2. Sparse Solutions of Underdetermined Linear Systems

2.1. **Vector Spaces.** We recall that a vector space V is a set that is closed under finite vector addition and scalar multiplication.

The basic example is the Euclidean space $\mathbb{R}^n$, where every element is represented by a list of n real numbers, scalars are real numbers, addition is componentwise, and scalar multiplication is multiplication on each term separately. The canonical unit vectors in $\mathbb{R}^n$ are denoted by $e_1, \dots, e_n$. They have entries

$$(e_i)_j = \delta_{i,j} = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

We can consider the standard inner product in $\mathbb{R}^n$, which we denote

$$\langle x, y \rangle = x^\top y = \sum_{i=1}^{n} x_i y_i. \tag{2.1}$$

With the inner product from above $\mathbb{R}^n$ becomes a (real) Hilbert space. This gives us the notion of orthogonality. When two vectors $x, y \in \mathbb{R}^n$ are orthogonal we often write

$$x \perp y \qquad \Leftrightarrow \qquad \langle x, y \rangle = 0.$$

Remark 2.1. The notions which we will introduce during this lecture, can be transferred to the n-dimensional complex space. The differences in statement and proofs come from the fact that the scalar product of $x, y \in \mathbb{C}^n$ is defined by

$$\langle x, y \rangle = x^\top y = \sum_{i=1}^{n} x_i \overline{y}_i. \tag{2.2}$$

where $\overline{z}$ is the complex conjugate of $z \in \mathbb{C}$. For the sake of simplicity we consider finite-dimensional signals in $\mathbb{R}^n$ and sensor matrices $A \in \mathbb{R}^{m \times n}$.

2.1.1. *Bases, ONB.* For simplicity we will be concerned with signals which are sparse with respect to the canonical basis $\{e_i\}$. In practice, however, the signal has a sparse representation with respect to a different basis, e.g. to the wavelet basis. Let us recall some terminology. A set of vectors $\{\phi^i\} \in \mathbb{R}^n$, which is linearly independent and which spans the whole space $\mathbb{R}^n$ is called a basis. It follows that such a set necessarily has n elements. Furthermore, every $x \in \mathbb{R}^n$ can be expressed *uniquely* as a linear combination of the basis vectors, i.e. there is a unique $z = (z_1, \dots, z_n)^\top$, such that

$$x = \sum_{i=1}^{n} \phi^i z_i. \tag{2.3}$$

Note that if we denote by $\Phi \in \mathbb{R}^{n \times n}$ the matrix with columns given by ϕ^i, then we can write $x = \Phi z$.

A basis is called *orthonormal*, if it satisfies the orthogonality relations

$$\langle \phi^i, \phi^j \rangle = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j. \end{cases}$$

An orthonormal basis has the advantage that the coefficients z can be easily calculated as

$$z_i = \langle x, \phi^i \rangle, \qquad i \in [n]$$

or

$$z = \Phi^\top x.$$

Indeed, the orthonormality of the columns gives $\Phi^\top \Phi = I$, where I denotes the identity matrix in $R^{n \times n}$.

2.1.2. *Four Fundamental Subspaces, Singular Value Decomposition.* We will be concerned with linear operators between the finite-dimensional spaces $\mathbb{R}^n$ and $\mathbb{R}^m$. These can be represented with the help of matrices $A \in \mathbb{R}^{m\times n}$. We denote the transpose of A by $A^\top$. For a given matrix $A \in \mathbb{R}^{m\times n}$, $\mathcal{N}(A)$ denotes the *nullspace* or *kernel* of A, and $\mathcal{R}(A)$ the *range* or *image* of A, that is the linear subspace spanned by the column vectors of A. $\mathcal{N}(A)^\perp$, $\mathcal{R}(A)^\perp$ denote the orthogonal complements, that is the linear subspaces orthogonal to $\mathcal{N}(A)$ resp. $\mathcal{R}(A)$. These four subspaces are also called the *four fundamental subspaces*. We refer to Fig. 2.1 for an illustration.

