Wavelets, Sparsity and their Applications

Pier Luigi Dragotti

Communications and Signal Processing Group
Imperial College London

Session two: Mathematical Background

Mathematical Background

Definition 1. [Vector Space] *By a vector space we mean a nonempty set E with two operations:*

vector space $\begin{cases} \text{addition} \\ \text{multiply by numbers} \end{cases}$

- *a mapping $(x, y) \to x + y$ from $E \times E$ into E called addition,*
- *a mapping $(\lambda, x) \to \lambda x$ from $\mathbb{R} \times E$ into E.*

such that the following conditions are satisfied:

1. *Commutativity: $x + y = y + x$;*
2. *Associativity $(x + y) + z = x + (y + z)$;*
3. *Distributivity $(\alpha + \beta)x = \alpha x + \beta x$ and $\alpha(x + y) = \alpha x + \alpha y$;*
4. *For every $x, y \in E$ there exists $z \in E$ such that $x + z = y$;*
5. *$\alpha(\beta x) = (\alpha\beta)x$;*
6. *$1x = x$.*

Mathematical Background

$\underline{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$ $\underline{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$

Definition 2. [Inner Product Space] *Let E be a complex vector space. A mapping $\langle \cdot, \cdot \rangle$ is called an inner product in E if for any $x, y, z \in E$ and $\alpha, \beta \in \mathbb{C}$ the following conditions are satisfied:*

$\vec{u} \cdot \vec{v}$ column vector

inner product $\langle \underline{u}, \underline{v} \rangle = \underline{v}^H \underline{u} = (v_1\ v_2 \cdots v_n) \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} = \underbrace{\sum_i u_i v_i^*}_{1 \times 1 \text{ scalar}}$

outer product $\underline{u} \otimes \underline{v} = \underline{u}\, \underline{v}^H = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} (v_1\ v_2 \cdots v_n)$

$= \begin{bmatrix} u_1 v_1 & u_1 v_2 & \cdots & u_1 v_n \\ u_2 v_1 & u_2 v_2 & \cdots & u_2 v_n \\ \vdots & \vdots & \ddots & \vdots \\ u_n v_1 & u_n v_2 & \cdots & u_n v_n \end{bmatrix}$ $n \times n$ matrix

1. $\langle x, y \rangle = \overline{\langle y, x \rangle}$;
2. $\langle\ \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle$;
3. $\langle x, x \rangle \geq 0$ and $\langle x, x \rangle = 0$ implies $x = 0$.

A vector space E with an inner product is called Inner Product Space.

$(\underline{u}\,\underline{v}^H)_{ij} = u_i v_j$

Given the definition of inner product, we define the norm of a vector x as:

$$\|x\| = \sqrt{\langle x, x \rangle}.$$

norm of vector

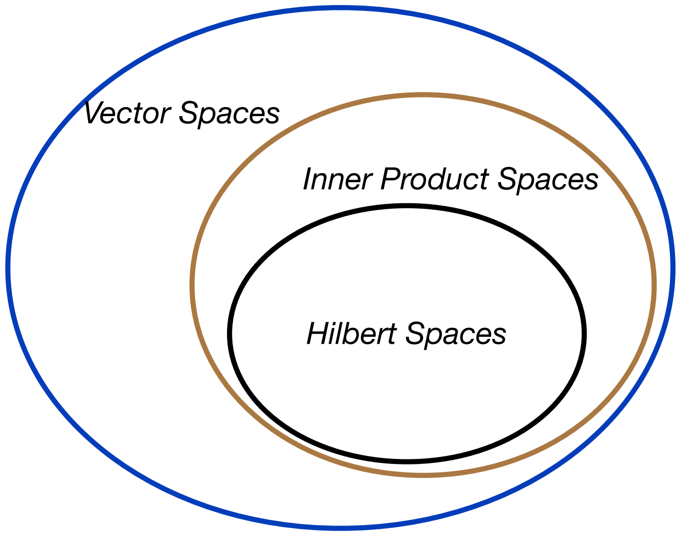
$\|x\| = \sqrt{\langle x, x \rangle} = \|x\|_2 = (\sum_i x_i^2)^{\frac{1}{2}}$

Mathematical Background

(safeguard)

Definition 3. [Hilbert Space] *A complete inner product space is called Hilbert Space.*

By completeness of the inner product space E, we mean that every Cauchy sequence in E converges to an element of E. A Cauchy sequence is defined as a sequence of vectors $x_n \in E$ such that for every $\epsilon > 0$ there exists a number M such that $\|x_m - x_n\| < \epsilon$ for all $m, n > M$.



$L_2(\mathbb{R})$

· inner product

$\langle f, g\rangle = \int_{-\infty}^{+\infty} f^*(t)g(t)dt$

· norm

$\|f(t)\| = \langle f, f\rangle^{\frac{1}{2}} = \sqrt{\int_{-\infty}^{+\infty} |f(t)|^2 dt}$

Mathematical Background

Examples of Hilbert Spaces

Example 1. [Square-integrable functions] The space of all complex-valued functions $f(t)$, $t \in \mathbb{R}$ such that

$$\int_{-\infty}^{\infty} |f(t)|^2 dt < \infty$$

$L_2(\mathbb{R})$ is composed of all $f(t)$ that satisfy this condition.

and it is denoted by $L_2(\mathbb{R})$. The inner product in $L_2(\mathbb{R})$ is given by $\langle f, g\rangle = \int_{-\infty}^{\infty} f^*(t)g(t)dt$, where $f^*(t)$ is the complex conjugate of $f(t)$ and the norm is $\|f(t)\| = \sqrt{\langle f, f\rangle} = \sqrt{\int_{-\infty}^{\infty} |f(t)|^2 dt}$.

$L_2(\mathbb{Z})$

Example 2. [Square-summable sequences] The space of all complex-valued sequences $x[n]$, $n \in \mathbb{Z}$ such that

· inner product

$\langle x, y\rangle = \sum_{n=-\infty}^{+\infty} x[n]y^*[n]$

· norm

$\|x\| = \langle x, x\rangle^{\frac{1}{2}} = \sqrt{\sum_{n=-\infty}^{+\infty} |x[n]|^2}$

$$\sum_{n=-\infty}^{\infty} |x[n]|^2 < \infty$$

$l_2(\mathbb{Z})$ is composed of all $x[n]$ that satisfy this condition.

and it is denoted by $l_2(\mathbb{Z})$. The inner product in $l_2(\mathbb{Z})$ is given by $\langle x, y\rangle = \sum_{n=-\infty}^{\infty} x[n]y^*[n]$ and the norm is $\|x\| = \sqrt{\langle x, x\rangle} = \sqrt{\sum_{n=-\infty}^{\infty} |x[n]|^2}$.

Example 3. N-dimensional vector spaces $\mathbb{R}^N$ or $\mathbb{C}^N$.

Mathematical Background

Definition 4. [Subspace] *A subset S of a vector space E is a subspace when it is closed under the operations of vector addition and scalar multiplication:*

1. *For all x and y in S, $x + y$ is in S.*
2. *For all x in S and α in $\mathbb{C}$, αx is in S.*

Definition 5. [Span] *The span of a set of vectors S is the set of all finite linear combinations of vectors in S:*

$$\mathit{span}(S) = \left\{ \sum_{k=0}^{N-1} \alpha_k \varphi_k \mid \alpha_k \in \mathbb{C}, \varphi_k \in S \text{ and } N \in \mathbb{N}. \right\}$$

Note that a span is always a subspace and that the sum has a finite number of terms even if the set S is infinite.

