Mathematics of Information

Andrea Ghirlanda

August 26, 2024

Mathematics of Information summary created by *aghirlanda@student.ethz.ch*

1 Frame Theory

1.1 Signal Expansion in Finite-D Hilbert Spaces

1.1.1 Orthonormal Bases (ONB):

ONB: Let $\{\mathbf{e}_k\}_{k=1}^M, \mathbf{e}_k \in \mathbb{C}^M$ ONB

Analysis Matrix: $\mathbf{T} \triangleq \begin{bmatrix} \mathbf{e}_1^{\mathrm{H}} \\ \vdots \\ \mathbf{e}_M^{\mathrm{H}} \end{bmatrix}$ **Synthesis Matrix:** $\mathbf{T}^{\mathrm{H}}$

$\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_M$ orthonormal $\Rightarrow$ $\mathbf{T}$ unitary $\Rightarrow$ $\mathbf{T}^H = \mathbf{T}^{-1}$ $\Rightarrow \mathbf{x} = \mathbf{T}^{\mathrm{H}}\mathbf{T}\mathbf{x} = \sum_{k=1}^M \langle \mathbf{x}, \mathbf{e}_k \rangle \mathbf{e}_k$

Orthonormal Bases are Norm-Preserving: Let $\mathbf{c} = \mathbf{T}\mathbf{x}$, then $\|\mathbf{c}\|^2 = \mathbf{c}^{\mathrm{H}}\mathbf{c} = \mathbf{x}^{\mathrm{H}}\mathbf{T}^{\mathrm{H}}\mathbf{T}\mathbf{x} = \mathbf{x}^{\mathrm{H}}\mathbf{I}_M\mathbf{x} = \|\mathbf{x}\|^2$

1.1.2 General Bases:

Basis: Let $\{\mathbf{e}_k\}_{k=1}^M, \mathbf{e}_k \in \mathbb{C}^M$ Basis

Analysis Matrix: $\mathbf{T} \triangleq \begin{bmatrix} \mathbf{e}_1^{\mathrm{H}} \\ \vdots \\ \mathbf{e}_M^{\mathrm{H}} \end{bmatrix}$

Synthesis matrix: $\tilde{\mathbf{T}}^{\mathrm{H}} \triangleq \begin{bmatrix} \tilde{\mathbf{e}}_1 & \cdots & \tilde{\mathbf{e}}_M \end{bmatrix}$

Goal: Find $\{\tilde{\mathbf{e}}_1, \ldots, \tilde{\mathbf{e}}_M\}, \tilde{\mathbf{e}}_k \in \mathbb{C}^M, k = 1, \ldots, M$ s.t $\mathbf{x} = \sum_{k=1}^M \langle \mathbf{x}, \mathbf{e}_k \rangle \tilde{\mathbf{e}}_k = \tilde{\mathbf{T}}^{\mathrm{H}}\mathbf{T}\mathbf{x}\ \forall \mathbf{x} \in \mathbb{C}^M$

Since $\{\mathbf{e}_k\}_{k=1}^M$ are lin. indep., $\mathbf{T}$ is Full-Rank $\Rightarrow$ The goal is equivalent to finding $\tilde{\mathbf{T}}^{\mathrm{H}} = \mathbf{T}^{-1}$ which is unique.

The sets $\{\mathbf{e}_k\}_{k=1}^M$ and $\{\tilde{\mathbf{e}}_k\}_{k=1}^M$ are biorthonormal bases

Biorthonormal Bases are not Norm-Preserving

1.1.3 Redundant Signal Expansion

Let $\{\mathbf{g}_1, \ldots, \mathbf{g}_N\}, \mathbf{g}_k, \mathbf{x} \in \mathbb{C}^M, k = 1, \ldots, N$ with $N > M$

Analysis Mat $\mathbf{T} \triangleq \begin{bmatrix} \mathbf{g}_1^{\mathrm{H}} \\ \vdots \\ \mathbf{g}_N^{\mathrm{H}} \end{bmatrix}$; synth Mat $\tilde{\mathbf{T}}^{\mathrm{H}} \triangleq [\tilde{\mathbf{g}}_1 \cdots \tilde{\mathbf{g}}_N]$

Goal: Find $\{\tilde{\mathbf{g}}_1, \ldots, \tilde{\mathbf{g}}_N\}, \tilde{\mathbf{g}}_k \in \mathbb{C}^M, k = 1, \ldots, N$ s.t. $\mathbf{x} = \sum_{k=1}^N \langle \mathbf{x}, \mathbf{g}_k \rangle \tilde{\mathbf{g}}_k = \tilde{\mathbf{T}}^{\mathrm{H}}\mathbf{T}\mathbf{x} \Rightarrow$ find left inverse $\tilde{\mathbf{T}}^{\mathrm{H}}$ of $\mathbf{T}$

$\mathbf{T}$ is left-invertible $\iff \mathbb{C}^M = \operatorname{span}\{\mathbf{g}_k\}_{k=1}^N$.

$\tilde{\mathbf{T}}^{\mathrm{H}}$ not unique, ∞-many left-inverses.

THEOREM 1.6: Let $\mathbf{A} \in \mathbb{C}^{N\times M}, N \geq M$. Assume that $\operatorname{rank}(\mathbf{A}) = M$. Then $\mathbf{A}^\dagger \triangleq \left(\mathbf{A}^{\mathrm{H}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{H}}$ is a left-inverse of $\mathbf{A}$, i.e., $\mathbf{A}^\dagger\mathbf{A} = \mathbf{I}_M$. Moreover, the general solution $\mathbf{L} \in \mathbb{C}^{M\times N}$ of the equation $\mathbf{L}\mathbf{A} = \mathbf{I}_M$ is given by

$$\mathbf{L} = \mathbf{A}^\dagger + \mathbf{M}\left(\mathbf{I}_N - \mathbf{A}\mathbf{A}^\dagger\right)$$

where $\mathbf{M} \in \mathbb{C}^{M\times N}$ is an arbitrary matrix.

Proof: $\mathbf{A}^\dagger\mathbf{A} = \left(\mathbf{A}^{\mathrm{H}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{H}}\mathbf{A} = \mathbf{I}_M$,

$\mathbf{L}\mathbf{A} = \left(\mathbf{A}^\dagger + \mathbf{M}\left(\mathbf{I}_N - \mathbf{A}\mathbf{A}^\dagger\right)\right)\mathbf{A} = \underbrace{\mathbf{A}^\dagger\mathbf{A}}_{\mathbf{I}_M} + \mathbf{M}\mathbf{A} - \mathbf{M}\mathbf{A}\underbrace{\mathbf{A}^\dagger\mathbf{A}}_{\mathbf{I}_M}$

$= \mathbf{I}_M + \mathbf{M}\mathbf{A} - \mathbf{M}\mathbf{A} = \mathbf{I}_M$ □

The previous theorem leads to $\begin{bmatrix} \tilde{\mathbf{g}}_1 & \cdots & \tilde{\mathbf{g}}_N \end{bmatrix} = \mathbf{L}$ where $\mathbf{L} = \mathbf{T}^\dagger + \mathbf{M}\left(\mathbf{I}_N - \mathbf{T}\mathbf{T}^\dagger\right)$

Canonical Dual: $[\tilde{\mathbf{g}}_1 \cdots \tilde{\mathbf{g}}_N] = \mathbf{T}^\dagger = \left(\mathbf{T}^{\mathrm{H}}\mathbf{T}\right)^{-1}\mathbf{T}^{\mathrm{H}}$

$\Rightarrow \tilde{\mathbf{g}}_k = \left(\mathbf{T}^{\mathrm{H}}\mathbf{T}\right)^{-1}\mathbf{g}_k, \quad k = 1, \ldots, N$

Similar to biorthonormal bases, redundant sets of vectors are, in general, not norm-preserving $\|\mathbf{c}\|^2 = \mathbf{x}^{\mathrm{H}}\mathbf{T}^{\mathrm{H}}\mathbf{T}\mathbf{x}$.

The following theorem holds for both redundant sets and biorthonormal basis

Rayleigh-Ritz theorem:

$\lambda_{\min}\left(\mathbf{T}^{\mathrm{H}}\mathbf{T}\right)\|\mathbf{x}\|^2 \leq \|\mathbf{c}\|^2 \leq \lambda_{\max}\left(\mathbf{T}^{\mathrm{H}}\mathbf{T}\right)\|\mathbf{x}\|^2$

1.2 Frames for General Hilbert Spaces

Let $\{g_k\}_{k\in\mathcal{K}}$ (with $\mathcal{K}$ countable set).

In this section we assume that $\{g_k\}_{k\in\mathcal{K}}$ is a

Bessel Sequence: $\sum_{k\in\mathcal{K}} |\langle x, g_k\rangle|^2 < \infty\ \forall x \in \mathcal{H}$

Definition 1.7: The linear operator $\mathbb{T}$ is defined as the operator that maps the Hilbert space $\mathcal{H}$ into the space l^2 of square-summable complex sequences $\mathbb{T} : \mathcal{H} \to l^2$, by assigning to each signal $x \in \mathcal{H}$ the sequence of inner products $\langle x, g_k\rangle$ according to

$$\mathbb{T} : x \to \{\langle x, g_k\rangle\}_{k\in\mathcal{K}}.$$

Note that $\|\mathbb{T}x\|^2 = \sum_{k\in\mathcal{K}} |\langle x, g_k\rangle|^2$, i.e., the energy $\|\mathbb{T}x\|^2$ of $\mathbb{T}x$ can be expressed as

$$\|\mathbb{T}x\|^2 = \sum_{k\in\mathcal{K}} |\langle x, g_k\rangle|^2$$

Properties that $\mathbb{T}$ should satisfy:

1. Signal x can be perfectly reconstructed from the coefficients (i.e $\mathbb{T}x = \mathbb{T}y \Rightarrow x = y$) $\{\langle x, g_k\rangle\}_{k\in\mathcal{K}} \Rightarrow$ $\mathbb{T}$ has to be left-invertible (i.e $\mathbb{T}$ invertible on $\mathcal{R}(\mathbb{T}) = \left\{y \in l^2 : y = \mathbb{T}x, x \in \mathcal{H}\right\} \iff A\|x - y\|^2 \leq \|\mathbb{T}x - \mathbb{T}y\|^2$ $\forall x, y \in \mathcal{H}$)

2. $A\|x\|^2 \leq \|\mathbb{T}x\|^2 \leq B\|x\|^2$ for finite constant A, B

Definition 1.8. A set of elements $\{g_k\}_{k\in\mathcal{K}}, g_k \in \mathcal{H}, k \in \mathcal{K}$ is called a frame for the Hilbert space $\mathcal{H}$ if

$$A\|x\|^2 \leq \sum |\langle x, g_k\rangle|^2 \leq B\|x\|^2, \forall x \in \mathcal{H}$$

Where $A, B \in \mathbb{R}$ s.t $0 < A \leq B < \infty$ are called frame bounds. The largest valid A and the smallest valid B are called **The** frame bounds. The existence of a lower frame bound $A > 0$ guarantees that $\mathbb{T}$ is left-invertible.

Definition 1.11. A set of elements $\{g_k\}_{k\in\mathcal{K}}, g_k \in \mathcal{H}, k \in \mathcal{K}$ is complete for the Hiblert space $\mathcal{H}$ if $\langle x, g_k\rangle = 0 \quad \forall k \in \mathcal{K}, x \in \mathcal{H} \Rightarrow x = 0$

Definition 1.12. The linear operator $\mathbb{T}^*$ is defined as

$$\mathbb{T}^* : l^2 \to \mathcal{H}$$

$$\mathbb{T}^* : \{c_k\}_{k\in\mathcal{K}} \to \sum_{k\in\mathcal{K}} c_k g_k.$$

Definition 1.13. Let $\mathbb{A} : \mathcal{H} \to \mathcal{H}'$, $\mathbb{A}$ bounded. The unique bounded linear operator $\mathbb{A}^* : \mathcal{H}' \to \mathcal{H}$ that satisfies $\langle \mathbb{A}x, y\rangle = \langle x, \mathbb{A}^* y\rangle \quad \forall x \in \mathcal{H}, y \in \mathcal{H}'$ is the adjoint of $\mathbb{A}$.

Note $\mathbb{T}^*$ is the generalization of $\mathbf{T}^{\mathrm{H}}$ to the ∞-Dim setting.

1.2.1 The Frame Operator

Definition 1.14. Let $\{g_k\}_{k\in\mathcal{K}}$ be a frame for the Hilbert space $\mathcal{H}$. The operator $\mathbb{S} : \mathcal{H} \to \mathcal{H}$ defined as

$$\mathbb{S} = \mathbb{T}^*\mathbb{T}$$

$$\mathbb{S}x = \sum_{k\in\mathcal{K}} \langle x, g_k\rangle g_k$$

is called the frame operator.

Note: $\|\mathbb{T}x\|^2 = \langle \mathbb{T}x, \mathbb{T}x\rangle = \langle \mathbb{T}^*\mathbb{T}x, x\rangle = \langle \mathbb{S}x, x\rangle$

$\Rightarrow A\|x\|^2 \leq \langle \mathbb{S}x, x\rangle \leq B\|x\|^2$

Theorem 1.15. $\mathbb{S}$ satisfies the follwing properties:

1. $\mathbb{S}$ is linear and bounded;

2. $\mathbb{S}$ is self-adjoint (i.e, $\mathbb{S}^* = \mathbb{S}$)

3. $\mathbb{S}$ is positive definite (i.e $\langle \mathbb{S}x, x\rangle > 0 \quad \forall x \in \mathcal{H}, x \neq 0$);

4. $\mathbb{S}$ a unique self-adjoint positive definite square root $\mathbb{S}^{1/2}$.

Proof.: 1. Linearity and boundedness of $\mathbb{S}$ follow from the fact that $\mathbb{S}$ is obtained by cascading a bounded linear operator and its adjoint

2. $\mathbb{S}^* = (\mathbb{T}^*\mathbb{T})^* = \mathbb{T}^*\mathbb{T} = \mathbb{S}$

3. $\langle \mathbb{S}x, x\rangle \geq A\|x\|^2 > 0 \quad \forall x \in \mathcal{H}, x \neq 0$

4. follows from 2., 3. and Lemma 1.16. (which states that all self-adjoint pos. def. operators have a self adjoint pos. def. sqrt.). □

Theorem 1.17. Let A,B tightest possible frame bounds for frame with frame op. $\mathbb{S}$. Then $A = \lambda_{\min}$ and $B = \lambda_{\max}$ where λ_{min} and λ_{max} denote the smallest and largest spectral values of $\mathbb{S}$ respectively.

1.2.2 The Canonical Dual Frame

We saw for finite case: $\tilde{\mathbf{g}}_k = \left(\mathbf{T}^{\mathrm{H}}\mathbf{T}\right)^{-1}\mathbf{g}_k, \quad k = 1,\ldots,N$
But for ∞-dim case $\mathbb{S} = \mathbb{T}^*\mathbb{T} \Rightarrow \tilde{\mathbf{g}}_k = \mathbb{S}^{-1}\mathbf{g}_k, k = 1,\ldots,N$
$\exists$ unique operator $\mathbb{S}^{-1}$ such that $\mathbb{S}\mathbb{S}^{-1} = \mathbb{S}^{-1}\mathbb{S} = \mathbb{I}_{\mathcal{H}}$

Theorem 1.18. The following properties hold:
1. $\mathbb{S}^{-1}$ is self-adjoint (i.e $(\mathbb{S}^{-1})^* = \mathbb{S}^{-1}$)
2. $\mathbb{S}^{-1}$ satisfites

$$\frac{1}{B} = \inf_{x\in\mathcal{H}} \frac{\langle \mathbb{S}^{-1}x, x\rangle}{\|x\|^2} \text{ and } \frac{1}{A} = \sup_{x\in\mathcal{H}} \frac{\langle \mathbb{S}^{-1}x, x\rangle}{\|x\|^2}$$

3. $\mathbb{S}^{-1}$ is positive definite

Theorem 1.19. Let $\{g_k\}_{k\in\mathcal{K}}$ frame for $\mathcal{H}$ with frame operator $\mathbb{S}$ and frame bounds A and B. Then the set $\{\tilde{g}_k\}_{k\in\mathcal{K}}$ given by $\tilde{g}_k = \mathbb{S}^{-1}g_k, \quad k \in \mathcal{K}$ is a frame for $\mathcal{H}$ with frame bounds $\tilde{A} = 1/B$ and $\tilde{B} = 1/A$
The analysis operator associated with $\{\tilde{g}_k\}_{k\in\mathcal{K}}$ defined as:

$$\tilde{\mathbb{T}} : \mathcal{H} \to l^2$$
$$\tilde{\mathbb{T}} : x \to \{\langle x, \tilde{g}_k\rangle\}_{k\in\mathcal{K}}$$

Satisfies $\tilde{\mathbb{T}} = \mathbb{T}\mathbb{S}^{-1} = \mathbb{T}\left(\mathbb{T}^*\mathbb{T}\right)^{-1}$

Proof. $\mathbb{S}^{-1}$ self-adj.$\Rightarrow \langle x, \tilde{g}_k\rangle = \left\langle x, \mathbb{S}^{-1}g_k\right\rangle = \left\langle \mathbb{S}^{-1}x, g_k\right\rangle$ $\forall x \in \mathcal{H} \Rightarrow \sum_{k\in\mathcal{K}} |\langle x, \tilde{g}_k\rangle|^2 = \sum_{k\in\mathcal{K}} \left|\left\langle x, \mathbb{S}^{-1}g_k\right\rangle\right|^2 = \sum_{k\in\mathcal{K}} \left|\left\langle \mathbb{S}^{-1}x, g_k\right\rangle\right|^2 = \left\langle \mathbb{S}\left(\mathbb{S}^{-1}x\right), \mathbb{S}^{-1}x\right\rangle = \left\langle x, \mathbb{S}^{-1}x\right\rangle = \left\langle \mathbb{S}^{-1}x, x\right\rangle$ Therefore:

$$\frac{1}{B}\|x\|^2 \le \sum_{k\in\mathcal{K}} |\langle x, \tilde{g}_k\rangle|^2 \le \frac{1}{A}\|x\|^2$$

i.e., the set $\{\tilde{g}_k\}_{k\in\mathcal{K}}$ constitutes a frame for $\mathcal{H}$ with frame bounds $\tilde{A} = 1/B$ and $\tilde{B} = 1/A$ which are **The** frame bounds (theorem 1.18). It remains to show $\tilde{\mathbb{T}} = \mathbb{T}\mathbb{S}^{-1}$:

$$\tilde{\mathbb{T}}x = \{\langle x, \tilde{g}_k\rangle\}_{k\in\mathcal{K}} = \left\{\left\langle x, \mathbb{S}^{-1}g_k\right\rangle\right\}_{k\in\mathcal{K}}$$
$$= \left\{\left\langle \mathbb{S}^{-1}x, g_k\right\rangle\right\}_{k\in\mathcal{K}} = \mathbb{T}^{-1}x \quad \square$$

We call $\{\tilde{g}_k\}_{k\in\mathcal{K}}$ the canonical dual frame associated to the frame $\{g_k\}_{k\in\mathcal{K}}$.

