Lecture 7: 1-D Gaussian Autoregressive Models

Chapter 3.3

Charles A. Bouman

Purdue University

Wednesday, September 9, 2026

Outline

1. **Homogeneous models and filtering**
   Time invariance, Toeplitz and circulant matrices, infinite signals.
2. **Autoregressive models**
   White noise driven model, AR order, generating an AR process.
3. **Autocovariance and power spectrum**
   What the prediction filter says about the statistics of X_n.
4. **Fitting an AR model**
   ML estimate of the prediction filter and variance.

Reading: Chapter 3.3, *Foundations of Computational Imaging*.

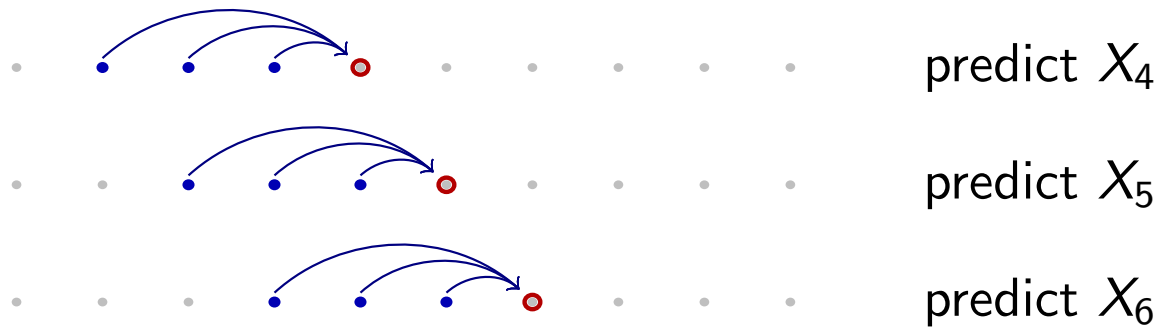
Today's Big Ideas

- Time-invariant model: causal prediction is convolution with a filter.
- An AR process is white noise passed through an IIR filter.
- The prediction filter and the prediction variance determine the autocovariance and the power spectrum of the random process.
- Maximum likelihood estimates the filter and variance from one signal.

Everything today assumes a natural order in time. Next lecture we impose an order on an image, where no natural one exists.

Time invariance and homogeneous models

Last lecture every time n could have its own coefficients $h_{n,i}$ and its own prediction variance σ_n^2. Today we assume they are the same at every time.