Mathematical Background

Definition 6. [Linear Independence] *The set of vectors $\{\varphi_0, \varphi_1, \ldots, \varphi_{N-1}\}$ is called linearly independent when $\sum_{k=0}^{N-1} \alpha_k \varphi_k = 0$ is true only if $\alpha_k = 0$ for all k. Otherwise, the set is linearly dependent.*

Definition 7. [Dimension] *A vector space E is said to have dimension N when it contains a linearly independent set with N elements and every set with $N+1$ or more elements is linearly dependent. If no such finite N exists, the vector space is infinite dimensional.*

On the Notion of Norm

Definition 8. [Norm] *A norm on a vector space V over $\mathbb{C}$ (or $\mathbb{R}$) is a real valued function with the following properties for any $x, y \in V$ and $\alpha \in \mathbb{C}$*

1. *Positive definiteness:* $\|x\| \geq 0$, *and* $\|x\| = 0$ *if and only if* $x = 0$.
2. *Homogeneity Property:* $\|\alpha x\| = |\alpha| \|x\|$.
3. *Triangle inequality:* $\|x + y\| \leq \|x\| + \|y\|$

Given $p \in [1, \infty)$, the l_p norm of the vector $\mathbf{x}$ is given by:

$$\|\mathbf{x}\|_p = (\sum_i |x_i|^p)^{1/p}. \quad (p \geq 1)$$

On the Notion of Norm (cont'd)

$$\|x\|_p = \begin{cases} \left(\sum_i |x_i|^p\right)^{\frac{1}{p}}, & p \ge 1 \\ \left(\frac{1}{n}\sum_i |x_i|^p\right)^{\frac{1}{p}}, & 0 \le p < 1 \end{cases}$$

draw the combination of (x_0, x_1)

s.t. $\|x\|_1, \|x\|_2 \cdots \|x\|_\infty = 1$, N.B. unit circle

$\|x\|_1 = \sum_i |x_i| = |x_0| + |x_1| = 1 \Rightarrow |x_1| = 1 - |x_0|$

Q_1: $x_1 = 1 - x_0$

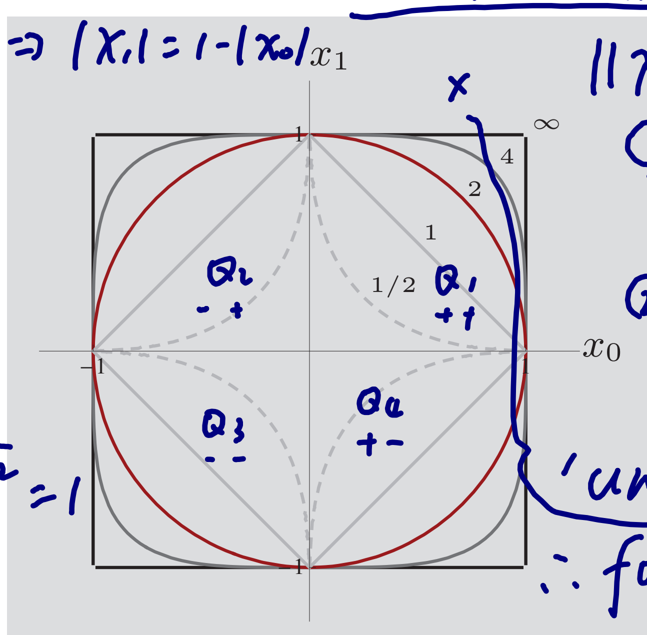
Q_2: $x_1 = 1 + x_0$

Q_3: $-x_1 = 1 + x_0 \Rightarrow x_1 = -1 - x_0$

Q_4: $-x_1 = 1 - x_0 \Rightarrow x_1 = x_0 - 1$

$\|x\|_2 = \left(\sum_i |x_i|^2\right)^{\frac{1}{2}} = \sqrt{x_0^2 + x_1^2} = 1$

unit circle



$\|x\|_\infty = \max |x_i| = 1$

① $|x_0| = 1 \Rightarrow \begin{cases} x_0 = \pm 1 \\ x_1 : \text{whatever abs} < 1 \end{cases}$

② $|x_1| = 1 \Rightarrow \begin{cases} x_0 : \text{whatever abs} < 1 \\ x_1 = \pm 1 \end{cases}$

'unit circle' expands as $p \nearrow$

$\therefore$ for vector x.

$\|x\|_1 \ge \|x\|_2 \ge \cdots \ge \|x\|_\infty$:

(coordinate or 'unit circle' expands but position remains)

Sets of unit-norm vectors for different p norms. Note that $\|x\|_\infty = \sup |x_n|$ and that $p = 1/2$ is not a norm. The $\|x\|_0$ corresponds to the number of non-zero entries in x and is not a norm.

$\|x\|_2 \triangleq \|x\| =$ distance to origin

Figure taken from 'Foundations of Signal Processing', Vetterli, Kovacevic and Goyal, Springer 2014.

Mathematical Background

Definition 9. [Basis] *Consider a set of elements* $\{\varphi_i\}_{i\in\mathbb{Z}} \in \mathbb{V}$*, this set is called a basis of* $\mathbb{V}$ *when it is complete, meaning that for any signal* $f \in \mathbb{V}$ *there exists a sequence* α_i *such that*

$$f = \sum_{i=-\infty}^{\infty} \alpha_i \varphi_i$$

and the sequence is unique.

Definition 10. [Orthogonal Basis] *The set* $\{\varphi_i\}$ *is an* **orthogonal** *basis of* $\mathbb{V}$ *if*

- *it is a basis*
- *it is orthogonal:* $\langle \varphi_i, \varphi_j \rangle = \delta_{ij}$

$\langle f, \varphi_i \rangle$: the projection length of f in the direction of φ_i.

In this case $\alpha_i = \langle f, \varphi_i \rangle$ and

$$f = \sum_{i=-\infty}^{\infty} \langle f, \varphi_i \rangle \varphi_i.$$

$\langle \varphi_i, \tilde{\varphi}_j \rangle = \delta_{ij} \Rightarrow \langle \varphi_i, \tilde{\varphi}_i \rangle = 1$ definition

Mathematical Background

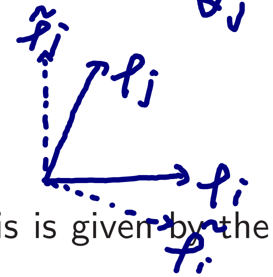
φ_j & $\hat{\varphi}_j$!

$\boxed{\alpha_j = \tilde{\alpha}_j}$

$\alpha_j = \langle f, \varphi_j \rangle = \langle \sum_n \tilde{\alpha}_n \tilde{\varphi}_n, \varphi_j \rangle = \sum_n \tilde{\alpha}_n \langle \tilde{\varphi}_n, \varphi_j \rangle = \tilde{\alpha}_j$

Definition 11. [Biorthogonal Basis] *The set $\{\varphi_i\}$ is a* **biorthogonal** *basis of $\mathbb{V}$ if* same scale for

- *it is a basis* ↓ unique dual basis
- *it is not orthogonal:* $\langle \varphi_i, \varphi_j \rangle \neq \delta_{ij}$



If $\langle \varphi_i, \varphi_j \rangle \neq \delta_{ij}$, one has to design the dual basis first. The dual basis is given by the set of elements $\{\tilde{\varphi}_i\}_{i \in \mathbb{Z}}$ satisfying

$$\langle \varphi_i, \tilde{\varphi}_j \rangle = \delta_{ij}.$$

$\tilde{\alpha}_j = \langle f, \tilde{\varphi}_j \rangle = \langle \sum_n \alpha_n \varphi_n, \tilde{\varphi}_j \rangle = \sum_n \alpha_n \langle \varphi_n, \tilde{\varphi}_j \rangle = \alpha_j$

Then the signal expansion formula becomes

$$f = \sum_{i=-\infty}^{\infty} \langle f, \tilde{\varphi}_i \rangle \varphi_i = \sum_{i=-\infty}^{\infty} \langle f, \varphi_i \rangle \tilde{\varphi}_i.$$

Definition 12. [Frames (informal)] *The set $\{\varphi_i\}_{i \in \mathbb{Z}} \in \mathbb{V}$ is a* **frame** *of $\mathbb{V}$ when it is overcomplete, meaning that for any signal $f \in \mathbb{V}$ there exists a sequence α_i such that*

$$f = \sum_{i=-\infty}^{\infty} \alpha_i \varphi_i$$

but the sequence is **not** *unique.*

$\therefore f = \sum_i \alpha_i \varphi_i = \sum_i \tilde{\alpha}_i \tilde{\varphi}_i$

$= \sum_i \langle f, \tilde{\varphi}_i \rangle \tilde{\varphi}_i = \sum_i \alpha_i \tilde{\varphi}_i$

or $= \sum_i \langle f, \varphi_i \rangle \varphi_i = \sum_i \tilde{\alpha}_i \varphi_i$

