Sinusoidal
Parametric
Modeling
Sinusoidal Parametric Modeling

Assume the observed signal is a linear combination of \( p \) complex sinusoids in white complex Gaussian noise:

\[
x[n] = \sum_{j=1}^{n} A_j \exp\left\{ i (2\pi f_j n + \phi_j) \right\} + Z[n]
\]

where

- \( A_j, f_j \) → unknown and desired
- \( \phi_j \) → unknown and possibly desired.
- \( Z[n] \) → CWGN with variance \( \sigma^2 \).

Like ARMA Modeling:

This is a parametric model. If we can estimate the parameters of the model, then the corresponding spectral estimate is

\[
\hat{P}_{\text{sin}}(f) = \]
\[ \hat{P}_{\sin}(f) = \sum_{j=1}^{P} A_j^2 \delta(f - f_j) + \sigma^2 \]

assuming \( \phi_i \) are \( \text{unif}[0,2\pi) \) and independent.

(The ACV is

\[ \gamma_x[k] = \sum_{j=1}^{P} A_j^2 \exp\left\{2\pi i f_j k\right\} + \sigma^2 \delta[k] \]

Unlike ARMA Modeling:

The model needs to be very accurate for the method to be effective.

Whereas ARMA models can approximate arbitrary spectra to any specified accuracy by taking sufficiently high order models, the same is not true for the sinusoidal model. An arbitrary spectrum cannot be written as a countably infinite \( (p = \infty) \) sum of delta functions.
Sinusoidal Parameter Estimation

Two cases:

A  \( f_1, \ldots, f_p \) known \( \Rightarrow \) MLE has nice form \( \Rightarrow \) easy

B  \( f_1, \ldots, f_p \) unknown \( \Rightarrow \) MLE requires difficult, nonlinear optimization \( \Rightarrow \) hard.

Case B is most common in practice. Therefore, the primary challenge is frequency estimation.

We will discuss three classes of methods for this:

1. Tackle the nonlinear MLE problem
2. Fit an AR model and find the peaks
3. If \( p < N \), then \( x[n] \) belongs to a signal subspace that is characterized by the eigenvectors of the ACV matrix.

Once the frequencies have been estimated, we're in the easy case A and can apply MLE.
Maximum Likelihood Estimation

Let's first consider the case of a single sinusoid

\[ x[n] = A \exp \left\{ i(2\pi fn + \phi) \right\} + z[n] \]

In vector notation,

\[ x = A_c \cdot e(f) + z \]

where

\[ A_c = A e^{i\phi} \]

\[ e(f) = [1 \ e^{2\pi if} \ e^{4\pi if} \ ... \ e^{2(N-1)\pi if}]^T \]

The likelihood function for \( \Theta = [A_c, f]^T \) is

\[ l(A_c, f) = \frac{1}{(\pi \sigma^2)^N} \exp \left\{ \frac{-(x - A_c e(f))^H (x - A_c e(f))}{\sigma^2} \right\} \]

Therefore, we need to minimize

\[ (x - A_c e(f))^H (x - A_c e(f)) \]

with respect to \( A_c \in \mathbb{C}, f \in [-\frac{1}{2}, \frac{1}{2}) \).
Previously we saw that for fixed $f$, the minimizing $A_c$ is

$$
\hat{A}_c(f) = \frac{e(f) \cdot \mathbf{\Gamma}^{-1} \cdot x}{\mathbf{e}(f) \cdot \mathbf{\Gamma}^{-1} \cdot \mathbf{e}(f)} = \frac{e(f) \cdot x}{\mathbf{e}(f) \cdot \mathbf{e}(f)}
$$

$$
= \frac{1}{N} \sum_{n=0}^{N-1} x[n] \exp(-2\pi ifn)
$$

Exercise

1. What are the MLEs of $A, \phi$, assuming $f$ is known?

2. What is $(x - \hat{A}_c e(f))^H (x - \hat{A}_c e(f))$?
Solution

1. \( \hat{A} = |\hat{A}_c| = \left| \frac{1}{N} \sum_{n=0}^{N-1} x[n] \exp \{-2\pi i \hat{f} n\} \right| \)

   \( \hat{\phi} = \arg \hat{A}_c = \tan^{-1} \left( \frac{\text{Im} \left( \frac{1}{N} \sum_{n=0}^{N-1} x[n] \exp \{-2\pi i \hat{f} n\} \right)}{\text{Re} \left( \frac{1}{N} \sum_{n=0}^{N-1} x[n] \exp \{-2\pi i \hat{f} n\} \right)} \right) \)

2. \( (x - \hat{A}_c \bar{e}(f))^H (x - \hat{A}_c \bar{e}(f)) \)
   
   \( = x^H (x - \hat{A}_c \bar{e}(f)) - \hat{A}_c^* \bar{e}(f)^H (x - \hat{A}_c \bar{e}(f)) \)

   \( = \hat{A}_c^* (N \cdot \hat{A}_c) - \hat{A}_c^* \hat{A}_c \cdot N \)
   
   \( = 0 \)

   \( = x^H x - \hat{A}_c \cdot x^H \bar{e}(f) \)

   \( = x^H x - \frac{1}{N} \left| \bar{e}(f)^H x \right|^2 \)
Conclusion: The MLE of $f$ is obtained by maximizing

$$\frac{1}{N} \left| \mathbb{E}(f)^H x \right|^2$$

$$= \frac{1}{N} \left| \sum_{n=0}^{N-1} x[n] \exp \left\{ -2\pi if_n \right\} \right|^2$$

$$= \hat{p}_{\text{PER}}(f).$$

In other words,

$$\hat{f}_{\text{MLE}} = \arg\min_{f \in [-\frac{1}{2}, \frac{1}{2}]} \hat{p}_{\text{PER}}(f)$$
Multiple Sinusoids

The preceding method can be generalized.

Assume

\[ x[n] = \sum_{j=1}^{P} A_c^j \exp\{2\pi i f_j n\} + \Xi[n] \]

In vector notation:

\[ \mathbf{x} = \sum_{j=1}^{P} A_c^j \mathbf{e} \begin{pmatrix} f_j \end{pmatrix} + \Xi \]

where

\[ A_c^j = A_j e^{i\phi_j} \]

\[ \mathbf{e} \begin{pmatrix} f_j \end{pmatrix} = [1 \quad e^{2\pi i f_j} \quad \cdots \quad e^{2\pi i f_j (N-1)}]^T \]

In matrix notation:

\[ \mathbf{x} = \mathbf{E} \cdot \mathbf{A}_c + \Xi \]
where

\[ E = \begin{bmatrix} \mathbf{e}(f_1) \\ \vdots \\ \mathbf{e}(f_p) \end{bmatrix} \]

\[ A_c = [A_c^1 \ldots A_c^p]^T. \]

Again the MLE is obtained by minimizing

\[(x - EA_c)^H (x - EA_c)\]

w.r.t. \(A_c\). This is a standard least squares problem and the solution is

\[ \hat{A}_c(f) = (E^H E)^{-1} E^H x \]
The minimum value of the squared error is

\[ x^H x - x^H E (E^H E)^{-1} E^H x \]

and therefore the MLE is obtained by maximizing

\[ x^H E (E^H E)^{-1} E^H x \]

w.r.t. \( f_1, \ldots, f_p \).

**Summary of discussion in Kay**:

- Objective function is highly nonlinear and very difficult to maximize when the true frequencies are very close together.

- When true frequencies are separated by \( \geq \frac{1}{N} \), a good approximate MLE is to take the \( p \) largest peaks of the periodogram, subject to those peaks being \( \geq \frac{1}{N} \) apart.
Frequency Estimation via High-Resolution S. E.

Since the periodogram cannot reliably resolve frequencies differing by \( \leq \frac{1}{N} \), an alternative is to identify the frequencies using a high resolution spectral estimator.

The AR SE is a popular choice.

---

Two approaches

1. Apply AR SE method of order \( p' \geq p \), such as the modified covariance method, and choose the frequencies corresponding to the \( p \) poles closest to the unit circle.
2. Constrain AR model so that poles lie on the unit circle, giving rise to pure sinusoidal components.

The resulting constrained optimization problem is difficult.

An efficient but suboptimal algorithm called the iterative filtering algorithm (IFA) is discussed in Kay.

Remarks: AR methods are good at low noise levels but suffer when noise is strong.
Idea: It turns out (we will show) that the linear combo of sinusoids lives in a subspace of dimension $p$.

This subspace is the span of the $p$ most significant components (the principal components) of the ACV $\Sigma$.

We can use this fact to our advantage, to modify existing spectral estimators and derive new ones.

In short, we can use the signal subspace property to do high resolution spectral estimation that has higher noise tolerance than AR and other methods studied previously.
To even talk about \(ACV\) matrices, we need \(X[n]\) to be WSS which means we need to assume

\[
\phi_j \sim \text{unif} [0, 2\pi), \quad j = 1, \ldots, p
\]
\[
\phi_j, \phi_k \text{ independent, } j \neq k
\]

Under this assumption, if

\[
X[n] = \sum_{j=1}^{p} A_j \exp \left\{ i (2\pi f_j n + \theta_j) \right\} + z[n]
\]

then

\[
\gamma_n[k] = \sum_{j=1}^{p} A_j^2 \exp \left\{ 2\pi f_j k \right\} + \sigma_z^2 \delta[k]
\]
The $M \times M$ autocorrelation matrix $\Gamma_{xx}$ is

$$
\Gamma_{xx} = \sum_{j=1}^{P} A_j^2 \mathbf{e}(f_j) \mathbf{e}(f_j)^H + \sigma_z^2 \mathbf{I}_M
$$

$$
= \Gamma_{ss} + \Gamma_{zz}
$$

**Exercise**

1. What are the eigenvalues/vectors of $\Gamma_{zz}$?
2. Show that if $\mathbf{v}$ is an eigenvector of $\Gamma_{ss}$ with eigenvalue $\lambda$, then $\mathbf{v}$ is also an eigenvector of $\Gamma_{xx}$ and find the corresponding eigenvalue.
3. Can you guess the eigenvectors of $\Gamma_{ss}$?
Solution

1. Every vector is an eigenvector of \( \Sigma \), and the eigenvalue is \( \frac{\sigma^2}{\Sigma} \):

\[
\Sigma \cdot \nu = \frac{\sigma^2}{\Sigma} \cdot I \cdot \nu = \frac{\sigma^2}{\Sigma} \cdot \nu
\]

\( \Sigma \)

2. If

\[
\Sigma \cdot \nu = \lambda \nu
\]

then

\[
\Sigma \cdot \nu = \Sigma \cdot \nu + \Sigma \cdot \nu
\]

\[
= \lambda \nu + \frac{\sigma^2}{\Sigma} \nu
\]

\[
= (\lambda + \frac{\sigma^2}{\Sigma}) \nu
\]

\( \uparrow \) eigenvalue
(3) In general, there is no nice formula for the eigenvectors of $\mathbf{P}_{ss}$.

**Eigenvectors of $\mathbf{P}_{ss}$**

Observe

$$
\mathbf{P}_{ss} = \sum_{j=1}^{p} A_j^2 e_j e_j^H
$$

= sum of $p$ rank 1 matrices

$$
\Rightarrow \text{rank}(\mathbf{P}_{ss}) \leq p.
$$

Note: $\mathbf{P}_{zz}$ is full rank, and hence so is $\mathbf{P}_{xx}$.

Let $v_1, \ldots, v_m$ be an orthonormal basis of $\mathbb{C}^m$ consisting of eigenvectors of $\mathbf{P}_{ss}$. Let $\lambda_1, \ldots, \lambda_m$ be the corresponding eigenvalues. Assume the eigenvectors are ordered such that

$$
\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \geq 0
$$

↑ why?
since \( \text{rank} (\Gamma_{ss}) \leq p \), we know
\[
\lambda_{p+1} = \lambda_{p+2} = \ldots = \lambda_m = 0.
\]
Assuming \( f_1, \ldots, f_p \) are distinct, we also have \( \text{rank} (\Gamma_{ss}) = p \), in which case
\[
\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p > 0.
\]

**Super Useful Fact**

\[
\text{span} \left\{ e(f_1), \ldots, e(f_p) \right\} = \text{span} \left\{ u_1, \ldots, u_p \right\}
\]

Now define the

- **signal subspace** := \( \text{span} \left\{ u_1, \ldots, u_p \right\} \)

- **noise subspace** := \( \text{span} \left\{ u_{p+1}, \ldots, u_m \right\} \).

Why is it called the "signal subspace?"
Because we are interested in signals of the form

\[ x = \sum_{j=1}^{p} A_c^j \mathbf{e}(f_j) \]

\[ \in \text{span} \{ \mathbf{e}(f_1), \ldots, \mathbf{e}(f_p) \} \]

\[ = \text{span} \{ \mathbf{u}_1, \ldots, \mathbf{u}_p \} \]

**Exercise**: Show that signal vectors are orthogonal to all vectors in the noise subspace.

**Solution**: Let

\[ z = \sum_{j=1}^{p} A_c^j \mathbf{e}(f_j) \]

\[ = \sum_{j=1}^{p} \alpha_j^j \mathbf{u}_j \in \text{signal subspace} \]

and

\[ t = \sum_{j=p+1}^{M} \beta_j^j \mathbf{u}_j \in \text{noise subspace} \]

Then

\[ z^H t = \sum_{j=1}^{p} \sum_{k=p+1}^{M} \alpha_j^j \beta_k \mathbf{u}_j^H \mathbf{u}_k \]

\[ = 0 \]
Key point

- There is no signal in the noise subspace.
- There is some noise in the signal subspace.

Terminology:

\(u_1, \ldots, u_p\) are called principal components or principal eigenvectors.
Basic Idea: Several spectral estimators can be expressed in terms of $\hat{\Pi}_{xx}$:

→ AR

$$\hat{a} = -\hat{\Pi}_{xx}^{-1} \hat{\xi}_x$$

→ Capon

$$\hat{P}_{\text{Capon}}(f) = \frac{M}{\hat{\xi}(f)^H \hat{\Pi}_{xx} \hat{\xi}(f)}$$

→ Blackman Tukey with Bartlett window:

$$\hat{P}_{\text{BAR}}(f) = \frac{1}{M} \hat{\xi}(f)^H \hat{\Pi}_{xx} \hat{\xi}(f)$$

Let's replace standard estimates of $\hat{\Pi}_{xx}$ with estimates that use the signal subspace property.
Exercise 1 Verify that

\[ \Gamma'_{xx} = \sum_{j=1}^{p} (\lambda_j + \sigma_x^2) v_j v_j^H + \sum_{j=p+1}^{M} \sigma_x^2 v_j v_j^H \]

Solution 1 We know \( \Gamma'_{xx} v_j = (\lambda_j + \sigma_x^2) v_j \)

Let

\[ \Lambda : = \begin{bmatrix} 1 & \cdots & 1 \\ v_1 & \cdots & v_M \\ 1 & \cdots & 1 \end{bmatrix}, \quad \Lambda = \begin{bmatrix} \lambda_1 + \sigma_x^2 & \cdots & \lambda_p + \sigma_x^2 \\ \sigma_x^2 v_1 & \cdots & \sigma_x^2 v_M \end{bmatrix} \]

Then \( \Gamma'_{xx} V = V \cdot \Lambda, \)

\[ \Longrightarrow \quad \Gamma'_{xx} = V \cdot \Lambda \cdot V^H \]

\[ = \sum_{j=1}^{M} (\lambda_j + \sigma_x^2) v_j v_j^H \]

\[ = \sum_{j=1}^{p} (\lambda_j + \sigma_x^2) v_j v_j^H + \sum_{j=p+1}^{M} \sigma_x^2 v_j v_j^H \]
Recall the Blackman-Tukey (lag window) s.e.

\[ \hat{P}_{BT}(f) = \sum_{k=-M}^{M} w[k] \delta_{xx}[k] e^{-2\pi i f k} \]

Using the Bartlett window \( w[k] = M - |k| \)

\[ w[k] = \begin{cases} \frac{M - |k|}{M} & |k| \leq M \\ 0 & |k| > M \end{cases} \]

we have

\[ \hat{P}_{BAR}(f) = \frac{1}{M} \sum_{k=-(M-1)}^{M-1} (M-|k|) \delta_{xx}[k] e^{-2\pi i f k} \]

\[ = \frac{1}{M} \sum_{M=0}^{M-1} \sum_{n=0}^{M-1} \delta_{xx}[M-n] e^{-2\pi i (M-n) f} \]

\[ = \frac{1}{M} \mathbf{e}(f)^H \sum_{M=0}^{M-1} \sum_{n=0}^{M-1} \mathbf{g}[M-n] \mathbf{e}(f) \]

where we have used the identity

\[ \sum_{k=-(M-1)}^{M-1} (M-|k|) g[k] = \sum_{M=0}^{M-1} \sum_{n=0}^{M-1} g[M-n] \]
The principal component Bartlett SE replaces

\[ \hat{\Gamma}_{xx} = \sum_{i=1}^{M} \hat{\lambda}_i \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^H \]

with

\[ \hat{\Gamma}_{xx} = \sum_{i=1}^{P} \hat{\lambda}_i \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^H \]

This results in

\[ \hat{P}_{\text{BAR-PC}}(f) = \frac{1}{M} \mathbf{e}(f)^H \left[ \sum_{i=1}^{P} \hat{\lambda}_i \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^H \right] \mathbf{e}(f) \]

\[ = \frac{1}{M} \sum_{i=1}^{P} \hat{\lambda}_i \left| \mathbf{e}(f)^H \hat{\mathbf{u}}_i \right|^2 \]

Contrast with the original Bartlett SE:

\[ \hat{P}_{\text{BAR}}(f) = \frac{1}{M} \sum_{i=1}^{P} \hat{\lambda}_i \left| \mathbf{e}(f)^H \hat{\mathbf{u}}_i \right|^2 + \frac{1}{M} \sum_{i=p+1}^{M} \hat{\lambda}_i \left| \mathbf{e}(f)^H \hat{\mathbf{u}}_i \right|^2 \]

\[ \Rightarrow \text{projections onto noise subspace} \]

\[ \Rightarrow \text{spurious peaks} \]
Algorithm

1. Estimate $\hat{\Gamma}_x$ using standard method, such as modified covariance method.

2. Find principal eigenvectors and eigenvalues of $\hat{\Gamma}_x$ (eig function in Matlab)

3. Plug in to $\Theta$

Note: Here I'm using $\lambda_i$ to denote an eigenvalue of $\hat{\Gamma}_x$. Previously, I used $\lambda_i$ to denote an eigenvalue of $\Gamma_{ss}$. Kay does the same.
Principal Component Capon Spectral Estimator

If

$$\Gamma_{xx} = V \cdot \Lambda \cdot V^H$$

then

$$\Gamma_{xx}^{-1} = V \Lambda^{-1} V^H$$

$$= \sum_{i=1}^{M} \frac{1}{\lambda_i} \hat{u}_i \hat{u}_i^H$$

This suggests the following modification of Capon's method:

$$\hat{P}_{\text{CAP-PC}}(f) = \frac{\sum_{i=1}^{M} \frac{1}{\lambda_i} \hat{u}_i \hat{u}_i^H \hat{e}(f)^H \hat{e}(f)}{\sum_{i=1}^{P} \left| \hat{e}(f)^H \hat{u}_i \right|^2}$$
Principal Component AR Spectral Estimator

The AR spectral estimator computes

\[ \hat{\gamma} = -\hat{\Pi}^{-1}_{xx} \hat{x}_x \]

and then

\[ \hat{P}_{AR}(f) = \frac{1}{|\hat{A}(f)|^2} \]

where

\[ \hat{A}(f) = 1 + \sum_{k=1}^{M} a[k] e^{-2\pi i fk} \]

The principal component version of the AR S.E. replaces \( \hat{\gamma} \) by

\[ \hat{\gamma}_{PC} = -\left[ \sum_{i=1}^{p} \frac{1}{\lambda_i} \hat{U}_i \hat{U}_i^H \right] \hat{x}_x \]
Noise Subspace (Nonprincipal Component) Frequency Estimation

The PC version of Capon’s method discussed earlier seems reasonable at first glance. Recall

\[ \hat{P}_{\text{CAP}}(f) = \frac{M}{\sum_{i=1}^{P} \frac{1}{\lambda_i} |e(f)^H \hat{\mathbf{v}}_i|^2 + \sum_{i=p+1}^{M} \frac{1}{\lambda_i} |e(f)^H \hat{\mathbf{v}}_i|^2} \]

We replaced (B) with 0. This was consistent with our knowledge that \( \Gamma_{xx} \) can be expressed using only the principal components. This approach reflects a desire to improve the spectral estimator as a whole.

Instead, what if we focus on frequency estimation. How should we modify \( \hat{P}_{\text{CAP}}(f) \) so that it has sharp peaks at \( f = f_i \), \( i=1, \ldots, n \)?
Eigenvector (EV) Method

\[
\hat{P}_{EV}(f) = \frac{M}{\sum_{\hat{\lambda}_i} \frac{1}{\hat{\lambda}_i} |\mathbf{e}(f)^H \hat{\mathbf{v}}_i|^2}
\]

If \( \hat{\mathbf{v}}_i \) are the true \( \mathbf{v}_i \), then

\[
\hat{P}_{EV}(f_i) =
\]

In practice, \( \hat{\mathbf{v}}_i \neq \mathbf{v}_i \) and \( \hat{P}_{EV}(f) \)

will have peaks at the \( f_i \).

**MUSIC** (Multiple Signal Classification)

\( \rightarrow \) clever acronym, but does not do classification

\( \rightarrow \) set \( \hat{\lambda}_i = 1 \) in EV method

\[
\hat{P}_{music}(f) = \frac{M}{\sum_{\hat{\lambda}_i} |\mathbf{e}(f)^H \hat{\mathbf{v}}|^2}
\]

Why is this reasonable?
Setting $\hat{\lambda}_i = 1$, $i = p+1, \ldots, n$, is reasonable because

(a) $\hat{\lambda}_i = \frac{1}{\hat{\sigma}^2}$ for $i = p+1, \ldots, n$

(b) We are only interested in frequency locations, not amplitudes

Figure 7.14 The eigenvector (EV) and MUSIC variations of truncating the eigenexpansions of the spatial correlation matrix’s inverse are shown in the context of the minimum variance algorithm. Two signals are propagating from $-5^\circ$ and $5^\circ$ in the presence of spherically isotropic noise. The signal-to-noise ratio for each signal is 0 dB, and the time-bandwidth product of the spatial correlation matrix’s estimate equals 100. The minimum variance algorithm alone does not have sufficient resolving power to distinguish the two signals. Either of the eigenanalysis-based methods can be at about the same level of performance. The greatest difference between the methods occurs in the invisible region; the eigenvector method yields the same results as minimum variance while MUSIC yields a flattened (whitened) result.


Comments:

- CAPON-PC does not have as sharp peaks as EV + MUSIC

- MUSIC has flat baseline which results from setting $\hat{\lambda}_i = 1$. 
If we choose $M = p+1$ then the dimension of the noise subspace is

Let $v_{p+1}$ be the non principal component spanning the noise subspace.

Then $v_{p+1} \perp e_{j}(f_{j})$ for $j=1,\ldots, p$.

$$0 = \sum_{n=1}^{p+1} [v_{p+1}]_{n} \cdot e^{-2\pi i f_{j}(n-1)/f_{j}}$$

$$= \sum_{n=1}^{p+1} [v_{p+1}]_{n} \cdot z_{j}^{(n-1)}$$

That is, $z_{1},\ldots,z_{p}$ are zeros of the polynomial

$$\sum_{n=1}^{p+1} [v_{p+1}]_{n} \cdot z^{n-1}$$

$$z_{j} = e^{-2\pi i f_{j}/f_{j}}$$
This suggests the following algorithm:

1. Compute $\hat{z}_1, \ldots, \hat{z}_{p+1}$

2. Find the zeros of the polynomial

$$\sum_{n=1}^{p+1} [\hat{v}_{p+n}]_n z^n$$

3. Let $\hat{z}_1, \ldots, \hat{z}_p$ be the zeros closest to the unit circle.

4. Set $f_j = \frac{1}{\text{arg} (\hat{z}_j^{-1})}$ (Avoid this Do this first)

Remarks

- If we compute $\hat{z}_j$ using the biased sample ACV for $F_{xx}$, then we retain the properties of the theoretical ACV matrix. In particular, $\hat{z}_1, \ldots, \hat{z}_p$ will lie on the unit circle.

- Extension to $M > p+1$ discussed in Kay.