Fourier Transform: Disributions and Generalizations

Wen Perng

NTUEE

2025 May 15

As you have seen from the table of contents, this talk covers *a lot* of topics.

That's usually not a good thing for those who wants to dive into a topic. However, bear in mind, that the topics introduced are specially picked from all that the speaker has accumulated from the four-year study in NTUEE, through courses like: 信號與系統, 泛函分析與逼近理論, 時頻分析與小波轉換, 高等數位訊號處理, 密碼學, 量子資訊與計算, 量子資訊專題, and guest talks. It is definitely not expected for one to understand all of the material in one go.

Instead, this talk serves as an introduction, quickly going over many of the results and their ideas and intuitions but not the proofs. In hopes that one day, when one run into similar topics, one will feel especially familiar and will not be afraid of the mathematics.

Though I tried my best to make this talk self-contained, some prior knowledge are assumed.

1. The four Fourier[1] transforms of CTFT, CTFS, DTFT, DTFS.
2. What a vector is.
3. What is a unitary matrix.
4. What is an eigenvalue.
5. Modular arithmetic.

The latter four is not exactly necessary, but acquaintance with them ensures that you can get the most out of this talk. Most important of all, however, you should have an open mind to new ideas, and don't hesitate to interrupt if you have any questions.

[1] Pun intended.

In the mandatory course of Signals & Systems of NTUEE, we are introduced to:

1. CTFT:
$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{f}(j\omega)\mathrm{e}^{j\omega x}\mathrm{d}\omega \xrightarrow{\mathscr{F}} \hat{f}(j\omega) = \int_{-\infty}^{\infty} f(x)\mathrm{e}^{-j\omega x}\mathrm{d}x, \tag{1}$$
2. CTFS:
$$f(x) = \sum_{n=-\infty}^{\infty} \hat{f}_k\mathrm{e}^{jkx} \xrightarrow{\mathscr{F}} \hat{f}_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\mathrm{e}^{-jkx}\mathrm{d}x, \tag{2}$$
3. DTFT:
$$f[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} \hat{f}(\mathrm{e}^{j\omega})\mathrm{e}^{j\omega n}\mathrm{d}\omega \xrightarrow{\mathscr{F}} \hat{f}(\mathrm{e}^{j\omega}) = \sum_{n=-\infty}^{\infty} f[n]\mathrm{e}^{-j\omega n}, \tag{3}$$
4. DTFS:
$$f[n] = \sum_{k=0}^{N-1} \hat{f}(\mathrm{e}^{j\omega})\mathrm{e}^{i\frac{2\pi}{N}kx} \xrightarrow{\mathscr{F}} \hat{f}(k) = \frac{1}{N}\sum_{n=0}^{N-1} f[n]\mathrm{e}^{-i\frac{2\pi}{N}kn}. \tag{4}$$

We will use a different notation today:

1. CTFT: for $x \in \mathbb{R}$ and $k \in \mathbb{R}$,

$$f(x) = \int_{-\infty}^{\infty} \hat{f}(k) \mathrm{e}^{\mathrm{i}2\pi kx} \mathrm{d}k \xrightarrow{\mathscr{F}} \hat{f}(k) = \int_{-\infty}^{\infty} f(x) \mathrm{e}^{-\mathrm{i}2\pi kx} \mathrm{d}x, \tag{5}$$

2. CTFS: for $x \in \mathbb{R}$ and $k \in \mathbb{Z}$,

$$f(x) = \sum_{k=-\infty}^{\infty} \hat{f}(k) \mathrm{e}^{\mathrm{i}2\pi kx} \xrightarrow{\mathscr{F}} \hat{f}(k) = \int_{-1/2}^{1/2} f(x) \mathrm{e}^{-\mathrm{i}2\pi kx} \mathrm{d}x, \tag{6}$$

This notation makes many of the formulas way more beautiful and easy to write down.

We will use a different notation today:

3. DTFT: for $n \in \mathbb{Z}$ and $k \in \mathbb{R}$,

$$f(n) = \int_{-1/2}^{1/2} \hat{f}(k) \mathrm{e}^{\mathrm{i}2\pi kn} \mathrm{d}k \xrightarrow{\mathscr{F}} \hat{f}(k) = \sum_{n=-\infty}^{\infty} f(n) \mathrm{e}^{-i2\pi kn}, \tag{7}$$

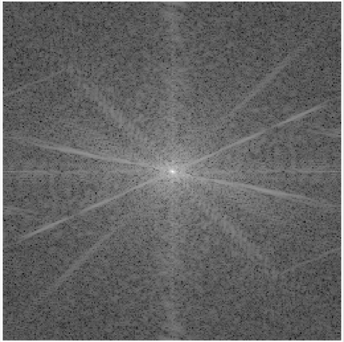
4. DTFS: for $n \in \mathbb{Z}$ and $k \in \mathbb{Z}$,

$$f(n) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \hat{f}(k) \mathrm{e}^{\mathrm{i}\frac{2\pi}{N}kn} \xrightarrow{\mathscr{F}} \hat{f}(k) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} f(n) \mathrm{e}^{-\mathrm{i}\frac{2\pi}{N}kn}. \tag{8}$$

This notation makes many of the formulas way more beautiful and easy to write down.

Fourier transform is really useful! Here are some examples

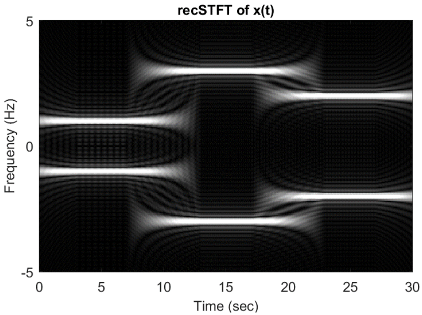
- **Frequency analysis**:



- **Solve partial differential equations**.

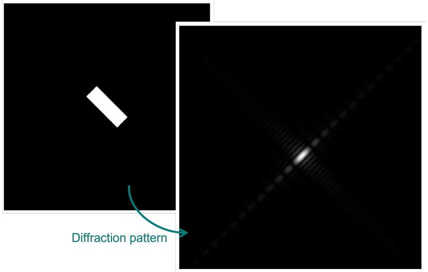
- **Time frequency analysis**:

$$\hat{f}(\omega, t) = \int_{\mathbb{R}} f(\tau) w(t-\tau) \mathrm{e}^{-\mathrm{i}\omega\tau} \mathrm{d}\tau$$



- **Optics** [1]:
Far-field diffraction = FT of aperture.

$$F(\boldsymbol{x}') = \frac{1}{i\lambda f} \mathscr{F}\{A(\lambda \boldsymbol{x})\}(\frac{\boldsymbol{x}'}{f})$$



- **Convolution theorem**: greatly reduces the computation time of the convolution operation. Applied in image processing and polynomial multiplication.
- **Solid state physics**: the relation between a lattice and its *reciprocal lattice*.
- **Hidden subgroup problem**: using the quantum Fourier transform to generate a uniformly distributed quantum state as input to the oracle, greatly reduces the query complexity of the algorithm.
- **Nonlinear Fourier transform** [2]: analogous to quantum signal processing, providing new insights and algorithm designs.
- **Spectral graph theory**: the eigenvectors to the Laplacian can be used to find clustering. Also useful in graph signal processing or even topological signal processing over simplicial complexes.
- and much much more...

So why is the Fourier transform of the following form?

$$\underbrace{f(x) = \sum_{k=-\infty}^{\infty} \hat{f}(k)\mathrm{e}^{\mathrm{i}2\pi kx}}_{\text{synthesis}} \xrightarrow{\mathscr{F}} \hat{f}(k) = \underbrace{\int_{-1/2}^{1/2} f(x)\mathrm{e}^{-\mathrm{i}2\pi kx}\mathrm{d}x}_{\text{analysis}}.$$

This is actually the same the coefficient for linear combination in a *vector space*! In essence, Fourier transform is just a change of basis.

$$\boldsymbol{v} = \sum_i c_i \boldsymbol{e}_i \to c_i = \frac{\boldsymbol{e}_i \cdot \boldsymbol{v}}{\boldsymbol{e}_i \cdot \boldsymbol{e}_i} = \frac{\langle \boldsymbol{e}_i, \boldsymbol{v} \rangle}{\langle \boldsymbol{e}_i, \boldsymbol{e}_i \rangle} \text{ with } \boldsymbol{e}_i \perp \boldsymbol{e}_j \text{ for } i \neq j.$$

The functions f and $\hat{f}$ simply represents the same information under different bases. The only difference is that we are working in an *infinite-dimensional vector space*.

What is a rigorous description of an infinite-dimensional vector space? It is a *Hilbert space*. To make sure we are on the same page, here is a definition of a vector space:

Def. (Vector Space)

A vector space V over the field $\mathbb{C}$ (can be other fields) is a set with vector-vector addition and scalar-vector multiplication satisfying: for all $\lambda, \mu \in \mathbb{C}$ and $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V$,

1. $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$.
2. $\boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w}) = (\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w}$.
3. $\exists \mathbf{0} \in V$ s.t. $\mathbf{0} + \boldsymbol{u} = \boldsymbol{u}$.
4. $\exists -\boldsymbol{v} \in V$ s.t. $\boldsymbol{v} + (-\boldsymbol{v}) = \mathbf{0}$.
5. $1 \cdot \boldsymbol{v} = \boldsymbol{v}$.
6. $\lambda \cdot (\boldsymbol{u} + \boldsymbol{v}) = \lambda \cdot \boldsymbol{u} + \lambda \cdot \boldsymbol{v}$.
7. $(\lambda + \mu) \cdot \boldsymbol{u} = \lambda \cdot \boldsymbol{u} + \mu \cdot \boldsymbol{u}$.
8. $(\lambda \cdot \mu) \cdot \boldsymbol{u} = \lambda \cdot (\mu \cdot \boldsymbol{u})$.

