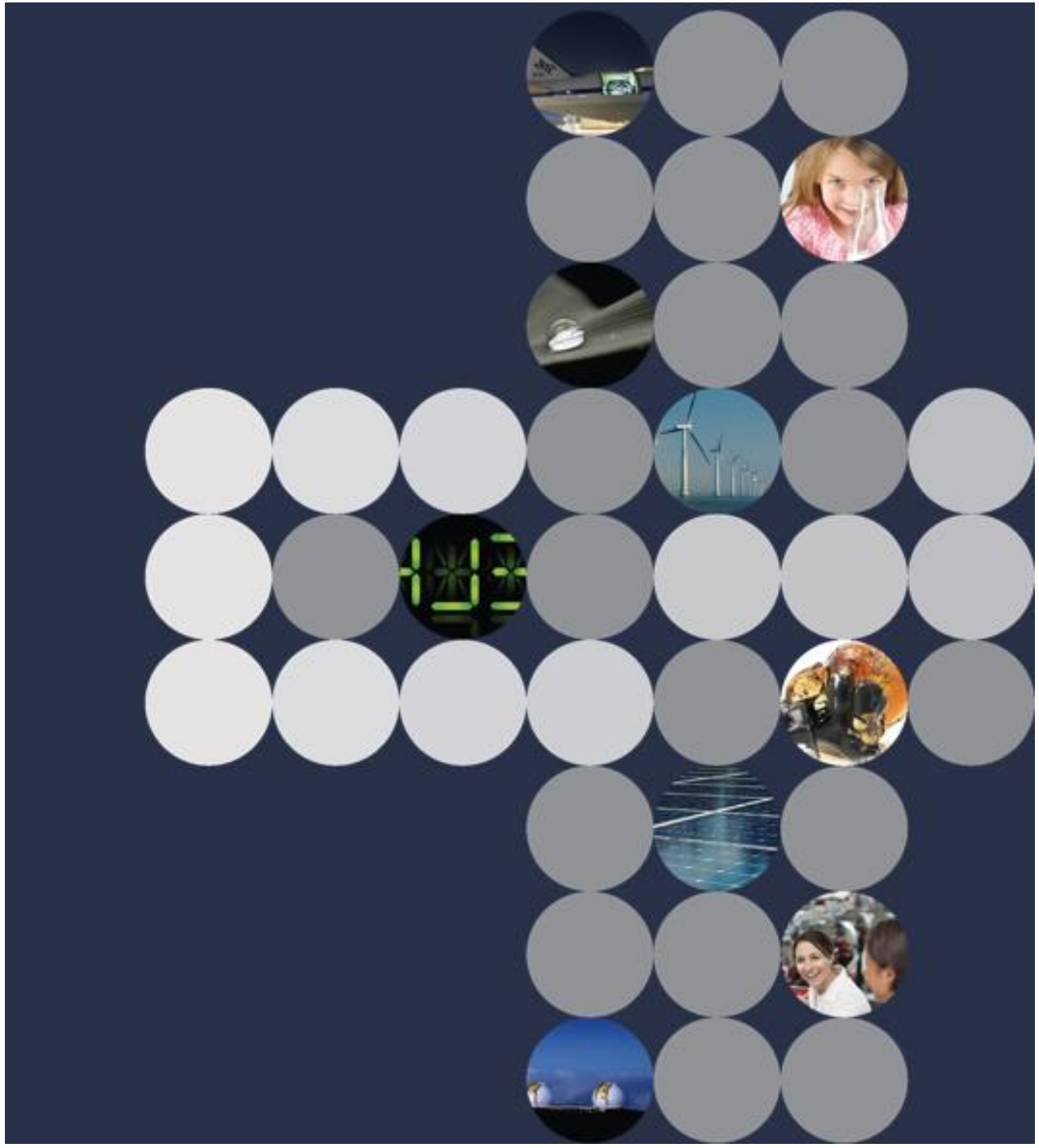
EE512 – Applied Biomedical Signal Processing

Linear Models I

João Jorge

CSEM Signal Processing Group



EPFL :: csem

Outline

Theory

- Intro/motivation
- Autoregressive (AR) models
  - Parameter estimation
  - Model design – order selection
  - Signal description – stability, structure

Practical session

- Based on Jupyter notebook
- See instructions in the notebook

Introduction

Why use models?

- Make use of **prior knowledge** (physics, maths, biology) of the problem
- More efficient and/or robust **estimation** of signal properties (e.g. power spectral density)
- Get a **compact representation** of the signal structure (i.e. the model parameters)

… why not?

- If the model is inappropriate, problems may follow

Introduction

What are linear models?

- Models that rely on linear combinations of variables and the model parameters

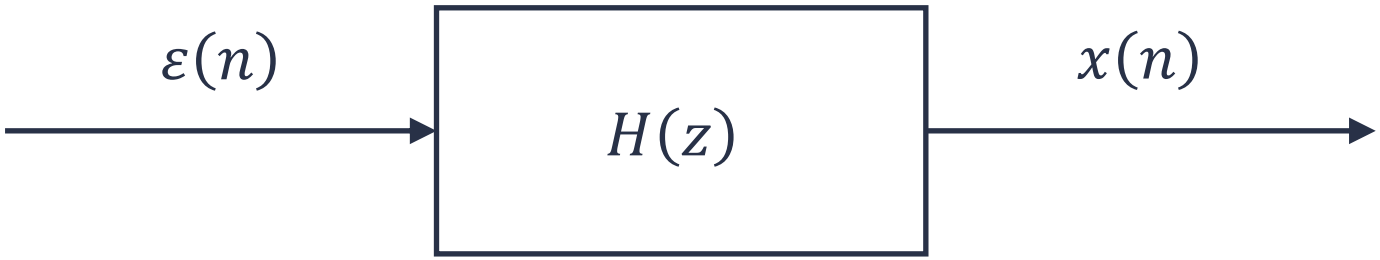
Why study linear models?