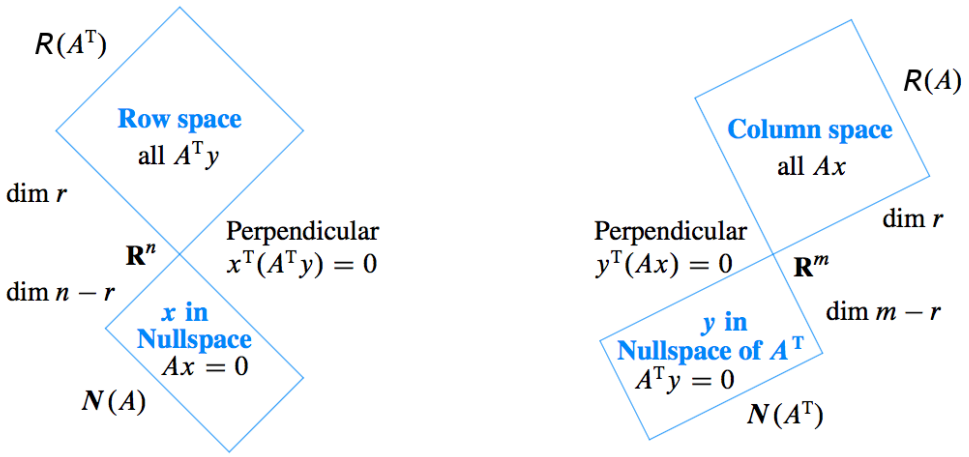

FIGURE 2.1. Dimensions and orthogonality of the four fundamental subspaces for any $m \times n$ matrix A of rank r.

The solution set to a linear system of equations

$$Ax = b\,, \quad A \in \mathbb{R}^{m\times n}, \qquad b \in \mathcal{R}(A), \tag{2.4}$$

is the affine subspace

$$x_0 + \mathcal{N}(A)\,, \tag{2.5}$$

where x_0 is any particular solution $Ax^0 = b$. Recall the formula

$$n = \dim \mathcal{R}(A) + \dim \mathcal{N}(A), \qquad \dim \mathcal{R}(A) = \operatorname{rank}(A) = \dim \mathcal{R}(A^\top). \tag{2.6}$$

Proposition 2.1 (Singular Value Decomposition (SVD)). *Let $A \in \mathbb{R}^{m\times n}$ with rank $r = \dim \mathcal{R}(A)$. Then there are orthogonal matrices*

$$U = (u^1, \dots, u^m) \in \mathcal{O}(m), \qquad U^\top U = UU^\top = I_m$$
$$V = (v^1, \dots, v^n) \in \mathcal{O}(n), \qquad V^\top V = VV^\top = I_n$$

such that

$$U^\top AV = \begin{pmatrix} U_1^\top AV_1 & U_1^\top AV_2 \\ U_2^\top AV_1 & U_2^\top AV_2 \end{pmatrix} = D = \begin{pmatrix} D_r & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{pmatrix},$$

and $D_r = \operatorname{Diag}\big(s_1(A), \dots, s_r(A)\big)$, where $U_1 := (u^1, \dots, u^r)$, $V_1 := (v^1, \dots, v^r)$, $U_2 := (u^{r+1}, \dots, u^m)$ and $V_2 := (v^{r+1}, \dots, v^n)$. The SVD is unique if the singular values $s_1(A) \geq \cdots \geq s_r(A) > 0$ are simple.

We denote by $P_{\mathcal{L}}$ the *orthogonal projection* onto a linear subspace $\mathcal{L}$. Recall that the orthogonal projection on a closed and convex set always exists.

Proposition 2.2. (Orthogonal Projection onto Closed Convex Sets) *Let $C \subset \mathbb{R}^n$ be nonempty, closed, convex. Then, for any $x_0 \in \mathbb{R}^n$, there exists a unique point $\hat{x}_0 = P_C x_0 \in C$, the **orthogonal projection** of x_0 onto C, such that*

$$\|x_0 - \hat{x}_0\|_2 = \inf_{x \in C} \|x_0 - x\|_2.$$

The vector $\hat{x}_0$ satisfies the variational inequality

$$\langle x_0 - \hat{x}_0, x - \hat{x}_0 \rangle \leq 0, \qquad \forall x \in C. \tag{2.7}$$

Conversely, if $y \in C$ satisfies (2.7), then $y = P_C x_0$.