Def. (Hilbert Space)

A Hilbert space $\mathcal{H}$ over $\mathbb{C}$ is a *complete* inner product space.

Let us introduce the terms inner product and complete.

Def. (Inner Product Space)

An inner product space is just a vector space V over $\mathbb{C}$ with an *inner product* $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}$, satisfying: for all $\lambda \in \mathbb{C}$ and $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V$,

1. $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \langle \boldsymbol{v}, \boldsymbol{u} \rangle^*$, where * is the complex conjugation.
2. $\langle \boldsymbol{v}, \lambda \boldsymbol{u} + \boldsymbol{w} \rangle = \lambda \langle \boldsymbol{v}, \boldsymbol{u} \rangle + \langle \boldsymbol{v}, \boldsymbol{w} \rangle$.
3. $\langle \boldsymbol{v}, \boldsymbol{v} \rangle \geq 0$, where equality is met only if $\boldsymbol{v} = 0$.

Some examples of inner product:

- For $\boldsymbol{v} = (v_1, \ldots, v_n)$ and $\boldsymbol{u} = (u_1, \ldots, u_n) \in \mathbb{C}^n$, we have

$$\langle \boldsymbol{v}, \boldsymbol{u} \rangle = \sum_{i=1}^{n} v_i^* u_i. \tag{9}$$

  If such finite dimensional $\boldsymbol{v}$ satisfies $\langle \boldsymbol{v}, \boldsymbol{v} \rangle < \infty$, we say $\boldsymbol{v} \in \ell^2(\mathbb{Z}_n)$ is square summable.

- Consider an infinite sequence $\boldsymbol{v} = (\ldots, v_{-1}, v_0, v_1, \ldots)$, if

$$\|v\|^2 := \langle \boldsymbol{v}, \boldsymbol{v} \rangle = \sum_{i=-\infty}^{\infty} |v_i|^2 < \infty,$$

  we have $\boldsymbol{v} \in \ell^2(\mathbb{Z})$.

- For complex-valued functions f, g defined on the interval $\mathcal{I} \subseteq \mathbb{R}$, we can define their inner product to be

$$\langle f, g\rangle = \int_{\mathcal{I}} f^*(x)\, g(x)\, \mathrm{d}x. \tag{10}$$

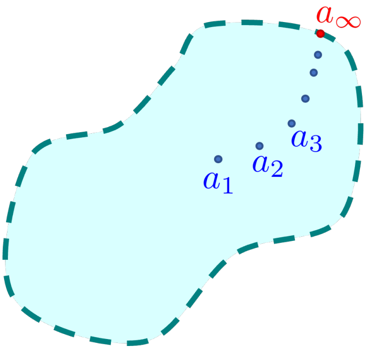
Again, if $\langle f, f\rangle < \infty$, we say $f \in L^2(\mathcal{I})$ is square integrable:

$$\|f\|^2 := \langle f, f\rangle = \int_{\mathcal{I}} |f(x)|^2\, \mathrm{d}x < \infty.$$

- Other possibilities exists, such as for $\mathcal{I} = [-1, 1]$,

$$\langle f, g\rangle = \int_{\mathcal{I}} f^*(x)\, g(x) \frac{\mathrm{d}x}{\sqrt{1 - x^2}}. \tag{11}$$

A rigorous definition of a *complete space* requires some knowledge in analysis, especially the meaning of the term *Cauchy sequence*. But it can be described easily by the illustration below:



For example:

$$\underbrace{\underbrace{3 + 0.1 + 0.04}_{a_3 \in \mathbb{Q}} + 0.001}_{a_4 \in \mathbb{Q}} + 0.0005 + \cdots = \pi \notin \mathbb{Q}.$$

Or consider all polynomials of real coefficient $\mathbb{R}[x]$:

$$\underbrace{1 + x + \frac{x^2}{2!} + \frac{x^3}{3!}}_{a_4 \in \mathbb{R}[x]} + \frac{x^4}{4!} + \cdots = \mathrm{e}^x \notin \mathbb{R}[x]$$

Finally, let us see how we can re-interpret Fourier transform as a linear combination of basis vectors in a Hilbert space. Let us take the CTFS as an example. Consider the basis $\{\mathrm{e}^{\mathrm{i}2\pi kx}\}_{k\in\mathbb{Z}} \subset L^2([-1/2, 1/2])$, we can see that they are orthonormal:

$$\langle \mathrm{e}^{\mathrm{i}2\pi k_1 x}, \mathrm{e}^{\mathrm{i}2\pi k_2 x}\rangle = \int_{-1/2}^{1/2} \mathrm{e}^{-\mathrm{i}2\pi k_1 x}\mathrm{e}^{\mathrm{i}2\pi k_2 x}\mathrm{d}x = \delta_{k_1 k_2}.$$

Hence we can decompose all functions[2] in $L^2([-1/2, 1/2])$ via the basis:

$$f(x) = \sum_{k=-\infty}^{\infty} \frac{\langle \mathrm{e}^{\mathrm{i}2\pi kx}, f(x)\rangle}{\langle \mathrm{e}^{\mathrm{i}2\pi kx}, \mathrm{e}^{\mathrm{i}2\pi kx}\rangle}\mathrm{e}^{\mathrm{i}2\pi kx} = \sum_{k=-\infty}^{\infty} \underbrace{\langle \mathrm{e}^{\mathrm{i}2\pi kx}, f(x)\rangle}_{\hat{f}(k)} \mathrm{e}^{\mathrm{i}2\pi kx}.$$

[2] The existence of an orthonormal basis that spans the whole Hilbert space one requires the Zorn's lemma, which we will not touch upon. For more details, see ref. [3].

Parseval's identity is trivial in this sense: since a vector does not change its length (norm) under a (unitary) coordinate transformation.

$$\langle f, f\rangle =: \|f\|^2 = \left\|\hat{f}\right\|^2 := \langle \hat{f}, \hat{f}\rangle. \tag{12}$$

We can also translate this into inner products via *polarization identity*:

Thm. (Plancherel Theorem)

For the Fourier transform pairs $f, \hat{f}$ and $g, \hat{g}$, we have

$$\langle f, g\rangle = \langle \hat{f}, \hat{g}\rangle. \tag{13}$$

→ *Proof.*

For example, in CTFS:

$$\begin{aligned}\langle f, g\rangle &= \int_{-1/2}^{1/2} f^*(x)g(x)\mathrm{d}x \\ &= \sum_{k=-\infty}^{\infty} \hat{f}^*(k)\hat{g}(k) = \langle \hat{f}, \hat{g}\rangle.\end{aligned}$$

The polarization identity induces an inner product from a norm that follows the parallelogram law.

Rmk. (Polarization Identity)

Given a norm $\|f\|$ that satisfies the *parallelogram law*:

$$\|f+g\|^2 + \|f-g\|^2 = 2\|f\|^2 + 2\|g\|^2, \tag{14}$$

it induces an inner product:

$$\langle f, g\rangle = \frac{1}{4}\left(\|f+g\|^2 - \|f-g\|^2\right). \tag{15}$$

As an example to an application of the Plancherel theorem, let us show a proof to the Poisson summation formula:

Thm. (Poisson Summation Formula)

Consider a function $f : \mathbb{R} \to \mathbb{C}$ that is *nice enough* (*Schwartz*), it satisfies

$$\sum_{n\in\mathbb{Z}} f(n) = \sum_{k\in\mathbb{Z}} \hat{f}(k). \tag{16}$$

Proof. Consider the impulse train:

$$\iota(x) = \sum_{n\in\mathbb{Z}} \delta(x-n) \xrightarrow{\mathscr{F}} \hat{\iota}(x) = \sum_{k\in\mathbb{Z}} \delta(x-k).$$

Then the proof follows directly by writing down $\langle \iota, f \rangle = \langle \hat{\iota}, \hat{f} \rangle$. ∎

When we learned Fourier transform of *functions*, some operations just seem ... illegal, weird, or unnatural, such as

- Why does $\int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}2\pi kx}\mathrm{d}k = \delta(x)$? Duality?
- What exactly is $\frac{\mathrm{d}^n}{\mathrm{d}x^n}\delta(x)$? Unit doublet?
- Is there a meaning to $u(x) \cdot \delta(x)$?
- Is there a meaning to $\delta(x) \cdot \delta(x)$?
- Why does $\mathscr{F}\{u(x)\}(k) = \frac{1}{\mathrm{i}2\pi k} + \frac{1}{2}\delta(k)$?

To answer these, we have to dive into the basics of *distribution theory*, explaining the meaning of sentences such as:

"The Dirac delta is not a function, it is a distribution."

