NTNU – Trondheim
Norwegian University of
Science and Technology

TTT4120 Digital Signal Processing
Fall 2018

Linear prediction

Prof. Stefan Werner
stefan.werner@ntnu.no
Office B329

Lecture in course book*

- Proakis, Manolakis Digital Signal Processing, 4th Ed.
  - 14.3.1 Forward linear prediction
  - 14.3.2 The Yule-Walker method for AR model parameters

- A comprehensive overview of topics treated in the lecture, see "Introdukjon til statistisk signalbehandling" on Blackboard

*Level of detail is defined by lectures and problem sets

Contents and learning outcomes

- How to find the AR parameters for a general process
- Linear prediction
- How many coefficients to choose? Model order estimation

Estimation in practice

- Only access to finite-length realization, $x[n]$, of process $X[n]$
  - True $\gamma_{XX}[l]$ must be estimated from $x[n] \Rightarrow \hat{\gamma}_{XX}[l]$
  - Parameter values computed using $\hat{\gamma}_{XX}[l]$ becomes parameter estimates $\{\hat{a}_k\} \Rightarrow$ Power spectrum estimate

$$\hat{\Gamma}_{XX}(f) = \frac{\hat{\sigma}_f^2}{|\hat{A}(f)|^2} = \frac{\hat{\sigma}_f^2}{\left|1+\sum_{k=1}^{p} \hat{a}_k e^{-j2\pi f k}\right|^2}$$

$$\boxed{\begin{array}{c} \gamma_{XX}[l] \longrightarrow \{a_k\} \longrightarrow \Gamma_{XX}(f) \\ \downarrow \text{(estimation)} \\ \hat{\gamma}_{XX}[l] \longrightarrow \{\hat{a}_k\} \longrightarrow \hat{\Gamma}_{XX}(f) \end{array}}$$

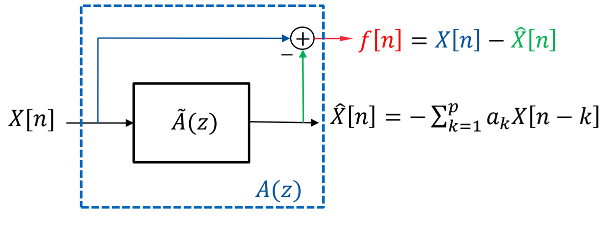
Estimation in practice…

- In practice process $X[n]$ may not be a true AR(p) process
  - How to choose parameters $\{\hat{a}_k\}$ to closely model $X[n]$ using an AR(p) process?
  - How do we measure closeness between model process and physical process?
- We will design pth-order linear predictor:
  - We observe/measure process $X[n]$
  - Store p prior values of $X[n]$, i.e., $\{X[n-1], \dots, X[n-p]\}$
  - Make linear combination of past values to estimate of $X[n]$

$$\hat{X}[n] = -\sum_{k=1}^{p} a_k X[n-k]$$

5

Linear prediction

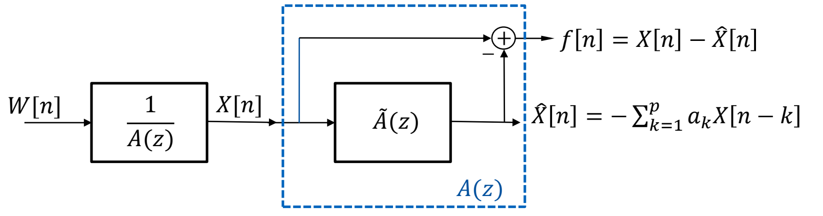


- Design a_k to match $X[n]$ as good as possible in some sense
  - We can compute the prediction error
  - Error $f[n]$ should be small
  - Find predictor coefficients that minimize mean-square error

$$\sigma_f^2 = E\left\{\left(X[n] - \hat{X}[n]\right)^2\right\} = E\left\{\left(X[n] + \sum_{k=1}^{p} a_k X[n-k]\right)^2\right\}$$

