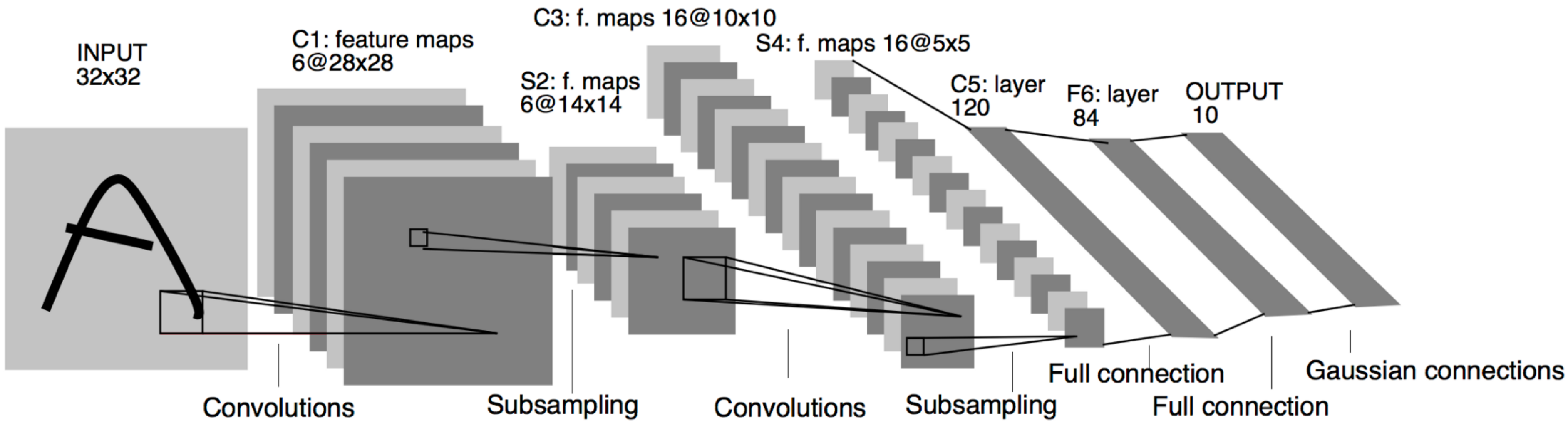
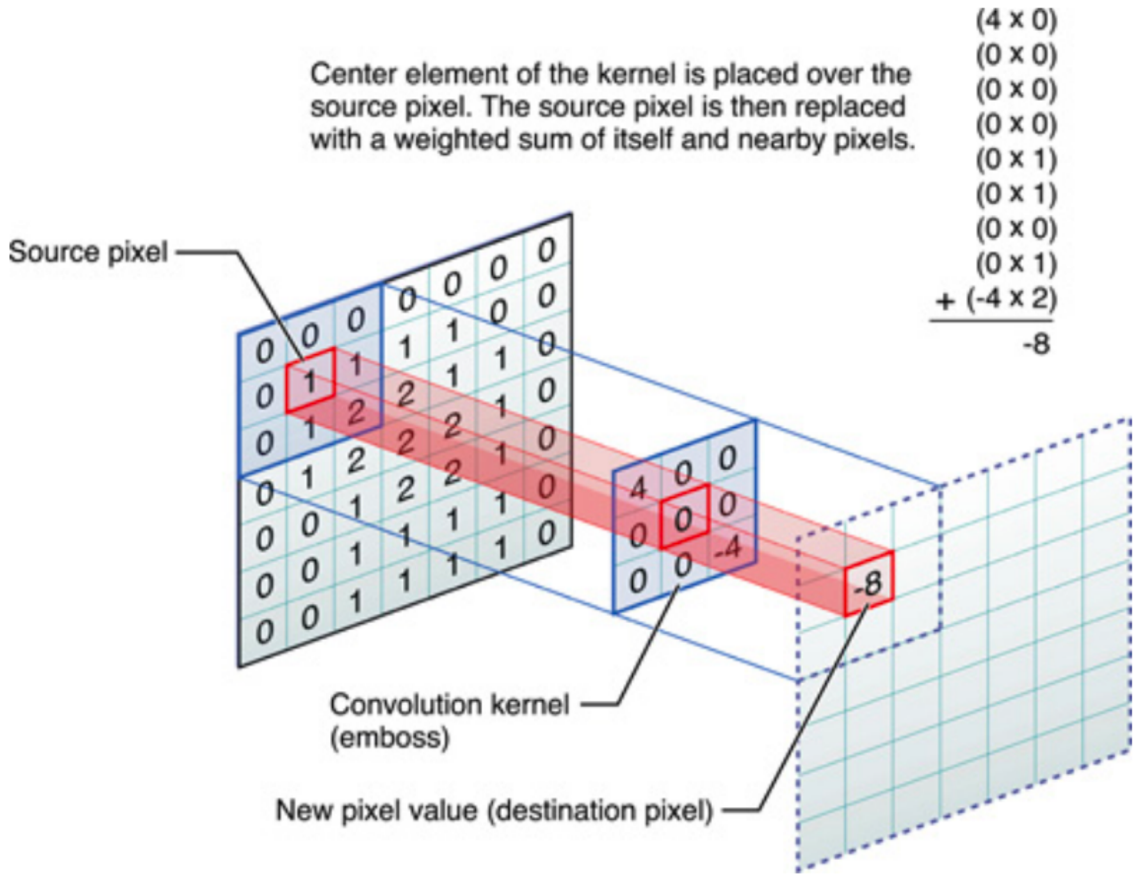
Convolution: basic definitions

Bamieh, arXiv:1805.05533



"Convolutional layers"