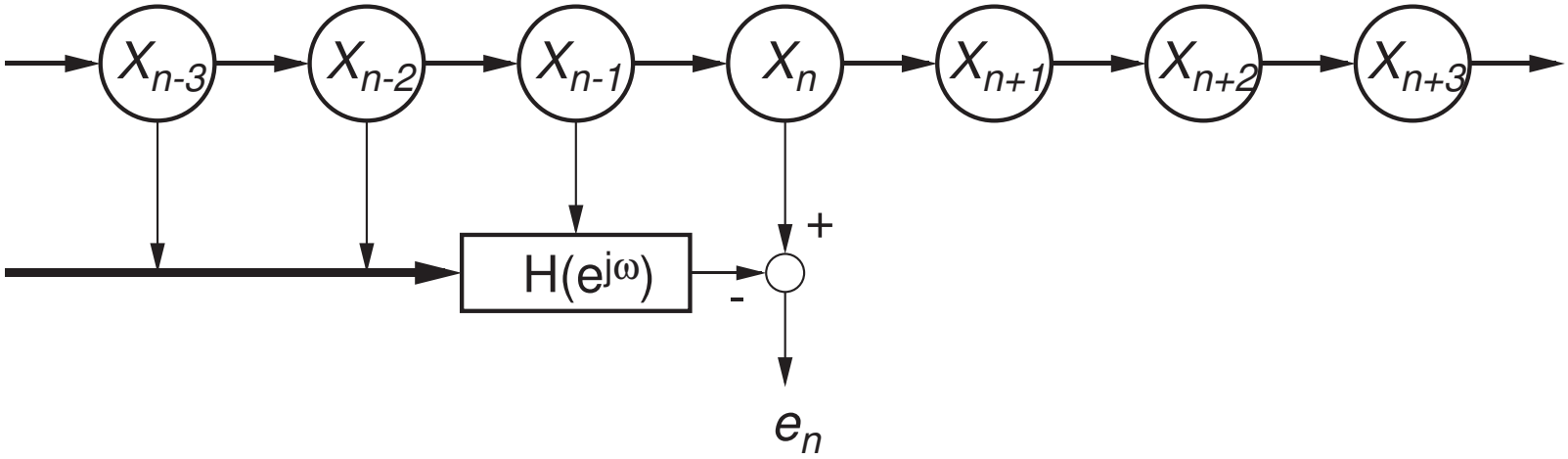


Homogeneous causal prediction model

A causal prediction model is **homogeneous** if the prediction filter and the prediction variance do not change with time.

The homogeneous causal predictor

Because the same coefficients are used at every time, prediction is convolution.



- The MMSE causal predictor must be a linear time-invariant filter, so the predictions are given by $\hat{X}_n = \sum_{i=1}^{N} h_{n-i} X_i$.
- h_n is the **causal prediction filter**; $h_n = 0$ for $n \leq 0$.
- Variance is constant, σ_C^2; errors are white if optimal.

Time invariance gives structured matrices

- A **Toeplitz** matrix satisfies $H_{i,j} = h_{i-j}$ for some function h_n, so every diagonal is constant. With a time-invariant prediction filter, H and A are Toeplitz.
- A **circulant** matrix is the special case $H_{i,j} = h_{(i-j) \bmod N}$, in which the filter wraps around at the boundary.

$$H = \begin{bmatrix} h_0 & h_{-1} & \cdots & h_{2-N} & h_{1-N} \\ h_1 & h_0 & \cdots & h_{3-N} & h_{2-N} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ h_{N-2} & h_{N-3} & \cdots & h_0 & h_{-1} \\ h_{N-1} & h_{N-2} & \cdots & h_1 & h_0 \end{bmatrix}$$

Toeplitz

$$H = \begin{bmatrix} h_0 & h_{N-1} & \cdots & h_2 & h_1 \\ h_1 & h_0 & \cdots & h_3 & h_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ h_{N-2} & h_{N-3} & \cdots & h_0 & h_{N-1} \\ h_{N-1} & h_{N-2} & \cdots & h_1 & h_0 \end{bmatrix}$$

Circulant

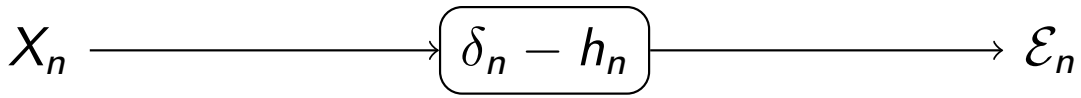
- Multiplying by a Toeplitz matrix is convolution with truncated boundaries. Multiplying by a circulant matrix is circular convolution, which the **discrete Fourier transform (DFT)** turns into multiplication.

Infinite extent: prediction becomes filtering

A simple way to get around problems with boundary conditions is to extend the random process X_n so that $n = -\infty, \cdots, -1, 0, 1, \cdots, \infty$. Then, in the standard notation of linear time-invariant systems,

$$\begin{aligned} \mathcal{E}_n &= X_n - \hat{X}_n \\ &= X_n - X_n * h_n \\ &= X_n * (\delta_n - h_n) \ , \end{aligned}$$

where $*$ denotes 1-D discrete-time convolution.