- Quite common in nature (e.g. superposition of EM fields)
- Simple, attractive properties
- Well-studied, plenty of algorithms for diverse needs

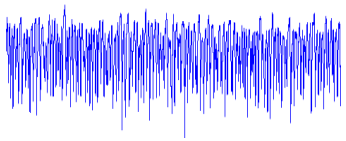
Autoregressive moving average (ARMA) models

General framework

- We consider a discrete white noise process $\varepsilon(n)$ that enters a causal, linear shift-invariant filter $H(z)$, resulting in a process $x(n)$:

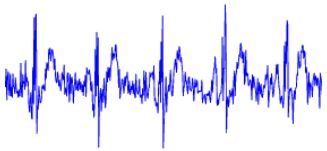


"Excitation signal"
Random



Filter
The model
Introduces temporal "structure"

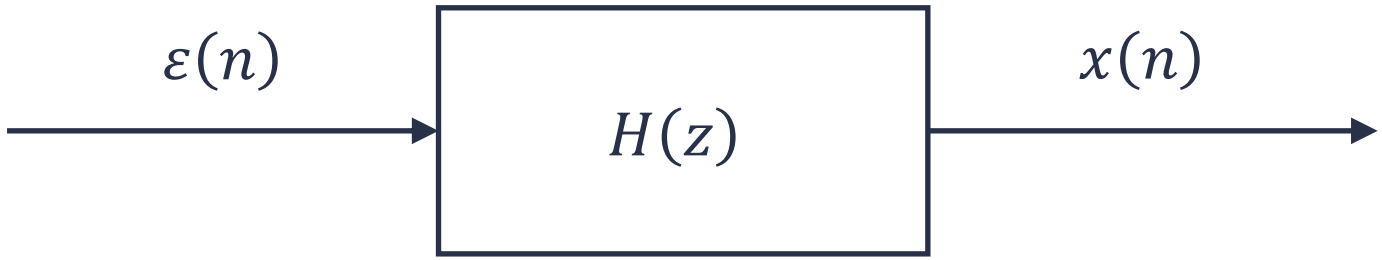
Measured signal
Has "structure"



Autoregressive moving average (ARMA) models

General framework

- We consider a discrete white noise process $\varepsilon(n)$ that enters a causal, linear shift-invariant filter $H(z)$, resulting in a process $x(n)$:



- Moreover, we define the filter $H(z)$ as a rational system function with p poles and q zeros:

$$H(z) = \frac{A(z)}{B(z)} = \frac{\sum_{k=0}^{q} b(k) z^{-k}}{1 + \sum_{k=1}^{p} a(k) z^{-k}}$$

Autoregressive moving average (ARMA) models

- Considering the relationship $X(z) = H(z)E(z) \Leftrightarrow X(z)B(z) = A(z)E(z)$, and returning to the time domain, we obtain:

$$x(n) + \sum_{l=1}^{p} a(l)x(n-l) = \sum_{l=0}^{q} b(l)\varepsilon(n-l)$$

- Or as a function of $x(n)$:

$$x(n) = \underbrace{\sum_{l=0}^{q} b(l)\varepsilon(n-l)}_{\text{Linear combination of present \& past inputs } \varepsilon} - \underbrace{\sum_{l=1}^{p} a(l)x(n-l)}_{\text{Linear combination of past outputs } x}$$

Autoregressive moving average (ARMA) models

- The ARMA model can be simplified when appropriate:

Autoregressive moving average (ARMA)

$$x(n) = \sum_{l=0}^{q} b(l)\varepsilon(n-l) - \sum_{l=1}^{p} a(l)x(n-l)$$

$a(l) = 0$ $\qquad$ $b(l) = \delta(l)$

Moving average (MA)

$$x(n) = \sum_{l=0}^{q} b(l)\varepsilon(n-l)$$

Autoregressive (AR)

$$x(n) = \varepsilon(n) - \sum_{l=1}^{p} a(l)x(n-l)$$

Autoregressive moving average (ARMA) models

- The ARMA model can be simplified when appropriate:

Autoregressive moving average (ARMA)

$$x(n) = \sum_{l=0}^{q} b(l)\varepsilon(n-l) - \sum_{l=1}^{p} a(l)x(n-l)$$

$a(l) = 0$ $\qquad$ $b(l) = \delta(l)$

Moving average (MA)

$$x(n) = \sum_{l=0}^{q} b(l)\varepsilon(n-l)$$

Autoregressive (AR)

$$x(n) = \varepsilon(n) - \sum_{l=1}^{p} a(l)x(n-l)$$

Autoregressive (AR) models

Why use AR models?

- Practical experience shows good performance in diverse situations;
- Well suited to highlight oscillations in a signal;
- Easy estimation of the AR parameters;
- Direct link with linear signal prediction.

Autoregressive (AR) models

In general:

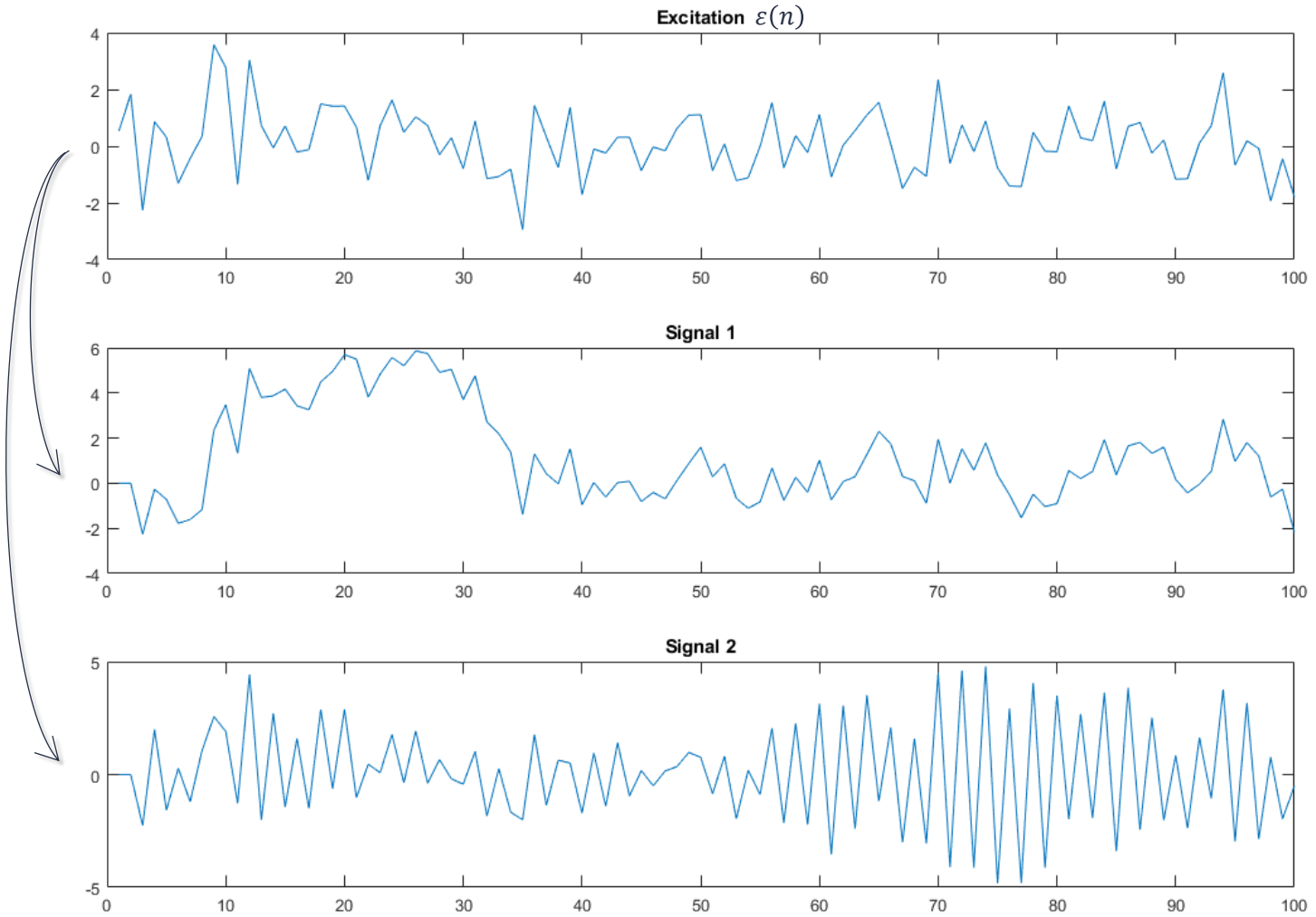
$$x(n) = -\sum_{l=1}^{p} a_l x(n-l) + \varepsilon(n)$$

