x is s-sparse, the recovery is exact.

Bibliography

[1] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l(1) minimization," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 5, pp. 2197–2202, Mar 2003.

[2] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *Information Theory, IEEE Transactions on*, vol. 47, no. 7, pp. 2845–2862, 2001.

[3] M. Elad and A. M. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases," *Ieee Transactions on Information Theory*, vol. 48, no. 9, pp. 2558–2567, Sep 2002.

[4] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *Ieee Transactions on Information Theory*, vol. 49, no. 12, pp. 3320–3325, Dec 2003.

[5] A. Cohen, W. Dahmen, and R. Devore, "Compressed sensing and best k-term approximation," *Journal of the American Mathematical Society*, vol. 22, no. 1, pp. 211–231, 2009.

[6] S. Foucart and R. Gribonval, "Real versus complex null space properties for sparse vector recovery," *Comptes Rendus Mathematique*, vol. 348, no. 15-16, pp. 863–865, Aug 2010.

[7] M. J. Lai and Y. Liu, "The null space property for sparse recovery from multiple measurement vectors," *Applied and Computational Harmonic Analysis*, vol. 30, no. 3, pp. 402–406, May 2011.

[8] E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, Aug 2006.

[9] R. Gribonval and A. Nielsen, "Highly sparse representations from dictionaries are unique and independent of the sparseness measure," *Applied and Computational Harmonic Analysis*, vol. 22, no. 3, pp. 335–355, May 2007.

[10] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *Information Theory, IEEE Transactions on*, vol. 52, no. 2, pp. 489–509, 2006.