Lecture 8: 2-D Gaussian AR Models

Chapter 3.4

Charles A. Bouman

Purdue University

Friday, September 11, 2026

Outline

1. **From 1-D to 2-D**
   The image lattice, raster ordering, and the causal prediction window.
2. **The same Gaussian AR machinery**
   Matrix form, block Toeplitz, autocovariance, power spectrum.
3. **Fitting a 2-D AR model**
   The same maximum-likelihood fit as in 1-D.
4. **The disadvantage of causality in images**
   Why raster order is artificial; why noncausal models come next.

Reading: Chapter 3.4, *Foundations of Computational Imaging*.

Today's Big Ideas

- An image has no natural causal order; raster order is usual.
- Raster order: all 1-D AR results hold, with 2-D convolution.
- A space-invariant filter gives a Toeplitz block Toeplitz matrix.
- The disadvantage is an asymmetric prediction window. The model uses a pixel above and to the right, but not the equally close pixel below and to the left.

That asymmetry has no basis in the image itself. It comes only from the ordering we chose, and it is the reason for Chapter 4.

From 1-D to 2-D

The analysis of 1-D AR models from the previous sections is easily generalized to a regular grid in two or more dimensions.

- Each point is $s = (s_1, s_2)$, with coordinates from one to N.
- When visualizing, s_1 is the row index and s_2 the column index.

The key issue

The key issue in generalizing the AR model to 2-D is the ordering of the points in the plane.

Raster ordering

Raster ordering scans the image left to right and top to bottom, in much the same way that one reads a page of English text. There is no truly natural ordering of the points in the plane, but this is the common choice:

1 →	2 →	3 →	4 →	5
6 →	7 →	8 →	9 →	10
11 →	12 →	13 →	14 →	15
16 →	17 →	18 →	19 →	20

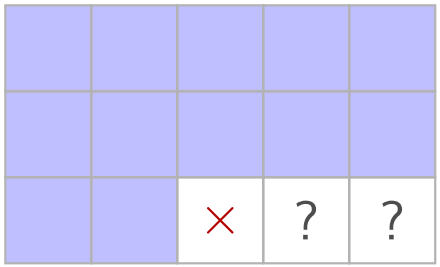
$$X = [X_{1,1}, \cdots, X_{1,N}, X_{2,1}, \cdots]^t$$

Raster ordering converts a 2-D image into the 1-D causal sequence required by Lecture 6. The causal prediction errors, $\mathcal{E}$, may be similarly ordered.

- Once pixels form a vector, every result of Sections 3.1–3.3 holds.

The 2-D causal prediction window

Which pixels count as "the past"? Under raster order, the previous rows and the pixels to the left:



shaded: available past (window W_P, $P = 2$)

$\times$: the pixel being predicted

?: future in raster order, so not allowed

Then the 2-D causal prediction error can again be written as

$$\mathcal{E}_s = X_s - \sum_{r \in W_P} h_r X_{s-r} \ ,$$

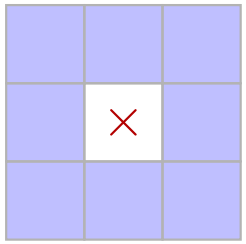
the same form as the 1-D case of Lecture 7, $\mathcal{E}_n = X_n - \sum_{i=1}^{P} h_i X_{n-i}$, with the sum over past times becoming a sum over the window of past pixels.

The prediction window W_P

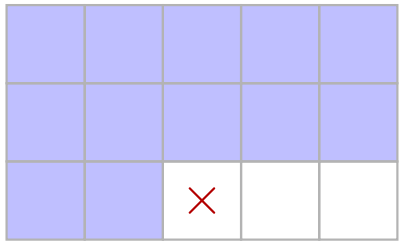
$$W_P = \left\{ r = (r_1, r_2) : \begin{array}{c} (r_1 = 0 \text{ and } 1 \le r_2 \le P) \\ \text{or} \\ (1 \le r_1 \le P \text{ and } -P \le r_2 \le P) \end{array} \right\} .$$

- Only earlier raster pixels are summed; the window holds $2P(P+1)$ pixels.

The window is not symmetric. Causality decides which neighbors may be used, and image geometry does not enter.



what geometry suggests



what raster causality allows

Why is the pixel above and to the right allowed, while the equally close pixel below and to the left is not? We return to this at the end of the lecture.

Matrix form

Using this convention, the prediction errors can be expressed in matrix form as

$$\mathcal{E} = (I - H)X \ ,$$

where H is the 2-D prediction matrix.