6

Linear prediction…



- If $X[n]$ is a true AR(p) process then $f[n] = W[n]$ whenever the prediction coefficients a_k match those of the AR(p) process
- In practice this assumption leads to an approximation

7

Linear prediction…

- Elaborate the MSE

$$\sigma_f^2 = E\left\{\left(X[n] - \hat{X}[n]\right)^2\right\} = E\left\{\left(X[n] + \sum_{k=1}^{p} a_k X[n-k]\right)^2\right\}$$

$$= E\{X^2[n] + 2\sum_{k=1}^{p} a_k X[n-k]X[n] + \sum_{l=1}^{p}\sum_{k=1}^{p} a_k a_l X[n-k]X[n-l]\}$$

$$= \gamma_{XX}[0] + 2\sum_{k=1}^{p} a_k \gamma_{XX}[k] + \sum_{l=1}^{p}\sum_{k=1}^{p} a_k a_l \gamma_{XX}[l-k]$$

- MSE is minimum if we choose a_k such that

$$\frac{d\sigma_f^2}{da_k} = 0, k = 1, 2, \dots, p$$

8

Linear prediction…

- Example: Find optimal predictor for $p = 1$, i.e., $\hat{X}[n] = -a_1 X[n-1]$

$$\sigma_f^2 = E\left\{\left(X[n] - \hat{X}[n]\right)^2\right\} = E\{(X[n] + a_1 X[n-1])^2\}$$
$$= \gamma_{XX}[0] + 2a_1\gamma_{XX}[1] + a_1^2\gamma_{XX}[0]$$
$$= \gamma_{XX}[0] - \frac{\gamma_{XX}^2[1]}{\gamma_{XX}[0]} + \gamma_{XX}[0]\left(a_1 + \frac{\gamma_{XX}[1]}{\gamma_{XX}[0]}\right)^2$$

- Prediction error variance minimized for value a_1 that gives $\frac{d\sigma_f^2}{da_1} = 0$:

$$\frac{d\sigma_f^2}{da_1} = 2\gamma_{XX}[1] + 2a_1\gamma_{XX}[0] = 0 \Rightarrow a_1 = -\frac{\gamma_{XX}[1]}{\gamma_{XX}[0]}$$

- Resulting prediction variance: $\sigma_f^2 = \gamma_{XX}[0] + a_1\gamma_{XX}[1]$

Linear prediction…

- In vector notation: $\sigma_f^2 = \gamma_{XX}[0] + 2\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\gamma}_{XX} + \boldsymbol{a}^{\mathrm{T}}\boldsymbol{\Gamma}_{XX}\boldsymbol{a}$

$$\boldsymbol{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}, \boldsymbol{\Gamma}_{XX} = \begin{bmatrix} \gamma_{XX}[0] & \gamma_{XX}[1] & \dots & \gamma_{XX}[p-1] \\ \gamma_{XX}[1] & \gamma_{XX}[0] & \dots & \gamma_{XX}[p-2] \\ \vdots & \ddots & \ddots & \vdots \\ \gamma_{XX}[p-1] & \gamma_{XX}[p-2] & \dots & \gamma_{XX}[0] \end{bmatrix}, \boldsymbol{\gamma}_{XX} = \begin{bmatrix} \gamma_{XX}[1] \\ \gamma_{XX}[2] \\ \vdots \\ \gamma_{XX}[p] \end{bmatrix}$$

- Set the gradient $\nabla_{\boldsymbol{a}}\sigma_f^2 = \boldsymbol{0}$, i.e.,

$$\nabla_{\boldsymbol{a}}\sigma_f^2 = \left[\frac{\partial\sigma_f^2}{\partial a_1} \quad \cdots \quad \frac{\partial\sigma_f^2}{\partial a_p}\right]^{\mathrm{T}} = [0 \quad \cdots \quad 0]^{\mathrm{T}}$$