Example 1

$$x(n) = 0.5\, x(n-1) + 0.4\, x(n-2) + \varepsilon(n)$$

Example 2

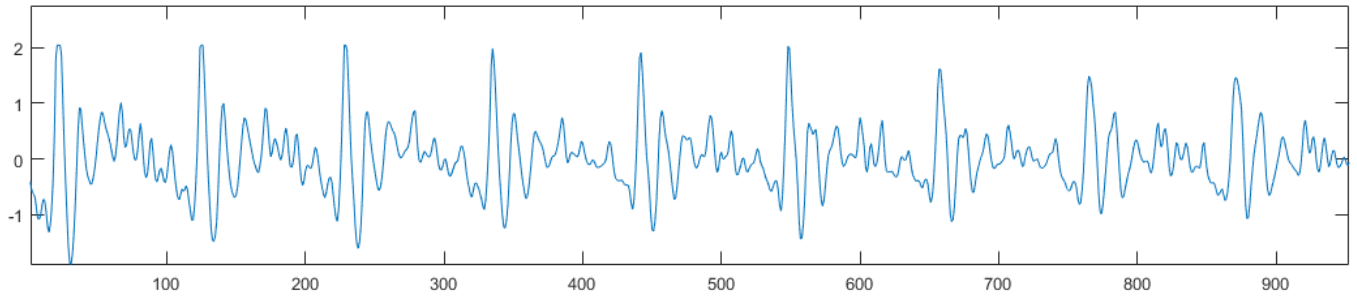
$$x(n) = -0.5\, x(n-1) + 0.4\, x(n-2) + \varepsilon(n)$$



Autoregressive (AR) models

- Typically, in practice, we go in the "opposite direction":

Measure signal as input



Model approximation

$$x(n) = -\sum_{l=1}^{p} a_l x(n-l) + \varepsilon(n)$$

Useful information / descriptions

Classification, decoding

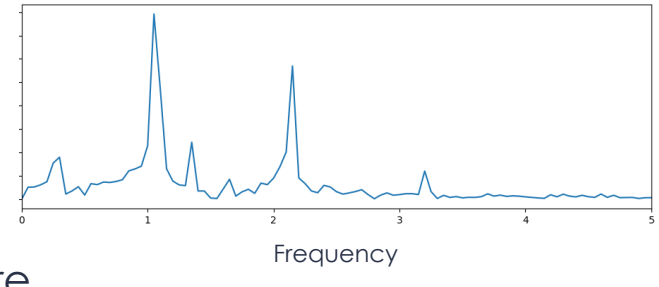
$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}$ Signal 1 $\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}$ Signal 4 $\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}$ Signal 7

Normal

$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}$ Signal 2 $\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}$ Signal 3 $\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}$ Signal 5

Pathological

Spectral properties



and more

EPFL :: csem

AR models: Yule-Walker equations

How can we estimate the parameters?

- An important relationship governing AR models is expressed in the so-called Yule-Walker equations.
- We start by noting that:
  - The **autocorrelation** of a signal x is defined as: $R_{xx}(n,k) = E\{x(n)x(n-k)\}$
  - In a **stationary signal**, $R_{xx}(n,k)$ does not depend on n : $R_{xx}(n,k) \rightarrow R_{xx}(k)$
  - In an AR process, since ε is stationary and $H(z)$ is time-invariant, then x is also stationary
  - For now we assume we know $\boldsymbol{R_{xx}(k)}$ and the desired model order $\boldsymbol{p}$

AR models: Yule-Walker equations

- Starting from the base relationship: $x(n) = -\sum_{l=1}^{p} a_l x(n-l) + \varepsilon(n)$
- Then, with $k \geq 0$:

$$R_{xx}(k) = E\{x(n)x(n-k)\}$$

$$= -\sum_{l=1}^{p} a_l E\{x(n-l)x(n-k)\} + E\{\varepsilon(n)x(n-k)\}$$

$$= -\sum_{l=1}^{p} a_l R_{xx}(k-l) + \underbrace{E\{\varepsilon(n)x(n-k)\}}$$

What can we say about this term?

