# Signal and information processing in time

Alejandro Ribeiro
Dept. of Electrical and Systems Engineering
University of Pennsylvania
aribeiro@seas.upenn.edu
http://www.seas.upenn.edu/users/~aribeiro/

April 25, 2016

# Two dimensional (2D) DFT

Two dimensional (2D) discrete Fourier transform (DFT)

Discrete Cosine Transform

The discrete Fourier transform with Hermitian matrices

Principal Component Analysis (PCA) transform

Graph Signals

Graph Fourier Transform (GFT)

Information sciences at ESE

# Definition of 2D DFT

- 2D signal $x$ With $N$ rows and $M$ columns. Elements $x(m, n)$
- We will focus on signals with $M = N$. To simplify notation.

- Signal $X$ is the 2D DFT of $x$ if its elements $X(k, l)$ are

$$X(k,l) := \frac{1}{N} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} x(m,n) e^{-j2\pi(km+ln)/N}$$

- As in 1D we write $X = \mathcal{F}(x)$.
- $X$ may be complex even for real 2D signals $x$. Focus on magnitude.

- Argument $k$ is horizontal frequency and $l$ is the vertical frequency

# The 2D DFT and the (regular, 1D) DFT

- Separate terms in the exponent and regroup factors to write

$$X(k,l) := \frac{1}{\sqrt{N}} \sum_{m=0}^{N-1} \left[ \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x(m,n) e^{-j2\pi l n/N} \right] e^{-j2\pi k m/N}$$

- For fixed $m$, the term between parentheses is the DFT of $x(m,\cdot)$
- We then take the DFT of the resulting DFTs with respect to $m$
- The 2D DFT of $x$ is the column-wise DFT of the row-wise DFTs
- Or the row-wise DFT of the column-wise DFTs. Just the same

# Discrete complex exponentials

- 2D Complex exponential of horizontal freq. $k$ and vertical freq. $l$

$$e_{klN}(m,n) = \frac{1}{N} e^{j2\pi(km+ln)/N} = \frac{1}{\sqrt{N}} e^{j2\pi(km/N)} \frac{1}{\sqrt{N}} e^{j2\pi(ln/N)}$$

- Separate the exponential into two factors to write

$$e_{klN}(m,n) = \frac{1}{\sqrt{N}} e^{j2\pi(km/N)} \frac{1}{\sqrt{N}} e^{j2\pi(ln/N)} = e_{kN}(m) e_{lN}(n)$$

- 2D complex exponential is product of two 1D complex exponentials

Theorem

*Complex exponentials with nonequivalent frequencies are orthogonal*

$$\langle e_{klN}, e_{\tilde{k}\tilde{l}N} \rangle = \delta(k-\tilde{k})\delta(l-\tilde{l})$$

# Definition of 2D iDFT

- Given a Fourier transform $X$, the inverse (i)DFT $x = \mathcal{F}^{-1}(X)$ is

$$x(m,n) := \frac{1}{N}\sum_{k=0}^{N-1}\sum_{l=0}^{N-1} X(k,l) e^{j2\pi(km+ln)/N}$$

- Sum is over horizontal and vertical frequencies dimensions

- Recall that 2D DFT has period $N$ in vertical and horizontal freqs.
- Any summation over $M \times N$ adjacent frequencies works as well. E.g.,

$$x(m,n) = \frac{1}{N}\sum_{k=-N/2+1}^{N/2}\sum_{l=-N/2+1}^{N/2} X(k,l) e^{j2\pi(km+ln)/N}$$

# iDFT is, indeed, the inverse of the DFT

Theorem

*The 2D inverse DFT* $\tilde{x} = \mathcal{F}^{-1}(X)$ *of the 2D DFT* $X = \mathcal{F}(x)$ *of any given signal* $x$ *is the original signal* $x$

$$\tilde{x} \equiv \mathcal{F}^{-1}(X) \equiv \mathcal{F}^{-1}(\mathcal{F}(x)) \equiv x$$

- Every 2D signal can be written as a sum of 2D complex exponentials

$$x(m, n) := \frac{1}{N} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} X(k, l) e^{j2\pi(km+ln)/N}$$

- The coefficient for horizontal frequency $k$ and vertical frequency $f$ is

$$X(k, l) := \frac{1}{N} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} x(m, n) e^{-j2\pi(km+ln)/N}$$

# Discrete Cosine Transform

Two dimensional (2D) discrete Fourier transform (DFT)

Discrete Cosine Transform

The discrete Fourier transform with Hermitian matrices

Principal Component Analysis (PCA) transform

Graph Signals

Graph Fourier Transform (GFT)

Information sciences at ESE

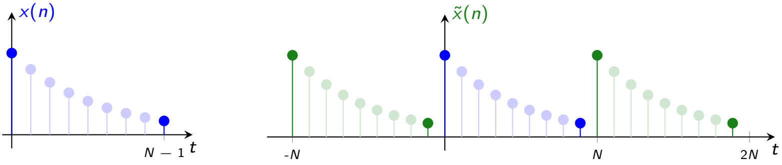
# Border effects in image compression

- Patches are well approximated by a subset of 2D DFT coefficients
- Except for borders. And still a problem if we retain most coefficients


- Although didn't mention, also a problem with (1D) DFTs $\Rightarrow$ Why?

