Mathematics for Machine Learning

Vazgen Mikayelyan

September 19, 2020

FAST DISCOVERING THE FUTURE

Signal Processing

Definition of Signal Processing

- What is a signal?

Definition of Signal Processing

- What is a signal?
  - A signal is mathematically just a function.

Definition of Signal Processing

- What is a signal?
    - A signal is mathematically just a function.
- What is a signal processing?

Definition of Signal Processing

- What is a signal?
  - A signal is mathematically just a function.
- What is a signal processing?
  - Convert one signal to another.

Definition of Signal Processing

- What is a signal?
  - A signal is mathematically just a function.
- What is a signal processing?
  - Convert one signal to another.
  - Information extraction and interpretation.

Fourier Transform

Fourier Transform

Definition

Let $f:\mathbb{R}\to\mathbb{C}$. The Fourier transform of f is defined as follows

$$\mathcal{F}\left[f\right](x)=\hat{f}(x)=\int\limits_{-\infty}^{+\infty}f(t)\,e^{-2\pi ixt}dt.$$

Fourier Transform

Definition

Let $f : \mathbb{R} \to \mathbb{C}$. The Fourier transform of f is defined as follows

$$\mathcal{F}[f](x) = \hat{f}(x) = \int_{-\infty}^{+\infty} f(t)\, e^{-2\pi i x t} dt.$$

Definition

For all $p \in [1, \infty)$ denote by L^p the set of all functions f, such that

$$\|f\|_p = \left(\int_{-\infty}^{+\infty} |f(x)|^p \, dx \right)^{\frac{1}{p}} < +\infty.$$

Fourier Transform

Definition

Let $f : \mathbb{R} \to \mathbb{C}$*. The Fourier transform of* f *is defined as follows*

$$\mathcal{F}[f](x) = \hat{f}(x) = \int_{-\infty}^{+\infty} f(t)\, e^{-2\pi i x t} dt.$$

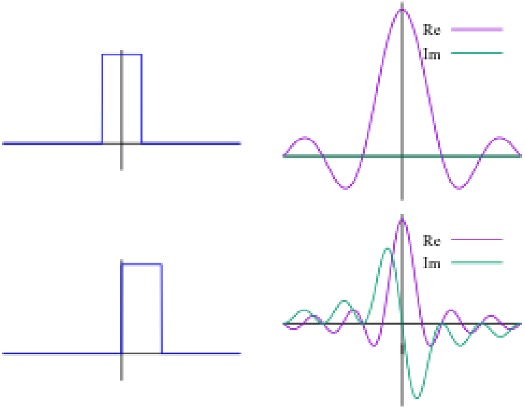
Definition

For all $p \in [1, \infty)$ *denote by* L^p *the set of all functions* f*, such that*

$$\|f\|_p = \left(\int_{-\infty}^{+\infty} |f(x)|^p \, dx \right)^{\frac{1}{p}} < +\infty.$$

Theorem

If $f \in L^1$*, then the Fourier transform of* f *exists.*


Re
Im
Re
Im

Fourier Transform

Properties

1. $\mathcal{F}\left[\alpha f + \beta g\right] = \alpha \mathcal{F}\left[f\right] + \beta \mathcal{F}\left[g\right]$, $\alpha, \beta \in \mathbb{R}$.

Fourier Transform

Properties

1. $\mathcal{F}\left[\alpha f + \beta g\right] = \alpha \mathcal{F}\left[f\right] + \beta \mathcal{F}\left[g\right]$, $\alpha, \beta \in \mathbb{R}$.
2. If $g\left(t\right) = f\left(t - a\right)$, $t, a \in \mathbb{R}$, then $\hat{g}\left(x\right) = e^{-2\pi i a x} \hat{f}\left(x\right)$, $x \in \mathbb{R}$.