Corollary 2.3. (Orthogonal Projection onto Subspaces) *Let $\mathcal{L} \subset \mathbb{R}^n$ be a nonempty linear subspace. Then the orthogonal projection $P_{\mathcal{L}} x_0$ of $x_0 \in \mathbb{R}^n$ onto $\mathcal{L}$ is characterized by*

$$\langle x_0 - P_{\mathcal{L}} x_0, x - P_{\mathcal{L}} x_0 \rangle = 0, \quad \forall x \in \mathcal{L}. \tag{2.8}$$

Moreover, $x_0 = P_{\mathcal{L}} x_0 + P_{\mathcal{L}^\perp} x_0$.

Proposition 2.4 (Properties of the SVD, Pseudoinverse). *Let $A = UDV^\top \in \mathbb{R}^{m \times n}$. Then,*

$$\begin{aligned}
A^\top A v^i &= s_i^2(A) v^i, & A A^\top u^i &= s_i^2(A) u^i, \\
A v^i &= s_i(A) u^i, & A^\top u^i &= s_i(A) v^i, \\
\operatorname{span}\{v^1, \dots, v^r\} &= \mathcal{N}(A)^\perp = \mathcal{R}(A^\top), & \operatorname{span}\{v^{r+1}, \dots, v^n\} &= \mathcal{N}(A), \\
\operatorname{span}\{u^1, \dots, u^r\} &= \mathcal{R}(A), & \operatorname{span}\{u^{r+1}, \dots, u^m\} &= \mathcal{R}(A)^\perp, \\
V_1 V_1^\top &= P_{\mathcal{N}(A)^\perp}, & V_2 V_2^\top &= P_{\mathcal{N}(A)}, \\
U_1 U_1^\top &= P_{\mathcal{R}(A)}, & U_2 U_2^\top &= P_{\mathcal{R}(A)^\perp}, \\
A = UDV^\top &= \sum_{i=1}^r s_i(A) u^i (v^i)^\top, & A^+ = VD^+U^\top &= \sum_{i=1}^r \frac{1}{s_i(A)} v^i (u^i)^\top,
\end{aligned}$$

where A^+ is called the pseudoinverse of A.

We have

$$\|A\|_F^2 = \sum_{i=1}^r s_i^2(A). \qquad \textit{(Frobenius norm)} \tag{2.9}$$

and

$$\|A\|_2 := s_1(A) = \sup_{\|x\|_2 = 1} \|Ax\|_2 = \sup_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2} \qquad \textit{(spectral norm)}, \tag{2.10}$$

i.e. the largest singular value is equal to the spectral norm of A, which is the *operator norm induced by the Euclidean vector norm*. Another matrix norm is

$$\|A\|_* = \sum_{i=1}^r s_i(A)$$

that is called *nuclear* norm and plays an important role in recovering low rank matrices by convex optimization.

2.1.3. *Norms and Quasinorms.* Throughout this lecture, we will treat signals as vectors in an n-dimensional Euclidean space, denoted by $\mathbb{R}^n$. We will typically be concerned with normed vector spaces, i.e., vector spaces endowed with a norm.

Definition 2.1 (Norm)**.** A *norm* on $\mathbb{R}^n$ is a real-valued function $x \mapsto \|x\|$ such that for all elements $x, y \in \mathbb{R}^n$ and any scalars $\lambda \in \mathbb{R}$ the following condition

(i) (*homogeneity*) $\|\lambda x\| = |\lambda| \|x\|$,
(ii) (*triangle inequality*) $\|x + y\| \leq \|x\| + \|y\|$,
(iii) (*definiteness*) $\|x\| = 0$ if and only if $x = 0$

hold. If only (i) and (ii) hold, so that $\|x\| = 0$ does not necessarily imply $x = 0$, then $\|\cdot\|$ is called *seminorm*. If (i) and (iii) hold, but (ii) is replaced by the weaker quasi-triangle inequality

$$\|x + y\| \leq C\left(\|x\| + \|y\|\right),$$

for some constant $C \geq 1$, then $\|\cdot\|$ is called *quasinorm*. The smallest constant C is called *quasinorm constant*.