Linear prediction…

- $\nabla_{\boldsymbol{a}}\sigma_f^2 = \mathbf{0}$:

$$\nabla_{\boldsymbol{a}}\sigma_f^2 = 2\boldsymbol{\gamma}_{XX} + 2\boldsymbol{\Gamma}_{XX}\boldsymbol{a} = \mathbf{0}$$
$$\Rightarrow \boldsymbol{a} = -\boldsymbol{\Gamma}_{XX}^{-1}\boldsymbol{\gamma}_{XX}$$

- Minimum MSE:

$$\sigma_f^2 = \gamma_{XX}[0] + 2\boldsymbol{a}^T\boldsymbol{\gamma}_{XX} + \boldsymbol{a}^T\boldsymbol{\Gamma}_{XX}\boldsymbol{a}$$
$$= \gamma_{XX}[0] + 2\boldsymbol{a}^T\boldsymbol{\gamma}_{XX} - \boldsymbol{a}^T\boldsymbol{\gamma}_{XX}$$
$$= \gamma_{XX}[0] + \boldsymbol{a}^T\boldsymbol{\gamma}_{XX} = \gamma_{XX}[0] + \sum_{k=1}^{p} a_k\gamma_{XX}[k]$$

- Same solution as we had for a pure AR(p) process

11

Linear prediction…

- Alternative approach by completing the square:

$$\sigma_f^2 = \gamma_{XX}[0] + 2\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\gamma}_{XX} + \boldsymbol{a}^{\mathrm{T}}\boldsymbol{\Gamma}_{XX}\boldsymbol{a}$$
$$= \gamma_{XX}[0] - \boldsymbol{\gamma}_{XX}^{\mathrm{T}}\boldsymbol{\Gamma}_{XX}^{-1}\boldsymbol{\gamma}_{XX} + (\boldsymbol{a} + \boldsymbol{\Gamma}_{XX}^{-1}\boldsymbol{\gamma}_{XX})^{\mathrm{T}}\boldsymbol{\Gamma}_{XX}(\boldsymbol{a} + \boldsymbol{\Gamma}_{XX}^{-1}\boldsymbol{\gamma}_{XX})$$

- The above holds true whenever $\boldsymbol{\Gamma}_{XX}$ is positive definite, i.e.,

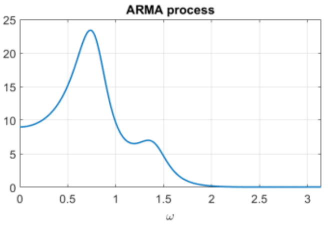
$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\Gamma}_{XX}\boldsymbol{x} > 0, \forall \boldsymbol{x} \neq \mathbf{0}$$

- Consequently, σ_f^2 is minimized when last term equals zero

$$\boldsymbol{a} = -\boldsymbol{\Gamma}_{XX}^{-1}\boldsymbol{\gamma}_{XX}$$

12

Linear prediction…



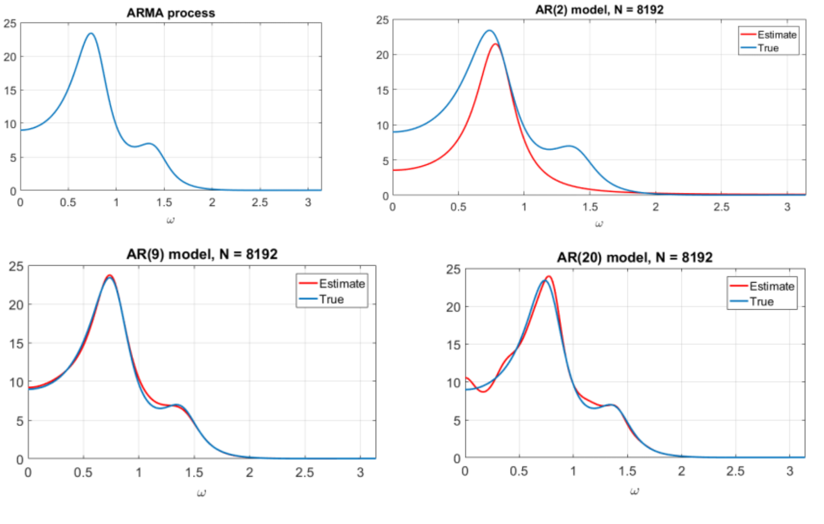
- Example: Estimate $\Gamma_{XX}(f)$ from a realization of an N-point ARMA process,