# The periodic extension of the iDFT

- First sample $x(0)$ and last sample $x(N-1)$ can be very different
  - $\Rightarrow$ Most likely are. Unless signal has some structure, e.g., symmetry
- This is a problem for the periodic extension
  - $\Rightarrow$ The value $x(0) = \tilde{x}(N)$ appears next to $x(N-1) = \tilde{x}(N-1)$




- It's tough to approximate a jump/discontinuity $\Rightarrow$ High frequency
- Never mind. We're more than Fourier people. We're fearless transformers

# Inverse discrete cosine transform

- Say that we have a transform $X$ so that we can write signal $\tilde{x}$ as

$$\tilde{x}(n) := \frac{1}{\sqrt{N}} X(0) + \sqrt{\frac{2}{N}} \sum_{k=1}^{N-1} X(k) \cos\left[\frac{\pi k(2n+1)}{2N}\right]$$

- No complex numbers involved. Signals and transforms assumed real

- Haven't said how to find $X$ so that $\tilde{x}(n) = x(n)$ for $n \in [0, N-1]$
- This is done with discrete cosine transform (DCT). We'll see later

- Details are different but this is still $x$ written as a sum of oscillations
  - $\Rightarrow$ Still expect low frequency components to be most significant
  - $\Rightarrow$ But have written cosine in a way that avoids border discontinuities

# The even extension of the iDCT

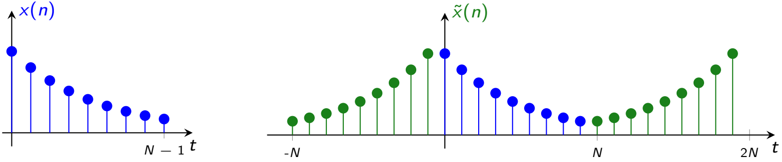
- Formalize argument to prove that the iDCT yields an even extension

$$\tilde{x}\Big[N + (n-1)\Big] = x\Big[N - n\Big]$$

- Or, to better visualize the symmetry

$$\tilde{x}\Big[(N - 1/2) + (n - 1/2)\Big] = x\Big[(N - 1/2) - (n - 1/2)\Big]$$

- Signal $x$ written as sum of oscillations without border effects



# The discrete cosine transform (DCT)

- Still have to find out a way of computing the coefficients $X(k)$
- Given a real signal $x$, the DCT $X = \mathcal{C}(x)$ is the real signal with

$$X(0) := \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x(n) \cos\left[\frac{\pi 0(2n+1)}{2N}\right]$$

$$X(k) := \sqrt{\frac{2}{N}} \sum_{n=0}^{N-1} x(n) \cos\left[\frac{\pi k(2n+1)}{2N}\right]$$

- Normalization constants are different for $k = 0$ and $k \in [1, N-1]$
- No complex numbers involved. Signals and transforms are real

# DCT basis

- Define the elements of the DCT basis as the signals $c_{kN}$ with

$$c_{0N}(n) := \frac{1}{\sqrt{N}} \qquad c_{kN}(n) := \sqrt{\frac{2}{N}} \cos\left[\frac{\pi k(2n+1)}{2N}\right]$$

- Akin to the DFT basis defined by the $N$ complex exponentials $e_{kN}$
- With basis defined can write DCT of $x$ as $\Rightarrow X(k) = \langle x, c_{kN} \rangle$
- Inner product implies the usual interpretation

  $\Rightarrow$ X(k) is how much $x(n)$ resembles oscillation of frequency $k$

# iDCT is the inverse of the DCT

**Theorem**

*The iDCT $\tilde{x} = \mathcal{C}^{-1}(X)$ of the DCT $X = \mathcal{C}(x)$ of any given signal $x$ is the original signal $x$, i.e.,*

$$\tilde{x} \equiv \mathcal{C}^{-1}(X) \equiv \mathcal{C}^{-1}(\mathcal{C}(x)) \equiv x$$

- Equivalence means $\tilde{x}(n) = x(n)$ for $n \in [0, N-1]$.
  - $\Rightarrow$ Otherwise, inverse transform $\tilde{x}$ is an even extension of original $x$
- To prove theorem, use DCT definition, iDCT definition, reverse summation order, and invoke orthogonality of the DCT basis.
- Conservation of energy (Parseval's) also holds $\Rightarrow$ orthogonality

# The DFT and iDFT with Hermitian matrices

# The discrete Fourier transform, again

- It is time to write and understand the DFT in a more abstract way
- Write signal $x$ and complex exponential $e_{kN}$ as vectors $\mathbf{x}$ and $\mathbf{e}_{kN}$

$$\mathbf{x} = \begin{pmatrix} x(0) \\ x(1) \\ \vdots \\ x(N-1) \end{pmatrix} \qquad \mathbf{e}_{kN} = \frac{1}{\sqrt{N}} \begin{pmatrix} e^{j2\pi k0/N} \\ e^{j2\pi k1/N} \\ \vdots \\ e^{j2\pi k(N-1)/N} \end{pmatrix}$$

- Use vectors to write the $k$th DFT component as ($\mathbf{e}_{kN}^H = (\mathbf{e}_{kN}^*)^T$)

$$X(k) \;=\; \mathbf{e}_{kN}^H \mathbf{x} \;=\; \langle \mathbf{x}, \mathbf{e}_{kN} \rangle \;=\; \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x(n) e^{-j2\pi kn/N}$$