When dealing with vectors in $\mathbb{R}^n$, we will make frequent use of the ℓ_p-norms, which for $p \in [1, \infty]$ are defined as

$$\|x\|_p = \begin{cases} \left(\sum_{i\in[n]} |x_i|^p\right)^{1/p}, & p \in [1, \infty), \\ \max_{i\in[n]} |x_i|, & p = \infty. \end{cases} \tag{2.11}$$

The inner product 2.1 leads to the ℓ_2-norm: $\|x\|_2 = \sqrt{\langle x, x\rangle}$. In some contexts it is useful to extend the notion of ℓ_p-norms to the case where $p < 1$. In this case, the "norm" defined in (2.11) fails to satisfy the triangle inequality, so it is actually a *quasinorm*. Indeed, we will see that

$$\|x + y\|_p \leq 2^{1/p-1}\left(\|x\|_p + \|y\|_p\right),$$
$$\|x + y\|_p^p \leq \left(\|x\|_p^p + \|y\|_p^p\right).$$

hold (Exercises). Therefore, the ℓ_p-quasinorm induces a metric via $d(x, y) = \|x - y\|_p^p$ for $0 < p < 1$. We recall the definition of a metric.

Definition 2.2. Let X be a set. A function $d : X \times X \to [0, \infty)$ is called *metric* if

(i) $d(x, y) = 0$ if and only if $x = y$,
(ii) $d(x, y) = d(y, x)$ for all $x, y \in X$,
(iii) $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in X$.

The ℓ_p (quasi-)norms have different properties for different values of p. To illustrate this, in Fig. 2.2 we show the unit sphere, i.e., $\{x\colon \|x\|_p = 1\}$ induced by each of these norms in $\mathbb{R}^2$. We use norms as a measure

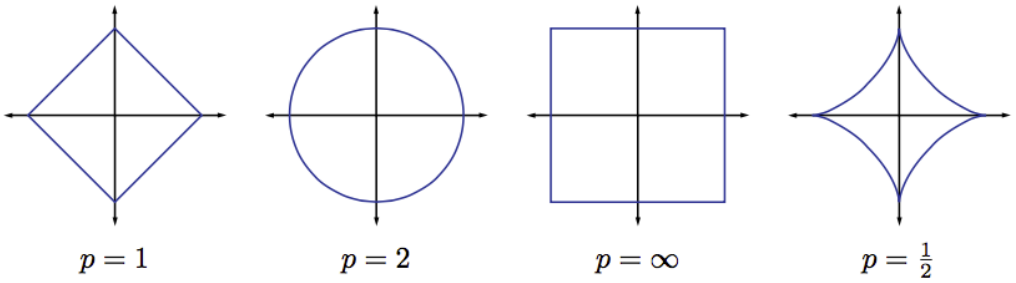

FIGURE 2.2. Unit spheres in R^2 for the ℓ_p norms with $p = 1, 2, \infty$, and for the ℓ_p quasinorm with $p = 1/2$.

of the length of a vector (strength of a signal), or the size of an error. For example, suppose we are given a signal $x \in \mathbb{R}^2$ and wish to approximate it using a point in a one-dimensional affine subspace $\mathcal{A}$. If we measure the approximation error using an ℓ_p-norm, then our task is to find the $\hat{x} \in \mathcal{A}$ that minimizes $\|x - \hat{x}\|_p$. The choice of p will have a significant effect on the properties of the resulting approximation error. An example is illustrated in Fig. 2.3. We observe that for larger p the error tends to spread out more evenly among the two coefficients, while smaller p leads to an error that is more unevenly distributed and tends to be sparse. This intuition generalizes to higher dimensions.