$$X[n] = -\sum_{k=1}^{p} a_k X[n-k] + \sum_{k=0}^{q} b_k W[n-k],\ W[n] \sim N(0, \sigma_W^2)$$

- Approximate with an AR(p) process and estimate model coefficients, $\hat{a}_k$, by minimizing prediction error variance, σ_f^2
  - What model order should I use?

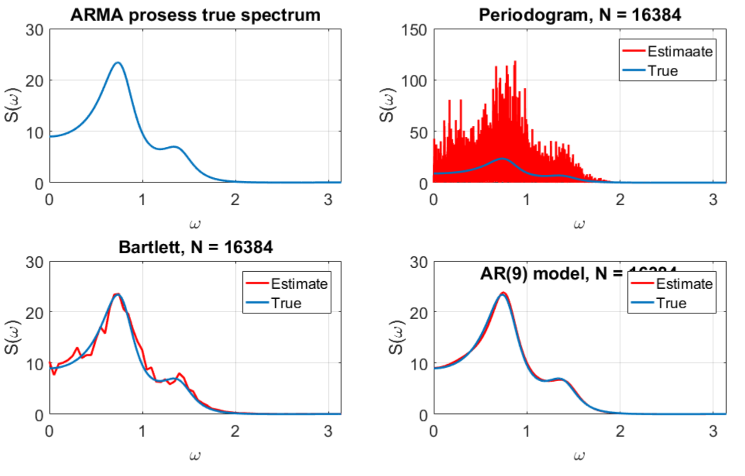
13

Linear prediction…



14

Linear prediction…



15

Determining model order *p*

- Model order not known when we shall model a physical process
- Proper choice of order p is necessary for good modelling capability
  - Too small p leads to smoothened spectrum
  - Too large p leads to spurious low-level peaks in the spectrum
- Prediction variance $\sigma_f^2(p)$ could be an indicator
  - Monotonically decreasing with p
  - Need to decide when changes are sufficiently small
  - Usually imprecise: in general no clear knee visible in plot $\sigma_f^2(p)$

16

Determining model order *p*

- Different criteria that penalizes high model order p:

$$\mathrm{FPE}(p) = \sigma_f^2(p)\frac{N+p+1}{N-p-1}$$

$$\mathrm{MDL}(p) = N\log\sigma_f^2(p) + p\log N$$

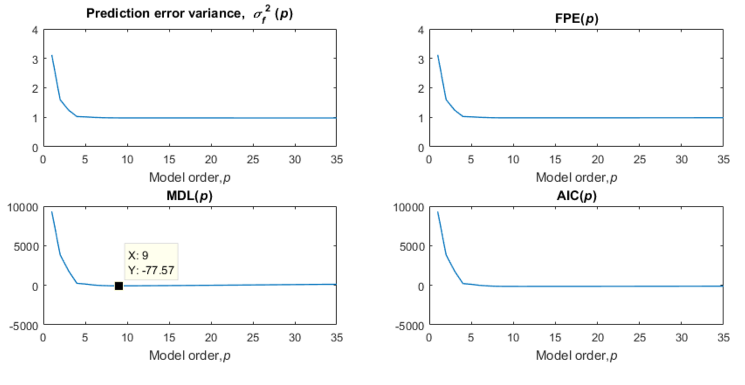
$$\mathrm{AIC}(p) = N\log\sigma_f^2(p) + 2p$$

