Adaptive SP & Machine Intelligence

Lecture 4: Subspace Methods for Spectral Estimation

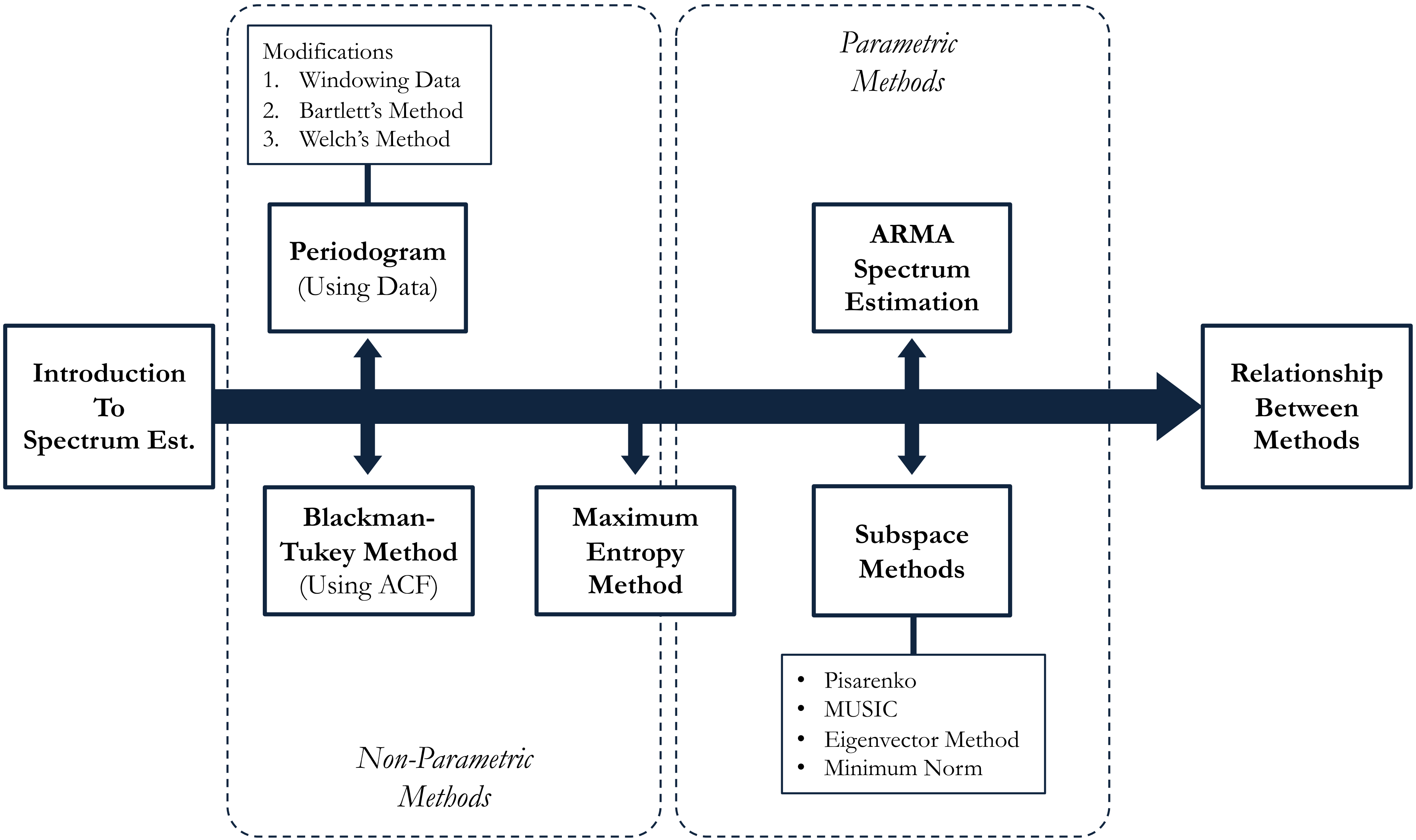
Danilo P. Mandic

Room: 813 Ext: 46271

Department of Electrical and Electronic Engineering
Imperial College London, UK

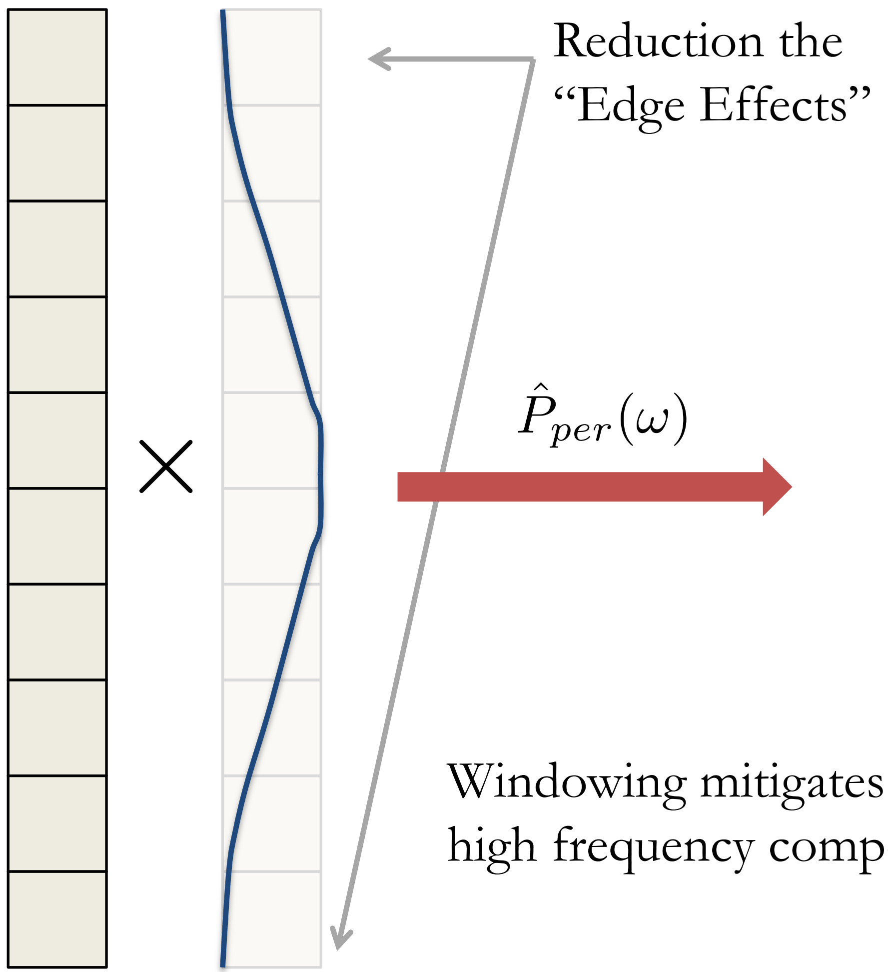
d.mandic@imperial.ac.uk URL: www.commsp.ee.ic.ac.uk/~mandic

