# Lecture 6: Causal Prediction in Gaussian Models

Chapter 3.1–3.2

Charles A. Bouman

Purdue University

Friday, September 4, 2026

# Outline

1. **Causal prediction**
   Past, present, future; the MMSE causal predictor and its error.
2. **Properties of Gaussian prediction errors**
   Linear predictor; errors independent of the past and each other.
3. **Matrix form**
   The causal prediction matrix $H$, and $\mathcal{E} = (I - H)X$.
4. **Density functions from causal prediction**
   Error density, change of variables, data density and covariance.

Reading: Chapter 3.1–3.2, *Foundations of Computational Imaging*.

# Today's Big Ideas

- An accurate predictor of a signal is an accurate model of the signal.
- Zero-mean Gaussian: the best causal predictor is linear in the past.
- Causal prediction removes all the correlation, because the prediction errors are independent of the past and of each other.
- The predictor coefficients and the prediction error variances are a complete description of the process. They give its density and its covariance.

Causal order is natural for a 1-D signal. An image has no natural causal order, and imposing one can produce artifacts. That problem is what Chapter 4 is about.

# Prediction is the most basic tool in modeling

- Effective prediction requires an accurate model of its behavior.
- Clearly, an accurate model can enable accurate prediction.
- Converse: **an accurate predictor can create an accurate model**.

## Where this chapter is headed

We predict each sample from its past, show that the prediction errors are independent, and then use them to write down the density of the data.

# The order of prediction

Prediction requires an order. **Causal order** runs from the past to the future, which rules out circular dependencies.



1-D: causal order is natural

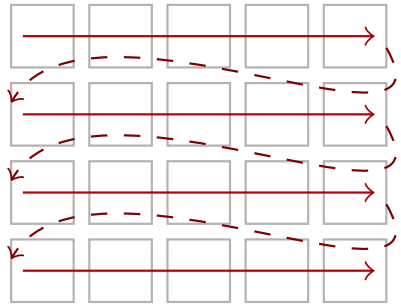
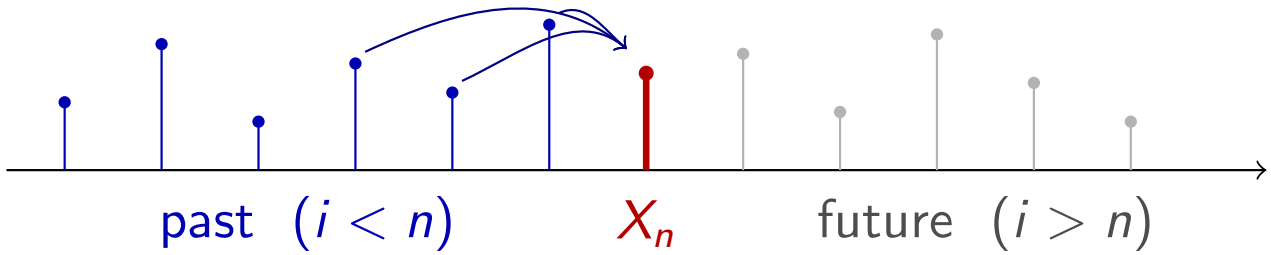


image: order must be *imposed*

For a 1-D signal the order is natural. For an image it is not, and the imposed raster order can produce artifacts.

# Past, present, and future

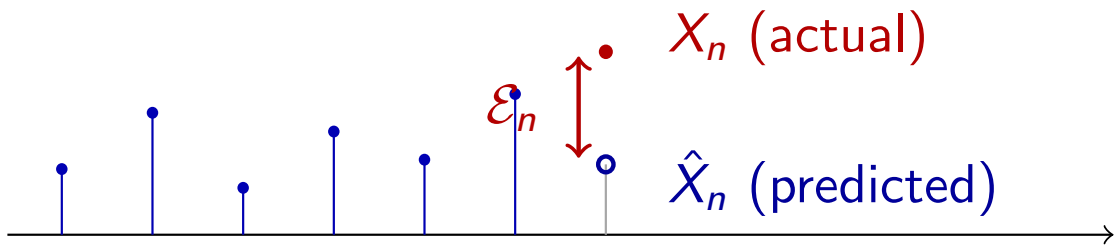
Let $X_1, X_2, \cdots, X_N$ be a portion of a discrete-time Gaussian random process, assumed **zero mean** (we may always subtract the mean in a preprocessing step). At any time $n$, the process partitions into three portions:




- **The Past**: $X_k$ for $1 \leq k < n$.
- **The Present**: $X_n$.
- **The Future**: $X_k$ for $n < k \leq N$.
- Our objective is to predict the present value $X_n$ from the past alone.