Mathematical Background

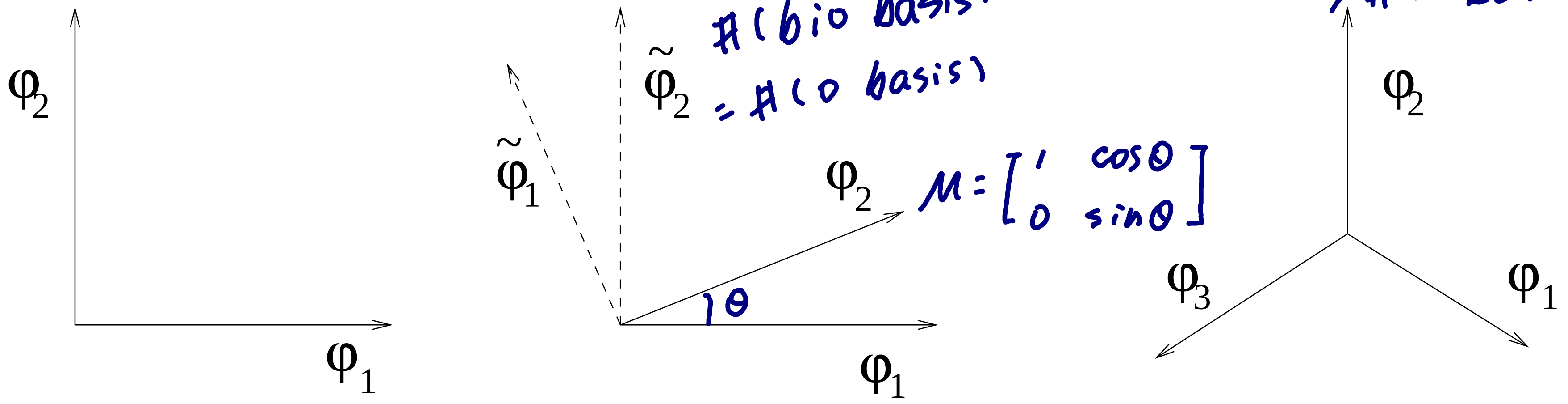
$\tilde{\varphi}_i \perp$ all φ_j when $i \neq j$.

Bases:

$\#(\text{frame}) > \#(\text{o basis})$

$\#(\text{bio basis}) = \#(\text{o basis})$

$M = \begin{bmatrix} 1 & \cos\theta \\ 0 & \sin\theta \end{bmatrix}$



a) Orthogonal Basis b) Biorthogonal Basis c) Frame

$$M = \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ f_1 & f_2 & \cdots & f_N \\ \downarrow & \downarrow & & \downarrow \end{bmatrix} \quad M^H = \begin{bmatrix} \leftarrow f_1^H \rightarrow \\ \leftarrow f_2^H \rightarrow \\ \vdots \\ \leftarrow f_N^H \rightarrow \end{bmatrix}$$

Bases and Frames: Matrix Interpretation

when f_i are orthogonal.

$$M^H M = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \quad \therefore \boxed{M^H = M^{-1}}$$

- Assume the 'atoms' $\{\varphi_i\}$ are finite dimensional column vectors of size *N*
- Stack them one next to the other to form the *synthesis* matrix *M*:

$$M = \begin{bmatrix} \uparrow & \cdots & \uparrow & \cdots \\ \varphi_1 & \cdots & \varphi_i & \cdots \\ \downarrow & \cdots & \downarrow & \cdots \end{bmatrix}$$

- If *M* is square and invertible then $\{\varphi_i\}$ is a basis (of R^N or C^N)

$MM^H = I = M^H M$

- If the inverse of *M* satisfies $M^{-1} = M^H$ the basis is orthogonal
- In the case of frames, *M* is invertible but is *'fat'* ($m<n$)

Bases and Frames: Matrix Interpretation

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = M\alpha = \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ \varphi_1 & \varphi_2 & \cdots & \varphi_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix}$$

$$\alpha = M^{-1}x = M^H x \quad (\text{o. basis})$$

$$\begin{pmatrix} \langle x, \tilde{\varphi}_1 \rangle \\ \langle x, \tilde{\varphi}_2 \rangle \\ \vdots \\ \langle x, \tilde{\varphi}_n \rangle \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} = \begin{bmatrix} \leftarrow & \tilde{\varphi}_1 & \rightarrow \\ \leftarrow & \varphi_2 & \rightarrow \\ & \vdots & \\ \leftarrow & \varphi_n & \rightarrow \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

- Remember the synthesis formula $x(t) = \sum_{i=-\infty}^{\infty} \alpha_i \varphi_i(t)$

$$x = M\alpha \rightleftharpoons x(t) = \sum_{i=-\infty}^{\infty} \alpha_i \varphi_i(t)$$

- In the finite dimensional case the inverse of *M* is the *analysis* matrix and we have: $\alpha = M^{-1}x$

$$\langle x, \tilde{\varphi}_1 \rangle = \tilde{\varphi}_1^H x$$

- Expanded:

$$\begin{pmatrix} \langle x, \tilde{\varphi}_1 \rangle \\ \vdots \\ \langle x, \tilde{\varphi}_i \rangle \\ \vdots \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_i \\ \vdots \end{pmatrix} = \underbrace{\begin{bmatrix} \leftarrow & \tilde{\varphi}_1 & \rightarrow \\ \cdots & \cdots & \cdots \\ \leftarrow & \tilde{\varphi}_i & \rightarrow \\ \cdots & \cdots & \cdots \end{bmatrix}}_{M^{-1}} \begin{pmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \end{pmatrix}$$

- The synthesis formula is $x = M\alpha \quad \Leftrightarrow \quad x(t) = \sum_{i=-\infty}^{\infty} \alpha_i \varphi_i(t)$

Dual Basis

Important: Orthogonal bases satisfy Parseval theorem:

$$\|f\|^2 = \sum_i |\langle f, \varphi_i \rangle|^2 \qquad a^2 + b^2 = c^2$$

$$\|f\|^2 = f_1^2 + f_2^2 + \cdots + f_n^2 = \sum_i |\langle f, \varphi_i \rangle|^2$$

In the case of biorthogonal bases we have:

$$0 < A\|f\|^2 \leq \sum_i |\langle f, \varphi_i \rangle|^2 \leq B\|f\|^2 < \infty$$

$\tilde{\varphi}_k = \sum_{i=1}^{N} a_{k,i}\varphi_i$ $\qquad$ $\langle \tilde{\varphi}_k, \varphi_j \rangle = \sum_{i=1}^{N} a_{k,i} \langle \varphi_i, \varphi_j \rangle$

$\tilde{\Phi} = \Phi A$

Dual Basis

$$\begin{bmatrix} \leftarrow \varphi_1 \rightarrow \\ \leftarrow \varphi_2 \rightarrow \\ \vdots \\ \leftarrow \varphi_n \rightarrow \end{bmatrix} \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ \tilde{\varphi}_1 & \tilde{\varphi}_2 & \cdots & \tilde{\varphi}_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix} = \begin{bmatrix} \leftarrow \varphi_1 \rightarrow \\ \leftarrow \varphi_2 \rightarrow \\ \vdots \\ \leftarrow \varphi_n \rightarrow \end{bmatrix} \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ \varphi_1 & \varphi_2 & \cdots & \varphi_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix} A$$

$\langle u, v \rangle = \underline{u} \cdot \underline{v} = \underline{v}^* \underline{u}$

$\therefore \varphi_i \perp \tilde{\varphi}_j,\ i \neq j$

It is often useful to find the dual of $\{\varphi_i\}$ in terms of the basis elements. It follows that:

Theorem 1. *Given a basis $\{\varphi_i\}_{i\in I}$ for a Hilbert space $\mathbb{H}$, then dual set $\{\tilde{\varphi}_i\}_{i\in I} \in \mathbb{H}$ is given by the synthesis operator $\tilde{\Phi} = \Phi(\Phi^*\Phi)^{-1}$, where $G = \Phi^*\Phi$ is the Gram matrix.* $\therefore I = \Phi^* \Phi A$

Proof:

Assume an Hilbert space of dimension N

$$\tilde{\varphi}_k = \sum_{i=1}^{N} a_{k,i}\varphi_i, \qquad k = 1, 2, .., N. \tag{1}$$