- $k$th DFT component $X(k)$ is the product of $\mathbf{x}$ with exponential $\mathbf{e}_{kN}^H$

# DFT in matrix notation

- Write DFT $\mathbf{X}$ as a stacked vector and stack individual definitions

$$\mathbf{X} = \begin{bmatrix} X(0) \\ X(1) \\ \vdots \\ X(N-1) \end{bmatrix} = \begin{bmatrix} \mathbf{e}_{0N}^H \mathbf{x} \\ \mathbf{e}_{1N}^H \mathbf{x} \\ \vdots \\ \mathbf{e}_{(N-1)N}^H \mathbf{x} \end{bmatrix} = \begin{bmatrix} \mathbf{e}_{0N}^H \\ \mathbf{e}_{1N}^H \\ \vdots \\ \mathbf{e}_{(N-1)N}^H \end{bmatrix} \mathbf{x}$$

- Define the DFT matrix $\mathbf{F}^H$ so that we can write $\mathbf{X} = \mathbf{F}^H \mathbf{x}$

$$\mathbf{F}^H = \begin{bmatrix} \mathbf{e}_{0N}^H \\ \mathbf{e}_{1N}^H \\ \vdots \\ \mathbf{e}_{(N-1)N}^H \end{bmatrix} = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & e^{-j2\pi(1)(1)/N} & \cdots & e^{-j2\pi(1)(N-1)/N} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-j2\pi(N-1)(1)/N} & \cdots & e^{-j2\pi(N-1)(N-1)/N} \end{bmatrix}$$

- The DFT of signal $x$ is a matrix multiplication $\Rightarrow \mathbf{X} = \mathbf{F}^H \mathbf{x}$

# The Hermitian of the DFT matrix $\mathbf{F}^H$

- Let $\mathbf{F} = \left(\mathbf{F}^H\right)^H$ be conjugate transpose of $\mathbf{F}^H$. We can write $\mathbf{F}$ as

$$\mathbf{F} = \begin{bmatrix} \mathbf{e}_{0N}^T \\ \mathbf{e}_{1N}^T \\ \vdots \\ \mathbf{e}_{(N-1)N}^T \end{bmatrix} \quad \Leftarrow \quad \mathbf{F}^H = \begin{bmatrix} \mathbf{e}_{0N}^* & \mathbf{e}_{1N}^* & \cdots & \mathbf{e}_{(N-1)N}^* \end{bmatrix}$$

- We say that $\mathbf{F}^H$ and $\mathbf{F}$ are Hermitians of each other (that's why $\mathbf{F}^H$)
- The $n$th row of $\mathbf{F}$ is the $n$th complex exponential $\mathbf{e}_{nN}^T$
- The $k$th column of $\mathbf{F}^H$ is the $k$th conjugate complex exponential $\mathbf{e}_{kN}^*$

# The product of $\mathbf{F}$ and its Hermitian $\mathbf{F}^H$

- The product between the DFT matrix $\mathbf{F}$ and its Hermitian $\mathbf{F}^H$ is

$$\begin{array}{c} \\ \begin{bmatrix} \mathbf{e}_{0N}^T \\ \vdots \\ \mathbf{e}_{kN}^T \\ \vdots \\ \mathbf{e}_{(N-1)N}^T \end{bmatrix} \end{array} \begin{array}{c} \begin{bmatrix} \mathbf{e}_{0N}^* & \cdots & \mathbf{e}_{kN}^* & \cdots & \mathbf{e}_{(N-1)N}^* \end{bmatrix} \\ \begin{bmatrix} \mathbf{e}_{0N}^T\mathbf{e}_{0N}^* & \cdots & \mathbf{e}_{0N}^T\mathbf{e}_{kN}^* & \cdots & \mathbf{e}_{0N}^T\mathbf{e}_{(N-1)N}^* \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{e}_{kN}^T\mathbf{e}_{0N}^* & \cdots & \mathbf{e}_{kN}^T\mathbf{e}_{kN}^* & \cdots & \mathbf{e}_{kN}^T\mathbf{e}_{(N-1)N}^* \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{e}_{(N-1)N}^T\mathbf{e}_{0N}^* & \cdots & \mathbf{e}_{(N-1)N}^T\mathbf{e}_{kN}^* & \cdots & \mathbf{e}_{(N-1)N}^T\mathbf{e}_{(N-1)N}^* \end{bmatrix} \end{array} = \mathbf{F}^H\mathbf{F}$$

- The $(n, k)$ element of product matrix is the inner product $\mathbf{e}_{nN}^T\mathbf{e}_{kN}^*$
- Orthonormality of complex exponentials $\Rightarrow \mathbf{e}_{nN}^T\mathbf{e}_{kN}^* = \delta(n-k)$

  $\Rightarrow$ Only the diagonal elements survive in the matrix product

# The matrix $\mathbf{F}$ and its inverse

- The DFT matrix $\mathbf{F}$ and its Hermitian are inverses of each other

$$\mathbf{F}^H\mathbf{F} = \begin{bmatrix} 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 1 \end{bmatrix} = \mathbf{I}$$

- Matrices whose inverse is its Hermitian, are said Hermitian matrices
- Have proved the following fundamental theorem. Orthonormality