Convolution = filtering

Continuous signals

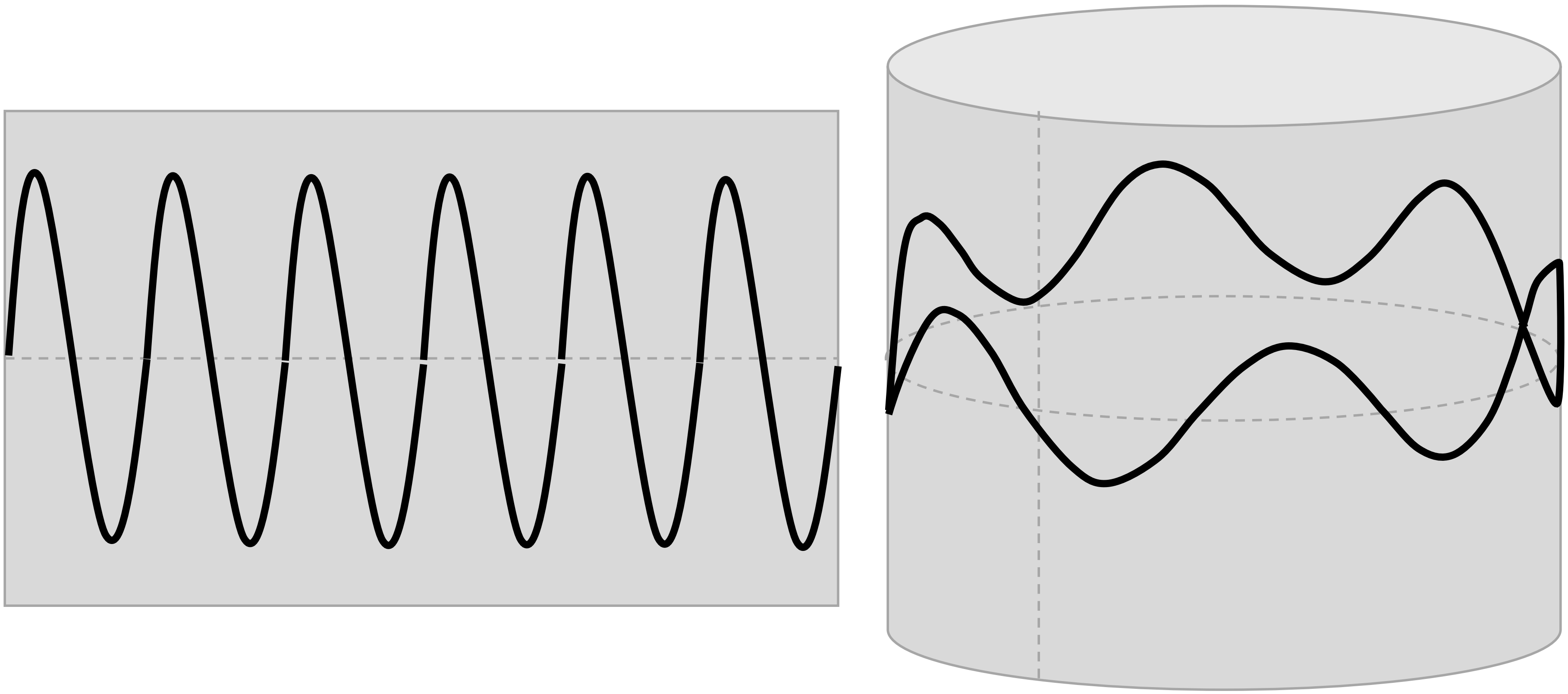
$$s(t) = \int x(a)w(t-a)da. \tag{9.1}$$

This operation is called **convolution**. The convolution operation is typically denoted with an asterisk:

$$s(t) = (x * w)(t). \tag{9.2}$$

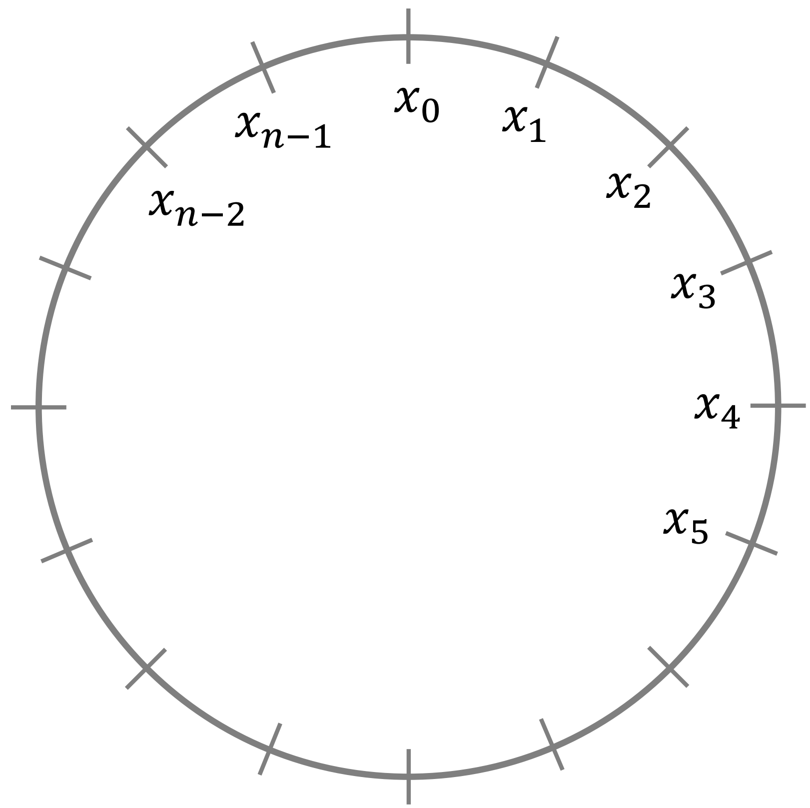
Goodfellow et al. 2016

Periodic signals



Periodic discrete signals

$$\mathbf{x} = \begin{bmatrix} x_0 \\ \vdots \\ x_{n-1} \end{bmatrix}$$



All indices are mod n

Circulant matrices

- **Circulant matrix**

$$\mathbf{C}(\mathbf{x}) = \begin{bmatrix} x_0 & x_{n-1} & \cdots & x_1 \\ x_1 & x_0 & & x_2 \\ \vdots & & \ddots & \vdots \\ x_{n-1} & x_{n-2} & \cdots & x_0 \end{bmatrix}$$

obtained by shifting the vector $\mathbf{x} = [x_0 \quad \cdots \quad x_{n-1}]^{\mathrm{T}}$