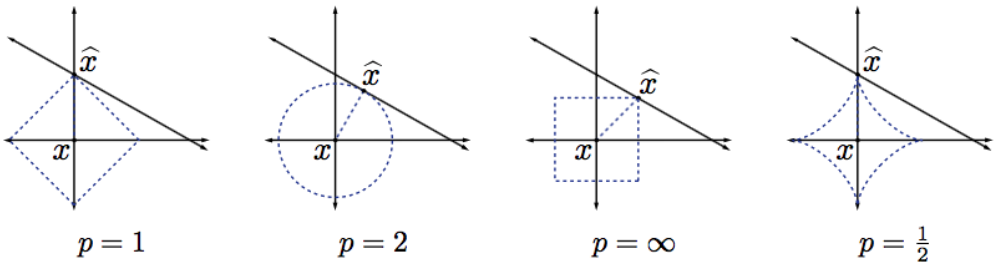

FIGURE 2.3. Best approximation of a point in $\mathbb{R}^2$ by a one-dimensional affine subspace $\mathcal{A}$ using the ℓ_p-norms for $p = 1, 2, \infty$, and the ℓ_p-quasinorm with $p = 1/2$. To compute the closest point to x in $\mathcal{A}$ using each ℓ_p "norm", we can imagine "inflating" an ℓ_p-sphere centered in x until it intersects $\mathcal{A}$. This will be the point $\hat{x} \in \mathcal{A}$ that is closest to x in the corresponding ℓ_p "norm".

2.2. **Sparsity and Compressibility.**

Definition 2.3 (Support, sparsity)**.** The **support** of a vector $x \in \mathbb{R}^n$ is the index set of its nonzero entries, i.e.

$$\operatorname{supp}(x) := \{i \in [n] \colon x_i \neq 0\} \, .$$

The vector $x \in \mathbb{R}^n$ is called s**-sparse** if at most s of its entries are nonzero, i.e. if

$$\|x\|_0 := |\operatorname{supp}(x)| \leq s.$$

The set of all s-sparse vectors is denoted by

$$\Sigma_s := \{x \in \mathbb{R}^n \colon \|x\|_0 \leq s\}.$$

Remark 2.2. $\|\cdot\|_0$ is not a norm and Σ_s is not a subspace. We observe that $\|\lambda x\|_0 = \|x\|_0$. Hence homogeneity fails to hold for any $\lambda \in \mathbb{R}$, $|\lambda| \neq 1$. The latter can be seen by observing that given a pair of s-sparse signals, a linear combination of the two signals will in general no longer be s-sparse, since their supports may not coincide. That is, for any $x, y \in \Sigma_s$, we do not necessarily have that $x + y \in \Sigma_s$ (although we do have that $x + y \in \Sigma_{2s}$). Σ_s consists of the union of all possible $\binom{n}{s}$ canonical subspaces. This is illustrated in Fig. 2.4 for $n = 3$ and $s = 2$.

Further note that $\|\cdot\|_0$ is not even a quasinorm, but one can show that

$$\lim_{p \to 0} \|x\|_p^p = |\operatorname{supp}(x)|,$$

justifying this choice of notation.

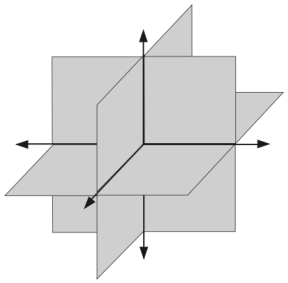

FIGURE 2.4. Illustration of the set $\Sigma_2 \subset \mathbb{R}^3$ of all 2-sparse signals in R^3, which is can be seen as a *union* of subspaces. Thus Σ_s is *not* a subspace.

In real applications, sparsity can be a strong assumption to impose, and we may prefer the weaker concept of *compressibility*. E.g., we can consider vectors that are nearly s-sparse, as measured by the *best s-term approximation*.

