AR Spectral Estimation In Practice

The Big Picture

If a linear prediction filter $A(z)$ can approximately whiten the data $x[n]$

$$x[n] \longrightarrow \boxed{A(z)} \longrightarrow e[n]$$

$\approx$ white

then the all-pole system

$$w[n] \longrightarrow \boxed{\frac{1}{A(z)}} \longrightarrow \tilde{x}[n]$$

white

is an approximation of $x[n]$ by an AR process

$$\Longrightarrow \hat{P}_{AR}(f) = \frac{\sigma_w^2}{|A(f)|^2}$$

is a reasonable estimate of the PSD of $x[n]$.

KEY ISSUES

- how to choose p
- how to estimate $a[1], \ldots, a[p]$ given only data.

The Autocovariance Method

The error sequence $e[n]$ is

$$e[n] = x[n] + \sum_{k=1}^{p} a[k]\, x[n-k].$$

Invoking ergodicity, an estimate of the prediction error power ρ based on an observed sequence $x[0], \dots, x[N-1]$ is

$$\hat{\rho} = \frac{1}{N} \sum_{n=-\infty}^{\infty} \left| x[n] + \sum_{k=1}^{p} a[k]\, x[n-k] \right|^2$$

So, let's choose $a[1], \dots, a[p]$ to minimize this estimate. For convenience, assume a real RP.

<u>Note</u>: Assume $x[n] = 0$ for $n < 0$, $n > N-1$.

<u>Exercise</u> Compute the derivative of $\hat{\rho}$ with respect to $a[\ell]$.

Solution/

$$\frac{\partial \hat{\rho}}{\partial a[\ell]} = \frac{1}{N} \sum_{n=-\infty}^{\infty} \left(x[n] + \sum_{k=1}^{p} a[k]\, x[n-k] \right) \cdot x[n-\ell]$$

$$\ell = 1, 2, \ldots, p$$

But

$$\frac{1}{N} \sum_{n=-\infty}^{\infty} x[n]\, x[n-\ell] = \hat{\gamma}[\ell]$$

sample autocovariance

and

$$\frac{1}{N} \sum_{k=-\infty}^{\infty} x[n-k]\, x[n-\ell] = \hat{\gamma}[\ell-k]$$

So, the optimal $a[1], \ldots, a[k]$ must satisfy

$$\hat{\gamma}[\ell] + \sum_{k=1}^{p} a[k]\, \hat{\gamma}[\ell-k] = 0, \quad \ell = 1, \ldots, p$$

or, in matrix form

$$\underbrace{\begin{bmatrix} \hat{\gamma}[0] & \hat{\gamma}[-1] & \cdots & \hat{\gamma}[-(p-1)] \\ \hat{\gamma}[1] & \hat{\gamma}[0] & \cdots & \hat{\gamma}[-(p-2)] \\ \vdots & & & \\ \hat{\gamma}[p-1] & \hat{\gamma}[p-2] & \cdots & \hat{\gamma}[0] \end{bmatrix}}_{\hat{\Gamma}_p} \begin{bmatrix} a[1] \\ a[2] \\ \vdots \\ a[p] \end{bmatrix} = \begin{bmatrix} \hat{\gamma}[1] \\ \hat{\gamma}[2] \\ \vdots \\ \hat{\gamma}[p] \end{bmatrix}$$

In other words, the autocovariance method is the same as solving the Yule-Walker equations with the sample autocovariances. (can use Levinson alg.)

Since $\hat{\gamma}[k]$ is a positive definite sequence, the solution will result in a causal, stable AR model.

For obvious reasons, this method is also called the Yule-Walker method.

The autocovariance method does have drawbacks. Can you think of any?

Matrix notation:

$$
\begin{array}{c}
\\ \\ \\ \\ \\ N-p \\ \\ \\ \\ \\ \\
\end{array}
\begin{bmatrix}
x[0] & 0 & 0 & \cdots & 0 & 0 \\
x[1] & x[0] & 0 & & 0 & 0 \\
\vdots & & & & & \\
x[p-1] & x[p-2] & & & x[0] & 0 \\
x[p] & x[p-1] & & \cdots & x[1] & x[0] \\
x[p+1] & x[p] & & & x[2] & x[1] \\
\vdots & & & & & \\
x[N-1] & x[N-2] & & \cdots & x[N-p] & x[N-p-1] \\
0 & x[N-1] & & & x[N-p-1] & x[N-p-2] \\
\vdots & & & & & \\
0 & & & & x[N-1] & x[N-2] \\
0 & & & & 0 & x[N-1]
\end{bmatrix}
\begin{bmatrix}
1 \\ \hat{a}[1] \\ \hat{a}[2] \\ \vdots \\ \hat{a}[p-1] \\ \hat{a}[p]
\end{bmatrix}
=
\begin{bmatrix}
e[0] \\ e[1] \\ \vdots \\ \\ e[N+p-1]
\end{bmatrix}
$$

Only $N-p$ of the $N+p$ noise terms $e[k]$ are computed without zero-padding. When N is small, this can be a problem.