- Elements have the form $c_{ij}(\mathbf{x}) = x_{i-j \bmod n}$

All indices are $\mathrm{mod}\ n$

Convolution

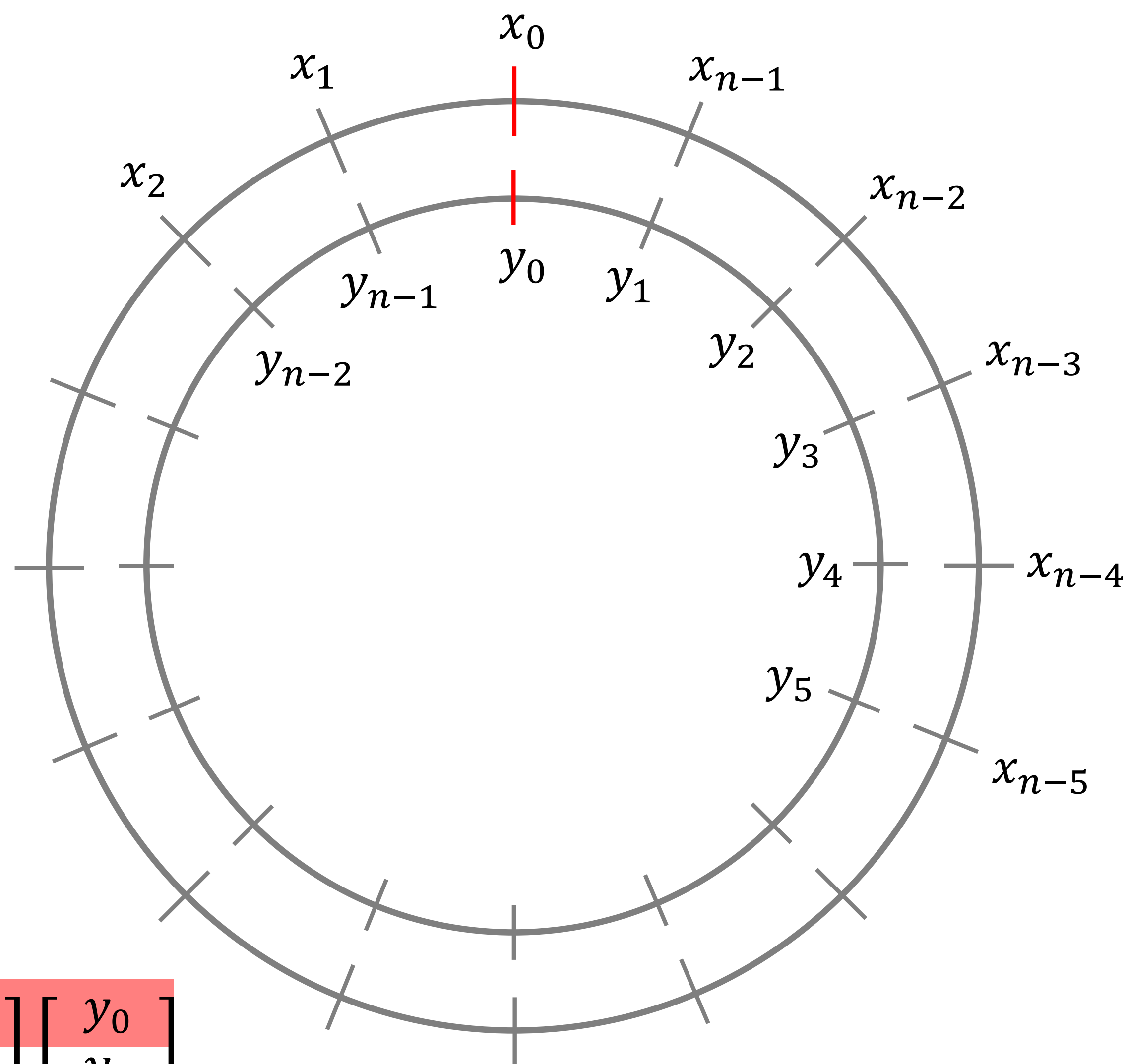
- Circulant matrices describe **circular convolution** operations

$$\mathbf{x} \star \mathbf{y} = \mathbf{C}(\mathbf{x})\mathbf{y}$$

$$(\mathbf{x} \star \mathbf{y})_i = \sum_{k=0}^{n-1} c_{ik}(\mathbf{x}) y_k = \sum_{k=0}^{n-1} x_{i-k} y_k$$