Theorem

*The DFT matrix* $\mathbf{F}$ *is Hermitian* $\Rightarrow \mathbf{F}^H\mathbf{F} = \mathbf{I} = \mathbf{F}\mathbf{F}^H$

# The iDFT in matrix form

- We can retrace methodology to also write the iDFT in matrix form
- No new definitions are needed. Use vectors $\mathbf{e}_{nN}$ and $\mathbf{X}$ to write

$$\tilde{x}(n) = \mathbf{e}_{nN}^T \mathbf{X} = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X(k) e^{j2\pi kn/N}$$

- Define stacked vector $\tilde{\mathbf{x}}$ and stack definitions. Use expression for $\mathbf{F}$

$$\tilde{\mathbf{x}} = \begin{bmatrix} \tilde{x}(0) \\ \tilde{x}(1) \\ \vdots \\ \tilde{x}(N-1) \end{bmatrix} = \begin{bmatrix} \mathbf{e}_{0N}^T \mathbf{X} \\ \mathbf{e}_{1N}^T \mathbf{X} \\ \vdots \\ \mathbf{e}_{(N-1)N}^T \mathbf{X} \end{bmatrix} = \begin{bmatrix} \mathbf{e}_{0N}^T \\ \mathbf{e}_{1N}^T \\ \vdots \\ \mathbf{e}_{(N-1)N}^T \end{bmatrix} \mathbf{X} = \mathbf{F}\mathbf{X}$$

- The iDFT is, as the DFT, just a matrix product $\Rightarrow \tilde{\mathbf{x}} = \mathbf{F}\mathbf{X}$

# Inverse theorem, like a pro

- When we proved theorems we had monkey steps and one smart step

  $\Rightarrow$ That was orthonormality $\Rightarrow$ matrix $\mathbf{F}$ is Hermitian $\Rightarrow \mathbf{F}^H\mathbf{F} = \mathbf{I}$

**Theorem**

*The iDFT is, indeed, the inverse of the DFT*

*Proof.*

- Write $\tilde{\mathbf{x}} = \mathbf{F}\mathbf{X}$ and $\mathbf{X} = \mathbf{F}^H\mathbf{x}$ and exploit fact that $\mathbf{F}$ is Hermitian

$$\tilde{\mathbf{x}} = \mathbf{F}\mathbf{X} = \mathbf{F}\mathbf{F}^H\mathbf{x} = \mathbf{I}\mathbf{x} = \mathbf{x} \qquad \square$$

- Actually, this theorem would be true for any transform pair

$$\mathbf{X} = \mathbf{T}^H\mathbf{x} \qquad \Longleftrightarrow \qquad \tilde{\mathbf{x}} = \mathbf{T}\mathbf{X}$$

- As long as the transform matrix $\mathbf{T}$ is Hermitian $\Rightarrow \mathbf{T}^H\mathbf{T} = \mathbf{I}$

# Energy conservation (Parseval) theorem, like a pro

Theorem

*The DFT preserves energy* $\Rightarrow \|\mathbf{x}\|^2 = \mathbf{x}^H\mathbf{x} = \mathbf{X}^H\mathbf{X} = \|\mathbf{X}\|^2$

Proof.

- Use iDFT to write $\mathbf{x} = \mathbf{F}\mathbf{X}$ and exploit fact that $\mathbf{F}$ is Hermitian

$$\|\mathbf{x}\|^2 = \mathbf{x}^H\mathbf{x} = (\mathbf{F}\mathbf{X})^H\mathbf{F}\mathbf{X} = \mathbf{X}^H\mathbf{F}^H\mathbf{F}\mathbf{X} = \mathbf{X}^H\mathbf{X} = \|\mathbf{X}\|^2 \qquad \square$$

- This theorem would also be true for any transform pair

$$\mathbf{X} = \mathbf{T}^H\mathbf{x} \qquad \Longleftrightarrow \qquad \tilde{\mathbf{x}} = \mathbf{T}\mathbf{X}$$

- As long as the transform matrix $\mathbf{T}$ is Hermitian $\Rightarrow \mathbf{T}^H\mathbf{T} = \mathbf{I}$

# Designing transformations adapted to signals

- A basic information processing theory can be built for any $\mathbf{T}$
- Then, why do we specifically choose the DFT? Or the DCT?
  - $\Rightarrow$ Oscillations represent different rates of change
  - $\Rightarrow$ Different rates of change represent different aspects of a signal
- Not a panacea, though. E.g., $\mathbf{F}^H$ is independent of the signal
- If we know something about signal, should use it to build better $\mathbf{T}$
- A way of "knowing something" is a stochastic model of the signal
- PCA: Principal component analysis
  - $\Rightarrow$ Use the eigenvectors of the covariance matrix to build $\mathbf{T}$

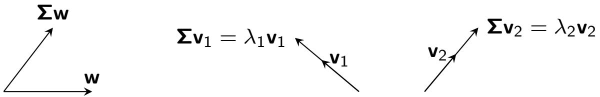
# Principal Component Analysis (PCA) transform

