Teaching Image Processing with Geometry

Vivek Goyal and Jelena Kovačević

Based upon textbooks with Martin Vetterli
Includes slides by Andrea Ridolfi and Amina Chebira

September 30, 2012
Outline

1. Introduction
   - Why a tutorial on teaching?
   - Overview

2. Basic Principles in Teaching SP with Geometry
   - Unified view
   - Hilbert space tools—Part I: Basics through projections
   - Basic results I: Best approximation
   - Hilbert space tools—Part II: Bases through discrete-time systems

3. Example Lecture: Sampling and Interpolation
   - Motivation
   - Classical View and Historical Notes
   - Operator View
   - Basis Expansion View
   - Extensions
   - Summary

4. Teaching Materials
   - Textbooks
   - Supplementary materials

5. Wrap up
Teaching about teaching?

Teaching should not remain static

- Applications change (and we should want them to)
  - Signal processing thinking should be applied broadly
  - Global Fourier techniques relatively less important than in the past

- Computing platforms change (and we should want them to)
  - Classical DSP architectures relatively less important than in the past
  - Likely to use high-level programming languages

- Students change (and we should want them to)
  - Different base of knowledge
    - Biology, economics, social sciences...

- Eternal challenge of the educator
  - Knowledge grows, time in school does not
  - Must be willing to cull details to convey big picture
  - Should teach what is most reusable and generalizable
Teaching about teaching?

Teaching should not remain static

- Applications change (and we should want them to)
  - Signal processing thinking should be applied broadly
  - Global Fourier techniques relatively less important than in the past

- Computing platforms change (and we should want them to)
  - Classical DSP architectures relative less important than in the past
  - Likely to use high-level programming languages

- Students change (and we should want them to)
  - Different base of knowledge
  - Biology, economics, social sciences, . . .

- Eternal challenge of the educator
  - Knowledge grows, time in school does not
  - Must be willing to call details to convey big picture
  - Should teach what is most reusable and generalizable
Teaching about teaching?

Teaching should not remain static

- Applications change (and we should want them to)
  - Signal processing thinking should be applied broadly
  - Global Fourier techniques relatively less important than in the past

- Computing platforms change (and we should want them to)
  - Classical DSP architectures relative less important than in the past
  - Likely to use high-level programming languages

- Students change (and we should want them to)
  - Different base of knowledge
  - Biology, economics, social sciences, . . .

- Eternal challenge of the educator
  - Knowledge grows, time in school does not
  - Must be willing to call details to convey big picture
  - Should teach what is most reusable and generalizable
Teaching about teaching?

Teaching should not remain static

- Applications change (and we should want them to)
  - Signal processing thinking should be applied broadly
  - Global Fourier techniques relatively less important than in the past

- Computing platforms change (and we should want them to)
  - Classical DSP architectures relative less important than in the past
  - Likely to use high-level programming languages

- Students change (and we should want them to)
  - Different base of knowledge
  - Biology, economics, social sciences, . . .

- Eternal challenge of the educator
  - Knowledge grows, time in school does not
  - Must be willing to cull details to convey big picture
  - Should teach what is most reusable and generalizable
Overview

Goals of the tutorial:
- See that geometric notions unify (simplify!) signal processing
- Learn/review basics of Hilbert space view
- See Hilbert space view in action
- Learn about textbooks *Foundations of Signal Processing* and *Fourier and Wavelet Signal Processing*

Structure of the tutorial:
- Developing unified view of signal processing
- Hilbert space tools—Part I: Basics through projections
  - A few key results (best approximation and the projection theorem)
- Hilbert space tools—Part II: Bases through discrete-time systems
- Example lecture: Sampling made easy
- Textbooks and related materials
Overview

Goals of the tutorial:

- See that geometric notions unify (simplify!) signal processing
- Learn/review basics of Hilbert space view
- See Hilbert space view in action
- Learn about textbooks *Foundations of Signal Processing* and *Fourier and Wavelet Signal Processing*

Structure of the tutorial:

- Developing unified view of signal processing
- Hilbert space tools—Part I: Basics through projections
  - A few key results (best approximation and the projection theorem)
- Hilbert space tools—Part II: Bases through discrete-time systems
- Example lecture: Sampling made easy
- Textbooks and related materials
Introduction

Overview

Goals of the tutorial:
- See that geometric notions unify (simplify!) signal processing
- Learn/review basics of Hilbert space view
- See Hilbert space view in action
- Learn about textbooks *Foundations of Signal Processing* and *Fourier and Wavelet Signal Processing*

Structure of the tutorial:
- Developing unified view of signal processing
- Hilbert space tools—Part I: Basics through projections
  - A few key results (best approximation and the projection theorem)
- Hilbert space tools—Part II: Bases through discrete-time systems
- Example lecture: Sampling made easy
- Textbooks and related materials
Overview

Goals of the tutorial:

- See that geometric notions unify (simplify!) signal processing
- Learn/review basics of Hilbert space view
- See Hilbert space view in action
- Learn about textbooks *Foundations of Signal Processing* and *Fourier and Wavelet Signal Processing*

Structure of the tutorial:

- Developing unified view of signal processing
- Hilbert space tools—Part I: Basics through projections
  - A few key results (best approximation and the projection theorem)
- Hilbert space tools—Part II: Bases through discrete-time systems
- Example lecture: Sampling made easy
- Textbooks and related materials
Unifying principles

Signal processing has various dichotomies
- continuous time vs. discrete time
- infinite intervals vs. finite intervals
- periodic vs. aperiodic
- deterministic vs. stochastic

Each can be placed in a common framework featuring geometry

Example payoffs:
- Unified understanding of best approximation (projection theorem)
- Unified understanding of Fourier domains
- Unified understanding of signal expansions (including sampling)
Unifying principles

Signal processing has various dichotomies
- continuous time vs. discrete time
- infinite intervals vs. finite intervals
- periodic vs. aperiodic
- deterministic vs. stochastic

Each can placed in a common framework featuring geometry

Example payoffs:
- Unified understanding of best approximation (projection theorem)
- Unified understanding of Fourier domains
- Unified understanding of signal expansions (including sampling)
Unifying principles

Signal processing has various dichotomies
- continuous time vs. discrete time
- infinite intervals vs. finite intervals
- periodic vs. aperiodic
- deterministic vs. stochastic

Each can be placed in a common framework featuring geometry

Example payoffs:
- Unified understanding of best approximation (projection theorem)
- Unified understanding of Fourier domains
- Unified understanding of signal expansions (including sampling)
Unifying principles

Signal processing has various dichotomies
- continuous time vs. discrete time
- infinite intervals vs. finite intervals
- periodic vs. aperiodic
- deterministic vs. stochastic

Each can placed in a common framework featuring geometry

Example payoffs:
- Unified understanding of best approximation (projection theorem)
- Unified understanding of Fourier domains
- Unified understanding of signal expansions (including sampling)
Unifying framework: Hilbert spaces

Examples of Hilbert spaces:
- finite-dimensional vectors (basic linear algebra)
- sequences on \{-1, 0, 1, \ldots\} (discrete-time signals)
- sequences on \{0, 1, \ldots, N-1\} (N-periodic discrete-time signals)
- functions on \(-\infty, \infty\) (continuous-time signals)
- functions on \([0, T]\) (T-periodic continuous-time signals)
- scalar random variables

More abstraction. More mathematics. More difficult?
- With framework in place, can go farther, faster
- Leverage "real world" geometric intuition
Unifying framework: Hilbert spaces

Examples of Hilbert spaces:

- finite-dimensional vectors (basic linear algebra)
- sequences on \{\ldots, -1, 0, 1, \ldots\} (discrete-time signals)
- sequences on \{0, 1, \ldots, N - 1\} (N-periodic discrete-time signals)
- functions on \((\infty, \infty)\) (continuous-time signals)
- functions on \([0, T]\) (T-periodic continuous-time signals)
- scalar random variables

More abstraction. More mathematics. More difficult?

- With framework in place, can go farther, faster
- Leverage “real world” geometric intuition
Mathematical rigor

*Everything should be made as simple as possible, but no simpler.*

– Common paraphrasing of Albert Einstein

*Make everything as simple as possible without being wrong.*

– Our variant for teaching

- Correct intuitions are separate from functional analysis details
- Teach the difference among
  - rigorously true, with elementary justification
  - rigorously true, justification not elementary (e.g., Poisson sum formula)
  - convenient and related to rigorous statements (e.g., uses of Dirac delta)

… if whether an airplane would fly or not depended on whether some function … was Lebesgue but not Riemann integrable, then I would not fly in it.

– Richard W. Hamming
Mathematical rigor

*Everything should be made as simple as possible, but no simpler.*

– Common paraphrasing of Albert Einstein

*Make everything as simple as possible without being wrong.*

– Our variant for teaching

- Correct intuitions are separate from functional analysis details
- Teach the difference among
  - rigorously true, with elementary justification
  - rigorously true, justification not elementary (e.g., Poisson sum formula)
  - convenient and related to rigorous statements (e.g., uses of Dirac delta)

...if whether an airplane would fly or not depended on whether some function ... was Lebesgue but not Riemann integrable, then I would not fly in it.

– Richard W. Hamming
Vector spaces

A vector space generalizes easily beyond the $\mathbb{R}^2$ Euclidean plane

**Axioms**

- A vector space over a field of scalars $\mathbb{C}$ (or $\mathbb{R}$) is a set of vectors $V$ together with operations
  - vector addition: $V \times V \rightarrow V$
  - scalar multiplication: $\mathbb{C} \times V \rightarrow V$

that satisfy the following axioms:

1. $x + y = y + x$
2. $(x + y) + z = x + (y + z)$
3. $\exists 0 \in V $ s.t. $x + 0 = x$ for all $x \in V$
4. $\alpha(x + y) = \alpha x + \alpha y$
5. $(\alpha + \beta)x = \alpha x + \beta x$
6. $(\alpha \beta)x = \alpha(\beta x)$
7. $0x = 0$ and $1x = x$
Basic Principles in Teaching SP and Geometry

Hilbert space tools—Part I: Basics through projections

Vector spaces

Examples

- \( \mathbb{C}^N \): complex (column) vectors of length \( N \)
- \( \mathbb{C}^\mathbb{Z} \): sequences – discrete-time signals
  - (write as infinite column vector)
- \( \mathbb{C}^\mathbb{R} \): functions – continuous-time signals
- polynomials of degree at most \( K \)
- scalar random variables
- discrete-time stochastic processes
Vector spaces

Key notions

- **Subspace**
  - $S \subseteq V$ is a subspace when it is closed under vector addition and scalar multiplication:
    - For all $x, y \in S$, $x + y \in S$
    - For all $x \in S$ and $\alpha \in \mathbb{C}$, $\alpha x \in S$

- **Span**
  - $S$: set of vectors (could be infinite)
  - $\text{span}(S)$: set of all finite linear combinations of vectors in $S$:
    \[
    \text{span}(S) = \left\{ \sum_{k=1}^{N} n_k x_k \mid n_k \in \mathbb{C}, x_k \in S \text{ and } N \in \mathbb{N} \right\}
    \]
  - $\text{span}(S)$ is always a subspace
Vector spaces

Key notions

- **Subspace**
  - $S \subseteq V$ is a subspace when it is closed under vector addition and scalar multiplication:
    - For all $x, y \in S$, $x + y \in S$
    - For all $x \in S$ and $\alpha \in \mathbb{C}$, $\alpha x \in S$

- **Span**
  - $S$: set of vectors (could be infinite)
  - $\text{span}(S) = \text{set of all finite linear combinations of vectors in } S$:
    \[
    S = \left\{ \sum_{k=0}^{N-1} \alpha_k \varphi_k \mid \alpha_k \in \mathbb{C}, \varphi_k \in S \text{ and } N \in \mathbb{N} \right\}
    \]
  - $\text{span}(S)$ is always a subspace
Vector spaces

Key notions

- Linear independence
  - \( S = \{\varphi_k\}_{k=0}^{N-1} \) is linearly independent when:
    \[
    \sum_{k=0}^{N-1} \alpha_k \varphi_k = 0 \quad \text{only when} \quad \alpha_k = 0 \quad \text{for all} \quad k
    \]
  - If \( S \) is infinite, we need every finite subset to be linearly independent