1.2.3 Signal Expansion

Theorem 1.22. Let $\{g_k\}_{k\in\mathcal{K}}$ and $\{\tilde{g}_k\}_{k\in\mathcal{K}}$ be canonical dual frames for $\mathcal{H}$. Every signal $x \in \mathcal{H}$ can be decomposed as follows:

$$x = \mathbb{T}^*\tilde{\mathbb{T}}x = \sum_{k\in\mathcal{K}} \langle x, \tilde{g}_k\rangle\, g_k$$
$$x = \tilde{\mathbb{T}}^*\mathbb{T}x = \sum_{k\in\mathcal{K}} \langle x, g_k\rangle\, \tilde{g}_k.$$

Proof:

$$\mathbb{T}^*\tilde{\mathbb{T}}x = \sum_{k\in\mathcal{K}} \langle x, \tilde{g}_k\rangle\, g_k = \sum_{k\in\mathcal{K}} \langle x, \mathbb{S}^{-1}g_k\rangle\, g_k$$
$$= \sum_{k\in\mathcal{K}} \langle \mathbb{S}^{-1}x, g_k\rangle\, g_k = \mathbb{S}\mathbb{S}^{-1}x = x$$

This proves that $\mathbb{T}^*\tilde{\mathbb{T}} = \mathbb{I}_{\mathcal{H}}$.
The proof of $\tilde{\mathbb{T}}^*\mathbb{T} = \mathbb{I}_{\mathcal{H}}$ is similar. $\square$

Definition 1.25. A frame $\{g_k\}_{k\in\mathcal{K}}$ with tightest possible frame bounds $A = B$ is called **tight frame**.

Theorem 1.26. Let $\{g_k\}_{k\in\mathcal{K}}$ be a frame for $\mathcal{H}$. $\{g_k\}_{k\in\mathcal{K}}$ is tight with frame bound A $\iff$ its corresponding frame operattor satisfies $\mathbb{S} = A\mathbb{I}_{\mathcal{H}}$, or equivalently, if:

$$x = \frac{1}{A}\sum_{k\in\mathcal{K}} \langle x, g_k\rangle\, g_k \quad \forall x \in \mathcal{H}$$

Theorem 1.28. Let $\{g_k\}_{k\in\mathcal{K}}$ be a frame for $\mathcal{H}$ with frame operator $\mathbb{S}$. Denote the positive definite sqrt of $\mathbb{S}^{-1}$ by $\mathbb{S}^{-1/2}$. Then $\left\{\mathbb{S}^{-1/2}g_k\right\}_{k\in\mathcal{K}}$ is a tight frame for $\mathcal{H}$ with frame bound $A = 1$, i.e.

$$x = \sum_{k\in\mathcal{K}} \left\langle x, \mathbb{S}^{-1/2}g_k\right\rangle \mathbb{S}^{-1/2}g_k \quad \forall x \in \mathcal{H}$$

Proof: Since $\mathbb{S}^{-1}$ is self-adjoint and positive definite by Theorem 1.18, it has by Lemma 1.16, a unique self-adjoint positive definite sqrt $\mathbb{S}^{-1/2}$ that commutes with $\mathbb{S}^{-1}$ (and with $\mathbb{S}$ by the following $\mathbb{S}^{-1/2}\mathbb{S}^{-1} = \mathbb{S}^{-1}\mathbb{S}^{-1/2}\mathbb{S}\mathbb{S}^{-1/2}\mathbb{S}^{-1} = \mathbb{S}^{-1/2}\mathbb{S}\mathbb{S}^{-1/2} = \mathbb{S}^{-1/2}\mathbb{S}$). Note the following:

$$x = \mathbb{S}^{-1}\mathbb{S}x = \mathbb{S}^{-1/2}\mathbb{S}^{-1/2}\mathbb{S}x$$
$$= \mathbb{S}^{-1/2}\mathbb{S}\mathbb{S}^{-1/2}x = \sum_{k\in\mathcal{K}} \left\langle \mathbb{S}^{-1/2}x, g_k\right\rangle \mathbb{S}^{-1/2}g_k$$
$$= \sum_{k\in\mathcal{K}} \left\langle x, \mathbb{S}^{-1/2}g_k\right\rangle \mathbb{S}^{-1/2}g_k \quad \square$$

Theorem 1.30 Let $N > M$. Suppose $\{\mathbf{g}_1,\ldots,\mathbf{g}_N\}, \mathbf{g}_k \in \mathcal{H}, k = 1,\ldots,N$ is a tight frame for an M-Dimensional $\mathcal{H}$ with frame bound $A = 1$. Then, there exists an N-Dimensional $\mathcal{K} \supset \mathcal{H}$ and ONB $\{\mathbf{e}_1,\ldots,\mathbf{e}_N\}$ for $\mathcal{K}$ s.t. $\mathbb{P}\mathbf{e}_k = \mathbf{g}_k, k = 1,\ldots,N$, where $\mathbb{P} : \mathcal{K} \to \mathcal{K}$ is the orthogonal projection onto $\mathcal{H}$. (no proof provided, only example).

Example 1.31 Consider Hilbert space $\mathcal{K} = \mathbb{R}^3$ and assume $\mathcal{H} \subset \mathcal{K}$ with plane
$\mathcal{H} = \text{span}\left\{\begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^\top, \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^\top\right\}$

1.2.4 Exact Frames and Biorthonormality

Definition 1.32. Let $\{g_k\}_{k\in\mathcal{K}}$ frame for $\mathcal{H}$. We call $\{g_k\}_{k\in\mathcal{K}}$ **exact** if $\forall m \in \mathcal{K}\, \{g_k\}_{k\neq m}$ is incomplete for $\mathcal{H}$. $\{g_k\}_{k\in\mathcal{K}}$ **inexact** if $\exists g_m$ that can be removed from the frame, s.t $\{g_k\}_{k\neq m}$ is still a frame for $\mathcal{H}$.

Lemma 1.33 $\{g_k\}_{k\in\mathcal{K}}$ frame for $\mathcal{H}$ and $\{\tilde{g}_k\}_{k\in\mathcal{K}}$ its canonical dual frame. For fixed $x \in \mathcal{H}$, let $c_k = \langle x, \tilde{g}_k\rangle$ s.t $x = \sum_{k\in\mathcal{K}} c_k g_k$. If $\exists$ scalars $\{a_k\}_{k\in\mathcal{K}} \neq \{c_k\}_{k\in\mathcal{K}}$ s.t. $x = \sum_{k\in\mathcal{K}} a_k g_k$ then we must have

$$\sum_{k\in\mathcal{K}} |a_k|^2 = \sum_{k\in\mathcal{K}} |c_k|^2 + \sum_{k\in\mathcal{K}} |c_k - a_k|^2$$

Proof. $c_k = \langle x, \tilde{g}_k\rangle = \left\langle x, \mathbb{S}^{-1}g_k\right\rangle = \left\langle \mathbb{S}^{-1}x, g_k\right\rangle = \langle \tilde{x}, g_k\rangle$ with $\tilde{x} = \mathbb{S}^{-1}x \Rightarrow \langle x, \tilde{x}\rangle = \left\langle \sum_{k\in\mathcal{K}} c_k g_k, \tilde{x}\right\rangle = \sum_{k\in\mathcal{K}} c_k \langle g_k, \tilde{x}\rangle =$
$\sum_{k\in\mathcal{K}} c_k c_k^* = \sum_{k\in\mathcal{K}} |c_k|^2 = \left\langle \sum_{k\in\mathcal{K}} a_k g_k, \tilde{x}\right\rangle = \sum_{k\in\mathcal{K}} a_k \langle g_k, \tilde{x}\rangle$
$= \sum_{k\in\mathcal{K}} a_k c_k^* \Rightarrow \sum_{k\in\mathcal{K}} |c_k|^2 = \sum_{k\in\mathcal{K}} a_k c_k^* = \sum_{k\in\mathcal{K}} a_k^* c_k \Rightarrow$
$= \sum_{(k\in\mathcal{K}} |c_k|^2 + \sum_{k\in\mathcal{K}} |c_k|^2 - \sum_{k\in\mathcal{K}} c_k a_k^* - \sum_{k\in\mathcal{K}} c_k^* a_k + \sum_{k\in\mathcal{K}} |a_k|^2$

$\Rightarrow \sum_{k\in\mathcal{K}} |c_k|^2 + \sum_{k\in\mathcal{K}} |c_k - a_k|^2 = \sum_{k\in\mathcal{K}} |a_k|^2$ $\square$

Lemma 1.34. Let $\{g_k\}_{k\in\mathcal{K}}$ frame for $\mathcal{H}$ and $\{\tilde{g}_k\}_{k\in\mathcal{K}}$ its canonical dual frame. Then $\forall m \in \mathcal{K}$ we have:

$$\sum_{k\neq m} |\langle g_m, \tilde{g}_k\rangle|^2 = \frac{1 - |\langle g_m, \tilde{g}_m\rangle|^2 - |1 - \langle g_m, \tilde{g}_m\rangle|^2}{2}$$

Proof. Can represent g_m in two different ways. **1)** $g_m = \sum_{k\in\mathcal{K}} a_k g_k$ with $a_m = 1$ and $a_k = 0$ for $k \neq m$ s.t. $\sum_{k\in\mathcal{K}} |a_k|^2 = 1$ **2)** $g_m = \sum_{k\in\mathcal{K}} c_k g_k$ with $c_k = \langle g_m, \tilde{g}_k\rangle$

$$1 = \sum_{k\in\mathcal{K}} |a_k|^2 = \sum_{k\in\mathcal{K}} |c_k|^2 + \sum_{k\in\mathcal{K}} |c_k - a_k|^2$$
$$= \sum_{k\in\mathcal{K}} |c_k|^2 + |c_m - a_m|^2 + \sum_{k\neq m} |c_k - a_k|^2$$
$$= \sum_{k\in\mathcal{K}} |\langle g_m, \tilde{g}_k\rangle|^2 + |\langle g_m, \tilde{g}_m\rangle - 1|^2 + \sum_{k\neq m} |\langle g_m, \tilde{g}_k\rangle|^2$$
$$= 2\sum_{k\neq m} |\langle g_m, \tilde{g}_k\rangle|^2 + |\langle g_m, \tilde{g}_m\rangle|^2 + |1 - \langle g_m, \tilde{g}_m\rangle|^2$$

and hence:

$$\sum_{k\neq m} |\langle g_m, \tilde{g}_k\rangle|^2 = \frac{1 - |\langle g_m, \tilde{g}_m\rangle|^2 - |1 - \langle g_m, \tilde{g}_m\rangle|^2}{2}$$

Theorem 1.35. Let $\{g_k\}_{k\in\mathcal{K}}$ frame for $\mathcal{H}$ and $\{\tilde{g}_k\}_{k\in\mathcal{K}}$ its canonical dual frame. Then
1. $\{g_k\}_{k\in\mathcal{K}}$ is exact $\iff \langle g_m, \tilde{g}_m\rangle = 1 \forall\, m \in \mathcal{K}$;
2. $\{g_k\}_{k\in\mathcal{K}}$ is inexact $\iff$ $\exists$ at least one $m \in \mathcal{K}$ s.t. $\langle g_m, \tilde{g}_m\rangle \neq 1$

Proof. Recall $\sum_{k\neq m} |\langle g_m, \tilde{g}_k\rangle|^2 = \frac{1-|\langle g_m,\tilde{g}_m\rangle|^2-|1-\langle g_m,\tilde{g}_m\rangle|^2}{2}$

Since $\langle g_m, \tilde{g}_m \rangle = 1$, we have $\sum_{k \neq m} |\langle g_m, \tilde{g}_k \rangle|^2 = 0$ so that $\langle g_m, \tilde{g}_k \rangle = \langle \tilde{g}_m, g_k \rangle = 0$ for all $k \neq m$. But $\tilde{g}_m \neq 0$ since $\langle g_m, \tilde{g}_m \rangle = 1$. Therefore, $\{g_k\}_{k \neq m}$ is incomplete for $\mathcal{H}$, because $\tilde{g}_m \neq 0$ is orthogonal to all elements of the set $\{g_k\}_{k \neq m}$.

Corollary 1.36. Let $\{g_k\}_{k \in \mathcal{K}}$ frame for $\mathcal{H}$.
$\{g_k\}_{k \in \mathcal{K}}$ exact $\iff$ $\{g_k\}_{k \in \mathcal{K}}$ and $\{\tilde{g}_k\}_{k \in \mathcal{K}}$ biorthonormal.

Proof. If $\{g_k\}_{k \in \mathcal{K}}$ is exact, then biorthonormality follows by noting that Theorem 1.35 implies $\langle g_m, \tilde{g}_m \rangle = 1$ for all $m \in \mathcal{K}$, and Lemma 1.34 implies $\sum_{k \neq m} |\langle g_m, \tilde{g}_k \rangle|^2 = 0$ for all $m \in \mathcal{K}$ and thus $\langle g_m, \tilde{g}_k \rangle = 0$ for all $k \neq m$. To show that, conversely, biorthonormality of $\{g_k\}_{k \in \mathcal{K}}$ and $\{\tilde{g}_k\}_{k \in \mathcal{K}}$ implies that the frame $\{g_k\}_{k \in \mathcal{K}}$ is exact, we simply note that $\langle g_m, \tilde{g}_m \rangle = 1$ for all $m \in \mathcal{K}$, by Theorem 1.35, implies that $\{g_k\}_{k \in \mathcal{K}}$ is exact. □

Theorem 1.37. If $\{g_k\}_{k \in \mathcal{K}}$ is an exact frame for $\mathcal{H}$ and $x = \sum_{k \in \mathcal{K}} c_k g_k$, $x \in \mathcal{H}$ then coeffs $\{c_k\}_{k \in \mathcal{K}}$ are unique and given by $c_k = \langle x, \tilde{g}_k \rangle$.

Proof. x can be written as $x = \sum_{k \in \mathcal{K}} \langle x, \tilde{g}_k \rangle g_k$. Now assume that $\exists$ $\{c_k\}_{k \in \mathcal{K}}$ s.t. $x = \sum_{k \in \mathcal{K}} c_k g_k$. Taking the inner product on both sides gives and using biorthonormality relation gives $\langle x, \tilde{g}_m \rangle = \sum_{k \in \mathcal{K}} c_k \langle g_k, \tilde{g}_m \rangle = c_m$. Thus $c_m = \langle x, \tilde{g}_m \rangle \, \forall m \in \mathcal{K}$ □

1.3 The Sampling Theorem

Def. $x(t)$ is bandlimited to B Hz if $\widehat{x}(f) = 0$ for $|f| > B$. Note that this implies implies that the total bandwidth of x(t), counting positive and negative frequencies, is 2B.

Consider sequence of samples $\{x[k] \triangleq x(kT)\}_{k \in \mathbb{Z}}$ of the signal $x(t) \in \mathcal{L}^2(B)$ and compute its discrete-time fourier transform(DTFT):

$$\widehat{x}_d(f) \triangleq \sum_{k=-\infty}^{\infty} x[k] e^{-\mathrm{i}2\pi kf} = \sum_{k=-\infty}^{\infty} x(kT) e^{-\mathrm{i}2\pi kf} = \frac{1}{T} \sum_{k=-\infty}^{\infty} \widehat{x}\left(\frac{f+k}{T}\right)$$

$\Rightarrow \widehat{x}_d(f)$ is simply a periodized version of $\widehat{x}(f)$ Therefore:

$$\widehat{x}(f/T) = \widehat{x}_d(f) T \widehat{h}_{\mathrm{LP}}(f) \text{ with } \widehat{h}_{\mathrm{LP}}(f) = \begin{cases} 1, & |f| \leq BT \\ 0, & \text{otherwise} \end{cases}$$

$\Rightarrow \widehat{x}(f) = T \widehat{h}_{\mathrm{LP}}(fT) \sum_{k=-\infty}^{\infty} x[k] e^{-\mathrm{i}2\pi kfT}$
and we can recover x(t) as follows: (Note: $\mathrm{sinc}(x) \triangleq \frac{\sin(\pi x)}{\pi x}$)

$$x(t) = \int_{-\infty}^{\infty} \widehat{x}(f) e^{\mathrm{i}2\pi tf} df = \int_{-\infty}^{\infty} T \widehat{h}_{\mathrm{LP}}(fT) \sum_{k=-\infty}^{\infty} x[k] e^{-\mathrm{i}2\pi kfT} e^{\mathrm{i}2\pi ft} df$$

$$= \sum_{k=-\infty}^{\infty} x[k] \int_{-\infty}^{\infty} \widehat{h}_{\mathrm{LP}}(fT) e^{\mathrm{i}2\pi fT(t/T-k)} d(fT) = \sum_{k=-\infty}^{\infty} x[k] h_{\mathrm{LP}}\left(\frac{t}{T} - k\right)$$

$$= 2BT \sum_{k=-\infty}^{\infty} x[k] \, \mathrm{sinc}(2B(t-kT)) \text{ where } h_{\mathrm{LP}}(t) = \int_{-\infty}^{\infty} \widehat{h}_{\mathrm{LP}}(f) e^{\mathrm{i}2\pi tf} df$$

Theorem 1.38. (Sampling Theorem) Let Let $x(t) \in \mathcal{L}^2(B)$. Then $x(t)$ is uniquely specified by its samples $x(kT), k \in \mathbb{Z}$, if $1/T \geq 2B$. Specifically, we can reconstruct $x(t)$ from $x(kT)$ according to

$$x(t) = 2BT \sum_{k=-\infty}^{\infty} x(kT) \, \mathrm{sinc}(2B(t-kT))$$

1.3.1 Sampling Theorem as a Frame Expansion

Let $g_k(t) = 2B \, \mathrm{sinc}(2B(t-kT)), k \in \mathbb{Z}$ Then x can be Written as:

$$x(kT) = \int_{-B}^{B} \widehat{x}(f) e^{\mathrm{i}2\pi kfT} df = \langle \widehat{x}, \widehat{g}_k \rangle$$

Where $\widehat{g}_k(f) = \begin{cases} e^{-\mathrm{i}2\pi kfT}, |f| \leq B \\ 0, \text{ otherwise} \end{cases} \Rightarrow x(t) = T \sum_{k=-\infty}^{\infty} \langle x, g_k \rangle \, g_k(t)$

Let's prove that $\{g_k(t)\}_{k \in \mathbb{Z}}$ is a frame for the space $\mathcal{L}^2(B)$:
$\|x\|^2 = \langle x, x \rangle = \left\langle T \sum_{k=-\infty}^{\infty} \langle x, g_k \rangle \, g_k(t), x \right\rangle$
$= T \sum_{k=-\infty}^{\infty} |\langle x, g_k \rangle|^2 \Rightarrow \frac{1}{T} \|x\|^2 = \sum_{k=-\infty}^{\infty} |\langle x, g_k \rangle|^2$
$\Rightarrow \{g_k(t)\}_{k \in \mathbb{Z}}$ is tight frame for $\mathcal{L}^2(B)$ with $A = \frac{1}{T}$
The analysis operator for this frame $\mathbb{T} : \mathcal{L}^2(B) \to l^2$

$$\mathbb{T} : x \to \{\langle x, g_k \rangle\}_{k \in \mathbb{Z}}$$

i.e $\mathbb{T}$ maps $x(t)$ to sequence of samples $\{x(kT)\}_{k \in \mathbb{Z}}$
The action of the adjoint analysis operator
$\mathbb{T}^* : l^2 \to \mathcal{L}^2(B)$ is to perform interpolation according to

$$\mathbb{T}^* : \{c_k\}_{k \in \mathbb{Z}} \to \sum_{k=-\infty}^{\infty} c_k g_k$$