AR models: Yule-Walker equations

- Considering the transfer function:

$$x(n) = \sum_{j=0}^{\infty} g(j)\,\varepsilon(n-j)$$

- Then, for the previous term:

$$E\{\varepsilon(n)x(n-k)\} = \sum_{j=0}^{\infty} g(j)E\{\varepsilon(n)\varepsilon(n-k-j)\}$$

$$= \sum_{j=0}^{\infty} g(j)R_{\varepsilon\varepsilon}(k+j)$$

$$= \begin{cases} g(0)\sigma_\varepsilon^2 & if\ k=0 \\ 0 & otherwise \end{cases}$$

Note that with:

$$G(z) = \sum_{j=0}^{\infty} g(j)z^{-k}$$

Then:

$$g(0) = \lim_{z\to\infty} G(z)$$

$$= \lim_{z\to\infty} \frac{1}{1+\sum_{l=1}^{p} a(l)z^{-l}} = 1$$

AR models: Yule-Walker equations

- We then obtain the **Yule-Walker equations**:

$$R_{xx}(k) = \begin{cases} -\sum_{l=1}^{p} a_l R_{xx}(k-l) & \text{for } k > 0 \\ -\sum_{l=1}^{p} a_l R_{xx}(k-l) + \sigma_\varepsilon^2 & \text{for } k = 0 \end{cases}$$

- These linear equations relate the signal autocorrelation R_{xx} with the AR model parameters a_l ;
- If R_{xx} is known, the parameters $\boldsymbol{a_l}$ can be estimated with only p equations;
- Once the a_l are estimated, the equation for $k = 0$ will give $\boldsymbol{\sigma_\varepsilon^2}$.

AR models: Yule-Walker equations

- The p first Yule-Walker equations ($k = 1, \dots, p$) can be expressed in matrix form (benefitting from the symmetry of R_{xx}) as:

$$\underbrace{\begin{bmatrix} R_{xx}(0) & R_{xx}(1) & \cdots & R_{xx}(p-1) \\ R_{xx}(1) & R_{xx}(0) & \cdots & R_{xx}(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ R_{xx}(p-1) & R_{xx}(p-2) & \cdots & R_{xx}(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = - \begin{bmatrix} R_{xx}(1) \\ R_{xx}(2) \\ \vdots \\ R_{xx}(p) \end{bmatrix}}_{R\boldsymbol{a} = -\boldsymbol{b}}$$

- The matrix R is the **autocorrelation matrix** of x:
  - In practice, usually invertible ⇨ allows us to solve for $\boldsymbol{a}$;
  - Hermitian/symmetric and Toeplitz (same value along each diagonal) matrix;

AR models: Yule-Walker equations

- A simple example: $\boldsymbol{p = 1}$ **model**

Y.W. for $k = 1$: $\quad [R_{xx}(0)][a_1] = -[R_{xx}(1)] \quad \Leftrightarrow \quad a_1 = -\dfrac{R_{xx}(1)}{R_{xx}(0)}$

Y.W. for $k = 0$: $\quad R_{xx}(0) = -a_1 R_{xx}(1) + \sigma_\varepsilon^2 \quad \Leftrightarrow \quad \sigma_\varepsilon^2 = (1 - a_1^2) R_{xx}(0)$

Thus giving a_1 and σ_ε^2.

AR models: Levinson-Durbin algorithm

In general, how to solve the Yule-Walker equations for a?

- The Toeplitz property of R allows applying the **Levinson-Durbin algorithm**;
- It computes the AR parameters recursively for increasing order p.
- We denote by a_{pi} the i^{th} coefficient, and by σ_p^2 the variance of ε for the p-order model.

$$\begin{bmatrix} R_{xx}(0) & R_{xx}(1) & \cdots & R_{xx}(p-1) \\ R_{xx}(1) & R_{xx}(0) & \cdots & R_{xx}(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ R_{xx}(p-1) & R_{xx}(p-2) & \cdots & R_{xx}(0) \end{bmatrix} \begin{bmatrix} a_{p1} \\ a_{p2} \\ \vdots \\ a_{pp} \end{bmatrix} = - \begin{bmatrix} R_{xx}(1) \\ R_{xx}(2) \\ \vdots \\ R_{xx}(p) \end{bmatrix}$$

AR models: Levinson-Durbin algorithm

- **The algorithm:**

1. Initialize with the 1st-order model:
   a) $a_{11} = -R_{xx}(1)/R_{xx}(0)$
   b) $\sigma_1^2 = (1 - a_{11}^2)R_{xx}(1)$
2. For increasing order $j = 2, 3, \ldots, p$:
   a) $\gamma_j = -\left(R_{xx}(j) + \sum_{l=1}^{j-1} a_{j-1,l} R_{xx}(j-l)\right)/\sigma_{j-1}^2$
   b) For $l = 1, 2, \ldots, j-1 : a_{j,l} = a_{j-1,l} + \gamma_j a_{j-1,j-l}$
   c) $a_{j,j} = \gamma_j$
   d) $\sigma_j^2 = \sigma_{j-1}^2(1 - \gamma_j^2)$

By the end we obtain the $\boldsymbol{a}_p$ parameters and the σ_p^2 value.

A derivation can be found in:
Hayes, *Statistical Digital Signal Processing and Modeling*, Wiley, 1996, chapter 5.2

AR models: Levinson-Durbin algorithm

Notes:

- At each iteration (order j), the estimated parameters are exactly those that would be obtained directly from Y.W. for that order;
- It can be shown that $|\gamma_j| \leq 1$ for a stable signal; as a consequence, $\sigma_j^2 \leq \sigma_{j-1}^2$ – the excitation variance decreases when j increases;
- If x is generated by an AR(m) process, then for $j > m$, we will obtain: $\gamma_j = 0,\ \sigma_j^2 = \sigma_m^2,\ a_{j,m+i} = 0$, i.e. it is useless to go beyond m.

AR models: autocorrelation

What about the autocorrelation $R_{xx}(k)$?

- In most applications, the **statistical autocorrelation** $R_{xx}(k)$ is unknown, and must be estimated from a sample realization of the process;
- Given a sample $x(n)$ with length N $(1 \leq n < N)$, we may estimate the **sample autocorrelation** $\hat{R}_{xx}(k)$:

$$\hat{R}_{xx}(k) = \frac{1}{N} \sum_{n=1}^{N-k} x(n)x(n+k)$$

Using this biased version has some convenient properties and guarantees the estimated AR model is stable.

Exam-like question

Suppose we have obtained 3 measurements of heart rate variation over time:

$$x_1 = 2,\ x_2 = 1,\ x_3 = -2 \qquad \text{(bpm)}$$

1. Calculate the sample autocorrelation (biased estimate) of x, $\hat{R}_{xx}(k)$, for $k = 0,1,2$, based on these samples.