# Eigenvectors and eigenvalues of covariance matrix

- Consider a vector with $N$ elements $\Rightarrow \mathbf{v} = [v(0), v(1), \ldots, v(N-1)]$
- We say that $\mathbf{v}$ is an eigenvector of $\mathbf{\Sigma}$ if for some scalar $\lambda \in \mathbb{R}$

$$\mathbf{\Sigma v} = \lambda \mathbf{v}$$

- We say that $\lambda$ is the eigenvalue associated to $\mathbf{v}$




- In general, non-eigenvectors $\mathbf{w}$ and $\mathbf{\Sigma w}$ point in different directions
- But for eigenvectors $\mathbf{v}$, the product vector $\mathbf{\Sigma v}$ is collinear with $\mathbf{v}$
- We use normalized eigenvectors with unit energy $\Rightarrow \|\mathbf{v}\|^2 = 1$

# Ordering

Theorem

*The eigenvalues of* $\mathbf{\Sigma}$ *are real and nonnegative* $\Rightarrow \lambda \in \mathbb{R}$ *and* $\lambda \geq 0$

- Order eigenvalues from largest to smallest $\Rightarrow \lambda_0 \geq \lambda_1 \geq \ldots \geq \lambda_{N-1}$
- Eigenvectors inherit order $\Rightarrow \mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_{N-1}$
- The $n$th eigenvector of $\mathbf{\Sigma}$ is associated with its $n$th largest eigenvalue

# Eigenvectors are orthonormal

Theorem

*Eigenvectors of $\mathbf{\Sigma}$ associated with different eigenvalues are orthogonal*

- Define the matrix $\mathbf{T} = [\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_{N-1}]$
- Since the eigenvectors $\mathbf{v}_k$ are orthonormal, the product $\mathbf{T}^H\mathbf{T}$ is

$$\begin{bmatrix} \mathbf{v}_0 & \cdots & \mathbf{v}_k & \cdots & \mathbf{v}_{N-1} \end{bmatrix}$$

$$\mathbf{T}^H\mathbf{T} = \begin{bmatrix} \mathbf{v}_0^H \\ \vdots \\ \mathbf{v}_k^H \\ \vdots \\ \mathbf{v}_{N-1}^H \end{bmatrix} \begin{bmatrix} \mathbf{v}_0^H\mathbf{v}_0 & \cdots & \mathbf{v}_1^H\mathbf{v}_k & \cdots & \mathbf{v}_0^H\mathbf{v}_{N-1} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{v}_k^H\mathbf{v}_0 & \cdots & \mathbf{v}_k^H\mathbf{v}_k & \cdots & \mathbf{v}_k^H\mathbf{v}_{N-1} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{v}_{N-1}^H\mathbf{v}_{N-1} & \cdots & \mathbf{v}_{N-1}^H\mathbf{v}_k & \cdots & \mathbf{v}_{N-1}^H\mathbf{v}_{N-1} \end{bmatrix} = \begin{bmatrix} 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 1 \end{bmatrix}$$

- The eigenvector matrix **T is Hermitian** $\Rightarrow \mathbf{T}^H\mathbf{T} = \mathbf{I}$

# Principal component analysis transform

- Any Hermitian $\mathbf{T}$ can be used to define an info processing transform

- Define principal component analysis (PCA) transform $\Rightarrow \mathbf{y} = \mathbf{T}^H\mathbf{x}$
- And the inverse (i)PCA transform $\Rightarrow \tilde{\mathbf{x}} = \mathbf{T}\mathbf{y}$

- Since $\mathbf{T}$ is Hermitian, iPCA is, indeed, the inverse of the PCA

$$\tilde{\mathbf{x}} \;=\; \mathbf{T}\mathbf{y} \;=\; \mathbf{T}\left(\mathbf{T}^H\mathbf{x}\right) \;=\; \mathbf{T}\mathbf{T}^H\mathbf{x} \;=\; \mathbf{I}\mathbf{x} \;=\; \mathbf{x}$$

- Thus $\mathbf{y}$ is an equivalent representation of $\mathbf{x}$ $\Rightarrow$ Back and forth

- And, also because $\mathbf{T}$ is Hermitian, Parseval's theorem holds

$$\|\mathbf{x}\|^2 \;=\; \mathbf{x}^H\mathbf{x} \;=\; (\mathbf{T}\mathbf{y})^H\,\mathbf{T}\mathbf{y} \;=\; \mathbf{y}^H\mathbf{T}^H\mathbf{T}\mathbf{y} \;=\; \mathbf{y}^H\mathbf{y} \;=\; \|\mathbf{y}\|^2$$

- Modifying elements $y_k$ means altering energy composition of signal

# Compression with the PCA

- Transform signal $\mathbf{x}$ into eigenvector domain with PCA $\mathbf{y} = \mathbf{T}^H\mathbf{x}$
- Recover $\mathbf{x}$ from $\mathbf{y}$ through iPCA matrix multiplication $\mathbf{x} = \mathbf{T}\mathbf{y}$
- We compress by retaining $K < N$ PCA coefficients to write

$$\tilde{\mathbf{x}}(n) = \sum_{k=0}^{K-1} y(k)\mathbf{v}_k(n)$$

- Equivalently, we define the compressed PCA as

$$\tilde{\mathbf{y}}(k) = y(k) \quad \text{for} \quad k < K, \qquad \tilde{\mathbf{y}}(k) = 0 \text{ otherwise}$$