The frame operator $\mathbb{S} : \mathcal{L}^2(B) \to \mathcal{L}^2(B)$ is given by $\mathbb{S} = \mathbb{T}^* \mathbb{T}$ and acts as follows:

$$\mathbb{S} : x(t) \to \sum_{k=-\infty}^{\infty} \langle x, g_k \rangle \, g_k(t)$$

Since $\{g_k(t)\}_{k \in \mathbb{Z}}$ is a tight frame for $\mathcal{L}^2(B)$ with frame bound $A = \frac{1}{T}$, it follows that $\mathbb{S} = (1/T) \mathbb{I}_{\mathcal{L}^2(B)}$. Therefore

$$\tilde{g}_k(t) = \mathbb{S}^{-1} g_k(t) = T \mathbb{I}_{\mathcal{L}^2(B)} g_k(t) = T g_k(t), \quad k \in \mathbb{Z}$$

Critical Sampling, Exact Frame and ONB:

$$\langle g_k, \tilde{g}_k \rangle = T \langle g_k, g_k \rangle = T \|g_k\|^2 = T \|\widehat{g}_k\|^2 = 2BT, \quad \text{for all } k \in \mathbb{Z}$$

For critical sampling $2BT = 1$ hence $\langle g_k, \tilde{g}_k \rangle = 1 \forall k \in \mathbb{Z}$
Now we show that when properly normalized, $\{g_k(t)\}_{k \in \mathbb{Z}}$ is an ONB for $\mathcal{L}^2(B)$: Let $g'_k(t) = \sqrt{T} g_k(t)$
$\Rightarrow x(t) = \sum_{k=-\infty}^{\infty} \langle x, g'_k \rangle \, g'_k(t) \Rightarrow \{g'_k(t)\}_{k \in \mathbb{Z}}$ is a tight frame for $\mathcal{L}^2(B)$ with $A = 1$. Moreover
$\|g'_k\|^2 = T \|g_k\|^2 = 2BT \Rightarrow$For Critical Sampling $\|g'_k\|^2 = 1$
$\forall k \in \mathbb{Z} \Rightarrow \{g'_k(t)\}_{k \in \mathbb{Z}}$ ONB for $\mathcal{L}^2(B)$

Oversampling (inexact frames) In case of oversampling $(1/T > 2B)$ $\{g_k(t)\}_{k \in \mathbb{Z}}$ is inexact frame. This can be seen from: $\langle g_m, \tilde{g}_m \rangle = 2BT < 1, \forall m \in \mathbb{Z}]$
$\Rightarrow \{g_k(t)\}_{k \in \mathbb{Z}}$ is inexact for $\frac{1}{T} > 2B \Rightarrow$the removal of any sample $x(mT)$ from $\{x(kT)\}_{k \in \mathbb{Z}}$ still leaves us with a frame decomposition s.t. $x(t)$ can in theory be recovered from $\{x(kT)\}_{k \neq m}$ but the resulting frame will no longer be tight, which makes the computation of the canonical dual frame complicated in general.

2 Uncertainty Relations and Sparse Signal Recovery

2.1 Introduction

Notation: For $\mathcal{A} \subseteq \{1, \ldots, m\}$, $\mathbf{D}_{\mathcal{A}}$ is mxm Diag. Mat. with $(\mathbf{D}_{\mathcal{A}})_{i,i} = 1$ for $i \in \mathcal{A}$ and $(\mathbf{D}_{\mathcal{A}})_{i,i} = 0$ otherwise.
With $\mathbf{U} \in \mathbb{C}^{m \times m}$ unitary, define orthogonal projection $\mathbf{P}_{\mathcal{A}}(\mathbf{U}) = \mathbf{U} \mathbf{D}_{\mathcal{A}} \mathbf{U}^{\mathrm{H}}$ and set $\mathcal{W}^{\mathbf{U},\mathcal{A}} = \mathcal{R}(\mathbf{P}_{\mathcal{A}}(\mathbf{U}))$. For $\mathbf{x} \in \mathbb{C}^m$, $\mathbf{x}_{\mathcal{A}} = \mathbf{D}_{\mathcal{A}} \mathbf{x}$. For $\mathbf{A} \in \mathbb{C}^{m \times m}$, $|||\mathbf{A}|||_1 = \max_{\mathbf{x}:\|\mathbf{x}\|_1 = 1} \|\mathbf{A}\mathbf{x}\|_1$, $|||\mathbf{A}|||_2 = \max_{\mathbf{x}:\|\mathbf{x}\|_2 = 1} \|\mathbf{A}\mathbf{x}\|_2$ where $\|\mathbf{A}\|_2 = \sqrt{\mathrm{tr}(\mathbf{A}\mathbf{A}^{\mathrm{H}})}$ and $\|\mathbf{A}\|_1 = \sum_{i,j=1}^{m} |A_{i,j}|$. $\mathbf{x} \in \mathbb{C}^m$ is said to be s-sparse if it has at most s nonzero entries.

2.2 Uncertainty Relations in $(\mathbb{C}^m, \|\cdot\|_2)$

Let $\mathbf{U} \in \mathbb{C}^{m \times m}$ be a unitary matrix, $\mathcal{P}, \mathcal{Q} \subseteq \{1, \ldots, m\}$. Let $\Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{U}) = |||\mathbf{D}_{\mathcal{P}} \mathbf{P}_{\mathcal{Q}}(\mathbf{U})|||_2$.By Lemma 2.20 we have:

$$\Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{U}) = \max_{\mathbf{x} \in \mathcal{W}^{\mathbf{U},\mathcal{Q}} \setminus \{\mathbf{0}\}} \frac{\|\mathbf{x}_{\mathcal{P}}\|_2}{\|\mathbf{x}\|_2}$$

An uncertainty relation in $(\mathbb{C}^m, \|\cdot\|_2)$ is an upper bound of the form $\Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{U}) \leq c$ with $c \geq 0 \Rightarrow \|\mathbf{x}_{\mathcal{P}}\|_2 \leq c \|\mathbf{x}\|_2 \forall \mathbf{x} \in \mathcal{W}^{\mathbf{U},\mathcal{Q}}$. By Lemma 2.21 we have:

$$\frac{\|\mathbf{D}_{\mathcal{P}} \mathbf{P}_{\mathcal{Q}}(\mathbf{U})\|_2}{\sqrt{\mathrm{rank}(\mathbf{D}_{\mathcal{P}} \mathbf{P}_{\mathcal{Q}}(\mathbf{U}))}} \leq \Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{U}) \leq \|\mathbf{D}_{\mathcal{P}} \mathbf{P}_{\mathcal{Q}}(\mathbf{U})\|_2$$

Note that $\|\mathbf{D}_{\mathcal{P}} \mathbf{P}_{\mathcal{Q}}(\mathbf{U})\|_2 = \sqrt{\mathrm{tr}(\mathbf{D}_{\mathcal{P}} \mathbf{P}_{\mathcal{Q}}(\mathbf{U}))}$ and $\mathrm{rank}(\mathbf{D}_{\mathcal{P}} \mathbf{P}_{\mathcal{Q}}(\mathbf{U})) = \mathrm{rank}(\mathbf{D}_{\mathcal{P}} \mathbf{U} \mathbf{D}_{\mathcal{Q}} \mathbf{U}^{\mathrm{H}}) \leq \min(|\mathcal{P}|, |\mathcal{Q}|)$
$\Rightarrow \sqrt{\frac{\mathrm{tr}(\mathbf{D}_{\mathcal{P}} \mathbf{P}_{\mathcal{Q}}(\mathbf{U}))}{\min(|\mathcal{P}|,|\mathcal{Q}|)}} \leq \Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{U}) \leq \sqrt{\mathrm{tr}(\mathbf{D}_{\mathcal{P}} \mathbf{P}_{\mathcal{Q}}(\mathbf{U}))}$. If $U = F$: $\sqrt{\mathrm{tr}(\mathbf{D}_{\mathcal{P}} \mathbf{P}_{\mathcal{Q}}(\mathbf{F}))} = \sqrt{\mathrm{tr}(\mathbf{D}_{\mathcal{P}} \mathbf{F} \mathbf{D}_{\mathcal{Q}} \mathbf{F}^{\mathrm{H}})} = \sqrt{\sum_{i \in \mathcal{P}} \sum_{j \in \mathcal{Q}} |\mathbf{F}_{i,j}|^2} = \sqrt{\frac{|\mathcal{P}||\mathcal{Q}|}{m}}$
$\Rightarrow \sqrt{\frac{\max(|\mathcal{P}|,|\mathcal{Q}|)}{m}} \leq \Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{F}) \leq \sqrt{\frac{|\mathcal{P}||\mathcal{Q}|}{m}}$
$\exists \mathcal{P}, \mathcal{Q} \subseteq \{1, \ldots, m\} : \sqrt{\max(|\mathcal{P}|, |\mathcal{Q}|)/m} = \sqrt{|\mathcal{P}||\mathcal{Q}|/m} = 1$ and therefore $\Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{F}) = 1$ For example:

$\mathcal{P} = \left\{\frac{m}{n}, \frac{2m}{n}, \ldots, \frac{(n-1)m}{n}, m\right\}$ and $\mathcal{Q} = \{l+1, \ldots, l+n\}$ with $l \in \{1, \ldots, m\}$ and $\mathcal{Q}$ interpreted circularly in $\{1, \ldots, m\} \Rightarrow \Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{F}) = \sqrt{n/m}$ with n divisor of m.

2.2.1 Coherence-based Uncertainty Relation

Def 2.3. For $\mathbf{A} = (\mathbf{a}_1 \ldots \mathbf{a}_n) \in \mathbb{C}^{m\times n}$ with columns $\|\cdot\|_2$-normalized to 1, the coherence $\mu(\mathbf{A}) \triangleq \max_{i\neq j} \left|\mathbf{a}_i^{\mathrm{H}}\mathbf{a}_j\right|$

Lemma 2.4. Let $\mathbf{U} \in \mathbb{C}^{m\times m}$ unitary and $\mathcal{P}, \mathcal{Q} \subseteq \{1, \ldots, m\}$. Then $\Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{U}) \leq \sqrt{|\mathcal{P}||\mathcal{Q}|}\mu\left(\begin{bmatrix}\mathbf{I} & \mathbf{U}\end{bmatrix}\right)$

Proof.

$$\Delta^2_{\mathcal{P},\mathcal{Q}}(\mathbf{U}) \leq \mathrm{tr}\left(\mathbf{D}_{\mathcal{P}}\mathbf{U}\mathbf{D}_{\mathcal{Q}}\mathbf{U}^{\mathrm{H}}\right) = \sum_{k\in\mathcal{P}}\sum_{l\in\mathcal{Q}} |\mathbf{U}_{k,l}|^2$$

$$\leq |\mathcal{P}||\mathcal{Q}| \max_{k,l} |\mathbf{U}_{k,l}|^2 = |\mathcal{P}||\mathcal{Q}|\mu^2([\mathbf{I} \quad \mathbf{U}])$$

Since $\mu\left(\begin{bmatrix}\mathbf{I} & \mathbf{F}\end{bmatrix}\right) = 1/\sqrt{m}$, this Lemma particularized to $U = F$ recovers $\sqrt{\frac{\max(|\mathcal{P}|,|\mathcal{Q}|)}{m}} \leq \Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{F}) \leq \sqrt{\frac{|\mathcal{P}||\mathcal{Q}|}{m}}$

2.2.2 Concentration inequalities

Def 2.5. Let $\mathcal{P} \subseteq \{1, \ldots, m\}, \varepsilon_{\mathcal{P}} \in [0, 1]$. The vector $\mathbf{x} \in \mathbb{C}^m$ is said to be $\epsilon_{\mathcal{P}}$-concentrated if $\|\mathbf{x} - \mathbf{x}_{\mathcal{P}}\|_2 \leq \varepsilon_{\mathcal{P}}\|\mathbf{x}\|_2$

Lemma 2.6. Let $\mathrm{U} \in \mathbb{C}^{m\times m}$ be unitary and $\mathcal{P}, \mathcal{Q} \subseteq \{1, \ldots, m\}$. Sps. $\exists$ nonzero $\epsilon_{\mathcal{P}}$ concentrated $\mathbf{p} \in \mathbb{C}^m$ and nonzero $\epsilon_{\mathcal{Q}}$-concentrated $\mathbf{q} \in \mathbb{C}^m$ s.t. $\mathbf{p} = \mathrm{Uq}$. Then

$$\Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{U}) \geq [1 - \varepsilon_{\mathcal{P}} - \varepsilon_{\mathcal{Q}}]_+$$

Proof.

$\|\mathbf{p} - \mathbf{P}_{\mathcal{Q}}(\mathrm{U})\mathbf{p}_{\mathcal{P}}\|_2 \leq \|\mathbf{p} - \mathbf{P}_{\mathcal{Q}}(\mathrm{U})\mathbf{p}\|_2 + \|\mathbf{P}_{\mathcal{Q}}(\mathrm{U})\mathbf{p}_{\mathcal{P}} - \mathbf{P}_{\mathcal{Q}}(\mathrm{U})\mathbf{p}\|_2$

$\leq \|\mathbf{p} - \mathbf{P}_{\mathcal{Q}}(\mathrm{U})\mathbf{p}\|_2 + \|\mathbf{P}_{\mathcal{Q}}(\mathrm{U})\|_2 \|\mathbf{p}_{\mathcal{P}} - \mathbf{p}\|_2$

$\leq \left\|\mathbf{p} - \mathbf{U}\mathbf{D}_{\mathcal{Q}}\mathbf{U}^{\mathrm{H}}\mathbf{p}\right\|_2 + \|\mathbf{p}_{\mathcal{P}} - \mathbf{p}\|_2 = \|\mathbf{q} - \mathbf{q}_{\mathcal{Q}}\|_2 + \|\mathbf{p}_{\mathcal{P}} - \mathbf{p}\|_2$

$\leq \varepsilon_{\mathcal{Q}}\|\mathbf{q}\|_2 + \varepsilon_{\mathcal{P}}\|\mathbf{p}\|_2 = (\varepsilon_{\mathcal{P}} + \varepsilon_{\mathcal{Q}})\|\mathbf{p}\|_2$

$\Rightarrow \|\mathbf{P}_{\mathcal{Q}}(\mathrm{U})\mathbf{p}_{\mathcal{P}}\|_2 \geq \left[\|\mathbf{p}\|_2 - \|\mathbf{p} - \mathbf{P}_{\mathcal{Q}}(\mathrm{U})\mathbf{p}_{\mathcal{P}}\|_2\right]_+ \geq \|\mathbf{p}\|_2 [1 - \varepsilon_{\mathcal{P}} - \varepsilon_{\mathcal{Q}}]_+$

$\Rightarrow \left\|\mathbf{P}_{\mathcal{Q}}(\mathrm{U})\mathbf{D}_{\mathcal{P}}\frac{\mathbf{p}}{\|\mathbf{p}\|_2}\right\|_2 \geq [1 - \varepsilon_{\mathcal{P}} - \varepsilon_{\mathcal{Q}}]_+$

$\Rightarrow \|\mathbf{P}_{\mathcal{Q}}(\mathrm{U})\mathbf{D}_{\mathcal{P}}\|_2 \geq [1 - \varepsilon_{\mathcal{P}} - \varepsilon_{\mathcal{Q}}]_+ \Rightarrow \Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{U}) \geq [1 - \varepsilon_{\mathcal{P}} - \varepsilon_{\mathcal{Q}}]_+$ □

Corollary 2.7. Let $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{m\times m}$ be unitary and $\mathcal{P}, \mathcal{Q} \subseteq \{1, \ldots, m\}$. Sps $\exists$ nonzero $\varepsilon_{\mathcal{P}}$-concentrated $\mathbf{p} \in \mathbb{C}^m$ and nonzero $\varepsilon_{\mathcal{Q}}$-concentrated $\mathbf{q} \in \mathbb{C}^m : \mathbf{Ap} = \mathbf{Bq}$. Then

$$|\mathcal{P}||\mathcal{Q}| \geq \frac{[1 - \varepsilon_{\mathcal{P}} - \varepsilon_{\mathcal{Q}}]_+^2}{\mu^2\left(\begin{bmatrix}\mathbf{AB}\end{bmatrix}\right)}$$

Proof. Let $\mathbf{U} = \mathbf{A}^{\mathrm{H}}\mathbf{B}$, then by lemma 2.4 & 2.6 we have

$$[1 - \varepsilon_{\mathcal{P}} - \varepsilon_{\mathcal{Q}}]_+ \leq \Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{U}) \leq \sqrt{|\mathcal{P}||\mathcal{Q}|}\mu\left(\begin{bmatrix}\mathbf{I} & \mathbf{U}\end{bmatrix}\right)$$

Since $\mu\left(\begin{bmatrix}\mathbf{I} & \mathbf{U}\end{bmatrix}\right) = \mu\left(\begin{bmatrix}\mathbf{A} & \mathbf{B}\end{bmatrix}\right)$ the proof is finished □

Corollary 2.8. Let $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{m\times m}$ unitary.
If $\mathbf{Ap} = \mathbf{Bq}$ for nonzero $\mathbf{p}, \mathbf{q} \in \mathbb{C}^m$,
then $\|\mathbf{p}\|_0\|\mathbf{q}\|_0 \geq 1/\mu^2\left(\begin{bmatrix}\mathbf{A} & \mathbf{B}\end{bmatrix}\right)$

2.2.3 Noisy Recovery in $(\mathbb{C}^m, \|\cdot\|_2)$

Lemma 2.9. Let $\mathbf{U} \in \mathbb{C}^{m\times m}$ Unitary, $\mathcal{Q} \subseteq \{1, \ldots, m\}, \mathbf{p} \in \mathcal{W}^{\mathbf{U},\mathcal{Q}}$ and consider $\mathbf{y} = \mathbf{p}_{\mathcal{P}^c} + \mathbf{n}$ where $\mathbf{n} \in \mathbb{C}^m$ and $\mathcal{P}^c = \{1, \ldots, m\}\backslash\mathcal{P}$ with $\mathcal{P} \subseteq \{1, \ldots, m\}$. If $\Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{U}) < 1$ then $\exists \mathbf{L} \in \mathbb{C}^{m\times m}$ s.t.:

$$\|\mathbf{Ly} - \mathbf{p}\|_2 \leq C\|\mathbf{n}_{\mathcal{P}^c}\|_2$$

with $C = 1/\left(1 - \Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{U})\right)$. In particular,

$$|\mathcal{P}||\mathcal{Q}| < \frac{1}{\mu^2([\mathbf{I} \quad \mathbf{U}])}$$

is sufficient for $\Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{U}) < 1$.