2. Suppose we seek to model this signal with an AR(2) model. Write its Yule-Walker equations in terms of the AR model parameters, and with the numerical values obtained in **1**.

3. Solve the Yule-Walker equations to estimate all the AR model parameters.

Exam-like question

$$x_1 = 2,\ x_2 = 1,\ x_3 = -2 \qquad \text{(bpm)}$$

1. Calculate the sample autocorrelation (biased estimate) of x, $\hat{R}_{xx}(k)$, for $k = 0,1,2$, based on these samples.

The sample autocorrelation (biased form) is given by:

$$\hat{R}_{xx}(k) = \frac{1}{N}\sum_{n=1}^{N-k} x(n)x(n+k)$$

Therefore, for this sample, we have:

$$\hat{R}_{xx}(0) = \frac{1}{3}(4+1+4) = 3$$

$$\hat{R}_{xx}(1) = \frac{1}{3}(2-2) = 0$$

$$\hat{R}_{xx}(2) = -\frac{1}{3}4 = -\frac{4}{3}$$

Exam-like question

2. Suppose we seek to model this signal with an AR(2) model. Write its Yule-Walker equations in terms of the AR model parameters, and with the numerical values obtained in 1.

The Yule-Walker equations are given by:

$$R_{xx}(k) = \begin{cases} -\sum_{l=1}^{p} a_l R_{xx}(k-l) & \text{for } k > 0 \\ -\sum_{l=1}^{p} a_l R_{xx}(k-l) + \sigma_\varepsilon^2 & \text{for } k = 0 \end{cases}$$

For an AR(2) model, this becomes (note the minus sign was moved to the left side):

$k = 2$: $\quad -R_{xx}(2) = a_1 R_{xx}(1) + a_2 R_{xx}(0)$

$$\boldsymbol{\frac{4}{3} = 3a_2}$$

$k = 1$: $\quad -R_{xx}(1) = a_1 R_{xx}(0) + a_2 R_{xx}(-1)$

$$\leftrightarrow -R_{xx}(1) = a_1 R_{xx}(0) + a_2 R_{xx}(1)$$

$$\boldsymbol{0 = 3a_1}$$

$k = 0$: $\quad -R_{xx}(0) = a_1 R_{xx}(-1) + a_2 R_{xx}(-2) - \sigma_\varepsilon^2$

$$\leftrightarrow -R_{xx}(0) = a_1 R_{xx}(1) + a_2 R_{xx}(2) - \sigma_\varepsilon^2$$

$$\boldsymbol{-3 = -\frac{4}{3} a_2 - \sigma_\varepsilon^2}$$

Exam-like question

3. Solve the Yule-Walker equations to estimate all the AR model parameters.

From the previous answer, we have:

$k = 2$: $\frac{4}{3} = \mathbf{3}a_2$

$k = 1$: $\mathbf{0} = \mathbf{3}a_1$

$k = 0$: $-\mathbf{3} = -\frac{4}{3}a_2 - \sigma_\varepsilon^2$

- From the equation of $k = 1$ we see that $a_1 = 0$.
- From the eq. of $k = 2$ we can derive $a_2 = \frac{4}{9}$.
- From the eq. of $k = 0$, and having obtained a_2 above, we can derive:

$$\sigma_\varepsilon^2 = 3 - \frac{4}{3}a_2 = 3 - \frac{4}{3}\frac{4}{9} = 3 - \frac{16}{27} = \frac{81}{27} - \frac{16}{27} = \frac{65}{27}.$$

AR models: linear prediction

- AR model estimation can also be seen as a **linear prediction** problem, where we are trying to predict $x(n)$ based on p past samples:

$$\hat{x}(n) = -\sum_{l=1}^{p} a_l x(n-l) = -\boldsymbol{a}^T \boldsymbol{x}$$

with: $a = [a_1 \quad \dots \quad a_p]^T$ $x = [x(n-1) \quad \dots \quad x(n-p)]^T$

- What do we consider a good prediction? A common choice is based on the **mean squared error (MSE)**:

$$MSE = E\left\{\left(x(n) - \hat{x}(n)\right)^2\right\}$$

AR models: linear prediction

- Expanding this expression:

$$MSE = E\left\{\left(x(n) - \hat{x}(n)\right)^2\right\}$$

$$= E\{(x(n) + \boldsymbol{a}^T\boldsymbol{x})^2\}$$

$$= E\{x(n)^2 + 2x(n)\boldsymbol{a}^T\boldsymbol{x} + \boldsymbol{a}^T\boldsymbol{x}\boldsymbol{x}^T\boldsymbol{a}\}$$

$$= E\{x(n)^2\} + 2\boldsymbol{a}^T\underbrace{E\{x(n)\boldsymbol{x}\}}_{\text{Relates to the autocorrelation}} + \boldsymbol{a}^T\underbrace{E\{\boldsymbol{x}\boldsymbol{x}^T\}}_{\text{Also relates to the autocorrelation}}\boldsymbol{a}$$

$$= E\{x(n)^2\} + 2\boldsymbol{a}^T\boldsymbol{b} + \boldsymbol{a}^T R\boldsymbol{a}$$ (see definitions next)

AR models: linear prediction

$$\boldsymbol{b} = E\{x(n)\boldsymbol{x}\}$$

$$= E\left\{x(n)\begin{bmatrix} x(n-1) \\ \vdots \\ x(n-p) \end{bmatrix}\right\}$$

$$= \begin{bmatrix} E\{x(n)x(n-1)\} \\ \vdots \\ E\{x(n)x(n-p)\} \end{bmatrix}$$

$$= \begin{bmatrix} R_{xx}(1) \\ \vdots \\ R_{xx}(p) \end{bmatrix}$$

$$R = E\{\boldsymbol{x}\boldsymbol{x}^T\}$$

$$= E\left\{\begin{bmatrix} x(n-1)^2 & \cdots & x(n-1)x(n-p) \\ \vdots & \ddots & \vdots \\ x(n-p)x(n-1) & \cdots & x(n-p)^2 \end{bmatrix}\right\}$$

$$= \begin{bmatrix} R_{xx}(0) & R_{xx}(1) & \cdots & R_{xx}(p-1) \\ R_{xx}(1) & R_{xx}(0) & \cdots & R_{xx}(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ R_{xx}(p-1) & R_{xx}(p-2) & \cdots & R_{xx}(0) \end{bmatrix}$$

(which is the autocorrelation matrix)

AR models: linear prediction

$$MSE = E\{x(n)^2\} + 2\boldsymbol{a}^T\boldsymbol{b} + \boldsymbol{a}^T R\boldsymbol{a}$$

- It can be shown that R is a semi-positive definite matrix (i.e. $\boldsymbol{a}^T R\boldsymbol{a} \geq \boldsymbol{0}$), and typically it is positive definite;
- It can also be shown that this implies that the MSE has a **global minimum**;
- The global minimum can then be obtained by zero-ing the **gradient** of MSE with respect to $\boldsymbol{a}$.