Fourier Transform

Properties

1. $\mathcal{F}\left[\alpha f+\beta g\right]=\alpha\mathcal{F}\left[f\right]+\beta\mathcal{F}\left[g\right]$, $\alpha,\beta\in\mathbb{R}$.
2. If $g\left(t\right)=f\left(t-a\right)$, $t,a\in\mathbb{R}$, then $\hat{g}\left(x\right)=e^{-2\pi iax}\hat{f}\left(x\right)$, $x\in\mathbb{R}$.
3. If $g\left(t\right)=f\left(at\right)$, $t,a\in\mathbb{R}$, $a\neq 0$, then $\hat{g}\left(x\right)=\frac{1}{a}\hat{f}\left(\frac{x}{a}\right)$, $x\in\mathbb{R}$.

Fourier Transform

Properties

1. $\mathcal{F}\left[\alpha f+\beta g\right]=\alpha\mathcal{F}\left[f\right]+\beta\mathcal{F}\left[g\right]$, $\alpha,\beta\in\mathbb{R}$.
2. If $g\left(t\right)=f\left(t-a\right)$, $t,a\in\mathbb{R}$, then $\hat{g}\left(x\right)=e^{-2\pi iax}\hat{f}\left(x\right)$, $x\in\mathbb{R}$.
3. If $g\left(t\right)=f\left(at\right)$, $t,a\in\mathbb{R}$, $a\neq 0$, then $\hat{g}\left(x\right)=\frac{1}{a}\hat{f}\left(\frac{x}{a}\right)$, $x\in\mathbb{R}$.
4. If $g=\overline{f}$, then $\hat{g}\left(x\right)=\overline{\hat{f}\left(-x\right)}$.

Fourier Transform

Properties

1. $\mathcal{F}\left[\alpha f+\beta g\right]=\alpha\mathcal{F}\left[f\right]+\beta\mathcal{F}\left[g\right]$, $\alpha,\beta\in\mathbb{R}$.
2. If $g\left(t\right)=f\left(t-a\right)$, $t,a\in\mathbb{R}$, then $\hat{g}\left(x\right)=e^{-2\pi iax}\hat{f}\left(x\right)$, $x\in\mathbb{R}$.
3. If $g\left(t\right)=f\left(at\right)$, $t,a\in\mathbb{R}$, $a\neq 0$, then $\hat{g}\left(x\right)=\frac{1}{a}\hat{f}\left(\frac{x}{a}\right)$, $x\in\mathbb{R}$.
4. If $g=\overline{f}$, then $\hat{g}\left(x\right)=\overline{\hat{f}\left(-x\right)}$.

Plancherel's Theorem

If $f\in L^1\cap L^2$, then $\hat{f}\in L^2$ and

$$\int\limits_{-\infty}^{+\infty}\left|\hat{f}\left(x\right)\right|^2dx=\int\limits_{-\infty}^{+\infty}\left|f\left(t\right)\right|^2dt.$$

Convolution

Convolution

Definition

Convolution of the functions $f, g : \mathbb{R} \to \mathbb{R}$, $f, g \in L^1$, *is defined as the integral of the product of the two functions after one is reversed and shifted:*

$$(f * g)(t) =: \int_{-\infty}^{+\infty} f(x)\, g(t-x)\, dx.$$

Convolution

Definition

Convolution of the functions $f, g : \mathbb{R} \to \mathbb{R}$, $f, g \in L^1$, is defined as the integral of the product of the two functions after one is reversed and shifted:

$$(f * g)(t) =: \int_{-\infty}^{+\infty} f(x)\, g(t-x)\, dx.$$

Properties

1. $f * g = g * f$,

Convolution

Definition

Convolution of the functions $f, g : \mathbb{R} \to \mathbb{R}$, $f, g \in L^1$, is defined as the integral of the product of the two functions after one is reversed and shifted:

$$(f * g)(t) =: \int_{-\infty}^{+\infty} f(x)\, g(t - x)\, dx.$$

Properties

1. $f * g = g * f$,
2. $(f * g) * h = f * (g * h)$,

Convolution

Young's Convolution Inequality

If $f \in L^p$, $g \in L^q$ and $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$, with $1 \leq p$ and $1 \leq q \leq r < \infty$, then $\|f * g\|_r \leq \|f\|_p \|g\|_q$.