- This is exactly the form from Lecture 6, with $A = I - H$.
- X and $\mathcal{E}$ have length N^2, so H is $N^2 \times N^2$.
- 1-D results carry over: diagonal covariance, $\det(A) = 1$, $p_x(x)$.

Once the image has been raster ordered, the mathematics cannot tell that it came from a 2-D image. The geometry is hidden inside the structure of H.

Toeplitz block Toeplitz structure

The matrix H is called a **Toeplitz block Toeplitz** matrix. It has this special structure because H represents the application of a linear space-invariant (LSI) 2-D filter.

$$H = \begin{bmatrix} \boxed{B_0} & \boxed{B_{-1}} & \cdots & \boxed{B_{1-N}} \\ \boxed{B_1} & \boxed{B_0} & \cdots & \boxed{B_{2-N}} \\ \vdots & \vdots & \ddots & \vdots \\ \boxed{B_{N-1}} & \boxed{B_{N-2}} & \cdots & \boxed{B_0} \end{bmatrix} \qquad \boxed{B_k} = \begin{bmatrix} h_{(k,0)} & h_{(k,-1)} & \cdots & h_{(k,1-N)} \\ h_{(k,1)} & h_{(k,0)} & \cdots & h_{(k,2-N)} \\ \vdots & \vdots & \ddots & \vdots \\ h_{(k,N-1)} & h_{(k,N-2)} & \cdots & h_{(k,0)} \end{bmatrix}$$

- Horizontal shifts of the filter give Toeplitz structure inside each $N \times N$ block. Vertical shifts give the Toeplitz arrangement of the blocks. Compactly, $H_{mN+k,\,nN+l} = h_{(m-n,\,k-l)}$.

A Toeplitz block Toeplitz matrix results whenever a space-invariant filter is written as a matrix operator and the pixels are organized in raster order.

Formally identical results

The 2-D AR model has properties quite similar to the 1-D case. In fact, the results are formally identical, with 2-D convolution in place of 1-D convolution and the 2-D DSFT in place of the 1-D DTFT.

1-D	**2-D**
time n	pixel $s = (s_1, s_2)$
1-D convolution	2-D convolution
$H(\omega)$ (DTFT)	$H(\mu, \nu)$ (DSFT)
$S_X(\omega)$	$S_X(\mu, \nu)$

Autocovariance ($* =$ 2-D convolution) and power spectral density:

$$R_X(s) * (\delta_s - h_s) * (\delta_s - h_{-s}) = \sigma_C^2 \delta_s \ , \qquad S_X(\mu, \nu) = \frac{\sigma_C^2}{|1 - H(\mu, \nu)|^2} \ .$$

- Nothing new happens to the Gaussian model. The new issue is geometry. The pair (h_s, σ_C^2) is still a complete model of the process, by the result of Lecture 5.

Example 3.2: the 2-D density

The 2-D likelihood is the 1-D likelihood with n replaced by pixel location s. Consider the P^{th}-order 2-D AR random process X_s for $s \in S = \{(s_1, s_2) : 1 \leq s_1, s_2 \leq N\}$ with prediction variance σ_C^2 and prediction errors $\mathcal{E}_s = X_s - \sum_{r \in W_P} h_r X_{s-r}$. As before, assume $X_s = 0$ when $s_1 \leq 0$ or $s_2 \leq 0$.

Following along the lines of Example 3.1, we compute the probability density of X using the matrix-form PDF of Section 3.2:

$$p(x) = \frac{1}{(2\pi\sigma_C^2)^{N^2/2}} \exp\left\{ -\frac{1}{2\sigma_C^2} \sum_{s \in S} \left(x_s - \sum_{r \in W_P} h_r x_{s-r} \right)^2 \right\} .$$

- N^2 pixels replace N samples; nothing else changes.

Example 3.2: parameter and statistic vectors

In 2-D, we can define the parameter vector and statistic vector

$$\begin{aligned} h &= \left[h_{(0,1)}, \cdots, h_{(0,P)}, h_{(1,-P)}, \cdots, h_{(1,P)}, \cdots, h_{(P,P)}\right]^t , \\ Z_s &= \left[X_{s-(0,1)}, \cdots, X_{s-(0,P)}, X_{s-(1,-P)}, \cdots, X_{s-(1,P)}, \cdots, X_{s-(P,P)}\right]^t . \end{aligned}$$

Using these definitions, the prediction error can be compactly written as

$$\mathcal{E}_s = X_s - h^t Z_s .$$

- Both vectors have length $2P(P+1)$, one entry per window pixel.
- Z_s collects the window pixels around location s, and h holds their prediction coefficients. This is the 2-D version of $Z_n = [X_{n-1}, \cdots, X_{n-P}]^t$ from Lecture 7.