In matrix vector form:

$$\tilde{\Phi} = \Phi A$$

$$\begin{cases} \tilde{\Phi} = \Phi A \\ I = \Phi^* \Phi A \end{cases} \Rightarrow \boxed{\tilde{\Phi} = \Phi(\Phi^*\Phi)^{-1}}$$

dual set from orthogonal set

Because of (1), we can write:

inner product

$$\langle \tilde{\varphi}_k, \varphi_j \rangle = \sum_{i=1}^{N} a_{k,i} \langle \varphi_i, \varphi_j \rangle \quad k = 1, 2, .., N.$$

$A = (\Phi^*\Phi)^{-1}$

In matrix vector form and using $\langle \tilde{\varphi}_k, \varphi_j \rangle = \delta_{k-j}$, we obtain: $I = (\Phi^*\Phi)A$ which then leads to:

$$\tilde{\Phi} = \Phi(\Phi^*\Phi)^{-1}$$

$\alpha = Ax$

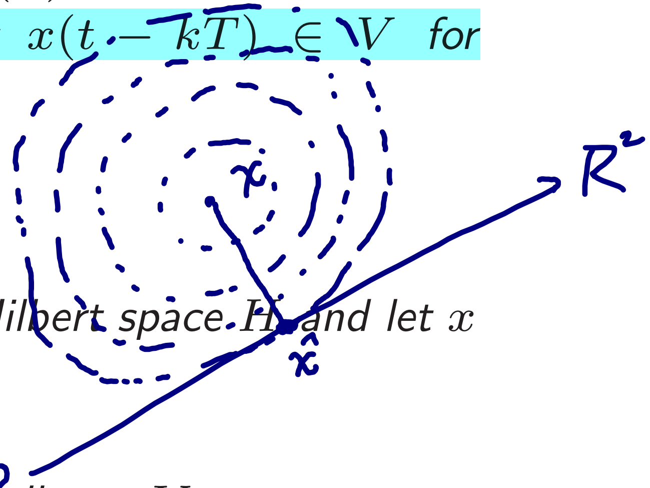
$\|\alpha\|^2 = \alpha^T\alpha = (Ax)^T(Ax) = x^T A^T A x \overset{0.}{=\!=} x^T x$

Mathematical Background

$A^TA \;\Big\{$ I if orthogonal; has a bound if biorthogonal

Definition 13. [Shift-Invariant Subspaces] *A subspace $V \in L_2(\mathbb{R})$ is a shift-invariant subspace with respect to shift $T \in \mathbb{R}^+$ when $x(t) \in V$ implies $x(t-kT) \in V$ for every integer k. In addition, $\varphi(t) \in V$ is called a generator of V when*

$$V = span(\{\varphi(t-nT)\}_{n\in\mathbb{Z}}).$$



Theorem 2. [Projection Theorem] *Let V be a closed subspace of Hilbert space H and let x be a vector in H.*

distance

1. *Existence: There exists $\hat{x} \in V$ such that $\|x-\hat{x}\| \leq \|x-v\|$ for all $v \in V$.*
2. *Orthogonality: $(x-\hat{x})\perp V$ is necessary and sufficient for determining $\hat{x}$.*
3. *Uniqueness: The vector $\hat{x}$ is unique.*
4. *Linearity: $\hat{x} = Px$ where P is a linear operator that depends on V and not on x.*
5. *Idempotency: $P(Px) = Px$ for all $x \in H$.*
6. *Self-adjointness: $P = P^*$.*

for orthogonal projection operator.

$\hat{x} = [3\ 5] \quad x = \begin{bmatrix} 2 & 5 \\ 4 & 5 \end{bmatrix}$

$\hat{x} = Px \quad P = \hat{x}x^{-1} = \hat{x}\dfrac{x^*}{|x|} = [3\ 5]\dfrac{1}{-2}\begin{bmatrix} 5 & -5 \\ -4 & 2 \end{bmatrix}$

$= -\frac{1}{2}[-5\ \ 1] = [\frac{5}{2}\ \ -\frac{1}{2}]$

Mathematical Background - Projections and Approximations

Definition 14. [Orthogonal Projection Operator, Oblique Projection Operator] *Consider the idempotent operator P. An idempotent operator P is such that $P^2 = P$.*

1. *A projection operator is a bounded linear operator that is idempotent.*
2. *An orthogonal projection operator is a projection operator that is selfadjoint.*
3. *An oblique projection operator is a projection operator that is not selfadjoint.*

Mathematical Background - Projections

Projections and Approximations
Consider a set of linearly independent functions $\{\varphi_i\}_{i=1,..,N}$ that covers the sub-space V and a function $f(t) \in L_2(\mathbb{R})$.
The orthogonal projection of $f(t)$ onto V is:

$$\hat{f}(t) = \sum_{i=1}^{N} \langle f(t), \tilde{\varphi}_i(t) \rangle \varphi_i(t).$$

Sketch of the proof:

For simplicity we consider a finite-dimensional vector spaces H of size M; $x \in H$ and we want to find its orthogonal projection in $V = span(\{\varphi_k\}_{k=1,2,...N})$ with $N < M$.

Orthogonal projecting operator

$\boxed{P = A(A^H A)^{-1} A^H}$

Projections and Approximations

x - dimension M

$\hat{x}$ - dimension N (projected onto space V (dim N))

We want:

$\langle x - Ac, \varphi_k \rangle = 0, \ k = 1 \cdots N.$ $\quad \langle x, \varphi_k \rangle = \langle Ac, \varphi_k \rangle$

A^H ($N \times M$): map M dimensions to N. $\quad A^H x = A^H A c$

$$\hat{x} = \sum_{k=1}^{N} c_k \varphi_k = [\ \varphi_1 \ \ \varphi_2 \ \ \ldots \ \ \varphi_N \ ] \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{bmatrix} = \underset{M \times N}{A} c, \tag{2}$$

$\because \hat{x} = Ac$

$\therefore A^H (x - Ac) = 0 \Rightarrow A^H x = A^H A c \Rightarrow c = (A^H A)^{-1} A^H x$

$\therefore \tilde{x} = A(A^H A^{-1}) A^H x$

such that $\|x - \hat{x}\|$ is minimized. The matrix A is of size $M \times N$ and its columns are the vectors φ_k, $k = 1, 2, \ldots N$. We need to determine the coefficients c_i. Using the projection theorem, we can use the orthogonality condition of the error vector which yields:

$$\langle x - Ac, \varphi_k \rangle = 0, \quad k = 1, 2, \ldots, N.$$

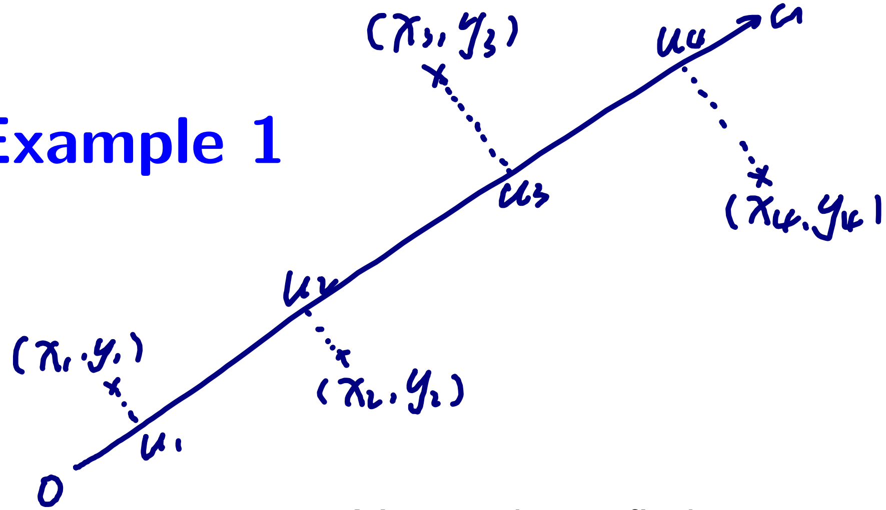
In matrix vector form this is equivalent to imposing

$$\underset{N \times M}{A^H} \underset{M \times N}{A} \underset{N \times 1}{c} = \underset{N \times M}{A^H} \underset{M \times 1}{x}, \tag{3}$$

where A^H is the conjugate transpose of A. By combining (2) and (3) we obtain

$$P = \underbrace{A(A^H A)^{-1}}_{\text{dual basis}} A^H.$$

Orthogonal Projection: Example 1



Line best fitting data:

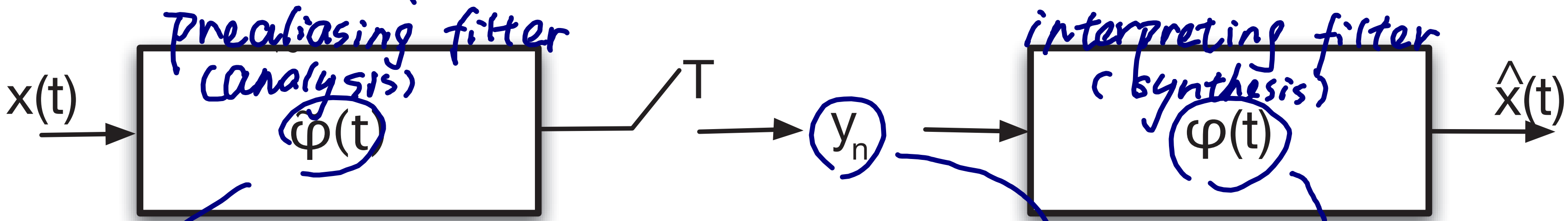
Assume that you are given the data points (x_i, y_i) for $i = 1, 2, \ldots N$. You wish to find the line $y = c_0 x + c_1$ that fits the data best, i.e., you wish to find the coefficients c_0, c_1 that minimize $\sum_{i=1}^{N} |y_i - c_0 x_i - c_1|^2$.
This is equivalent to finding the orthogonal projection of the N dimensional vector $Y = (y_1 \quad y_2 \quad \cdots \quad y_N)^T$ onto the subspace spanned by $X = (x_1 \quad x_2 \quad \cdots \quad x_N)^T$ and $U = (1 \quad 1 \quad \cdots \quad 1)^T$.
Using the orthogonality principle and the results of the previous slide we have that $c = (A^T A)^{-1} A^T Y$, where $c = (c_0 \quad c_1)^T$ and A is the $N \times 2$ matrix given by $A = (X \quad U)$.
Here we are taking only the transpose of A since the matrix is real.

Orthogonal Projection: Example 2

$\langle \mathrm{sinc}(\frac{t}{T}), \mathrm{sinc}(\frac{t}{T}-n)\rangle = T\delta[n]$ { orthogonal; generate shift-invariant subspace.

Shannon Sampling Theorem:

Consider the traditional sampling and reconstruction scheme shown below where we want to reconstruct signals bandlimited to $f_s = 1/2T$.



bandwidth limited → sinc function

piecewise constant → box function

$\tilde{\varphi}(t) = \dfrac{\sin(\frac{\pi t}{T})}{\frac{\pi t}{T}} = \mathrm{sinc}(\frac{t}{T})$

$y_n = \langle x(t), \mathrm{sinc}(\frac{t}{T}-n)\rangle$

$= \sum_n x(nT)\,\tilde{\varphi}(t-nT)$

$\tilde{\varphi}(t-n)$

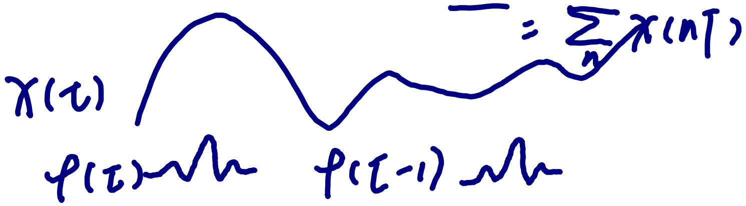
$\varphi(t) = \mathrm{sinc}(\frac{t}{T})$

$\hat{x}(t) = \langle y[n], \mathrm{sinc}(\frac{t}{T}-n)\rangle$

$= \sum_n y_n \delta(t-nT) * \varphi(t)$

$= \sum_n y_n \varphi(t-nT)$

(linear combination of shifted filter (basis))



$y_0 = \langle x(t), \tilde{\varphi}(\frac{t}{T})\rangle$

$y_1 = \langle x(t), \tilde{\varphi}(\frac{t}{T}-1)\rangle$

Orthogonal Projection: Example 2

Shannon Sampling Theorem (cont'd):
The analysis filter is the sinc function: $\text{sinc}_T(t) = \sin(\pi t/T)/\pi t/T$, therefore the samples are given by: $y_n = \langle x(t), \text{sinc}(t/T - n)\rangle$.
The synthesis filter is again the sinc function. Therefore the reconstructed signal is:

$$\hat{x}(t) = \frac{1}{T}\sum_n y_n \text{sinc}(t/T - n). \tag{4}$$

The set $\{\text{sinc}_T(t - nT)\}_{n\in\mathbb{Z}}$ is orthogonal and it generates the shift invariant subspace $V = \text{span}(\{\text{sinc}_T(t - nT)\}_{n\in\mathbb{Z}})$. From Eq. (4), we therefore realize that the reconstructed signals always belong to V. In other words the space of bandlimited functions (bandlimited to $f_s = 1/2T$) is shift invariant for shifts multiple of T.
Moreover the whole sampling and reconstruction process can be seen as an orthogonal projection onto the subspace $V = \text{span}(\{\text{sinc}_T(t - nT)\}_{n\in\mathbb{Z}})$. When $x(t)$ is bandlimited, then $x(t) \in V$ and the reconstruction is perfect. Otherwise $\hat{x}(t)$ is the closest bandlimited signal to $x(t)$.

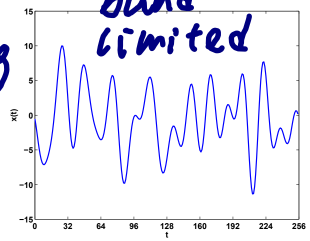
Classical sampling theorem

- A bandlimited signal $x(t)$ $BL_{[-\omega_M,\omega_M]}$ is perfectly recovered from samples $x(nT)$ taken T apart if the sampling rate $\omega_s = 2\pi/T$ is such that $\omega_s \geq 2\omega_M \Leftrightarrow T \leq \pi/\omega_M$.

Time domain

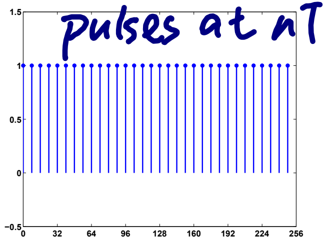
pre-aliasing filter $\tilde{\varphi}(t)$

band limited



$x(t)$

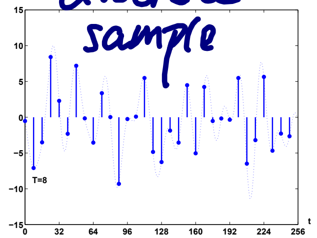
pulses at nT



$\times \sum_{n\in\mathbb{Z}} \delta(t-nT)$

sample

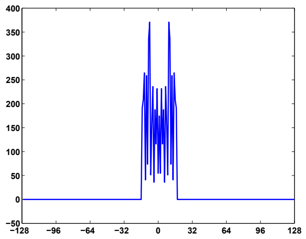
discrete sample



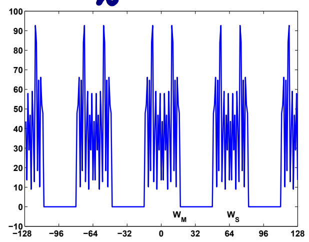
$x_s(t) = \sum_{n\in\mathbb{Z}} x(nT)\,\delta(t-nT)$ if no pre-aliasing

$= \sum_n x(nT)\,\tilde{\varphi}(t-nT)$ with pre-aliasing

Frequency domain



$X(\omega)$



$X_s(\omega)$

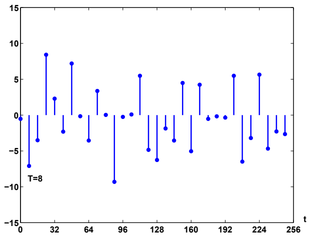
Reconstruction scheme (linear)

basis design

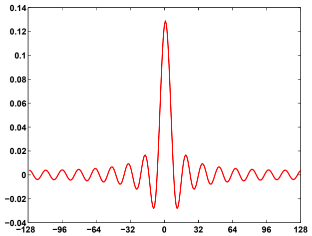
orthogonal signals can do. (box. tri...)

use o. basis to recover.