- Reconstructed signal is obtained with iPCA $\Rightarrow \tilde{\mathbf{x}} = \mathbf{T}\tilde{\mathbf{y}}$

# Expected reconstruction error

- PCA dimensionality reduction minimizes the expected error energy
- To see that this is true, define the error signal as $\Rightarrow \mathbf{e} := \mathbf{x} - \tilde{\mathbf{x}}$
- The energy of the error signal is $\Rightarrow \|\mathbf{e}\|^2 = \|\mathbf{x} - \tilde{\mathbf{x}}\|^2$
- The expected value of the energy of the error signal is

$$\mathbb{E}\left[\|\mathbf{e}\|^2\right] = \mathbb{E}\left[\|\mathbf{x} - \tilde{\mathbf{x}}\|^2\right]$$

- Keeping the first $K$ PCA coefficients minimizes $\mathbb{E}\left[\|\mathbf{e}\|^2\right]$

  $\Rightarrow$ Among all reconstructions that use, at most, $K$ coefficients

# Dimensionality reduction expected error

Theorem

*The expectation of the reconstruction error is the sum of the eigenvalues corresponding to the eigenvectors of the coefficients that are discarded*

$$\mathbb{E}\left[\|\mathbf{e}\|^2\right] = \sum_{k=K}^{N-1} \lambda_k$$

- It follows that keeping the first $K$ PCA coefficients is optimal
  - $\Rightarrow$ In the sense that it minimizes the Expected error energy
- Good on average. Across realizations of the stochastic signal $\mathbf{X}$
- Need not be good for given realization (but we expect it to be good)

# Proof of expected error expression

Proof.

- Error signal signal is $\mathbf{e} := \mathbf{x} - \tilde{\mathbf{x}}$. Define error PCA transform as $\mathbf{f} = \mathbf{T}^H\mathbf{x}$
- Using Parseval's (energy conservation) we can write the energy of $\mathbf{e}$ as

$$\|\mathbf{e}\|^2 = \|\mathbf{f}\|^2 = \sum_{k=K}^{N-1} y^2(k)$$

- In the last equality we used that $\mathbf{f} = \mathbf{y} - \tilde{\mathbf{y}} = [0, \ldots, 0, y(K), \ldots, y(N-1)]$
- Here, we are interested in the expected value of the error's energy
- Take expectation on both sides of equality $\Rightarrow \mathbb{E}\left[\|\mathbf{e}\|^2\right] = \sum_{k=K}^{N-1} \mathbb{E}\left[y^2(k)\right]$
- Used the fact that expectations are linear operators

# Proof of expected error expression

Proof.

- Compute expected value $\mathbb{E}\left[y^2(k)\right]$ of the squared PCA coefficient $y(k)$
- As per PCA transform definition $y(k) = \mathbf{v}^H\mathbf{x}$, which implies

$$\mathbb{E}\left[y^2(k)\right] = \mathbb{E}\left[(\mathbf{v}_k^H\mathbf{x})^2\right] = \mathbb{E}\left[\mathbf{v}_k^H\mathbf{x}\mathbf{x}^T\mathbf{v}_k\right] = \mathbf{v}_k^H\mathbb{E}\left[\mathbf{x}\mathbf{x}^T\right]\mathbf{v}_k$$

- Covariance matrix: $\boldsymbol{\Sigma} := \mathbb{E}\left[\mathbf{x}\mathbf{x}^T\right]$. Eigenvector definition $\boldsymbol{\Sigma}\mathbf{v}_k = \lambda_k$. Thus

$$\mathbb{E}\left[y^2(k)\right] = \mathbf{v}_k^H\boldsymbol{\Sigma}\mathbf{v}_k = \mathbf{v}_k^H\lambda_k\mathbf{v}_k = \lambda_k\|\mathbf{v}_k\|^2$$

- Substitute into expression for $\mathbb{E}\left[\|\mathbf{e}\|^2\right]$ to write $\Rightarrow \mathbb{E}\left[\|\mathbf{e}\|^2\right] = \sum_{k=K}^{N-1}\lambda_k$ □

# Discussions

- The PCA transform is defined for any signal (vector) $\mathbf{x}$
  - $\Rightarrow$ But we expect to work well only when $\mathbf{x}$ is a realization of $\mathbf{X}$
- Write the iPCA in expanded form and compare with the iDFT

$$x(n) = \sum_{k=0}^{N-1} y(k) v_k(n) \quad \Leftrightarrow \quad x(n) = \sum_{k=0}^{N-1} X(k) e_{kN}(n)$$

- The same except that they use different bases for the expansion
- Still, like developing a new sense.
- But not one that is generic. Rather, adapted to the random signal $\mathbf{X}$

# Graph Signals

# Graphs and graph signals

- A graph (network) is a triplet $(\mathcal{V}, \mathcal{E}, W)$. Vertices, edges, weights
- (In) Neighborhood $\Rightarrow \mathcal{N}(n) = \{m \in \mathcal{V} : (m, n) \in \mathcal{E}\}$
- $W : \mathcal{E} \to \mathbb{R}$ is a map from the set of edges to scalar values, $w_{nm}$
  - $\Rightarrow$ Represents the level of relationship from $n$ to $m$
  - $\Rightarrow$ Unweighted $\Rightarrow w_{nm} \in \{0, 1\}$. Undirected $\Rightarrow w_{nm} = w_{mn}$
  - $\Rightarrow$ Most often weights are strictly positive, $W : \mathcal{E} \to \mathbb{R}_{++}$
- Graph signals are mappings defined on vertices of graph $x : \mathcal{V} \to \mathbb{R}$
  - $\Rightarrow$ Vector $\mathbf{x} \in \mathbb{R}^N$ where $x_n$ represents signal value at the $n$th vertex