3 Compressive Sensing

3.1 Discrete Fourier Transform

$\hat{x}(\theta) = \sum_{n=-\infty}^{\infty} x[n]e^{-2\pi i\theta n} = \sum_{n=0}^{N-1} x[n]e^{-2\pi i\theta n}$ Since $x[n]$ has length N, N samples of $\hat{x}(\theta)$ suffice to specify $x[n] \Rightarrow \widehat{x}[k] := \frac{1}{\sqrt{N}}\widehat{x}\left(\frac{k}{N}\right) = \frac{1}{\sqrt{N}}\sum^{N-1} x[n]\omega_N^{kn}$, for $k = 0, 1, .., N-1$ where $\omega_N = e^{-2\pi \mathrm{i}/N}$ & $\widehat{x}[k+N] = \widehat{x}[k] \forall k \in \mathbb{Z}$

$$\underbrace{\begin{bmatrix}\widehat{x}[0]\\ \widehat{x}[1]\\ \vdots\\ \widehat{x}[N-1]\end{bmatrix}}_{=:\widehat{\mathbf{x}}} = \underbrace{\frac{1}{\sqrt{N}}\begin{bmatrix}1 & 1 & 1 & \cdots & 1\\ 1 & \omega_N & \omega_N^2 & \cdots & \omega_N^{N-1}\\ 1 & \omega_N^2 & \omega_N^4 & \cdots & \omega_N^{2(N-1)}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 1 & \omega_N^{N-1} & \omega_N^{2(N-1)} & \cdots & \omega_N^{(N-1)^2}\end{bmatrix}}_{\mathbf{F}_N} \underbrace{\begin{bmatrix}x[0]\\ x[1]\\ \vdots\\ x[N-1]\end{bmatrix}}_{=:\mathbf{x}}.$$

$\Rightarrow \widehat{\mathbf{x}} = \mathbf{F}_N\mathbf{x}$, and since $\mathbf{F}_N$ unitary $\Rightarrow \mathbf{x} = \mathbf{F}_N^H\widehat{\mathbf{x}}$

$$\widehat{x}[k] = \frac{1}{\sqrt{N}}\sum_{n=0}^{N-1} x[n]e^{-2\pi ikn/N}, \quad \text{with } \widehat{x}[k+N] = \widehat{x}[k]$$
$$x[n] = \frac{1}{\sqrt{N}}\sum_{k=0}^{N-1} \widehat{x}[k]e^{2\pi inn/N}, \quad \text{with } x[n+N] = x[n]$$

Let $\mathbf{F}_N^H = \begin{bmatrix}\mathbf{f}_0 & \mathbf{f}_1 & \cdots & \mathbf{f}_{N-1}\end{bmatrix}$ then

$$\mathbf{x} = \mathbf{F}_N^H\mathbf{F}_N\mathbf{x} = \begin{bmatrix}\mathbf{f}_0 & \mathbf{f}_1 & \cdots & \mathbf{f}_{N-1}\end{bmatrix}\begin{bmatrix}\mathbf{f}_0^H\\ \mathbf{f}_1^H\\ \vdots\\ \mathbf{f}_{N-1}^H\end{bmatrix}\mathbf{x} = \sum_{\ell=0}^{N-1}\langle\mathbf{x}, \mathbf{f}_\ell\rangle\mathbf{f}_\ell$$

Thus the DFT is the expansion of x into an ONB for $\mathbb{C}^N$.

3.1.1 Oversampling

$\widehat{x}[k] = \frac{1}{\sqrt{M}}\widehat{x}\left(\frac{k}{M}\right) = \frac{1}{\sqrt{M}}\sum_{n=0}^{N-1} x[n]e^{-2\pi ikn/M}$,
$k = 0, .., M-1, M > N$

$$\underbrace{\begin{bmatrix}\widehat{x}(0)\\ \widehat{x}[1/M]\\ \vdots\\ \widehat{x}[(M-1)/M]\end{bmatrix}}_{\widehat{\mathbf{x}}} = \underbrace{\frac{1}{\sqrt{M}}\begin{bmatrix}1 & 1 & 1 & \cdots & 1\\ 1 & \omega_M & \omega_M^2 & \cdots & \omega_M^{N-1}\\ 1 & \omega_M^2 & \omega_M^4 & \cdots & \omega_M^{2(N-1)}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 1 & \omega_M^{M-1} & \omega_M^{2(M-1)} & \cdots & \omega_M^{(N-1)(M-1)}\end{bmatrix}}_{\mathbf{F_o}\in\mathbb{C}^{M\times N}} \underbrace{\begin{bmatrix}x[0]\\ x[1]\\ \vdots\\ x[N-1]\end{bmatrix}}_{\mathbf{x}}$$

$\left(\mathbf{F_o}^H\mathbf{F_o}\right)^{-1}\mathbf{F_o}^H\widehat{\mathbf{x}} = \mathbf{F_o}^H\widehat{\mathbf{x}} = x \Rightarrow x[n] = \frac{1}{\sqrt{M}}\sum_{k=0}^{M-1}\widehat{x}[k]\omega_M^{-kn}$

3.1.2 Undersampling

Formula for $\widehat{x}[k]$ same as for oversampling but now $M < N$.
DFT Matrix $\mathbf{F_u} \in \mathbb{C}^{M\times N}$ is wide therefore not invertible.

3.2 Compressive Sensing

Let $\mathbf{y}$ be the measurement vector with $\dim\mathbf{y} = m \times 1$ obtained when compressing the signal $\mathbf{x}$ with $\dim\mathbf{x} = n \times 1$ by means of the matrix $\mathbf{D}$ with $\dim\mathbf{D} = m \times n$ according to $\mathbf{y} = \mathbf{Dx}$. We would like to recover the original vector $\mathbf{x}$.
When set of signals is not constrained, this problem has infinitely many solutions. We consider the case where $\mathbf{x}$ is s-sparse (i.e $\|\mathbf{x}\|_0 \leq s$ which means that x has at most s nonzero entries).
If the support set of $\mathbf{x}$ is **known**, we can choose the first $m = s$ rows of $\mathbf{F}^H$. The resulting D-matrix is Vandermonde, hence full-rank, irrespectively of the support set.
If the support set of x is **unknown** we can choose the first $m = 2s$ rows of $\mathbf{F}^H$. Any $\mathbf{x}_1 - \mathbf{x}_2$ with $\mathbf{x}_1, \mathbf{x}_2 s$-sparse and $\mathbf{x}_1 \neq \mathbf{x}_2$ satisfies $\|\mathbf{D}(\mathbf{x}_1 - \mathbf{x}_2)\|^2 > 0$ since $\mathbf{x}_1 - \mathbf{x}_2$ is $2s$-sparse and the resulting $\mathbf{D}$-matrix is Vandermonde ad hence of rank 2s.
For a given sparsity basis (e.g., wavelets), find a sampling basis such that s-sparse vectors are distinguishable, i.e. $\forall s$-sparse $\mathbf{x}_1, \mathbf{x}_2$ with $\mathbf{x}_1 \neq \mathbf{x}_2, \|\mathbf{D}(\mathbf{x}_1 - \mathbf{x}_2)\|_2^2 > 0$. Hence all collections of 2s columns of $\mathbf{D}$ have to be linearly independent, which is the case only if $m \geq 2s$
Def. 3.1. The spark of a matrix $\mathbf{A}$ denoted by $spark(\mathbf{A})$ is defined as the cardinality of the smallest set of linearly dependend columns.
For a given matrix $\mathbf{D}$ of dimension $m \times n$, uniqueness of recovery of s-sparse vectors $\mathbf{x}$ from the observation $\mathbf{y} = \mathbf{Dx}$ is guaranteed for $s < \frac{\mathrm{spark}(\mathbf{D})}{2}$

3.3 The Recovery Problem (P0)

If $\mathbf{D}$ is a (known) ONB, recovering x from y is simply $\mathbf{D}^{\mathrm{H}}\mathbf{y} = \mathbf{D}^{\mathrm{H}}\mathbf{D}\mathbf{x} = \mathbf{x}$. If instead D a (known) basis: $\mathbf{D}^{-1}\mathbf{y} = \mathbf{D}^{-1}\mathbf{D}\mathbf{x} = \mathbf{x}$. If $\mathbf{D} = \begin{bmatrix} \mathbf{A} & d \end{bmatrix}$ where $\mathbf{A}$ is an ONB and d is an extra column we cannot uniquely determine x from $\mathbf{y} = \mathbf{D}\mathbf{x}$. But if $\mathbf{x}$ is s-sparse and $s < \frac{\mathrm{spark}(\mathbf{D})}{2}$ we can recover $\mathbf{X}$ through a combinatorial search:

$$(P0) \text{ find } argmin\|\hat{\mathbf{x}}\|_0 \text{ subject to } \mathbf{y} = \mathbf{D}\hat{\mathbf{x}}$$

sps. $\|\mathbf{x}\|_0 \leq s$ and $s < \frac{\mathrm{spark}(\mathbf{D})}{2}$. Let $\tilde{\mathbf{x}} \neq \mathbf{x}$ with $\|\tilde{\mathbf{x}}\|_0 \leq s$ and $\mathbf{y} = \mathbf{D}\tilde{\mathbf{x}}$, then $0 = \mathbf{y} - \mathbf{y} = \mathbf{D}\mathbf{x} - \mathbf{D}\tilde{\mathbf{x}} = \mathbf{D}\underbrace{(\mathbf{x} - \tilde{\mathbf{x}})}_{\|\mathbf{x}-\tilde{\mathbf{x}}\|_0 \leq 2s}$ since $2s < \mathrm{spark}(\mathbf{D})$, we know, however, that $\|\mathbf{D}(\mathbf{x}-\tilde{\mathbf{x}})\| > 0, \mathbf{x} - \tilde{\mathbf{x}} \neq 0$ as any set of 2s columns of $\mathbf{D}$ is linearly independent. Therefore (P0) recovers x uniquely. Determining $spark(D)$ is a combinatorial problem and leads to huge computational complexity even for small problem size. Specifically, every set of a columns out of the $\begin{pmatrix} n \\ a \end{pmatrix}$ possible sets has to be checked for linear independence and the parameter a has to be increased starting from two.

Theorem 3.2. (P0) applied to $\mathbf{y} = \mathrm{D}\mathrm{x}$ recovers x if

$$\|\mathrm{x}\|_0 \leq s < \frac{1}{2}\left(1 + \frac{1}{\mu(\mathbf{D})}\right)$$

Proof. Proving this is equivalent to proving $\mathrm{spark}(\mathbf{D}) \geq 1 + 1/\mu(\mathbf{D})$ Consider $x \in \mathbb{C}^n$, $\|x\|_0 = \mathrm{spark}(\mathbf{D})$ and $\mathrm{D}\mathrm{x} = 0$. Then

$\forall l \in \{1, \ldots, n\}$ it holds:

$d_l \mathrm{x}_l = -\sum d_r \mathrm{x}_r$

$\mathbf{x}_l = -\sum_{r \neq l} d_l^{\mathrm{H}} d_r \mathbf{x}_r$

$|\mathrm{x}_l| = \left|\sum_{r \neq l} d_l^{\mathrm{H}} d_r \mathrm{x}_r\right| \leq \sum_{r \neq l} \left|d_l^{\mathrm{H}} d_r\right| |\mathrm{x}_r| \leq \mu(\mathbf{D}) \sum_{r \neq l} |\mathrm{x}_r|$

$(1 + \mu(\mathbf{D}))\,|\mathrm{x}_l| \leq \mu(\mathbf{D})\|\mathrm{x}\|_1$ (added $\mu(\mathbf{D})\,|\mathrm{x}_l|$ on both sides)

$(1 + \mu(\mathbf{D}))\|\mathbf{x}\|_1 \leq \mu(\mathbf{D})\|\mathbf{x}\|_1\,\mathrm{spark}(\mathbf{D})$ (Sum over all l for which $\mathrm{x}_l \neq 0$)

$\Rightarrow \mathrm{spark}(\mathbf{D}) \geq 1 + \frac{1}{\mu(\mathbf{D})}$. □

Notice: Determining $\mu(\mathbf{D})$ has the complexity of doing the first step in the computation of spark(D)

3.4 Basis Pursuit (P1)

In this section we consider the following recovery problem:

$$\text{(P1) find } \arg\min \|\hat{\mathbf{x}}\|_1 \text{ subject to } \mathbf{y} = \mathbf{D}\hat{\mathbf{x}}$$

Def. 3.3.

$$P_1(\mathcal{S}, \mathbf{D}) \triangleq \max_{\alpha \in \mathcal{N}(\mathbf{D}), \boldsymbol{\alpha} \neq 0} \frac{\sum_{k \in \mathcal{S}} |\alpha_k|}{\sum_k |\alpha_k|}$$

Theorem 3.4. Arbitrary fix x with support set $\mathcal{S}$, let $\mathbf{y} = \mathbf{D}x$. If $P_1(\mathcal{S}, \mathbf{D}) < 1/2$, then x is unique sol. to (P1).

Proof. We need to prove that $\forall \alpha \in \mathcal{N}(D)$

$$\sum_k |\mathbf{x}_k + \alpha_k| > \sum_k |\mathbf{x}_k|$$

(reminder: reverse triangle ineq: $|a + b| \geq |a| - |b|$)

$$\sum_k |\mathbf{x}_k + \alpha_k| = \sum_{k \notin \mathcal{S}} |\mathbf{x}_k + \alpha_k| + \sum_{k \in \mathcal{S}} |\mathbf{x}_k + \alpha_k|$$
$$= \sum_{k \notin \mathcal{S}} |\alpha_k| + \sum_{k \in \mathcal{S}} |\mathbf{x}_k + \alpha_k| \geq \sum_{k \notin \mathcal{S}} |\alpha_k| + \sum_{k \in \mathcal{S}} |\mathbf{x}_k| - \sum_{k \in \mathcal{S}} |\alpha_k|$$

Therefore the theorem follows from

$$\sum_{k \notin \mathcal{S}} |\alpha_k| > \sum_{k \in \mathcal{S}} |\alpha_k|$$

Adding $\sum_{k \in \mathcal{S}} |\alpha_k|$ to both sides of the above equation results in

$$\sum_k |\alpha_k| > 2 \sum_{k \in \mathcal{S}} |\alpha_k| \Rightarrow \underbrace{\frac{\sum_{k \in \mathcal{S}} |\alpha_k|}{\sum_k |\alpha_k|}}_{P_1(\mathcal{S}, \mathbf{D})} < \frac{1}{2}$$

But this is satisfied $\forall \alpha \in \mathcal{N}(\mathbf{D})$ since $P_1(\mathcal{S}, \mathbf{D}) < \frac{1}{2}$ by assumption□

Theorem 3.5. (P1) applied to $\mathbf{y} = \mathbf{D}\mathbf{x}$ recovers x if

$$\|\mathrm{x}\|_0 < \frac{1}{2}\left(1 + \frac{1}{\mu(\mathbf{D})}\right)$$

Proof. Consider $\alpha \in \mathcal{N}(\mathbf{D})$. Due to the proof of Theorem 3.2 we know that

$$(1 + \mu(\mathbf{D}))\,|\alpha_l| \leq \mu(\mathbf{D})\|\alpha\|_1, \quad \forall l = 1, \ldots, n.$$

Summing over all $l \in \mathcal{S}$ we get

$$(1 + \mu(\mathbf{D})) \sum_{l \in \mathcal{S}} |\alpha_l| \leq \mu(\mathbf{D})\|\alpha\|_1 |\mathcal{S}|$$
$$(1 + \mu(\mathbf{D})) \frac{\sum_{l \in \mathcal{S}} |\alpha_l|}{\sum_l |\alpha_l|} \leq \mu(\mathbf{D})|\mathcal{S}|$$
$$(1 + \mu(\mathbf{D}))P_1(\mathcal{S}, \mathbf{D}) \leq \mu(\mathbf{D})|\mathcal{S}|$$
$$P_1(\mathcal{S}, \mathbf{D}) \leq \frac{1}{1 + 1/\mu(\mathbf{D})}|\mathcal{S}|.$$
$$\text{therefore } |\mathcal{S}| < \frac{1}{2}\left(1 + \frac{1}{\mu(\mathbf{D})}\right) \Rightarrow P_1(\mathcal{S}, \mathbf{D}) < 1/2 \quad \square$$

Theorem 3.8. Let $\mathbf{D} \in \mathbb{C}^{m \times n}$be a dictionary with coherence $\mu(\mathbf{D})$, then

$$\mu(\mathbf{D}) \geq \sqrt{\frac{n - m}{m(n - 1)}}, \text{ where } m \leq n$$

4 Finite Rate of Innovation

sequence consisting of K Dirac impulses with unknown locations and weights:

$$x(t) = \sum_{k=0}^{K-1} c_k \delta(t - t_k) \quad , \quad 0 \leq t_k \leq \tau$$

bandwidth of $x(t)$ is $\infty \Rightarrow$ according to classical sampling theorem we would have to sample at rate ∞ if we wanted to reconstruct the signal from its samples. But signal has only 2K parameters, namely $\{t_k, c_k\}_{k=0}^{K-1}$, therefore the signal should be recoverable in a finite number of measurements.
Lowpass measurements in form of fourier series coeffs:

$$d_n = \frac{1}{\tau}\int_0^\tau \sum_{k=0}^{K-1} c_k \delta(t - t_k)\, e^{-i2\pi n \frac{t}{\tau}} dt = \frac{1}{\tau}\sum_{k=0}^{K-1} c_k e^{-i2\pi n \frac{t_k}{\tau}}$$

Periodized version of x(t):

$$x(t) = \sum_{n \in \mathbb{Z}} \left(\frac{1}{\tau}\sum_{k=0}^{K-1} c_k e^{-i2\pi n \frac{t_k}{\tau}}\right) e^{i2\pi n \frac{t}{\tau}}$$

Algorithm for recovering x(t) from finite number of fourier coefficients (Berlekamp-Massy):

$$A(z) = \sum_{m=0}^{K} a_m z^{-m}, \text{ with K zeros at } u_k = e^{-i2\pi \frac{t_k}{\tau}}$$
$$\Rightarrow A(z) = \prod_{k=0}^{K-1}\left(1 - e^{-i2\pi \frac{t_k}{\tau}} z^{-1}\right)$$
$$\left(e^{-i2\pi \frac{t_k}{\tau} n} \star \left(\delta[n] - e^{-i2\pi \frac{t_k}{\tau}}\delta[n-1]\right)\right) = 0$$
$$\Rightarrow \text{since the fourier series coeffs } d_n \text{ are linear combinations}$$
$$\text{of exponentials } e^{-i2\pi \frac{t_k}{\tau} n}, \text{ we have: } (d_l)_{l \in \mathbb{Z}} \star (a_l)_{l \in \mathbb{Z}} = 0$$

In summary, for a given Fourier series coefficient sequence d_n, if we can find the corresponding annihilating filter impulse response a_n, the zeros of $A(z)$ yield the locations t_k through $A\left(e^{-i2\pi \frac{t_k}{\tau}}\right) = 0$, as $t_k = -\frac{\tau}{2\pi}\arg\left(e^{-i2\pi \frac{t_k}{\tau}}\right)$. Once we have the t_k, the corresponding weights c_k can be obtained by solving a linear system of equations given by $d_n = \frac{1}{\tau}\sum_{k=0}^{K-1} c_k u_k^n$.
In the following subsections we will describe all the steps involved in finding the annihilating filter.