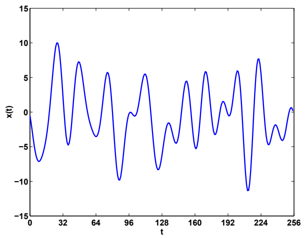
- Time domain



$x_n = x(nT) = < x(t), \delta(t - nT) >$

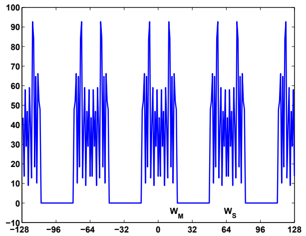


$*h(t)(= \mathrm{sinc})$

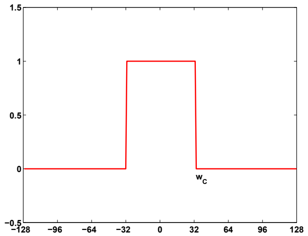


$x_{rec}(t) = \sum_n x_n \,\mathrm{sinc}(\frac{t}{T} - n)$

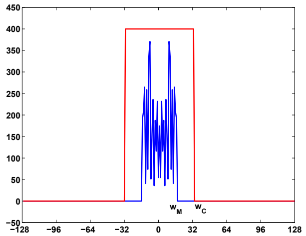
- Frequency domain



$X_s(\omega)$

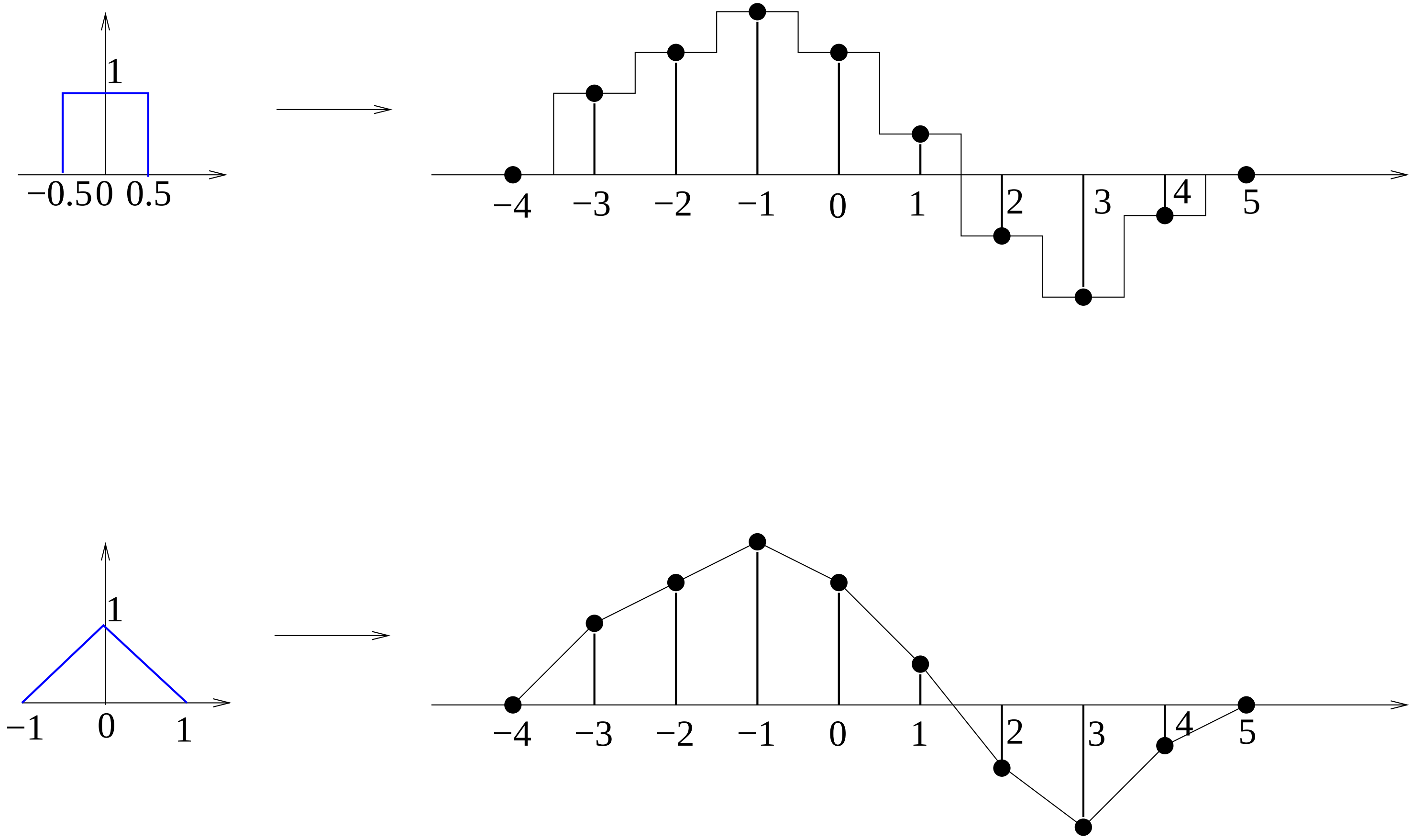


$\times H(\omega)(= \text{Rect})$



$X(\omega)$

Sampling as Projecting into Shift-Invariant Sub-Spaces



Fourier Theory

Given a function $f(t) \in L_2(\mathbb{R})$ the Fourier transform is defined by $\hat{f}(\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t}dt$.
The inverse Fourier transform is given by $f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{f}(\omega)e^{j\omega t}d\omega$.
The moment theorem states that

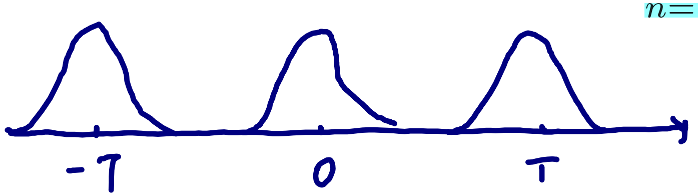
$$(-j)^n m_n = \left.\frac{\partial^n \hat{f}(\omega)}{\partial \omega^n}\right|_{\omega=0}$$

where $m_n = \int_{-\infty}^{\infty} t^n f(t)dt, \quad n = 0, 1, 2, \ldots$
Assume a function $f(t)$ with sufficient smoothness and decay, the Poisson summation formula states that

$$\sum_{n=-\infty}^{\infty} f(t - nT) = \frac{1}{T}\sum_{k=-\infty}^{\infty} \hat{f}\left(\frac{2\pi k}{T}\right)e^{j2\pi kt/T}.$$

For $T = 1$ and $t = 0$, we have that

$$\sum_{n=-\infty}^{\infty} f(n) = \sum_{k=-\infty}^{\infty} \hat{f}(2\pi k).$$



Sum of function at all integer points = sum of its Fourier transform at all integer points.

z-transform

Given a discrete-time signal $f[n]$ the z-transform is defined as

$$F(z) = \sum_{n \in \mathbb{Z}} f[n] z^{-n} \quad z \in \text{ ROC} .$$

Notice that the z-transform is equal to the discrete-time Fourier transform for $z = e^{j\omega}$.
useful properties of the z-transform:

sequence (vector) f:
f^T- time reverse.

- Delay

$$f[n - n_0] \longleftrightarrow z^{-n_0} F(z)$$

- Time-reversal

$$f^T[n] = f[-n] \longleftrightarrow F(z^{-1})$$

- Modulation

$$(-1)^n f[n] \longleftrightarrow F(-z)$$

- Up-sampling and down-sampling.

$x[n] \to \uparrow 2 \to y[n]$

Upsampling by 2

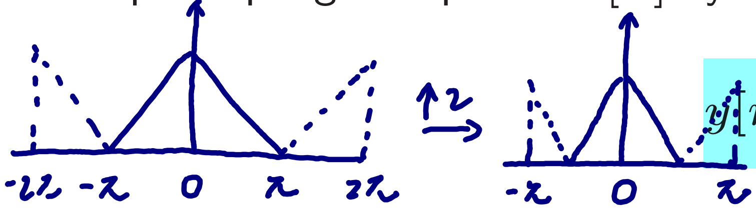
expand in time

↕

shrink in frequency

	z^0	z^{-1}	z^{-2}	z^{-3}	z^{-4}	
x -	$x[0]$	$x[1]$	$x[2]$	$x[3]$	$x[4]$	..
y -	$x[0]$	0	$x[1]$	0	$x[2]$	..

