Contents

1 Causal Gaussian Models **3**

1.1 Causal Prediction in Gaussian Models 4

1.2 Density Functions Based on Causal Predication 7

1.3 1-D Gaussian Autoregressive (AR) Models 8

1.4 2-D Gaussian AR Models 13

2 Non-Causal Gaussian Models **19**

2.1 Non-Causal Prediction in Gaussian Models 19

2.2 Density Functions Based on Non-Causal Prediction 21

2.3 1-D Gaussian Markov Random Fields (GMRF) 22

2.4 2-D Gaussian Markov Random Fields (GMRF) 25

2.5 Relation Between GMRF and Gaussian AR Models 26

2.6 GMRF Models on General Lattices 30

Chapter 1

Causal Gaussian Models

- $n \in [1, \cdots, N]$ - 1-D index set
- $s \in S$ - arbitrary index set
- X_n - random process in 1 or more dimensions
- $\mathcal{E}_s$ - causal or non-causal prediction error
- Causal notation
  - H - causal predictor matrix
  - $A = I - H$ - causal predictor matrix
  - $\mathcal{E} = AX = (I - H)X$ - causal predictor matrix
  - Λ - diagonal matrix of causal prediction variances
  - $h_{i,j}$ or h_{i-j} - causal prediction filter
  - $\mathcal{F}_n = \{X_i : i < n\}$ - past observations
  - $H(\omega)$ - DTFT of causal prediction filter
- Non-causal notation
  - G - non-causal predictor matrix
  - $B = I - G$ - non-causal predictor matrix
  - $\mathcal{E} = BX = (I - G)X$ - non-causal predictor matrix
  - Γ - diagonal matrix of non-causal prediction variances
  - $g_{i,j}$ or g_{i-j} - non-causal prediction filter

– $G(\omega)$ - DTFT of causal prediction filter

Perhaps the most basic tool in modeling is prediction. Intuitively, if one can effectively predict the behavior of something, then one must have an accurate model of its behavior. Clearly, an accurate model can enable accurate prediction, but we will demonstrate that the converse is also true: an accurate predictor can be used to create an accurate model.

In order to model data using prediction, we must decide the order in which prediction will occur. The simplest approach is to predict values in causal order, starting in the past and proceeding toward the future. In this chapter, we will show that causal prediction leads to many interesting and powerful tools; and, perhaps most importantly, it eliminates the possibility of circular dependencies in the prediction model.

However, the price we pay for causal prediction is that it requires that we impose a causal ordering on the data. For some types of data, causal ordering is quite natural. However, for images, which are the primary subject of this book, this is typically not the case, and imposing a causal order can often lead to artifacts in the results of processing.

1.1 Causal Prediction in Gaussian Models

Let $X_1, X_2, \cdots, X_N$ be a discrete-time Gaussian random process. Without loss of generality, we will assume that X_n is zero-mean, since we may always subtract the mean from X in a preprocessing step. At any particular time n, we may partition the random processing into three distinct 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 current value, X_n, from the past. As we saw in Chapter **??**, one reasonable predictor is the MMSE estimate of X_n given by

$$\hat{X}_n \triangleq E\left[X_n | X_i \text{ for } i < n\right] \ .$$

We will refer to $\hat{X}$ as a **causal predictor** since it only uses the past to predict the present value, and we define the **causal prediction error** as

$$\mathcal{E}_n = X_n - \hat{X}_n \ .$$

In order to simplify notation, let $\mathcal{F}_n$ denote the set of past observations given by $\mathcal{F}_n = \{X_i \text{ for } i < n\}$. Then the MMSE causal predictor of X_n can be succinctly expressed as

$$\hat{X}_n = \mathrm{E}[X_n|\mathcal{F}_n] \ .$$

Causal prediction leads to a number of very interesting and useful properties, the first of which is listed below.

Property 1.1: *Linearity of Gaussian predictor* - The MMSE causal predictor for a zero-mean Gaussian random process is a linear function of the past, i.e.

$$\hat{X}_n = E\left[X_n|\mathcal{F}_n\right] = \sum_{i=1}^{n-1} h_{n,i} X_i \tag{1.1}$$

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

This property is a direct result of the linearity of conditional expectation for zero-mean Gaussian random vectors (Property **??**.5). From this, we also know that the prediction errors must be a linear function of the past and present values of X.

$$\mathcal{E}_n \ = \ X_n - \sum_{i=1}^{n-1} h_{n,i} X_i \tag{1.2}$$

Another important property of causal prediction is that the prediction error, $\mathcal{E}_n$, is uncorrelated from all past values of X_i for $i < n$.

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

Notice that the fourth equality is a result of Property **??**.4, and the fifth equality is a result of Property **??**.3. Since both X and $\mathcal{E}$ are jointly Gaussian,

this result implies that the prediction errors are independent of past values of X, which is stated in the following property.