Beside Plancherel's, many of the properties are parallel across the four transforms:

	Time / Freq. Shift	Differentiation	Convolution
CTFT	$f(x-x_0) \xrightarrow{\mathscr{F}} \mathrm{e}^{-\mathrm{i}2\pi k x_0}\hat{f}(k)$ $\mathrm{e}^{\mathrm{i}2\pi k_0 x} f(x) \xrightarrow{\mathscr{F}} \hat{f}(k+k_0)$	$f'(x) \xrightarrow{\mathscr{F}} \mathrm{i}2\pi k \hat{f}(k)$ $-\mathrm{i}2\pi x f(x) \xrightarrow{\mathscr{F}} \hat{f}'(k)$	$(f*g)(x) \xrightarrow{\mathscr{F}} \hat{f}(k)\hat{g}(k)$ $f(x)g(x) \xrightarrow{\mathscr{F}} (\hat{f}*\hat{g})(k)$
CTFS	$f(x-x_0) \xrightarrow{\mathscr{F}} \mathrm{e}^{-\mathrm{i}2\pi k x_0}\hat{f}(k)$ $\mathrm{e}^{\mathrm{i}2\pi k_0 x} f(x) \xrightarrow{\mathscr{F}} \hat{f}(k+k_0)$	$f'(x) \xrightarrow{\mathscr{F}} \mathrm{i}2\pi k \hat{f}(k)$	$(f*g)(x) \xrightarrow{\mathscr{F}} \hat{f}(k)\hat{g}(k)$ $f(x)g(x) \xrightarrow{\mathscr{F}} (\hat{f}*\hat{g})(k)$
DTFT	$f(n-n_0) \xrightarrow{\mathscr{F}} \mathrm{e}^{-\mathrm{i}2\pi k n_0}\hat{f}(k)$ $\mathrm{e}^{\mathrm{i}2\pi k_0 x} f(x) \xrightarrow{\mathscr{F}} \hat{f}(k+k_0)$	$-\mathrm{i}2\pi n f(n) \xrightarrow{\mathscr{F}} \hat{f}'(k)$	$(f*g)(x) \xrightarrow{\mathscr{F}} \hat{f}(k)\hat{g}(k)$ $f(x)g(x) \xrightarrow{\mathscr{F}} (\hat{f}*\hat{g})(k)$
DTFS	$f(n-n_0) \xrightarrow{\mathscr{F}} \mathrm{e}^{-\mathrm{i}\frac{2\pi}{N} k n_0}\hat{f}(k)$ $\mathrm{e}^{\mathrm{i}\frac{2\pi}{N} k_0 n} f(n) \xrightarrow{\mathscr{F}} \hat{f}(k+k_0)$	$\times$	$(f*g)(x) \xrightarrow{\mathscr{F}} \hat{f}(k)\hat{g}(k)$ $f(x)g(x) \xrightarrow{\mathscr{F}} (\hat{f}*\hat{g})(k)$

Note that the convolutions over different domains are different:

- Over $\mathbb{R}$:

$$(f * g)(x) = \int_{-\infty}^{\infty} f(\xi) g(x - \xi) \mathrm{d}\xi. \tag{17}$$

- Over $[-1/2, 1/2]$:

$$(f * g)(x) = \int_{-1/2}^{1/2} f(\xi) g(x - \xi) \mathrm{d}\xi. \tag{18}$$

- Over $\mathbb{Z}$:

$$(f * g)(n) = \sum_{m=-\infty}^{\infty} f(m) g(n - m). \tag{19}$$

- Over $\{0, 1, \ldots, N - 1\}$:

$$(f * g)(n) = \sum_{m=0}^{N-1} f(m) g(n - m). \tag{20}$$

Furthermore, if we repetitively apply Fourier transform on a function: (only works for CTFT and DTFS)

$$
\begin{aligned}
\mathscr{F}\{f\}(x) &= \int_{-\infty}^{\infty} f(\xi)\mathrm{e}^{-\mathrm{i}2\pi\xi x}\mathrm{d}\xi = \hat{f}(x), \\
\mathscr{F}^2\{f\}(x) &= \int_{-\infty}^{\infty} \hat{f}(\xi)\mathrm{e}^{-\mathrm{i}2\pi\xi x}\mathrm{d}\xi = f(-x), \\
\mathscr{F}^3\{f\}(x) &= \mathscr{F}^2\{\hat{f}\}(x) = \hat{f}(-x), \\
\mathscr{F}^4\{f\}(x) &= f(x).
\end{aligned}
$$

We can see that the *Fourier transform operator* satisfies $\mathscr{F}^4 = \mathbb{1}$, the identity operator; while $\mathscr{F}^2 = \mathrm{refl}$, the reflection operator about the y-axis.

This can be seen even more clearly through DTFS, often also called the *discrete Fourier transform* (*DFT*):

$$\hat{f}(k) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \mathrm{e}^{-\mathrm{i}2\pi kn} f(n)$$

$$\begin{bmatrix} \hat{f}(0) \\ \hat{f}(1) \\ \vdots \\ \hat{f}(N-1) \end{bmatrix} = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & \mathrm{e}^{-\mathrm{i}\frac{2\pi}{N}} & \cdots & \mathrm{e}^{-\mathrm{i}\frac{2\pi}{N}(N-1)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \mathrm{e}^{-\mathrm{i}\frac{2\pi}{N}(N-1)} & \cdots & \mathrm{e}^{-\mathrm{i}\frac{2\pi}{N}(N-1)^2} \end{bmatrix} \begin{bmatrix} f(0) \\ f(1) \\ \vdots \\ f(N-1) \end{bmatrix} \tag{21}$$

$$\hat{\boldsymbol{f}} = F_N \boldsymbol{f},$$

where $[F_N]_{kn} = \frac{1}{\sqrt{N}} \mathrm{e}^{-\mathrm{i}\frac{2\pi}{N}kn}$ $(k, n = 0 \sim N-1)$. The DFT matrix is $\sqrt{N} F_N$.

The matrix F_N satisfies:

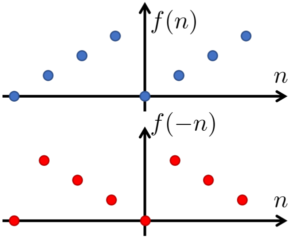
- $F_N^{-1} = F_N^{\dagger} \equiv F_N^{\mathsf{H}}$, hence F_N is a unitary matrix!
- The eigenvalue of F_N is in $\{1, -1, \mathrm{i}, -\mathrm{i}\}$.
- $F_N^4 = \mathbb{1}$, the identity matrix.
- The reflection operator is of the form:

$$F_N^2 = \left[\begin{array}{c:cccc} 1 & & & & \\ \hdashline & & & & 1 \\ & & & & \\ & & & \iddots & \\ & 1 & & & \end{array}\right]. \tag{22}$$

- Taking the limit of $N \to \infty$, we can obtain the CTFT.

$\rightarrow$ Unitary: $UU^{\mathsf{H}} = U^{\mathsf{H}}U = \mathbb{1}$, where ${}^{\mathsf{H}}$ is the *Hermitian conjugation*, or the complex conjugate transpose.

$\rightarrow$ Eigenvalue λ: $Av = \lambda v$.



Fourier transform can be done in higher dimensions, many interesting properties arises in higher dimensions.

Furthermore, we have seen the Fourier transforms between

$$\begin{aligned} \mathbb{R} &\longleftrightarrow \mathbb{R} \\ \mathbb{Z} &\longleftrightarrow \mathbb{T} \ (= \text{unit circle} \cong [-1/2, 1/2]\ ) \\ \mathbb{Z}_N &\longleftrightarrow \mathbb{Z}_N \ (= \{0, 1, \ldots, N-1\} = \text{mod } N \text{ integers}) \end{aligned}$$

with functions of complex values. Why do we have such dualities? Can we have "Fourier transforms" on other wierd mathematical objects?

We will introduce the Fourier transform on prime fields (體), abelian groups (群), and graphs (圖)!

The materials starting from here will become only more difficult. So let us start things easy.

Given a single-variable function, we can apply Fourier transform on its single variable. For a multivariate function, we can apply Fourier transform on each of the coordinates.

Def. (Multi-Dimensional Fourier Transform)

Given a function $f(\boldsymbol{x}) = f(x_1, x_2, \ldots, x_n)$, we can define its Fourier transform into the coordinates $\boldsymbol{k} = (k_1, k_2, \ldots, k_n)$ by

$$\hat{f}(\boldsymbol{k}) = \int \cdots \int_{-\infty}^{\infty} f(\boldsymbol{x}) \mathrm{e}^{-\mathrm{i}2\pi(k_1 x_1 + \cdots k_n x_n)} \mathrm{d}x_1 \cdots \mathrm{d}x_n = \int_{\mathbb{R}^n} f(\boldsymbol{x}) \mathrm{e}^{-\mathrm{i}2\pi \boldsymbol{k}^\mathsf{T} \boldsymbol{x}} \mathrm{d}^n \boldsymbol{x}. \quad (23)$$

All the properties have counterparts in the multi-dimensional case:

- Translation: $f(\boldsymbol{x}-\boldsymbol{x}_0) \xrightarrow{\mathscr{F}} \mathrm{e}^{-\mathrm{i}2\pi \boldsymbol{k}^\mathsf{T}\boldsymbol{x}} \hat{f}(\boldsymbol{k})$.
- Phase modulation: $\mathrm{e}^{\mathrm{i}2\pi \boldsymbol{k}_0^\mathsf{T}\boldsymbol{x}} f(\boldsymbol{x}) \xrightarrow{\mathscr{F}} \hat{f}(\boldsymbol{k}-\boldsymbol{k}_0)$.
- Differentiation:
$$\frac{\partial}{\partial \boldsymbol{x}^\mathsf{T}} f(\boldsymbol{x}) = \nabla_{\boldsymbol{x}} f(\boldsymbol{x}) \xrightarrow{\mathscr{F}} \mathrm{i}2\pi \boldsymbol{k} \hat{f}(\boldsymbol{k}).$$
- Scaling: $f(A\boldsymbol{x}) \xrightarrow{\mathscr{F}} |\det(A)|^{-1}\, \hat{f}(A^{-1}\boldsymbol{k})$.
- Convolution:
$$f(\boldsymbol{x}) * \boldsymbol{g}(\boldsymbol{x}) \xrightarrow{\mathscr{F}} \hat{f}(\boldsymbol{k})\hat{f}(\boldsymbol{k}).$$
- Plancherel:
$$\langle f(\boldsymbol{x}), g(\boldsymbol{x})\rangle = \langle \hat{f}(\boldsymbol{k}), \hat{g}(\boldsymbol{k})\rangle.$$

For a two-dimensional signal, we have

$$\hat{f}(k_x, k_y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\mathrm{e}^{-\mathrm{i}2\pi(k_x x + k_y y)}\,\mathrm{d}x\,\mathrm{d}y.$$