4.1 Finding the a_k

$$\sum_{l=0}^{K} a_l d_{n-l} = 0, \quad \forall n \in \mathbb{Z} \Rightarrow \overbrace{\begin{bmatrix} d_0 & d_{-1} & \dots & d_{-K} \\ d_1 & d_0 & \dots & d_{-K+1} \\ \vdots & & & \vdots \\ d_K & d_{K-1} & \dots & d_0 \end{bmatrix}}^{\mathbf{S}} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_K \end{bmatrix} = 0 \quad d_n = \frac{1}{\tau}\sum_{k=0}^{K-1} c_k u_k^n \text{ which in matrix-vector form, reads:}$$

$$\frac{1}{\tau}\begin{bmatrix} 1 & 1 & \dots & 1 \\ u_0 & u_1 & \dots & u_{K-1} \\ \vdots & \vdots & & \vdots \\ u_0^{K-1} & u_1^{K-1} & \dots & u_{K-1}^{K-1} \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{K-1} \end{bmatrix} = \begin{bmatrix} d_0 \\ d_1 \\ \vdots \\ d_{K-1} \end{bmatrix}$$

To solve this linear system of equations, we need at least 2K + 1 Fourier series coefficients, namely $\{d_{-K}, \dots, d_0, \dots, d_K\}$. In practive, this linear system of equations is solved by identifying the singular vector of $\mathbf{S}$ corresponding to the smallest singular value.

4.1.1 Uniqueness

Rewriting S as a linear combination of rank-1 matrices:

$$\mathbf{S} = \frac{1}{\tau}\sum_{k=0}^{K-1} c_k \begin{bmatrix} 1 & u_k^{-1} & \dots & u_k^{-K} \\ u_k & 1 & \dots & u_k^{-K+1} \\ u_k^2 & u_k & \dots & u_k^{-K+2} \\ \vdots & \vdots & & \vdots \\ u_k^K & u_k^{K-1} & \dots & 1 \end{bmatrix}$$

$$= \frac{1}{\tau}\sum_{k=0}^{K-1} c_k \begin{bmatrix} 1 \\ u_k \\ u_k^2 \\ \vdots \\ u_k^K \end{bmatrix} \begin{bmatrix} 1 & u_k^{-1} & u_k^{-2} & \dots & u_k^{-K} \end{bmatrix}$$

The individual rank-one matrices are linearly independent as the Vandermonde matrices

$$\begin{bmatrix} 1 & 1 & \dots & 1 \\ u_0 & u_1 & \dots & u_K^1 \\ u_0^2 & u_1^2 & \dots & u_K^2 \\ \vdots & \vdots & & \vdots \\ u_0^K & u_1^K & \dots & u_K^K \end{bmatrix}, \quad \begin{bmatrix} 1 & 1 & \dots & 1 \\ u_0^{-1} & u_1^{-1} & \dots & u_K^{-1} \\ u_0^{-2} & u_1^{-2} & \dots & u_K^{-2} \\ \vdots & \vdots & & \vdots \\ u_0^{-K} & u_1^{-K} & \dots & u_K^{-K} \end{bmatrix}$$

are full-rank, provided all the u_k are different. Therefore, provided all $c_k \neq 0$, the $(K+1)\times(K+1)$ matrix $\mathbf{S}$ is of rank K and hence the system $\mathbf{Sa}$ has a unique solution.

4.2 Finding the Zeros

Once $a_0, \dots, a_K$ are found, we write:

$$A(z) = \sum_{m=0}^{K} a_m z^{-m} = \prod_{k=0}^{K-1} (1 - \alpha_k z^{-1})$$

and identify the zeros α_k which yielda $u_k = e^{-i2\pi\frac{t_k}{\tau}}$

4.3 Finding the c_k

To determine the weights c_k, it suffices to take K equations among

and has a unique solution when $u_k \neq u_l, \forall k \neq l$ (since the system matrix is a Vandermonde matrix). We hence have a method that retrieves the 2K unknowns $\{t_k, c_k\}$ from $\geq 2K+1$ Fourier series coefficients.

5 Sampling of Multi-Band Signals

5.1 Sampling Spectrally Sparse Signals

A sparse signal $x(t)$ can be perfectly reconstructed even though we undersample it. The minimum sampling rate required in order for exact recovery to be possible equals the support set size of the nonzero spectral components.

Theorem 5.1. (Landau) Consider a signal with spectral occupancy $\mathcal{I} \subset [-f_0, f_0]$ and assume the sampling set $\mathcal{P} = \{t_n\}$ (i.e. we are given the signal values $\{x(t_n)\}$). Then to reconstruct stably, we need

$$\mathcal{D}^-(\mathcal{P}) = \lim_{r\to\infty} \inf_{t\in\mathbb{R}} \frac{|\mathcal{P} \cap [t, t+r]|}{r} \geq |\mathcal{I}|$$

where $\mathcal{D}^-(\mathcal{P})$ denotes the lower Beurling density.

5.1.1 Stable Sampling

Def. 5.2. A set of points $\mathcal{P} = \{t_n\}$ is called a stable sampling set if $\forall x_1, x_2 \in \mathcal{H}$ if

$$A\|x_1 - x_2\|_{\mathcal{H}}^2 \leq \|x_1(\mathcal{P}) - x_2(\mathcal{P})\|_2^2 \leq B\|x_1 - x_2\|_{\mathcal{H}}^2$$

for some $A > 0$ and $B < \infty$. If $\mathcal{H}$ is a vector space we have $x_1 - x_2 \in \mathcal{H} \Rightarrow \forall x \in \mathcal{H} \quad A\|x\|_{\mathcal{H}}^2 \leq \|\mathbb{T}x\|_2^2 \leq B\|x\|_{\mathcal{H}}^2$ where $\mathbb{T} : x(t) \to \{x(\mathcal{P})\}$ denotes the sampling operator.
We go back to sampling of multi-band signals and consider

$$\mathcal{B}(\mathcal{I}) \triangleq \{x(t) \in \mathcal{L}^2(\mathbb{R}) : \hat{x}(f) = 0, \forall f \notin \mathcal{I}\}$$

The space $\mathcal{B}(\mathcal{I})$ of signals with Fourier transform supported on a given interval $\mathcal{I}$ is a vector space, since every linear combination of signals in $\mathcal{B}(\mathcal{I})$ is also in $\mathcal{B}(\mathcal{I})$.

5.2 Multicoset Sampling

We partition the overall spectral support region into L cells $\mathcal{F}_i$ of equal length $\frac{f_0}{L}$ i.e.

$$\mathcal{F}_i = \left[i\frac{f_0}{L}, (i+1)\frac{f_0}{L}\right), \quad \text{for } i \in \{0, \dots, L-1\}$$

For $L \to \infty$ this setup becomes the general setup considered previously. For L finite, we approximate $\mathcal{I}$ by s intervals of length $\frac{f_0}{L}$ i.e. $|\mathcal{I}| \approx \frac{sf_0}{L}$.
The signal $x(t)$ is sampled on a periodic nonuniform grid

$$\Psi = \Psi_1 \cup \dots \cup \Psi_K \text{ with } \Psi_k = \{(mL+k)T : m \in \mathbb{Z}\}, \quad \text{for } k = 1, \dots, K$$

So for every subgrid Ψ_k the sampling rate is $\frac{1}{LT} = \frac{f_0}{L}$
The samples corresponding to Ψ_k are

$$x_k[m] \triangleq x((mL+k)T), m \in \mathbb{Z}$$

And the overall sampling rate is $\mathcal{D}^-(\mathcal{P}) = \frac{K}{LT} = \frac{K}{L}f_0$// $\forall$ coset $\{x_k[m]\}_{m\in\mathbb{Z}}$ we compute the DFT

$$\begin{aligned} x_d^{(k)}(f) &= \sum_{m\in\mathbb{Z}} x_k[m] e^{-i2\pi f m TL} \\ &= \sum_{m\in\mathbb{Z}} x((mL+k)T) e^{-i2\pi f m TL} \\ &= e^{i2\pi fkT} \sum_{m\in\mathbb{Z}} x(mTL + \underbrace{kT}_{t}) e^{-i2\pi f(mTL+kT)} \\ &= e^{i2\pi fkT} \frac{1}{TL} \sum_{m\in\mathbb{Z}} \widehat{x}\left(f + \frac{m}{TL}\right) e^{i2\pi\frac{mk}{L}}, \quad f \in [0, 1), \end{aligned}$$

where the last equality follows from the Poisson summation formula

$$\sum_{l\in\mathbb{Z}} s(t + lT) = \frac{1}{T}\sum_{l\in\mathbb{Z}} \widehat{s}\left(\frac{l}{T}\right) e^{i2\pi\frac{l}{T}t}$$

Let

$$\begin{aligned} v_k(f) &:= x_d^{(k)}(f) e^{-i2\pi fkT} TL \\ &= \sum_{m\in\mathbb{Z}} \widehat{x}\left(f + \frac{m}{TL}\right) e^{i2\pi\frac{mk}{L}}, \quad f \in [0, 1/(LT)) \end{aligned}$$

and write (with $f \in [0, 1/(LT))$)

$$\underbrace{\begin{bmatrix} v_1(f) \\ v_2(f) \\ \vdots \\ v_K(f) \end{bmatrix}}_{=:\mathbf{v}(f)\in\mathbb{C}^K} = \underbrace{\begin{bmatrix} 1 & e^{i2\pi\frac{1}{L}} & \dots & e^{i2\pi\frac{L-1}{L}} \\ 1 & e^{i2\pi\frac{2}{L}} & \dots & e^{i2\pi\frac{2(L-1)}{L}} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{i2\pi\frac{K}{L}} & \dots & e^{i2\pi\frac{K(L-1)}{L}} \end{bmatrix}}_{=:\mathbf{A}\in\mathbb{C}^{K\times L}} \underbrace{\begin{bmatrix} \widehat{x}(f) \\ \widehat{x}\left(f + \frac{1}{TL}\right) \\ \vdots \\ \widehat{x}\left(f + \frac{L-1}{TL}\right) \end{bmatrix}}_{=:\mathbf{x}(f)\in\mathbb{C}^L}$$

It holds that $K \leq L$. If $K = s$ (which is ideal), then sampling is performed at Landau rate.
$\forall$ set $\mathcal{S} = \{s_0, s_1, \dots, s_{K-1}\} \subset \{1, \dots, L\}$ with $|\mathcal{S}| = K$,

let $\mathbf{A}_{\mathcal{S}}$ be the submatrix which contains the columns of $\mathbf{A}$ indexed by $\mathcal{S}$. Let $z_i \triangleq a_{s_i}$ for $i \in \{0, \dots, K-1\}$ then

$$\mathbf{A}_{\mathcal{S}}^T = \underbrace{\operatorname{diag}\left((z_0, z_1, \dots, z_{K-1})\right)}_{\text{full-rank}} \underbrace{\begin{bmatrix} 1 & z_0 & z_0^2 & \dots & z_0^{K-1} \\ 1 & z_1 & z_1^2 & \dots & z_1^{K-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & z_{K-1} & z_{K-1}^2 & \dots & z_{K-1}^{K-1} \end{bmatrix}}_{\mathbf{V}(z_0, z_1, \dots, z_{K-1})}$$

Hence every set of $s \leq K$ columns of $\mathbf{A}$ is linearly indep. If $\hat{\mathbf{x}}(f)$ is sparse and we know the spectral support $\mathcal{I}$ of the original signal $x(t)$ and hence the support set of $\hat{\mathbf{x}}(f)$, we can recover $\hat{\mathbf{x}}(f)$ according to

$$\mathbf{v}(f) = \mathbf{A}\hat{\mathbf{x}}(f) = \mathbf{A}_\gamma \hat{\mathbf{x}}_\gamma(f) \Rightarrow \hat{\mathbf{x}}_\gamma(f) = \mathbf{A}_\gamma^\dagger \mathbf{v}(f)$$

where γ is set of indices corresponding to nonzero entries of $\hat{\mathbf{x}}(f)$. Given support set of $\hat{\mathbf{x}}(f)$ the minimum K we need is s. This corresponds to $\mathcal{D}^-(\mathcal{P}) = \frac{K}{LT} \geq \frac{s}{LT} \approx |\mathcal{I}|$. Multicoset sampling therefore allows recovery from samples taken at the Landau rate and is universal in the sense that it is applicable irrespective of the spectral occupancy $\mathcal{I}$ provided that number of occupied cells $\mathcal{F}_i$ is at most s.

5.3 Spectrum-Blind Sampling

Now we consider the case where the support set γ is not known a priori, but we know that $|\gamma| \leq s$. This amounts to considering the set $\mathcal{X}(C) = \bigcup_{|\mathcal{I}| \leq C} \mathcal{B}(\mathcal{I})$.
$\forall x_1, x_2 \in \mathcal{X}(C)$ $x_1 - x_2 \notin \mathcal{X}(C)$ in general but $x_1 - x_2 \in \mathcal{X}(2C)$. Therefore from def. of stable sampling: $A\|x\|^2_{\mathcal{X}(2C)} \leq \|\mathbb{T}x\|_2^2 \leq B\|x\|^2_{\mathcal{X}(2C)}, \quad \forall x \in \mathcal{X}(2C)$.
To satisfy $\|\mathbb{T}x\|_2^2 \geq A\|x\|^2_{\mathcal{X}(2C)}, \quad \forall x \in \mathcal{X}(2C)$ we have the following necessary condition $\mathcal{D}^-(\mathcal{P}) \geq 2C$.
This condition is also sufficient. This can be shown in the following way: If $\|\hat{\mathbf{x}}_1(f) - \hat{\mathbf{x}}_2(f)\|_0 \leq 2s \leq K, \forall f \in \left[0, \frac{1}{LT}\right]$, the Vandermonde structure implies that $\forall x_1 - x_2$ multicoset sampling is stable since $\|\hat{\mathbf{x}}(f)\|_0 = s \leq \frac{K}{2}$ implies that $|\mathcal{I}| = \frac{s}{LT} \leq \frac{1}{LT}\frac{K}{2} = \frac{1}{2}\frac{K}{LT} = \frac{\mathcal{D}^-(\mathcal{P})}{2}$. Since $|\mathcal{I}| \leq C$ holds $\forall x \in \mathcal{X}(C)$, $\mathcal{D}^-(\mathcal{P}) \geq 2C$ is also sufficient for stable sampling.

6 The ESPRIT Algorithm

The problem we consider is as follows. Recover the complex numbers $z_1, z_2, \dots, z_K$ with $\leq 1, k \in \{1, 2, \dots, K\}$ (henceforth referred to as "nodes") and the corresponding complex weights $\alpha_1, \alpha_2, \dots, \alpha_K$ from the measurements

$$x_n = \sum_{k=1}^{K} \alpha_k z_k^n, \quad n \in \{0, 1, \dots, N-1\}$$

where the number of samples N, satisfies $N \geq 2K$. The nodes can be written as $z_k = e^{-d_k} e^{2\pi i \xi_k}, k \in \{1, 2, \dots, K\}$, where $d_k \geq 0$ is referred to as the damping factor and ξ_k as the normalized frequency of the k-th cisoid. In the following chapter we assume $z_1, z_2, \dots, z_K$ and $\alpha_1, \alpha_2, \dots, \alpha_K$ to be nonzero and $z_1, z_2, \dots, z_K$ to be pairwise distinct.

6.1 Signal and Noise Subspaces

We start by describing the subspace ESPRIT relies on. To this end, we first construct the data matrix $\mathbf{X} \triangleq \mathscr{H}_L(x_0, x_1, \dots, x_{N-1}) \in \mathbb{C}^{L \times (N-L+1)}$ from the signal samples $\{\sigma_n\}_{n=0}^{N-1}$ given by the measurements x_n. Note that $\mathbf{X}$ can be factorized according to

$$\mathbf{X} = \mathbf{V}_L \mathbf{D}_{\boldsymbol{\alpha}} \mathbf{V}_{N-L+1}^T \in \mathbb{C}^{L \times (N-L+1)}$$
$$\text{where } \mathbf{V}_L \triangleq \mathscr{V}_{L \times K}(z_1, z_2, \dots, z_K) \in \mathbb{C}^{L \times K}$$
$$\mathbf{D}_{\boldsymbol{\alpha}} \triangleq \operatorname{diag}(\alpha_1, \alpha_2, \dots, \alpha_K)$$
$$\mathbf{V}_{N-L+1} \triangleq \mathscr{V}_{(N-L+1) \times K}(z_1, z_2, \dots, z_K) \in \mathbb{C}^{(N-L+1) \times K}$$

Here, $L \in \mathbb{N}$ is a parameter, satifying $K+1 \leq L \leq N-K-1$ that controls the aspect ration of the data matrix $\mathbf{X}$. Since the nodes are nonzero and pairwise distinct, and $K+1 \leq L \leq N-K-1$, the Vandermonde matrices V_L and V_{N-L+1} both have full rank K. As the weights $\alpha_1, \alpha_2, \dots, \alpha_K$ are nonzero, $\mathbf{D}_\alpha$ is invertible, and hence $\mathbf{X}$ has rank K $\Rightarrow$ $\mathbf{X}$ has K nonzero singular values $\sigma_1, \sigma_2, \dots, \sigma_K$ and can be decomposed according to

$$\mathbf{X} \triangleq \begin{pmatrix} \mathbf{S} & \mathbf{S}_\perp \end{pmatrix} \begin{pmatrix} \boldsymbol{\Lambda} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{R}^H \\ \mathbf{R}_\perp^H \end{pmatrix} = \mathbf{S}\boldsymbol{\Lambda}\mathbf{R}^H$$

$\begin{pmatrix} \mathbf{S} & \mathbf{S}_\perp \end{pmatrix}$ and $\begin{pmatrix} \mathbf{R}^H \\ \mathbf{R}_\perp^H \end{pmatrix}$ unitary, $\boldsymbol{\Lambda} \triangleq \operatorname{diag}(\sigma_1, \sigma_2, \dots, \sigma_K)$
Let $\mathcal{S} \triangleq$ "column span of $\mathbf{S}$". Due to the decomposition above and the fact that $\mathbf{X}$ and $\mathbf{S}$ both have rank K, we conclude that $\mathcal{S}$ coincides with the column span of $\mathbf{X}$.