Property 1.2: *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

$$\mathcal{E}_n \perp\!\!\!\perp (X_0, \cdots, X_{n-1}) \ ,$$

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

A similar approach can be used to compute the correlation between predictions errors themselves. If we assume that $i < n$, then the correlation of prediction errors is given by

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

By symmetry, this result must also hold for $i > n$. Also, since we know that the prediction errors are jointly Gaussian, we can therefore conclude joint independence from this result.

Property 1.3: *Joint independence of causal Gaussian prediction errors* - The MMSE prediction errors for a zero-mean Gaussian random process are jointly independent which implies that for all $i \neq j$, $\mathcal{E}_i \perp\!\!\!\perp \mathcal{E}_j$.

The causal prediction errors are independent, but we do not know their variance. So we denote the **causal prediction variance** for X_n as

$$\sigma_n^2 \triangleq \mathrm{E}\big[\mathcal{E}_n^2\big] \ .$$

The prediction equations of (1.1) and (1.2) can be compactly expressed using vector-matrix notation. To do this, we let X, $\hat{X}$, and $\mathcal{E}$ denote column vectors with elements indexed form 1 to N. So for example, $X = [X_1, \cdots, X_N]^t$.

Then the causal prediction equation of (1.1) becomes

$$\hat{X} = HX \tag{1.3}$$

where H is an $N \times N$ **causal prediction matrix** containing the prediction coefficients, $h_{i,j}$. By relating the entries of H to the coefficients of (1.1), we can see that H is a lower triangular matrix with zeros on the diagonal and with the following specific form.

$$H = \begin{bmatrix} 0 & 0 & \cdots & & 0 \\ h_{2,1} & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ h_{N-1,1} & h_{N-1,2} & \cdots & 0 & 0 \\ h_{N,1} & h_{N,2} & \cdots & h_{N,N-1} & 0 \end{bmatrix}$$

Using this notation, the causal prediction error is then given by

$$\mathcal{E} = (I - H)X = AX \ , \tag{1.4}$$

where $A = I - H$.

1.2 Density Functions Based on Causal Predication

We can derive compact expressions for the density of both the prediction error, $\mathcal{E}$, and the data, X, by using the vector-matrix notation of the previous section. To do this, first define $\Lambda = \text{diag}\left\{\sigma_1^2, \cdots, \sigma_N^2\right\}$ to be a diagonal matrix containing the causal prediction variances. Then the covariance of $\mathcal{E}$ is given by

$$\text{E}\big[\mathcal{E}\mathcal{E}^t\big] = \Lambda$$

due to the independence of the prediction errors. Using the general form of the density function for a zero-mean multivariate Gaussian random vector, we then can write the density function for $\mathcal{E}$ as

$$p_{\mathcal{E}}(e) = \frac{1}{(2\pi)^{N/2}} |\Lambda|^{-1/2} \exp\left\{-\frac{1}{2} e^t \Lambda^{-1} e\right\} \ . \tag{1.5}$$

Since $\mathcal{E}$ and X are related through a bijective transformation, it can be easily shown that the density of X is proportional to the density of $\mathcal{E}$, with

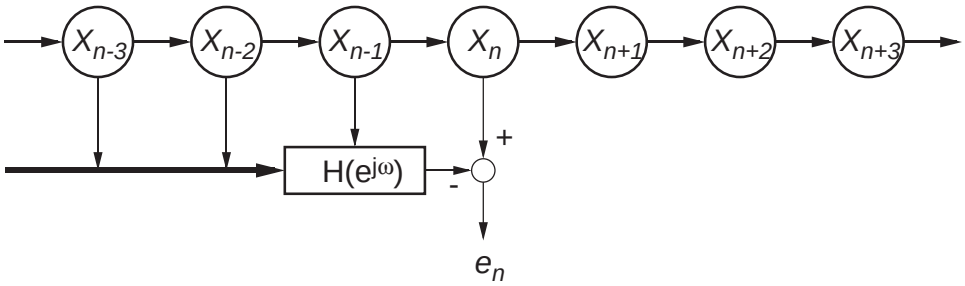

Figure 1.1: Diagram of a causal predictor for a P^{th} order Gaussian AR process. The linear time-invariant prediction filter, h_n, has frequency response, $H(\omega)$. The resulting prediction errors, $\mathcal{E}_n$, are white when the predictor is optimal.

the absolute value of the Jacobian determinant of the transformation serving as the constant of proportionality. For this particular linear relationship between X and $\mathcal{E}$, the probability densities are related by

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

where $|\det(A)|$ is the absolute value of the determinant of the matrix A. Fortunately, because A is a causal predictor, it is constrained to be lower triangular with 1's on its diagonal. Therefore, its determinant is one. Applying this result, and using the form of the density function for $p_{\mathcal{E}}(e)$ of (1.5) yields

$$p_x(x) = \frac{1}{(2\pi)^{N/2}}|\Lambda|^{-1/2}\exp\left\{-\frac{1}{2}x^t A^t \Lambda^{-1} A x\right\}\ . \tag{1.6}$$

From this, we can also see that the covariance of X is given by

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

where A is the causal prediction matrix and Λ is the diagonal matrix of causal prediction variances.