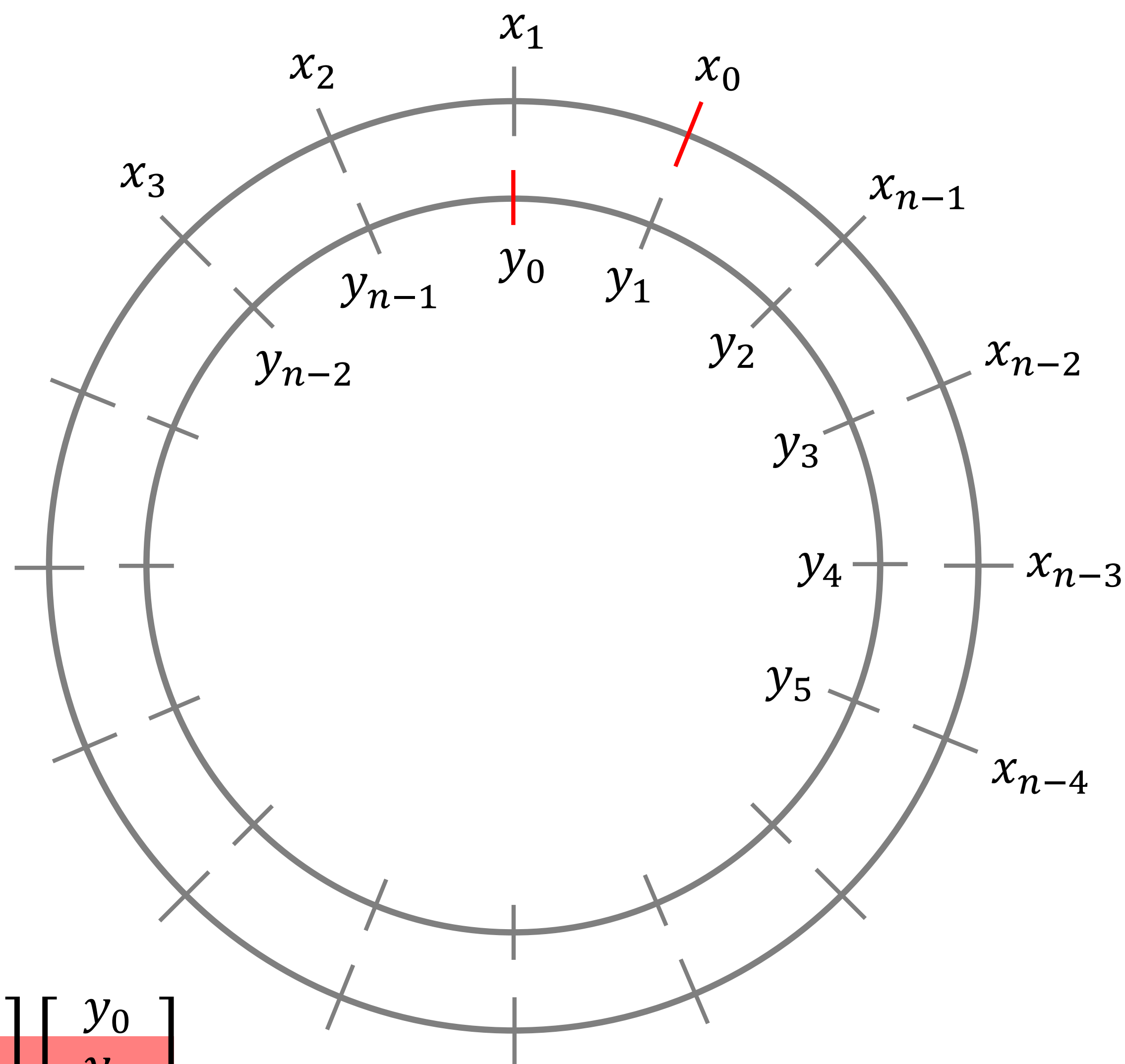
All indices are mod n

Convolution



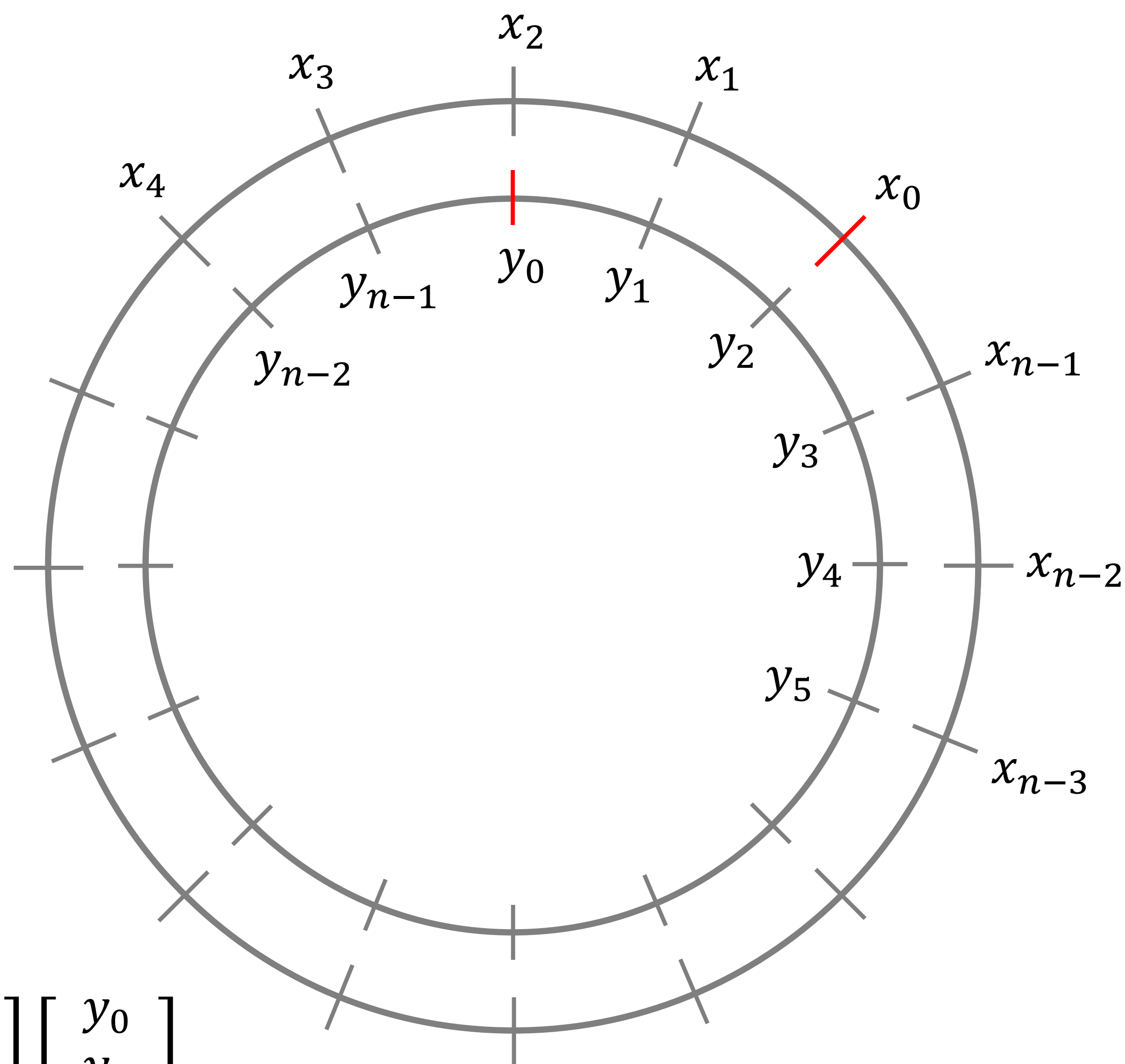
$$\begin{bmatrix} x_0 & x_{n-1} & \cdots & x_1 \\ x_1 & x_0 & & x_2 \\ \vdots & & \ddots & \vdots \\ x_{n-1} & x_{n-2} & \cdots & x_0 \end{bmatrix} \begin{bmatrix} y_0 \\ y_2 \\ \vdots \\ y_{n-1} \end{bmatrix}$$

Convolution



$$\begin{bmatrix} x_0 & x_{n-1} & \cdots & x_1 \\ x_1 & x_0 & & x_2 \\ & \vdots & \ddots & \vdots \\ x_{n-1} & x_{n-2} & \cdots & x_0 \end{bmatrix} \begin{bmatrix} y_0 \\ y_2 \\ \vdots \\ y_{n-1} \end{bmatrix}$$

Convolution



$$\begin{bmatrix} x_0 & x_{n-1} & \cdots & x_1 \\ x_1 & x_0 & & x_2 \\ & \vdots & \ddots & \vdots \\ x_{n-1} & x_{n-2} & \cdots & x_0 \end{bmatrix} \begin{bmatrix} y_0 \\ y_2 \\ \vdots \\ y_{n-1} \end{bmatrix}$$

Convolution

- Circulant matrices describe **circular convolution** operations

$$\mathbf{x} \star \mathbf{y} = \mathbf{C}(\mathbf{x})\mathbf{y}$$

$$(\mathbf{x} \star \mathbf{y})_i = \sum_{k=0}^{n-1} c_{ik}(\mathbf{x}) y_k = \sum_{k=0}^{n-1} x_{i-k} y_k$$

All indices are mod n

Convolution

- Circulant matrices describe **circular convolution** operations

$$\mathbf{y} \star \mathbf{x} = \mathbf{C}(\mathbf{y})\mathbf{x}$$

$$(\mathbf{y} \star \mathbf{x})_i = \sum_{k=0}^{n-1} c_{ik}(\mathbf{y}) x_k = \sum_{k=0}^{n-1} y_{i-k} x_k$$

by changing the index $l = i - k$:

$$= \sum_{l=0}^{n-1} y_{i-(i-l)}\, x_{i-l} = \sum_{l=0}^{n-1} y_i\, x_{i-l} = (\mathbf{x} \star \mathbf{y})_i$$

All indices are mod n

Properties of convolution

- **Commutative** $\mathbf{x} \star \mathbf{y} = \mathbf{y} \star \mathbf{x}$

- **Associative** $(\mathbf{x} \star \mathbf{y}) \star \mathbf{z} = \mathbf{x} \star (\mathbf{y} \star \mathbf{z})$

- Product of circulant matrices is also circulant

$$\mathbf{C}(\mathbf{x} \star \mathbf{y})\mathbf{z} = (\mathbf{x} \star \mathbf{y}) \star \mathbf{z} = \mathbf{x} \star (\mathbf{y} \star \mathbf{z}) = \mathbf{C}(\mathbf{x})\mathbf{C}(\mathbf{y})\mathbf{z}$$

All indices are $\mathrm{mod}\ n$

Shift operator

- **Right shift**

$$\mathbf{S} = \begin{bmatrix} 0 & & \cdots & & 1 \\ 1 & & & & \\ \vdots & & \ddots & \vdots & \\ & & \cdots & 1 & 0 \end{bmatrix}$$