AR models: linear prediction

- The gradient is: $\frac{\partial MSE}{\partial \boldsymbol{a}} = 2R\boldsymbol{a} + 2\boldsymbol{b}$

- Equaling to zero: $R\boldsymbol{a} = -\boldsymbol{b}$

- This result matches the Yule-Walker equations. Thus, seeking the MSE-optimal predictor is the same as estimating the AR model via Y.W.

- Also, it can be shown that the prediction error $e(n) = x(n) - \hat{x}(n)$ obeys $E\{e(n)\boldsymbol{x}\} = \boldsymbol{0}$. Intuitively, we can interpret that the prediction error at sample n does not share any information with past samples of x.

AR models: linear prediction

Linear prediction with a real, finite sample

- Say we have a sample $x(n)$ with length N $(0 \leq n < N-1)$;
- From MSE, we derive an appropriate criterion:

$$MSE = E\left\{\left(x(n) - \hat{x}(n)\right)^2\right\}$$

$$E = \sum_{k=k_0}^{k_1} \left(x(k) - \hat{x}(k)\right)^2 = \sum_{k=k_0}^{k_1} \left(x(k) + \sum_{l=1}^{p} a_l x(k-l)\right)^2$$

where k_0 and k_1 are chosen indices (more on that later)

AR models: linear prediction

- This expression is again quadratic with respect to the coefficients. We estimate the global minimum by zero-ing the gradient:

$$\frac{\partial E}{\partial a_j} = 2 \sum_{k=k_0}^{k_1} x(k-j) \left(x(k) + \sum_{l=1}^{p} a_l x(k-l) \right) = 0 \qquad \text{for } j = 1, \dots, p$$

- This yields the equation set:

$$\sum_{l=1}^{p} a_l \phi(j,l) = -\phi(j,0) \qquad \text{for } j = 1, \dots, p$$

with: $\phi(j,l) = \sum_{k=k_0}^{k_1} x(k-j)x(k-l)$

And, the system can be written in matrix form:

$$\Phi \boldsymbol{a} = -\boldsymbol{\psi}$$

$$[\Phi]_{j,l} = \phi(j,l)$$

$$[\psi]_j = \phi(j,0)$$

AR models: linear prediction

The choice of bounds k_0, k_1

- Two main approaches exist:
  - **"Autocorrelation"** : $k_0 = 0, \quad k_1 = N + p - 1$
  - **"Covariance"** : $k_0 = p, \quad k_1 = N - 1$

  (the designations are historical)

AR models: linear prediction

Autocorrelation approach $(k_0 = 0,\ k_1 = N + p - 1)$

- It is equivalent to considering the sample as part of a signal s of infinite duration: $\ldots \quad 0 \quad 0 \quad 0 \quad 0 \quad x_0 \quad x_1 \quad \ldots \quad x_{N-2} \quad x_{N-1} \quad 0 \quad 0 \quad 0 \quad 0 \quad \ldots$
- We can in fact consider k_0 and k_1 as expanding to $\pm\infty$, since the prediction error is zero for $k < 0$ and $k > N - 1 + p$. So we can write:

$$E = \sum_{k=-\infty}^{+\infty} \left(x(k) + \sum_{l=1}^{p} a_l x(k-l) \right)^2 \qquad \phi(j,l) = \sum_{k=-\infty}^{+\infty} x(k)x(k+j-l)$$

- Taking us back to Y.W.: $\sum_{l=1}^{p} a_l R_{ss}(j-l) = -R_{ss}(l)$

AR models: linear prediction

Advantages:

- In this case the matrix Φ is Hermitian Toeplitz, and one can use Levinson-Durbin to solve the system. Additionally, one gets a stable model.
- Fully equivalent to solving the Y.W. equations using the biased sample autocorrelation estimates.

Disadvantages:

- Boundary effects: prediction error will be larger for $k = 0, \dots, p-1$ (trying to predict signal samples using null samples), and for $k = N, \dots, N+p-1$ (predict null samples with signal samples), which will then affect the parameter estimation.

AR models: linear prediction

Covariance approach $(k_0 = p, \quad k_1 = N - 1)$

- It includes in the estimation only the prediction errors that are "fully supported", i.e. errors in predicting signal samples using signal samples.
- We then use:

$$\phi(j,l) = \sum_{k=p}^{N-1} x(k-j)x(k-l)$$

AR models: linear prediction

Advantages:

- The matrix Φ is symmetric and positive-definite.
- The AR estimation is more accurate than with the autocorrelation approach.

Disadvantages:

- The matrix Φ is not Toeplitz, and Levinson-Durbin cannot be applied.
- The AR model is not guaranteed to be stable.

AR models: linear prediction

Matrix formulation

- Provides an equivalent way to pursue the solution to the MSE criterion.
- For example for the covariance method, we would write:

$$\boldsymbol{e} = \boldsymbol{x} - \hat{\boldsymbol{x}}$$

$$\begin{bmatrix} e(p) \\ e(p+1) \\ \vdots \\ e(N-1) \end{bmatrix} = \begin{bmatrix} x(p) \\ x(p+1) \\ \vdots \\ x(N-1) \end{bmatrix} - \begin{bmatrix} x(p-1) & x(p-2) & \cdots & x(0) \\ x(p) & x(p-1) & \cdots & x(1) \\ \vdots & \vdots & \ddots & \vdots \\ x(N-2) & x(N-3) & \cdots & x(N-p-1) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}$$

$$\boldsymbol{e} \qquad \boldsymbol{c} \qquad X\boldsymbol{a}$$

AR models: linear prediction

- We then minimize $\boldsymbol{e}$ in the least-squares sense:

$$\|\boldsymbol{e}\|^2 = (X\boldsymbol{a} + \boldsymbol{c})^T(X\boldsymbol{a} + \boldsymbol{c}) = \boldsymbol{a}^T X^T X\boldsymbol{a} + 2\boldsymbol{c}^T X\boldsymbol{a} + \boldsymbol{c}^T\boldsymbol{c}$$