- Dimension
  - \( \dim(V) = N \) if \( V \) contains a linearly independent set with \( N \) vectors and every set with \( N + 1 \) or more vectors is linearly dependent
  - \( V \) is infinite dimensional if no such finite \( N \) exists
Vector spaces

**Key notions**

- **Linear independence**
  - \( S = \{ \varphi_k \}_{k=0}^{N-1} \) is linearly independent when:
  \[
  \sum_{k=0}^{N-1} \alpha_k \varphi_k = 0 \text{ only when } \alpha_k = 0 \text{ for all } k
  \]
  - If \( S \) is infinite, we need every finite subset to be linearly independent

- **Dimension**
  - \( \dim(V) = N \) if \( V \) contains a linearly independent set with \( N \) vectors and every set with \( N + 1 \) or more vectors is linearly dependent
  - \( V \) is infinite dimensional if no such finite \( N \) exists
Inner products

Inner products generalize angles (especially right angles) and orientation

**Definition (Inner product)**

- An inner product on vector space $V$ is a function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}$ satisfying:
  - Distributivity: $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$
  - Linearity in the first argument: $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$
  - Hermitian symmetry: $\langle x, y \rangle^* = \langle y, x \rangle$
  - Positive definiteness: $\langle x, x \rangle \geq 0$ and $\langle x, x \rangle = 0$ if and only if $x = 0$

- Note: $\langle x, \alpha y \rangle = \alpha^* \langle x, y \rangle$
Inner products

Examples

- On $\mathbb{C}^N$: $\langle x, y \rangle = \sum_{n=0}^{N-1} x_n y_n^* = y^* x$ (writing $x$ and $y$ as column vectors)

- On $\mathbb{C}^\mathbb{Z}$: $\langle x, y \rangle = \sum_{n \in \mathbb{Z}} x_n y_n^* = y^* x$ (writing $x$ and $y$ as column vectors)

- On $\mathbb{C}^\mathbb{R}$: $\langle x, y \rangle = \int_{-\infty}^{\infty} x(t)y^*(t) \, dt$

- On $\mathbb{C}$-valued random variables: $\langle x, y \rangle = E[xy^*]$
Geometry in inner product spaces

Drawn in $\mathbb{R}^2$ and true in general:

\[
\langle x, y \rangle = x_1 y_1 + x_2 y_2 \\
= \|x\| \|y\| \cos \alpha \\
= \text{product of 2-norms times the cos of the angle between the vectors}
\]

\[
\langle x, e_1 \rangle = x_1 = \|x\| \cos \alpha
\]
Geometry in inner product spaces

Drawn in $\mathbb{R}^2$ and true in general:

1. $\langle x, y \rangle = x_1y_1 + x_2y_2$
   $= \|x\| \|y\| \cos \alpha$
   $= \text{product of 2-norms times the cosine of the angle between the vectors}$

2. $\langle x, e_1 \rangle = x_1 = \|x\| \cos \alpha_x$
Orthogonality

Let \( S = \{ \varphi_i \}_{i \in I} \) be a set of vectors

**Definition (Orthogonality)**

- \( x \) and \( y \) are orthogonal when \( \langle x, y \rangle = 0 \) written \( x \perp y \)
- \( S \) is orthogonal when for all \( x, y \in S, x \neq y \) we have \( x \perp y \)
- \( S \) is orthonormal when it is orthogonal and for all \( x \in S, \langle x, x \rangle = 1 \)
- \( x \) is orthogonal to \( S \) when \( x \perp s \) for all \( s \in S \), written \( x \perp S \)
- \( S_0 \) and \( S_1 \) are orthogonal when every \( s_0 \in S_0 \) is orthogonal to \( S_1 \), written \( S_0 \perp S_1 \)

Right angles (perpendicularity) extends beyond Euclidean geometry
Norms generalize length in ordinary Euclidean space

**Definition (Norm)**

- A norm on $V$ is a function $\|\cdot\| : V \to \mathbb{R}$ satisfying
  - Positive definiteness: $\|x\| \geq 0$ and $\|x\| = 0$ if and only if $x = 0$
  - Positive scalability: $\|\alpha x\| = |\alpha| \|x\|
  - Triangle inequality: $\|x + y\| \leq \|x\| + \|y\|$ with equality if and only if $y = \alpha x$

- Any inner product induces a norm
  $$\|x\| = \sqrt{\langle x, x \rangle}$$

- Not all norms are induced by an inner product
Norm

Norms generalize length in ordinary Euclidean space

**Definition (Norm)**

- A norm on $V$ is a function $\| \cdot \| : V \rightarrow \mathbb{R}$ satisfying
  - Positive definiteness: $\| x \| \geq 0$ and $\| x \| = 0$ if and only if $x = 0$
  - Positive scalability: $\| \alpha x \| = |\alpha| \| x \|$
  - Triangle inequality: $\| x + y \| \leq \| x \| + \| y \|$ with equality if and only if $y = \alpha x$

- Any inner product induces a norm
  $$\| x \| = \sqrt{\langle x, x \rangle}$$

- Not all norms are induced by an inner product
Norms induced by inner products

Any inner product induces a norm: $\|x\| = \sqrt{\langle x, x \rangle}$

**Examples**

- On $\mathbb{C}^N$: $\|x\| = \sqrt{\langle x, x \rangle} = \left( \sum_{n=0}^{N-1} |x_n|^2 \right)^{1/2}$

- On $\mathbb{C}^\mathbb{Z}$: $\|x\| = \sqrt{\langle x, x \rangle} = \left( \sum_{n \in \mathbb{Z}} |x_n|^2 \right)^{1/2}$

- On $\mathbb{C}^\mathbb{R}$: $\|x\| = \sqrt{\langle x, x \rangle} = \left( \int_{-\infty}^{\infty} |x(t)|^2 \, dt \right)^{1/2}$

- On $\mathbb{C}$-valued random variables: $\|x\| = \sqrt{\langle x, x \rangle} = \left( \mathbb{E}[|x|^2] \right)^{1/2}$
Norms induced by inner products

Properties

- **Pythagorean theorem**
  - $x \perp y \Rightarrow \|x + y\|^2 = \|x\|^2 + \|y\|^2$
  - $\{x_k\}_{k \in K}$ orthogonal $\Rightarrow \left\| \sum_{k \in K} x_k \right\|^2 = \sum_{k \in K} \|x_k\|^2$
Norms induced by inner products

Properties

- Cauchy–Schwarz inequality
  \[ |\langle x, y \rangle| \leq \|x\| \|y\| \]

Examples

- On \( \mathbb{C}^N \):
  \[ \sum_{n=0}^{N-1} x_n y_n^* \leq \left( \sum_{n=0}^{N-1} |x_n|^2 \right)^{1/2} \left( \sum_{n=0}^{N-1} |y_n|^2 \right)^{1/2} \]

- On \( \mathbb{C} \)-valued random variables:
  \[ |E(xy^*)| \leq \left( E[|x|^2] E[|y|^2] \right)^{1/2} \]
  \[ \implies \text{correlation coefficient } \rho \text{ satisfies } |\rho| \leq 1 \]

\[ \cos \theta = \frac{\langle x, y \rangle}{\|x\| \|y\|} \] defines angle \( \theta \) between vectors
Norms not necessarily induced by inner products

Examples

- On $\mathbb{C}^N$: $\|x\|_p = \left(\sum_{n=0}^{N-1} |x_n|^p\right)^{1/p}$, $p \in [1, \infty)$

- On $\mathbb{C}^\mathbb{Z}$: $\|x\|_p = \left(\sum_{n \in \mathbb{Z}} |x_n|^p\right)^{1/p}$, $p \in [1, \infty)$

- On $\mathbb{C}^\mathbb{R}$: $\|x\|_p = \left(\int_{-\infty}^{\infty} |x(t)|^p \, dt\right)^{1/p}$, $p \in [1, \infty)$

$\|x\|_\infty = \sup_{n \in \mathbb{Z}} |x_n|

$\|x\|_\infty = \text{ess sup}_{t \in \mathbb{R}} |x(t)|$

Only induced by inner products for $p = 2$
Geometry of $\ell^p$: Unit balls

Valid norm (and convex unit ball) for $p \geq 1$; ordinary geometry for $p = 2$
Normed vector spaces

- A normed vector space is a set satisfying axioms of a vector space where the norm is finite.

- $\ell^2(\mathbb{Z})$: square-summable sequences ("finite-energy discrete-time signals")
  \[ \|x\| = \left( \sum_{n \in \mathbb{Z}} |x_n|^2 \right)^{1/2} < \infty \]

- $L^2(\mathbb{R})$: square-integrable functions ("finite-energy continuous-time signals")
  \[ \|x\| = \left( \int_{-\infty}^{\infty} |x(t)|^2 \, dt \right)^{1/2} < \infty \]

$x$ and $y$ are the same when $\|x - y\| = 0$. No harm in considering only functions with finitely-many discontinuities.
Normed vector spaces

- A normed vector space is a set satisfying axioms of a vector space where the norm is finite

- $\ell^2(\mathbb{Z})$: square-summable sequences ("finite-energy discrete-time signals")

$$
\|x\| = \left( \sum_{n \in \mathbb{Z}} |x_n|^2 \right)^{1/2} < \infty
$$

- $L^2(\mathbb{R})$: square-integrable functions ("finite-energy continuous-time signals")

$$
\|x\| = \left( \int_{-\infty}^{\infty} |x(t)|^2 \, dt \right)^{1/2} < \infty
$$

$x$ and $y$ are the same when $\|x - y\| = 0$

No harm in considering only functions with finitely-many discontinuities
Hilbert spaces: Convergence

Definition
A sequence of vectors \( x_0, x_1, \ldots \) in a normed vector space \( V \) is said to converge to \( v \in V \) when \( \lim_{k \to \infty} \| v - x_k \| = 0 \), or for any \( \varepsilon > 0 \), there exists \( K_\varepsilon \) such that \( \| v - x_k \| < \varepsilon \) for all \( k > K_\varepsilon \).

Choice of the norm in \( V \) is key.

Example
For \( k \in \mathbb{Z}^+ \), let
\[
\alpha(t) = \begin{cases} 
1, & \text{for } t \in [0, 1/k], \\
0, & \text{otherwise}.
\end{cases}
\]
\( v(t) = 0 \) for all \( t \). Then for \( p \in [1, \infty) \),
\[
\| v - \alpha \|_p = \left( \int_{-\infty}^{\infty} |v(t) - \alpha(t)|^p \, dt \right)^{1/p} = \left( \frac{1}{k} \right)^{1/p} \xrightarrow{k \to \infty} 0.
\]
For \( p = \infty \), \( \| v - \alpha \|_\infty = 1 \) for all \( k \).
Hilbert spaces: Convergence

Definition
A sequence of vectors $x_0, x_1, \ldots$ in a normed vector space $V$ is said to converge to $v \in V$ when $\lim_{k \to \infty} \|v - x_k\| = 0$, or for any $\varepsilon > 0$, there exists $K_\varepsilon$ such that $\|v - x_k\| < \varepsilon$ for all $k > K_\varepsilon$.

Choice of the norm in $V$ is key

Example
For $k \in \mathbb{Z}^+$, let

\[ x_k(t) = \begin{cases} 1, & \text{for } t \in [0, \frac{1}{k}] \\ 0, & \text{otherwise.} \end{cases} \]

$v(t) = 0$ for all $t$. Then for $p \in [1, \infty)$,

\[ \|v - x_k\|_p = \left( \int_{-\infty}^{\infty} |v(t) - x_k(t)|^p \, dt \right)^{1/p} = \left( \frac{1}{k} \right)^{1/p} \xrightarrow{k \to \infty} 0. \]

For $p = \infty$: $\|v - x_k\|_\infty = 1$ for all $k$. 
Hilbert spaces: Convergence

Definition
A sequence of vectors $x_0, x_1, \ldots$ in a normed vector space $V$ is said to converge to $v \in V$ when $\lim_{k \to \infty} \|v - x_k\| = 0$, or for any $\varepsilon > 0$, there exists $K_\varepsilon$ such that $\|v - x_k\| < \varepsilon$ for all $k > K_\varepsilon$.