We can *discretize* this equation to obtain the two-dimensional DFT:

$$\hat{f}(k_x, k_y) = \frac{1}{N^2}\sum_{n_x=0}^{N-1}\sum_{n_y=0}^{N-1} f(n_x, n_y)\mathrm{e}^{-\mathrm{i}2\pi\left(\frac{k_x}{N}n_x + \frac{k_y}{N}n_y\right)}$$

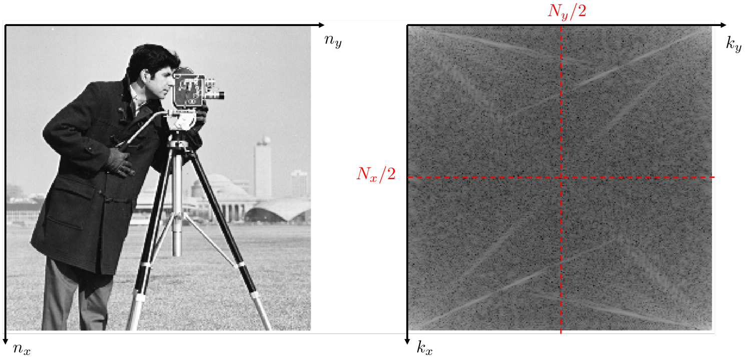
$$\propto \frac{1}{\sqrt{N}}\sum_{n_x=0}^{N-1}\frac{1}{\sqrt{N}}\sum_{n_y=0}^{N-1}\mathrm{e}^{-\mathrm{i}\frac{2\pi}{N}k_x n_x} f(n_x, n_y)\mathrm{e}^{-\mathrm{i}\frac{2\pi}{N}k_y n_y} = [F_N \underbrace{X}_{N\times N} F_N^{\mathsf{T}}]_{k_x,k_y},$$

where $[X]_{n_x,n_y} = f(n_x, n_y)$, $[\widehat{X}]_{k_x,k_y} = \hat{f}(k_x, k_y)$. Remember, $[F_N]_{kn} = \mathrm{e}^{-\mathrm{i}\frac{2\pi}{N}kn}$.

When given an image $X \in \mathbb{R}^{N_x \times N_y}$, we have its Fourier transform be of the form

$$\widehat{X} = F_{N_x} X F_{N_y}^{\mathsf{T}}. \tag{24}$$

In MATLAB, we have the functions: $\widehat{X} = \text{fft2}(X)$ and $\text{fftshift}(\cdot)$.



Def. (Vectorization)

The vectorization of a matrix is created by stacking its columns together. For example,

$$\operatorname{vec}\left(\begin{bmatrix} | & & | \\ \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_n \\ | & & | \end{bmatrix}\right) = \begin{bmatrix} \boldsymbol{v}_1 \\ \vdots \\ \boldsymbol{v}_n \end{bmatrix}. \qquad (25)$$

Rmk. (Vectorization)

The following relation holds:

$$\operatorname{vec}(AXB) = (B^{\mathsf{T}} \otimes A)\operatorname{vec}(X), \qquad (26)$$

where $\otimes$ is the Kronecker product.

Kronecker product:

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots \\ a_{21}B & a_{22}B & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}. \qquad (27)$$

Hence, the Fourier transform of an image becomes

$$\begin{aligned} \operatorname{vec}(\widehat{X}) &= \operatorname{vec}(F_{N_x} X F_{N_y}^{\mathsf{T}}) \\ &= (F_{N_y} \otimes F_{N_x})\operatorname{vec}(X). \end{aligned} \qquad (28)$$

A nice result of Fourier transform is

$$\underbrace{\hat{f}(0)}_{\text{slice}} = \underbrace{\int_{-\infty}^{\infty} f(x) \mathrm{e}^{-\mathrm{i}2\pi 0} \mathrm{d}x}_{\text{projection}} .$$

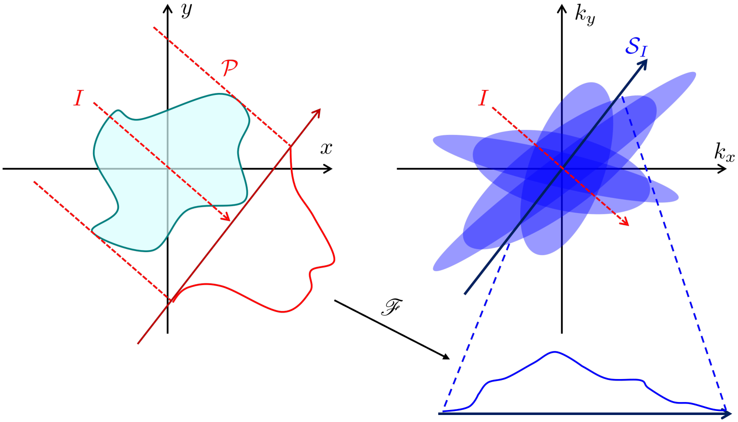
We can see that a slice in Fourier space corresponds to a projection in real space. This can be generalized to:

Thm. (Fourier Slice Theorem)

Consider the function $f(\boldsymbol{x})$ with $\boldsymbol{x} \in \mathbb{R}^n$. Let $\mathcal{F}_i$ denote the Fourier transform over the ith coordinate, let $\mathcal{P}_i$ denote the projection over the ith coordinate, let $\mathcal{S}_i$ denote the slice of the ith coordinate at the origin. Then for coordinates $I \subseteq [n] = \{1, 2, \ldots, n\}$,

$$\mathcal{S}_I \circ \mathcal{F}_{[n]}(f) = \mathcal{F}_{[n]\setminus I} \circ \mathcal{P}_I(f). \tag{29}$$

This is utilized in CT scans: We can reconstruct the Fourier spectrum from the Fourier transform of projections (constituting the slices).



The main goal of this section is to answer the following question:

What exactly is $\delta(x)$?

And also some other miscellaneous questions such as:

- What are some valid operations on $\delta(x)$ and other distributions?
- How about their Fourier transform?
- What do we mean by the formula below?

$$f(x) = \int_{-\infty}^{\infty} \underbrace{\left(\int_{-\infty}^{\infty} f(\xi)\mathrm{e}^{-\mathrm{i}2\pi\xi k}\mathrm{d}\xi\right)}_{\hat{f}(k)} \mathrm{e}^{\mathrm{i}2\pi kx}\mathrm{d}k = \int_{-\infty}^{\infty} f(\xi) \underbrace{\left(\int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}2\pi k(x-\xi)}\mathrm{d}k\right)}_{???} \mathrm{d}\xi$$

Heavy influence of the material from 泛函分析與逼近理論 by prof. Blu [4].

In course, our first introduction to the Dirac delta is the weird function

$$\delta(x) = \begin{cases} 0, & x \neq 0; \\ \infty, & x = 0. \end{cases}$$

This definition has no meaning at all!

Instead, we were offered with the operational definition of: the Dirac delta is a *function* (?) that satisfies

$$\int_{-\infty}^{\infty} \delta(x) f(x) \, \mathrm{d}x := f(0) \tag{30}$$

for all suitable function $f(x)$.

The Dirac delta only makes sense under an integral with suitable functions.

Def. (Test Function)

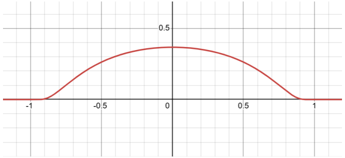
A test function is a C^∞ (infinitely differentiable) function with *bounded support*. The space of all such test functions is denoted by $\mathscr{D}$.

Def. (Distribution)

A distribution $\psi \in \mathscr{D}'$ is a linear operator acting on $\mathscr{D}$ characterized by

- the scalar product $\langle \psi, \varphi \rangle$ for all $\varphi \in \mathscr{D}$,
- its order, which ensures the continuity of the scalar product.

$$\varphi(x) = \begin{cases} \exp\left(-\frac{1}{1-x^2}\right), & |x| \leq 1; \\ 0, & |x| > 1. \end{cases}$$



$$\langle \psi, \varphi \rangle \overset{\text{symbolic}}{=} \int_{\mathbb{R}} \psi(x)\varphi(x)\,\mathrm{d}x \tag{31}$$

For distributions, instead of defining them via their values on each x, we *characterize* them by their effect on test functions! They are not functions, but can be seen as *linear operators* on functions.

The Dirac delta is fully characterized by

$$\langle \delta, \varphi \rangle := \varphi(0) \tag{32}$$

for all test functions $\varphi \in \mathscr{D}$. Similarly, we have the operational definition of the *unit doublet* as

$$\langle u_1, \varphi \rangle := \frac{\mathrm{d}\varphi(0)}{\mathrm{d}x}. \tag{33}$$

Note that since distributions are in fact linear operators, some operations are not valid. Luckily, we can define the derivatives of distributions!

Def. (Derivatives of Distributions)

For a distribution ψ, its derivative is the distribution satisfying

$$\langle \psi', \varphi \rangle = -\langle \psi, \varphi' \rangle \tag{34}$$

for all test functions φ.

The minus sign is from

$$\int_{\mathbb{R}} \psi' \varphi \mathrm{d}x = \cancelto{0}{[\psi\varphi]_{-\infty}^{\infty}} - \int_{\mathbb{R}} \psi \varphi' \mathrm{d}x. \qquad \blacksquare$$

For example, we have $u_1 = -\delta'$ and $\delta = u'$.