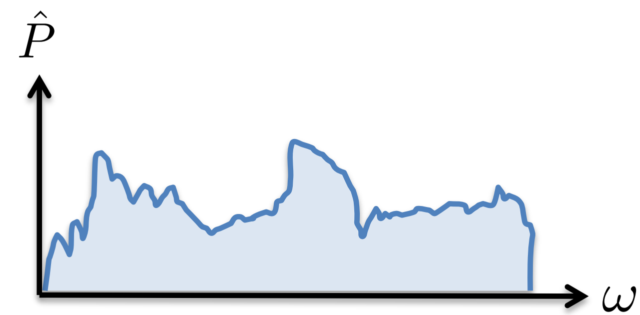
Overview of Spectral Estimation Methods



Modified Periodogram

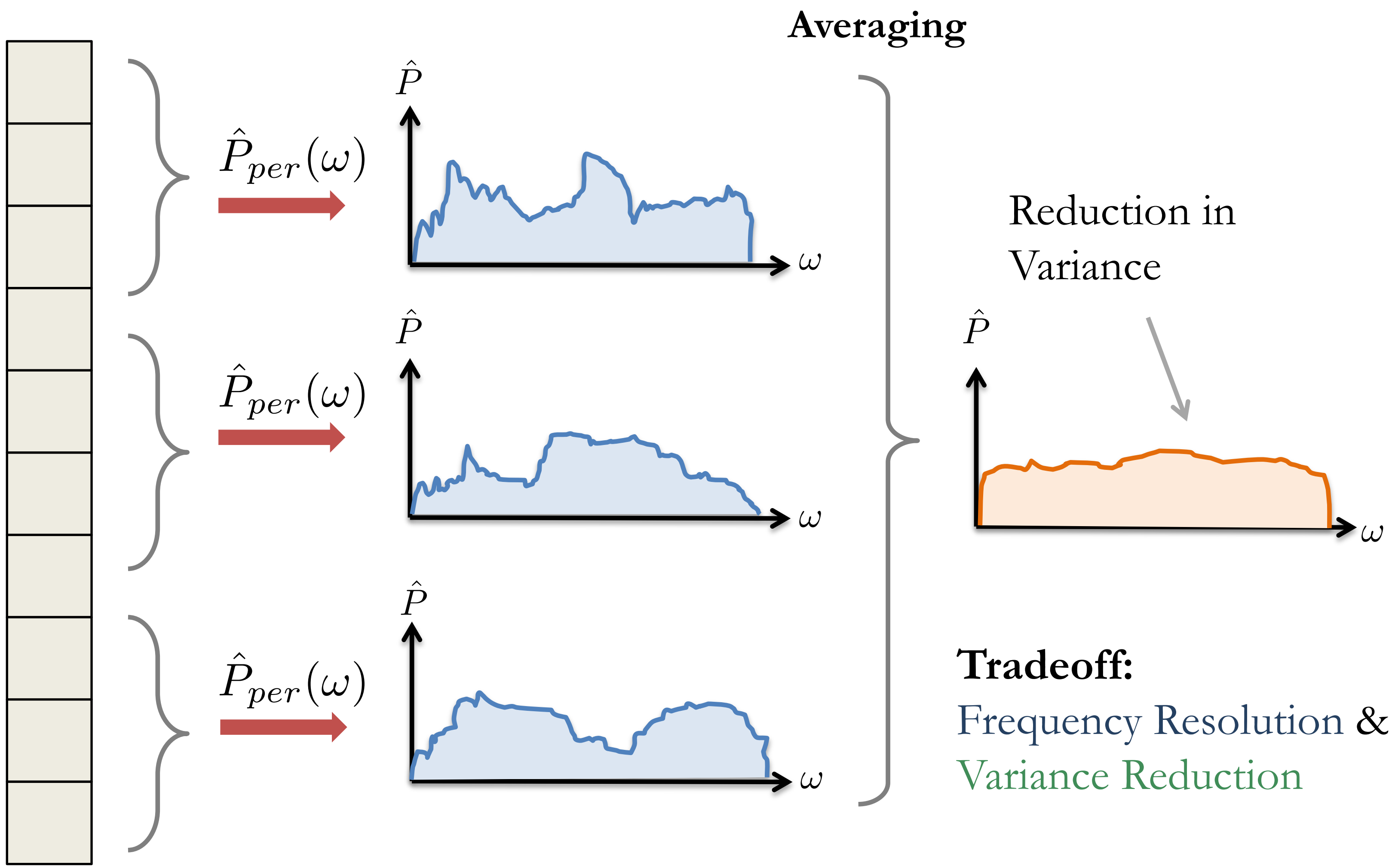
Windowing





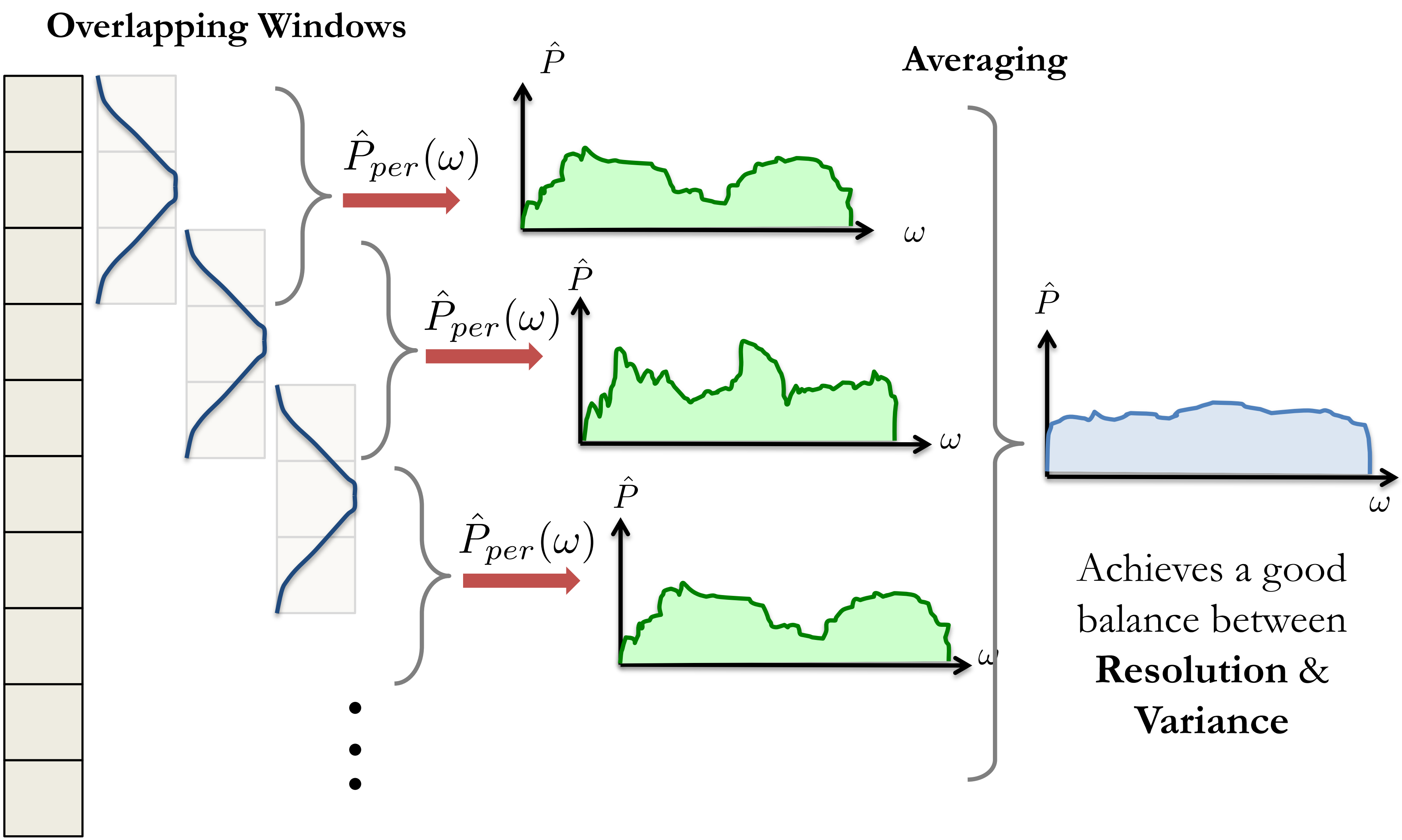
Windowing mitigates the problem of spurious high frequency components in the spectrum.

Bartlett's Method



Tradeoff:
Frequency Resolution &
Variance Reduction

Welch's Method



Achieves a good balance between **Resolution** & **Variance**

Periodogram Based Methods

Periodogram

$$\hat{P}_{per}(\omega_m) = \frac{1}{N}\left|\sum_{k=0}^{N-1} x[k]e^{-\jmath\omega_m k}\right|^2$$

Windowing
Modified Periodogram

$$\hat{P}_{mod}(\omega_m) = \frac{1}{NU}\left|\sum_{k=0}^{N-1} w[k]x[k]e^{-\jmath\omega_m k}\right|^2$$

Averaging
Bartlett's Method

$$\hat{P}_{B}(\omega_m) = \frac{1}{N}\sum_{i=0}^{K-1}\left|\sum_{k=0}^{N-1} x[k+iL]e^{-\jmath\omega_m k}\right|^2$$

+ Overlapping windows
Welch's Method

$$\hat{P}_{W}(\omega_m) = \frac{1}{KLU}\sum_{i=0}^{K-1}\left|\sum_{k=0}^{N-1} w[k]x[k+iD]e^{-\jmath\omega_m k}\right|^2$$

Blackman-Tukey Method

The Periodogram can also be expressed as:

$$\hat{P}_{per}(\omega_m) = \sum_{k=-N+1}^{N-1} \hat{\mathbf{r}}_{xx}[k] e^{-\jmath \omega_m k}$$

Autocorrelation Estimates at large lags are **unreliable**

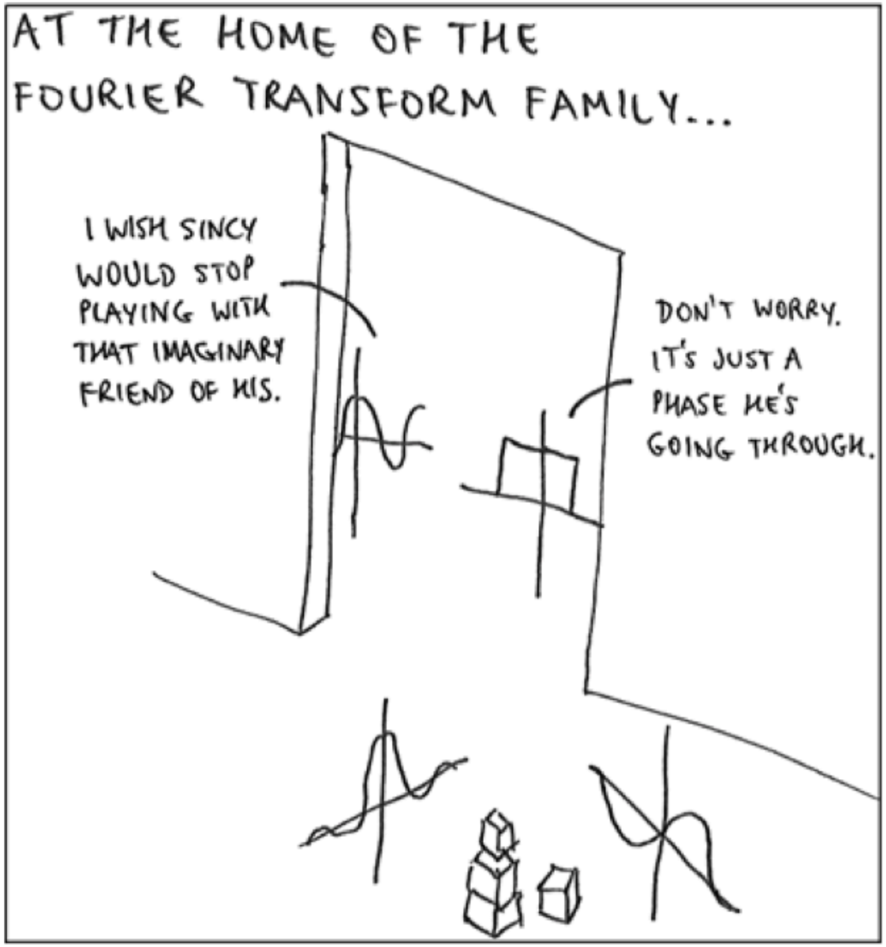
$$\hat{P}_{BT}(\omega_m) = \sum_{k=-M}^{M} w[k] \hat{\mathbf{r}}_{xx}[k] e^{-\jmath \omega_m k}$$

Lags: $M < N - 1$

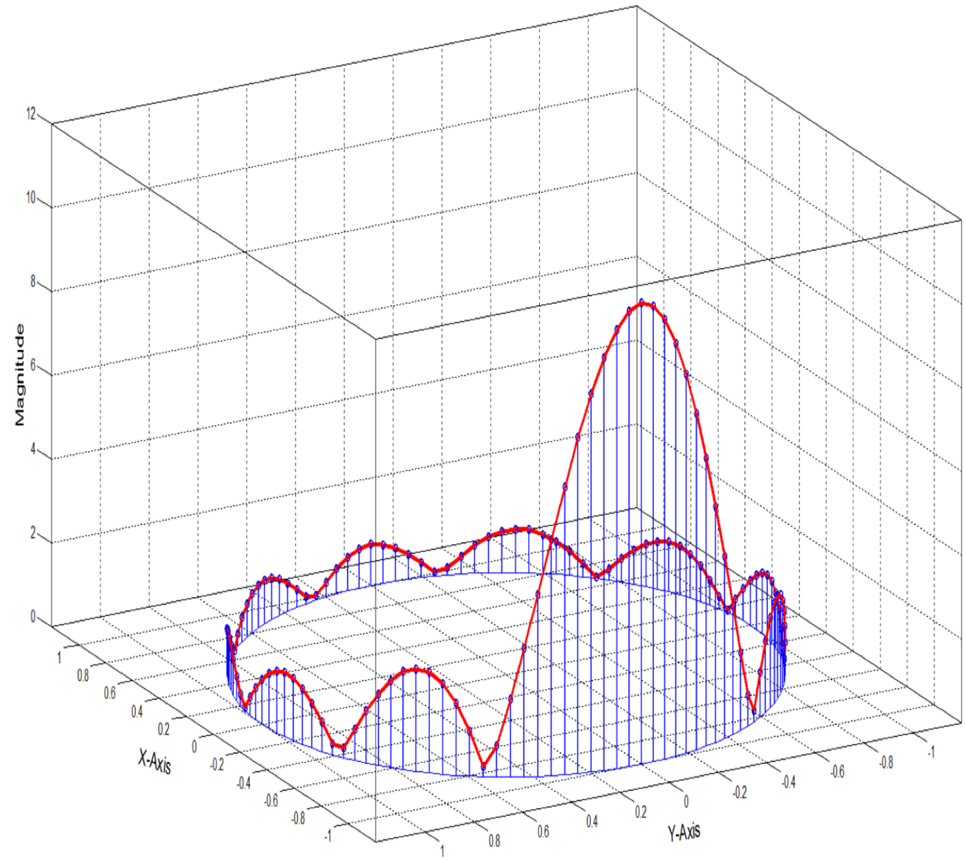
Windowing

Next: Can we **extrapolate the autocorrelation** estimates for lags $k > M$?