Left shift

$$\mathbf{S}^{\mathrm{T}} = \begin{bmatrix} 0 & 1 & & \cdots & \\ & \vdots & & \ddots & \vdots \\ & & & & 1 \\ & & & \cdots & \\ 1 & & & & 0 \end{bmatrix}$$

- Elements have the form $s_{ij} = \delta_{i-j-1 \bmod n}$ and

$$(\mathbf{S}\mathbf{x})_i = x_{i-1 \bmod n} \qquad \left(\mathbf{S}^{\mathrm{T}}\mathbf{x}\right)_i = x_{i+1 \bmod n}$$

- Orthogonal matrix $\mathbf{S}^{\mathrm{T}}\mathbf{S} = \mathbf{S}\mathbf{S}^{\mathrm{T}} = \mathbf{I}$

- $\boldsymbol{S}$ has orthogonal eigenvectors

All indices are mod n

Commutativity with shift

$\mathbf{A}$ is circulant iff it commutes with the shift operator $\mathbf{S}$, i.e.

$$\mathbf{AS} = \mathbf{SA}$$

Proof:

- Assume $\mathbf{A} = \mathbf{C}(\mathbf{x})$ is circulant. Then

$$\big(\mathbf{SC}(\mathbf{x})\big)_{ij} = \sum_k s_{ik} c_{kj}(\mathbf{x}) = \sum_k \delta_{i-j-1}\, x_{k-j} = x_{i-j-1}$$

$$(\mathbf{C}(\mathbf{x})\mathbf{S})_{ij} = \sum_k c_{ik}(\mathbf{x}) s_{kj} = \sum_k x_{i-k}\, \delta_{k-j-1} = x_{i-j-1}$$

All indices are mod n

Commutativity with shift

> $\mathbf{A}$ is circulant iff it commutes with the shift operator $\mathbf{S}$, i.e.
> $$\mathbf{AS} = \mathbf{SA}$$

Proof:

- Assume $\mathbf{AS} = \mathbf{SA}$ for some $\mathbf{A}$. Then

$$(\mathbf{SA})_{ij} = \sum_k s_{ik} a_{kj} = \sum_k \delta_{i-j-1} a_{kj} = a_{i-1,j}$$

$$(\mathbf{AS})_{ij} = \sum_k a_{ik} s_{kj} = \sum_k a_{ik}\, \delta_{k-j-1} = a_{i,j+1}$$

- The equality $a_{i-1,j} = a_{i,j+1}$ (all modulo n) implies $\mathbf{A}$ is circulant (alternatively, $a_{i,j} = a_{i+k,j+k}$ for any k)

All indices are mod n

Shift ~~invariance~~ equivariance

$$(\mathbf{Sx}) \star \mathbf{y} = \mathbf{y} \star (\mathbf{Sx}) = \mathbf{C}(\mathbf{y})\mathbf{Sx} = \mathbf{SC}(\mathbf{y})\mathbf{x} = \mathbf{S}(\mathbf{x} \star \mathbf{y})$$

convolution = linear shift-equivariant operator

Joint diagonalization

$\mathbf{A}$ and $\mathbf{B}$ are **jointly diagonalizable** (diagonalized by the same set of eigenvectors) iff they commute, i.e

$$\mathbf{AB} = \mathbf{BA}$$

shift equivariance $\Rightarrow$

convolution commutes with shift $\Rightarrow$

convolution is diagonalized by the eigenvectors of the shift operator

Eigenvalues of the shift operator $\mathbf{S}^{\mathrm{T}}$

$$\mathbf{S}^{\mathrm{T}}\mathbf{v} = \lambda\mathbf{v} \quad \Leftrightarrow \quad v_{i+1} = \lambda v_i \;\text{ for all } i \in \mathbb{Z}^n$$

$$\left(\mathbf{S}^{\mathrm{T}}\right)^2\mathbf{v} = \lambda^2\mathbf{v} \quad \Leftrightarrow \quad v_{i+2} = \lambda^2 v_i$$

$$\vdots$$

$$\left(\mathbf{S}^{\mathrm{T}}\right)^n\mathbf{v} = \lambda^n\mathbf{v} \quad \Leftrightarrow \quad v_{i+n} = v_i = \lambda^n v_i$$

since $\mathbf{v} \neq \mathbf{0}$, eigenvalues must be roots of unity $\lambda^n = 1$

$$\lambda = e^{\iota\frac{2\pi}{n}k} \qquad \text{for } k \in \mathbb{Z}^n$$

All indices are mod n

Eigenvectors of the shift operator $\mathbf{S}^{\mathrm{T}}$

$$v_k = \lambda^k v_0 = e^{\iota \frac{2\pi}{n} k} v_0$$

Matrix of eigenvectors

$$\mathbf{V} = v_0 \begin{bmatrix} | & | & & | \\ \boldsymbol{\varphi}^0 & \boldsymbol{\varphi}^1 & & \ldots \boldsymbol{\varphi}^{n-1} \\ | & | & & | \end{bmatrix}$$

where $\boldsymbol{\varphi} = \begin{bmatrix} 1 & e^{\iota \frac{2\pi}{n}} & e^{\iota \frac{2\pi}{n} 2} & \ldots & e^{\iota \frac{2\pi}{n}(n-1)} \end{bmatrix}^{\mathrm{T}}$

Eigenvectors of the shift operator $\mathbf{S}^{\mathrm{T}}$

$$v_k = \lambda^k v_0 = e^{\iota \frac{2\pi}{n} k} v_0$$

Matrix of eigenvectors

$$\mathbf{V} = \frac{1}{\sqrt{n}} \begin{bmatrix} | & | & & | \\ \mathbf{1} & \boldsymbol{\varphi}^1 & \ldots & \boldsymbol{\varphi}^{n-1} \\ | & | & & | \end{bmatrix}$$

where $\boldsymbol{\varphi} = \begin{bmatrix} 1 & e^{\iota \frac{2\pi}{n}} & e^{\iota \frac{2\pi}{n} 2} & \ldots & e^{\iota \frac{2\pi}{n}(n-1)} \end{bmatrix}^{\mathrm{T}}$ and eigenvectors are unit length

$$\|\mathbf{v}_k\| = \left\| \frac{1}{\sqrt{n}} \boldsymbol{\varphi}^k \right\| = \left(\sum_{k=0}^{n-1} \left| \frac{1}{\sqrt{n}} e^{\iota \frac{2\pi}{n} k} \right|^2 \right)^{1/2} = 1$$

Fourier transform

$$\mathbf{V} = \frac{1}{\sqrt{n}} \begin{bmatrix} | & | & | \\ \mathbf{1} & \boldsymbol{\varphi}^1 & \ldots \boldsymbol{\varphi}^{n-1} \\ | & | & | \end{bmatrix}$$