Choice of the norm in $V$ is key

Example
For $k \in \mathbb{Z}^+$, let
\[
x_k(t) = \begin{cases} 
1, & \text{for } t \in [0, 1/k]; \\
0, & \text{otherwise}.
\end{cases}
\]
$v(t) = 0$ for all $t$. Then for $p \in [1, \infty)$,
\[
\|v - x_k\|_p = \left( \int_{-\infty}^{\infty} |v(t) - x_k(t)|^p \, dt \right)^{1/p} = \left( \frac{1}{k} \right)^{1/p} \xrightarrow{k \to \infty} 0.
\]
For $p = \infty$, $\|v - x_k\|_\infty = 1$ for all $k$.
Hilbert spaces: Completeness

Definitions

- A sequence \( \{x_n\} \) is a **Cauchy sequence** in a normed space when for any \( \varepsilon > 0 \), there exists \( k_\varepsilon \) such that \( \|x_k - x_m\| < \varepsilon \) for all \( k, m > k_\varepsilon \).

- A normed vector space \( V \) is **complete** if every Cauchy sequence converges in \( V \).

- A complete normed vector space is called a **Banach space**.

- A complete inner product space is called a **Hilbert space**.
Hilbert spaces

Examples

- \( \mathbb{Q} \) is not a complete space

\[
\sum_{n=1}^{\infty} \frac{1}{n^2} \rightarrow \frac{\pi^2}{6} \in \mathbb{R}, \not\in \mathbb{Q}
\]

\[
\sum_{n=0}^{\infty} \frac{1}{n!} \rightarrow e \in \mathbb{R}, \not\in \mathbb{Q}
\]

- \( \mathbb{R} \) is a complete space
Hilbert spaces

Examples

- $\mathbb{Q}$ is not a complete space
  
  $$\sum_{n=1}^{\infty} \frac{1}{n^2} \to \frac{\pi^2}{6} \in \mathbb{R}, \notin \mathbb{Q}$$

  $$\sum_{n=0}^{\infty} \frac{1}{n!} \to e \in \mathbb{R}, \notin \mathbb{Q}$$

- $\mathbb{R}$ is a complete space
Hilbert spaces

Examples

- All finite-dimensional spaces (over field of scalars $\mathbb{C}$ or $\mathbb{R}$) are complete
- $l^p(\mathbb{Z})$ and $L^p(\mathbb{R})$ are complete
  - $l^2(\mathbb{Z})$ and $L^2(\mathbb{R})$ are Hilbert spaces
- $C^q([a, b])$, functions on $[a, b]$ with $q$ continuous derivatives, are not complete except for $q = 0$ under $L^\infty$ norm
Hilbert spaces

Examples

- All finite-dimensional spaces (over field of scalars \( \mathbb{C} \) or \( \mathbb{R} \)) are complete
- \( \ell^p(\mathbb{Z}) \) and \( L^p(\mathbb{R}) \) are complete
  - \( \ell^2(\mathbb{Z}) \) and \( L^2(\mathbb{R}) \) are Hilbert spaces
- \( C^q([a, b]) \), functions on \([a, b]\) with \( q \) continuous derivatives, are not complete except for \( q = 0 \) under \( L^\infty \) norm
Hilbert spaces

**Examples**

- All finite-dimensional spaces (over field of scalars $\mathbb{C}$ or $\mathbb{R}$) are complete
- $\ell^p(\mathbb{Z})$ and $\mathcal{L}^p(\mathbb{R})$ are complete
  - $\ell^2(\mathbb{Z})$ and $\mathcal{L}^2(\mathbb{R})$ are Hilbert spaces
- $C^q([a, b])$, functions on $[a, b]$ with $q$ continuous derivatives, are not complete except for $q = 0$ under $\mathcal{L}^\infty$ norm
Summary on spaces

Vector spaces

Normed vector spaces

Inner product spaces

Banach spaces

Hilbert spaces

• $\mathbb{Q}^N$
• $C([a, b])$
• $\ell^1(\mathbb{Z})$
• $\ell^\infty(\mathbb{Z})$
• $\mathcal{C}^1([a, b])$, $\|\cdot\|_\infty$
• $(V, d)$
• $\ell^0(\mathbb{Z})$
Linear operators

Linear operators generalize matrices

Definitions

- $A : H_0 \to H_1$ is a **linear operator** when for all $x, y \in H_0, \alpha \in \mathbb{C}$:
  - Additivity: $A(x + y) = Ax + Ay$
  - Scalability: $A(\alpha x) = \alpha(Ax)$
- Null space (subspace of $H_0$): $\mathcal{N}(A) = \{x \in H_0, Ax = 0\}$
- Range space (subspace of $H_1$): $\mathcal{R}(A) = \{Ax \in H_1, x \in H_0\}$
- Operator norm: $\|A\| = \sup_{\|x\| = 1} \|Ax\|$
- A is bounded when: $\|A\| < \infty$
- Inverse: Bounded $B : H_1 \to H_0$ inverse of bounded $A$ if and only if:
  - $B Ax = x$, for every $x \in H_0$
  - $ABy = y$, for every $y \in H_1$
Linear operators

Linear operators generalize matrices

Definitions

- $A : H_0 \to H_1$ is a **linear operator** when for all $x, y \in H_0, \alpha \in \mathbb{C}$:
  - **Additivity**: $A(x + y) = Ax + Ay$
  - **Scalability**: $A(\alpha x) = \alpha(Ax)$
- **Null space** (subspace of $H_0$): $\mathcal{N}(A) = \{x \in H_0, Ax = 0\}$
- **Range space** (subspace of $H_1$): $\mathcal{R}(A) = \{Ax \in H_1, x \in H_0\}$
- **Operator norm**: $\|A\| = \sup_{\|x\|=1} \|Ax\|$
- $A$ is **bounded** when: $\|A\| < \infty$
- **Inverse**:
  - Bounded $B : H_1 \to H_0$ inverse of bounded $A$ if and only if:
    - $BAx = x$, for every $x \in H_0$
    - $ABy = y$, for every $y \in H_1$
## Linear operators

Linear operators generalize matrices

<table>
<thead>
<tr>
<th>Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A : H_0 \to H_1$ is a <strong>linear operator</strong> when for all $x, y \in H_0, \alpha \in \mathbb{C}$:</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Null space</strong> (subspace of $H_0$): $\mathcal{N}(A) = {x \in H_0, Ax = 0}$</td>
</tr>
<tr>
<td><strong>Range space</strong> (subspace of $H_1$): $\mathcal{R}(A) = {Ax \in H_1, x \in H_0}$</td>
</tr>
<tr>
<td><strong>Operator norm:</strong> $|A| = \sup_{|x|=1} |Ax|$</td>
</tr>
<tr>
<td><strong>A is bounded when:</strong> $|A| &lt; \infty$</td>
</tr>
<tr>
<td><strong>Inverse:</strong> Bounded $B : H_1 \to H_0$ inverse of bounded $A$ if and only if:</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Linear operators

Linear operators generalize matrices

Definitions

- **A**: $H_0 \rightarrow H_1$ is a **linear operator** when for all $x, y \in H_0, \alpha \in \mathbb{C}$:
  - Additivity: $A(x + y) = Ax + Ay$
  - Scalability: $A(\alpha x) = \alpha(Ax)$
- **Null space** (subspace of $H_0$): $\mathcal{N}(A) = \{x \in H_0, Ax = 0\}$
- **Range space** (subspace of $H_1$): $\mathcal{R}(A) = \{Ax \in H_1, x \in H_0\}$
- **Operator norm**: $\|A\| = \sup_{\|x\| = 1} \|Ax\|$
- **A is bounded** when: $\|A\| < \infty$
- **Inverse**: Bounded $B : H_1 \rightarrow H_0$ inverse of bounded $A$ if and only if:
  - $BAx = x$, for every $x \in H_0$
  - $ABy = y$, for every $y \in H_1$
Linear operators: Illustration

\[ R(A) \text{ is the plane } 5y_1 + 2y_2 + 8y_3 = 0 \]
Adjoint operators

Adjoint generalizes Hermitian transposition of matrices

**Definition (Adjoint and self-adjoint operators)**

- $A^* : H_1 \to H_0$ is the adjoint of $A : H_0 \to H_1$ when
  $$\langle Ax, y \rangle_{H_1} = \langle x, A^* y \rangle_{H_0} \text{ for every } x \in H_0, y \in H_1$$

- If $A = A^*$, $A$ is self-adjoint or Hermitian

- Note that $\mathcal{N}(A^*) = \mathcal{R}(A)^\perp$
Adjoint operator: Illustration

\[ \mathcal{N}(A^*) \text{ is the line } \frac{1}{2} y_1 = \frac{1}{2} y_2 = \frac{1}{2} y_3 \]
Adjoint operators

Theorem (Adjoint properties)

Let $A : H_0 \rightarrow H_1$ be a bounded linear operator

- $A^*$ exists and is unique
- $(A^*)^* = A$
- $AA^*$ and $A^*A$ are self-adjoint
- $\|A^*\| = \|A\|
- If $A$ is invertible, $(A^{-1})^* = (A^*)^{-1}$
- If $B : H_0 \rightarrow H_1$ is bounded, $(A + B)^* = A^* + B^*$
- If $B : H_1 \rightarrow H_2$ is bounded, $(BA)^* = A^*B^*$
Adjoint operators: Local averaging

\[ A : L^2(\mathbb{R}) \to \ell^2(\mathbb{Z}) \quad (Ax)_k = \int_{k-1/2}^{k+1/2} x(t) \, dt \]

\[
\langle Ax, y \rangle_{\ell^2} = \sum_{n \in \mathbb{Z}} (Ax)_n y^*_n = \sum_{n \in \mathbb{Z}} \left( \int_{n-1/2}^{n+1/2} x(t) \, dt \right) y^*_n = \sum_{n \in \mathbb{Z}} \int_{n-1/2}^{n+1/2} x(t) y^*_n \, dt \\
= \sum_{n \in \mathbb{Z}} \int_{n-1/2}^{n+1/2} x(t)(A^*y)^*_n(t) \, dt = \int_{-\infty}^{\infty} x(t)(A^*y)^*_n(t) \, dt = \langle x, A^*y \rangle_{\ell^2}
\]
Unitary operators

Definition (Unitary operators)

- A bounded linear operator $A : H_0 \to H_1$ is unitary when:
  - $A$ is invertible
  - $A$ preserves inner products: $\langle Ax, Ay \rangle_{H_1} = \langle x, y \rangle_{H_0}$ for every $x, y \in H_0$

- If $A$ is unitary, then $\|Ax\|^2 = \|x\|^2$
- $A$ is unitary if and only if $A^{-1} = A^*$
Unitary operators

**Definition (Unitary operators)**

- A bounded linear operator $A : H_0 \to H_1$ is unitary when:
  - $A$ is invertible
  - $A$ preserves inner products: $\langle Ax, Ay \rangle_{H_1} = \langle x, y \rangle_{H_0}$ for every $x, y \in H_0$

- If $A$ is unitary, then $\|Ax\|^2 = \|x\|^2$
- $A$ is unitary if and only if $A^{-1} = A^*$
Projection operators

Definition (Projection, orthogonal projection, oblique projection)

- \( P \) is idempotent when \( P^2 = P \)
- A projection operator is a bounded linear operator that is idempotent
- An orthogonal projection operator is a self-adjoint projection operator
- An oblique projection operator is not self-adjoint
Projection operators

**Definition (Projection, orthogonal projection, oblique projection)**

- **$P$ is idempotent** when $P^2 = P$
- **A projection operator** is a bounded linear operator that is idempotent
- **An orthogonal projection** operator is a self-adjoint projection operator
- **An oblique projection** operator is not self adjoint

**Theorem**

- If $A : H_0 \rightarrow H_1$, $B : H_1 \rightarrow H_0$ bounded and $A$ is a left inverse of $B$, then $BA$ is a projection operator. If $B = A^*$ then, $BA = A^*A$ is an orthogonal projection

$$(BA)^2 = BA BA = B(AB)A = BA$$
Best approximation: Euclidean geometry

- $\mathbf{x}$ is a point in Euclidean space
- $S$ is a line in Euclidean space
Best approximation: Euclidean geometry

- $x$ is a point in Euclidean space
- $S$ is a line in Euclidean space

Nearest point problem: Find $\hat{x} \in S$ that is closest to $x$
Best approximation: Euclidean geometry

- $x$ is a point in Euclidean space
- $S$ is a line in Euclidean space

Nearest point problem: Find $\hat{x} \in S$ that is closest to $x$
Best approximation: Euclidean geometry

- $x$ is a point in Euclidean space
- $S$ is a line in Euclidean space

Nearest point problem: Find $\hat{x} \in S$ that is closest to $x$

Solution uniquely determined by $x - \hat{x} \perp S$
  - Circle must touch $S$ in one point, radius $\perp$ tangent
Best approximation: Euclidean geometry

- $x$ is a point in Euclidean space
- $S$ is a line in Euclidean space

Nearest point problem: Find $\hat{x} \in S$ that is closest to $x$

Solution uniquely determined by $x - \hat{x} \perp S$
  - Circle must touch $S$ in one point, radius $\perp$ tangent
Best approximation: Hilbert space geometry

- \( S \) closed subspace of a Hilbert space
- Best approximation problem:

\[
\hat{x} \in S \text{ that is closest to } x
\]

\[
\hat{x} = \arg \min_{s \in S} \|x - s\|
\]
Best approximation by orthogonal projection

**Theorem (Projection theorem)**

Let $S$ be a closed subspace of Hilbert space $H$ and let $x \in H$.