But, the main problem remains the same …

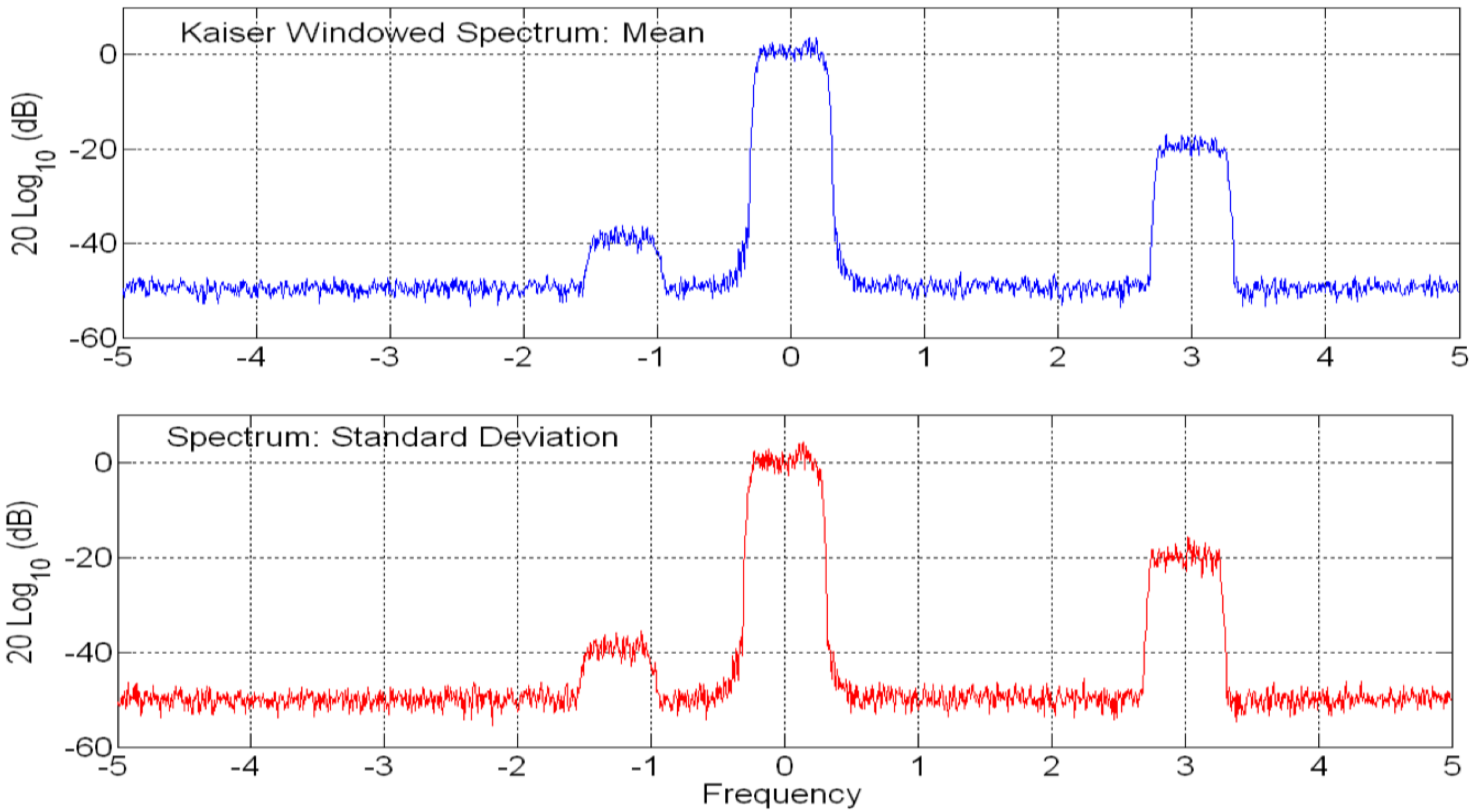


(credit, Fred Harris)



Sinc on a unit circle

Fourier methods yield inconsistent estimates



Spectral estimation as a linear estimation model

Problem: Fitting data $x[n]$ with a linear model with $[N-1]$ complex sinusoids:

$$\hat{x}[n] = \frac{1}{N}\sum_{k=0}^{N-1} w[k]e^{\jmath\frac{2\pi}{N}nk} \qquad (1)$$

Eq (1) can be formulated in vector notation as $\hat{\mathbf{x}} = \frac{1}{N}\mathbf{F}\mathbf{w}$, where

$$\begin{bmatrix} \hat{x}[0] \\ \hat{x}[1] \\ \hat{x}[2] \\ \hat{x}[3] \\ \vdots \\ \hat{x}[N-1] \end{bmatrix} = \frac{1}{N} \underbrace{\begin{bmatrix} 1 & 1 & 1 & 1 & \cdots & 1 \\ 1 & \alpha & \alpha^2 & \alpha^3 & \cdots & \alpha^{N-1} \\ 1 & \alpha^2 & \alpha^4 & \alpha^6 & \cdots & \alpha^{2(N-1)} \\ 1 & \alpha^3 & \alpha^6 & \alpha^9 & \cdots & \alpha^{3(N-1)} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha^{N-1} & \alpha^{2(N-1)} & \alpha^{3(N-1)} & \cdots & \alpha^{(N-1)(N-1)} \end{bmatrix}}_{\mathbf{F}} \begin{bmatrix} w[0] \\ w[1] \\ w[2] \\ w[3] \\ \vdots \\ w[N-1] \end{bmatrix}$$

where $\alpha = e^{\jmath\omega} = e^{\jmath\frac{2\pi}{N}}$.
Each column of $\mathbf{F}$ represents a sinusoid with a different frequency.

Spectral estimation as a Least Squares problem

The least squares solution to $\mathbf{w}$ is found by (CW question):

$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\text{argmin}} \, \|\mathbf{x} - \mathbf{F}\mathbf{w}\|^2 = \mathbf{F}^H \mathbf{x}$$

$\Longrightarrow$ DFT coefficient at bin k is $w[k] = \sum_{n=0}^{N-1} x[n] e^{-\jmath \frac{2\pi}{N} nk}$

What are the properties of the Fourier matrix?

- Is it unitary? $(\mathbf{F}^H \mathbf{F} \overset{?}{=} \mathbf{I})$
- Is it Hermitian? $(\mathbf{F}^H \overset{?}{=} \mathbf{F})$

$\rightarrow$ **Can you prove these properties?**

What happens if your signal $\mathbf{x}$ cannot be represented as a sum of the uniformly spaced sinusoids?

Example: What if $\mathbf{x} = \begin{bmatrix} 1 & \alpha^{\frac{1}{2}} & \alpha^{2\frac{1}{2}} & \dots & \alpha^{(N-1)\frac{1}{2}} \end{bmatrix}^T$?

Incoherent sampling $\Longrightarrow$ **A limitation of the DFT for a small N.**

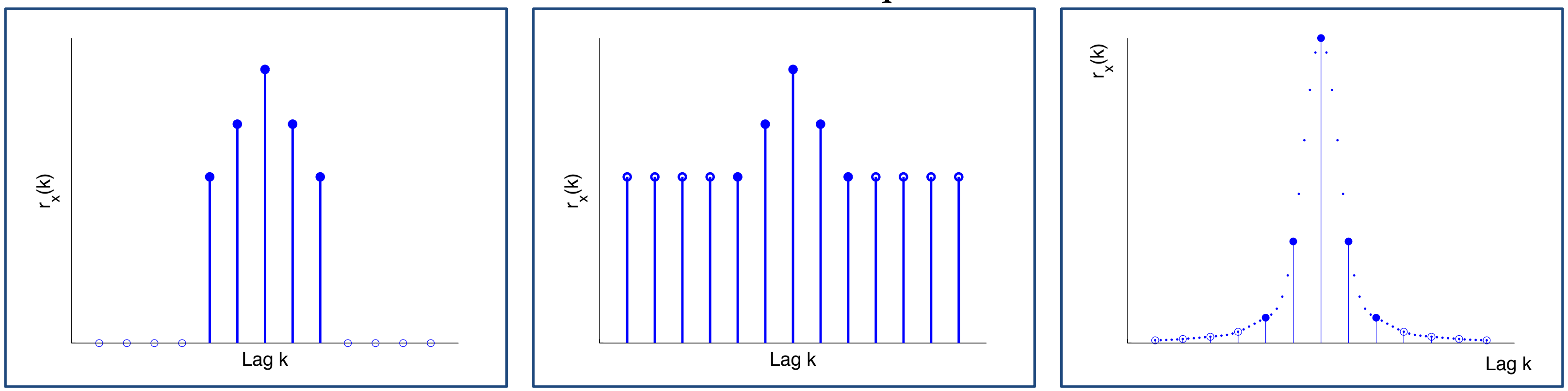
The Maximum Entropy Method: No DFT Legacy

How can we **extrapolate the autocorrelation** estimates with imposing the **least amount of structure on the data**?

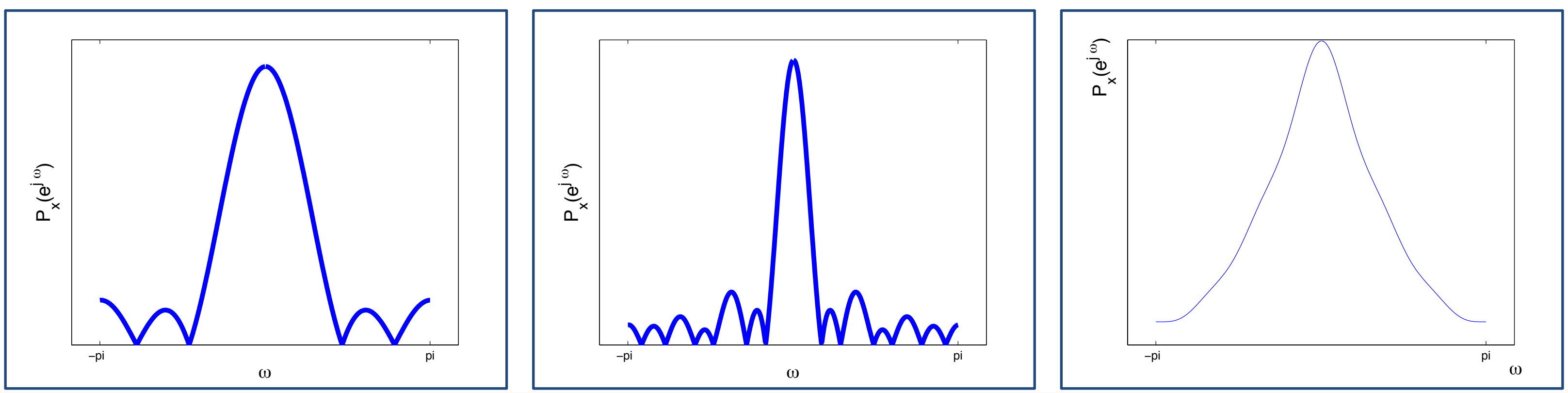
$\Longrightarrow$ Maximize the randomness $\Longrightarrow$ **Maximize Entropy**

Which one has the "flattest" PSD?

Autocorrelation Sequences



Power Spectral Density (PSD)



Maximum Entropy Method (MEM)

Entropy of Gaussian random process $x(n)$ with PSD $P_{xx}(\omega)$:

$$H(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln P_{xx}(\omega)\, d\omega$$

Goal: Find extrapolated autocorrelation values $r_e(k)$ to maximize the entropy:

$$\frac{\partial H(x)}{\partial r_e^*(k)} = 0, \quad \text{for: } |k| > p$$

**Refer to handout for the full derivation*

$$\boxed{\hat{P}_{mem}(\omega) = \frac{\sigma_w^2}{\left|1 + \sum_{k=1}^{p} \hat{a}_k e^{-\jmath k\omega}\right|^2}}$$

Estimated using the Yule-Walker Method

The MEM method is **identical to the all-pole AR(p) spectrum** although **no assumptions were made** about the model of the data (except Gaussianity).

MEW, derivation

$\Rightarrow$ for a Gaussian process with a given autocorr. sequence $r_x(k)$ for $|k| \leq p$ the Maximum Entropy Power Spectrum minimises entropy $H(x)$

subject to the constraint that the inverse DFT of $P_{xx}(\omega)$ equals the given set of autocorrelations for $|k| \leq p$, that is

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} P_{xx}(\omega)e^{\jmath k\omega}d\omega = r_x(k) \qquad |k| \leq p$$

The values of $r_e(k)$ that maximize the entropy may be found by setting the derivative of $H(x)$ wrt $r_e^*(k)$ equal to zero:

$$\frac{\partial H(x)}{\partial r_e^*(k)} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1}{P_{xx}(\omega)}\frac{\partial P_{xx}(\omega)}{\partial r_e^*}d\omega = 0 \qquad |k| > p$$

Notice that $\frac{\partial P_{xx}(\omega)}{\partial r_e^*} = e^{\jmath k\omega} \quad \Rightarrow \quad \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1}{P_{xx}(\omega)}e^{\jmath k\omega}d\omega = 0, \quad |k| > p.$

MEM, spectrum

Therefore:

$$Q_{xx}(\omega) = \frac{1}{P_{xx}(\omega)} = \sum_{k=-p}^{p} q_{xx}(k) e^{-\jmath k \omega}$$

$\Rightarrow \hat{P}_{mem}$ is an all–pole spectrum, given by

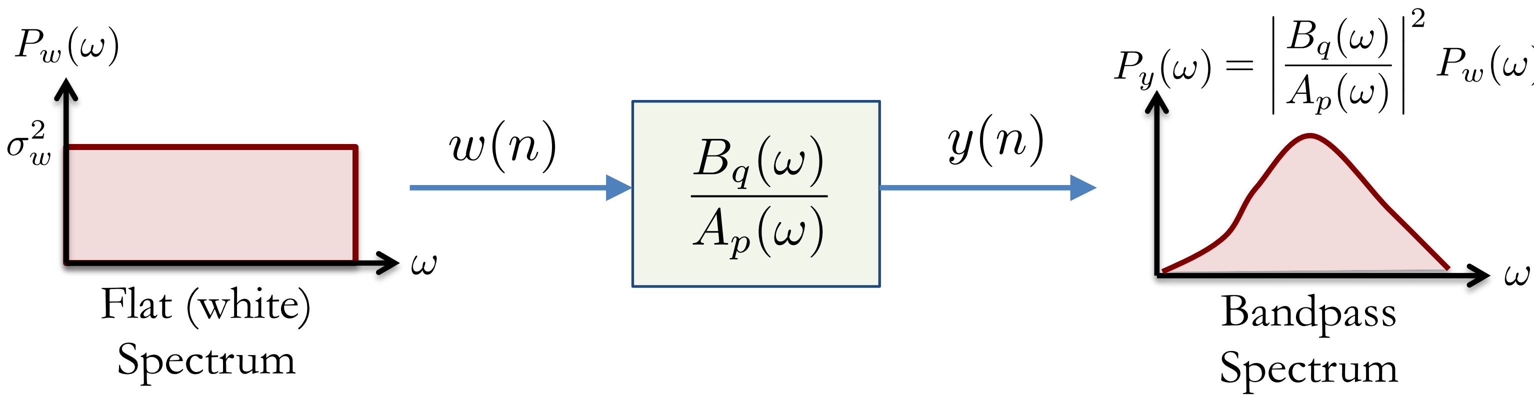
$$\hat{P}_{mem}(\omega) = \frac{|b(0)|^2}{A_p(\omega) A_p^*(\omega)} = \frac{|b(0)|^2}{|1 + \sum_{k=1}^{p} a_p(k) e^{-\jmath k \omega}|^2}$$

Alternatively

$$\hat{P}_{mem}(\omega) = \frac{|b(0)|^2}{|\mathbf{e}^H \mathbf{a}_p|^2}$$

Coefficients $\mathbf{a}[1, a_p(1), \ldots, a_p(p)]^T$ and $b(0)$ are found from the normal equations (Yule–Walker).

MEM power is the same as ARMA power!

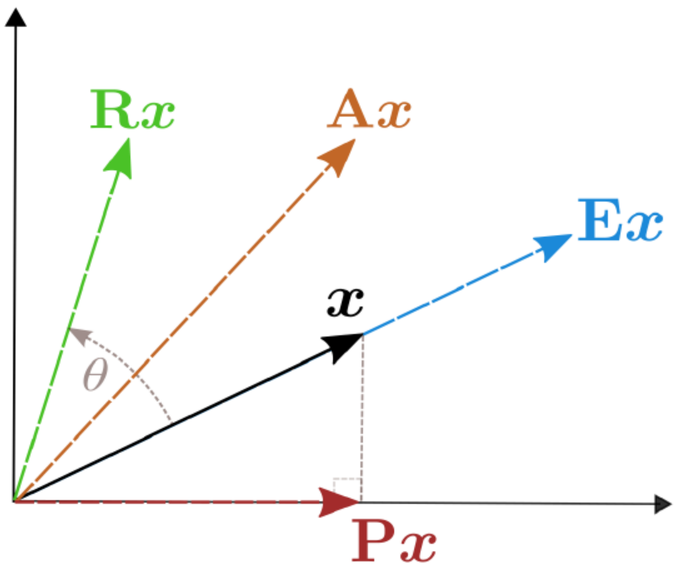


$$y(n) = -\underbrace{\sum_{k=1}^{p} a_k y(n-k)}_{\substack{\textbf{Autoregressive} \\ \text{AR(p)}}} + \underbrace{\sum_{k=0}^{q} b_k w(n-k)}_{\substack{\textbf{Moving Average} \\ \text{MA(q)}}}$$

$$\boxed{\hat{P}_{ARMA}(\omega) = \frac{\left|\sum_{k=0}^{q} \hat{b}_k e^{-\jmath k\omega}\right|^2}{\left|1 + \sum_{k=1}^{p} \hat{a}_k e^{-\jmath k\omega}\right|^2}}$$

Subspace Methods: Introduction

What is that a matrix does to a vector?



Ampli-twist

A matrix $\mathbf{A}$ which multiplies a vector $\mathbf{x}$
(i) stretches or shortens the vector
(ii) rotates the vector

A $\leadsto$ any general matrix

R $\leadsto$ a rotation matrix ($\mathbf{R}^T = \mathbf{R}^{-1}$ and $\det \mathbf{R} = 1$)

$\mathbf{E}\mathbf{x} = \lambda \mathbf{x}$ $\leadsto$ eigenanalysis

P $\leadsto$ projection matrix

An example of a rotation matrix

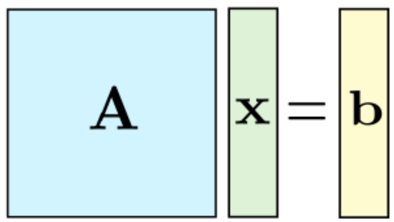
$$\mathbf{R} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

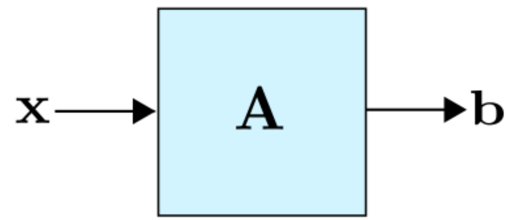
What can we say about the properties of the matrix $\mathbf{A}$, matrix $\mathbf{E}$ and the projection matrix $\mathbf{P}$ (rank, invertibility, ...)?

Is the projection matrix invertible?

The meaning of eigenanalysis

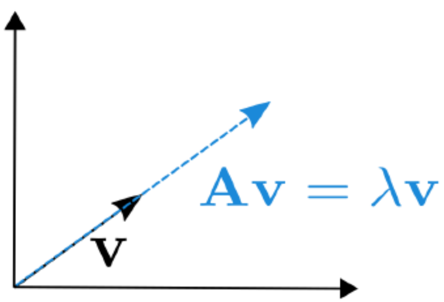
Let $\mathbf{A}$ be an $n \times n$ matrix, where $\mathbf{A}$ is a linear operator on vectors in $\mathbb{R}^n$, such that $\mathbf{A}\,\mathbf{x} = \mathbf{b}$



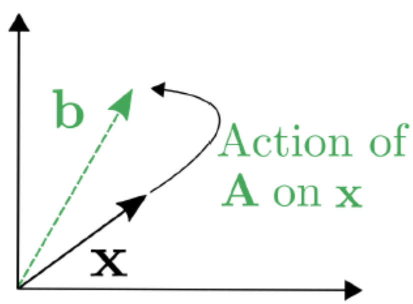


An **eigenvector** of $\mathbf{A}$ is a vector $\mathbf{v} \in \mathbb{R}^n$ such that $\mathbf{A}\,\mathbf{v} = \lambda\mathbf{v}$, where λ is called the corresponding eigenvalue.

Matrix $\mathbf{A}$ only changes the length of $\mathbf{v}$, not its direction!



Equation $\mathbf{A}\,\mathbf{v} = \lambda\mathbf{v}$



Equation $\mathbf{A}\,\mathbf{x} = \mathbf{b}$.

Eigenvalues

For an $n \times n$ matrix $\mathbf{A}$, its **eigenvalues** are found from the n-th order polynomial in λ defined by

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x} \quad \Rightarrow (\mathbf{A} - \lambda\mathbf{I})\,\mathbf{x} = \mathbf{0} \quad \Rightarrow \quad \text{nontrivial solution} \;\; det\,(\mathbf{A} - \lambda\mathbf{I}) = 0$$

where $\mathbf{I}$ is the $n \times n$ identity matrix and $\lambda_1, \ldots, \lambda_n$ the eigenvalues.

The corresponding n **eigenvectors**, $\mathbf{v}_1, \ldots, \mathbf{v}_n$, satisfy $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ and are generally normalised to have unit norm $\| \mathbf{v} \|_2 = 1$.

For distinct eigenvalues, these eigenvectors are **linearly independent**.

A symmetric matrix is positive definite iff all its eigenvalues are positive

The **Spectral Theorem** allows for a symmetric matrix to be written as

$$\mathbf{A} = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i^T$$

and the

$$\text{Trace}(\mathbf{A}) = \sum_{i=1}^{n} \lambda_i$$

Any connection with signal power?

More about eigenvalues

Let $\mathbf{A} = \begin{bmatrix} 2 & -4 \\ -1 & -1 \end{bmatrix}$. The characteristic polynomial $det(\mathbf{A} - \lambda\mathbf{I})$ is

$$\begin{aligned} p(\lambda) &= det\left(\begin{bmatrix} 2-\lambda & -4 \\ -1 & -1-\lambda \end{bmatrix}\right) \\ &= (2-\lambda)(-1-\lambda) - (-4)(-1) = \lambda^2 - \lambda - 6 = (\lambda-3)(\lambda+2) \end{aligned}$$

Thus the **eigenvalues** of $\mathbf{A}$ are $\lambda_1 = 3$ and $\lambda_2 = -2$.

To find **eigenvectors** $\mathbf{v} = [v_1, \ldots, v_n]^T$ corresponding to an eigenvalue λ

$$solve \qquad (\mathbf{A} - \lambda\mathbf{I})\,\mathbf{v} = 0 \qquad for\;\; \mathbf{v}$$

For $\lambda_1 = 3$ we thus have

$$\begin{bmatrix} 2-3 & -4 \\ -1 & -1-3 \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Rightarrow \quad \mathbf{v}_1 = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} -4 \\ 1 \end{bmatrix}$$

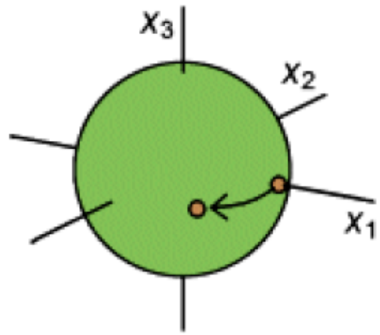
Similarly, for $\lambda_2 = -2$ we have $\mathbf{v}_2 = [1,\ 1]^T$, and thus $\mathbf{v}_1 \perp \mathbf{v}_2$.

$\mathbf{v}_1$ and $\mathbf{v}_2$ are **bases of the eigenspace spanned by these vectors**

Eigenanalysis, an intuitive example

A sphere of unit radius is positioned at the centre of a three–dimensional coordinate system. It is rotating about the x_3 axis. This rotation[1] can be described by the matrix

$$\mathbf{C} = \begin{bmatrix} 0.707 & 0.707 & 0 \\ -0.707 & 0.707 & \\ 0 & 0 & 1 \end{bmatrix}$$



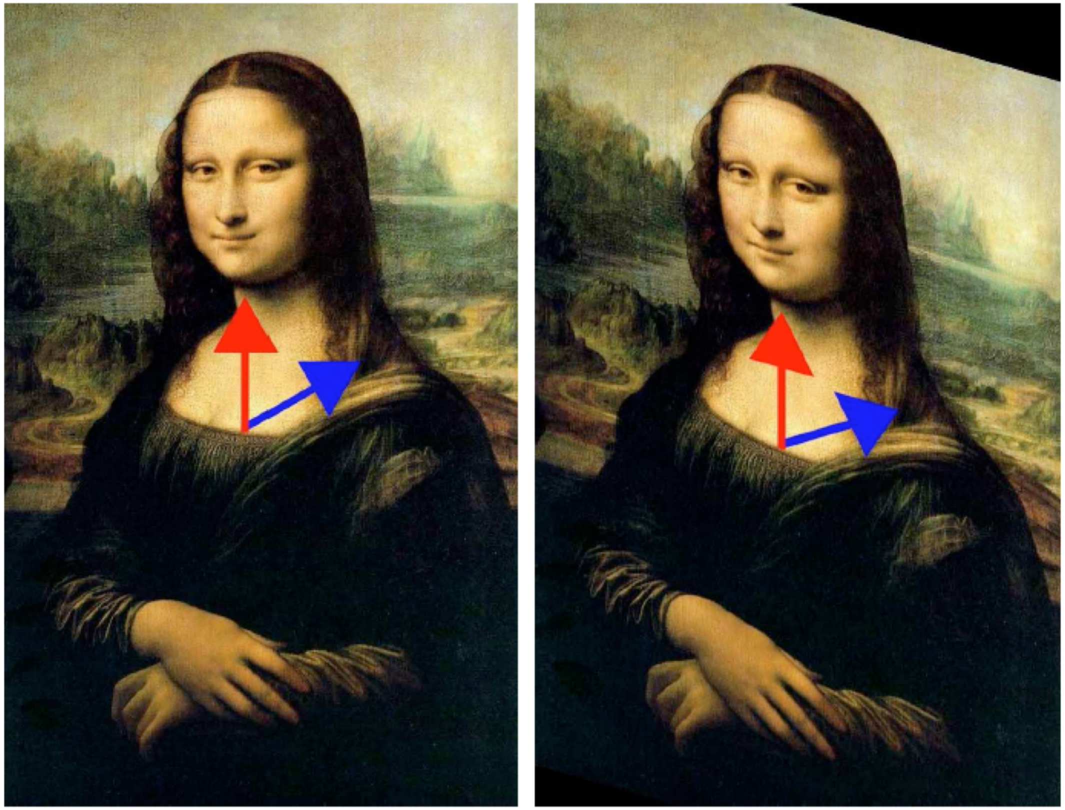
Intuitively, what is the eigenvector of $\mathbf{C}$? **Is there a point on the unit sphere that a 45^o rotation transforms into a multiple of itself?**

This is the north pole $[0, 0, 1] \Rightarrow$ an eigenvector of $\mathbf{C}$ is the vector $[0, 0, 1]$.

[1] By 45^o in this case. For example, multiplying $\mathbf{C}$ by the vector $[1, 0, 0]$ yields the vector $[0.707, 0.707.0]$, which is rotated 45^o.

Eigenanalysis, image repreentation

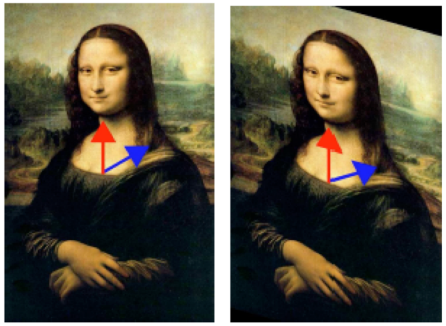
Eigenvector of a transformation $\Rightarrow$ a vector which, in the transformation, is multiplied by a constant factor, called the **eigenvalue of that vector**



The red vector is an eigenvector of the transformation, and the blue is not

Mona Lisa, finding eigenvectors

The red vector was neither stretched nor compressed $\Rightarrow$ its eigenvalue is 1



$$\text{Transformation} \qquad A = \begin{bmatrix} 1 & 0 \\ -\frac{1}{2} & 1 \end{bmatrix}$$

Eigenvectors:

$$\mathbf{Ax} = \lambda \mathbf{x} \quad \Rightarrow \quad (\mathbf{A} - \lambda \mathbf{I})\, \mathbf{x} = \mathbf{0}$$

For non–trivial solutions $\rightarrow det(\mathbf{A} - \lambda \mathbf{I}) = 0 \quad \Rightarrow$

$$det\left(\begin{bmatrix} 1 & 0 \\ -\frac{1}{2} & 1 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) = det \begin{bmatrix} 1-\lambda & 0 \\ -\frac{1}{2} & 1-\lambda \end{bmatrix} = 0 \Rightarrow \lambda = 1$$

Finally, eigenvectors for Mona Lisa

We have found $\lambda = 1$, the eigenvalue of matrix $\mathbf{A}$.

We can now solve for eigenvectors

$$\begin{bmatrix} 1-\lambda & 0 \\ -\frac{1}{2} & 1-\lambda \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \mathbf{0}$$

Substituting $\lambda = 1$ we have

$$\mathbf{v} = \begin{bmatrix} 0 \\ c \end{bmatrix}$$

where c is an arbitrary constant.

All vectors of this form, pointing straight up or down, are eigenvectors of $\mathbf{A}$

In general $\mathbf{A}$ will have two distinct eigenvalues, and thus two distinct eigenvectors.

Most vectors will have both their lengths and direction changed by $\mathbf{A}$ whereas eigenvectors will have only their lengths changed.

Dynamical systems and eigenanalysis

South Ken has two pizza places and $N \to \infty$ of pizza–loving students.

- Suppose that 5000 people buy one pizza each every week.

Tony's Pizza place has the better pizza and 80% of people who buy pizza each week at Tony's return the following week. Mike's Pizza does not have a good sauce and only 40% of the customers return the following week.

We can represent this situation by a discrete dynamical system

$$\mathbf{x}_{n+1} = \mathbf{A}\,\mathbf{x}_n \quad where \qquad \mathbf{A} = \begin{bmatrix} 0.8 & 0.6 \\ 0.2 & 0.4 \end{bmatrix}$$

Let us start from $\mathbf{x}_0 = [2500,\ 2500]^T$, then we have

$$\mathbf{x}_1 = \begin{bmatrix} 3500 \\ 1500 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} 3700 \\ 1300 \end{bmatrix}, \mathbf{x}_3 = \begin{bmatrix} 3740 \\ 1260 \end{bmatrix}, \mathbf{x}_4 = \begin{bmatrix} 3748 \\ 1252 \end{bmatrix}, \mathbf{x}_5 = \begin{bmatrix} 3750 \\ 1250 \end{bmatrix}$$

Also $\mathbf{x}_6 = \cdots = \mathbf{x}^* = \cdots = x_\infty = [3750, 1250]^T$

Clearly, $\mathbf{x}^* = [3750, 1250]^T$ is the **eigenvector of $\mathbf{A}$, and $\mathbf{A}\,\mathbf{x}^* = \mathbf{x}^*$**

Will Mike closed down – eigenanalysis answers

In Matlab

```
[v,lambda] = eig(A)          %also look at the demo 'eigshow(A)'

      0.9487   -0.7071
v =
      0.3162    0.7071

           1.0000          0
lambda =
           0               0.2000
```

$\looparrowright$ $\mathbf{v}_1 = [0.9487, 0.3162]^T, \ \mathbf{v}_2 = [-0.7071, 0.7071]^T, \quad \lambda_1 = 1, \ \lambda_2 = 0.2.$

Notice that the elements of $\mathbf{v}_1$ are related as $3 \div 1$, the same as the ratio of Tony's and Mike's customers

Since $\mathbf{x}_n = \mathbf{A}^n\mathbf{x}_0 \rightsquigarrow \mathbf{A}^n_{n\rightarrow\infty} = [\mathbf{v}_1 \vdots \mathbf{v}_2]$ $\looparrowright$ **equilibrium!**

This is closely related to fixed point theory since $\mathbf{A}$ is a Markov matrix

Spectral estimation as an eigenanalysis problem

Def: A function which remains unchanged when passed through a system, apart from a scaling by a constant, is called an **eigenfunction**, and the scaling constant is called an **eigenvalue**.

For a digital filter with the imp. resp. h_k, the eigenfunction e_k must satisfy

$$\lambda e_k = \sum_{i=-\infty}^{\infty} h_i e_{k-i} \qquad \text{no general method for deriving } \; e_k$$

Consider a candidate eigenfunction $e_k = \cos(\omega k)$, then

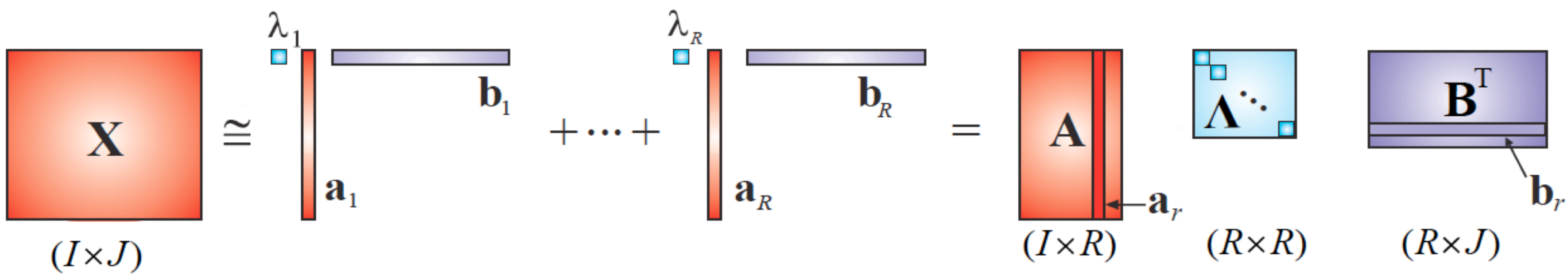
$$y_k = \sum_{i=-\infty}^{\infty} h_i \cos[\omega(k-i)] = \cos(\omega k)\Big[\sum_{i=-\infty}^{\infty} h_i \cos\omega i\Big] + \sin(\omega k)\Big[\sum_{i=-\infty}^{\infty} h_i \sin\omega i\Big]$$

- Clearly $\cos$ comes close, but is not suitable due to the $\sin$ terms.
- A sum $a\cos\omega k + b\sin\omega k = c\cos(\omega k + \Phi)$ is therefore not suitable either

On the other hand, for $e^{\jmath\omega k} = \cos\omega k + \jmath\sin\omega k$, we have

$$y_k = \sum_{i=-\infty}^{\infty} h_i e^{\jmath\omega(k-i)} = e^{\jmath\omega k}\Big[\sum_{i=-\infty}^{\infty} h_i e^{-\jmath\omega i}\Big] = e^{\jmath\omega k} H(\omega) \quad \text{clearly an eigenfunction}$$

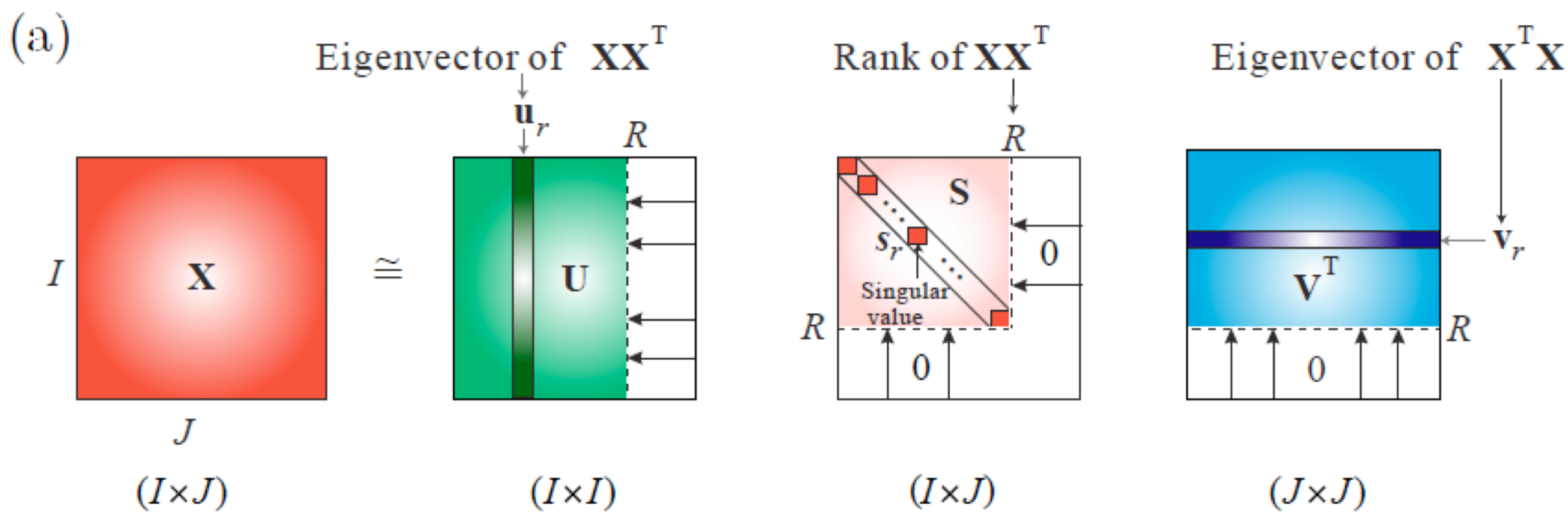
Matrix decomposition into rank-1 terms



Using SVD, we can decompose any matrix into a sum of "simpler" and "easier to swallow" rank-1 terms. The vectors **a** and **b** can be arranged into the matrices **A** and **B**

The eigenvalues (scaling factors) can be arranged into a diagonal matrix

This gives us the freedom to find **latent variables** in data (more about this at a later stage), and to perform **dimensionality reduction,** as illustrated below



Principal Component Analysis (PCA)

- Many signal processing, control and machine learning tasks employ multivariate data which often exhibit dependencies and redundancies.

- For example, it is often useful to reduce the dimensionality of a signal while maintaining the useful information.

- This reduces the computational complexity of any algorithm while preserving the physical meaning of the data.

- Besides dimensionality reduction, we often would like to transform the multi-channel data such each channel is orthogonal to each other (the data covariance matrix is diagonal)

- We use the PCA to accomlish this goal $\rightarrow$ The PCA has been called one of the most valuable results from applied linear algebra.

PCA – derivation

- Consider a general data vector, $\mathbf{x}_k \in \mathbb{C}^{M\times 1}$, with the empirical (sample) covariance matrix defined as

$$\mathrm{cov}(\mathbf{x}_k) \stackrel{\text{def}}{=} \mathbf{R}_\mathbf{x} = \frac{1}{N}\sum_{i=0}^{N-1} \mathbf{x}_k \mathbf{x}_k^H .$$

- Also, if we define a matrix $\mathbf{X} \in \mathbb{C}^{N\times M}$:

$$\mathbf{X}^T = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_N \end{bmatrix} \implies \mathbf{R}_\mathbf{x} = \tfrac{1}{N}\mathbf{X}^H\mathbf{X}$$

- The symmetric covariance matrix $\mathbf{R}_\mathbf{x}$ admits the following eigenvalue decomposition: $\mathbf{Q}^H\mathbf{R}_\mathbf{x}\mathbf{Q} = \mathbf{\Lambda}$
- The diagonal eigenvalue matrix, $\mathbf{\Lambda} = \mathrm{diag}\{\lambda_1, \lambda_2, \dots, \lambda_M\}$, indicates the power of each component of $\mathbf{x}_k$.
- The matrix of eigenvectors, $\mathbf{Q}_r = [\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_M]$, designates the principal directions of the data.

PCA as a prewhitening transform

- Suppose $\mathbf{x}_k$ is to be transformed into a vector, $\mathbf{u}_k \in \mathbb{C}^{M\times 1}$, using a linear transformation matrix $\mathbf{W}$, so that

$$\mathbf{u}_k = \mathbf{W}\mathbf{x}_k, \quad \text{where} \quad \text{cov}(\mathbf{u}_k) = \mathbf{I}.$$

- The PCA states that $\mathbf{W} = \mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{Q}^H$ can be obtained from the eigenvector and eigenvalue matrices.

- Proof:

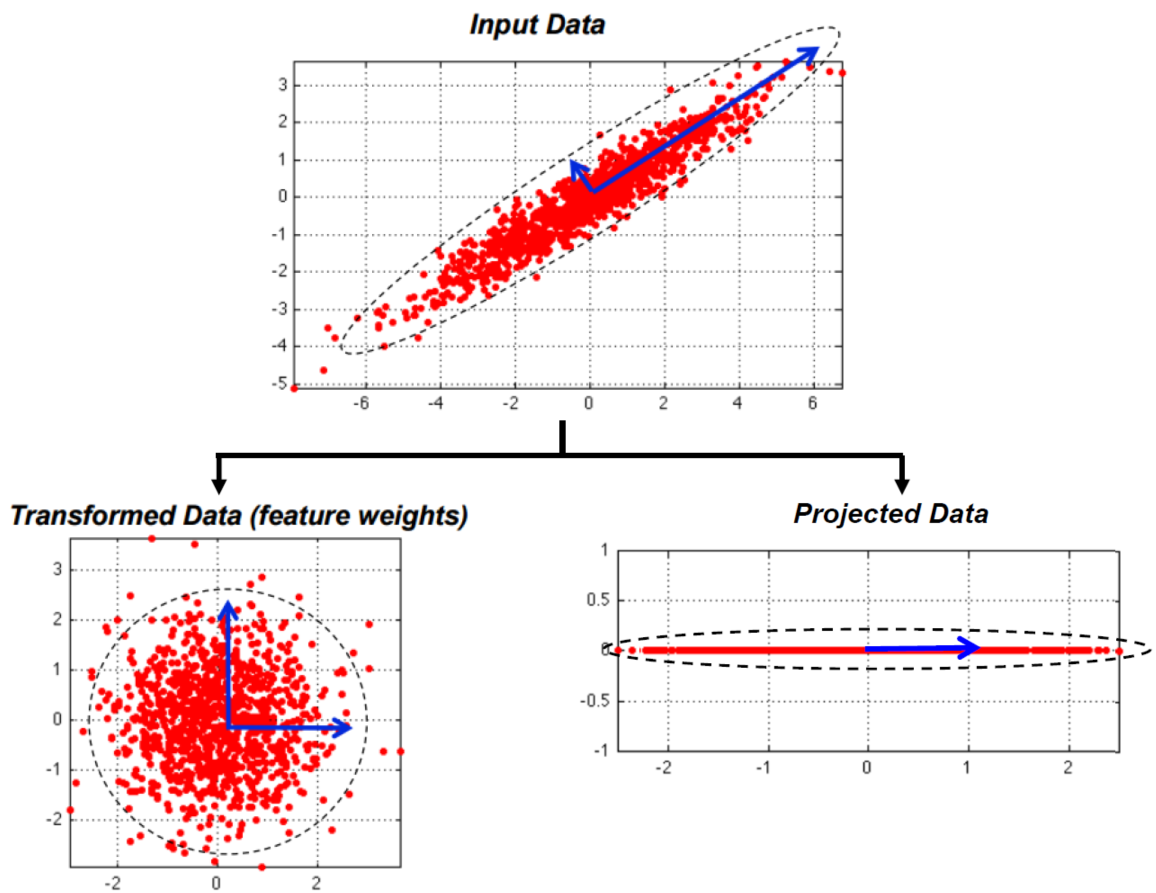
$$\begin{aligned}
\text{cov}(\mathbf{u}_k) &= \frac{1}{N}\sum_{i=0}^{N-1} \mathbf{u}_k\mathbf{u}_k^H \\
&= \mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{Q}^H \left(\frac{1}{N}\sum_{i=0}^{N-1} \mathbf{x}_k\mathbf{x}_k^H\right) \mathbf{Q}\mathbf{\Lambda}^{-\frac{1}{2}} \\
&= \mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{Q}^H\mathbf{R}_{\mathrm{x}}\mathbf{Q}\mathbf{\Lambda}^{-\frac{1}{2}} = \mathbf{I}.
\end{aligned}$$

PCA – dimensionality reduction

- PCA selects the **directions in which the data expresses the maximal variance**, that is, the directions of the principal eigenvectors of the data matrix.
- If we choose the r-leading principal components, $r < M$, we can reduce the dimensionality of the data onto the axes which exhibit the r-largest variances.
- To perform dimensionality reduction, the PCA can be applied to obtain the transformed data vector $\mathbf{u}_{r,k} \in \mathbb{C}^r$ with dimensions $r < M$ as

$$\mathbf{u}_{r,k} = \mathbf{W}_r \mathbf{x}_k = \mathbf{\Lambda}_{1:r}^{-\frac{1}{2}} \mathbf{Q}_{1:r}^T \mathbf{x}_k$$

- The PCA matrix $\mathbf{W}_r$ can be interpreted as a projection matrix as we are unable to recover $\mathbf{x}_k$ from the "reduced" data vector $\mathbf{u}_{r,k}$.



Back to spectral estimation – general bases

Consider a signal $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \dots & x_M \end{bmatrix}^T \in \mathbb{C}^{M \times 1}$. It can be represented with a different basis using

$$\mathbf{x} = \mathbf{F}\mathbf{w}$$

where $\mathbf{w} \in \mathbb{C}^{M \times 1}$The matrix $\mathbf{F} \in \mathbb{C}^{M \times M}$ can represent many different representations

- **Fourier basis:** Columns of $\mathbf{F}$ are sinusoids with different frequencies.
- **Wavelet basis:** Columns of $\mathbf{F}$ are wavelets.
- **Principal Component Analysis (PCA):** Rows of $\mathbf{F}$ are the eigenvectors of the covariance matrix of $\mathbf{F}$.

There are many other bases to represent signals which reside in a certain subspace of e.g. $\mathbb{R}^N$ or $\mathbb{C}^N$.

Rank of covariance matrix for sinusoidal data

Consider a single complex sinusoid with no noise

$$z_k = Ae^{\jmath\omega k} = A\cos(\omega k + \phi) + \jmath A\sin(\omega k + \phi)$$

There are two possible representations of the signal: A univariate complex-valued vector *or* bivariate real-valued matrix:

1. $\mathbf{z} = [z_0, z_1, \ldots, z_{N-1}]^T = A[1, e^{j\omega}, \ldots, e^{j(N-1)\omega}]^T \stackrel{\text{def}}{=} A\mathbf{e}$

2. $\mathbf{Z} = \begin{bmatrix} \text{Re}\{\mathbf{z}\} \\ \text{Im}\{\mathbf{z}\} \end{bmatrix} = A \begin{bmatrix} 1 & \cos(\omega+\phi) & \ldots & \cos(\omega(N-1)+\phi) \\ 0 & \sin(\omega+\phi) & \ldots & \sin(\omega(N-1)+\phi) \end{bmatrix}^T$

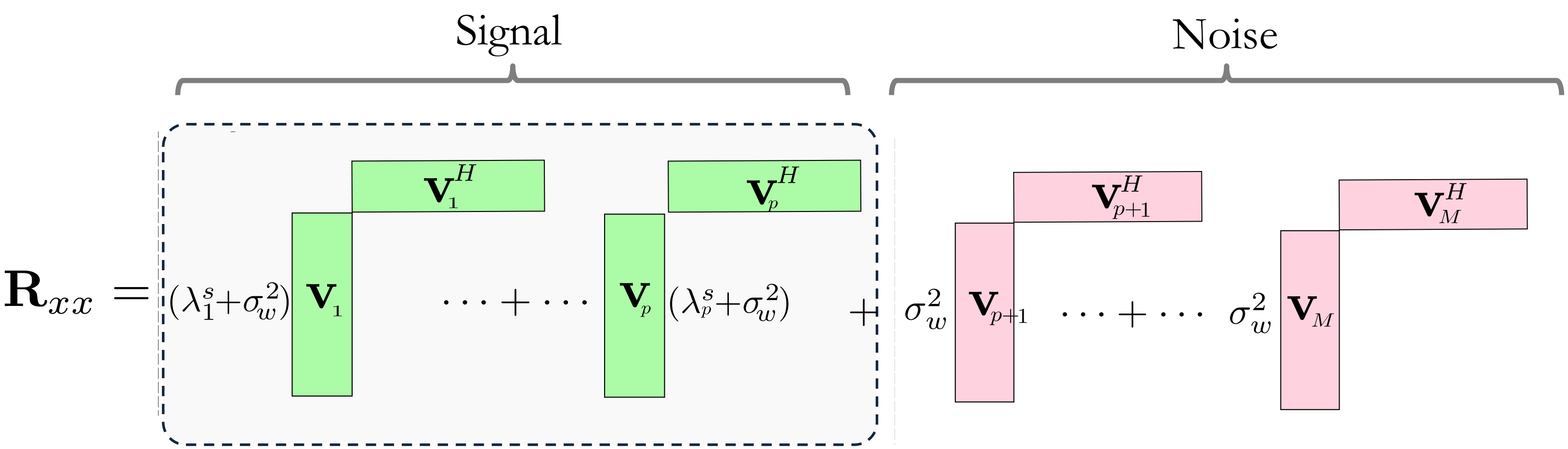
The corresponding covariance matrices exhibit a very interesting property:

- $\mathbf{C}_{zz} = E\{\mathbf{z}\mathbf{z}^H\} = |A|^2\mathbf{e}\mathbf{e}^H \rightarrow \text{Rank} = 1.$
- $\mathbf{C}_{ZZ} = E\{\mathbf{Z}\mathbf{Z}^T\} \rightarrow \text{Rank} = 2.$

What would happen with p sinusoids?

Principal Components Spectral Estimation

- PCA can be applied to the **autocorrelation matrix** of a univariate time series to find orthogonal directions of maximal variance, also known as **Singular Spectrum Analysis**.
- PCA **can also be used with Blackman–Tukey, maximum entropy method** and **AR spectrum estimation.**



$$\hat{\mathbf{R}}_{xx} \approx \hat{\mathbf{R}}_s = \sum_{i=1}^{p} \lambda_i \mathbf{v}_i \mathbf{v}_i^H$$

Linear Algebra terms: **We impose a rank p constraint** on $\mathbf{R}_{xx}$

Can we de-noise the signal by discarding the noise eigenvectors? $[\mathbf{v}_{p+1}, \ldots, \mathbf{v}_M]$

Subspace Methods for Spectral Estimation

- Eigenanalysis can be applied to the **autocorrelation matrix** of a univariate time series to find orthogonal directions of maximal variance, also known as **Singular Spectrum Analysis (SSA)**.

- Consider $x(n) = A_1 e^{\jmath n \omega_1} + w(n)$

$$\boxed{A_1 = |A_1| e^{\jmath \Phi} \qquad w(n) \sim \mathcal{N}(0, \sigma_w^2)}$$

- On vector notation: $\boxed{\mathbf{x} = A_1 \mathbf{e}_1 + \mathbf{w}}$

$$\mathbf{x} = [x(0),\, x(1),\, \ldots,\, x(M-1)]^T$$
$$\mathbf{e}_1 = [1,\, e^{\jmath\omega_1},\, \ldots,\, e^{\jmath\omega_1(M-1)}]^T$$