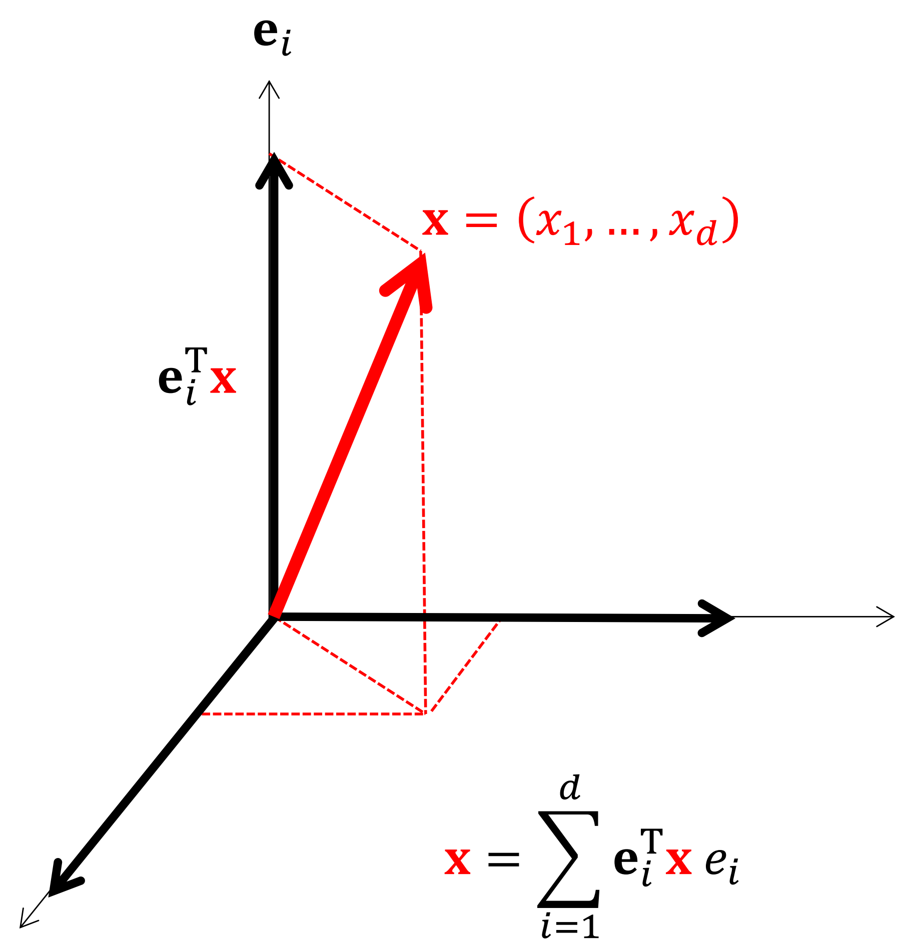
- The matrix is orthogonal $\mathbf{V}^*\mathbf{V} = \mathbf{I}$
- Its columns form an orthonormal basis called **Fourier basis**
- **Discrete Fourier transform (DFT)**

$$\hat{\mathbf{x}} = \mathbf{V}^*\mathbf{x}$$

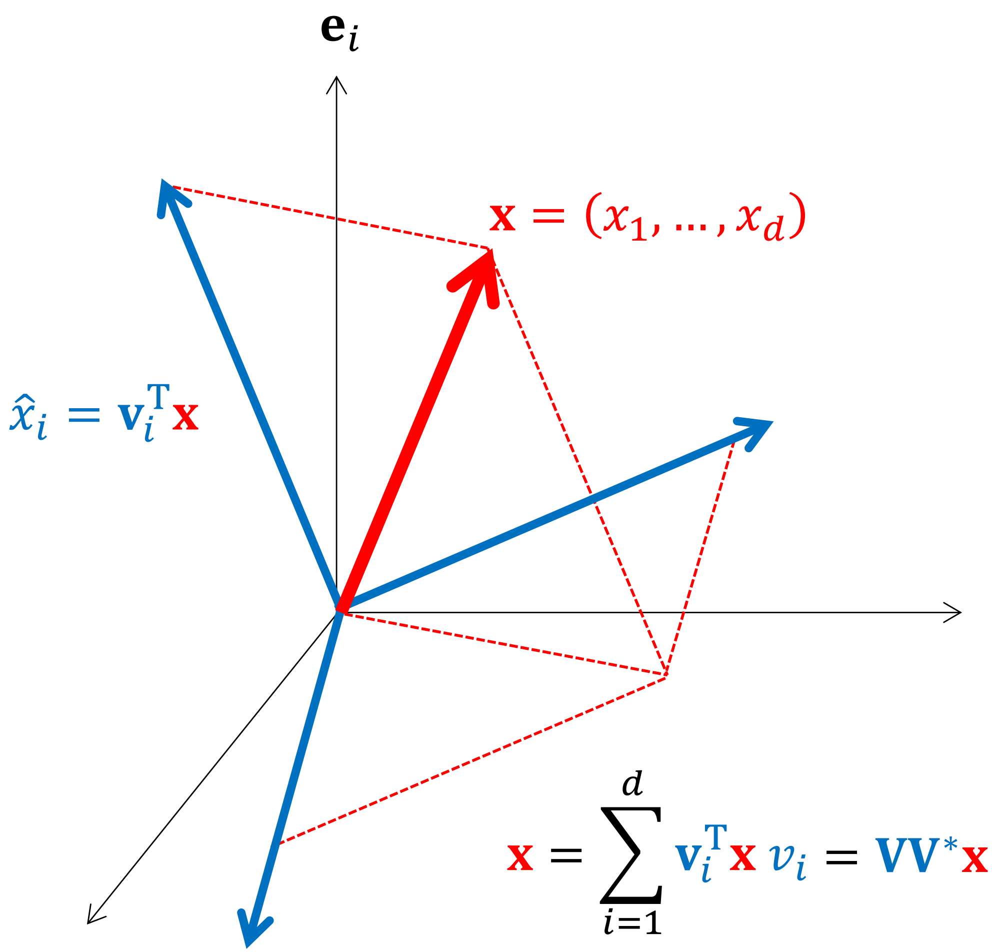
$$\hat{x}_k = \frac{1}{\sqrt{n}} \sum_{l=0}^{n-1} x_l e^{-\iota \frac{2\pi}{n} kl}$$