6.2 Review of Similarity Principle

Def. 6.1. The matrices $\mathbf{X} \in \mathbb{C}^{n \times n}$ and $\mathbf{Y} \in \mathbb{C}^{n \times n}$ are similar if $\exists$ an invertible $n \times n$ matrix $\mathbf{P}$: $\mathbf{X} = \mathbf{P}^{-1}\mathbf{Y}\mathbf{P}$
Theorem 6.2. Let $\mathbf{A}$ and $\mathbf{B}$ be similar matrices. Then, $\mathbf{A}$ and $\mathbf{B}$ have the same eigenvalues with the same geometric multiplicities.
Proof. $\mathbf{A} = \mathbf{P}^{-1}\mathbf{B}\mathbf{P} \implies \mathbf{B} = \mathbf{P}\mathbf{A}\mathbf{P}^{-1}$ If $\mathbf{A}\mathbf{u} = \lambda\mathbf{u}$ then $\mathbf{P}^{-1}\mathbf{B}\mathbf{P} = \lambda\mathbf{u}$, thus $\mathbf{B}\mathbf{P}\mathbf{u} = \lambda\mathbf{P}\mathbf{u}$. (and the same procedure applies for $\mathbf{B}\mathbf{u} = \lambda\mathbf{u} \Rightarrow \mathbf{A}\mathbf{P}^{-1}\mathbf{u} = \lambda\mathbf{P}^{-1}\mathbf{u}$)

6.3 The ESPRIT Algorithm

Let $\mathbf{V}_\downarrow \in \mathbb{C}^{(L-1) \times K}$ be the matrix consisting of the top $L-1$ rows of $\mathbf{V}_L$ and $\mathbf{V}_\uparrow \in \mathbb{C}^{(L-1) \times K}$ the matrix consisting of the bottom $L-1$ rows of $\mathbf{V}_L$. We have

$$\mathbf{V}_\uparrow = \mathbf{V}_\downarrow \mathbf{D}_{\boldsymbol{z}}, \quad \text{where} \quad \mathbf{D}_{\boldsymbol{z}} \triangleq \operatorname{diag}(z_1, z_2, \dots, z_K)$$

Since the columns of both $\mathbf{S}$ and $\mathbf{V}_L$ form bases for the signal subspace $\mathcal{S}$, there exists an invertible matrix $\mathbf{P} \in \mathbb{C}^{K \times K}$ s.t. $\mathbf{S} = \mathbf{V}_L\mathbf{P}$. From $\mathbf{V}_\uparrow = \mathbf{V}_\downarrow \mathbf{D}_z$ follows

$$\mathbf{S}_\uparrow = \mathbf{S}_\downarrow \boldsymbol{\Phi} \text{ where } \Phi \triangleq \mathbf{P}^{-1}\mathbf{D}_z\mathbf{P}$$

As $\mathbf{D}_{\boldsymbol{z}} = \operatorname{diag}(z_1, z_2, \dots, z_K)$ and $\mathbf{P} \in \mathbb{C}^{K \times K}$ is invertible, $z_1, z_2, \dots, z_K$ are the eigenvalues of $\boldsymbol{\Phi} = \mathbf{S}_\downarrow^\dagger \mathbf{S}_\uparrow$. This is a direct consequence of the similarity principle. Moreover since $\mathbf{S}_\downarrow$ has full rank, $\boldsymbol{\Phi} = \mathbf{S}_\downarrow^\dagger \mathbf{S}_\uparrow$ is the unique solution of $\mathbf{S}_\downarrow \mathbf{Y} = \mathbf{S}_\uparrow$.

6.4 Finding the Zeros of a Polynomial

Consider polynomial $p(z) = \alpha_0 + \alpha_1 z + \dots + \alpha_K z^K$. For simplicity of exposition we assume that all zeros of $p(z)$ have multiplicity one. First we write

$$p(z) = \begin{pmatrix} \alpha_0 \alpha_1 & \dots & \alpha_K \end{pmatrix} \begin{pmatrix} 1 \\ z \\ \vdots \\ z^K \end{pmatrix}$$

and note that for the zeros $z_0, \dots, z_{K-1}$ we have

$$\underbrace{(\alpha_0 \alpha_1 \dots \alpha_K)}_{=:\boldsymbol{\alpha} \in \mathbb{C}^{1 \times (K+1)}} \underbrace{\begin{pmatrix} 1 & 1 & \dots & 1 \\ z_0 & z_1 & \dots & z_{K-1} \\ \vdots & \vdots & & \vdots \\ z_0^K & z_1^K & \dots & z_{K-1}^K \end{pmatrix}}_{=:\mathbf{V} \in \mathbb{C}^{(K+1) \times K}} = \mathbf{0}^T$$

where $\mathbf{V}$ is a vandermonde matrix with nodes given by the zeros of $p(z)$. As the zeros z_i are all of multiplicity one, by assumption it follows that the Vandermonde matrix $\mathbf{V}$ is of rank K (i.e. full rank).
Next, we note that $\mathbf{V}$ is a basis for the right null-space of α. This null-space has dimension K (the ambient space has dimension $K+1$). We determine a basis for this null-space according to

$$\boldsymbol{\alpha} = \underbrace{\begin{bmatrix} u_1 & \mathbf{u}_2 \end{bmatrix}}_{1 \times (K+1)} \underbrace{\begin{bmatrix} \sigma & \mathbf{0}_{1 \times K} \\ \mathbf{0}_{K \times 1} & \mathbf{0}_{K \times K} \end{bmatrix}}_{(K+1) \times (K+1)} \underbrace{\begin{bmatrix} \mathbf{w}^T \\ \mathbf{Z}^T \end{bmatrix}}_{(K+1) \times (K+1)}$$

where $\mathbf{Z}$ is a $(K+1)\times K$ matrix containing the basis for the null-space. As both $\mathbf{V}$ and $\mathbf{Z}$ are bases for the right null-space of α, they must be related through a full rank matrix, denoted as $\mathbf{T}$: $\mathbf{V}=\mathbf{ZT}$. And since $\mathbf{V}_\uparrow=\mathbf{V}_\downarrow\mathbf{D_z}$

$$\left.\begin{array}{l}\mathbf{V}_\uparrow=\mathbf{Z}_\uparrow\mathbf{T}\\ \mathbf{V}_\downarrow=\mathbf{Z}_\downarrow\mathbf{T}\end{array}\right\}\Rightarrow\mathbf{Z}_\uparrow\mathbf{T}=\mathbf{Z}_\downarrow\mathbf{T}D_\mathbf{z}$$
$$\Longrightarrow\mathbf{Z}_\uparrow=\mathbf{Z}_\downarrow\underbrace{\mathbf{T}\mathbf{D_z}\mathbf{T}^{-1}}_{\mathbf{\Phi}}$$

Just like in the ESPRIT algorithm, by similarity, the zeros z_i are given by the eigenvalues of $\mathbf{\Phi}=\mathbf{Z}_\downarrow^\dagger\mathbf{Z}_\uparrow$.

7 RIP and JL Lemma

7.1 Restricted Isometry Property

Assume that we observe $\mathbf{y}=\mathbf{\Phi x}\in\mathbb{C}^m$ where $\mathbf{x}\in\mathbb{C}^n$ is a signal unknown to us that we want to reconstruct, and $\mathbf{\Phi}\in\mathbb{C}^{m\times n}$ is a known measurement matrix. Here we considere the case $m<n$.

Def. 7.1. For $s=1,2,\ldots$, the isometry constant δ_s of a matrix $\mathbf{\Phi}$ is defined as the smallest number s.t.

$$(1-\delta_s)\left\|\mathbf{x}\right\|_2^2\leq\left\|\mathbf{\Phi x}\right\|_2^2\leq(1+\delta_s)\left\|\mathbf{x}\right\|_2^2$$

holds for all s-sparse vectors $\mathbf{x}$.

Theorem 7.2. Let $\mathbf{y}=\mathbf{\Phi x}$. Assume $\delta_{2s}<\sqrt{2}-1$. Then the solution $\mathbf{x}^*$ to

$$\underset{\widetilde{\mathbf{x}}\in\mathbb{R}^n}{\text{minimize}}\|\widetilde{\mathbf{x}}\|_1\quad\text{subject to }\mathbf{\Phi}\widetilde{\mathbf{x}}=\mathbf{y}$$

obeys $\left\|\mathbf{x}^*-\mathbf{x}\right\|_1\leq C_0\left\|\mathbf{x}-\mathbf{x}_s\right\|_1$ and $\left\|\mathbf{x}^*-\mathbf{x}\right\|_2\leq C_0 s^{-1/2}\left\|\mathbf{x}-\mathbf{x}_s\right\|_1$ for some constant C_0 specified in the proof. Here x_s is the vector obtained by setting all but the s largest entries of $\mathbf{x}$ equal to zero. In particular, if $\mathbf{x}$ is s-sparse, recovery is exact.

Theorem 7.3. Let $\mathbf{y}=\mathbf{\Phi x}+\mathbf{n}$. Assume $\delta_{2s}<\sqrt{2}-1$ and $\left\|\mathbf{n}\right\|_2\leq\varepsilon$. Then, the solution $\mathbf{x}^*$ to

$$\underset{\widetilde{\mathbf{x}}\in\mathbb{R}^n}{\text{minimize}}\|\widetilde{\mathbf{x}}\|_1\quad\text{subject to }\left\|\mathbf{y}-\mathbf{\Phi}\widetilde{\mathbf{x}}\right\|_2\leq\varepsilon$$

$$\text{obeys }\left\|\mathbf{x}^*-\mathbf{x}\right\|_2\leq C_0 s^{-1/2}\left\|\mathbf{x}-\mathbf{x}_s\right\|_1+C_1\varepsilon$$

with the same C_0 constant as before and some constant C_1 given explicitly in the proof.

Note that the 2nd bound in theorem 7.2 follows directly from theorem 7.3 by setting $\mathbf{n}=0$ and therefore $\varepsilon=0$.

7.2 Johnson-Lindenstrauss Lemma

Suppose we are given a set of $\mathcal{U}$ of m points in $\mathbb{R}^n$. We would like to embed these points into a lower dimensional euclidean space (i.e $\mathbb{R}^k$ with $k<n$) while approximately preserving the distances between the points in $\mathcal{U}$. The JL Lemma, shows that any set of m points can be embedded in $k=\mathcal{O}\left(\log(m)/\epsilon^2\right)$ dimensions while the distances between any two points change by at most a factor of $1\pm\epsilon$.

Lemma 8.1. (JL Lemma) Choose ϵ with $0<\epsilon<1$ and sps. k satisfies $k\geq\frac{8}{\epsilon^2-\epsilon^3}\log(2m)$. Then, for every set $\mathcal{U}$ of m points, $\exists$ a map $f:\mathbb{R}^n\to\mathbb{R}^k$ s.t. $\forall\mathbf{u},\mathbf{u}'\in\mathcal{U}$, we have

$$(1-\epsilon)\left\|\mathbf{u}-\mathbf{u}'\right\|^2\leq\left\|f(\mathbf{u})-f\left(\mathbf{u}'\right)\right\|^2\leq(1+\epsilon)\left\|\mathbf{u}-\mathbf{u}'\right\|^2$$

Lemma 8.2. Let $\mathbf{A}\in\mathbb{R}^{k\times n}$ be a random matrix with i.i.d. $\mathcal{N}(0,1/k)$. Then, for $\epsilon\in(0,1)$ and fixed $\mathbf{u}\in\mathbb{R}^n$.

$$\mathbb{P}\left(|\ \|\mathbf{Au}\|^2-\mathbb{E}\left[\|\mathbf{Au}\|^2\right]\ |\geq\epsilon\|\mathbf{u}\|^2\right)<2e^{-k\frac{\epsilon^2-\epsilon^3}{4}}$$
$$\text{with }\mathbb{E}\left[\|\mathbf{Au}\|^2\right]=\|\mathbf{u}\|^2$$

In words, Lemma 8.2 states that the r.v. $\|\mathbf{Au}\|^2$ is concentrated around its mean. Relations of this form are called "concentration of measure inequality". Note that Lemma 8.2. is not restricted to gaussian random matrices, but generalizes to other random matrices $\mathbf{A}$ where each $a_{i,j}$ is sub-Gaussian (i.e. its tail probability satisfies $\mathbb{P}\left(|a_{i,j}|>t\right)\leq c_1e^{-c_2t^2}$).

Proof that Lemma 8.2⇒JL Lemma:

The proof is done by showing lin. map $f(\mathbf{u})=\mathbf{Au}$ with $\mathbf{A}\in\mathbb{R}^{k\times n}$ rand. mat with i.i.d. $\mathcal{N}(0,1/k)$ entries, satisfies JL $\forall\mathbf{u},\mathbf{u}'\in\mathcal{U}$ with non-zero prob. Applying union bound over all $m(m-1)/2<m^2$ pairs of points in $\mathcal{U}$, it follows from Lemma 8.2 that JL violated for one or more pairs of point with prob. $<m^2 2e^{-k\frac{\epsilon^2-\epsilon^3}{4}}$.

$$m^2 2e^{-k\frac{\epsilon^2-\epsilon^3}{4}}\leq 1/2\Leftrightarrow -k\frac{\epsilon^2-\epsilon^3}{4}\leq\log\left(1/\left(4m^2\right)\right)$$
$$\Leftrightarrow k\geq\frac{4}{\epsilon^2-\epsilon^3}2\log(2m).$$

Proof of Lemma 8.2

$$\mathbb{E}\left[\|\mathbf{Au}\|^2\right]=\mathbb{E}\left[\mathbf{u}^T\mathbf{A}^T\mathbf{Au}\right]=\mathbf{u}^T\mathbb{E}\left[\mathbf{A}^T\mathbf{A}\right]\mathbf{u}=\mathbf{u}^T\mathbf{Iu}=\|\mathbf{u}\|^2$$

next let $\mathbf{a}_j^T$ be the j-th row of $\mathbf{A}$ and $X_j:=\frac{\sqrt{k}}{\|\mathbf{u}\|}\mathbf{a}_j^T\mathbf{u}$.

$\mathbf{a}_j^T\mathbf{u}\sim\mathcal{N}\left(0,\|\mathbf{u}\|^2/k\right)\Rightarrow X_j\sim\mathcal{N}(0,1)$.

$$\sum_{j=1}^k X_j^2=\frac{k}{\|\mathbf{u}\|^2}\sum_{j=1}^k\left|\mathbf{a}_j^T\mathbf{u}\right|^2=\frac{k}{\|\mathbf{u}\|^2}\|\mathbf{Au}\|^2$$

thus for $\lambda\geq 0$

$$\mathbb{P}\left(\|\mathbf{Au}\|^2\geq(1+\epsilon)\|\mathbf{u}\|^2\right)=\mathbb{P}(X\geq(1+\epsilon)k)=\mathbb{P}\left(e^{\lambda X}\geq e^{\lambda(1+\epsilon)k}\right)$$
$$\overbrace{\leq}^{\text{Markov inequality}}\frac{1}{e^{(1+\epsilon)k\lambda}}\mathbb{E}\left[e^{\lambda X}\right]=\frac{1}{e^{(1+\epsilon)k\lambda}}\prod_{j=1}^k\mathbb{E}\left[e^{\lambda X_j^2}\right]$$
$$=\frac{1}{e^{(1+\epsilon)k\lambda}}\left(\mathbb{E}\left[e^{\lambda X_1^2}\right]\right)^k$$

moment generating function $\mathbb{E}\left[e^{\lambda X_1^2}\right]$:

$$\mathbb{E}\left[e^{\lambda X_1^2}\right]=\int_{-\infty}^{\infty}e^{\lambda x^2}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx$$
$$=\frac{1}{\sqrt{1-2\lambda}}\int_{-\infty}^{\infty}\frac{\sqrt{1-2\lambda}}{\sqrt{2\pi}}e^{-\frac{x^2}{2}(1-2\lambda)}dx=\frac{1}{\sqrt{1-2\lambda}},$$

Putting all together yields

$$\mathbb{P}\left(\|\mathbf{Au}\|^2\geq(1+\epsilon)\|\mathbf{u}\|^2\right)\leq\left(\frac{e^{-2(1+\epsilon)\lambda}}{1-2\lambda}\right)^{\frac{k}{2}}$$

the λ that minimizes the right-hand side is $\lambda=\frac{\epsilon}{2(1+\epsilon)}$, which gives:

$$\mathbb{P}\left(\|\mathbf{Au}\|^2\geq(1+\epsilon)\|\mathbf{u}\|^2\right)\leq\left((1+\epsilon)e^{-\epsilon}\right)^{\frac{k}{2}}\overbrace{<}^{1+\epsilon<e^{\epsilon-\frac{\epsilon^2-\epsilon^3}{2}}}e^{-\left(\epsilon^2-\epsilon^3\right)\frac{k}{4}}$$

Similarly we obtain

$$\mathbb{P}\left(\|\mathbf{Au}\|^2\leq(1-\epsilon)\|\mathbf{u}\|^2\right)<e^{-\left(\epsilon^2-\epsilon^3\right)\frac{k}{4}}$$

Combining these two last results yields

$$\mathbb{P}\left(|\ \|\mathbf{Au}\|^2-\mathbb{E}\left[\|\mathbf{Au}\|^2\right]\ |\geq\epsilon\|\mathbf{u}\|^2\right)<2e^{-k\frac{\epsilon^2-\epsilon^3}{4}}$$

7.3 Verifying the RIP through the JL Lemma

Lemma 9.1. Let $\mathbf{\Phi}\in\mathbb{R}^{m\times n}$ rand. mat. with i.i.d. $\mathcal{N}(0,1/m)$ entries. Then, for any set $\mathcal{S}$ with $|\mathcal{S}|=k<m$ and any $0<\delta<1$, we have

$$(1-\delta)\|\mathbf{x}\|\leq\|\mathbf{\Phi x}\|\leq(1+\delta)\|\mathbf{x}\|,\quad\forall\mathbf{x}\in\mathcal{X}_\mathcal{S}$$

with probability

$$\geq 1-2(12/\delta)^k e^{-c_0(\delta/2)m},$$

where $c_0(x)=\frac{1}{4}\left(x^2-x^3\right)$.

Theorem 9.2. Suppose m,n, and $0<\delta<1$ are given. If pdf generating $\lesseqgtr$ satisfies the concentration inequality in Lemma 8.2, then $\exists$ constants $c_1,c_2>0$ depending only on δ s.t. RIP holds for Φ with the prescribed δ for any $k\leq c_1m/\log(n/k)$ with prob. $\geq 1-2e^{-c_2m}$

8 Approximation Theory

8.1 Min-Max (Kolmogorov) Rate Distortion Theory

Let $d\in\mathbb{N},\Omega\subset\mathbb{R}^a$ and consider the function class $\mathcal{C}\subset L^2(\Omega)$. Then, for each $\ell\in\mathbb{N}$, we denote by

$$\mathfrak{E}^\ell:=\left\{E:\mathcal{C}\to\{0,1\}^\ell\right\}\text{ set of binary encoders of length }\ell$$

$$\mathfrak{D}^\ell:=\left\{D:\{0,1\}^\ell\to L^2(\Omega)\right\}\text{set of binary decoders of length }\ell$$

An encoder-decoder pair $(E,D)\in\mathfrak{E}^\ell\times\mathfrak{D}^\ell$ is said to achieve uniform error ε over the function class $\mathcal{C}$ if

$$\sup_{f\in\mathcal{C}}\|D(E(f))-f\|_{L^2(\Omega)}\leq\varepsilon$$

The minimax code length is

$$L(\varepsilon,\mathcal{C}) := \min\{\ell \in \mathbb{N} : \exists (E,D) \in \mathfrak{E}^\ell \times \mathfrak{D}^\ell : \sup_{f\in\mathcal{C}} \|D(E(f)) - f\|_{L^2(\Omega)} \leq \varepsilon\}$$

Morever, the optimal exponent is defined as:

$$\gamma^*(\mathcal{C}) := \sup\left\{\gamma \in \mathbb{R} : L(\varepsilon,\mathcal{C}) \in \mathcal{O}\left(\varepsilon^{-1/\gamma}\right), \varepsilon \to 0\right\}$$

8.2 Metric Entropy, Covering, and Packing

Def. 10.2.(covering) Let $(\mathcal{X},\rho)$ be a metric space. An ϵ-covering of a compact set $\mathcal{C} \subseteq \mathcal{X}$ with respect to the metric ρ is a set $\{x_1,\dots,x_N\} \subset \mathcal{C}$ s.t. for each $x \in \mathcal{C}$, $\exists i \in [1,N]$ s.t. $\rho(x,x_i) \leq \epsilon$. The ϵ-covering number $N(\epsilon;\mathcal{C},\rho)$ is the cardinality of the smallest ϵ-covering.