- **Existence**: There exists $\hat{x} \in S$ such that $\|x - \hat{x}\| \leq \|x - s\|$ for all $s \in S$.

- **Orthogonality**: $x - \hat{x} \perp S$ is necessary and sufficient to determine $\hat{x}$.

- **Uniqueness**: $\hat{x}$ is unique.

- **Linearity**: $\hat{x} = Px$ where $P$ is a linear operator.

- **Idempotency**: $P(Px) = Px$ for all $x \in H$.

- **Self-adjointness**: $P = P^*$

All "nearest vector in a subspace" problems in Hilbert spaces are the same.
Best approximation by orthogonal projection

**Theorem (Projection theorem)**

Let $S$ be a closed subspace of Hilbert space $H$ and let $x \in H$.

- **Existence:** There exists $\hat{x} \in S$ such that $\|x - \hat{x}\| \leq \|x - s\|$ for all $s \in S$

- **Orthogonality:** $x - \hat{x} \perp S$ is necessary and sufficient to determine $\hat{x}$

- **Uniqueness:** $\hat{x}$ is unique

- **Linearity:** $\hat{x} = Px$ where $P$ is a linear operator

- **Idempotency:** $P(Px) = Px$ for all $x \in H$

- **Self-adjointness:** $P = P^*$

All “nearest vector in a subspace” problems in Hilbert spaces are the same
Best approximation by orthogonal projection

Theorem (Projection theorem)

Let $S$ be a closed subspace of Hilbert space $H$ and let $x \in H$.

- **Existence:** There exists $\hat{x} \in S$ such that $\|x - \hat{x}\| \leq \|x - s\|$ for all $s \in S$
- **Orthogonality:** $x - \hat{x} \perp S$ is necessary and sufficient to determine $\hat{x}$
- **Uniqueness:** $\hat{x}$ is unique
- **Linearity:** $\hat{x} = Px$ where $P$ is a linear operator
- **Idempotency:** $P(Px) = Px$ for all $x \in H$
- **Self-adjointness:** $P = P^*$

All “nearest vector in a subspace” problems in Hilbert spaces are the same
Example 1: Least-square polynomial approximation

- Consider: \( x(t) = \cos(\frac{3}{4} \pi t) \in L^2([0,1]) \)
- Find the degree-1 polynomial closest to \( x \) (in \( L^2 \) norm)
- Solution: Use orthogonality
Example 1: Least-square polynomial approximation

- Consider: \( x(t) = \cos\left(\frac{3\pi}{2}t\right) \in L^2([0,1]) \)
- Find the degree-1 polynomial closest to \( x \) (in \( L^2 \) norm)
- Solution: Use orthogonality
Example 1: Least-square polynomial approximation

- Consider: \( x(t) = \cos\left(\frac{3}{2}\pi t\right) \in L^2([0,1]) \)
- Find the degree-1 polynomial closest to \( x \) (in \( L^2 \) norm)
- Solution: Use orthogonality

\[
0 = \langle x(t) - \hat{x}(t), 1 \rangle_t = \int_0^1 \left( \cos\left(\frac{3}{2}\pi t\right) - (a_0 + a_1 t) \right) dt = -\frac{2}{3\pi} - a_0 - \frac{1}{2}a_1
\]

\[
0 = \langle x(t) - \hat{x}(t), t \rangle_t = \int_0^1 \left( \cos\left(\frac{3}{2}\pi t\right) - (a_0 + a_1 t) \right) t dt = -\frac{4 + 6\pi}{9\pi^2} - \frac{1}{2}a_0 - \frac{1}{3}a_1
\]

Approx. with degree-1 polynomial

Approx. with higher-degree polynomials
Example 2: MMSE estimate

- **Consider:** Real-valued random variable $x$
- Find the constant $c$ that minimizes $\mathbb{E}[(x - c)^2]$

**Note:**
- Expected square is a squared Hilbert space norm
- Constants are a closed subspace in vector space of random variables

**Solution:** Use orthogonality
- $c$ determined uniquely by $\mathbb{E}[(x - c)\alpha c] = 0$ for all $\alpha \in \mathbb{R}$
- $c = \mathbb{E}[x]$

**Alternative:**
- Expand into quadratic function of $c$ and minimize with calculus
- Not too difficult, but lacks insight
Example 2: MMSE estimate

- Consider: Real-valued random variable $x$
- Find the constant $c$ that minimizes $E[(x - c)^2]$

Note:
- Expected square is a squared Hilbert space norm
- Constants are a closed subspace in vector space of random variables

Solution: Use orthogonality
- $c$ determined uniquely by $E[(x - c)\alpha c] = 0$ for all $\alpha \in \mathbb{R}$
- $c = E[x]$

Alternative:
- Expand into quadratic function of $c$ and minimize with calculus
- Not too difficult, but lacks insight
Example 2: MMSE estimate

- Consider: Real-valued random variable $x$
- Find the constant $c$ that minimizes $E[(x - c)^2]$  
- Note:
  - Expected square is a squared Hilbert space norm
  - Constants are a closed subspace in vector space of random variables
- Solution: Use orthogonality
  - $c$ determined uniquely by $E[(x - c)\alpha c] = 0$ for all $\alpha \in \mathbb{R}$
  - $c = E[x]$
- Alternative:
  - Expand into quadratic function of $c$ and minimize with calculus
  - Not too difficult, but lacks insight
Example 2: MMSE estimate

- Consider: Real-valued random variable $x$
- Find the constant $c$ that minimizes $E[(x - c)^2]$

**Note:**
- Expected square is a squared Hilbert space norm
- Constants are a closed subspace in vector space of random variables

**Solution:** Use orthogonality
- $c$ determined uniquely by $E[(x - c)\alpha c] = 0$ for all $\alpha \in \mathbb{R}$
- $c = E[x]$

**Alternative:**
- Expand into quadratic function of $c$ and minimize with calculus
- Not too difficult, but lacks insight
Example 2: MMSE estimate

- Consider: Real-valued random variable $x$
- Find the constant $c$ that minimizes $E[(x - c)^2]$

**Note:**
- Expected square is a squared Hilbert space norm
- Constants are a closed subspace in vector space of random variables

**Solution:** Use orthogonality
- $c$ determined uniquely by $E[(x - c)\alpha c] = 0$ for all $\alpha \in \mathbb{R}$
- $c = E[x]$

**Alternative:**
- Expand into quadratic function of $c$ and minimize with calculus
- Not too difficult, but lacks insight
Example 3: Wiener filter

- Consider: Jointly wide-sense stationary discrete-time stochastic processes $x$ and $y$
- Find the linear shift-invariant filter $h$ that minimizes $E[|x_n - \hat{x}_n|^2]$ where $\hat{x} = h \ast y$

Note:
- Expected squared absolute value is a squared Hilbert space norm
- LSI filtering puts $\hat{x}_n$ in a closed subspace
- Solution: Use orthogonality (extended for processes)
  - $h$ determined uniquely by relation between cross- and autocorrelations:
    \[ c_{x,y,k} = a_{x,k}, \quad k \in \mathbb{Z} \]
  - DTFT-domain version: $H(e^{i\omega}) = \frac{C_{x,y}(e^{i\omega})}{A_y(e^{i\omega})}$, $\omega \in \mathbb{R}$
- Alternative:
  - Expand into quadratic function of $h$ and minimize with calculus
  - Messy (!)—especially in complex case—and lacks insight
Example 3: Wiener filter

- Consider: Jointly wide-sense stationary discrete-time stochastic processes $x$ and $y$
- Find the linear shift-invariant filter $h$ that minimizes $E[|x_n - \hat{x}_n|^2]$ where $\hat{x} = h * y$

Note:
- Expected squared absolute value is a squared Hilbert space norm
- LSI filtering puts $\hat{x}_n$ in a closed subspace

Solution: Use orthogonality (extended for processes)
- $h$ determined uniquely by relation between cross- and autocorrelations:
  $$c_{x,y,k} = a_{x,k}, \quad k \in \mathbb{Z}$$
- DTFT-domain version: $H(e^{j\omega}) = \frac{C_{x,y}(e^{j\omega})}{A_y(e^{j\omega})}, \quad \omega \in \mathbb{R}$

Alternative:
- Expand into quadratic function of $h$ and minimize with calculus
- Messy (!)—especially in complex case—and lacks insight
Example 3: Wiener filter

- Consider: Jointly wide-sense stationary discrete-time stochastic processes $x$ and $y$
- Find the linear shift-invariant filter $h$ that minimizes $E[|x_n - \hat{x}_n|^2]$ where $\hat{x} = h * y$

**Note:**
- Expected squared absolute value is a squared Hilbert space norm
- LSI filtering puts $\hat{x}_n$ in a closed subspace

**Solution:** Use orthogonality (extended for processes)
- $h$ determined uniquely by relation between cross- and autocorrelations:
  $$c_{x,\hat{x},k} = a_{\hat{x},k}, \quad k \in \mathbb{Z}$$
- DTFT-domain version: $H(e^{j\omega}) = \frac{C_{x,\hat{x}}(e^{-j\omega})}{A_Y(e^{j\omega})}, \quad \omega \in \mathbb{R}$

**Alternative:**
- Expand into quadratic function of $h$ and minimize with calculus
- Messy (!)—especially in complex case—and lacks insight
Example 3: Wiener filter

- Consider: Jointly wide-sense stationary discrete-time stochastic processes $x$ and $y$
- Find the linear shift-invariant filter $h$ that minimizes $E[|x_n - \hat{x}_n|^2]$ where $\hat{x} = h * y$

Note:
- Expected squared absolute value is a squared Hilbert space norm
- LSI filtering puts $\hat{x}_n$ in a closed subspace
- Solution: Use orthogonality (extended for processes)
  - $h$ determined uniquely by relation between cross- and autocorrelations:
    $$c_{x,\hat{x},k} = a_{\hat{x},k}, \quad k \in \mathbb{Z}$$
  - DTFT-domain version: $H(e^{j\omega}) = \frac{C_{x,y}(e^{j\omega})}{A_y(e^{j\omega})}, \quad \omega \in \mathbb{R}$

Alternative:
  - Expand into quadratic function of $h$ and minimize with calculus
  - Messy (!)—especially in complex case—and lacks insight
Example 4: Best piecewise-constant approximation

- Local averaging
  \[ A : L^2(\mathbb{R}) \to \ell^2(\mathbb{Z}) \quad (Ax)_k = \int_{k - \frac{1}{2}}^{k + \frac{1}{2}} x(t)dt \]
  has adjoint \( A^* : \ell^2(\mathbb{Z}) \to L^2(\mathbb{R}) \) that produces staircase function

- \( AA^* \) is identity, so \( A^*A \) is orthogonal projection

\[
\]
Example 5: Approximations of “All is vanity” image—Haar
Example 5: Approximations of “All is vanity” image—Haar
Example 5: Approximations of “All is vanity” image—sinc
Bases

Definition (Basis)

- \( \Phi = \{ \varphi_k \}_{k \in K} \subset \mathcal{V} \) is a basis when
  - \( \Phi \) is linearly independent
  - \( \Phi \) is complete in \( \mathcal{V} \): \( \mathcal{V} = \text{span}(\Phi) \)

Expansion formula: for any \( x \in \mathcal{V} \), \( x = \sum_{k \in K} \alpha_k \varphi_k \)