17

Determining model order *p*…

- Example: Estimate $\Gamma_{XX}(f)$ from a realization of an ARMA process

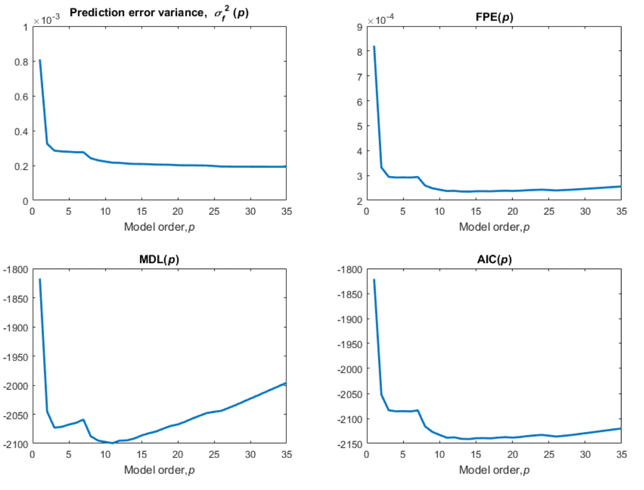
$$X[n] = -\sum_{k=1}^{p} a_k X[n-k] + \sum_{k=0}^{q} b_k W[n-k],\ W[n] \sim N(0,\sigma_W^2)$$



18

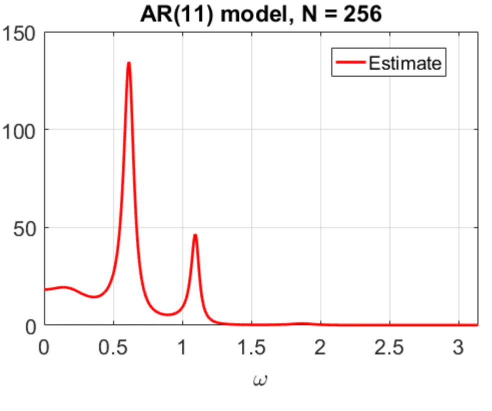
Determining model order *p*…

- Example: Vowel 'æ', $N = 256$:

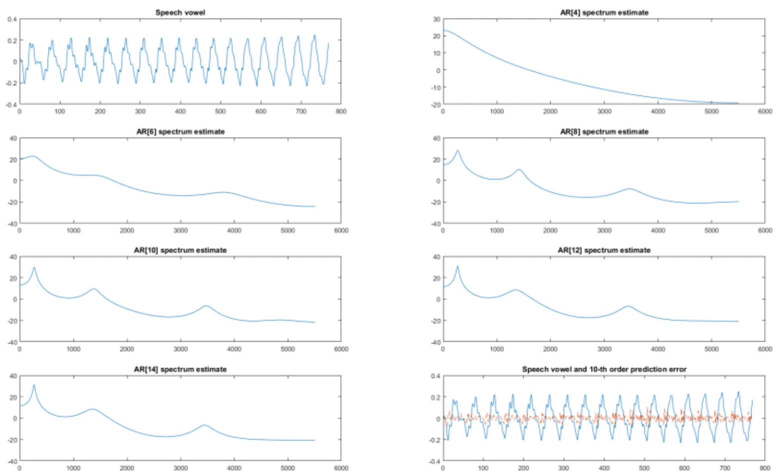


Determining model order *p*…

- Example: Vowel 'æ', $N = 256$:



Determining model order *p*…



Final notes on estimation in practice

- All methods looked at so far assume
  - Random processes to be stationary and ergodic
  - Random processes are autoregressive (AR)
- In practice, all physical processes of interest are nonstationary
  - Short-time stationarity: process varies slowly and within a certain time window, statistical properties are constant
  - Assume stationarity over M times and we need $N < M$ points
- Other methods for finding estimates
  - Usually lead to similar performance. Main differences are in the performance with few data points

Summary

- Today we discussed:
  - Linear prediction
  - Model order
- Next time:
  - FIR filter design