Example 3.2: log likelihood and statistics

Only the indexing changes from the 1-D derivation. The log likelihood can be written as

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

where

$$\begin{aligned}
\hat{\sigma}_x^2 &= \frac{1}{N^2}\sum_{s\in S} X_s^2 \ , \\
\hat{b} &= \frac{1}{N^2}\sum_{s\in S} Z_s X_s \ , \\
\hat{R} &= \frac{1}{N^2}\sum_{s\in S} Z_s Z_s^t \ .
\end{aligned}$$

- Same three statistics as 1-D; sums over n become sums over $s \in S$.

Example 3.2: the ML estimates

A calculation similar to that in Example 3.1 then results in the ML estimates of the model parameters h and σ_C^2:

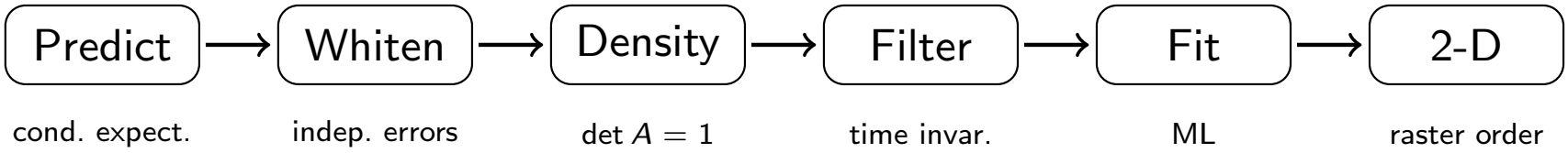
$$\begin{aligned} \hat{h} &= \hat{R}^{-1}\hat{b}\ , \\ \hat{\sigma}_C^2 &= \hat{\sigma}_x^2 - \hat{b}^t\hat{R}^{-1}\hat{b}\ , \end{aligned}$$

or, equivalently,

$$\hat{\sigma}_C^2 = \frac{1}{N^2}\sum_{s\in S}\left(X_s - \hat{h}^t Z_s\right)^2\ .$$

Nothing new had to be invented. The optimization and the answer are the same as in 1-D. Only the geometry is new.

Review of Chapter 3



1. **Predict.** $\hat{X}_n = \mathbb{E}[X_n | F_n]$ is the MMSE estimator; for jointly Gaussian data it is linear.
2. **Whiten.** Causal errors are independent, so $R_{\mathcal{E}} = \Lambda$ is diagonal.
3. **Density.** $\mathcal{E} = AX$ with A lower triangular and unit diagonal, so $\det(A) = 1$ and $p_x(x) \propto \exp\{-\frac{1}{2} x^t A^t \Lambda^{-1} A x\}$.
4. **Filter.** Time invariance makes prediction convolution: the AR process is white noise driving an IIR filter, with $S_X(\omega) = \sigma_C^2 / |1 - H(\omega)|^2$.
5. **Fit.** $\hat{h} = \hat{R}^{-1} \hat{b}$; $\hat{\sigma}_C^2 =$ average squared prediction error.
6. **2-D.** Impose raster order and every step repeats.

Every step works in 2-D. The disadvantage of the imposed ordering is the subject of the next lecture.

Summary

- An image has no natural causal order. Raster order supplies one, and it defines the asymmetric prediction window W_P, which holds $2P(P+1)$ pixels.
- The prediction error keeps its 1-D form, $\mathcal{E}_s = X_s - \sum_{r \in W_P} h_r X_{s-r}$, and in matrix form $\mathcal{E} = (I - H)X$.
- A space-invariant filter on a raster-ordered image gives a Toeplitz block Toeplitz matrix, $H_{mN+k,\, nN+l} = h_{(m-n,\, k-l)}$.
- The statistical results are formally identical to the 1-D case, with $R_X(s) * (\delta_s - h_s) * (\delta_s - h_{-s}) = \sigma_C^2 \delta_s$ and $S_X(\mu, \nu) = \sigma_C^2 / |1 - H(\mu, \nu)|^2$.
- ML fitting is unchanged as well, giving $\hat{h} = \hat{R}^{-1} \hat{b}$ and $\hat{\sigma}_C^2 = \frac{1}{N^2} \sum_{s \in S} (X_s - \hat{h}^t Z_s)^2$.
- The one real disadvantage is geometric. The model uses a pixel above and to the right but not the equally close pixel below and to the left.

Reading: Chapter 3.4. **Next lecture:** Chapter 4.1–4.2, noncausal Gaussian models.