Example

- The standard basis for \( \mathbb{R}^N \)
  \[ e_k = [0 \ldots 0 1 0 \ldots 0]^T \quad k = 0, \ldots, N-1 \]

- For any \( x \in \mathbb{R}^N \), \( x = \sum_{k=0}^{N-1} x_k e_k \)
Bases

**Definition (Basis)**

- \( \Phi = \{\varphi_k\}_{k \in K} \subseteq V \) is a basis when
  - \( \Phi \) is linearly independent and
  - \( \Phi \) is complete in \( V \): \( V = \text{span}(\Phi) \)

- **Expansion formula**: for any \( x \in V \), \( x = \sum_{k \in K} \alpha_k \varphi_k \)
  - \( \{\alpha_k\}_{k \in K} \) is unique
  - \( \alpha_k \) are expansion coefficients

**Example**

- The standard basis for \( \mathbb{R}^N \)
  \( e_k = [0 \ 0 \ \ldots \ 0 \ 1 \ 0 \ \ldots \ 0] \) for \( k = 0, \ldots, N-1 \)
  - any \( x \in \mathbb{R}^N \), \( x = \sum_{k=0}^{N-1} x_k e_k \)
Bases

**Definition (Basis)**

- $\Phi = \{\varphi_k\}_{k \in K} \subset V$ is a **basis** when
  - $\Phi$ is linearly independent and
  - $\Phi$ is complete in $V$: $V = \text{span}(\Phi)$

**Expansion formula:** for any $x \in V$, 

$$x = \sum_{k \in K} \alpha_k \varphi_k$$

- $\{\alpha_k\}_{k \in K}$: is unique
- $\alpha_k$: expansion coefficients

**Example**

- The standard basis for $\mathbb{R}^N$
  $$e_k = [0 \ 0 \ \cdots \ 0 \ 1 \ 0 \ \cdots \ 0]^T, \ k = 0, \ldots, N - 1$$

  any $x \in \mathbb{R}^N, \ x = \sum_{k=0}^{N-1} x_k e_k$
Bases

Examples

- This tutorial concentrates on orthonormal case
- Full study should include use of frames (overcomplete sets)
Operators associated with bases

**Definition (Basis synthesis operator)**

- **Synthesis operator**
  - $\Phi : \ell^2(K) \rightarrow H$
  - $\Phi \alpha = \sum_{k \in K} \alpha_k \varphi_k$

- Adjoint: Let $\alpha \in \ell^2(\mathbb{Z})$ and $y \in H$
  - $\langle \Phi \alpha, y \rangle = \left\langle \sum_{k \in K} \alpha_k \varphi_k, y \right\rangle = \sum_{k \in K} \alpha_k \langle \varphi_k, y \rangle = \sum_{k \in K} \alpha_k \langle y, \varphi_k \rangle^*$

**Definition (Basis analysis operator)**

- **Analysis operator**
  - $\Phi^* : H \rightarrow \ell^2(K)$
  - $(\Phi^* x)_k = \langle x, \varphi_k \rangle$

- Note that the analysis operator is the adjoint of the synthesis operator
Operators associated with bases

**Definition (Basis synthesis operator)**

- **Synthesis** operator
  - \( \Phi : \ell^2(\mathcal{K}) \rightarrow H \)
  - \( \Phi \alpha = \sum_{k \in \mathcal{K}} \alpha_k \varphi_k \)
  - Adjoint: Let \( \alpha \in \ell^2(\mathbb{Z}) \) and \( y \in H \)
    \[
    \langle \Phi \alpha, y \rangle = \left\langle \sum_{k \in \mathcal{K}} \alpha_k \varphi_k, y \right\rangle = \sum_{k \in \mathcal{K}} \alpha_k \langle \varphi_k, y \rangle = \sum_{k \in \mathcal{K}} \alpha_k \langle y, \varphi_k \rangle^* \]

**Definition (Basis analysis operator)**

- **Analysis** operator
  - \( \Phi^* : H \rightarrow \ell^2(\mathcal{K}) \)
  - \( (\Phi^* x)_k = \langle x, \varphi_k \rangle, \; k \in \mathcal{K} \)

Note that the analysis operator is the adjoint of the synthesis operator.
Orthonormal bases

Definition (Orthonormal basis)

- $\Phi = \{\varphi_k\}_{k \in K} \subset H$ is an **orthonormal basis** for $H$ when
  - $\Phi$ is a basis for $H$ and
  - $\Phi$ is an orthonormal set
    \[ \langle \varphi_i, \varphi_k \rangle = \delta_{i-k} \text{ for all } i, k \in K \]

- If $\Phi$ is an orthogonal set, then it is linearly independent

- If $\text{span}(\Phi) = H$ and $\Phi$ is an orthogonal set, then $\Phi$ is an orthogonal basis for $H$

- If we also have $\|\varphi_k\| = 1$, then $\Phi$ is an orthonormal basis
## Orthonormal basis expansions

### Definition (Orthonormal basis expansions)

- \( \Phi = \{ \varphi_k \}_{k \in \mathcal{K}} \) orthonormal basis for \( H \), then for any \( x \in H \):
  - \( \alpha_k = \langle x, \varphi_k \rangle \) for \( k \in \mathcal{K} \), or \( \alpha = \Phi^* x \), and \( \alpha \) is unique.
- **Synthesis:** \( x = \sum_{k \in \mathcal{K}} \langle x, \varphi_k \rangle \varphi_k = \Phi \alpha = \Phi \Phi^* x \)

### Example

\[ \varphi_0 = \phi_0 \]
\[ \varphi_1 = \phi_1 \]
\[ \varphi_2 = \phi_2 \]
\[ \langle x, \varphi_0 \rangle \]
\[ \langle x, \varphi_1 \rangle \]
\[ \langle x, \varphi_2 \rangle \]
Orthonormal basis: Parseval equality

Theorem (Parseval’s equalities)

- \( \Phi = \{ \phi_k \}_{k \in K} \) orthonormal basis for \( H \)

\[
\|x\|^2 = \sum_{k \in K} |(x, \phi_k)|^2 = \|\Phi^*x\|^2 = \|\alpha\|^2
\]

- In general:

\[
(x, y) = (\Phi^*x, \Phi^*y) = (\alpha, \beta)
\]

where \( \alpha_k = (x, \phi_k) \), \( \beta_k = (y, \phi_k) \)
Orthonormal basis: Parseval equality

Theorem (Parseval’s equalities)

- \( \Phi = \{ \varphi_k \}_{k \in \mathcal{K}} \) orthonormal basis for \( H \)
  \[
  \| x \|^2 = \sum_{k \in \mathcal{K}} |\langle x, \varphi_k \rangle|^2 = \| \Phi^* x \|^2 = \| \alpha \|^2
  \]
- In general:
  \[
  \langle x, y \rangle = \langle \Phi^* x, \Phi^* y \rangle = \langle \alpha, \beta \rangle
  \]
  where \( \alpha_k = \langle x, \varphi_k \rangle, \beta_k = \langle y, \varphi_k \rangle \)
Orthogonal projection and decomposition

**Theorem**

- \( \Phi = \{ \varphi_k \}_{k \in I} \subset H, \quad I \subset K \)

\[
P_I x = \sum_{k \in I} \langle x, \varphi_k \rangle \varphi_k = \Phi \Phi^*_I x
\]

is the **orthogonal projection** of \( x \) onto \( S_I = \text{span}(\{ \varphi_k \}_{k \in I}) \)

- \( \Phi \) induces an orthogonal decomposition

\[
H = \bigoplus_{k \in K} S_{\{k\}} \quad \text{where} \quad S_{\{k\}} = \text{span}(\varphi_k)
\]
Orthogonal projection and decomposition

**Theorem**

- \( \Phi = \{ \varphi_k \}_{k \in I} \subset H, \ I \subset K \)
- \( P_I x = \sum_{k \in I} \langle x, \varphi_k \rangle \varphi_k = \Phi \Phi^*_I x \)

is the **orthogonal projection** of \( x \) onto \( S_I = \text{span}(\{ \varphi_k \}_{k \in I}) \)

- \( \Phi \) induces an orthogonal decomposition

\[
H = \bigoplus_{k \in K} S_{(k)} \quad \text{where} \quad S_{(k)} = \text{span}(\varphi_k)
\]
Let \( y = Ax \) with \( A : H \to H \)

Matrix representation of operator: Orthonormal basis

- How are expansion coefficients of \( x \) and \( y \) related?
  - \( \{ \varphi_k \}_{k \in K} \) orthonormal basis of \( H \)
  - \( x = \Phi \alpha, \quad y = \Phi \beta \)

- Matrix representation allows computation of \( A \) directly on coefficient sequences

\[
\Gamma : \ell^2(K) \to \ell^2(K) \quad \text{s.t.} \quad \beta = \Gamma \alpha
\]

- As a matrix:

\[
\begin{bmatrix}
\vdots & \vdots & \vdots \\
\langle A\varphi_{-1}, \varphi_{-1} \rangle & \langle A\varphi_0, \varphi_{-1} \rangle & \langle A\varphi_1, \varphi_{-1} \rangle & \langle A\varphi_{-1}, \varphi_0 \rangle & \langle A\varphi_0, \varphi_0 \rangle & \langle A\varphi_1, \varphi_0 \rangle \\
\langle A\varphi_{-1}, \varphi_0 \rangle & \langle A\varphi_0, \varphi_0 \rangle & \langle A\varphi_1, \varphi_0 \rangle & \langle A\varphi_{-1}, \varphi_1 \rangle & \langle A\varphi_0, \varphi_1 \rangle & \langle A\varphi_1, \varphi_1 \rangle \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots
\end{bmatrix}
\]
Matrix representation of operator: Orthonormal basis

- Central example for signal processing: $H = \mathcal{B}L[-\pi/T, \pi/T] \subset \mathcal{L}^2(\mathbb{R})$
- When orthonormal bases are used, matrix representation of $A^*$ is $\Gamma^*$
Example: Derivative operator I

Let \( A : H_0 \rightarrow H_1 \) with \( y(t) = (Ax)(t) = x'(t) \)

\( H_0 \): piecewise-linear, continuous, finite-energy functions with breakpoints at integers

\( H_1 \): piecewise-constant, finite-energy functions with breakpoints at integers

- Let \( \varphi(t) = \begin{cases} 1 - |t|, & \text{for } |t| < 1; \\ 0, & \text{otherwise} \end{cases} \) and \( \varphi_k(t) = \varphi(t - k) \) for \( k \in \mathbb{Z} \).

\( \{\varphi_k\}_{k \in \mathbb{Z}} \) is a nonorthonormal basis for \( H_0 \).

- Let \( 1_I(t) = \begin{cases} 1, & \text{for } t \in I; \\ 0, & \text{otherwise} \end{cases} \) and \( \psi_k = 1_{[k,k+1)} \) for \( k \in \mathbb{Z} \).

\( \{\psi_k\}_{k \in \mathbb{Z}} \) is an orthonormal basis for \( H_1 \).
Example: Derivative operator I

- Let $A : H_0 \to H_1$ with $y(t) = (Ax)(t) = x'(t)$

  $H_0$: piecewise-linear, continuous, finite-energy functions with breakpoints at integers

  $H_1$: piecewise-constant, finite-energy functions with breakpoints at integers

- Let $\varphi(t) = \begin{cases} 
  1 - |t|, & \text{for } |t| < 1; \\
  0, & \text{otherwise}
\end{cases}$ and $\varphi_k(t) = \varphi(t - k)$ for $k \in \mathbb{Z}$.

  $\{\varphi_k\}_{k \in \mathbb{Z}}$ is a nonorthonormal basis for $H_0$.

- Let $1_I(t) = \begin{cases} 
  1, & \text{for } t \in I; \\
  0, & \text{otherwise}
\end{cases}$ and $\psi_k = 1_{[k,k+1)}$ for $k \in \mathbb{Z}$.