It is normally not possible to multiply distributions together, but we can multiply a distribution with an infinitely differentiable (C^∞) function f:

$$\langle f\psi, \varphi\rangle = \langle \psi, f\varphi\rangle. \tag{35}$$

A counterexample will be the unit-step function (Heaviside function): $u(x)$. We have $u(x)^2 = u(x)^3$, yet

$$\frac{\mathrm{d}}{\mathrm{d}x}u(x)^2 = 2u(x)\delta(x) \neq 3u(x)^2\delta(x) = \frac{\mathrm{d}}{\mathrm{d}x}u(x)^3.$$

Thm. (Limits in the Sense of Distributions)

If for all $\varphi \in \mathscr{D}$, we have

$$\lim_{n\to\infty} \langle \psi_n, \varphi \rangle = \langle \psi, \varphi \rangle, \tag{36}$$

then we say $\psi_n \xrightarrow{n\to\infty} \psi$ in the sense of distribution.

For example: $\mathrm{e}^{\mathrm{i}nx} \xrightarrow{n\to\infty} 0.$

Proof.

$$\left| \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}nx} \varphi(x) \mathrm{d}x \right| = \left| \cancelto{0}{\left[\frac{\mathrm{e}^{\mathrm{i}nx}}{\mathrm{i}n} \varphi(x) \right]_{-\infty}^{\infty}} - \frac{1}{\mathrm{i}n} \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}nx} \varphi'(x) \mathrm{d}x \right| \le \frac{1}{n} \left| \int_{\mathbb{R}} \varphi'(x) \mathrm{d}x \right| = \frac{c}{n} \to 0. \qquad \blacksquare$$

Rmk. (Regularizing Sequence)

For $\psi \in L^1(\mathbb{R})$, define the sequence $\psi_n(x) = n \cdot \psi(nx)$ and $\int_{\mathbb{R}} \psi(x) = 1$. Then $\psi_n(x) \to \delta(x)$ is a regularizing sequence.

Proof.

$$\begin{aligned}\langle \psi_n, \varphi \rangle &= \langle n\psi(nx), \varphi(x) \rangle = \langle \psi(x), \varphi(x/n) \rangle \\ &= \int_{\mathbb{R}} \psi(x)\varphi(x/n)\mathrm{d}x \xrightarrow[\text{DCT}]{n\to\infty} \int_{\mathbb{R}} \psi(x)\varphi(0)\mathrm{d}x = \psi(0) = \langle \delta, \varphi \rangle. \qquad \blacksquare\end{aligned}$$

DCT is the *dominated convergence theorem*, which we will not touch on. An example of the above remark is

$$\delta(x) = \lim_{n\to\infty} \frac{n}{\sqrt{2\pi}} \mathrm{e}^{-\frac{n^2x^2}{2}} = \lim_{\varepsilon\to 0^+} \frac{1}{\sqrt{2\pi\varepsilon^2}} \mathrm{e}^{-\frac{x^2}{2\varepsilon^2}}. \tag{37}$$

We would like to define the Fourier transform of distributions via the Plancherel property:

$$\langle \hat{\psi}, \varphi \rangle = \langle \psi, \left(\breve{\varphi^*}\right)^* \rangle,$$

where $\breve{}$ is the inverse Fourier transform. Note that

$$\left(\breve{\varphi^*}\right)^* = \left[\int_{\mathbb{R}} \varphi^*(k)\mathrm{e}^{\mathrm{i}2\pi kx}\mathrm{d}k\right]^* = \int_{\mathbb{R}} \varphi(k)\mathrm{e}^{-\mathrm{i}2\pi kx}\mathrm{d}k = \hat{\varphi}(x).$$

However, since φ is a test function, it has bounded support, then $\hat{\varphi}$ has *unbounded* support. $\hat{\varphi}$ is no longer a test function!

We need a less restrictive replacement to $\mathscr{D}$.

The test functions seem to be a bit restrictive. We have some alternatives:

Rmk. (Schwartz Space)

The Schwartz space $\mathscr{S}$ contains all functions with all derivatives that are decreasing faster than polynomials: for all m, n,

$$\lim_{x\to\infty} |x|^n \left| f^{(m)}(x) \right| = 0. \tag{38}$$

Obviously, we have $\mathscr{S} \supset \mathscr{D}$, and $\varphi \in \mathscr{S} \Leftrightarrow \hat{\varphi} \in \mathscr{S}$. The linear operators mapping to scalars on the Schwartz space form distributions:

Rmk. (Tempered Distribution)

The tempered distributions $\mathscr{S}' \subset \mathscr{D}'$ contains all distributions ψ such that $\langle \psi, \varphi \rangle$ is well-defined for all $\varphi \in \mathscr{S}$.

Def. (Fourier Transform pf Distributions)

For a distribution $\psi \in \mathscr{D}$, its Fourier transform is defined by the operation

$$\langle \hat{\psi}, \varphi \rangle = \langle \psi, \hat{\varphi} \rangle \tag{39}$$

for all $\varphi \in \mathscr{S}$ if it is well-defined.

For example, consider the distribution $1 \in \mathscr{D}$, its Fourier transform is characterized by

$$\langle 1, \hat{\varphi} \rangle = \int_{\mathbb{R}} \hat{\varphi}(x) \mathrm{d}x = \varphi(0) = \langle \hat{1}, \varphi \rangle.$$

Hence, we have $\hat{1} = \delta(k) \overset{\text{symbolic}}{=} \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}2\pi kx} \mathrm{d}x$.

Since we have

$$\mathscr{F}\{\delta(x)\} = 1$$
$$\mathscr{F}\{u'(x)\} = \mathrm{i}2\pi k \cdot \mathscr{F}\{u(x)\},$$

it is natural to think that $\mathscr{F}\{u(x)\} = \frac{1}{\mathrm{i}2\pi k}$. However, $\mathrm{i}2\pi k \cdot \mathscr{F}\{u(x)\}$ effectively removes all DC component to $\mathscr{F}\{u(x)\}$.

The DC value to $\mathscr{F}\{u(x)\}$ is $\frac{1}{2}$, hence we add it back:

$$\mathscr{F}\{u(x)\} = \frac{1}{\mathrm{i}2\pi k} + \frac{1}{2}\delta(k). \tag{40}$$

For example:

$$\sum_{n=0}^{\infty} \mathrm{e}^{-\mathrm{i}2\pi nk} = \frac{1}{1 - \mathrm{e}^{-\mathrm{i}2\pi k}} + \frac{1}{2}\sum_{n=-\infty}^{\infty} \delta(k-n). \tag{41}$$

In this subsection, we will consider the Fourier transform over integers (prime fields).

This is a useful thing to have since if we want to consider the multiplication of polynomials with integer coefficients:

$$(x^2+3x+1)(3x^2+x+2) = \begin{array}{rrrrrr} & 3 & 9 & 3 & & \\ & & 1 & 3 & 1 & \\ +) & & & 2 & 6 & 2 \\ \hline & 3 & 10 & 8 & 7 & 2 \end{array} = 3x^4+10x^3+8x^2+7x+2.$$

This is equal to the convolution of the two sequences! We can use the convolution theorem to calculate the convolution via multiplication in the Fourier domain.

However, for *numerical accuracy*, one should opt for working only in the integers. Is Fourier transform still possible then?

If we let $\omega = \mathrm{e}^{-\mathrm{i}\frac{2\pi}{N}}$, then the DFT matrix can be rewritten as

$$F = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega & \omega^2 & \cdots & \omega^{N-1} \\ 1 & \omega^2 & \omega^4 & \cdots & \omega^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{N-1} & \omega^{2(N-1)} & \cdots & \omega^{(N-1)^2} \end{bmatrix} \qquad (\omega^N = 1). \tag{42}$$

Key idea: find an integer $\omega \in \mathbb{Z}$ such that $\omega^N \overset{!}{=} 1$, if that's even possible.

Here we start with some fundamentals of number theory to aid us in our discussion.

Def. (Prime Field)

Given a prime number p, we can define the prime field $\mathbb{Z}_p$ as the set

$$\mathbb{Z}_p = \{0, 1, 2, \dots, p-1\}$$

with modulo p addition, subtraction, multiplication, and division.

For example, in $\mathbb{Z}_7$, we have

- $1 + 5 = 6$, $6 + 6 = 12 = 5$.
- $1 - 5 = -4 = 3$.
- $4 \times 3 = 12 = 5$.
- $5 \div 3 = ???$

Please allow my abuse of notation:
in $\mathbb{Z}_p$, I simply write $a = b$ as equivalent to

$$a \equiv b \mod p.$$

How can we have division, this is a non-trivial task.

Thm. (Bézout's Lemma)

For integer a and b, there exists integers x and y such that

$$ax + by = \gcd(a, b). \tag{43}$$

Proof. By the well-ordering principle, What we used in our proof above is called the *Euclidean algorithm* (輾轉相除法). It can be used to find inverses on prime fields.

Def. (Inverse on Prime Field)

For all non-zero $a \in \mathbb{Z}_p$ we have the inverse of a defined as the unique $a^{-1} \in \mathbb{Z}_p$ satisfying

$$a^{-1}a = 1 \mod p. \tag{44}$$

For example: Find the inverse of $a = 12$ in $\mathbb{Z}_p$ where $p = 31$.

This is why we only work on *prime* fields.

Now that we have $\mathbb{Z}_p$ at our disposal, it is high time to address the big problem: does there exist an element in $\omega \in \mathbb{Z}_p$ such that

$$\omega^N = 1?$$

The following theorem gives us a hint of its existence:

Rmk. (Fermat's Little Theorem)

For all $a \in \mathbb{Z}_p$ that is non-zero, we have

$$a^{p-1} = 1. \tag{45}$$

An immediate corollary is that a sufficient condition is $N \mid p-1$. It is also necessary.

Def. (Primitive Root)

For the prime field $\mathbb{Z}_p$, there exists $\varphi(p-1)$ primitive elements a that satisfies

$$\mathbb{Z}_p = \left\{0, a^0, a^1, a^2, \dots, a^{p-2}\right\}, \tag{46}$$

i.e., the powers of a primitive root (and 0) generate the whole $\mathbb{Z}_p$. $\varphi(n)$ is the Euler's totient function, counting the numbers from 1 to n that are coprime to n.