Definition 2.4 (Best s-term approximation)**.** For $p > 0$, the ℓ_p-error of best s-term approximation to a vector $x \in \mathbb{R}^n$ is defined by

$$\sigma_s(x)_p = \inf_{y \in \Sigma_s} \|y - x\|_p. \tag{2.12}$$

The best s-term approximation is the vector $\hat{x}^s$ that minimizes this error

$$\hat{x}^s = \arg\min_{y \in \Sigma_s} \|y - x\|_p. \tag{2.13}$$

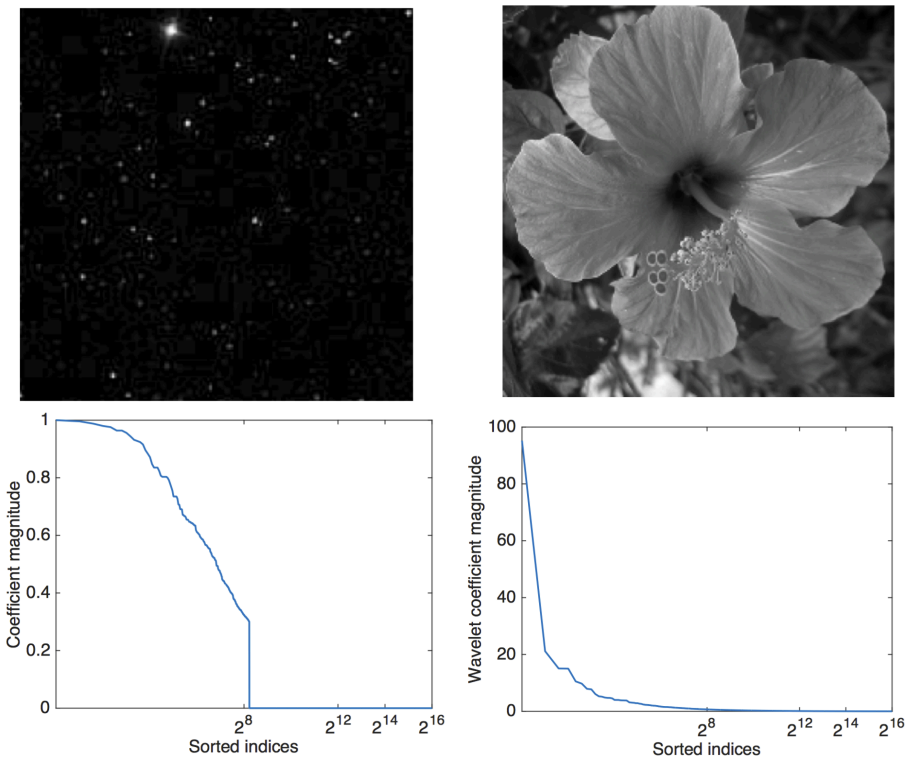

FIGURE 2.5. Sparsity versus compressibility. The 256×256 sky image (*top left*) is a signal that is compressible in space. If we sort the pixel in decreasing order, there is a sharp descent (*bottom left*) from nonzero values to zero values. The 256×256 hibiscus image (*top right*) is not compressible in space, but it is compressible in the wavelet domain since its wavelet coefficients sorted in decreasing corresponding to their absolute values exhibit a power law decay (*bottom right*).

Example 2.1. The best 3-term approximation $\hat{x}^3$ of $x = (1, 3, 0, 2, -2)$ is $\hat{x}^3 = (0, 3, 0, 2, -2)$ and $\sigma_3(x)_p = 1$. What is the 2-term approximation of x? Is this unique?

If $x \in \Sigma_s$, then clearly $\sigma_s(x)_p = 0$ for any p. Moreover, one can show that the infimum is achieved and the procedure of keeping only the s largest coefficients in magnitude - also called *thresholding* - results in the optimal approximation as measured by (2.12) for all ℓ_p-norms. It is useful to introduce the following notion.