  $\{\psi_k\}_{k \in \mathbb{Z}}$ is an orthonormal basis for $H_1$. 
Example: Derivative operator II

Example (Cont.)

\[ \varphi'(t) = \begin{cases} 1, & \text{for } -1 < t < 0; \\ -1, & \text{for } 0 < t < 1; \\ 0, & \text{for } |t| > 1, \end{cases} \]

so \( \langle A\varphi_0, \tilde{\psi}_i \rangle = \begin{cases} 1, & \text{for } i = -1; \\ -1, & \text{for } i = 0; \\ 0, & \text{otherwise}. \end{cases} \)

Shifting \( \varphi \) by \( k \) shifts the derivative:

\[ \langle A\varphi_k, \tilde{\psi}_i \rangle = \begin{cases} 1, & \text{for } i = k - 1; \\ -1, & \text{for } i = k; \\ 0, & \text{otherwise.} \end{cases} \)

Then \( \Gamma = \begin{bmatrix} \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \]

Simplicity of matrix representation depends on the basis!
Discrete-time systems

- A linear system $A : \ell^2(\mathbb{Z}) \to \ell^2(\mathbb{Z})$ has a matrix representation $H$ with respect to the standard basis.

- For a linear shift-invariant (LSI) system, the matrix $H$ is Toeplitz:

$$
\begin{bmatrix}
\vdots \\
y_{-2} \\
y_{-1} \\
y_0 \\
y_1 \\
y_2 \\
\vdots \\
\end{bmatrix}
= 
\begin{bmatrix}
\vdots \\
\vdots \\
\vdots \\
\vdots \\
\vdots \\
\vdots \\
\vdots \\
\end{bmatrix} 
= H
\begin{bmatrix}
\vdots \\
x_{-2} \\
x_{-1} \\
x_0 \\
x_1 \\
x_2 \\
\vdots \\
\end{bmatrix} = Hx
$$
Discrete-time systems

- Matrix representation of $A^*$ is $H^*$  [Note: using orthonormal basis]

$$H^* = \begin{bmatrix}
  \cdots & \cdots & \cdots & \cdots & \cdots \\
  h_0^* & h_1^* & h_2^* & h_3^* & h_4^* \\
  h_{-1}^* & h_0^* & h_1^* & h_2^* & h_3^* \\
  h_{-2}^* & h_{-1}^* & h_0^* & h_1^* & h_2^* \\
  h_{-3}^* & h_{-2}^* & h_{-1}^* & h_0^* & h_1^* \\
  h_{-4}^* & h_{-3}^* & h_{-2}^* & h_{-1}^* & h_0^* \\
  \cdots & \cdots & \cdots & \cdots & \cdots
\end{bmatrix}$$

- Adjoint of filtering by $h_n$ is filtering by $h_{-n}^*$

Note: Hilbert space tools—Part II: Bases through discrete-time systems

Goyal & Kovačević
www.FourierAndWavelets.org
September 30, 2012 58 / 90
DTFT and other Fourier representations

- Eigensequences lead to diagonal representation of $H$
- Discrete-time Fourier transform follows logically from the class of operators
- Convolution theorem follows logically from the definition of the DTFT
- Closely-parallel reasoning for all cases:
  - Sequences (convolution, discrete-time Fourier transform)
  - Periodic sequences (circular convolution, discrete Fourier transform)
  - Functions (convolution, Fourier transform)
  - Periodic functions (circular convolution, Fourier series)
Example Lecture:

Sampling and Interpolation
Sampling and Interpolation

Sampling and interpolation bridge the analog and digital worlds
- Sampling: discrete-time sequence from a continuous-time function
- Interpolation: continuous-time function from a discrete-time sequence

Doing all computation in discrete time is the essence of digital signal processing:

\[ x(t) \xrightarrow{\text{Sampling}} y_n \xrightarrow{\text{DT processing}} w_n \xrightarrow{\text{Interpolation}} v(t) \]

Interpolation followed by sampling occurs in digital communication:

\[ x_n \xrightarrow{\text{Interpolation}} y(t) \xrightarrow{\text{CT channel}} v(t) \xrightarrow{\text{Sampling}} \hat{x}_n \]
Why Study Sampling and Interpolation Further?

- Real-world sampling not pure mathematical idealization
  - Don’t/can’t sample at one point
  - Causal, non-ideal filters
- Many practical architectures different from classical structure
  - Multichannel, time-interleaved
- Most information acquisition is intimately related to sampling
  - Digital photography
  - Computational imaging (magnetic resonance, space-from-time, ultrasound, computed tomography, synthetic aperture radar, . . . )
  - Reflection seismology, acoustic tomography, . . .

Goals from this lecture:

- Understand classical sampling as a special case of a Hilbert space theory
- Gain a generalizable understanding of sampling
Why Study Sampling and Interpolation Further?

- Real-world sampling not pure mathematical idealization
  - Don’t/can’t sample at one point
  - Causal, non-ideal filters

- Many practical architectures different from classical structure
  - Multichannel, time-interleaved

- Most information acquisition is intimately related to sampling
  - Digital photography
  - Computational imaging (magnetic resonance, space-from-time, ultrasound, computed tomography, synthetic aperture radar, ...)
  - Reflection seismology, acoustic tomography, ...

Goals from this lecture:

- Understand classical sampling as a special case of a Hilbert space theory
- Gain a generalizable understanding of sampling
Why Study Sampling and Interpolation Further?

- Real-world sampling not pure mathematical idealization
  - Don’t/can’t sample at one point
  - Causal, non-ideal filters
- Many practical architectures different from classical structure
  - Multichannel, time-interleaved
- Most information acquisition is intimately related to sampling
  - Digital photography
  - Computational imaging (magnetic resonance, space-from-time, ultrasound, computed tomography, synthetic aperture radar, . . .)
  - Reflection seismology, acoustic tomography, . . .

Goals from this lecture:

- Understand classical sampling as a special case of a Hilbert space theory
- Gain a generalizable understanding of sampling
Why Study Sampling and Interpolation Further?

- Real-world sampling not pure mathematical idealization
  - Don’t/can’t sample at one point
  - Causal, non-ideal filters
- Many practical architectures different from classical structure
  - Multichannel, time-interleaved
- Most information acquisition is intimately related to sampling
  - Digital photography
  - Computational imaging (magnetic resonance, space-from-time, ultrasound, computed tomography, synthetic aperture radar, ...)
  - Reflection seismology, acoustic tomography, ...

Goals from this lecture:

- Understand classical sampling as a special case of a Hilbert space theory
- Gain a generalizable understanding of sampling
Bandwidth

Definition (Bandwidth of sequence)

A sequence \( x \) is \textit{bandlimited} when there exists \( \omega_0 \in [0, 2\pi) \) such that the discrete-time Fourier transform \( X \) satisfies

\[
X(e^{j\omega}) = 0 \quad \text{for all } \omega \text{ with } |\omega| \in (\omega_0/2, \pi].
\]

The smallest such \( \omega_0 \) is called the \textit{bandwidth} of \( x \).
Bandwidth

**Definition (Bandwidth of function)**

A function $x$ is **bandlimited** when there exists $\omega_0 \in [0, \infty)$ such that the Fourier transform $X$ satisfies

$$X(\omega) = 0 \quad \text{for all } \omega \text{ with } |\omega| > \omega_0/2.$$ 

The smallest such $\omega_0$ is called the **bandwidth** of $x$.

**Definition (Bandlimited sets)**

The set of sequences in $\ell^2(\mathbb{Z})$ with bandwidth at most $\omega_0$ and the set of functions in $L^2(\mathbb{R})$ with bandwidth at most $\omega_0$ are denoted $\text{BL}[-\omega_0/2, \omega_0/2]$.

- If $\omega_0 < \omega_1$ then $\text{BL}[-\omega_0/2, \omega_0/2] \subset \text{BL}[-\omega_1/2, \omega_1/2]$.
- A bandlimited set is always a subspace (Subspace is closed in Hilbert space $\ell^2(\mathbb{Z})$ or $L^2(\mathbb{R})$).
Bandwidth

**Definition (Bandwidth of function)**

A function $x$ is **bandlimited** when there exists $\omega_0 \in [0, \infty)$ such that the Fourier transform $X$ satisfies

$$X(\omega) = 0 \quad \text{for all } \omega \text{ with } |\omega| > \omega_0/2.$$ 

The smallest such $\omega_0$ is called the **bandwidth** of $x$.

**Definition (Bandlimited sets)**

The set of sequences in $\ell^2(\mathbb{Z})$ with bandwidth at most $\omega_0$ and the set of functions in $L^2(\mathbb{R})$ with bandwidth at most $\omega_0$ are denoted $\text{BL}[−\omega_0/2, \omega_0/2]$.

- If $\omega_0 < \omega_1$ then $\text{BL}[−\omega_0/2, \omega_0/2] \subset \text{BL}[−\omega_1/2, \omega_1/2]$. 
- A bandlimited set is always a subspace of $\ell^2(\mathbb{Z})$ or $L^2(\mathbb{R})$. 
Bandwidth

**Definition (Bandlimited function)**

A function $x$ is **bandlimited** when there exists $\omega_0 \in [0, \infty)$ such that the Fourier transform $X$ satisfies

$$X(\omega) = 0 \quad \text{for all } \omega \text{ with } |\omega| > \omega_0/2.$$  

The smallest such $\omega_0$ is called the **bandwidth** of $x$.

**Definition (Bandlimited sets)**

The set of sequences in $l^2(\mathbb{Z})$ with bandwidth at most $\omega_0$ and the set of functions in $L^2(\mathbb{R})$ with bandwidth at most $\omega_0$ are denoted $\text{BL}[-\omega_0/2, \omega_0/2]$.

- If $\omega_0 < \omega_1$ then $\text{BL}[-\omega_0/2, \omega_0/2] \subset \text{BL}[-\omega_1/2, \omega_1/2]$.

- A bandlimited set is always a subspace.
  - Subspace is closed in Hilbert space $l^2(\mathbb{Z})$ or $L^2(\mathbb{R})$. 

Classical Sampling

Recall: \( \text{sinc}(t) = \begin{cases} \frac{\sin(t)}{t}, & \text{for } t \neq 0; \\ 1, & \text{for } t = 0 \end{cases} \)

**Theorem (Sampling theorem)**

Let \( x \) be a function and let \( T > 0 \). Define

\[
\hat{x}(t) = \sum_{n \in \mathbb{Z}} x(nT) \text{sinc}(\frac{\pi}{T}(t - nT)).
\]

If \( x \in \text{BL}[\omega_0 / 2, \omega_0 / 2] \) with \( \omega_0 \leq 2\pi / T \), then \( \hat{x} = x \).

- Exact recovery for (sufficiently) bandlimited signals
- Nyquist period for bandwidth \( \omega_0 \): \( T = 2\pi / \omega_0 \)
- Nyquist rate for bandwidth \( \omega_0 \): \( T^{-1} = \omega_0 / 2\pi \)
- Easier in cycles/sec rather than radians/sec:
  
  Need two samples per cycle of fastest component
History

Many names:
- Shannon sampling theorem
- Nyquist–Shannon sampling theorem
- Nyquist–Shannon–Kotel’nikov sampling theorem
- Whittaker–Shannon–Kotel’nikov sampling theorem
- Whittaker–Nyquist–Shannon–Kotel’nikov sampling theorem

Well-known people associated with sampling (but less often so):
- Cauchy (1841) – apparently not true
- Borel (1897)
- de la Vallée Poussin (1908)
- E. T. Whittaker (1915)
- J. M. Whittaker (1927)
- Gabor (1946)

Less-known people, almost lost to history:
- Ogura (1920)
- Küpfmüller (Küpfmüller filter)
- Raabe [PhD 1939] (assistant to Küpfmüller)
- Someya (1949)
- Weston (1949)

Best approach: use no names?
- Cardinal Theorem of interpolation theory
History

Many names:
- Shannon sampling theorem
- Nyquist–Shannon sampling theorem
- Nyquist–Shannon–Kotel’nikov sampling theorem
- Whittaker–Shannon–Kotel’nikov sampling theorem
- Whittaker–Nyquist–Shannon–Kotel’nikov sampling theorem

Well-known people associated with sampling (but less often so):
- Cauchy (1841) – apparently not true
- Borel (1897)
- de la Vallée Poussin (1908)
- E. T. Whittaker (1915)
- J. M. Whittaker (1927)
- Gabor (1946)

Less-known people, almost lost to history:
- Ogura (1920)
- Küpfmüller (Küpfmüller filter)
- Raabe [PhD 1939] (assistant to Küpfmüller)
- Someya (1949)
- Weston (1949)

Best approach: use no names?
- Cardinal Theorem of interpolation theory
History

Many names:
- Shannon sampling theorem
- Nyquist–Shannon sampling theorem
- Nyquist–Shannon–Kotel’nikov sampling theorem
- Whittaker–Shannon–Kotel’nikov sampling theorem
- Whittaker–Nyquist–Shannon–Kotel’nikov sampling theorem