# Adjacency matrices

- Given a graph $G = (\mathcal{V}, \mathcal{E}, W)$ of $N$ vertices,
- Its adjacency matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$ is defined as

$$A_{nm} = \begin{cases} w_{nm}, & \text{if}(n,m) \in \mathcal{E} \\ 0, & \text{otherwise} \end{cases}$$

- A matrix representation incorporating all information about $G$

  $\Rightarrow$ For unweighted graphs, positive entries represent connected pairs

  $\Rightarrow$ For weighted graphs, also denote proximities between pairs

# Laplacian of a graph

- Given a graph $G$ with adjacency matrix $\mathbf{A}$ and degree matrix $\mathbf{D}$
- We define the Laplacian matrix $\mathbf{L} \in \mathbb{R}^{N \times N}$ as

$$\mathbf{L} = \mathbf{D} - \mathbf{A}$$

- Equivalently, $\mathbf{L}$ can be defined elementwise as

$$L_{ij} = \begin{cases} \deg(i) & \text{if } i = j \\ -w_{ij} & \text{if } (i, j) \in \mathcal{E} \\ 0 & \text{otherwise} \end{cases}$$

- We assume undirected $G \Rightarrow \deg(i)$ is well-defined

# Laplacian as a measure of variability

- Given a graph $\mathcal{G}$ with Laplacian $\mathbf{L}$ and a signal $\mathbf{x}$ define signal $\mathbf{y} = \mathbf{Lx}$

$$y_i = [\mathbf{Lx}]_i = \sum_{j \in \mathcal{N}(i)} w_{ij}(x_i - x_j)$$

- Summand $w_{ij}(x_i - x_j)$ large $\Rightarrow$ Weight $w_{ij}$ large. Values $x_i$ and $x_j$ different
- Signal component $y_i$ measures difference between $x_i$ and neighbor's values
- We can also define the Laplacian quadratic form of $\mathbf{x}$

$$\mathbf{x}^T\mathbf{Lx} = \frac{1}{2} \sum_{(i,j) \in \mathcal{E}} w_{ij}(x_i - x_j)^2$$

- Quantifies variation of signal $\mathbf{x}$ with respect to the graph's structure

# Graph Fourier Transform (GFT)

Two dimensional (2D) discrete Fourier transform (DFT)

Discrete Cosine Transform

The discrete Fourier transform with Hermitian matrices

Principal Component Analysis (PCA) transform

Graph Signals

Graph Fourier Transform (GFT)

Information sciences at ESE

# Graph-shift operator

- Given an arbitrary graph $G = (\mathcal{V}, \mathcal{E}, W)$
- A graph-shift operator $\mathbf{S} \in \mathbb{R}^{N \times N}$ of graph $G$ ia a matrix satisfying
  - $\Rightarrow S_{ij} = 0$ for $i \neq j$ and $(i, j) \notin \mathcal{E}$
- $\mathbf{S}$ can take nonzero values in the edges of $G$ or in its diagonal
- We have already seen some possible graph-shift operators
  - $\Rightarrow$ Adjacency $\mathbf{A}$, Degree $\mathbf{D}$ and Laplacian $\mathbf{L}$ matrices
- We restrict our attention to normal shifts $\mathbf{S} = \mathbf{V} \Lambda \mathbf{V}^H$
  - $\Rightarrow$ Columns of $\mathbf{V} = [\mathbf{v}_1 \mathbf{v}_2 \ldots \mathbf{v}_N]$ correspond to the eigenvectors of $\mathbf{S}$
  - $\Rightarrow$ $\Lambda$ is a diagonal matrix containing the eigenvalues of $\mathbf{S}$

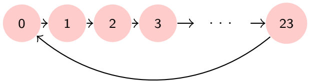
# Graph Fourier Transform (GFT)

- Given a graph $G$ and a graph signal $\mathbf{x} \in \mathbb{R}^N$ defined on $G$
  - $\Rightarrow$ Consider a normal graph-shift $\mathbf{S} = \mathbf{V}\Lambda\mathbf{V}^H$
- The Graph Fourier Transform (GFT) of $\mathbf{x}$ is defined as

$$\tilde{\mathbf{x}}(k) = \langle \mathbf{x}, \mathbf{v}_k \rangle = \sum_{n=1}^{N} \mathbf{x}(n)\mathbf{v}_k^*(n)$$

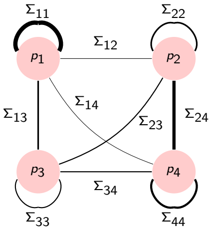
- In matrix form, $\tilde{\mathbf{x}} = \mathbf{V}^H\mathbf{x}$
- Given that the columns of $\mathbf{V}$ are the eigenvectors $\mathbf{v}_i$ of $\mathbf{S}$
  - $\Rightarrow$ $\tilde{\mathbf{x}}(k) = \mathbf{v}_k^H\mathbf{x}$ is the inner product between $\mathbf{v}_k$ and $\mathbf{x}$
  - $\Rightarrow$ $\tilde{\mathbf{x}}(k)$ is how similar $\mathbf{x}$ is to $\mathbf{v}_k$
  - $\Rightarrow$ In particular, GFT $\equiv$ DFT when $\mathbf{V}^H = \mathbf{F}$, i.e. $\mathbf{v}_k = \mathbf{e}_{kN}$