correlated process prediction-error filter white noise

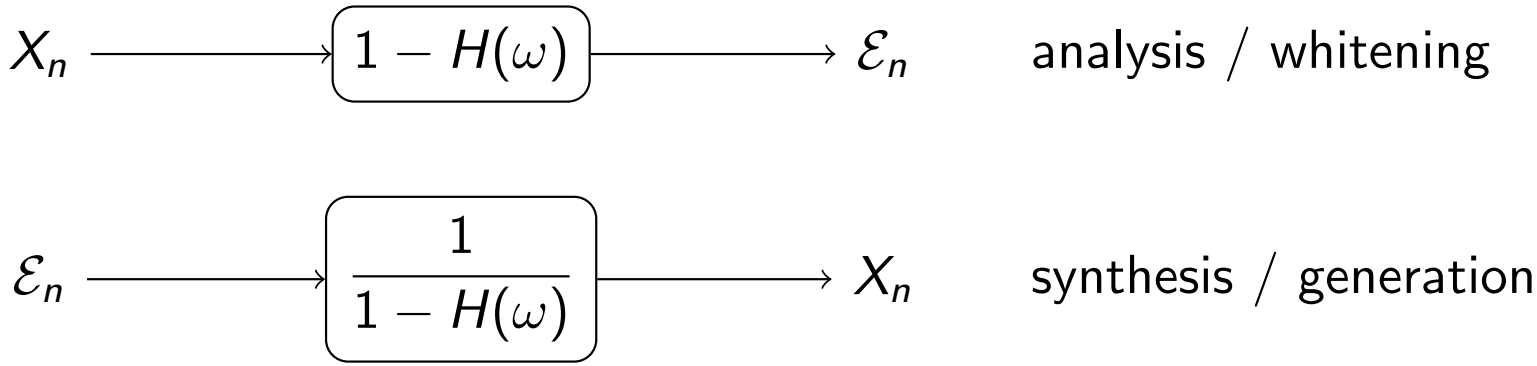
The prediction error filter $\delta_n - h_n$ is a **whitening filter**.

The white noise driven model

Turning things around, we may also write

$$X_n \;=\; \mathcal{E}_n + X_n * h_n \;=\; \mathcal{E}_n + \sum_{i=1}^{\infty} X_{n-i} h_i \;,$$

which is the form of an **infinite impulse response (IIR)** filter.

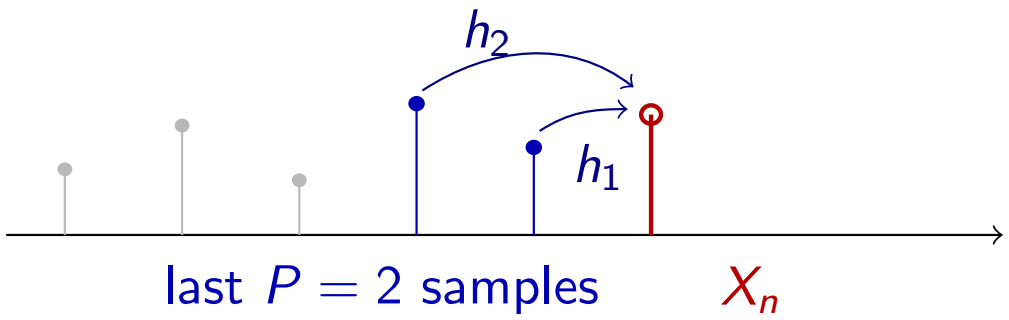


where $H(\omega)$ is the DTFT of the prediction filter h_n.

- This is the **white noise driven model** for the AR process. It is a **linear time-invariant (LTI) system** driven by the white noise $\mathcal{E}_n$.
- The two views are inverses of each other. Filter X_n to whiten it, or filter white noise to make X_n.

AR(P) processes and the prediction window

- In general, the MMSE prediction may use all past values of X_n.
- If $h_i = 0$ for $i > P$, then the prediction depends only on the last P observations, and we call X_n a P^{th}-order **AR process**.