Convolution

Young's Convolution Inequality

If $f \in L^p$, $g \in L^q$ and $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$, with $1 \leq p$ and $1 \leq q \leq r < \infty$, then $\|f * g\|_r \leq \|f\|_p \|g\|_q$.

Convolution Theorem

If $f, g \in L^1$, then $\mathcal{F}[f * g] = \mathcal{F}[f]\,\mathcal{F}[g]$.

Inverse Fourier Transform

Inverse Fourier Transform

Definition

Let $g : \mathbb{R} \to \mathbb{C}$. *The inverse Fourier transform of* f *is defined as follows*

$$\mathcal{F}^{-1}[g](x) = \int_{-\infty}^{+\infty} g(t)\, e^{2\pi i x t} dt.$$

Inverse Fourier Transform

Definition

Let $g : \mathbb{R} \to \mathbb{C}$. *The inverse Fourier transform of* f *is defined as follows*

$$\mathcal{F}^{-1}[g](x) = \int_{-\infty}^{+\infty} g(t)\, e^{2\pi i x t} dt.$$

Theorem

If $f, \hat{f} \in L^1$ *then* $\mathcal{F}^{-1}\left[\hat{f}\right] = f$, *almost everywhere on* $\mathbb{R}$.

Digital Signal Processing

Digital Signal Processing

- What does digital mean?

Digital Signal Processing

- What does digital mean?
- Discrete representation.

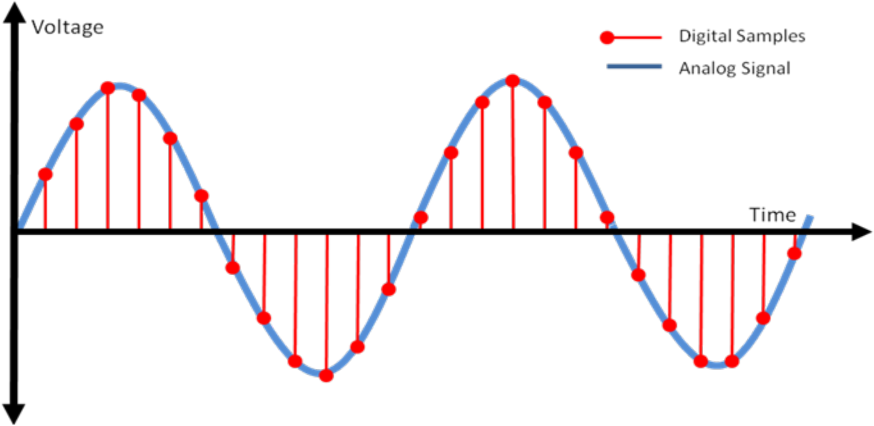
Digital Signal Processing

- What does digital mean?
- Discrete representation.
- Can be samples of a continuous signal.

Digital Signal Processing

- What does digital mean?
- Discrete representation.
- Can be samples of a continuous signal.
- Can be discrete by nature.

Example of Sampling



Examples

- **Audio recording:**

Examples

- **Audio recording:** an analog pressure wave is sampled and converted to a one-dimensional discrete-time signal.

Examples

- **Audio recording:** an analog pressure wave is sampled and converted to a one-dimensional discrete-time signal.
- **Photos:**

Examples

- **Audio recording:** an analog pressure wave is sampled and converted to a one-dimensional discrete-time signal.
- **Photos:** the analog scene of light is sampled using a CCD array and stored as a two-dimensional discrete-space signal

Examples

- **Audio recording:** an analog pressure wave is sampled and converted to a one-dimensional discrete-time signal.
- **Photos:** the analog scene of light is sampled using a CCD array and stored as a two-dimensional discrete-space signal
- **Ratings:**

Examples

- **Audio recording:** an analog pressure wave is sampled and converted to a one-dimensional discrete-time signal.
- **Photos:** the analog scene of light is sampled using a CCD array and stored as a two-dimensional discrete-space signal
- **Ratings:** for books (Goodreads), movies (Netflix), vacation rentals (Air bnb) are stored using the integers 0-5