* In fact, any orthonormal basis is a Fourier basis

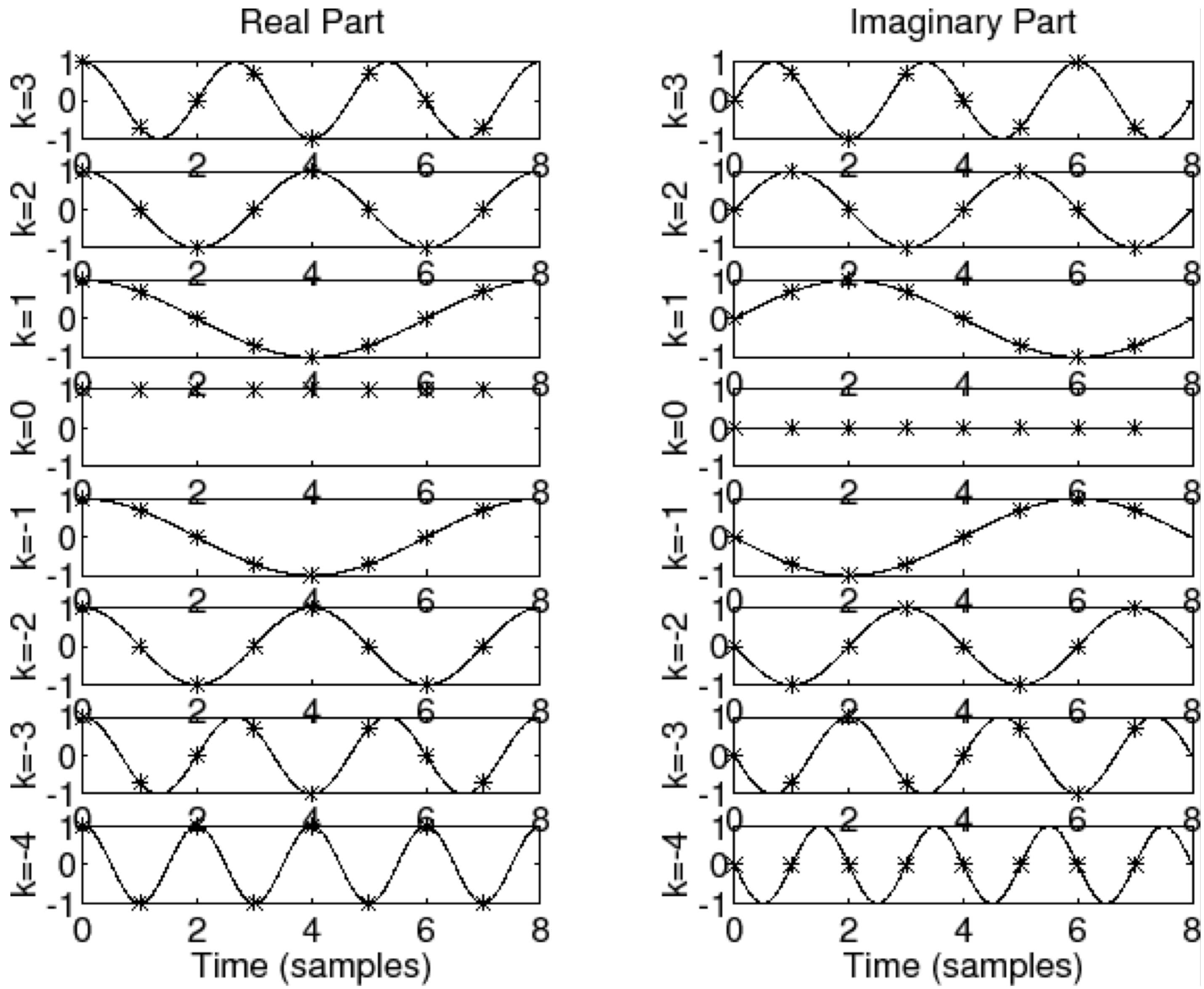
Fourier transform: geometric intuition



Fourier transform: geometric intuition



Fourier basis



Fourier transform: physical intuition

$$\boldsymbol{\Delta} = \begin{bmatrix} 1 & -2 & \cdots & 0 & 1 \\ & \vdots & \ddots & \vdots & \\ & & \cdots & & \\ -2 & 1 & & 0 & 1 \end{bmatrix}$$

- Second-order derivative (finite difference)
- Commutes with shift (diagonalized by the Fourier basis)
- **Dirichlet energy** = smoothness of a signal

$$\mathcal{E}(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\boldsymbol{\Delta}\mathbf{x}$$

Fourier transform: physical intuition

$$\min_{\mathrm{X}} \mathbf{X}^{\mathrm{T}} \boldsymbol{\Delta} \mathbf{X} = \min_{\mathrm{X}} \sum_{i=1}^{n} \mathcal{E}(\mathbf{x}_i) \;\; \mathrm{s.t.} \quad \mathbf{X}^{\mathrm{T}} \mathbf{X} = \mathbf{I}$$

- Solution given by the eigenvectors of $\boldsymbol{\Delta}$ = Fourier basis
- Fourier basis = “smoothest orthogonal basis”

Fast Fourier Transform (FFT)

$$\sqrt{n}\hat{x}_k = \sum_{l=0}^{n-1} x_l e^{-\iota\frac{2\pi}{n}kl} = \sum_{m=0}^{n/2-1} x_{2m} e^{-\iota\frac{2\pi}{n}k(2m)} + \sum_{l=0}^{n-1} x_{2m+1} e^{-\iota\frac{2\pi}{n}k(2m+1)}$$

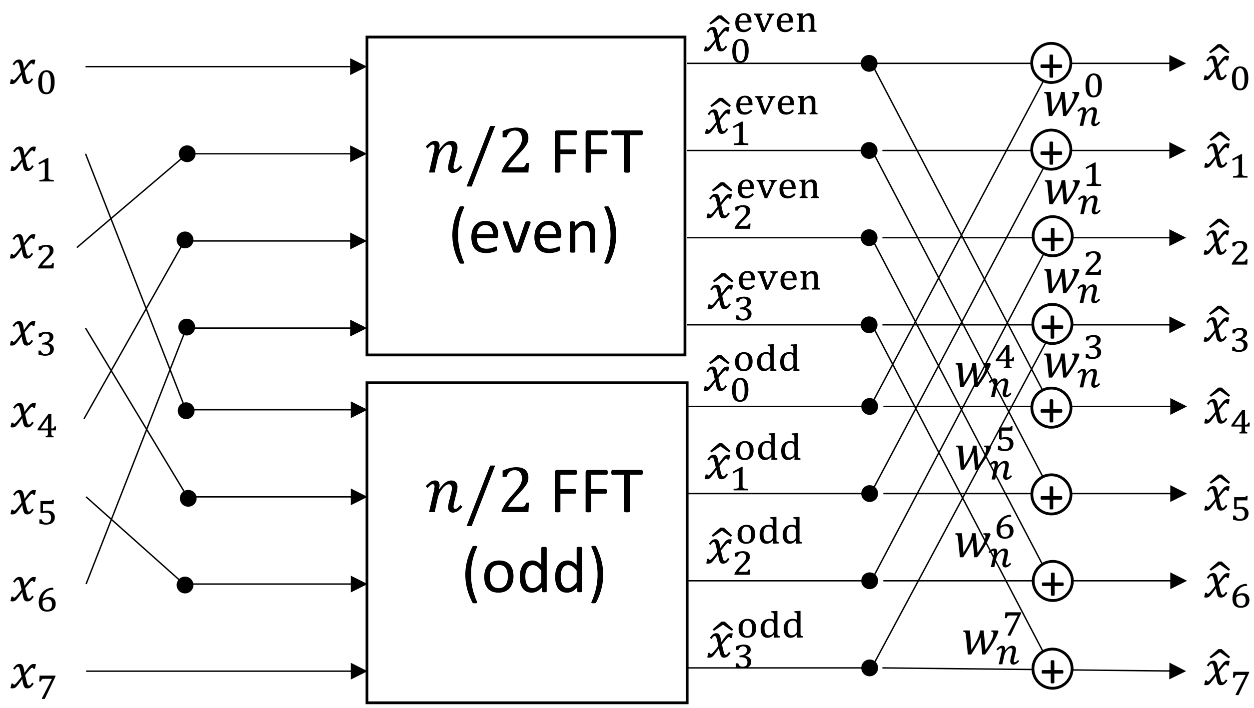
$$= \sum_{m=0}^{n/2-1} x_{2m} e^{-\iota\frac{2\pi}{n/2}km} + e^{-\iota\frac{2\pi}{n}k} \sum_{l=0}^{n-1} x_{2m+1} e^{-\iota\frac{2\pi}{n/2}k}$$