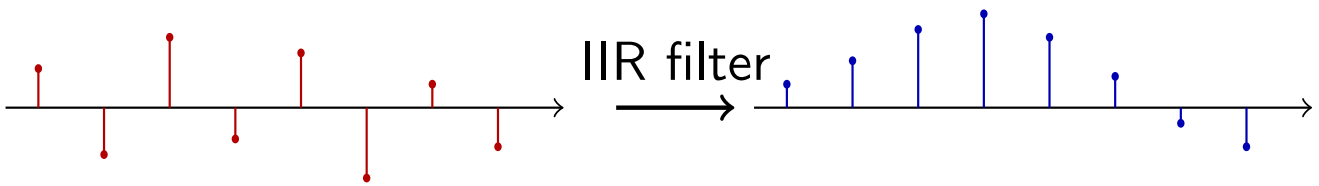
- AR(P): only the most recent P samples are needed to predict.

Generating an AR process

The white noise driven model provides an easy method for generating an AR process X_n.

Recipe

1. Draw i.i.d. Gaussian random variables, each $N(0, \sigma_C^2)$.
2. Filter them with the IIR filter $X_n = \mathcal{E}_n + \sum_{i=1}^{P} X_{n-i} h_i$.



$\mathcal{E}_n$: white noise $\qquad$ X_n: correlated AR process

- If the IIR filter is stable, then its output X_n is a stationary random process. The feedback in the filter is what creates the correlation.

The autocovariance of an AR process

Since the prediction errors $\mathcal{E}_n$ are i.i.d., their time autocovariance is

$$R_{\mathcal{E}}(i-j) = \mathbb{E}[\mathcal{E}_i \mathcal{E}_j] = \sigma_C^2 \delta_{i-j} \ .$$

Combining this with $\mathcal{E}_n = X_n * (\delta_n - h_n)$, the autocovariance of X_n obeys the important relationship

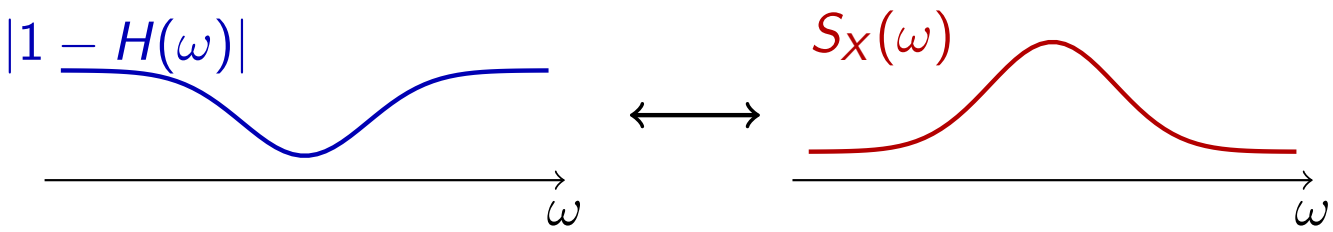
$$R_X(n) * (\delta_n - h_n) * (\delta_n - h_{-n}) = R_{\mathcal{E}}(n) = \sigma_C^2 \delta_n \ .$$

- Whitening in time is a statement about the autocovariance of X_n.
- Taking the DTFT of this relationship gives the power spectrum.

The power spectrum of an AR process

Taking the DTFT of the time autocovariance gives the power spectral density of the AR process:

$$S_X(\omega) = \frac{\sigma_C^2}{|1 - H(\omega)|^2} \ .$$



Where $|1 - H(\omega)|$ is small, the process has large power.

Why this is a complete model

From Lecture 5, a zero-mean stationary Gaussian process is completely specified by its autocovariance, and equivalently by its power spectral density. So $(h_n, \sigma_C^2) \Rightarrow R_X(n) \Leftrightarrow S_X(\omega) \Rightarrow p(x)$, and the pair (h_n, σ_C^2) is a complete model of the process.