Covariance Method

Minimize

$$\hat{\rho} = \frac{1}{N-p} \sum_{n=p}^{N-1} \left| x[n] + \sum_{k=1}^{p} a[k]\, x[n-k] \right|^2$$

Note: no zero-padding is necessary

Taking the derivative w.r.t. $a[\ell]$ yields

$$\frac{\partial \hat{\rho}}{\partial a[\ell]} = \frac{1}{N-p} \sum_{n=p}^{N-1} \left(x[n] + \sum_{k=1}^{p} a[k]\, x[n-k] \right) x[n-\ell]$$

Define

$$c[k,\ell] = \frac{1}{N-p} \sum_{n=p}^{N-1} x[n-k]\, x[n-\ell]$$

Setting the derivative to zero yields:

$$\underbrace{\begin{bmatrix} c[1,1] & c[1,2] & \cdots & c[1,p] \\ c[2,1] & c[2,2] & \cdots & c[2,p] \\ \vdots & \vdots & \ddots & \vdots \\ c[p,1] & c[p,2] & \cdots & c[p,p] \end{bmatrix}}_{\hat{C}_p} \begin{bmatrix} a[1] \\ a[2] \\ \vdots \\ a[p] \end{bmatrix} = - \begin{bmatrix} c[1,0] \\ c[2,0] \\ \vdots \\ c[p,0] \end{bmatrix}$$

Remarks

- $\hat{C}_p$ is generally symmetric/Hermitian and positive definite (good exercise)
- $\hat{C}_p$ is <u>not</u> Toeplitz
  - $\Longrightarrow$ Can't use Levinson alg.
  - $\Longrightarrow$ AR model might be unstable (poles outside unit circle)
- An alternate algorithm of order $O(p^2)$ has been derived by Morf et. al (1977)

We may view $c[i,j]$ as an estimate of $\gamma[i-j]$:

Example $\quad p = 3, \quad N = 6$

$$\begin{bmatrix} x_2^2 + x_3^2 + x_4^2 & x_2x_1 + x_3x_2 + x_4x_3 & x_2x_0 + x_3x_1 + x_4x_2 \\ x_1x_2 + x_2x_3 + x_3x_4 & x_1^2 + x_2^2 + x_3^2 & x_1x_0 + x_2x_1 + x_3x_2 \\ x_0x_2 + x_1x_3 + x_2x_4 & x_0x_1 + x_1x_2 + x_2x_3 & x_0^2 + x_1^2 + x_2^2 \end{bmatrix}$$

Estimates of $\gamma[0]$

Reflections on matrices

Consider

$$X = \left[\begin{array}{cccccc}
x_0 & 0 & 0 & \cdots & 0 & 0 \\
x_1 & x_0 & 0 & \cdots & 0 & 0 \\
\vdots & & & & & \\
x_{p-1} & x_{p-2} & x_{p-3} & \cdots & x_0 & 0 \\
\hline
x_p & x_{p-1} & x_{p-2} & \cdots & x_1 & x_0 \\
x_{p+1} & x_p & x_{p-1} & \cdots & x_2 & x_1 \\
\vdots & & & & & \\
x_{N-1} & x_{N-2} & x_{N-3} & \cdots & x_{N-p} & x_{N-p-1} \\
\hline
0 & x_{N-1} & x_{N-2} & \cdots & x_{N-p-1} & x_{N-p-2} \\
\vdots & & & & & \\
0 & 0 & 0 & \cdots & x_{N-1} & x_{N-2} \\
0 & 0 & 0 & \cdots & 0 & x_{N-1}
\end{array}\right]$$

X is $(N+p) \times p$

$\tilde{X}$ is $(N-p) \times p$

Observe:

$$\hat{\Gamma}_p = \frac{1}{N} X^T X$$

$$\hat{C}_p = \frac{1}{N-p} \tilde{X}^T \tilde{X}$$

Modified Covariance Method

Forward prediction error:

$$e^f[n] = x[n] + \sum_{k=1}^{p} a^f[k]\, x[n-k]$$

Backward prediction error:

$$e^b[n] = x[n] + \sum_{k=1}^{p} a^b[k]\, x[n+k]$$

Exercise: $\quad a^b[k] = \left(a^f[k]\right)^*$

$\boxed{\text{IDEA}}$ Minimize the average of forward and backward prediction errors

$$\hat{\rho} = \frac{1}{2}\left(\hat{\rho}^f + \hat{\rho}^b\right)$$

where

$$\hat{\rho}^f = \frac{1}{N-p} \sum_{n=p}^{N-1} \left| x[n] + \sum_{k=1}^{p} a[k]\, x[n-k] \right|^2$$

$$\hat{\rho}^b = \frac{1}{N-p} \sum_{n=0}^{N-p-1} \left| x[n] + \sum_{k=1}^{p} a^*[k]\, x[n+k] \right|^2$$