Def. 10.3.(packing) Let $(\mathcal{X},\rho)$ be a metric space. An ϵ-packing of a compact set $\mathcal{C} \subset \mathcal{X}$ w.r.t. metric ρ is a set $\{x_1,\dots,x_N\} \subset \mathcal{C}$ s.t. $\rho(x_i,x_j) > \epsilon\ \forall i \neq j$. The ϵ-packing number $M(\epsilon;\mathcal{X},\rho)$ is the cardinality of the largest ϵ-packing.

Lemma 10.4. Let $(\mathcal{X},\rho)$ be a metric space and $\mathcal{C}$ a compact set in $\mathcal{X}$. $\forall \epsilon > 0$, the packing and the covering number are related according to

$$M(2\epsilon;\mathcal{C},\rho) \leq N(\epsilon;\mathcal{C},\rho) \leq M(\epsilon;\mathcal{C},\rho)$$

Proof. First, choose a minimal ϵ-covering and a maximal 2ϵ-packing of $\mathcal{C}$. Since no two centers of the 2ϵ-packing can lie in the same ball of the ϵ-covering, it follows that $M(2\epsilon;\mathcal{C},\rho) \leq N(\epsilon;\mathcal{C},\rho)$. To establish $N(\epsilon;\mathcal{C},\rho) \leq M(\epsilon;\mathcal{C},\rho)$, we note that, given a maximum packing $M(\epsilon;\mathcal{C},\rho)$, for any x $\in \mathcal{C}$, we have the center of at least one of the balls in the packing within distance less than ϵ. If this were not the case, we could add another ball to the packing thereby violating its maximality. This maximal packing hence also provides an ϵ-covering and since $N(\epsilon;\mathcal{C},\rho)$ is a minimal covering, we must have $N(\epsilon;\mathcal{C},\rho) \leq M(\epsilon;\mathcal{C},\rho)$. □

Lemma 10.5. Consider a pair of norms $\|\cdot\|$ and $\|\cdot\|'$ on $\mathbb{R}^d$, and let $\mathcal{B}$ and $\mathcal{B}'$ be their corresponding unit balls, i.e., $\mathcal{B} = \left\{x \in \mathbb{R}^d \mid \|x\| \leq 1\right\}$ and $\mathcal{B}' = \left\{x \in \mathbb{R}^d \mid \|x\|' \leq 1\right\}$. Then the ϵ-covering number of $\mathcal{B}$ in the $\|\cdot\|'$-norm satisfies

$$\left(\frac{1}{\epsilon}\right)^d \frac{\mathrm{vol}(\mathcal{B})}{\mathrm{vol}(\mathcal{B}')} \leq N(\epsilon;\mathcal{B},\|\cdot\|') \leq \frac{\mathrm{vol}\left(\frac{2}{\epsilon}\mathcal{B}+\mathcal{B}'\right)}{\mathrm{vol}(\mathcal{B}')}$$

Proof. Let $\left\{x_1,\dots,x_{N(\epsilon;\mathcal{B},\|\cdot\|\|\cdot\|')}\right\}$ be an ϵ-covering of $\mathcal{B}$ in $\|\cdot\|'$-norm. Then we have $\mathcal{B} \subseteq \bigcup_{j=1}^{N(\epsilon;\mathcal{B},\|\cdot\|\|\cdot\|')} \{x_j + \epsilon\mathcal{B}'\}$, which implies $\mathrm{vol}(\mathcal{B}) \leq N(\epsilon;\mathcal{B},\|\cdot\|')\,\epsilon^d\,\mathrm{vol}(\mathcal{B}')$ thus establishing the lower bound. The upper bound is obtained by starting with maximal ϵ-packing $\left\{x_1,\dots,x_{M(\epsilon;\mathcal{B},\|\cdot\|\|\cdot\|')}\right\}$ of $\mathcal{B}$ in the $\|\cdot\|'$-norm. The balls $\left\{x_j + \frac{\epsilon}{2}\mathcal{B}', j = 1,\dots,M(\epsilon;\mathcal{B},\|\cdot\|')\right\}$ are all disjoint and contained within $\mathcal{B} + \frac{\epsilon}{2}\mathcal{B}'$. Taking volumes, we can therefore conclude that $\sum_{j=1}^{M(\epsilon;\mathcal{B},\|\cdot\|')} \mathrm{vol}\left(x_j + \frac{\epsilon}{2}\mathcal{B}'\right) \leq \mathrm{vol}\left(\mathcal{B} + \frac{\epsilon}{2}\mathcal{B}'\right)$ and hence $M(\epsilon;\mathcal{B},\|\cdot\|')\,\mathrm{vol}\left(\frac{\epsilon}{2}\mathcal{B}'\right) \leq \mathrm{vol}\left(\mathcal{B} + \frac{\epsilon}{2}\mathcal{B}'\right)$. Finally, we have $\mathrm{vol}\left(\frac{\epsilon}{2}\mathcal{B}'\right) = \left(\frac{\epsilon}{2}\right)^d \mathrm{vol}(\mathcal{B}')$ and $\mathrm{vol}\left(\mathcal{B} + \frac{\epsilon}{2}\mathcal{B}'\right) = \left(\frac{\epsilon}{2}\right)^d \mathrm{vol}\left(\frac{2}{\epsilon}\mathcal{B} + \mathcal{B}'\right)$, which together with $M(\epsilon;\mathcal{B},\|\cdot\|') \geq N(\epsilon;\mathcal{B},\|\cdot\|')$, yields the upper bound. □

So far we studied the metric entropy of various subsets of $\mathbb{R}^d$. We now turn to the metric entropy of function classes, begining with a simple one-parameter class. For a fixed θ, define $f_\theta(x) = 1 - e^{-\theta x}$ and consider the class $\mathcal{P} = \{f_\theta : [0,1] \to \mathbb{R} \mid \theta \in [0,1]\}$. The set $\mathcal{P}$ constitutes a metric space under the sup-norm given by $\|f-g\|_\infty = \sup_{x\in[0,1]} \mid f(x) - g(x) \mid$. We show (in lecure notes) that the covering number of $\mathcal{P}$ satisfies

$$1 + \left\lfloor \frac{1 - 1/e}{2\epsilon} \right\rfloor \leq N(\epsilon;\mathcal{P},\|\cdot\|_\infty) \leq \frac{1}{2\epsilon} + 2$$

which leads to the scaling behavior $N(\epsilon;\mathcal{P},\|\cdot\|_\infty) \asymp \epsilon^{-1}$, hence metric entropy scaling according to $\log_2(N(\epsilon;\mathcal{P},\|\cdot\|_\infty)) = \log_2\left(\epsilon^{-1}\right)$.

8.3 Approximation with Representation Systems

optimal approximation on hilbert spaces: Let $\mathcal{H}$ be a HS equipped with inner product and induced norm $\|\cdot\|_\mathcal{H}$, and let $e_k, k = 1,2,\dots$ ONB for $\mathcal{H}$. For linear approximation we use the linear space $\mathcal{H}_M := \mathrm{span}\{e_k : 1 \leq k \leq M\}$ to approximate a given element $f \in \mathcal{H}$. We measure the approximation error by $E_M(f)_\mathcal{H} := \inf_{g\in\mathcal{H}_M} \|f-g\|_\mathcal{H}$. In nonlinear approximation, we consider M-term approximation which replaces $\mathcal{H}_M$ by the space Σ_M consisting of all elements $g \in \mathcal{H}$ that can be expressed as $g = \sum_{k\in\Lambda} c_k e_k$ where $\Lambda \subset \mathbb{N}$ is a set of indices with $|\Lambda| \leq M$. Note that, in contrast to $\mathcal{H}_M$, the space Σ_M is not linear. A linear combination of two elements in Σ_M will, in general, need 2M terms in its representation by the e_k. Analogous to E_M, we define the error of M-term approximation $\sigma_M(f)_\mathcal{H} := \inf_{g\in\Sigma_M} \|f-g\|_\mathcal{H}$. We now proceed to M-term approximation in general dictionaries, i.e., we replace the orthonormal basis $\{e_k\}$ by a general, possibly redundant set of functions.

Def 10.6. Given $d \in \mathbb{N}, \Omega \subset \mathbb{R}^d$, a function class $\mathcal{C} \subset L^2(\Omega)$ and a representation system $\mathcal{D} = (\varphi_i)_{i\in I} \subset L^2(\Omega)$, we define, for $f \in \mathcal{C}$ and $M \in \mathbb{N}$

$$\Gamma_M^\mathcal{D}(f) := \inf_{I_M\subseteq I, \#I_M = M, (c_i)_{i\in I_M}} \left\| f - \sum_{i\in I_M} c_i\varphi_i \right\|_{L^2(\Omega)}$$

We call $\Gamma_M^\mathcal{D}(f)$ the best M-term approximation error of f in $\mathcal{D}$. Every $f_M = \sum_{i\in I_M} c_i\varphi_i$ attaining the infimum is referred to as a best M-term approximation of f in $\mathcal{D}$. The supremal $\gamma > 0$ s.t.

$$\sup_{f\in\mathcal{C}} \Gamma_M^\mathcal{D}(f) \in \mathcal{O}\left(M^{-\gamma}\right), M \to \infty$$

will be denoted by $\gamma^*(\mathcal{C},\mathcal{D})$. We say that the best M-term approximation rate of $\mathcal{C}$ in the representation system $\mathcal{D}$ is $\gamma^*(\mathcal{C},\mathcal{D})$

Def. 10.7. Given $d \in \mathbb{N}$, $\Omega \subset \mathbb{R}^d$, a function class $\mathcal{C} \subset L^2(\Omega)$, and a representation system $\mathcal{D} = (\varphi_i)_{i\in I} \subset L^2(\Omega)$, the supremal $\gamma > 0$ so that there exists a polynomial π

$$\sup_{f\in\mathcal{C}} \inf_{\substack{I_M\subset\{1,2,\dots,\pi(M)\} \\ \#I_M = M, (c_i)_{i\in I_M}}} \left\| f - \sum_{i\in I_M} c_i\varphi_i \right\|_{L^2(\Omega)} \in \mathcal{O}\left(M^{-\gamma}\right), M \to \infty$$

will be denoted by $\gamma^{*,\mathrm{eff}}(\mathcal{C},\mathcal{D})$ and referred to as effective best M-term approximation rate of $\mathcal{C}$ in the representation system $\mathcal{D}$.

Theorem 10.8. Let $d \in \mathbb{N}$ and $\Omega \subset \mathbb{R}^d$. The effective best M-term approximation rate of the function class $\mathcal{C} \subset L^2(\Omega)$ in the representation system $\mathcal{D} \subset L^2(\Omega)$ satisfies

$$\gamma^{*,eff}(\mathcal{C},\mathcal{D}) \leq \gamma^*(\mathcal{C})$$

9 Sparsity in Redundant Dictionaries

9.0.1 Linear Approximation Error

Let $\mathcal{B} = \{g_j\}_{j\in\mathbb{N}_0}$ be an ONB for $\mathcal{H}$.

$$\mathbf{x} = \sum_{j=0}^{\infty} \langle \mathbf{x}, g_j\rangle\, g_j \text{ and } \mathbf{x}_m = \sum_{j=0}^{M-1} \langle \mathbf{x}, g_j\rangle\, g_j$$

$$\mathbf{x} - \mathbf{x}_M = \sum_{j=M}^{\infty} \langle \mathbf{x}, g_j\rangle\, g_j \Rightarrow \mathcal{E}_l[M] = \|\mathbf{x} - \mathbf{x}_M\|^2 = \sum_{j=M}^{\infty} |\langle \mathbf{x}, g_j\rangle|^2$$

Theorem 11.1. Let $s > 1/2$ and $\sum_{j=0}^{\infty} j^{2s} |\langle \mathbf{x}, g_j\rangle|^2 < \infty$. There exist constants $A, B > 0$ s.t.

$$A\sum_{j=0}^{\infty} j^{2s} |\langle \mathbf{x}, g_j\rangle|^2 \leq \sum_{M=0}^{\infty} M^{2s-1}\mathcal{E}_l[M] \leq B\sum_{j=0}^{\infty} j^{2s} |\langle \mathbf{x}, g_j\rangle|^2$$

and hence $\mathcal{E}_l[M] = o\left(M^{-2s}\right)$ i.e. $\lim_{M\to\infty} \mathcal{E}_l[M]M^{2s} = 0$

Theorem 11.2. If $\mathbf{x}$ is integrable, then its FT is a uniformly continuous function satisfying

$$|\hat{x}(f)| \leq \int_{-\infty}^{\infty} |x(t)|dt < \infty, \quad f \in \mathbb{R} \text{ and } \lim_{|f|\to\infty} \hat{x}(f) = 0$$

From the theorem we can also see that if $\hat{x}$ is integrable, then $\mathbf{x}$ is uniformly continuous and bounded and satisfies $\lim_{|t|\to\infty} x(t) = 0$.

Proposition 11.3. A function x is bounded and p times continuously differentiable with bounded derivative if $\int_{-\infty}^{\infty} |\hat{x}(f)|(1+|f|)^p df < \infty$(proof use $x^{(k)}(t) \overset{\mathcal{F}}{\Longleftrightarrow} (2\pi\mathrm{i}f)^k\hat{x}(f)$ The proposition implies that if $\exists$ const. k and $\epsilon > 0$ s.t $|\hat{x}(f)| \leq \frac{k}{(1+|f|)^{p+1+\epsilon}}$ then $\mathbf{x} \in C^p$

9.1 Nonlinear Approximations

9.1.1 Approximation Error

Let $\mathbf{x} \in \mathcal{H}$ approximated with M elements in an ONB $\mathcal{B} = \{g_j\}_{j\in\mathbb{N}}$ of $\mathcal{H}$. $\mathbf{X}_M$ projection of $\mathbf{x}$ over M elements whose indices are collected in $\mathcal{I}_M$

$$\mathbf{x}_M = \sum_{j\in\mathcal{I}_M} \langle \mathbf{x}, g_j\rangle\, g_j \text{ and } \mathcal{E}_n[M] = \|\mathbf{x}-\mathbf{x}_M\|^2 = \sum_{j\notin\mathcal{I}_M} |\langle \mathbf{x}, g_j\rangle|^2$$

To minimize this error, the indices in $\mathcal{I}_M$ must correspond to the M elements having the largest inner product amplitudes since

$$\|\mathbf{x}\|^2 = \sum_j |\langle \mathbf{x}, g_j\rangle|^2 = \sum_{j\in\mathcal{I}_M} |\langle \mathbf{x}, g_j\rangle|^2 + \sum_{j\notin\mathcal{I}_M} |\langle \mathbf{x}, g_j\rangle|^2 .$$

We also have $\mathcal{E}_n[M] \le \mathcal{E}_l[M]$. Let $x^r_{\mathcal{B}}[k] = \langle \mathbf{x}, g_{j_k}\rangle$, where $|x^r_{\mathcal{B}}[k]| \ge |x^r_{\mathcal{B}}[k+1]|, k \ge 1$. The best nonlinear approx. is $\mathbf{x}_M = \sum_{k=1}^{M} x^r_{\mathcal{B}}[k] g_{j_k}$ & $\mathcal{E}_n[M] = \|\mathbf{x}-\mathbf{x}_M\|^2 = \sum_{k=M+1}^{\infty} |x^r_{\mathcal{B}}[k]|^2$

Theorem 11.4. Let $s > \frac{1}{2}$. If $\exists X > 0$ s.t. $x^r_{\mathcal{B}}[k] \le Ck^{-s}$, then $\mathcal{E}_n[M] \le \frac{C^2}{2s-1} M^{1-2s}$ and
$\mathcal{E}_n[M] \le \frac{C^2}{2s-1} M^{1-2s} \Rightarrow |x^r_{\mathcal{B}}[k]| \le \left(1-\frac{1}{2s}\right)^{-s} Ck^{-s}$

Theorem 11.5. Let p ¡ 2. If $\|\mathbf{x}\|_{\mathcal{B},p} < \infty$, then

$$|x^r_{\mathcal{B}}[k]| \le \|\mathbf{x}\|_{\mathcal{B},p} k^{-1/p} \text{ and } \mathcal{E}_n[M] = o\left(M^{1-2/p}\right)$$

i.e. $\lim_{M\to\infty} \mathcal{E}_n[M] M^{-1+2/p} = 0$.

9.1.2 Best M-term Approximation in Redundant Dictionaries

$\mathcal{D} = \{g_j\}_{j\in\Gamma}$ dictionary with $\|g_j\| = 1$, $g_j \in \mathbb{C}^N$, $|\Gamma| \ge N$. We study sparse approximations if $x \in \mathbb{C}^N$ with vectors selected in $\mathcal{D}$. Let Δ s.t. $|\Delta| \le N$. $x_\Delta = \sum_{j\in\Delta} a[j] g_j$.

$$\begin{aligned}
&\mathcal{E}_n(x,\Delta) = \|x - x_\Delta\|^2 = \left\|x - \textstyle\sum_{j\in\Delta} a[j] g_j\right\|^2 \\
&= \left\|\textstyle\sum_{j\in\Gamma} \langle x, h_j\rangle\, g_j - \sum_{j\in\Delta} a[j] g_j\right\|^2 \\
&= \left\|\textstyle\sum_{j\in\Delta} \left(\langle x, h_j\rangle - a[j]\right) g_j + \sum_{j\in\Gamma\setminus\Delta} \langle x, h_j\rangle\, g_j\right\|^2
\end{aligned}$$

Where $\{h_j\}$ denotes the dual frame of $\{g_j\}$.
If $\{g_j\}_{j\in\Gamma}$ is an ONB, we get $h_j = g_j$, hence

$$\mathcal{E}_n(x,\Delta) = \| \sum_{j\in\Delta} \left(\langle x, h_j\rangle - a[j]\right) g_j + \sum_{j\in\Gamma\setminus\Delta} \langle x, h_j\rangle\, g_j\|^2$$

For given support set Δ, $\mathcal{E}_n(x,\Delta)$ is minimized if

$$a[j] = \langle x, g_j\rangle \Rightarrow \mathcal{E}_n(x,\Delta) = \sum_{j\in\Gamma\setminus\Delta} |\langle x, g_j\rangle|^2$$

From parseval: $\|x\|^2 = \sum_{j\in\Delta} |\langle x, g_j\rangle|^2 + \sum_{j\in\Gamma\setminus\Delta} |\langle x, g_j\rangle|^2$
Hence optimum support set of a given cardinality (say M), is obtained by maximizing $\sum_{j\in\Delta} |\langle x, g_j\rangle|^2$ over all support sets of $|\Delta| = M$. In absence of ortonormality of the vectors g_j, we have

$$\mathcal{E}_n(x,\Delta) = \left\| \sum_{j\in\Delta} \left(\langle x, h_j\rangle - a[j]\right) g_j + \sum_{j\in\Gamma\setminus\Delta} \langle x, h_j\rangle\, g_j \right\|^2$$

This optimization problem requires enumeration of all possibilities, which is NP-Hard.