Up-sampling a sequence $x[n]$ by 2 results in a new sequence $y[n]$ given by



$$y[n] = \begin{cases} 0 & n \text{ odd} \\ x[l] & n = 2l \text{ even} \end{cases}$$

$$Y(z) = \sum_n y[n] z^{-n} = \sum_l x[l] z^{-2l} \overset{u=z^2}{=} \sum_l x[l] u^{-l} = X(u) = X(z^2)$$

In matrix notation:

U: 'thin' matrix

$$y = U_2 x = \begin{bmatrix} \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & 1 & 0 & 0 & \cdots \\ \cdots & 0 & 0 & 0 & \cdots \\ \cdots & 0 & 1 & 0 & \cdots \\ \cdots & 0 & 0 & 0 & \cdots \\ \cdots & 0 & 0 & 1 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix} \begin{bmatrix} \cdots \\ x[0] \\ x[1] \\ x[2] \\ \cdots \end{bmatrix}$$

In the frequency domain

$$Y(z) = X(z^2) \quad \text{or} \quad Y(e^{j\omega}) = X(e^{j2\omega}).$$

Upsampling by 2 : (add redundent 0s)

$Y(z) = X(z^2)$. $Y(e^{j\omega}) = X(e^{j2\omega})$

$x[n] \longrightarrow (\downarrow 2) \longrightarrow y[n]$

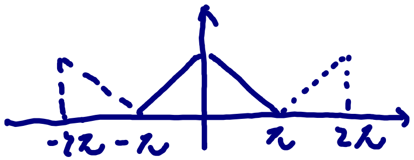
shrink in time

↕

expand in frequency

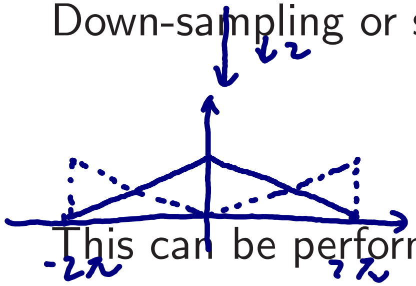
Down-sampling by 2

$$\begin{array}{llllll} & z^{-0} & z^{-1} & z^{-2} & & \\ x - & x[0] & x[1] & x[2] & x[3] & x[4] \cdots \\ y - & x[0] & x[2] & x[4] & \cdots & \end{array}$$



Down-sampling or sub-sampling a sequence $x[n]$ by 2 results in a new sequence $y[n]$ given by

$$y[n] = x[2n].$$



$$\begin{aligned} Y(z) &= \sum_n y[n] z^{-n} \\ &= \sum_n x[2n] z^{-n} \\ &= \sum_n x[n] z^{-\frac{n}{2}} \cdot \frac{1}{2}[1 + (-1)^n] \\ &= \frac{1}{2}\Big(\sum_n x[n] z^{-\frac{n}{2}} + \sum_n (-1)^n x[n] z^{-\frac{n}{2}}\Big) \\ &= \frac{1}{2}\Big(X(z^{\frac{1}{2}}) + X(-z^{\frac{1}{2}})\Big) \end{aligned}$$

This can be performed using the sub-sampling matrix $D_2 = U_2^T$:

D: 'fat' matrix

$D_2 = U_2^T$

$$y = D_2 x = \begin{bmatrix} \cdots & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdots \\ \cdots & 1 & 0 & 0 & 0 & 0 & 0 & \cdots \\ \cdots & 0 & 0 & 1 & 0 & 0 & 0 & \cdots \\ \cdots & 0 & 0 & 0 & 0 & 1 & 0 & \cdots \\ \cdots & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdots \end{bmatrix} \begin{bmatrix} \cdots \\ x[0] \\ x[1] \\ x[2] \\ \cdots \end{bmatrix}$$

In the frequency domain we have that

$$Y(z) = \frac{1}{2}[X(z^{1/2}) + X(-z^{1/2})] \quad \text{or} \quad Y(e^{j\omega}) = \frac{1}{2}[X(e^{j\omega/2}) + X(e^{j(\omega-2\pi)/2})].$$

Downsampling by 2: (drop some elements)

$$Y(z) = \frac{1}{2}[X(z^{\frac{1}{2}}) + X(-z^{\frac{1}{2}})]. \quad Y(e^{jw}) = \frac{1}{2}[X(e^{\frac{jw}{2}}) + X(e^{-\frac{jw}{2}})]$$

$$\begin{cases} D_2U_2 = I & \text{up - down: no info loss} \\ U_2D_2 \neq I & \text{down-up: info loss (replaced by 0)} \end{cases}$$

Filtering before downsampling

$$D_2 = \begin{bmatrix} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & 1 & 0 & 0 & 0 & 0 & 0 & \cdot \\ \cdot & 0 & 0 & 1 & 0 & 0 & 0 & \cdot \\ \cdot & 0 & 0 & 0 & 0 & 1 & 0 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}$$

$$D_2 \cdot H = \begin{bmatrix} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & h[2] & h[1] & h[0] & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & h[4] & h[3] & h[2] & h[1] & h[0] & \cdot & \cdot & \cdot \\ \cdot & h[6] & h[5] & h[4] & h[3] & h[2] & h[1] & h[0] & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}$$

down-sample filter

$x[n] \to \boxed{h[n]} \to \downarrow 2 \to y[n]$

$$H = \begin{bmatrix} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & h[2] & h[1] & h[0] & \cdot & \cdot & \cdot & \cdot \\ \cdot & h[3] & h[2] & h[1] & h[0] & \cdot & \cdot & \cdot \\ \cdot & h[4] & h[3] & h[2] & h[1] & h[0] & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}$$

D_2: drag →

- $$H = \begin{bmatrix} \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & h[2] & h[1] & h[0] & 0 & 0 & \cdots \\ \cdots & h[3] & h[2] & h[1] & h[0] & 0 & \cdots \\ \cdots & h[4] & h[3] & h[2] & h[1] & h[0] & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$ Toeplitz

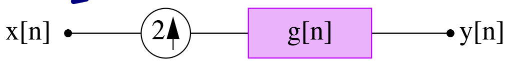
- $y = D_2 \cdot H \cdot x$ or $y[n] = \sum_k h[2n-k]x[k] = \langle h[2n-k], x[k] \rangle_k$

$$y = \begin{bmatrix} \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & h[2] & h[1] & h[0] & 0 & 0 & 0 & 0 & \cdots \\ \cdots & h[4] & h[3] & h[2] & h[1] & h[0] & 0 & 0 & \cdots \\ \cdots & h[6] & h[5] & h[4] & h[3] & h[2] & h[1] & h[0] & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix} x$$

Filtering after upsampling

$$(G\cdot U_2)^T = U_2^T G^T = D_2 \cdot H = \begin{bmatrix} \cdot & h[0] & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & h[2] & h[1] & h[0] & \cdot & \cdot & \cdot \\ \cdot & h[4] & h[3] & h[2] & h[1] & h[0] & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}$$

$$G = \begin{bmatrix} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & g[2] & g[1] & g[0] & \cdot & \cdot & \cdot \\ \cdot & g[3] & g[2] & g[1] & g[0] & \cdot & \cdot \\ \cdot & g[4] & g[3] & g[2] & g[1] & g[0] & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}$$



transpose

$$U_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$G \cdot U_2 = \begin{bmatrix} \cdot & g[0] & \cdot & \cdot & \cdot & \cdot \\ \cdot & g[1] & \cdot & \cdot & \cdot & \cdot \\ \cdot & g[2] & g[0] & \cdot & \cdot & \cdot \\ \cdot & g[3] & g[1] & \cdot & \cdot & \cdot \\ \cdot & g[4] & g[2] & g[0] & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}$$