Well-known people associated with sampling (but less often so):
- Cauchy (1841) – apparently not true
- Borel (1897)
- de la Vallée Poussin (1908)
- E. T. Whittaker (1915)
- J. M. Whittaker (1927)
- Gabor (1946)

Less-known people, almost lost to history:
- Ogura (1920)
- Küpfmüller (Küpfmüller filter)
- Raabe [PhD 1939] (assistant to Küpfmüller)
- Someya (1949)
- Weston (1949)

Best approach: use no names?
- Cardinal Theorem of interpolation theory
History

Many names:
- Shannon sampling theorem
- Nyquist–Shannon sampling theorem
- Nyquist–Shannon–Kotel’nikov sampling theorem
- Whittaker–Shannon–Kotel’nikov sampling theorem
- Whittaker–Nyquist–Shannon–Kotel’nikov sampling theorem

Well-known people associated with sampling (but less often so):
- Cauchy (1841) – apparently not true
- Borel (1897)
- de la Vallée Poussin (1908)
- E. T. Whittaker (1915)
- J. M. Whittaker (1927)
- Gabor (1946)

Less-known people, almost lost to history:
- Ogura (1920)
- Küpfmüller (Küpfmüller filter)
- Raabe [PhD 1939] (assistant to Küpfmüller)
- Someya (1949)
- Weston (1949)

Best approach: use no names?
- Cardinal Theorem of interpolation theory
Sampling theorem: Conventional justification

- Let \( \tilde{x}(t) = \sum_{n \in \mathbb{Z}} x(nT) g(t - nT) \)  
  [will deduce that \( g \) should be sinc]
Sampling theorem: Conventional justification

- Let $\hat{x}(t) = \sum_{n \in \mathbb{Z}} x(nT) g(t - nT)$ [will deduce that $g$ should be sinc]
- Since $x(nT)g(t - nT) = \int_{-\infty}^{\infty} x(\tau) g(t - \tau) \delta(\tau - nT) \, d\tau$,

$$\hat{x}(t) = \int_{-\infty}^{\infty} g(t - \tau) x(\tau) \sum_{n \in \mathbb{Z}} \delta(\tau - nT) \, d\tau$$
Sampling theorem: Conventional justification

- Let $\hat{x}(t) = \sum_{n \in \mathbb{Z}} x(nT) g(t - nT)$  
  [will deduce that $g$ should be sinc]
- Since $x(nT) g(t - nT) = \int_{-\infty}^{\infty} x(\tau) g(t - \tau) \delta(\tau - nT) \, d\tau,$
  \[ \hat{x}(t) = \int_{-\infty}^{\infty} g(t - \tau) x(\tau) \sum_{n \in \mathbb{Z}} \delta(\tau - nT) \, d\tau \]

- Recall: Fourier transform of Dirac comb
  \[ \sum_{n \in \mathbb{Z}} \delta(t - nT) \xlongmapsto{\text{FT}} \frac{2\pi}{T} \sum_{k \in \mathbb{Z}} \delta \left( \omega - \frac{2\pi}{T}k \right) \]
Sampling theorem: Conventional justification

- Let \( \hat{x}(t) = \sum_{n \in \mathbb{Z}} x(nT) g(t - nT) \) [will deduce that \( g \) should be sinc]
- Since \( x(nT) g(t - nT) = \int_{-\infty}^{\infty} x(\tau) g(t - \tau) \delta(\tau - nT) \, d\tau \),
  \[
  \hat{x}(t) = \int_{-\infty}^{\infty} g(t - \tau) x(\tau) \sum_{n \in \mathbb{Z}} \delta(\tau - nT) \, d\tau
  \]
- Recall: Fourier transform of Dirac comb
  \[
  \sum_{n \in \mathbb{Z}} \delta(t - nT) \leftrightarrow \frac{2\pi}{T} \sum_{k \in \mathbb{Z}} \delta(\omega - \frac{2\pi}{T} k)
  \]
- Take Fourier transforms, using convolution theorem for right side:
  \[
  \hat{X}(\omega) = G(\omega) \frac{1}{T} \sum_{k \in \mathbb{Z}} X \left( \omega - \frac{2\pi}{T} k \right) H(\omega)
  \]
Sampling theorem: Conventional justification

\[ \hat{X}(\omega) = \frac{1}{T} G(\omega) \sum_{k \in \mathbb{Z}} X \left( \omega - \frac{2\pi}{T} k \right) \]

- Reconstruction \( \hat{X} \) has “spectral replication”
- How can we have \( \hat{X}(\omega) = X(\omega) \) for all \( \omega \)?
  - \( x \in \text{BL}[-\pi/T, \pi/T] \) implies replicas do not overlap
  - \( G(\omega) = \begin{cases} T, & \text{for } |\omega| < \pi/T; \\
 0, & \text{for } t = 0 \end{cases} \)
    selects “desired” replica with correct gain
- Shows recovery and deduces correctness of sinc interpolator
Dissatisfaction

- Mathematical rigor of derivation:
  \[
  \sum_{n \in \mathbb{Z}} \delta(t-nT) \quad \overset{\text{FT}}{\leftrightarrow} \quad \frac{2\pi}{T} \sum_{k \in \mathbb{Z}} \delta(\omega - \frac{2\pi}{T} k)
  \]
  
  - Not a convergent Fourier transform (in elementary sense)

- Mathematical plausibility of use:
  \[
  \hat{x}(t) = \sum_{n \in \mathbb{Z}} x(nT) \ \text{sinc}(\frac{\pi}{T}(t-nT))
  \]
  
  - At each \( t \), reconstruction is an infinite sum
  - Very slow decay of terms makes truncation accuracy poor

- Technological implementability:
  - Point measurements difficult to approximate physically
  - Causality of reconstruction
Dissatisfaction

- Mathematical rigor of derivation:
  \[ \sum_{n \in \mathbb{Z}} \delta(t - nT) \leftrightarrow \frac{2\pi}{T} \sum_{k \in \mathbb{Z}} \delta \left( \omega - \frac{2\pi}{T} k \right) \]
  \[ \implies \text{Not a convergent Fourier transform (in elementary sense)} \]

- Mathematical plausibility of use:
  \[ \hat{x}(t) = \sum_{n \in \mathbb{Z}} x(nT) \text{sinc}(\frac{\pi}{T}(t - nT)) \]
  \[ \implies \text{At each } t, \text{ reconstruction is an infinite sum} \]
  \[ \implies \text{Very slow decay of terms makes truncation accuracy poor} \]

- Technological implementability:
  \[ \implies \text{Point measurements difficult to approximate physically} \]
  \[ \implies \text{Causality of reconstruction} \]
Dissatisfaction

• Mathematical rigor of derivation:

\[ \sum_{n \in \mathbb{Z}} \delta(t - nT) \xrightarrow{FT} \frac{2\pi}{T} \sum_{k \in \mathbb{Z}} \delta(\omega - \frac{2\pi}{T}k) \]

  ▷ Not a convergent Fourier transform (in elementary sense)

• Mathematical plausibility of use:

\[ \hat{x}(t) = \sum_{n \in \mathbb{Z}} x(nT) \text{sinc}(\frac{\omega}{2\pi T}(t - nT)) \]

  ▷ At each \( t \), reconstruction is an infinite sum
  ▷ Very slow decay of terms makes truncation accuracy poor

• Technological implementability:
  ▷ Point measurements difficult to approximate physically
  ▷ Causality of reconstruction
Operator view: Interpolation

**Definition (Interpolation operator)**

For a fixed positive $T$ and interpolation postfilter $g(t)$, let $\Phi : l^2(\mathbb{Z}) \to L^2(\mathbb{R})$ be given by

$$
(\Phi y)(t) = \sum_{n \in \mathbb{Z}} y_n g(t - n T), \quad t \in \mathbb{R}
$$

- Generalizes sinc interpolation
- For simplicity, we will consider only $T = 1$: $(\Phi y)(t) = \sum_{n \in \mathbb{Z}} y_n g(t - n)$
Reconstruction space

Range of interpolation operator has special form

**Definition (Shift-invariant subspace of $L^2(\mathbb{R})$)**

A subspace $W \subseteq L^2(\mathbb{R})$ is a **shift-invariant subspace** with respect to shift $T \in \mathbb{R}^+$ when $x(t) \in W$ implies $x(t - kT) \in W$ for every integer $k$. In addition, $w \in L^2(\mathbb{R})$ is called a **generator** of $W$ when $W = \operatorname{span}(\{w(t - kT)\}_{k \in \mathbb{Z}})$.

- Range of $\Phi$ is a shift-invariant subspace generated by $g$
Definition (Sampling operator)

For a fixed positive \( T \) and sampling prefilter \( g^*(-t) \), let \( \Phi^* : L^2(\mathbb{R}) \to \ell^2(\mathbb{Z}) \) be given by

\[
(\Phi^* x)_n = \langle x(t), g(t - nT) \rangle_t, \quad n \in \mathbb{Z}
\]

- For simplicity, we will consider only \( T = 1 \):

\[
(\Phi^* x)_n = \langle x(t), g(t - n) \rangle_t
\]
Adjoint relationship between sampling and interpolation

**Theorem**

*Sampling and interpolation operators are adjoints*

Let \( x \in \mathcal{L}^2(\mathbb{R}) \) and \( y \in \ell^2(\mathbb{Z}) \)

\[
\langle \Phi^* x, y \rangle = \sum_{n \in \mathbb{Z}} \langle x(t), g(t - n) \rangle y_n^*
\]

\[
= \sum_{n \in \mathbb{Z}} \left( \int_{-\infty}^{\infty} x(t) g^*(t - n) \, dt \right) y_n^*
\]

\[
= \int_{-\infty}^{\infty} x(t) \left( \sum_{n \in \mathbb{Z}} y_n^* g(t - n) \right) \, dt
\]

\[
= \int_{-\infty}^{\infty} x(t) \left( \sum_{n \in \mathbb{Z}} y_n^* g(t - n) \right)^* \, dt
\]

\[
= \langle x, \Phi y \rangle
\]
Adjoint relationship between sampling and interpolation

**Theorem**

Sampling and interpolation operators are adjoints

Let $x \in L^2(\mathbb{R})$ and $y \in \ell^2(\mathbb{Z})$

$$
\langle \Phi^* x, y \rangle = \sum_{n \in \mathbb{Z}} \langle x(t), g(t - n) \rangle; y_n^*
$$

$$
= \sum_{n \in \mathbb{Z}} \left( \int_{-\infty}^{\infty} x(t) g^*(t - n) \, dt \right) y_n^*
$$

$$
= \int_{-\infty}^{\infty} x(t) \left( \sum_{n \in \mathbb{Z}} g^*(t - n) y_n^* \right) \, dt
$$

$$
= \int_{-\infty}^{\infty} x(t) \left( \sum_{n \in \mathbb{Z}} y_n g(t - n) \right)^* \, dt
$$

$$
= \langle x, \Phi y \rangle
$$
Relationships between sampling and interpolation

Sampling followed by interpolation: $\hat{x} = \Phi \Phi^* x$

- $\hat{x}$ is best approximation of $x$ within shift-invariant subspace generated by $g$ if
- $P = \Phi \Phi^*$ is an orthogonal projection operator
- $P$ is automatically self-adjoint: $P^* = (\Phi \Phi^*)^* = P$
- Need $P$ idempotent: $P^2 = \Phi \Phi^* \Phi \Phi^* = P$
- Require $\Phi^* \Phi = I$ \implies study interpolation followed by sampling
Relationships between sampling and interpolation

Sampling followed by interpolation: \( \hat{x} = \Phi \Phi^* x \)

- \( \hat{x} \) is best approximation of \( x \) within shift-invariant subspace generated by \( g \) if 
  \( P = \Phi \Phi^* \) is an orthogonal projection operator
- \( P \) is automatically self-adjoint: \( P^* = (\Phi \Phi^*)^* = P \)
- Need \( P \) idempotent: \( P^2 = \Phi \Phi^* \Phi \Phi^* = P \)
- Require \( \Phi^* \Phi = I \)  \( \Rightarrow \) study interpolation followed by sampling
Relationships between sampling and interpolation