# DFT and PCA as particular cases of GFT




- For the directed cycle graph, GFT $\equiv$ DFT
  - $\Rightarrow$ if $\mathbf{S} = \mathbf{A}$ or
  - $\Rightarrow$ if $\mathbf{S} = \mathbf{L}$ for symmetrized graph
  - $\Rightarrow$ then $\mathbf{V}^H = \mathbf{F}$

- For the covariance graph, GFT $\equiv$ PCA
  - $\Rightarrow$ if $\mathbf{S} = \mathbf{A}$, then $\mathbf{V}^H = \mathbf{P}^H$



# Inverse graph Fourier transform

- Recall the graph Fourier transform $\mathbf{x}$
  - $\Rightarrow$ of any signal $\mathbf{x} \in \mathbb{R}^N$ on the vertices of graph $G$
  - $\Rightarrow$ is the expansion of $\mathbf{x}$ of the eigenvectors of the Laplacian

$$\tilde{\mathbf{x}}(k) = \langle \mathbf{x}, \mathbf{v}_k \rangle = \sum_{n=1}^{N} x(n) v_k^*(n)$$

- In matrix form, $\tilde{\mathbf{x}} = \mathbf{V}^H \mathbf{x}$
- The inverse graph Fourier transform is

$$\mathbf{x}(n) = \sum_{k=0}^{N-1} \tilde{\mathbf{x}}(k) v_k(n)$$

- In matrix form, $\mathbf{x} = \mathbf{V}\tilde{\mathbf{x}}$

# Inverse theorem, like a pro

- Recap in proving theorems we have monkey steps and one smart step
  $\Rightarrow$ That was orthonormality $\Rightarrow$ $\mathbf{V}^H$ is Hermitian $\Rightarrow$ $\mathbf{V}\mathbf{V}^H = \mathbf{I}$

Theorem

*The inverse graph Fourier transform (iGFT) is, indeed, the inverse of the GFT.*

Proof.

- Write $\mathbf{x} = \mathbf{V}\tilde{\mathbf{x}}$ and $\tilde{\mathbf{x}} = \mathbf{V}^H\mathbf{x}$ and exploit fact that $\mathbf{V}$ is Hermitian

$$\mathbf{x} \;=\; \mathbf{V}\tilde{\mathbf{x}} \;=\; \mathbf{V}\mathbf{V}^H\mathbf{x} \;=\; \mathbf{I}\mathbf{x} \;=\; \mathbf{x} \qquad \square$$

- This is the last inverse theorem we will see...

# Energy conservation (Parseval) theorem, like a pro

**Theorem**

*The GFT preserves energy* $\Rightarrow \|\mathbf{x}\|^2 = \mathbf{x}^H\mathbf{x} = \tilde{\mathbf{x}}^H\tilde{\mathbf{x}} = \|\tilde{\mathbf{x}}\|^2$

**Proof.**

- Use GFT to write $\tilde{\mathbf{x}} = \mathbf{V}^H\mathbf{x}$ and the fact that $\mathbf{V}$ is Hermitian

$$\|\tilde{\mathbf{x}}\|^2 = \tilde{\mathbf{x}}^H\tilde{\mathbf{x}} = \left(\mathbf{V}^H\mathbf{x}\right)^H\mathbf{V}^H\mathbf{x} = \mathbf{x}^H\mathbf{V}\mathbf{V}^H\mathbf{x} = \mathbf{x}^H\mathbf{x} = \|\mathbf{x}\|^2 \quad \square$$

# Courses to consider

- If you want to explore more about transforms and filters
  - ⇒ ESE210: Introduction to Dynamic Systems
  - ⇒ ESE303: Stochastic Systems Analysis and Simulation
  - ⇒ ESE325: Fourier Analysis and Applications ...
  - ⇒ ESE531: Digital Signal Processing

# More courses to consider

- Once you have information you may want to something with it
- Controlling the state of a system
  - ⇒ ESE406: Control of Systems
  - ⇒ ESE500: Linear Systems Theory
- Making decisions that are good in some sense (optimal)
  - ⇒ ESE204: Decision Models
  - ⇒ ESE304: Optimization of Systems
  - ⇒ ESE504: Introduction to Optimization Theory
  - ⇒ ESE605: Modern Convex Optimization

# Even more courses to consider

- At some point, you want to use what you've learned to do something
  - $\Rightarrow$ ESE290: Introduction to ESE Research Methodology
  - $\Rightarrow$ ESE350: Embedded Systems/Microcontroller Laboratory

# Research

- Most professors use about 5% of their time on teaching
- The other 95% of their time they use on research
- It is a pity to come to Penn and not spend a summer doing research

- Most of us are happy to have help
- Even if we are not, our doctoral students are desperate for help

# Thanks

- It has been my pleasure.
- If you need my help at some point in the next 29 years, let me know
- I will be retired after that