Example 3.1: ML estimation of an AR model

Consider the P^{th}-order AR random process $X_1, \cdots, X_N$ with prediction variance σ_C^2 and prediction errors

$$\mathcal{E}_n = X_n - \sum_{i=1}^{P} h_i X_{n-i} \ .$$

To simplify notation, we assume $X_n = 0$ when $n \leq 0$, so that we do not need special indexing at the boundaries of the signal.

The task is to compute the joint ML estimate of the prediction filter h_n and the prediction variance σ_C^2.

From the density of the AR process,

$$p(x) = \frac{1}{(2\pi\sigma_C^2)^{N/2}} \exp\left\{ -\frac{1}{2\sigma_C^2} \sum_{n=1}^{N} \left(x_n - \sum_{i=1}^{P} h_i x_{n-i} \right)^2 \right\} \ .$$

Example 3.1: the log likelihood in vector form

The likelihood depends on the data only through the squared prediction errors. Define the parameter vector and statistic vector

$$\begin{aligned} h &= [h_1, h_2, \cdots, h_P]^t \ , \\ Z_n &= [X_{n-1}, X_{n-2}, \cdots, X_{n-P}]^t \ . \end{aligned}$$

Then the log likelihood of the observations X is

$$\begin{aligned} \log p(X) &= -\frac{N}{2} \log\left(2\pi\sigma_C^2\right) - \frac{1}{2\sigma_C^2} \sum_{n=1}^{N} \left(X_n - \sum_{i=1}^{P} h_i X_{n-i} \right)^2 \\ &= -\frac{N}{2} \log\left(2\pi\sigma_C^2\right) - \frac{1}{2\sigma_C^2} \sum_{n=1}^{N} \left(X_n^2 - 2h^t Z_n X_n + h^t Z_n Z_n^t h \right) \ . \end{aligned}$$

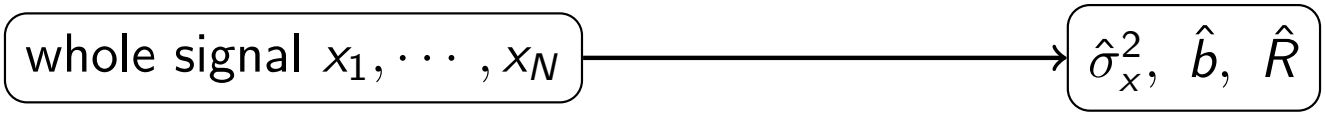
Example 3.1: three sample statistics

All the data needed for ML collapses to three sample statistics:

$$\log p(X) = -\frac{N}{2}\log\left(2\pi\sigma_C^2\right) - \frac{N}{2\sigma_C^2}\left(\hat{\sigma}_x^2 - 2h^t\hat{b} + h^t\hat{R}h\right) \ ,$$

where

$$\hat{\sigma}_x^2 \ = \ \frac{1}{N}\sum_{n=1}^{N} X_n^2 \ , \qquad \hat{b} \ = \ \frac{1}{N}\sum_{n=1}^{N} Z_n X_n \ , \qquad \hat{R} \ = \ \frac{1}{N}\sum_{n=1}^{N} Z_n Z_n^t \ .$$



- They estimate the variance of X_n, the covariance between Z_n and X_n, and the covariance of Z_n. These are sufficient statistics, in the sense of Lecture 3.

Example 3.1: ML estimate of the prediction filter

The ML predictor is obtained by solving a linear system. Take the gradient of the log likelihood with respect to h:

$$\begin{aligned} \nabla_h \log p(X) &= -\frac{N}{2\sigma_C^2} \nabla_h \left(h^t \hat{R} h - 2 h^t \hat{b} \right) \\ &= -\frac{N}{\sigma_C^2} \left(\hat{R} h - \hat{b} \right) \ , \end{aligned}$$

where we use the convention that the gradient is represented as a column vector.