$$= \sqrt{n/2}\ \hat{x}_k^{\text{even}} + e^{-\iota\frac{2\pi}{n}k}\sqrt{n/2}\ \hat{x}_k^{\text{odd}} \qquad k = 0, \ldots, \frac{n}{2} - 1$$

$$\sqrt{n}\hat{x}_{k+n/2} = \cdots = \sqrt{n/2}\ \hat{x}_k^{\text{even}} - e^{-\iota\frac{2\pi}{n}k}\sqrt{n/2}\ \hat{x}_k^{\text{odd}}$$

Cooley, Tukey 1965 (ideas date back to Gauss)

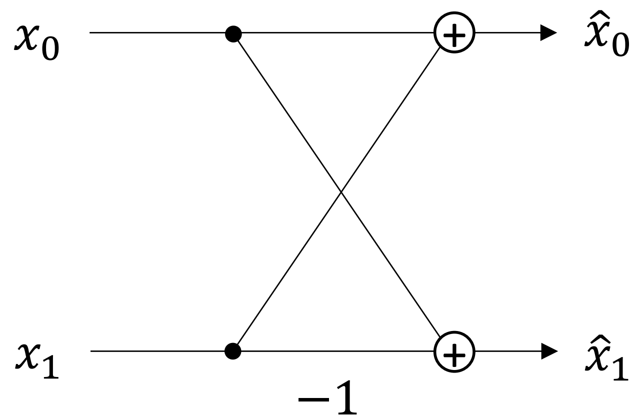
Fast Fourier Transform (FFT)



$$w_n^k = e^{-\iota \frac{2\pi}{n} k}$$

Cooley, Tukey 1965 (ideas date back to Gauss)

Fast Fourier Transform (FFT): butterfly



2 FFT

Cooley, Tukey 1965 (ideas date back to Gauss)

Joint diagonalization

$\mathbf{A}$ and $\mathbf{B}$ are **jointly diagonalizable** (diagonalized by the same set of eigenvectors) iff they commute, i.e

$$\mathbf{AB} = \mathbf{BA}$$

shift equivariance $\Rightarrow$ convolution is diagonalized by the eigenvectors of the shift operator

Eigenvalues of circulant matrix $\mathbf{C}(\mathbf{x})$

- $\mathbf{C}(\mathbf{x})$ and $\mathbf{S}^{\mathrm{T}}$ have same eigenvectors $\mathbf{V}$ but different eigenvalues

$$\mathbf{C}(\mathbf{x})\mathbf{v}_k = \mu_k \mathbf{v}_k$$

$$\begin{bmatrix} x_0 & x_{n-1} & & x_1 \\ x_1 & x_0 & \cdots & x_2 \\ & \vdots & \ddots & \vdots \\ x_{n-1} & x_{n-2} & \cdots & x_0 \end{bmatrix} \begin{bmatrix} 1 \\ e^{\iota \frac{2\pi}{n} k} \\ \vdots \\ e^{\iota \frac{2\pi}{n} k(n-1)} \end{bmatrix} = \mu_k \begin{bmatrix} 1 \\ e^{\iota \frac{2\pi}{n} k} \\ \vdots \\ e^{\iota \frac{2\pi}{n} k(n-1)} \end{bmatrix}$$

$$\mu_k = x_0 + x_{n-1} e^{\iota \frac{2\pi}{n} k} + \cdots + x_1 e^{\iota \frac{2\pi}{n} k(n-1)}$$

$$= x_0 + x_1 e^{-\iota \frac{2\pi}{n} k} + \cdots + x_{n-1} e^{-\iota \frac{2\pi}{n} k(n-1)}$$

$$= \sum_{l=0}^{n-1} x_l e^{-\iota \frac{2\pi}{n} kl} = \hat{x}_k$$

Convolution in the Fourier domain

$$\mathbf{C}(\mathbf{x}) = \mathbf{V}\begin{bmatrix} \hat{x}_0 & \cdots & \\ \vdots & \ddots & \vdots \\ & \cdots & \hat{x}_{n-1} \end{bmatrix}\mathbf{V}^{\mathrm{T}}$$

In order to compute convolution $\mathbf{x} \star \mathbf{y}$

- **Perform FFT:** $\hat{\mathbf{y}} = \mathbf{V}^{\mathrm{T}}\mathbf{y}$ $\quad O(n\log n)$
- **Apply filter:** $\hat{\mathbf{x}} \circ \hat{\mathbf{y}} = \begin{bmatrix} \hat{x}_0 & \cdots & \\ \vdots & \ddots & \vdots \\ & \cdots & \hat{x}_{n-1} \end{bmatrix}\hat{\mathbf{y}}$ $\quad O(n)$

  "Hadamard product"
- **Perform inverse FFT:** $\mathbf{x} \star \mathbf{y} = \mathbf{V}(\hat{\mathbf{x}} \circ \hat{\mathbf{y}})$ $\quad O(n\log n)$

Convolution in the Fourier domain

$$
\begin{tikzcd}[column sep=huge, row sep=huge]
\mathbf{y} \arrow[r, "\mathbf{C}(\mathbf{x})"] \arrow[d, shift left, "\mathbf{V}^{\mathrm{T}}"] & \mathbf{x} \star \mathbf{y} \arrow[d, shift left, "\mathbf{V}^{\mathrm{T}}"] \\
\hat{\mathbf{y}} \arrow[r, swap, "{\begin{bmatrix} \hat{x}_0 & \cdots & \\ \vdots & \ddots & \vdots \\ & \cdots & \hat{x}_{n-1} \end{bmatrix}}"] \arrow[u, shift left, "\mathbf{V}"] & \widehat{\mathbf{x} \star \mathbf{y}} \arrow[u, shift left, "\mathbf{V}"]
\end{tikzcd}
$$