SP and ML Examples

- Speaker Identification

SP and ML Examples

- Speaker Identification
- Speech to Text

SP and ML Examples

- Speaker Identification
- Speech to Text
- Source Separation

Basic Notation

- Continuous signals:

$$x(t) : \mathbb{R} \to \mathbb{R} \text{ or } x(t) : \mathbb{R} \to \mathbb{C}.$$

Basic Notation

- Continuous signals:

$$x(t) : \mathbb{R} \to \mathbb{R} \text{ or } x(t) : \mathbb{R} \to \mathbb{C}.$$

- Discrete signals:

$$x[n] : \mathbb{Z} \to \mathbb{R} \text{ or } x[n] : \mathbb{Z} \to \mathbb{C}.$$

Examples

- Delta Signal:

$$\delta_n = \begin{cases} 1, & \text{if } n = 0, \\ 0, & \text{if } n \neq 0 \end{cases}$$

Examples

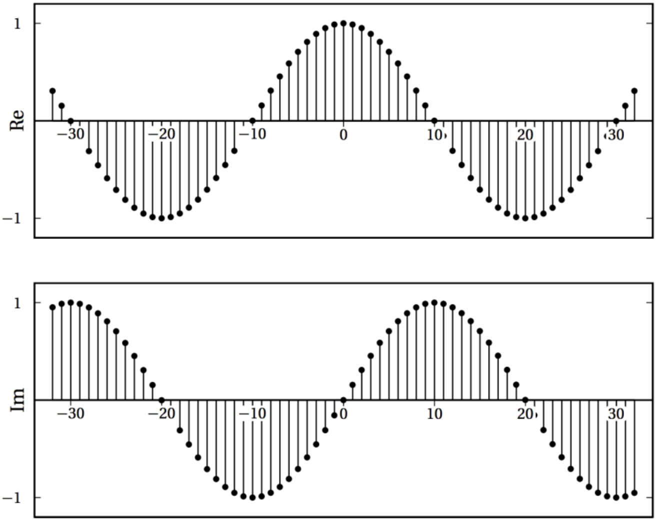
- Delta Signal:

$$\delta_n = \begin{cases} 1, & \text{if } n = 0, \\ 0, & \text{if } n \neq 0 \end{cases}$$

- Complex Exponential Signal:

$$x\left[n\right] = e^{i(\omega_0 n + \varphi)}$$

Complex Exponential Signal



Cosine Signal

$$x[n] = A\cos\left(\omega_0 n + \varphi\right)$$

Cosine Signal

$$x[n] = A\cos\left(\omega_0 n + \varphi\right) = \frac{A}{2}e^{i(\omega_0 n+\varphi)} + \frac{A}{2}e^{-i(\omega_0 n+\varphi)}$$

Cosine Signal

$$x[n] = A\cos(\omega_0 n + \varphi) = \frac{A}{2}e^{i(\omega_0 n + \varphi)} + \frac{A}{2}e^{-i(\omega_0 n + \varphi)}$$

- A is the amplitude.

Cosine Signal

$$x[n] = A\cos\left(\omega_0 n + \varphi\right) = \frac{A}{2}e^{i(\omega_0 n+\varphi)} + \frac{A}{2}e^{-i(\omega_0 n+\varphi)}$$

- A is the amplitude.
- φ is the phase in radians.

Cosine Signal

$$x[n] = A\cos\left(\omega_0 n + \varphi\right) = \frac{A}{2}e^{i(\omega_0 n+\varphi)} + \frac{A}{2}e^{-i(\omega_0 n+\varphi)}$$

- A is the amplitude.
- φ is the phase in radians.
- n is the sample number.

Cosine Signal

$$x[n] = A\cos\left(\omega_0 n + \varphi\right) = \frac{A}{2}e^{i(\omega_0 n+\varphi)} + \frac{A}{2}e^{-i(\omega_0 n+\varphi)}$$

- A is the amplitude.
- φ is the phase in radians.
- n is the sample number.
- ω_0 is the frequency in radians per sample.