# The causal predictor and the prediction error

The prediction error is the part of $X_n$ that the past does not explain.




One reasonable predictor is the MMSE estimate of $X_n$:

$$\hat{X}_n \stackrel{\triangle}{=} \mathbb{E}[X_n | X_i \text{ for } i < n] = \mathbb{E}[X_n | \mathcal{F}_n] \,, \qquad \mathcal{E}_n \stackrel{\triangle}{=} X_n - \hat{X}_n \,,$$

where $\mathcal{F}_n = \{X_i \text{ for } i < n\}$ denotes the set of past observations. We call $\hat{X}_n$ the **causal predictor** and $\mathcal{E}_n$ the **causal prediction error**.

# Property: linearity of the Gaussian predictor

## Linearity of Gaussian predictor

The MMSE causal predictor for a zero-mean Gaussian random process is a linear function of the past; that is,

$$\hat{X}_n = \mathbb{E}[X_n|\mathcal{F}_n] = \sum_{i=1}^{n-1} h_{n,i} X_i \ ,$$

where $h_{n,i}$ are scalar coefficients.

This follows from the linearity of conditional expectation for zero-mean Gaussian random vectors. The prediction error is then also linear,

$$\mathcal{E}_n = X_n - \sum_{i=1}^{n-1} h_{n,i} X_i \ .$$

So $\mathcal{E}_n$ and $X_1, \cdots, X_{n-1}$ are jointly Gaussian, and for jointly Gaussian quantities uncorrelated implies independent.

# Derivation: the error is uncorrelated with the past

We show that $\mathbb{E}[X_i \mathcal{E}_n] = 0$ for all $i < n$, so the error contains nothing that the past can predict.

$$
\begin{aligned}
\mathbb{E}[X_i \mathcal{E}_n] &= \mathbb{E}\Big[X_i (X_n - \hat{X}_n)\Big] \\
&= \mathbb{E}[X_i X_n] - \mathbb{E}\Big[X_i \hat{X}_n\Big] \\
&= \mathbb{E}[X_i X_n] - \mathbb{E}\Big[X_i \mathbb{E}[X_n | \mathcal{F}_n]\Big] \\
&= \mathbb{E}[X_i X_n] - \mathbb{E}\Big[\mathbb{E}[X_i X_n | \mathcal{F}_n]\Big] \\
&= \mathbb{E}[X_i X_n] - \mathbb{E}[X_i X_n] \\
&= 0 \, .
\end{aligned}
$$

- Only two steps are more than definition or linearity. Equality 4 uses the known quantities property, since $X_i$ is known given $\mathcal{F}_n$. Equality 5 uses the filtration property.

# Property: independence of the errors from the past

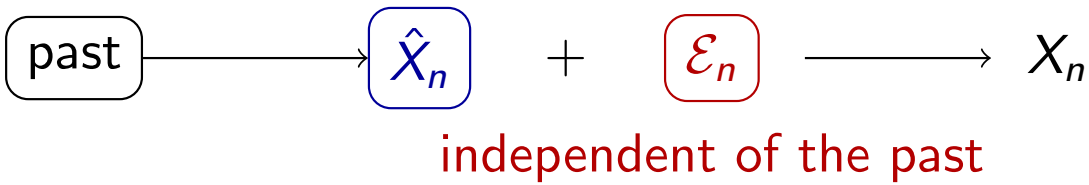
Since $\mathcal{E}_n$ and $X_1, \cdots, X_{n-1}$ are jointly Gaussian, being uncorrelated makes them independent.

## Independence of causal Gaussian prediction errors from past

The MMSE causal prediction errors for a zero-mean Gaussian random process are independent of the past of the random process. Formally, we write that for all $n$,

$$\mathcal{E}_n \perp\!\!\!\perp (X_1, \cdots, X_{n-1}) ,$$

where the symbol $\perp\!\!\!\perp$ indicates that the two quantities on the left and right are jointly independent of each other.




- The past fixes $\hat{X}_n$; $\mathcal{E}_n$ is new information.

# Property: joint independence of the prediction errors

**Joint independence of causal Gaussian prediction errors**

The MMSE causal prediction errors for a zero-mean Gaussian random process are jointly independent, which implies that for all $i \neq j$, $\mathcal{E}_i \perp\!\!\!\perp \mathcal{E}_j$.