For example, consider $\mathbb{Z}_{11}$, we have the powers of 2 as:

$$1, 2, 4, 8, 5, 10, 9, 7, 3, 6.$$

Thus, 2 is a primitive roots. However, the powers of 3 are:

$$1, 3, 9, 5, 4.$$

Def. (Number Theoretic Transform)

Given a prime field $\mathbb{Z}_p$ and $N \mid p-1$, we define ω to be a *primitive* N-root (the minimum integer $k > 0$ such that $\omega^k = 1$ is $k = N$). The NTT of a vector $\boldsymbol{f}$ and the inverse transform are defined as

$$\hat{\boldsymbol{f}} = \mathrm{NTT}(\boldsymbol{f}) = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & \omega & \cdots & \omega^{N-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{N-1} & \cdots & \omega^{(N-1)^2} \end{bmatrix} \boldsymbol{f}, \tag{47}$$

$$\boldsymbol{f} = \mathrm{NTT}^{-1}(\hat{\boldsymbol{f}}) = N^{-1} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & \omega^{-1} & \cdots & \omega^{-(N-1)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{-(N-1)} & \cdots & \omega^{-(N-1)^2} \end{bmatrix} \hat{\boldsymbol{f}}. \tag{48}$$

Back to the task of polynomial multiplication: the two degree $N-1$ polynomials $f(x)$ and $g(x)$ can be determined uniquely by N *coefficients*. Similarly, we can uniquely define them by their values on N different points, say

$$f(1),\ f(\omega),\ f(\omega^2),\ \ldots,\ f(\omega^{N-1}) \text{ and}$$
$$g(1),\ g(\omega),\ g(\omega^2),\ \ldots,\ g(\omega^{N-1}).$$

It is easy to multiply the two polynomials by knowing their values on these N points, we simply have

$$f(1)g(1),\ f(\omega)g(\omega),\ f(\omega^2)g(\omega^2),\ \ldots,\ f(\omega^{N-1})g(\omega^{N-1}).$$

Notice that the polynomial uniquely determined by these N numbers is of degree $N-1$. The polynomial obtained is equivalent to setting $x^N = 1$.

Thm. (Polynomial Multiplication)

Given degree $N-1$ polynomials $f(x)$ and $g(x)$ with $\mathbb{Z}_p$-coefficients, we can represent their coefficients via vectors $\boldsymbol{f}$ and $\boldsymbol{g}$. Then the polynomial $h(x)$ obtained by

$$\boldsymbol{h} = \mathrm{NTT}^{-1}\left(\hat{\boldsymbol{f}} \circ \hat{\boldsymbol{g}}\right) \tag{49}$$

is the polynomial $f(x) \cdot g(x)$ with $x^N = 1$ identified. The $\circ$ represents vector element-wise multiplication.

Some technicalities: The set (ring) of all polynomials with $\mathbb{Z}_p$-coefficient is $\mathbb{Z}_p[x]$. In NTT, we are working with the set of all such polynomials with $x^N = 1$ identified, i.e., the set (quotient ring)

$$\mathbb{Z}_p[x]/(x^N - 1). \tag{50}$$

For example, consider in $\mathbb{Z}_{17}$, we have $f(x) = x^2 + 3x + 1$ and $g(x) = 3x^2 + x + 2$.
We have $f(x) \cdot g(x) = 3x^4 + 10x^3 + 8x^2 + 7x + 2 \overset{x^4=1}{=} 10x^3 + 8x^2 + 7x + 5$.

Let us choose $N = 4 \mid 17 - 1$. Then we have a primitive root $\omega = 4$, $\omega^{-1} = 13$.

$$\hat{\boldsymbol{f}} = \mathrm{NTT}\left(\begin{bmatrix}1\\3\\1\\0\end{bmatrix}\right) = \begin{bmatrix}5\\12\\16\\5\end{bmatrix}, \hat{\boldsymbol{g}} = \begin{bmatrix}1 & 1 & 1 & 1\\1 & 4 & 16 & 13\\1 & 16 & 1 & 16\\1 & 13 & 16 & 4\end{bmatrix}\begin{bmatrix}2\\1\\3\\0\end{bmatrix} = \begin{bmatrix}6\\3\\4\\12\end{bmatrix}$$

Hence, we have $\hat{\boldsymbol{h}} = \hat{\boldsymbol{f}} \circ \hat{\boldsymbol{g}} = [13, 2, 13, 9]^{\mathsf{T}}$, and $\boldsymbol{h} = [5, 7, 8, 10]^{\mathsf{T}}$. Finally, we obtained $h(x) = 5 + 7x + 8x^2 + 10x^3$. The two results coincide.

Both NTT and DFT has the structure of

$$F = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega & \omega^2 & \cdots & \omega^{N-1} \\ 1 & \omega^2 & \omega^4 & \cdots & \omega^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{N-1} & \omega^{2(N-1)} & \cdots & \omega^{(N-1)^2} \end{bmatrix} \qquad (\omega^N = 1),$$

The same analyses above also works for FFT. Moreover, the two methods have the same realization complexity.

The fastest method is the by the fast Fourier transform (FFT) algorithm. Using the *butterfly network* architecture with $N = 2^n$, the complexity of such computation is $O(N \ln N)$.

What's the difference between using FFT and NTT?

- FFT works in $\mathbb{C}[x]/(x^N - 1)$. NTT works in $\mathbb{Z}_p[x]/(x^N - 1)$.
- If the polynomials have "positive integer coefficients" with values less than p, then the results from the two methods should coincide!

Why? Because no one can discriminate what ω actually is, all we know is that $\omega^N = 1$! It can be $\mathrm{e}^{\mathrm{i}\frac{2\pi}{N}}$, $\mathrm{e}^{-\mathrm{i}\frac{2\pi}{N}}$, or even integers.

For example, if someone changes all i in the universe to $-\mathrm{i}$, no one can notice a difference.

In this subsection, we will discuss the Fourier transform over abelian groups.

The goal of this section is to give you an idea of why the *duality* between $\mathbb{T}$ and $\mathbb{Z}$ holds. This also extends to give a general description of duality between the *spatial* and *frequency domains*.

$$f(x) = \sum_{k=-\infty}^{\infty} \hat{f}(k)\mathrm{e}^{i2\pi kx} \xrightarrow{\mathscr{F}} \hat{f}(k) = \int_{-1/2}^{1/2} f(x)\mathrm{e}^{-\mathrm{i}2\pi kx}\mathrm{d}x \tag{51}$$

$$f(n) = \int_{-1/2}^{1/2} \hat{f}(k)\mathrm{e}^{\mathrm{i}2\pi kn}\mathrm{d}k \xrightarrow{\mathscr{F}} \hat{f}(k) = \sum_{n=-\infty}^{\infty} f(n)\mathrm{e}^{-i2\pi kn} \tag{52}$$

Numbers measure quantities, groups measure symmetries.

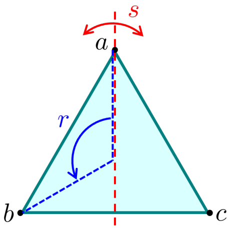
Def. (Group)

A group is a set G with operation $\odot : G \times G \to G$ satisfying: for all $a, b, c \in G$,

1. $a \odot (b \odot c) = (a \odot b) \odot c$.
2. There exists an identity e such that $e \odot a = a$.
3. For all $a \in G$, there exists an inverse a^{-1} such that $a^{-1} \odot a = e$.

Take the symmetries of a triangle, $(D_6, \cdot)$, for example. We ask the question, under what transformation does the triangle remain unchanged (symmetry)? Let s: reflection and r: rotation. We have

$$D_6 = \{e, s, r, r^2, rs, r^2 s\}. \tag{53}$$



Note that $rs = sr^2$, it is not *commutative*.

The discussion of non-commutative groups is very interesting but too difficult for today, so let us shift our attention to the special cases of commutative groups.

Def. (Abelian Group)

Abelian groups are commutative groups, i.e., for all g, h in an abelian group $(G, \cdot)$ (resp. $(G, +)$), we have $g \cdot h = h \cdot g$ (resp. $g + h = h + g$).

Some common examples:

1. Additive groups: $(\mathbb{R}, +)$, $(\mathbb{Z}, +)$, $(\mathbb{Z}_N, +)$. We denote the additive identity e by 0.
2. Multiplicative groups: circle group $(\mathbb{T}, \cdot)$, Nth roots of unity $(\{\mathrm{e}^{\mathrm{i}\frac{2\pi}{N}k}\}_{k=0}^{N-1}, \cdot)$. We denote the multiplicative identity e by 1.

Some abelian groups can be generated by a few elements:

- $(\mathbb{Z}, +)$ can be generated by iteratively adding / subtracting 1.
- $(\{\mathrm{e}^{\mathrm{i}\frac{2\pi}{N}k}\}_{k=0}^{N-1}, \cdot)$ can be generated by taking integer powers of $\mathrm{e}^{\mathrm{i}\frac{2\pi}{N}}$.
- $(\mathbb{Z}_p \times \mathbb{Z}_q, +)$ with p and q coprime can be generated by iteratively adding $(1, 1)$.

Rmk. (Fundamental Theorem of Finitely Generated Abelian Groups)

Every *finitely-generated* abelian group can be written as

$$G \cong \mathbb{Z}_{m_1} \times \cdots \mathbb{Z}_{m_k} \times \mathbb{Z}^s, \tag{54}$$

where $m_1 \mid m_2$, ..., $m_{k-1} \mid m_k$.

Now, things get a bit abstract from here on.