Cosine Signal

$$x[n] = A\cos(\omega_0 n + \varphi) = \frac{A}{2}e^{i(\omega_0 n+\varphi)} + \frac{A}{2}e^{-i(\omega_0 n+\varphi)}$$

- A is the amplitude.
- φ is the phase in radians.
- n is the sample number.
- ω_0 is the frequency in radians per sample.
- Frequency in cycles per sample $f = \dfrac{\omega_0}{2\pi}$.

Periodicity

Definition

A discrete time signal is periodic if there is an $n_0 \in \mathbb{Z}$ *such that* $x[n] = x[n + n_0]$, *for all* $n \in \mathbb{Z}$.

Periodicity

Definition

A discrete time signal is periodic if there is an $n_0 \in \mathbb{Z}$ *such that* $x[n] = x[n + n_0]$, *for all* $n \in \mathbb{Z}$.

Theorem

A discrete-domain or discrete-time sinusoid is periodic if and only if its frequency ω_0 *is* π *times a rational number, that is*

$$\omega_0 = \frac{M}{N}\pi, M, N \in \mathbb{Z}.$$

Energy and Power

Definition

Energy of a discrete signal $x[n], n \in \mathbb{Z}$ *is defined as follows*

$$E_x = \sum_{n=-\infty}^{+\infty} |x[n]|^2 .$$

Energy and Power

Definition

Energy of a discrete signal $x[n], n \in \mathbb{Z}$ *is defined as follows*

$$E_x = \sum_{n=-\infty}^{+\infty} |x[n]|^2 .$$

Definition

Power of a discrete signal $x[n], n \in \mathbb{Z}$ *is defined as follows*

$$P_x = \lim_{N \to \infty} \frac{1}{2N+1} \sum_{n=-N}^{N} |x[n]|^2 .$$

Sampling

Let $x_a\left(t\right)$ be a analog signal. If we sample once every T seconds, then the value of the n-th number in the sequence is equal to

$$x\left[n\right] = x_a\left(nT\right).$$

Sampling

Let $x_a(t)$ be a analog signal. If we sample once every T seconds, then the value of the n-th number in the sequence is equal to

$$x[n] = x_a(nT).$$

- T is called the sampling period,

Sampling

Let $x_a(t)$ be a analog signal. If we sample once every T seconds, then the value of the n-th number in the sequence is equal to

$$x[n] = x_a(nT).$$

- T is called the sampling period,
- $\frac{1}{T}$ is called the sampling frequency.

Sampling Theorem

Definition

In signal processing aliasing is an effect that causes different signals to become indistinguishable when sampled.

Sampling Theorem

Definition

In signal processing aliasing is an effect that causes different signals to become indistinguishable when sampled.

Let us suppose that any analog signal can be represented as a sum of sinusoids of different amplitudes, frequencies and phases, i.e.

$$x_a(t) = \sum_{i=1}^{n} A_i \cos\left(2\pi F_i t + \theta_i\right).$$

Sampling Theorem

Definition

In signal processing aliasing is an effect that causes different signals to become indistinguishable when sampled.

Let us suppose that any analog signal can be represented as a sum of sinusoids of different amplitudes, frequencies and phases, i.e.

$$x_a(t) = \sum_{i=1}^{n} A_i \cos\left(2\pi F_i t + \theta_i\right).$$

All signals such that speech and video, lend themselves to such a represantation over any short time segment.

Sampling Theorem

Definition

In signal processing aliasing is an effect that causes different signals to become indistinguishable when sampled.

Let us suppose that any analog signal can be represented as a sum of sinusoids of different amplitudes, frequencies and phases, i.e.

$$x_a(t) = \sum_{i=1}^{n} A_i \cos\left(2\pi F_i t + \theta_i\right).$$

All signals such that speech and video, lend themselves to such a represantation over any short time segment. **For speech the maximum frequency is approximately 3400 Hz.**