Setting the gradient to zero, and assuming $\hat{R}$ is invertible, gives the ML estimate of the prediction filter:

$$\hat{h} = \hat{R}^{-1} \hat{b} \ .$$

Example 3.1: ML estimate of the prediction variance

Substituting the ML estimate $\hat{h} = \hat{R}^{-1}\hat{b}$ and differentiating with respect to σ_C^2:

$$\begin{aligned}\frac{d}{d\sigma_C^2}\log p(X) &= \frac{d}{d\sigma_C^2}\left[-\frac{N}{2}\log\left(2\pi\sigma_C^2\right) - \frac{N}{2\sigma_C^2}\left(\hat{\sigma}_x^2 - \hat{b}^t\hat{R}^{-1}\hat{b}\right)\right] \\ &= -\frac{N}{2}\left[\frac{1}{\sigma_C^2} - \frac{1}{\sigma_C^4}\left(\hat{\sigma}_x^2 - \hat{b}^t\hat{R}^{-1}\hat{b}\right)\right] .\end{aligned}$$

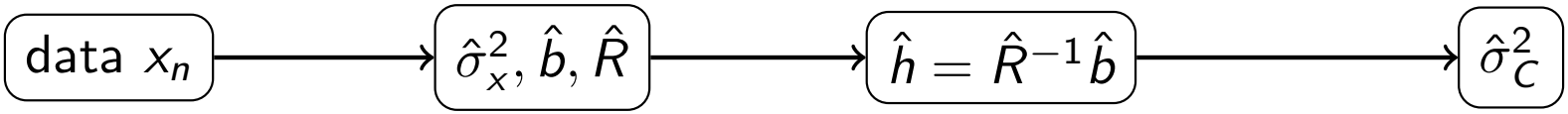
Setting the derivative to zero yields

$$\hat{\sigma}_C^2 = \hat{\sigma}_x^2 - \hat{b}^t\hat{R}^{-1}\hat{b} .$$

Example 3.1: the prediction variance as a mean square

A more intuitive form for $\hat{\sigma}_C^2$ is the average of the squared prediction errors of the fitted model,

$$\hat{\sigma}_C^2 = \frac{1}{N} \sum_{n=1}^{N} \left(X_n - \hat{h}^t Z_n \right)^2 \quad .$$



Fitting an AR model to data

Fitting takes one pass through the data and one linear solve. There is no iteration and no search. This is the ML estimation of Lecture 4, applied to a Gaussian density.

Summary

- A homogeneous model uses the same prediction filter and the same prediction variance at every time, so prediction becomes convolution and H becomes Toeplitz.
- Over an infinite-extent signal the prediction error is $\mathcal{E}_n = X_n * (\delta_n - h_n)$, so $\delta_n - h_n$ is a whitening filter.
- Reading the same equation the other way, $X_n = \mathcal{E}_n + \sum_{i=1}^{P} h_i X_{n-i}$, gives the white noise driven model. Filtering white noise through a stable IIR filter generates the process.
- The autocovariance satisfies $R_X(n) * (\delta_n - h_n) * (\delta_n - h_{-n}) = \sigma_C^2 \delta_n$, so the power spectrum is $S_X(\omega) = \sigma_C^2 / |1 - H(\omega)|^2$.
- The pair (h_n, σ_C^2) is therefore a complete model of the process.
- ML estimation from one observed signal uses the three sample statistics $\hat{\sigma}_x^2$, $\hat{b}$, and $\hat{R}$, and gives $\hat{h} = \hat{R}^{-1}\hat{b}$ and $\hat{\sigma}_C^2 = \frac{1}{N}\sum_n (X_n - \hat{h}^t Z_n)^2$.

Reading: Chapter 3.3. **Next lecture:** Chapter 3.4, 2-D Gaussian AR models.