Def. (Character)

Let us define a *homomorphism* on an abelian group G to $\mathbb{T}$: for all $x \in G$,

$$\chi_y(x) = \mathrm{e}^{\mathrm{i}2\pi yx} \in \mathbb{T}. \tag{55}$$

This is a *character* of the group, which is indexed by y.

Note that both x and y are merely indices.

A homomorphism φ on G is a function which preserves operation:

$$\varphi(g)\varphi(h) = \varphi(g \cdot h) \text{ or } \varphi(g)\varphi(h) = \varphi(g + h). \tag{56}$$

What are all the possible characters? Our task is to enumerate all possible and unique y's given the group G.

- If we let $G \cong \mathbb{Z}_N$, where $x \in \{0, 1, \ldots, N-1\}$, then the character $\chi_y(x) = \mathrm{e}^{\mathrm{i}2\pi yx}$ and its index y should satisfy:

  1. Homomorphism: since $Nx = 0$ for all $x \in G$. We have

$$\chi_y(Nx) = \mathrm{e}^{\mathrm{i}2\pi yNx} = 1 = \chi_y(0) \Rightarrow y \in \frac{1}{N} \cdot \mathbb{Z}.$$

  2. Uniqueness: note that for any given $y \in \frac{1}{N} \cdot \mathbb{Z}$, $\chi_y(x) = \chi_{y'}(x)$ for all x if $y' \in y + \mathbb{Z}$. Hence, the unique indices y are $\{0, 1/N, 2/N, \ldots, (N-1)/N\}$.

The structure of y's is the *same* as $\mathbb{Z}_N$.

- If we let $G \cong \mathbb{Z}$, where $x \in \{\ldots, -1, 0, 1, \ldots\}$. Then we have

    1. Homomorphism: since $c \cdot x = 0$ for $c \in G$ if and only if $c = 0$. This poses no constraints to the possible $y \in \mathbb{R}$.
    2. Uniqueness: note that $\mathrm{e}^{\mathrm{i}2\pi\mathbb{Z}} = 1$. Hence, $\chi_y = \chi_{y'}$ for all $y' \in y + \mathbb{Z}$, the only possible unique y's are $y \in [0, 1]$.

  Henceforth, we have that the structure of y's is the same as the interval $[0, 1]$ (or $[-1/2, 1/2]$ or $\mathbb{T}$).

- Similarly, try and show that if $G \cong \mathbb{T}$, with $x \in [-1/2, 1/2]$, the structure of y is the same as $\mathbb{Z}$.

- Similarly, try and show that if $G \cong \mathbb{R}$, with $x \in \mathbb{R}$, the structure of y is the same as $\mathbb{R}$.

Def. (Dual Group)

All unique indices y of the characters of an abelian group G forms its dual group $\widehat{G}$.

Dual group is a group. In brief, we have

$$\widehat{\mathbb{Z}_N} \cong \mathbb{Z}_N,\ \widehat{\mathbb{T}} \cong \mathbb{Z},\ \widehat{\mathbb{Z}} \cong \mathbb{T},\ \widehat{\mathbb{R}} \cong \mathbb{R}. \tag{57}$$

Further, we have $\widehat{G \times H} \cong \widehat{G} \times \widehat{H}$.

Thm. (Pontryagin Duality)

We have the natural isomorphism

$$\widehat{\widehat{G}} \cong G. \tag{58}$$

Thm. (Orthogonality of Characters)

If $\widehat{G}$ is discrete, then if $|G|$ is the number of elements in G, we have

$$\frac{1}{|\widehat{G}|}\sum_{y\in\widehat{G}}\chi_y(x)\chi_y^*(x') = \delta_{xx'}. \tag{59}$$

Due to the Pontryagin duality, we also have

$$\frac{1}{|G|}\sum_{x\in G}\chi_y(x)\chi_{y'}^*(x) = \delta_{yy'}. \tag{60}$$

If $\widehat{G}$ is continuous, then

$$\int_{\widehat{G}}\chi_y(x)\chi_y^*(x')\,\mathrm{d}y = \delta(x-x'). \tag{61}$$

For the following, all summation can be replaced by appropriate choices of integrals and measures. Some minor changes in the scaling constants should be noticed as well.

Any function on the group G can be decomposed into:

$$f(x) = \sum_{y \in \widehat{G}} \hat{f}(y) \chi_y(x). \tag{62}$$

Then by the orthogonality relation, one has

$$\hat{f}(y) = \frac{1}{|G|} \sum_{x \in G} f(x) \chi_y^*(x). \tag{63}$$

This is it! The duality between the four usual Fourier transforms are really just related via the dual relation between groups and its characters (dual group).

Besides extending to the product case: in the notation of quantum information science, since $\mathbb{Z}_{2^n} = \mathbb{Z}_2 \times \cdots \times \mathbb{Z}_2$,

$$F_{\mathbb{Z}_{2^n}} \left| x_{n-1} \dots x_1 x_0 \right\rangle = \frac{1}{\sqrt{N}} \sum_{y \in \mathbb{Z}_{2^n}} \mathrm{e}^{\mathrm{i}2\pi \frac{xy}{N}} \left| y_{n-1} \dots y_1 y_0 \right\rangle , \tag{64}$$

we can further extend this result to some nonabelian groups. Given a (special) group G with *representation* $\sigma(\cdot)$ of dimension d_σ, we can define the Fourier transform on G as

$$F_G \left| x \right\rangle = \sum_{\sigma,i,j} \sqrt{\frac{d_\sigma}{|G|}} \sigma_{ij}(x) \left| \sigma, i, j \right\rangle . \tag{65}$$

Table of Contents

In this subsection, we will introduce the Fourier transform on graphs. The graph Fourier transform is a useful tool in processing signals on a graph.

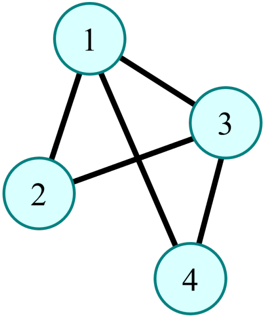
Def. (Graph)

An *unweighted* and *undirected* graph $G = (V, E)$ is defined by the set of vertices V and the set of edges E.

We often denote the number of vertices by $|V| = N$.

Def. (Degree)

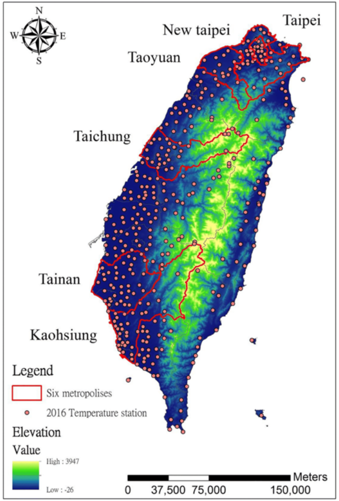
The degree of a node is the amount of edges it is connected to.



Normally in signal processing, our data is obtained from a sampling over a grid. This is, however, not as common as one may think. For example, the data from weather stations are scattered around the land.

How can one process the data / signals over the region? The *topology* of the graph must be included into the analysis. What are some *low and high spatial variation features* of the signal?

All these questions requires the knowledge of graph signal processing. The tool of graph Fourier transform proves to be important, too.



(Hsu et al., 2020)

If we can recall from Signals & Systems, we define the Fourier transform based on $\mathrm{e}^{\mathrm{i}2\pi kx}$ being the *eigenfunction* of an LTI system:

The shifting (lag) operator is defined as $\mathcal{T}_\delta$, with $\mathcal{T}_\delta f(x) = f(x+\delta)$, then

$$\mathcal{T}_\delta \mathrm{e}^{\mathrm{i}2\pi kx} = \mathrm{e}^{\mathrm{i}2\pi k\delta}\mathrm{e}^{\mathrm{i}2\pi kx}.$$

Taking the limit of

$$\lim_{\delta\to 0}\frac{1}{\delta}\left(\mathcal{T}_\delta - 1\right)\mathrm{e}^{\mathrm{i}2\pi kx} = \frac{\mathrm{d}}{\mathrm{d}x}\mathrm{e}^{\mathrm{i}2\pi kx} = \mathrm{i}2\pi k\,\mathrm{e}^{\mathrm{i}2\pi kx}. \tag{66}$$

We can define the Fourier transform by finding the eigenfunctions (an orthogonal basis) to the operators $\mathcal{T}_\delta$ or $\frac{\mathrm{d}}{\mathrm{d}x}$.

As we will see later on, it is a bit difficult to define both $\mathcal{T}_\delta$ or $\frac{\mathrm{d}}{\mathrm{d}x}$ on a graph. Instead, the eigenfunction to $\frac{\mathrm{d}}{\mathrm{d}x}$ coincides with both the gradient and the Laplacian operator:

$$\nabla_{\boldsymbol{x}} \mathrm{e}^{\mathrm{i}2\pi \boldsymbol{k}^\top \boldsymbol{x}} = \mathrm{i}2\pi \boldsymbol{k}\, \mathrm{e}^{\mathrm{i}2\pi \boldsymbol{k}^\top \boldsymbol{x}} \tag{67}$$

$$\nabla^2_{\boldsymbol{x}} \mathrm{e}^{\mathrm{i}2\pi \boldsymbol{k}^\top \boldsymbol{x}} = -\left\|2\pi \boldsymbol{k}\right\|^2 \mathrm{e}^{\mathrm{i}2\pi \boldsymbol{k}^\top \boldsymbol{x}}. \tag{68}$$

It is easy to define the Laplacian on a graph!

Our scheme will be as follows:

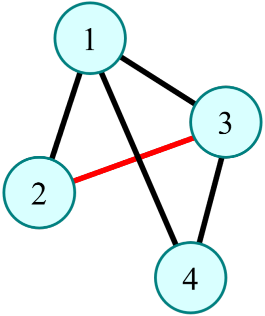
1. Find a suitable definition of the Laplacian L on a graph (in a matrix form).
2. Find its eigenvectors $L\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i$. They should be orthogonal.
3. Any signal $\boldsymbol{s}$ on the graph can be decomposed into $\boldsymbol{s} = \sum_i c_i \boldsymbol{v}_i$, with $c_i = \langle \boldsymbol{v}_i, \boldsymbol{s}\rangle / \langle \boldsymbol{v}_i, \boldsymbol{v}_i\rangle$.