Sampling Theorem

Nyquist–Shannon

If the highest frequency contained in an analog signal $x_a(t)$ is F_{max} and the signal is sampled at a rate $F_s > 2F_{max}$, then $x_a(t)$ can be exactly recovered from its sample values using the interpolation function

$$g(t) = \frac{\sin(2\pi F_{max} t)}{2\pi F_{max} t}.$$

Thus $x_a(t)$ can be expressed as

$$x_a(t) = \sum_{n=-\infty}^{+\infty} x_a\left(\frac{n}{F_s}\right) g\left(t - \frac{n}{F_s}\right).$$

Sampling Theorem

Nyquist–Shannon

If the highest frequency contained in an analog signal $x_a(t)$ is F_{max} and the signal is sampled at a rate $F_s > 2F_{max}$, then $x_a(t)$ can be exactly recovered from its sample values using the interpolation function

$$g(t) = \frac{\sin(2\pi F_{max} t)}{2\pi F_{max} t}.$$

Thus $x_a(t)$ can be expressed as

$$x_a(t) = \sum_{n=-\infty}^{+\infty} x_a\left(\frac{n}{F_s}\right) g\left(t - \frac{n}{F_s}\right).$$

The sampling rate $F_N = 2F_{max}$ is called the Nyquist rate.

Fourier Series

Fourier Series

Definition

The function $T : \mathbb{R} \to \mathbb{R}$ *is called a trigonometric polynomial of degree* n*, if there exist sequences of real numbers* $\{a_k\}_{k=0}^{n}$*,* $\{b_k\}_{k=1}^{n}$ *such that* $a_n^2 + b_n^2 > 0$ *and*

$$T(x) = a_0 + \sum_{k=1}^{n} (a_k \cos kx + b_k \sin kx).$$

Some Identities

1. $\int\limits_{-\pi}^{\pi} \sin mx dx = 0, \; m \in \mathbb{Z},$

Some Identities

1. $\int\limits_{-\pi}^{\pi} \sin mx dx = 0, \ m \in \mathbb{Z},$
2. $\int\limits_{-\pi}^{\pi} \cos mx dx = \begin{cases} 0, m \in \mathbb{Z} \setminus \{0\} \\ 2\pi, m = 0 \end{cases} \quad ,$

Some Identities

1. $$\int\limits_{-\pi}^{\pi} \sin mx dx = 0,\ m \in \mathbb{Z},$$
2. $$\int\limits_{-\pi}^{\pi} \cos mx dx = \begin{cases} 0, m \in \mathbb{Z} \setminus \{0\} \\ 2\pi, m = 0 \end{cases},$$
3. $$\int\limits_{-\pi}^{\pi} \sin mx \cos kx dx = 0,\ m, k \in \mathbb{N},$$

Some Identities

1. $$\int_{-\pi}^{\pi} \sin mx dx = 0,\ m \in \mathbb{Z},$$
2. $$\int_{-\pi}^{\pi} \cos mx dx = \begin{cases} 0, m \in \mathbb{Z} \setminus \{0\} \\ 2\pi, m = 0 \end{cases},$$
3. $$\int_{-\pi}^{\pi} \sin mx \cos kx dx = 0,\ m, k \in \mathbb{N},$$
4. $$\int_{-\pi}^{\pi} \sin mx \sin kx dx = \begin{cases} 0, m \neq k \\ \pi, m = k \end{cases},\ m, k \in \mathbb{N},$$

Some Identities

1. $$\int_{-\pi}^{\pi} \sin mx dx = 0, \; m \in \mathbb{Z},$$
2. $$\int_{-\pi}^{\pi} \cos mx dx = \begin{cases} 0, m \in \mathbb{Z} \setminus \{0\} \\ 2\pi, m = 0 \end{cases},$$
3. $$\int_{-\pi}^{\pi} \sin mx \cos kx dx = 0, \; m, k \in \mathbb{N},$$
4. $$\int_{-\pi}^{\pi} \sin mx \sin kx dx = \begin{cases} 0, m \neq k \\ \pi, m = k \end{cases}, \; m, k \in \mathbb{N},$$
5. $$\int_{-\pi}^{\pi} \cos mx \cos kx dx = \begin{cases} 0, m \neq k \\ \pi, m = k \end{cases}, \; m, k \in \mathbb{N}.$$