Taking the derivative, equating with zero, and arranging in matrix form, we have:

$$\begin{bmatrix} c[1,1] & c[1,2] & \cdots & c[1,p] \\ c[2,1] & c[2,2] & \cdots & c[2,p] \\ \vdots & \vdots & \ddots & \vdots \\ c[p,1] & c[p,2] & \cdots & c[p,p] \end{bmatrix} \begin{bmatrix} \hat{a}[1] \\ \hat{a}[2] \\ \vdots \\ \hat{a}[p] \end{bmatrix} = - \begin{bmatrix} c[1,0] \\ c[2,0] \\ \vdots \\ c[p,0] \end{bmatrix}$$

where

$$c[k,l] = \frac{1}{2(N-p)} \left(\sum_{n=p}^{N-1} x^*[n-j]\, x[n-k] + \sum_{n=0}^{N-p-1} x[n+j]\, x^*[n+k] \right)$$

another estimate of $\gamma[k-l]$

Model Order Selection

Choose p correctly is critical:

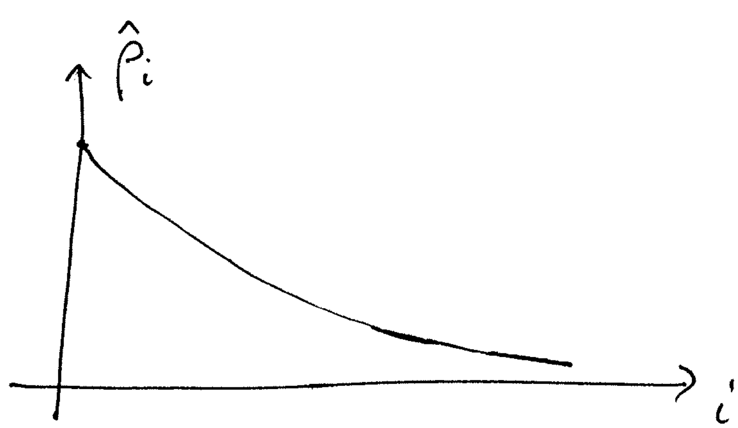
p too small $\Longrightarrow$ details of $P_{xx}(f)$ not captured

p too large $\Longrightarrow$ overfitting: model captures noise caused by limited data.

Consider the prediction error power (estimated)

$$\hat{\rho}_i = \widehat{Var}\{e_i[n]\}$$

based on the linear predictor of order i

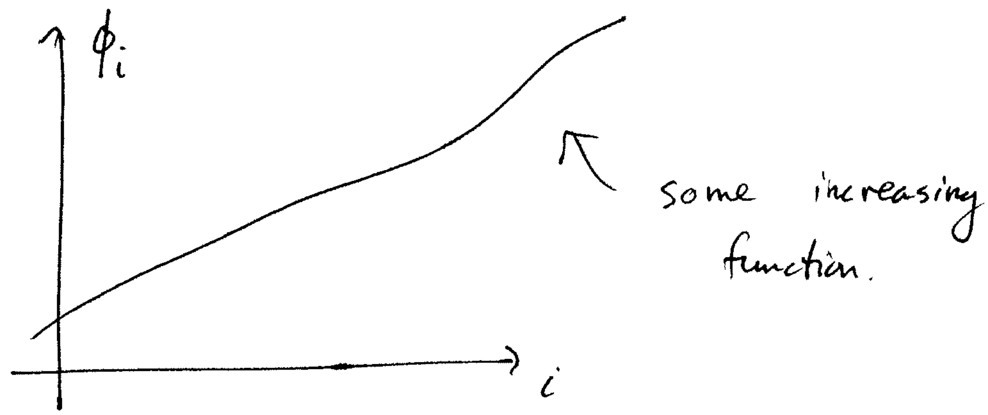


IDEA: Complexity regularization:

Minimize

$$\hat{\rho}_i + \phi_i$$

where



Examples:

Akaike information criterion (AIC):

$$N \ln \hat{\rho}_i + 2 \cdot i$$

Minimum description length (MDL)

$$N \ln \hat{\rho}_i + i \ln N$$

Theoretical methods tend to <u>underestimate</u> p in practice

Useful rule of thumb:

$$\frac{N}{3} \leq p \leq \frac{N}{2}$$

<u>Bottom line</u>: like much of spectral estimation, AR model order selection is <u>art</u> rather than <u>science</u>.