Definition 2.5 (Decreasing rearrangement)**.** For any vector $x = (x_1, \dots, x_n)^\top \in \mathbb{R}^n$, the *decreasing rearrangement* of x

$$x_\downarrow := (x_{[1]}, \dots, x_{[n]})^\top, \qquad x_{[1]} \geq \cdots \geq x_{[n]}, \tag{2.14}$$

denotes the vector with components of x rearranged in decreasing order.

Proposition 2.5. *For $x \in \mathbb{R}^n$, define $|x| \in \mathbb{R}^n_+$ as $|x|_i = |x_i|$, $i \in [n]$ and set $x^* := |x|_\downarrow$. Then we have*

(i)

$$\sigma_s(x)_p = \begin{cases} \left(\sum_{i=s+1}^n (x_i^*)^p\right)^{1/p}, & p \in (0,\infty), \\ x_{s+1}^*, & p = \infty. \end{cases}$$

(ii)

$$\sigma_s(x)_q \leq \frac{1}{s^{1/p-1/q}} \|x\|_p, \qquad \textit{for any} \quad q > p > 0.$$

Proof. (i) Exercise! Observe that sorting the components of x does not make a difference.
(ii) If $x^* \in \mathbb{R}^n_+$ is the decreasing rearrangement of $x \in \mathbb{R}^n$ we obtain in view of

$$x_i^* \leq x_s^*, \quad \forall i \geq s+1$$

and decrease of $p \mapsto (x_i^*/x_s^*)^p$ for each $i \geq s+1$,

$$\begin{aligned} \sigma_s(x)_q^q = \sum_{i=s+1}^n (x_i^*)^q &\leq (x_s^*)^{q-p} \sum_{i=s+1}^n (x_i^*)^p \leq \left(\frac{1}{s}\sum_{i=1}^s (x_i^*)^p\right)^{\frac{q-p}{p}} \sum_{i=s+1}^n (x_i^*)^p \\ &\leq \left(\frac{1}{s}\|x\|_p^p\right)^{\frac{q-p}{p}} \|x\|_p^p = \frac{1}{s^{q/p-1}}\|x\|_p^q. \end{aligned}$$

Now we can take the $1/q$ power of both sides of the above inequality.

□

Definition 2.6. Let $1 \leq q < \infty$ and $r > 0$. The signal $x \in \mathbb{R}^n$ is called *q-compressible* (or compressible w.r.t. the ℓ_q-norm) with constant C and rate r if there exist constants $C, r > 0$ such that

$$\sigma_s(x)_q \leq C \cdot s^{-r}$$

holds for all $s \in [n]$.

Informally, we call $x \in \mathbb{R}^n$ a compressible vector if $\sigma_s(x)_q$ decays quickly in s. According to Prop. 2.5(ii), this happens in particular if x belongs to the unit ℓ_p-ball for some small $p > 0$, where the unit ℓ_p-ball is defined by

$$\mathbb{B}_{\ell_p} := \{y \in \mathbb{R}^n \colon \|y\|_p \leq 1\}.$$

Remark 2.3. One can show [FR13, Thm. 2.5] that for any $q > p > 0$ and any $x \in \mathbb{R}^n$, the inequality

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

holds with

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

Note that for $p = 1$ and $q = 2$ this leads to

$$\sigma_s(x)_2 \leq \frac{1}{2\sqrt{s}} \|x\|_1. \tag{2.15}$$

An alternative way to think about compressible signals is to consider the rate of decay of their coefficients. For many important classes of signals there exist bases such that the coefficients obey a power law decay, compare Fig. 2.5. In such cases signals are highly compressible. Specifically, if $x = \Phi z$ and we sort the coefficients z_i in decreasing order of their absolute value according to Def. (2.5), then we say that the coefficients obey a power law decay if there exist constants $C, r > 0$ such that

$$|z|_{[i]} \leq Ci^{-r}, \quad \forall i \in [n].$$

The larger r is, the faster the magnitudes decay, and the more compressible the signal x is in Φ. Because the magnitudes of their coefficients decay so rapidly, compressible signals can be represented accurately by $s \ll n$ coefficients.