- To minimize, we search for the solution that nulls the gradient:

$$\frac{\partial\|\boldsymbol{e}\|^2}{\partial\boldsymbol{a}} = 0 \quad \Leftrightarrow \quad 2X^T X\boldsymbol{a} + 2X^T\boldsymbol{c} = \boldsymbol{0}$$

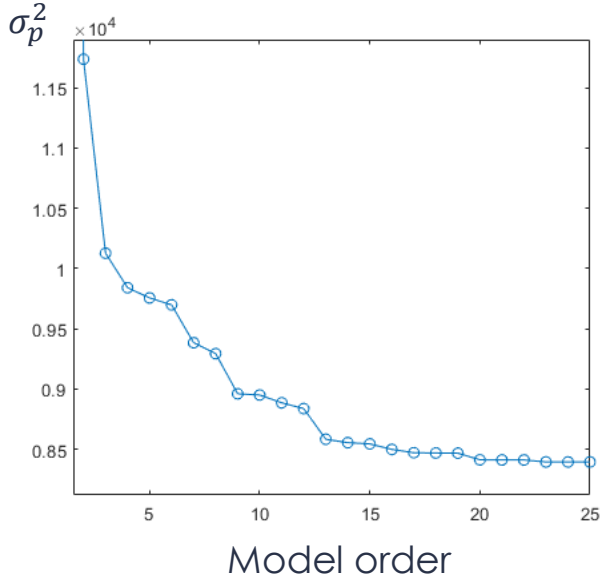
$$\Leftrightarrow \quad \boldsymbol{a} = -(X^T X)^{-1}X^T\boldsymbol{c}$$

If $X^T X$ is invertible.

AR models: order selection

How to choose the model order p?

- The larger the better? Not in practice:
  - Limited estimation accuracy of the autocorrelation due to limited sample size;
  - Computational resources are not infinite either.
- An empirical rule of thumb suggests that, with N samples, the order should not be larger than $N/3$.
- The rate of decrease of the excitation variance σ_p^2 with respect to p may help setting the stop point.



AR models: order selection

Selection criteria

Whiteness of the excitation term:

- Once the model is estimated, one can compute the excitation term:

$$\varepsilon(n) = x(n) + \sum_{l=1}^{p} a_l x(n-l)$$

- ε is supposed to be a white noise process. A whiteness test can therefore be applied: if it fails the test, the order can be increased.

AR models: order selection

Selection criteria

Several statistical criteria have been proposed as well:

- **Final prediction error (FPE):** $FPE(p) = \underbrace{\sigma_p^2}_{\text{Decreases with } \Uparrow p} \underbrace{\left(\frac{N+p+1}{N-p-1}\right)}_{\text{Increases with } \Uparrow p}$

- **Akaike information criterion (AIC):** $AIC(p) = \underbrace{N \ln\left(\sigma_p^2\right)}_{\text{Decreases with } \Uparrow p} + \underbrace{2(p+1)}_{\text{Increases with } \Uparrow p}$

AR models: order selection

Selection criteria

Several statistical criteria have been proposed as well:

- **Minimum description length (MDL):** $MDL(p) = \underbrace{N \ln\left(\sigma_p^2\right)}_{\text{Decreases with } \Uparrow p} + \underbrace{(p+1)\ln(N)}_{\text{Increases with } \Uparrow p}$

- MDL is probably the best criterion, especially with short signals.
- For purposes of spectral estimation, it is recommended to choose larger orders than what is given by the criteria.

AR models: sinusoids

- A sinusoid is a rather "organized" signal
- Sinusoids can be represented by an AR(2) process with no excitation:

$$x(n) = \sin(2\pi f n) = 2\cos(2\pi f)x(n-1) - x(n-2)$$

Using: $\sin(a+b) + \sin(a-b) = 2\sin(a)\cos(b)$

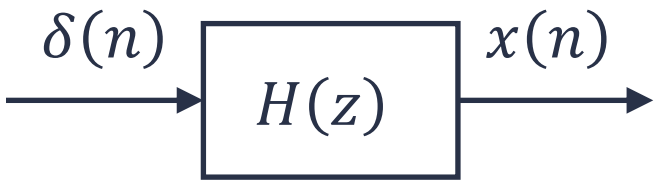
with $a = 2\pi f(n-1)$ and $b = 2\pi f$

- The denominator of the transfer function $H(z) = \frac{1}{1+\sum_{k=1}^{p} a(k)z^{-k}}$ is then:

$$1 - 2\cos(2\pi f)\, z^{-1} + z^{-2}$$

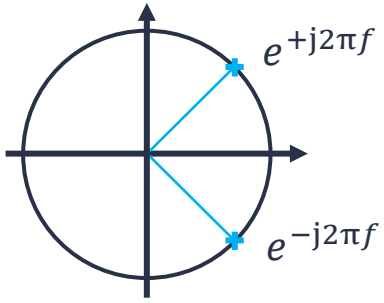
Which is equal to: $\left(1 - \underbrace{e^{+\mathrm{j}2\pi f}}_{\text{Pole 1}} z^{-1}\right)\left(1 - \underbrace{e^{-\mathrm{j}2\pi f}}_{\text{Pole 2}} z^{-1}\right)$

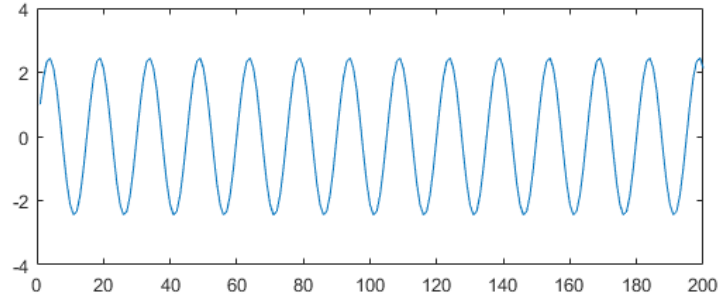
AR models: sinusoids



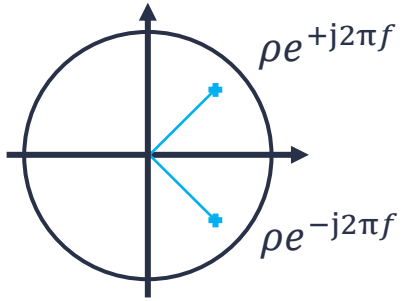
- As the poles get closer to the unit circle, the signal approaches a stationary sinusoid;
- It can be considered to become more "organized" in this sense;
- Analogous observations can be done for sums of sinusoids.