9.2 Matching Pursuit

The matching pursuit algorithm computes non-optimal yet good M-term approximations in a computationally efficient way. Let $\mathcal{D} = \{g_j\}_{j\in\Gamma}$ be a dicitonary of $|\Gamma| \ge N$ vectors having unit norm. The dictionary is assumed to be complete. We consider the orthogonal MP algo.
Let $R^0 x = x$. Suppose that $R^m x$ is already computed for $m \ge 1$. The algo stops if $R^m x = 0$. Otherwise, the algorithm computes $g_{j_m} \in \mathcal{D}$ according to
$j_m = \arg\max_{j\in\Gamma} |\langle R^m x, g_j\rangle|$. A Gram-Schmidt step orthogonalizes g_{j_m} w.r.t. $\{g_{j_e}\}_{0\le e\le m-1}$ and computes
$u_m = g_{j_m} - \sum_{l=0}^{m-1} \frac{\langle g_{j_m}, u_l\rangle}{\|u_l\|^2} u_l$.
$R^m x$ projected onto orthogonal complement of space spanned by u_m according to $R^{m+1}x = R^m x - \frac{\langle R^m x, u_m\rangle}{\|u_m\|^2} u_m$ and so on. Note

$$\begin{aligned}
u_m &= g_{j_m} - \sum_{l=0}^{m-1} \frac{\langle g_{j_m}, u_l\rangle}{\|u_l\|^2} u_l \\
&= \underbrace{\underbrace{\left(\mathbf{I} - \sum_{l=0}^{m-1} \frac{u_l u_l^H}{\|u_l\|^2}\right)}_{\mathbb{P}_{V_{m-1}}}}_{\substack{\text{projection onto} \\ \text{orthogonal complement} \\ \text{of } V_{m-1}=\text{span}\{u_0,\ldots,u_{m-1}\}}} g_{j_m}
\end{aligned}$$

Therefore $R^m x = x - \mathbb{P}_{V_{m-1}} x \Rightarrow x = \mathbb{P}_{V_{m-1}} x + R^m x = \sum_{l=0}^{m-1} \frac{\langle x, u_l\rangle}{\|u_l\|^2} u_l + R^m x$. The algo stops after $M \le N$ iterations and yields

$$x = \sum_{l=0}^{M-1} \frac{\langle x, u_l\rangle}{\|u_l\|^2} u_l \Rightarrow \|x\|^2 = \sum_{l=0}^{M-1} \frac{|\langle x, u_l\rangle|^2}{\|u_l\|^2}$$

10 Uniform Laws of Large Numbers

10.1 Uniform convergence of CDFs

RV X can be fully specified by its CDF $F(t) = \mathbb{P}[X \le t], t \in \mathbb{R}$. Sps we want to estimate $F(t)$ from a given collection $\{X_i\}_{i=1}^n$ of i.i.d. samples. A natural estimate of F is $\hat{F}_n(t) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{(-\infty;t]}[X_i]$ (since $F(t) = \mathbb{E}\left[\mathbb{1}_{(-\infty;t]}[X]\right]$). Note

$$\mathbb{E}\left[\hat{F}_n(t)\right] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[\mathbb{1}_{(-\infty;t]}[X_i]\right] = \frac{1}{n}\sum_{i=1}^n F(t) = F(t)$$

one can show $\mathbb{P}\left[\lim_{n\to\infty} \hat{F}_n(t) = F(t)\right] = 1, \quad \forall t \in \mathbb{R}$.

Def. 12.4. A sequence of RVs X_n converges in probability to tyhe RV X if $\lim_{n\to\infty} \mathbb{P}\{|X - X_n| > \varepsilon\} = 0, \quad \forall \varepsilon > 0$.
Def 12.5. The functional γ is continuous at F in the sup-norm if, $\forall \varepsilon > 0, \exists \delta > 0$ s.t. $\|G - F\|_\infty \le \delta$ implies $|\gamma(G) - \gamma(F)| \le \varepsilon$.
Theorem 12.6. (Glivenko-Cantelli) For any distribution, the empirical CDF $\hat{F}_n$ satisfies $\left\|\hat{F}_n - F\right\|_\infty \xrightarrow{\text{a.s.}} 0$
Using this theorem we can prove

$$\lim_{n\to\infty} \mathbb{P}\left[\left|\gamma\left(\hat{F}_n\right) - \gamma(F)\right| > \varepsilon\right] = 0, \quad \forall \varepsilon > 0$$

Now, by continuity of $\gamma(\cdot)$ w.r.t. $\|\cdot\|_\infty$-norm, we get that for each $\varepsilon > 0$, there exists a $\delta > 0$ such that $\left\|\hat{F}_n - F\right\|_\infty \le \delta$ implies that $\left|\gamma\left(\hat{F}_n\right) - \gamma(F)\right| \le \varepsilon$, i.e.,

$$\left\|\hat{F}_n - F\right\|_\infty \le \delta \Longrightarrow \left|\gamma\left(\hat{F}_n\right) - \gamma(F)\right| \le \varepsilon.$$

Hence, it follows that

$$\left|\gamma\left(\hat{F}_n\right) - \gamma(F)\right| > \varepsilon \Longrightarrow \left\|\hat{F}_n - F\right\|_\infty > \delta,$$

which implies

$$\mathbb{P}\left[\left|\gamma\left(\hat{F}_n\right) - \gamma(F)\right| > \varepsilon\right] \le \mathbb{P}\left[\left\|\hat{F}_n - F\right\|_\infty > \delta\right]$$

From the theorem we get

$$\begin{aligned}
&\left\|\hat{F}_n - F\right\|_\infty \xrightarrow{a.s.} 0 \Rightarrow \left\|\hat{F}_n - F\right\|_\infty \xrightarrow{p} 0 \\
&\Rightarrow \lim_{n\to\infty} \mathbb{P}\left[\left\|\hat{F}_n - F\right\|_\infty > \varepsilon'\right] = 0, \quad \forall \varepsilon' > 0 \\
&\Rightarrow \lim_{n\to\infty} \mathbb{P}\left[\left|\gamma\left(\hat{F}_n\right) - \gamma(F)\right| > \varepsilon\right] \le \lim_{n\to\infty} \mathbb{P}\left[\left\|\hat{F}_n - F\right\|_\infty > \delta\right] = 0 \\
&\Rightarrow \lim_{n\to\infty} \mathbb{P}\left[\left|\gamma\left(\hat{F}_n\right) - \gamma(F)\right| > \varepsilon\right] = 0, \quad \forall \varepsilon > 0 \quad \square
\end{aligned}$$

10.2 Uniform laws for more general function classes

Let $\mathcal{F}$ be a class of integrable real-valued functions with domain $\mathcal{X}$, and let $\{X_i\}_{i=1}^n$ be a collection of i.i.d. samples from some distribution $\mathbb{P}$ over $\mathcal{X}$. Consider the RV

$$\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}} = \sup_{f\in\mathcal{F}} \left| \frac{1}{n}\sum_{i=1}^n f(X_i) - \mathbb{E}[f(X)] \right|$$

Def. 12.7. We say that $\mathcal{F}$ is a Glivenko-Cantelli clas for $\mathbb{P}$ if $\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}}$ converges to zero in probability as $n \to \infty$.

A stronger notion of this concept requires almost sure convergence and leads to what is referred to as the strong Glivenko-Cantelli property. Recall that almost sure convergence implies convergence in probability.

Example 12.8. (Empirical CDFs and indicator functions) Consider the function class $\mathcal{F} = \left\{ \mathbb{1}_{(-\infty;t]}(\cdot) \mid t \in \mathbb{R} \right\}$ For each fixed $t \in \mathbb{R}$ $\mathbb{E}\left[\mathbb{1}_{(-\infty;t]}(X)\right] = \mathbb{P}[X \le t] = F(t)$. Hence theorem 12.6. i.e.

$$\mathbb{P}\left[\lim_{n\to\infty} \sup_{t\in\mathbb{R}} \left| \hat{F}_n(t) - F(t) \right| = 0 \right] = 1$$

can be interpreted as a strong uniform law for the class $\mathcal{F}$. specifically by rewriting the previous eq.

$$\mathbb{P}[\lim_{n\to\infty} \sup_{t\in\mathbb{R}} | \underbrace{\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{(-\infty;t]}(X_i)}_{\hat{F}_n(t)} - \mathbb{E}\left[\mathbb{1}_{(-\infty;t]}(X)\right] | = 0] = 1$$

which is simply

$$\mathbb{P}\left[\lim_{n\to\infty} \sup_{f\in\mathcal{F}} \left| \frac{1}{n}\sum_{i=1}^n f(X_i) - \mathbb{E}[f(X)] \right| = 0 \right] = 1$$

Variables of the form $\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}}$ are prevalent in statistics and lie at the heart of empirical risk minimization.

Def. Empirical Risk: $\hat{R}_n(\theta, \theta^*) = \frac{1}{n}\sum_{i=1}^n \mathcal{L}_\theta(X_i)$

Def. Population Risk: $R(\theta, \theta^*) = \mathbb{E}_{\theta^*}[\mathcal{L}_\theta(X)]$

In practice one minimizes the empirical risk over some subset Ω_0 of the full space Ω thus obtaining some estimate $\hat{\theta}$.

Def. Excess Risk: $E(\hat{\theta}, \theta^*) = R(\hat{\theta}, \theta^*) - \inf_{\theta\in\Omega_0} R(\theta, \theta^*)$

Example MLE: Cost Function $\mathcal{L}_\theta(x) = \log\left[\frac{p_{\theta^*}(x)}{p_\theta(x)}\right]$

$\hat{\theta} = \arg\min_{\theta\in\Omega_0} \mathcal{L}_\theta(x)$ and $R(\theta, \theta^*) = \mathbb{E}_{\theta^*}\left[\log\left[\frac{p_{\theta^*}(X)}{p_\theta(X)}\right]\right] = D(p_{\theta^*} \| p_\theta)$ the last term is the KL-Divergence.

If $\theta^* \in \Omega_0$, $\inf_{\theta\in\Omega_0} R(\theta, \theta^*) = 0 \Rightarrow E\left(\hat{\theta}, \theta^*\right) = R\left(\hat{\theta}, \theta^*\right)$ Generally, assume $\exists\theta_0 \in \Omega_0$ s.t. $R(\theta_0, \theta^*) = \inf_{\theta\in\Omega_0} R(\theta, \theta^*)$. Then Excess Risk can be decomposed as

$$\underbrace{R\left(\hat{\theta}, \theta^*\right) - \hat{R}_n\left(\hat{\theta}, \theta^*\right)}_{T_1} + \underbrace{\hat{R}_n\left(\hat{\theta}, \theta^*\right) - \hat{R}_n\left(\theta_0, \theta^*\right)}_{T_2} + \underbrace{\hat{R}_n\left(\theta_0, \theta^*\right) - R\left(\theta_0, \theta^*\right)}_{T_3}$$

Note that $T_2 \le 0$.

$T_1 = \mathbb{E}_X\left[\mathcal{L}_{\hat{\theta}}(X)\right] - \frac{1}{n}\sum_{i=1}^n \mathcal{L}_{\hat{\theta}}(X_i)$ and it can be bounded by setting $\mathcal{F} = \{\mathcal{L}_\theta(\cdot) \mid \theta \in \Omega_0\}$ and noting that

$$T_1 \le \sup_{\theta\in\Omega_0} \left| \frac{1}{n}\sum_{i=1}^n \mathcal{L}_\theta(X_i) - \mathbb{E}_X\left[\mathcal{L}_\theta(X)\right] \right| = \|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}}$$

We also have $T_3 \le \|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}}$. Therefore the excess risk is upper-bounded by $2\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}}$, which shows that the central challenge in analyzing estimates based on empirical risk minimization is to establish uniform laws of large numbers.

10.3 Uniform laws via Rademacher complexity

Def. Rademacher Complexity: For any fixed collection $x_1^n = (x_1, \ldots, x_n)$ of points consider the subset of $\mathbb{R}^n$ given by $\mathcal{F}(x_1^n) = \{(f(x_1), f(x_2), \ldots, f(x_n)) \mid f \in \mathcal{F}\}$ (i.e the set of all vectors in $\mathbb{R}^n$ that can be realized by applying a function $f \in \mathcal{F}$ to the collection $(x_1, \ldots, x_n)$. The Empirical Rademacher Complexity is defined as

$$\mathcal{R}(\mathcal{F}(x_1^n)/n) = \mathbb{E}_\varepsilon\left[\sup_{f\in\mathcal{F}} \left| \frac{1}{n}\sum_{i=1}^n \varepsilon_i f(x_i) \right| \right]$$

where $\{\varepsilon_k\}_{k=1}^n$ is an i.i.d. sequence of Rademacher random variables (i.e., taking values $\{-1, +1\}$). The empirical Rademacher Complexity is a RV, and taking its Expectation yields the Rademacher Complexity of the function class

$$\mathcal{R}_n(\mathcal{F}) = \mathbb{E}_X\left[\mathcal{R}(\mathcal{F}(X_1^n)/n)\right] = \mathbb{E}_{\varepsilon,X}\left[\sup_{f\in\mathcal{F}} \left| \frac{1}{n}\sum_{i=1}^n \varepsilon_i f(X_i) \right| \right]$$

The Rademacher Complexity is the average of the maximum correlation between the vector $(f(X_1), \ldots, f(X_n))$ and the "noise vector" $(\varepsilon_1, \ldots, \varepsilon_n)$

Theorem 12.10. For any b-uniformly bounded function class, i.e., $\|f\|_\infty \le b$, $\forall f \in \mathcal{F}$,any positive integer $n \ge 1$ and any scalar $\delta \ge 0$,we have

$$\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}} \le 2\mathcal{R}_n(\mathcal{F}) + \delta$$

with $\mathbb{P}$-probability at least $1 - e^{-\frac{n\delta^2}{2b^2}}$. Consequently, as long as $\mathcal{R}_n(\mathcal{F}) = o(1)$, we have $\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}} \xrightarrow{\text{a.s.}} 0$

Proposition 12.11. For any b-uniformly bounded function class $\mathcal{F}$, any integer $n \ge 1$ and any scalar $\delta \ge 0$, we have

$$\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}} \ge \frac{1}{2}\mathcal{R}_n(\mathcal{F}) - \frac{\sup_{f\in\mathcal{F}} |\mathbb{E}_{\mathbb{P}}[f]|}{2\sqrt{n}} - \delta$$

with $\mathbb{P}$-probability at least $1 - e^{-\frac{n\delta^2}{2b^2}}$

10.4 Function classes with polynomial discrimination

The expectation and the sup, needed to compute the Rademacher Complexity are often difficult to compute analytically. In this section we look at techniques for upper-bounding $\mathcal{R}_n(\mathcal{F})$ that are more accessible, yet preserve the growth behavior needed to infer the Glivenko-Cantelli property. Such techniques apply only to certain function classes. In particular, we will be concerned with classes with polynomial discrimination.

For a given collection of points $x_1^n = (x_1, \ldots, x_n)$, the "size" of the set $\mathcal{F}(x_1^n)$ provide a sample-dependent measure of the complexity of $\mathcal{F}$.

Def. 12.12. (Polynomial discrimination) A class $\mathcal{F}$ of functions with domain $\mathcal{X}$ j has polynomial discrimination of order $\nu \ge 1$ if, for each positive integer n and collection x_1^n of n points in $\mathcal{X}$, the set $\mathcal{F}(x_1^n)$ has cardinality upper-bounded according to

$$|\mathcal{F}(x_1^n)| \le (n+1)^\nu$$

Lemma 12.13. Suppose that $\mathcal{F}$ has polynomial discrimination of order ν. Then, for all positive integers n and any collection of points $x_1^n = (x_1, \ldots, x_n)$, we have

$$\underbrace{\mathbb{E}_\varepsilon\left[\sup_{f\in\mathcal{F}} \left| \frac{1}{n}\sum_{i=1}^n \varepsilon_i f(x_i) \right| \right]}_{\mathcal{R}(\mathcal{F}(x_1^n)/n)} \le 4D(x_1^n)\sqrt{\frac{\nu\log(n+1)}{n}}$$

where $D(x_1^n) = \sup_{f\in\mathcal{F}} \sqrt{\frac{\sum_{i=1}^n f^2(x_i)}{n}}$ is the ℓ_2-radius of the set $\mathcal{F}(x_1^n)/\sqrt{n}$. Even though this lemma provides us with an upper bound on the empirical Rademacher complexity only, taking the expectation w.r.t. X yields an upper-bound on $\mathcal{R}_n(\mathcal{F})$ according to

$$\mathcal{R}_n(\mathcal{F}) = \mathbb{E}_X\left[\mathcal{R}(\mathcal{F}(X_1^n)/n)\right] \le 4\mathbb{E}_{X_1^n}\left[D(X_1^n)\right]\sqrt{\frac{\nu\log(n+1)}{n}}$$

And if the functiion class is b-uniformly bounded s.t. $D(x_1^n) \le b\forall$samples, then $\mathcal{R}_n(\mathcal{F}) \le 4b\sqrt{\frac{\nu\log(n+1)}{n}}\forall n \ge 1$ and thanks to theorem12.10, we conclude that every bounded function class with polynomial discrimination is Glivenko-Cantelli.

Corollary 12.14 (Classical Glivenko-Cantelli) Let $F(t) = \mathbb{P}[X \le t]$ be the CDF of a RV $X \sim \mathbb{P}$, and let $\hat{F}_n$ be the empirical CDF based on n i.i.d. samples $X_i \sim \mathbb{P}$. Then

$$\mathbb{P}\left[\left\| \hat{F}_n - F \right\|_\infty \ge 8\sqrt{\frac{\log(n+1)}{n}} + \delta \right] \le e^{-\frac{n\delta^2}{2}}, \quad \delta \ge 0$$

hence $\left\| \hat{F}_n - F \right\|_\infty \xrightarrow{\text{a.s.}} 0$.

10.5 Vapnik-Chervonenkis dimension

Let us consider a function class in which each function is binary-valued, taking the values $\{0,1\}$. In this case, the set $\mathcal{F}(x_1^n) = \{f(x_1), \ldots, f(x_n) \mid f \in \mathcal{F}\}$ can have at most 2^n elements.

Definition 12.15. (Shattering and VC dimension) Given a class $\mathcal{F}$ of binary-valued functions, we say thet the set $x_1^n = (x_1, \ldots, x_n)$ is shattered by $\mathcal{F}$ if $|\mathcal{F}(x_1^n)| = 2^n$. The VC dimension $\nu(\mathcal{F})$ is the largest integer n for which there is some collection $x_1^n = (x_1, \ldots, x_n)$ of n points that is shattered by $\mathcal{F}$. When $\nu(\mathcal{F})$ is finite, $\mathcal{F}$ is said to be a VC class.

Theorem 12.17. (Vapnik-Chervonenkis, Sauer-Shelah) Consider a set class $\mathcal{S}$ with $\nu(\mathcal{S}) < \infty$. Then, for any collection of points $P = (x_1, \ldots, x_n)$ with $n \geq \nu(\mathcal{S})$

$$|\mathcal{S}(P)| \leq \sum_{i=0}^{\nu(\mathcal{S})} \binom{n}{i} \leq (n+1)^{\nu(\mathcal{S})}$$