Sampling followed by interpolation: \( \hat{x} = \Phi \Phi^* x \)

\[ x(t) \xrightarrow{g^*(-t)} y_n \xrightarrow{\uparrow \uparrow} g(t) \xrightarrow{\downarrow \downarrow} \hat{x}(t) \]

- \( \hat{x} \) is best approximation of \( x \) within shift-invariant subspace generated by \( g \) if
  - \( P = \Phi \Phi^* \) is an orthogonal projection operator
- \( P \) is automatically self-adjoint: \( P^* = (\Phi \Phi^*)^* = P \)
- Need \( P \) idempotent: \( P^2 = \Phi \Phi^* \Phi \Phi^* = P \)
- Require \( \Phi^* \Phi = I \) \( \Rightarrow \) study interpolation followed by sampling
Relationships between sampling and interpolation

Sampling followed by interpolation: \( \hat{x} = \Phi \Phi^* x \)

- \( \hat{x} \) is best approximation of \( x \) within shift-invariant subspace generated by \( g \) if \( P = \Phi \Phi^* \) is an orthogonal projection operator
- \( P \) is automatically self-adjoint: \( P^* = (\Phi \Phi^*)^* = P \)
- Need \( P \) idempotent: \( P^2 = \Phi \Phi^* \Phi \Phi^* = P \)
- Require \( \Phi^* \Phi = I \) \( \Rightarrow \) study interpolation followed by sampling
Relationships between sampling and interpolation

Interpolation followed by sampling: \( \hat{y} = \Phi^* \Phi y \)

- Consider output due to input \( y = \delta \)
  \[ \hat{y}_n = \langle g(t - n), g(t) \rangle \]
- Shifting input shifts output
- \( \Phi^* \Phi = I \) if and only if \( \langle g(t - n), g(t) \rangle = \delta_n \)
Relationships between sampling and interpolation

Interpolation followed by sampling: $\hat{y} = \Phi^* \Phi y$

- Consider output due to input $y = \delta$
  \[ \hat{y}_n = \langle g(t-n), g(t) \rangle_t \]

- Shifting input shifts output
- $\Phi^* \Phi = I$ if and only if $\langle g(t-n), g(t) \rangle_t = \delta_n$
Relationships between sampling and interpolation

Interpolation followed by sampling: \( \hat{y} = \Phi^* \Phi y \)

- Consider output due to input \( y = \delta \)
  \[ \hat{y}_n = \langle g(t-n), g(t) \rangle \]

- Shifting input shifts output
- \( \Phi^* \Phi = I \) if and only if \( \langle g(t-n), g(t) \rangle = \delta_n \)
Sampling for shift-invariant subspaces

**Theorem**

Let $g$ be orthogonal to its integer shifts: $\langle g(t - n), g(t) \rangle_t = \delta_n$. The system

$$x(t) \xrightarrow{g^*(-t)} \frac{y_n}{T} \xrightarrow{g(t)} \hat{x}(t)$$

yields $\hat{x} = P \ x$ where $P$ is the orthogonal projection operator onto the shift-invariant subspace $S$ generated by $g$.

**Corollaries:**
- If $x \in S$, then $x$ is recovered exactly from samples $y$
- If $x \notin S$, then $\hat{x}$ is the best approximation of $x$ in $S$
Reinterpreting classical sampling

\[ x(t) \xrightarrow{g^*(-t)} \frac{\bar{Y}}{\bar{y}} \xrightarrow{\bar{g}(t)} \hat{x}(t) \]

Case of \( g(t) = \text{sinc}(\pi t) \)

- \( \text{sinc}(\pi t) \) is orthogonal to its integer shifts
  - Immediately, orthogonal projection property holds

- Prefilter bandlimits ("anti-aliasing")
- \( g^*(-t) = g(t) \)
Reinterpreting classical sampling

$x(t) \rightarrow g^*(-t) \rightarrow Y_n \rightarrow g(t) \rightarrow \hat{x}(t)$

Case of $g(t) = \text{sinc}(\pi t)$

- $\text{sinc}(\pi t)$ is orthogonal to its integer shifts
  - Immediately, orthogonal projection property holds
- Prefilter bandlimits ("anti-aliasing")
- $g^*(-t) = g(t)$
Discrete-time version (downsampling)

**Definition (Sampling operator)**

For a fixed positive $N$ and sampling filter $g$, let $\Phi^* : \ell^2(\mathbb{Z}) \to \ell^2(\mathbb{Z})$ be given by

$$(\Phi^* x)_k = \langle x_n, g_{n-kN} \rangle, \quad k \in \mathbb{Z}$$

**Definition (Interpolation operator)**

For a fixed positive $N$ and interpolation filter $g$, let $\Phi : \ell^2(\mathbb{Z}) \to \ell^2(\mathbb{Z})$ be given by

$$(\Phi y)_n = \sum_{k \in \mathbb{Z}} y_k g_{n-kN}, \quad n \in \mathbb{Z}$$
Discrete-time version (downsampling)

**Definition (Shift-invariant subspace of $\ell^2(\mathbb{Z})$)**

A subspace $W \subset \ell^2(\mathbb{Z})$ is a shift-invariant subspace with respect to shift $L \in \mathbb{Z}^+$ when $x_n \in W$ implies $x_{n-kL} \in W$ for every integer $k$. In addition, $w \in \ell^2(\mathbb{Z})$ is called a generator of $W$ when $W = \text{span}(\{w_{n-kL}\}_{k \in \mathbb{Z}})$.

**Theorem**

Let $g$ be orthogonal to its shifts by multiples of $N$: $(g_{n-kN}, g_n)_n = \delta_k$. The system

\[
\begin{array}{c}
x_n \\
g^*_{-n} \downarrow N \\
y_n \downarrow N \\
g_n \\
\end{array} \rightarrow \tilde{x}_n
\]

yields $\tilde{x} = P x$ where $P$ is the orthogonal projection operator onto the shift-invariant subspace $S$ generated by $g$ with shift $N$. 
Geometric interpretation of general case

\[ x(t) \xrightarrow{\tilde{g}(t)} \tau_n \xrightarrow{\Phi^*} y_n \xrightarrow{\Phi} \hat{x}(t) \]

- Sampling operator $\Phi^*$, interpolation operator $\Phi$
Geometric interpretation of general case

- Sampling operator \( \tilde{\Phi}^* \), interpolation operator \( \Phi \)
- \( \Phi \tilde{\Phi}^* \) generally not self-adjoint, but can still be a projection operator
Geometric interpretation of general case

- Sampling operator $\tilde{\Phi}^*$, interpolation operator $\Phi$
- $\Phi \tilde{\Phi}^*$ generally not self-adjoint, but can still be a projection operator
- Let $\tilde{S} = \mathcal{N}(\tilde{\Phi}^*)^\perp$ and $S = \mathcal{R}(\Phi)$
- Check $\tilde{\Phi}^* \Phi = I$ for an oblique projection to $S$
Variations

Multichannel sampling

\[ x(t) \rightarrow \tilde{g}_0(t) \quad \frac{2T}{\_} \rightarrow y_{0,n} \]
\[ \tilde{g}_1(t) \quad \frac{2T}{\_} \rightarrow y_{1,n} \]

- Sample signal and derivatives
- Periodic nonuniform sampling (time-interleaved ADC)

- Many inverse problems have linear forward models, perhaps not shift-invariant
- Similar subspace geometry holds
- Provides foundation for recent sampling methods based on semilinear signal models (finite rate of innovation)
Variations

Multichannel sampling

Sample signal and derivatives
Periodic nonuniform sampling (time-interleaved ADC)

Many inverse problems have linear forward models, perhaps not shift-invariant
Similar subspace geometry holds
Provides foundation for recent sampling methods based on semilinear signal models (finite rate of innovation)
Summary

- **Adoints**
  - Time reversal between sampling and interpolation

- **Subspaces**
  - Shift-invariant, range of interpolator $\Phi$
  - Null space of sampler $\Phi^*$

- **Projection**
  - $\Phi \Phi^*$ always self adjoint
  - $\Phi^* \Phi = I$ implies $\Phi \Phi^*$ is a projection operator
  - Together, orthogonal projection operator, best approximation

- **Basis expansions**
  - Sampling produces analysis coefficients for basis expansion
  - Interpolation synthesizes from expansion coefficients
Textbooks

Two books:

- M. Vetterli, J. Kovačević, and V. K. Goyal, *Foundations of Signal Processing*
- J. Kovačević, V. K. Goyal, and M. Vetterli, *Fourier and Wavelet Signal Processing*

Manuscripts distributed in draft form online (originally as a single volume and with some variations in titles) since 2010 at http://www.fourierandwavelets.org

- Free, online versions have gray scale images, no PDF hyperlinks, no exercises with solutions or exercises
Textbooks

*Foundations of Signal Processing*
- On Rainbows and Spectra
- From Euclid to Hilbert
- Sequences and Discrete-Time Systems
- Functions and Continuous-Time Systems
- Sampling and Interpolation
- Approximation and Compression
- Localization and Uncertainty

Features:
- About 640 pages illustrated with more than 200 figures
- More than 200 exercises (more than 30 with solutions within the text)
- Solutions manual for instructors
- Summary tables, guides to further reading, historical notes
Textbooks

Fourier and Wavelet Signal Processing

- Filter Banks: Building Blocks of Time-Frequency Expansions
- Local Fourier Bases on Sequences
- Wavelet Bases on Sequences
- Local Fourier and Wavelet Frames on Sequences
- Local Fourier Transforms, Frames and Bases on Functions
- Wavelet Bases, Frames and Transforms on Functions
- Approximation, Estimation, and Compression
Prerequisites

- Textbook is a mostly self-contained treatment
- Mathematical maturity
  - Mechanical use of calculus not enough
  - Sophistication to read and write precise mathematical statements needed
    (or could be learned here)
- Linear algebra
  - Basic facility with matrix algebra very useful
  - Abstract view built carefully within the book
- Probability
  - Basic background (e.g., first half of *Introduction to Probability* by Bertsekas and Tsitsiklis) needed (else all stochastic material could be skipped)
- Signals and systems
  - Basic background (e.g., *Signals and Systems* by Oppenheim and Willsky)
    helpful but not necessary
Solutions manual

Convolution of Derivative and Primitive

Let $h$ and $x$ be differentiable functions, and let

$$h^{(1)}(t) = \int_{-\infty}^{t} h(\tau) \, d\tau \quad \text{and} \quad x^{(1)}(t) = \int_{-\infty}^{t} x(\tau) \, d\tau$$

be their primitives. Give a sufficient condition for $h * x = h^{(1)} * x'$ based on integration by parts.
Solutions manual

From the definition of convolution, (4.35),

\[ (h * x)(t) = \int_{-\infty}^{\infty} x(\tau) h(t - \tau) d\tau. \]

We wish to apply definite integration by parts, (2.204b), to get to a form involving \( h^{(1)} \) and \( x' \). With the associations

\[ u(\tau) = x(\tau) \quad \text{and} \quad v'(\tau) = h(t - \tau), \]

we obtain

\[ u'(\tau) = x'(\tau) \quad \text{and} \quad v(\tau) = -h^{(1)}(t - \tau). \]

Substituting these into (2.204b) gives

\[ (h * x)(t) = -x(\tau) h^{(1)}(t - \tau) \bigg|_{\tau=\infty}^{\tau=-\infty} + \int_{-\infty}^{\infty} h^{(1)}(t - \tau) x'(\tau) d\tau. \]

(1)

This yields the desired result of

\[ (h * x)(t) = (h^{(1)} * x')(t), \quad \text{for all } t \in \mathbb{R}, \]

provided that the first term of (1) is zero:

\[ \lim_{\tau \to \pm\infty} x(\tau) h^{(1)}(t - \tau) = 0, \quad \text{for all } t \in \mathbb{R}. \]
Mathematica figures and interactive CDF documents

Figure 3.8: Truncated DTFT of the sinc sequence, illustrating the Gibbs phenomenon. Shown are $|X_N(e^{j\omega})|$ from (3.84) with different $N$. Observe how oscillations narrow from (a) to (c), but their amplitude remains constant (the topmost grid line in every plot), $1.089 \sqrt{2}$.

- Computable Document Format (Wolfram, 2011)
- Free standalone CDF Player and browser plugins

Why rethink how signal processing is taught?

- Signal processing is an essential and vibrant field
- Geometry is key to gaining intuition and understanding

Thank you for your interest