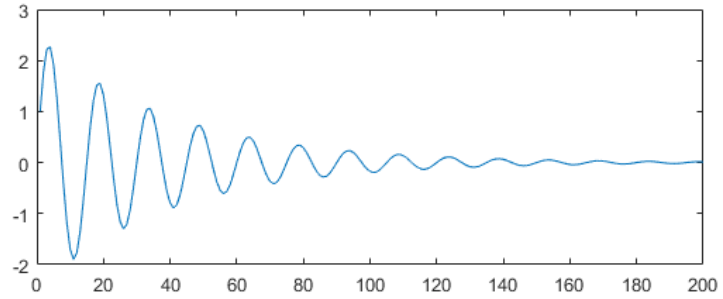
Neutrally stable



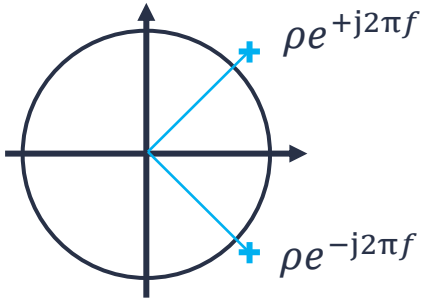


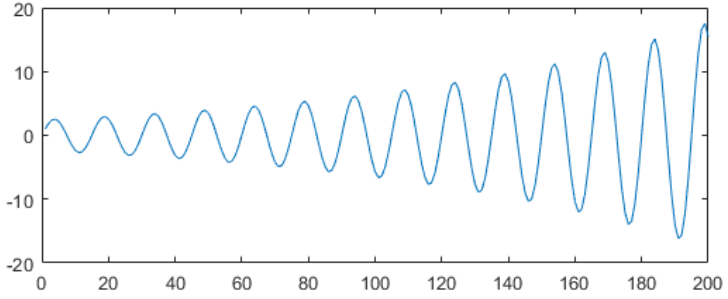
$\rho = 0.975$
Stable





$\rho = 1.010$
Unstable





Signal structure/organization

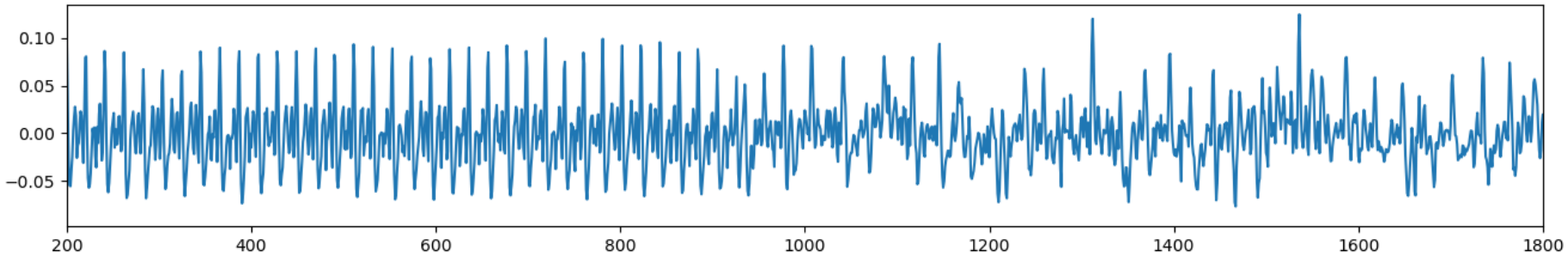
$$x(n) = -\sum_{l=1}^{p} a_l x(n-l) + \varepsilon(n)$$

Variance σ_x^2 contains everything

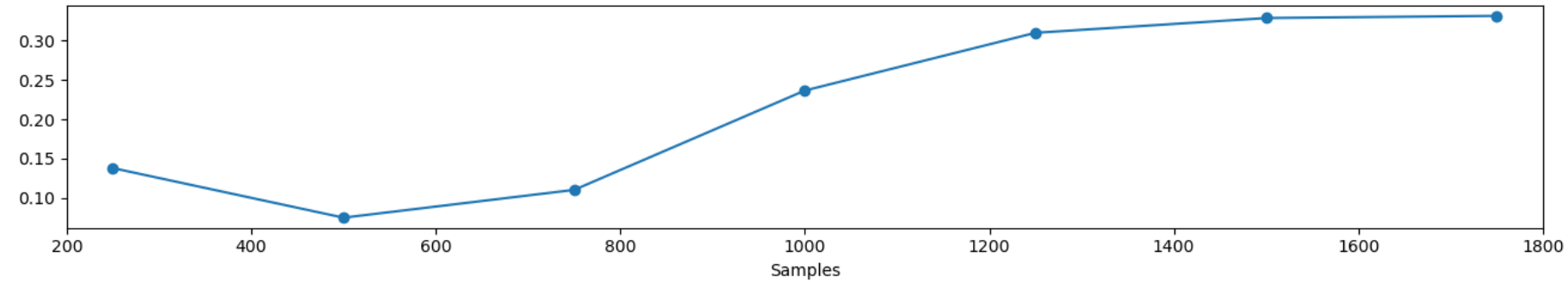
Variance σ_ε^2 contains the random contribution

- Once the model is estimated, both σ_x^2 and σ_ε^2 are known;
- The ratio $\sigma_\varepsilon^2/\sigma_x^2$ can inform on how "disorganized" or "unstructured" the signal is;
- It also depends on how "noisy" the measurement is.

Measured ECG signal



$\sigma_\varepsilon^2/\sigma_x^2$



Note: we estimated the AR model in a "sliding-window" approach – each dot corresponds to a segment (i.e. a model)

References

- M. H. Hayes, *Statistical Digital Signal Processing and Modeling*, Wiley, 1996.
- R. Shiavi, *Introduction to Applied Statistical Signal Analysis*, 2nd Ed., Academic Press 1999.
- S. M. Kay and S. L. Marple, *Spectrum analysis – A modern perspective*, Proc. IEEE, vol. 69, no. 11, Nov. 1981, pp. 1380-1419.