Autocorrelation: Matrix

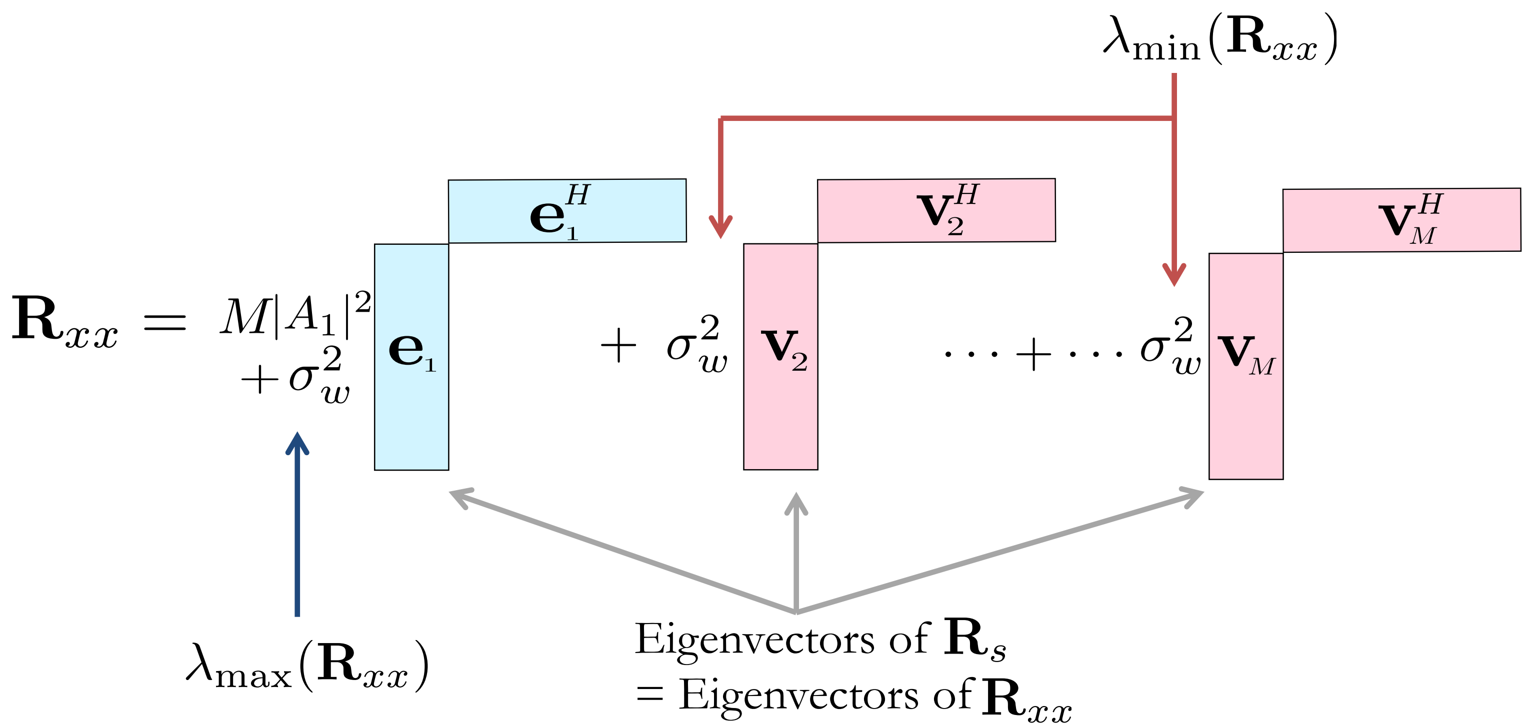
$$E(\mathbf{x}\mathbf{x}^H) = \mathbf{R}_{xx} = \underbrace{|A_1|^2 \mathbf{e}_1 \mathbf{e}_1^H}_{\mathbf{R}_s} + \underbrace{\sigma_w^2 \mathbf{I}}_{\mathbf{R}_n}$$

Signal Autocorrelation
Rank 1
Single non-zero Eigenvalue $= M|A_1|^2$

Noise Autocorrelation
Rank M
All Eigenvalues $= \sigma_w^2$

Decomposing the Autocorrelation Matrix

- $\mathbf{R}_s = |A_1|^2 \mathbf{e}_1 \mathbf{e}_1^H$ is Hermitian.
- Remaining M-1 eigenvectors are orthogonal to $\mathbf{e}_1$ $\rightsquigarrow$ $\mathbf{e}_1^H \mathbf{v}_i = 0, \;\; i = 2, \dots M$



Can we use the idea that $\mathbf{e}_1^H \mathbf{v}_i = 0$, to somehow estimate the power spectrum?

Multiple Sinusoids

Consider the signal $x(n) = A_1 e^{\jmath n\omega_1} + A_2 e^{\jmath n\omega_2} + w(n)$

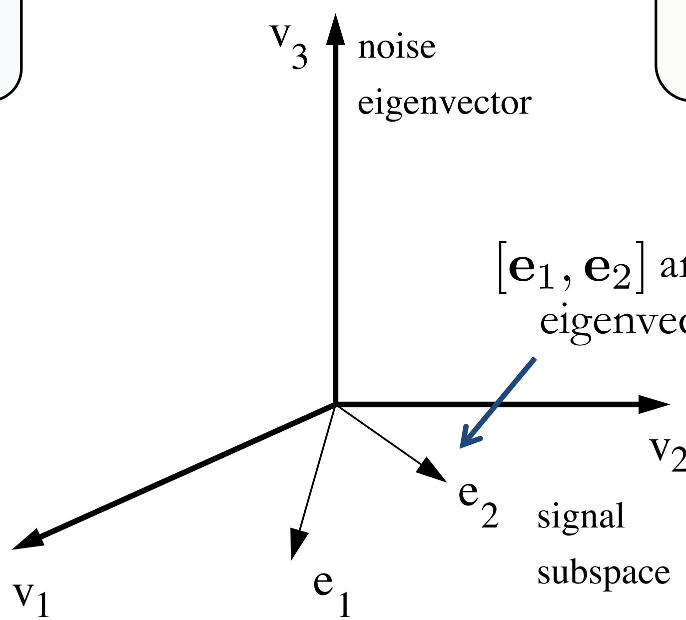
$$\mathbf{R}_{xx} = \mathbf{E}\mathbf{P}\mathbf{E}^H + \sigma_w^2 \mathbf{I} \qquad \mathbf{E} = [\mathbf{e}_1, \mathbf{e}_2]\,, \; \mathbf{P} = \mathsf{diag}(|A_1|^2, |A_2|^2)$$

Rank 2

The first 2 eigenvalues of $\mathbf{R}_{xx}$ are $\lambda_i^s + \sigma_w^2$
The remaining are σ_w^2

- Signal eigevectors span a 2D subspace
- Noise eigevectos span a (M-2) dim. subspace

The signal and noise subspaces are orthogonal



$[\mathbf{e}_1, \mathbf{e}_2]$ are not the signal eigenvectors in this case.

Subspace Methods

Extending to p sinusoids: $\mathbf{R}_{xx} = \mathbf{E}\mathbf{P}\mathbf{E}^H + \sigma_w^2 \mathbf{I}$

$$\mathbf{E} = [\mathbf{e}_1, \ldots, \mathbf{e}_p]\,, \; \mathbf{P} = \mathrm{diag}(|A_1|^2, \ldots, |A_p|^2)$$

Using $\mathbf{e}_i^H \mathbf{v}_k = 0 \quad \begin{cases} i = 1, \ldots, p \\ k = p+1, \ldots, M \end{cases}$

$\Longrightarrow$ PSD estimation can be performed as: $\hat{P}_{sub}(\omega) = \dfrac{1}{\sum_{i=p+1}^{M} \alpha_i \left|\mathbf{e}^H \mathbf{v}_i\right|^2}$

Pisarenko Harmonic Decomposition

$$\boxed{\hat{P}_{PHD}(\omega) = \frac{1}{\left|\mathbf{e}^H \mathbf{v}_{\min}\right|^2}}$$

MUltiple Signal Classification (MUSIC)

$$\boxed{\hat{P}_{MU}(\omega) = \frac{1}{\sum_{i=p+1}^{M} \left|\mathbf{e}^H \mathbf{v}_i\right|^2}}$$

EigenVector Method

$$\boxed{\hat{P}_{EV}(\omega) = \frac{1}{\sum_{i=p+1}^{M} \frac{1}{\lambda_i} \left|\mathbf{e}^H \mathbf{v}_i\right|^2}}$$

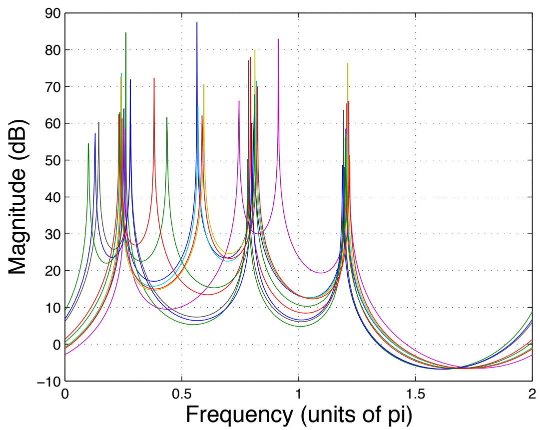
Minimum Norm Method

$$\boxed{\hat{P}_{MN}(\omega) = \frac{1}{\left|\mathbf{e}^H \mathbf{a}\right|^2}}$$

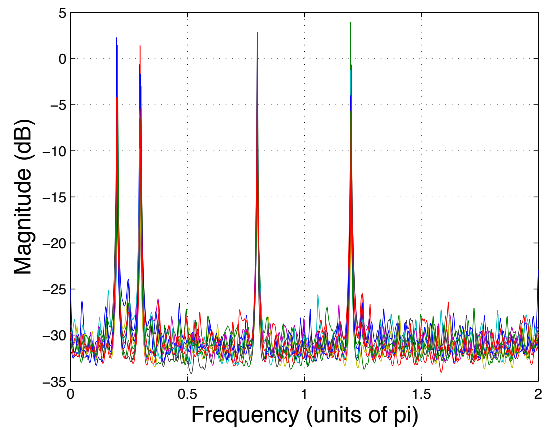
$\mathbf{a} \in$ Noise Subspace & has min. norm

Comparison of the 4 Subspace Methods

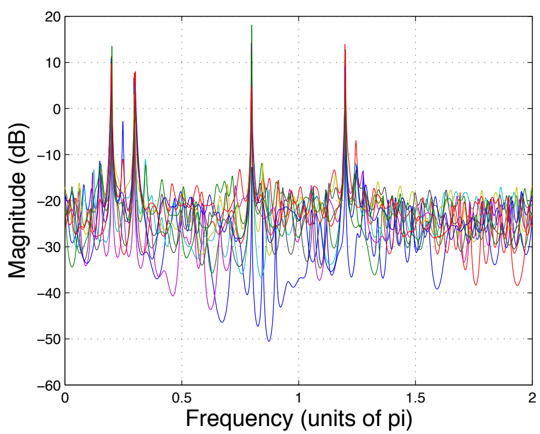
Pisarenko



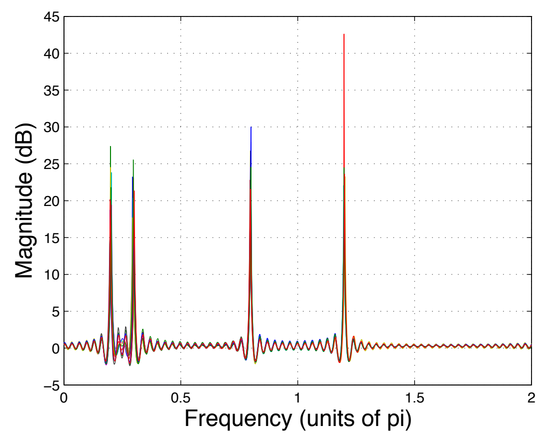
MUSIC



EigenVector



Minimum Norm



Overlay of 10 different realizations of 4 complex sinusoids in white noise.

Pisarenko only needs a 5 x 5 correlation matrix

A 64 x 64 correlation matrix was used for other methods

Except for Pisarenko's method, all other estimates are correct!

Summary of the Different Methods