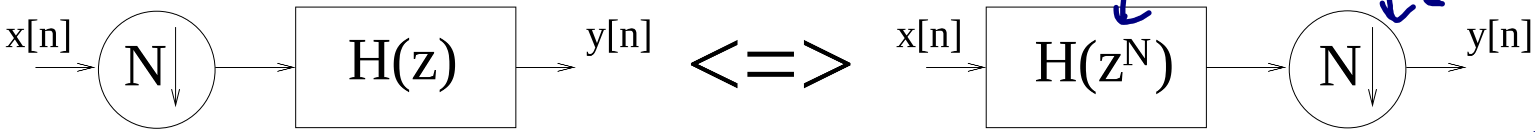
U_2: drag $\downarrow$

- $$G = \begin{bmatrix} \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & g[2] & g[1] & g[0] & 0 & 0 & \dots \\ \dots & g[3] & g[2] & g[1] & g[0] & 0 & \dots \\ \dots & g[4] & g[3] & g[2] & g[1] & g[0] & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix}$$ Toeplitz
- $y = G \cdot U_2 \cdot x$ or $y[n] = \sum_k g[n-2k]x[k] = \langle g[n-2k], x[k] \rangle_k$

$$y = \begin{bmatrix} \dots & \dots & \dots & \dots & \dots \\ \dots & g[0] & 0 & 0 & \dots \\ \dots & g[1] & 0 & 0 & \dots \\ \dots & g[2] & g[0] & 0 & \dots \\ \dots & g[3] & g[1] & 0 & \dots \\ \dots & g[4] & g[2] & g[0] & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix} x$$

Noble Identities

$Y(z) = \frac{1}{N}\sum_{k=0}^{N-1} X(W_N^k z^{\frac{1}{N}})\, H(z) =$ $Y(z) = \frac{1}{N}\sum_{k=0}^{N-1} X(\underline{W_N^k} z^{\frac{1}{N}})\, H(\underline{(W_N^k z^{\frac{1}{N}})^N})$

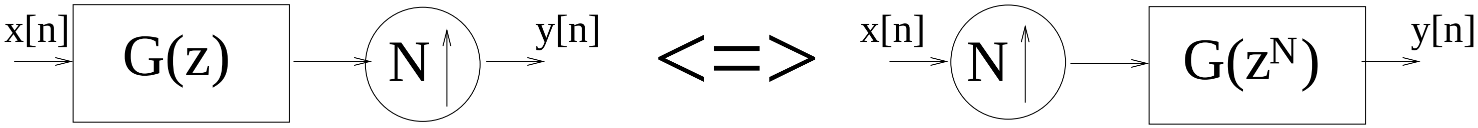


$\Rightarrow H(z^N)$

The first Noble Identity

$(W_N^k)^N = (e^{-j\frac{2\pi}{N}k})^N = 1$

$G(z^N)$ u



(no info loss)

The second Noble identity

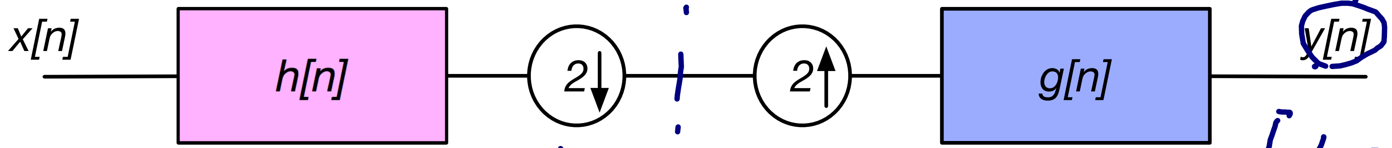
$Y(z) = \frac{1}{N}\sum_{k=0}^{N-1} X(W_N^k z^N)\, G(W_N^k z^N) = Y(z) = \frac{1}{N}\sum_{k=0}^{N-1} X(W_N^k z^N)\, G(z^N)$

Putting all together

$z[n] = \langle x[k], h[k-2n] \rangle$

$D \cdot U = I$

$H \cdot G = I$



projection onto $V = \text{span}\{g[n-2k]\}_{k\in\mathbb{Z}}$

$$H = \begin{bmatrix} \ddots & & & & \\ \cdots & h[2] & h[1] & h[0] & \\ & & h[2] & h[1] & h[0] & \cdots \\ & & & & \ddots \end{bmatrix} = \begin{bmatrix} \ddots & & & & \\ \cdots & g[0] & g[1] & g[2] & \\ & & g[0] & g[1] & g[2] & \cdots \\ & & & & \ddots \end{bmatrix}$$

designed!

where $\langle g[n], g[n-2k] \rangle = \delta_k$ and $h[n] = g[-n]$. Note that:

orthogonal

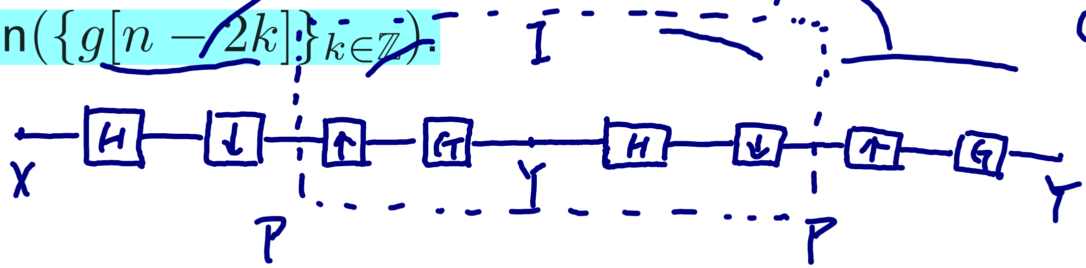
$$D_2 G^T \cdot G U_2 = I, \quad \text{and} \quad H = G^T.$$

Claim: The operator P mapping $x \to y$ is an orthogonal projection onto the subspace $V = \text{span}(\{g[n-2k]\}_{k\in\mathbb{Z}})$.

$$G = \begin{bmatrix} \ddots & & \\ g[0] & & \\ g[1] & g[0] & \cdots \\ g[2] & g[1] & \\ & g[2] & \\ & & \ddots \end{bmatrix}$$

Proof:

$P = GU_2D_2H$



$$P^2 = (GU_2D_2H)(GU_2D_2H) = GU_2ID_2H = P$$

$\because \langle g[n], g[n-2k] \rangle = \delta_k$

$$P^T = (GU_2D_2H)^T = G(U_2D_2)^T H = GU_2D_2H = P$$

$\because HG = I$.

$\langle g[n], g[n-2k]\rangle = \langle g[n], g[n-k]\rangle$ downsampling by 2.

$P[k] = \langle g[n], g[n-k]\rangle = \sum_n g[n]g[n-k]$

Appendix

$n-k=l$

$P(z) = \sum_k P[k]z^{-k} = \sum_k\sum_n g[n]g[n-k]z^{-k} = \sum_n g[n]\sum_k g[\overbrace{n-k}^{l}]z^{-k} = \sum_n g[n]\sum_l g[l]z^{l-n}$

$= \sum_n g[n]z^{-n}\sum_l g[l]z^{l} = G(z)G(z^{-1})$

$P'(z) =$

Claim: The equality $\langle g[n], g[n-2k]\rangle = \delta_k$ can be written in the z domain as follows:

$P(z) \xrightarrow{\downarrow 2} \frac{1}{2}[P(z^{\frac{1}{2}}) + P(-z^{\frac{1}{2}})]$

$\langle g[n], g[n-2k]\rangle = \delta_k$

$$G(z)G(z^{-1}) + G(-z)G(-z^{-1}) = 2.$$

$\therefore P'(z) = 1$ (z-transform)

$= \frac{1}{2}[G(z^{\frac{1}{2}})G(z^{-\frac{1}{2}}) + G(-z^{\frac{1}{2}})G(-z^{-\frac{1}{2}})]$

Proof: replace sum that. $\frac{1}{2}[G(z)G(z^{-1}) + G(-z)G(-z^{-1})]$

$\therefore G(z)G(z^{-1}) + G(-z)G(-z^{-1}) = 2$

Define the sequence $p[k] = \langle g[n], g[n-k]\rangle = \sum_n g[n]g[n-k]$. Then $\langle g[n], g[n-2k]\rangle$ is obtained by downsampling by 2 the sequence $p[k]$.

The z-transform of $p[k]$ is given by:

$$P(z) = \sum_k p[k]z^{-k} = \sum_k\sum_n g[n]g[n-k]z^{-k} = G(z)g(z^{-1}).$$

Applying the downsampling by two rule on $P(z)$ yields the desired equality.