There is a simple reason to expect this. For $i < n$,

$$\mathcal{E}_i = X_i - \mathbb{E}[X_i|\mathcal{F}_i]$$

is a function of $X_1, \cdots, X_i$, so $\mathcal{E}_i$ is part of the past relative to time $n$.

- We just proved that $\mathcal{E}_n$ is independent of the past, so $\mathcal{E}_n$ must be independent of $\mathcal{E}_i$ for $i < n$.

# Derivation: the errors are uncorrelated with each other

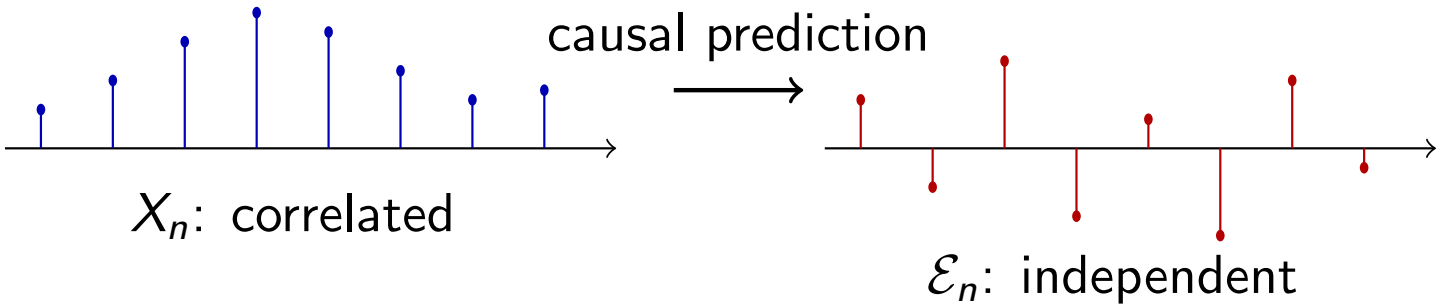
Assume $i < n$:

$$\begin{aligned}
\mathbb{E}[\mathcal{E}_n \mathcal{E}_i] &= \mathbb{E}\left[\left(X_n - \hat{X}_n\right)\left(X_i - \hat{X}_i\right)\right] \\
&= \mathbb{E}\Big[\left(X_n - \mathbb{E}[X_n|\mathcal{F}_n]\right)\left(X_i - \mathbb{E}[X_i|\mathcal{F}_i]\right)\Big] \\
&= \mathbb{E}\left[\mathbb{E}\Big[\left(X_n - \mathbb{E}[X_n|\mathcal{F}_n]\right)\left(X_i - \mathbb{E}[X_i|\mathcal{F}_i]\right)\Big|\,\mathcal{F}_n\right]\right] \\
&= \mathbb{E}\left[\left(X_i - \mathbb{E}[X_i|\mathcal{F}_i]\right)\mathbb{E}\Big[\left(X_n - \mathbb{E}[X_n|\mathcal{F}_n]\right)\Big|\,\mathcal{F}_n\Big]\right] \\
&= \mathbb{E}[\left(X_i - \mathbb{E}[X_i|\mathcal{F}_i]\right)\left(\mathbb{E}[X_n|\mathcal{F}_n] - \mathbb{E}[X_n|\mathcal{F}_n]\right)] \\
&= \mathbb{E}[\left(X_i - \mathbb{E}[X_i|\mathcal{F}_i]\right)0] \;=\; 0\,.
\end{aligned}$$

- Equality 3 is the filtration property. Equalities 4 and 5 are the known quantities property, since for $i < n$ the factor involving $X_i$ is known given $\mathcal{F}_n$. By symmetry the result also holds for $i > n$, and joint Gaussianity then gives joint independence.

# Prediction removes correlation

The errors are independent, but their variances are unknown. The **causal prediction variance** is $\sigma_n^2 \triangleq \mathbb{E}\left[\mathcal{E}_n^2\right]$.




- The $h_{n,i}$ and $\sigma_n^2$ together describe the process.

# Matrix form of causal prediction

The prediction equations can be compactly expressed using vector-matrix notation. Let $X$, $\hat{X}$, and $\mathcal{E}$ denote column vectors with elements indexed from one to $N$, so, for example, $X = [X_1, \cdots, X_N]^t$.

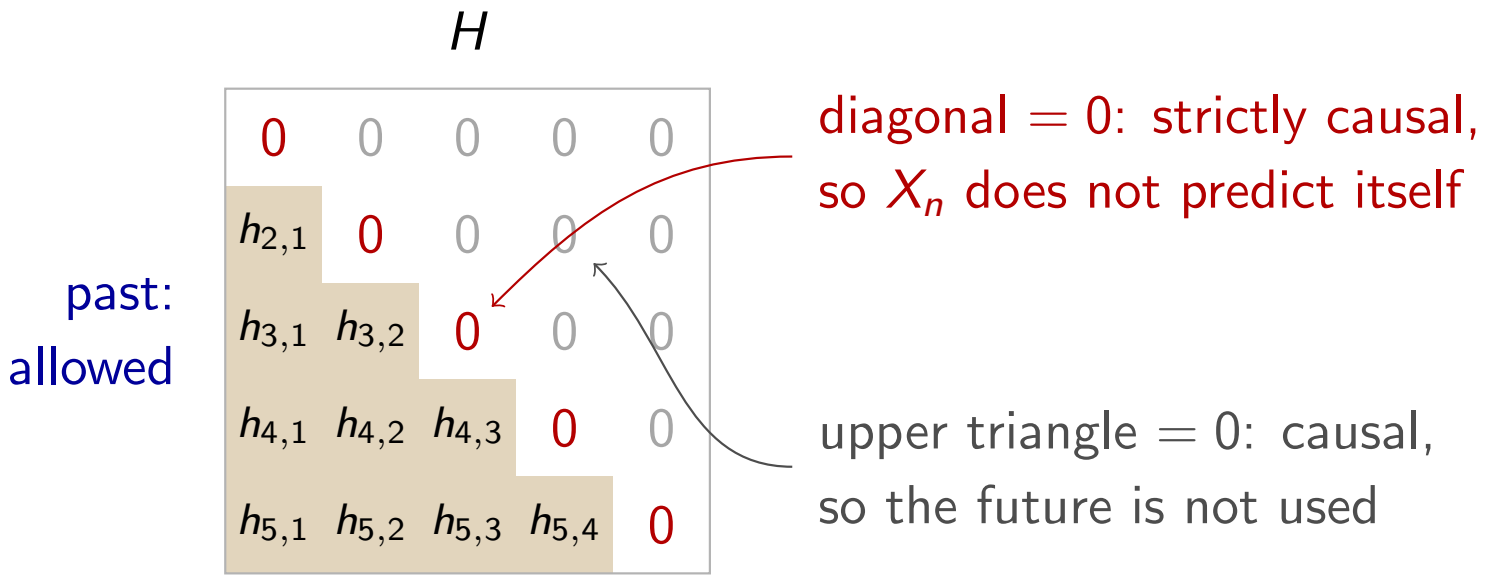
Then the causal prediction equation becomes

$$\hat{X} = HX \ ,$$

where $H$ is an $N \times N$ **causal prediction matrix** containing the prediction coefficients, $h_{i,j}$.

- Row $n$ of $H$ holds the coefficients $h_{n,i}$ that predict $X_n$ from the past.

# The causal prediction matrix is lower triangular



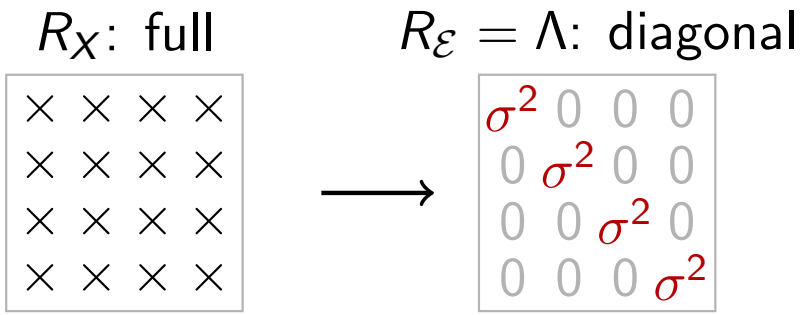

- Then $\mathcal{E} = (I - H)X = AX$, where $A = I - H$ is lower triangular with ones on its diagonal.

# The probability density of the prediction errors

For a zero-mean Gaussian process, the density of the causal prediction errors is given by

$$p_{\mathcal{E}}(e) = \frac{1}{(2\pi)^{N/2}} |\Lambda|^{-1/2} \exp\left\{ -\frac{1}{2} e^t \Lambda^{-1} e \right\} ,$$

where $\Lambda \stackrel{\triangle}{=} \text{diag}\left\{\sigma_1^2, \cdots, \sigma_N^2\right\}$.
This is the multivariate Gaussian density with covariance $\mathbb{E}[\mathcal{E}\mathcal{E}^t] = \Lambda$, which is diagonal because the errors are independent:




- A full $N \times N$ covariance has been replaced by $N$ numbers.
- Since $\Lambda$ is diagonal, the quadratic form separates, $e^t \Lambda^{-1} e = \sum_{n=1}^{N} e_n^2/\sigma_n^2$, so $p_{\mathcal{E}}$ is a product of $N$ scalar Gaussian densities.

# Can we write the density of $X$ in closed form?

## The question

Can we write down the probability density of $X$ in closed form from the prediction description, the coefficients $h_{n,i}$ and the variances $\sigma_n^2$?

Since $\mathcal{E} = AX$ is a bijective linear transformation, the density of $X$ is the density of $\mathcal{E}$ scaled by the absolute value of the **Jacobian determinant**,

$$p_x(x) = |\det(A)| \; p_{\mathcal{E}}(Ax) \; .$$

One important detail makes this a closed form. $A$ is lower triangular with ones on its diagonal, and the determinant of a triangular matrix is the product of its diagonal entries, so

$$\det(A) = 1 \; .$$

So $p_x(x) = p_{\mathcal{E}}(Ax)$, with no scaling factor at all.

# The density of the data, and its covariance

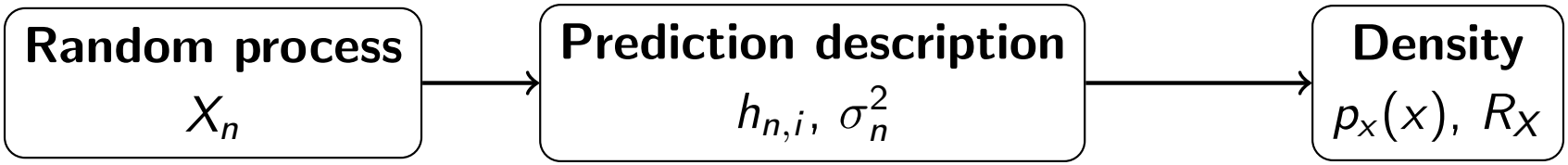
The answer is yes. Substituting the density of the errors gives the closed form

$$p_x(x) = \frac{1}{(2\pi)^{N/2}} |\Lambda|^{-1/2} \exp\left\{ -\frac{1}{2} x^t A^t \Lambda^{-1} A x \right\} ,$$

and therefore the covariance of $X$ is

$$R_X = (A^t \Lambda^{-1} A)^{-1} ,$$

where $A$ is the causal prediction matrix and $\Lambda$ is the diagonal matrix of causal prediction variances.




- The predictor coefficients and the $N$ prediction variances are a complete description of the process. This is the precise sense in which an accurate predictor is an accurate model.

# Summary

- Causal prediction splits each sample into a predicted part and an error, $X_n = \hat{X}_n + \mathcal{E}_n$, where $\hat{X}_n = \mathbb{E}[X_n|\mathcal{F}_n]$.
- For a zero-mean Gaussian process the predictor is linear, $\hat{X}_n = \sum_{i=1}^{n-1} h_{n,i} X_i$.
- The prediction errors are independent of the past and of each other, so $\mathbb{E}[\mathcal{E}\mathcal{E}^t] = \Lambda$ is diagonal with entries $\sigma_n^2$.
- In matrix form $\hat{X} = HX$ and $\mathcal{E} = (I - H)X = AX$, where $H$ is lower triangular with a zero diagonal.
- $A$ is lower triangular with ones on its diagonal, so $\det(A) = 1$ and $p_x(x) = p_{\mathcal{E}}(Ax)$.
- The coefficients $h_{n,i}$ and the variances $\sigma_n^2$ are therefore a complete model, giving $p_x(x) = (2\pi)^{-N/2}|\Lambda|^{-1/2} \exp\left\{-\frac{1}{2}x^t A^t \Lambda^{-1} A x\right\}$ and $R_X = (A^t \Lambda^{-1} A)^{-1}$.

**Reading:** Chapter 3.1–3.2. **Next lecture:** Chapter 3.3, 1-D Gaussian autoregressive models.