Def. (Adjacency Matrix)

The adjacency matrix A encodes which vertices are joined by an edge, where

$$A_{ij} = \begin{cases} 1, & (i,j) \in E; \\ 0, & (i,j) \notin E. \end{cases} \tag{69}$$

$$A = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & {\color{red}1} & 0 \\ 1 & {\color{red}1} & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

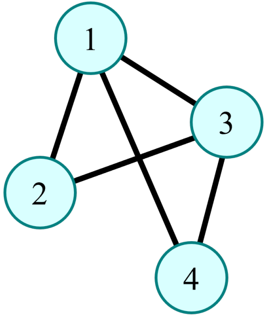


Def. (Laplacian Matrix)

Setting the degree matrix as the diagonal matrix D with $D_{ii} = \deg(i)$ the degree of the ith node, then the Laplacian matrix of the graph is defined as

$$L = D - A. \tag{70}$$

$$L = \begin{bmatrix} 3 & -1 & -1 & -1 \\ -1 & 2 & -1 & 0 \\ -1 & -1 & 3 & -1 \\ -1 & 0 & -1 & 2 \end{bmatrix}$$



Why is it *a* Laplacian?

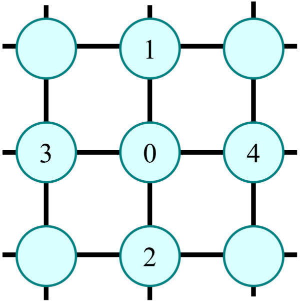
In the continuous case, the Laplacian is

$$\nabla^2 f(x) = \lim_{\delta \to 0} \frac{f(x+\delta) + f(x-\delta) - 2f(x)}{\delta^2}.$$

Let us consider a grid graph: the discrete Laplacian will be

$$-\nabla f(0) = 4f(0) - f(1) - f(2) - f(3) - f(4)$$

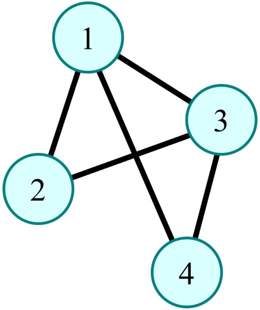
$$= \begin{bmatrix} & & \vdots & & \\ 4 & -1 & -1 & -1 & -1 \\ & & \vdots & & \end{bmatrix} \begin{bmatrix} f(0) \\ f(1) \\ f(2) \\ f(3) \\ f(4) \end{bmatrix} = [L\boldsymbol{f}]_0$$



Let us go on with the graph Fourier transform:

1. Find a suitable definition of the Laplacian L on a graph (in a matrix form).
2. Find its eigenvectors $L\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i$. They should be orthogonal.

$$L = \underbrace{\begin{bmatrix} \frac{1}{2} & 0 & \frac{-1}{2} & \frac{1}{\sqrt{2}} \\ \frac{1}{2} & \frac{1}{\sqrt{2}} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{-1}{2} & \frac{-1}{\sqrt{2}} \\ \frac{1}{2} & \frac{-1}{\sqrt{2}} & \frac{1}{2} & 0 \end{bmatrix}}_{V} \underbrace{\begin{bmatrix} 0 & & & \\ & 2 & & \\ & & 4 & \\ & & & 4 \end{bmatrix}}_{\Lambda} \underbrace{\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{\sqrt{2}} & 0 & \frac{-1}{\sqrt{2}} \\ \frac{-1}{2} & \frac{1}{2} & \frac{-1}{2} & \frac{1}{2} \\ \frac{1}{\sqrt{2}} & 0 & \frac{-1}{\sqrt{2}} & 0 \end{bmatrix}}_{V^{\mathsf{H}}} \tag{71}$$



3. Any signal $\boldsymbol{s}$ on the graph can be decomposed into $\boldsymbol{s} = \sum_i c_i \boldsymbol{v}_i$, with $c_i = \langle \boldsymbol{v}_i, \boldsymbol{s} \rangle / \langle \boldsymbol{v}_i, \boldsymbol{v}_i \rangle$.

Consider having a signal $\boldsymbol{s}$ on the vertices, we can decompose it via

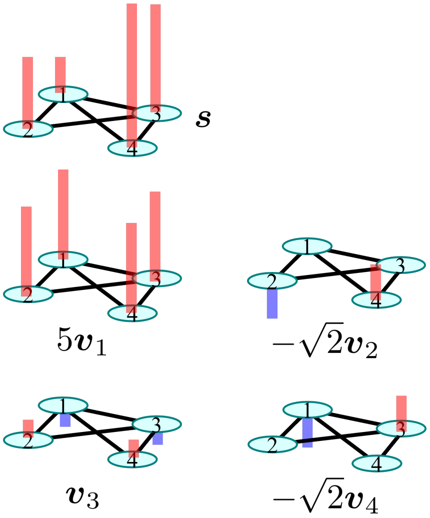
$$\boldsymbol{s} = \sum_i c_i \boldsymbol{v}_i = V\hat{\boldsymbol{s}} \tag{72}$$

$$\hat{\boldsymbol{s}} = V^{\mathsf{H}}\boldsymbol{s}. \tag{73}$$

For $\boldsymbol{s} = [1, 2, 3, 4]^{\mathsf{T}}$,

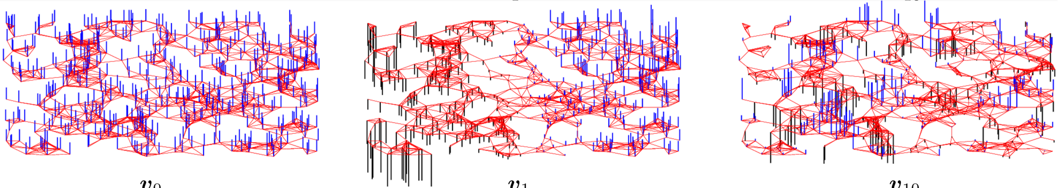
$$\hat{\boldsymbol{s}} = [5, -\sqrt{2}, 1, -\sqrt{2}]^{\mathsf{T}}.$$

The notion of *frequency* applies even on graphs!



From [5], one can see that the eigenvectors corresponding to smaller eigenvalues are those representing the low frequency features, and vice versa for the eigenvector corresponding to large eigenvalues.

$$\lambda_0 = 0 \quad < \quad \lambda_1 \quad < \quad \lambda_{10}$$



$\boldsymbol{v}_0$ $\boldsymbol{v}_1$ $\boldsymbol{v}_{10}$

Graph Fourier transform can also be utilized in noisy image filtering, gene relation analysis, and much more.

This lecture begins with reviewing the usual Fourier transforms and viewing them from the perspective of linear algebra. Then we digress to clarify some of the common misconceptions between distributions and functions.

Lastly, we see how Fourier transform is ubiquitous in the fields of engineering, especially in signal processing. Under different fields of study, it exhibits different forms. However, the main idea is the same:

representing a signal in the primal and dual coordinates interchangeably.

The kernels to the transform are just eigenfunctions / eigenvectors to the differential generator of the primal coordinate. In the case of linear differential equations, we have the complex exponentials; in the case of graph Fourier transform, we have the eigenvectors to the Laplacian.

[1] Joseph W. Goodman. *Introduction to Fourier optics*. 2nd. McGraw-Hill, 1969. URL: https://api.semanticscholar.org/CorpusID:118908270.

[2] Michel Alexis, Gevorg Mnatsakanyan, and Christoph Thiele. *Quantum signal processing and nonlinear Fourier analysis*. 2024. arXiv: 2310.12683 [quant-ph]. URL: https://arxiv.org/abs/2310.12683.

[3] *Fourier Series and their Convergence (Proof!) (Theory of Machine Learning)*. https://youtu.be/7KPaVLMiotE?si=xEozI8EhHk5eI9ut. Accessed: 2025-05-01.

[4] Thierry Blue. *Functional Analysis and Approximation Theory*. https://course.ntu.edu.tw/courses/113-2/13254. Accessed: 2025-05-09.

[5] David I Shuman et al. "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains". In: *IEEE Signal Processing Magazine* 30.3 (2013), pp. 83–98. DOI: 10.1109/MSP.2012.2235192.

[6] I.M. Gelfand and G. E. Shilov. *Generalized Functions*. Academic Press, 1968.

[7] Andrew Childs. *Lecture Notes on Quantum Algorithms*. https://www.cs.umd.edu/~amchilds/qa/, 2025.

[8] J.-P. Serre. *Linear Representations of Finite Groups*. Springer, 1977.

[9] Cheng-Wei Huang. *NTT_Python_Toolkit*. https://github.com/cheng-wei-huang0612/NTT_Python_Toolkit. Accessed: 2025-05-09.

[10] Cheng-Wei Huang. *ringfft*. https://github.com/RingFFT/ringfft. Accessed: 2025-05-09.

[11] Stefania Sardellitti, Sergio Barbarossa, and Paolo Di Lorenzo. "On the Graph Fourier Transform for Directed Graphs". In: *IEEE Journal of Selected Topics in Signal Processing* 11.6 (2017), pp. 796–811. DOI: 10.1109/JSTSP.2017.2726979.

[12] Stefania Sardellitti and Sergio Barbarossa. "Topological Signal Processing Over Generalized Cell Complexes". In: *IEEE Transactions on Signal Processing* 72 (2024), pp. 687–700. DOI: 10.1109/TSP.2023.3347069.

Notes: