

Lecture 10: 1-D and 2-D Gaussian Markov Random Fields

Chapter 4.3–4.4

Charles A. Bouman

Purdue University

Wednesday, September 16, 2026

① **From noncausal prediction to neighborhoods**

The Markov property and the 1-D GMRF window.

② **Homogeneous GMRFs**

Local prediction, symmetry, and the precision matrix.

③ **What changes from the AR model**

Prediction-error correlation and the GMRF power spectrum.

④ **From 1-D to 2-D**

Symmetric 2-D neighborhoods and the same spectral theory.

Reading: Chapter 4.3–4.4, *Foundations of Computational Imaging*.

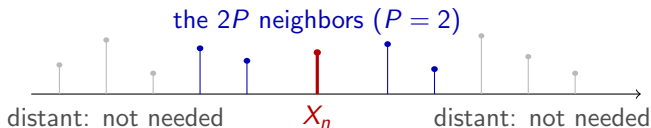
Today's Big Ideas

- A GMRF predicts each sample from a window of neighbors, so once those neighbors are known the distant samples add nothing.
- The neighborhood determines the precision matrix, because a sample that is left out of the predictor gives a zero entry.
- The noncausal prediction errors stay correlated with each other, and that changes the shape of the power spectrum.
- In two dimensions the neighborhood is a symmetric window around the pixel, so the raster order only indexes the pixels and every one-dimensional result carries over.

Restricting noncausal prediction to a neighborhood gives the GMRF, which is the noncausal counterpart of the AR model.

The neighborhood window ∂n

In Lecture 9 the noncausal predictor used *all* of the other samples. An important special case limits the predictor to a window of $n \pm P$ about the point being predicted, where P is the order of the model.



In order to simplify the notation, we define the window

$$\partial n \triangleq \{i \in [1, \dots, N] : i \neq n \text{ and } |i - n| \leq P\} .$$

- The window is symmetric, with P samples on either side of n , and it is truncated at the boundary. The point n itself is excluded, which is the condition $g_{n,n} = 0$ of Lecture 9.
- In contrast, the $\text{AR}(P)$ window held only the P *past* samples.

Definition of a 1-D GMRF

1-D GMRF

A **1-D GMRF** is any Gaussian random process with the property that

$$\mathbb{E}[X_n | X_i \text{ for } i \neq n] = \mathbb{E}[X_n | X_i \text{ for } i \in \partial n] \text{ .}$$

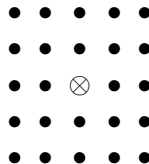
In words, the MMSE noncausal predictor for any pixel in a GMRF depends only on the pixel's neighbors.

- This is the **Markov property** in its noncausal form, and it is the reason for the name Gaussian *Markov* random field.
- The concept of a Markov random field grew out of the study of objects of two or more dimensions, where it is most useful. It applies perfectly well in 1-D also.

GMRF prediction windows for order $P = 2$



1-D GMRF: $2P = 4$ inputs



2-D GMRF: $(2P + 1)^2 - 1 = 24$ inputs



1-D AR (Lec. 7)



2-D AR (Lec. 8)

- The pixel $\otimes$ is predicted using past *and* future values $\bullet$.
- The AR neighborhood holds the P past samples. The GMRF neighborhood holds every sample within distance P .

The GMRF prediction error

The GMRF is a bit like an AR process, but the causal predictor of the AR process is replaced with a noncausal predictor. For the most general case of a zero-mean GMRF,

$$\mathcal{E}_n = X_n - \sum_{i \in \partial n} g_{n,i} X_i .$$

- The sum now runs over the $2P$ neighbors instead of all $N - 1$ other samples.
- Lecture 9 gave $B_{n,i} = -g_{n,i}/\sigma_n^2$ for $i \neq n$. So if $i \notin \partial n$, then $g_{n,i} = 0$, and therefore $B_{n,i} = 0$:

neighbors of X_n : $X_{n-1}, X_{n+1} \iff$ row n of B : $\cdots 0 \quad * \quad * \quad * \quad 0 \cdots$

- To reduce the number of model parameters, it is often useful to assume that the prediction coefficients are not a function of the position n . This is the same step that took us from a general causal predictor to the AR model.

Definition of a homogeneous GMRF

Homogeneous GMRF

A GMRF is said to be **homogeneous** if both the MMSE noncausal predictor and the MMSE prediction variance are invariant to position.

The MMSE noncausal prediction error for a homogeneous GMRF then has the form

$$\mathcal{E}_n = X_n - \sum_{i \in \partial n} g_{n-i} X_i ,$$

with the noncausal prediction variance, σ_{NC}^2 , taking on a constant value.

- The prediction coefficients g_{n-i} now depend only on the displacement $n - i$, and one number σ_{NC}^2 replaces the N prediction variances.

The density of a homogeneous GMRF

In this case, we can write the density function for the homogeneous GMRF as

$$p(x) = \frac{1}{(2\pi)^{N/2}} |B|^{1/2} \exp \left\{ -\frac{1}{2} x^t B x \right\}, \quad \text{where} \quad B_{i,j} = \frac{\delta_{i-j} - g_{i-j}}{\sigma_{NC}^2}.$$

Here B is the precision matrix of Lecture 9, assumed positive definite.

*	*	*	0	0	0	0
*	*	*	*	0	0	0
*	*	*	*	*	0	0
0	*	*	*	*	*	0
0	0	*	*	*	*	*
0	0	0	*	*	*	*
0	0	0	0	*	*	*

B is banded and Toeplitz:
only neighbors within distance P
produce nonzero entries

- This is the Lecture 9 relation $B = \Gamma^{-1}(I - G)$ with $\Gamma = \sigma_{NC}^2 I$.
- Since B must be symmetric, the prediction filter must satisfy $g_n = g_{-n}$. The model is then described by the $P + 1$ numbers $g_1, \dots, g_P$ and σ_{NC}^2 .

Noncausal prediction as a convolution

Once again, extending X_n to infinite extent, the relation between X_n and the noncausal prediction errors becomes convolution:

$$\mathcal{E}_n = X_n * (\delta_n - g_n) . \quad (1)$$

Solving the definition of $\mathcal{E}_n$ for X_n also gives

$$X_n = \mathcal{E}_n + \sum_{i \in \partial n} g_{n-i} X_i . \quad (2)$$

What we will prove

Equation (1) has the same form as the AR filter equation of Lecture 7, but the noncausal prediction errors are not white. The next two slides compute the autocovariance of the errors and show exactly how they are correlated.

Derivation 1: the covariance of $\mathcal{E}_n$ with X_{n+k}

The goal is to compute the covariance between the prediction error $\mathcal{E}_n$ and the samples X_{n+k} . We start with $k = 0$ and substitute equation (2),

$$\begin{aligned}\mathbb{E}[\mathcal{E}_n X_n] &= \mathbb{E}\left[\mathcal{E}_n \left(\mathcal{E}_n + \sum_{i \in \partial n} g_{n-i} X_i\right)\right] \\ &= \mathbb{E}[\mathcal{E}_n^2] + \sum_{i \in \partial n} g_{n-i} \mathbb{E}[\mathcal{E}_n X_i] \\ &= \mathbb{E}[\mathcal{E}_n^2] + 0 = \sigma_{NC}^2 .\end{aligned}$$

The last line uses the Lecture 9 property $\mathcal{E}_n \perp\!\!\!\perp \{X_i\}_{i \neq n}$. That same property gives $\mathbb{E}[\mathcal{E}_n X_{n+k}] = 0$ for $k \neq 0$, so the two cases combine into

$$\mathbb{E}[\mathcal{E}_n X_{n+k}] = \sigma_{NC}^2 \delta_k . \tag{3}$$

The prediction errors are uncorrelated with the values used in the prediction.

Derivation 2: the autocovariance of the errors

Lecture 9 showed that the noncausal prediction errors are not independent of each other. We now compute their autocovariance, $R_{\mathcal{E}}(k) \triangleq \mathbb{E}[\mathcal{E}_n \mathcal{E}_{n+k}]$, by substituting equation (2) into equation (3).

$$\begin{aligned}\sigma_{NC}^2 \delta_k &= \mathbb{E}[\mathcal{E}_n \mathcal{X}_{n+k}] \\&= \mathbb{E} \left[\mathcal{E}_n \left(\mathcal{E}_{n+k} + \sum_{i \in \partial(n+k)} g_{n+k-i} \mathcal{X}_i \right) \right] \\&= \mathbb{E}[\mathcal{E}_n \mathcal{E}_{n+k}] + \sum_{i \in \partial(n+k)} g_{n+k-i} \mathbb{E}[\mathcal{E}_n \mathcal{X}_i] \\&= R_{\mathcal{E}}(k) + \sum_{i \in \partial(n+k)} g_{n+k-i} \sigma_{NC}^2 \delta_{i-n} \\&= R_{\mathcal{E}}(k) + \sigma_{NC}^2 g_k .\end{aligned}$$

Rearranging terms results in

$$R_{\mathcal{E}}(n) = \sigma_{NC}^2 (\delta_n - g_n) . \quad (4)$$

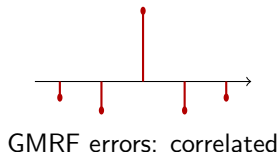
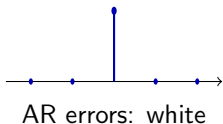
The prediction errors are not white

The key difference from the AR model

The autocovariance of the noncausal prediction errors is

$$R_{\mathcal{E}}(n) = \sigma_{NC}^2 (\delta_n - g_n) ,$$

so in general the errors are **not white**.



- For $n \neq 0$ the autocovariance is $R_{\mathcal{E}}(n) = -\sigma_{NC}^2 g_n$. So the errors are white only when $g_n = 0$ for all $n \neq 0$.
- For causal prediction the errors *were* independent, and that is what made the Chapter 3 density derivation work. Dropping causality loses that property.

Relating the two autocovariances

We now have two expressions for the autocovariance of the prediction errors, and comparing them gives the power spectrum of X . From equation (1) and the autocovariance of a filtered random process (Chapter 2 homework),

$$R_{\mathcal{E}}(n) = R_X(n) * (\delta_n - g_n) * (\delta_n - g_{-n}) .$$

Equating this with equation (4) gives

$$\sigma_{NC}^2 (\delta_n - g_n) = R_X(n) * (\delta_n - g_n) * (\delta_n - g_{-n}) .$$

Now take the DTFT of both sides, and use $g_n = g_{-n}$ so that $G(\omega)$ is real:

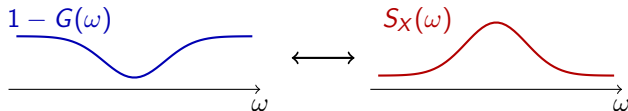
$$\sigma_{NC}^2 (1 - G(\omega)) = S_X(\omega) (1 - G(\omega))^2 . \quad (5)$$

Here $G(\omega)$ is the DTFT of g_n and $S_X(\omega)$ is the power spectrum of X_n . One factor of $(1 - G(\omega))$ cancels from the two sides, at every ω where $1 - G(\omega) \neq 0$.

The GMRF power spectrum

From equations (4) and (5), the power spectral densities of the homogeneous GMRF and its noncausal prediction errors are

$$S_X(\omega) = \frac{\sigma_{NC}^2}{1 - G(\omega)} , \quad S_{\mathcal{E}}(\omega) = \sigma_{NC}^2 (1 - G(\omega)) .$$



Where $1 - G(\omega)$ is small, the process has large power. Taking the inverse DTFT of the first expression,

$$R_X(n) * (\delta_n - g_n) = \sigma_{NC}^2 \delta_n .$$

- The power spectral density completely specifies a zero-mean stationary Gaussian process, so g_n and σ_{NC}^2 are a complete model.

The AR model versus the GMRF

AR

The predictor h_n is causal and the errors are white.

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

$$S_{\mathcal{E}}(\omega) = \sigma_C^2$$

$$R_{\mathcal{E}}(n) = \sigma_C^2 \delta_n$$

GMRF

The predictor $g_n = g_{-n}$ is symmetric and the errors are not white.

$$S_X(\omega) = \frac{\sigma_{NC}^2}{1 - G(\omega)}$$

$$S_{\mathcal{E}}(\omega) = \sigma_{NC}^2 (1 - G(\omega))$$

$$R_{\mathcal{E}}(n) = \sigma_{NC}^2 (\delta_n - g_n)$$

- Here σ_C^2 is the causal prediction variance of Lecture 7 and $H(\omega)$ is the DTFT of the causal predictor h_n .
- The difference in the two power spectra comes entirely from the difference in the prediction errors. White errors give a squared magnitude in the denominator, and correlated errors give the first power.

Raster order no longer defines the model

All of the 1-D derivations generalize to regular grids in two or more dimensions. Each lattice point is written $s = (s_1, s_2)$, and the set of all lattice points is $S = \{1, \dots, N\}^2$. The pixels are ordered in **raster order**,

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

with $\mathcal{E}$ ordered similarly.

2-D AR: raster order determines *which neighbors are allowed*.

2-D GMRF (today): raster order only says *where pixels sit in the vector*.

- This resolves the problem that Lecture 9 raised at the start of Chapter 4. A causal order on an image is artificial, and the GMRF does not need one.

The 2-D neighborhood and prediction error

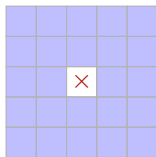
If the GMRF is homogeneous, then the 2-D noncausal prediction error is again given by

$$\mathcal{E}_s = X_s - \sum_{r \in \partial s} g_{s-r} X_r ,$$

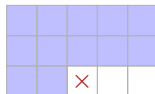
where the 2-D neighborhood is defined as

$$\partial s \triangleq \{r = (r_1, r_2) : r \neq s \text{ and } |r_1 - s_1| \leq P \text{ and } |r_2 - s_2| \leq P\} ,$$

a $(2P + 1) \times (2P + 1)$ window about the point being predicted.



GMRF: symmetric



AR: causal

The prediction errors are again $\mathcal{E} = (I - G)X$, where the 2-D prediction matrix G is **Toeplitz block Toeplitz**.

The 2-D results are formally identical

The results for the 2-D stationary GMRF are formally identical to the 1-D results, with 2-D convolution replacing 1-D convolution and the 2-D DSFT replacing the 1-D DTFT. More specifically,

$$\begin{aligned}R_{\mathcal{E}}(s) &= \sigma_{NC}^2 (\delta_s - g_s) , \\ R_X(s) * (\delta_s - g_s) &= \sigma_{NC}^2 \delta_s ,\end{aligned}$$

where $*$ denotes 2-D convolution of 2-D functions, and

$$\begin{aligned}S_X(\mu, \nu) &= \frac{\sigma_{NC}^2}{1 - G(\mu, \nu)} , \\ S_{\mathcal{E}}(\mu, \nu) &= \sigma_{NC}^2 (1 - G(\mu, \nu)) ,\end{aligned}$$

where $G(\mu, \nu)$ is the DSFT of g_s .

- A symmetric local image neighborhood therefore defines a full stationary Gaussian image model.

Summary

- A GMRF makes noncausal prediction local, so that $\mathbb{E}[X_n | X_i \text{ for } i \neq n] = \mathbb{E}[X_n | X_i \text{ for } i \in \partial n]$ and $\mathcal{E}_n = X_n - \sum_{i \in \partial n} g_{n,i} X_i$.
- For a homogeneous GMRF the neighborhood determines a sparse precision matrix, because $B_{i,j} = \frac{1}{\sigma_{NC}^2}(\delta_{i-j} - g_{i-j})$ is banded and zero beyond distance P . The symmetry of B forces the symmetric filter $g_n = g_{-n}$.
- Unlike AR models, the prediction errors are correlated, $R_{\mathcal{E}}(n) = \sigma_{NC}^2(\delta_n - g_n)$, and that is why the GMRF power spectrum has a first-power denominator, $S_X = \sigma_{NC}^2/(1 - G)$, instead of the AR $\sigma_C^2/|1 - H|^2$.
- In 2-D the neighborhood is symmetric rather than causal, so the raster order only indexes the pixels and every formula carries over with 2-D convolution and the DSFT, $S_X(\mu, \nu) = \sigma_{NC}^2/(1 - G(\mu, \nu))$.

Reading: Chapter 4.3–4.4. **Next lecture:** Chapter 4.5–4.6, the relation between GMRF and AR models, and GMRFs on general lattices.