1.3 1-D Gaussian Autoregressive (AR) Models

Time invariance is a very important concept that plays an essential role in the modeling of data. This is because in many practical cases it is reasonable to assume that the characteristic behavior of data does not change with time. One method for enforcing time-invariance in a random processes is to specify that the parameters of the model not change with time. When this is the

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

Toeplitz Circulant

Figure 1.2: Diagram illustrating the structure of $N \times N$ Toeplitz and circulant matrices. Toeplitz matrices represent truncated convolution in time, and circulant matrices represent circular convolution in time.

case, we say that the model is **homogeneous**. So for example, a causal prediction model is homogeneous if the prediction filter and the prediction variance do not change with time. In this case, the MMSE causal predictor must be a linear time-invariant filter, so that the predictions are given by

$$\hat{X}_n = \sum_{i=1}^{N} h_{n-i} X_i \ ,$$

where h_n is a **causal prediction filter** and the causal prediction variances take on a constant value of σ_C^2.

When the prediction filter is time invariant, then the prediction matrices, H and A, of equations (1.3) and (1.4) are said to be Toeplitz. A matrix is **Toeplitz** if there is a function, h_n, so that $H_{i,j} = h_{i-j}$. The structure of a Toeplitz matrix is illustrated in Fig. 1.2. Intuitively, each row of a Toeplitz matrix is a shifted version of a single 1-D function, and multiplication by a Toeplitz matrix is essentially time-invariant convolution, but using truncated boundary conditions. Toeplitz matrices arise in many applications, and their spatial structure sometimes presents computational advantages.

In the same way that Toeplitz matrices arise from convolution with truncated boundaries, **circulant matrices** arise when convolution is implemented with circular boundary conditions. In this case, $H_{i,j} = h_{n \bmod N}$ where $n \bmod N$ refers to modulo arithmetic. Multiplication by a circulant matrix, H, is equivalent to circular convolution with the function h_n. Circulant matrices have many useful properties because we know that circular convolution can be implemented with multiplication after taking the **discrete Fourier transform (DFT)** of a signal.

A simple way to get around problems with boundary conditions is to

extend the random process X_n so that $n = -\infty, \cdots, -1, 0, 1, \cdots, \infty$. When the signal has infinite extent, then we can use the standard notation of linear time-invariant systems. In this case,

$$\begin{aligned} \mathcal{E}_n &= X_n - \hat{X}_n \\ &= X_n - X_n * h_n \\ &= X_n * (\delta_n - h_n) \ , \end{aligned} \tag{1.7}$$

where $*$ denotes 1-D discrete-time convolution. Turning things around, we may also write

$$\begin{aligned} X_n &= \mathcal{E}_n + X_n * h_n \\ &= \mathcal{E}_n + \sum_{i=1}^{P} X_{n-i} h_i \ , \end{aligned} \tag{1.8}$$

which is the form of a P^{th} order **infinite impulse response (IIR)** filter.

When $P = \infty$, then all past values of X_n can be used in the MMSE prediction. However if $P < \infty$, then the prediction only depends on the last P observations, and we call X an **autoregressive (AR) random process.** Figure 1.3 shows how the predictor in an order P AR model only depends on the P past neighbors.

Equation (1.8) is sometimes called a **white noise driven model** for the AR process because it has the form of an LTI system with a white noise input, $\mathcal{E}_n$. The white noise driven model is particularly useful because it provides an easy method for generating an AR process, X_n. To do this, one simply generates a sequence of i.i.d. of $N(0, \sigma_C^2)$ Gaussian random variables, and filters them with the IIR filter of equation (1.8). If the IIR filter is stable, then its output, X_n, will be a stationary random process.

We can calculate the autocorrelation of the AR process, X_n, by using the relationship of equation (1.8). Since the prediction errors, $\mathcal{E}_n$, are i.i.d., we know that their time autocorrelation is given by

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

From the results of Chapter **??** and the relationship of equation (1.7), we know that the autocorrelation of X_n obeys the following important relationship

$$R_X(n) * (\delta_n - h_n) * (\delta_n - h_{-n}) = R_{\mathcal{E}}(i-j) = \sigma_C^2 \delta_n \ . \tag{1.9}$$

From equation (1.9) we can calculate the power spectrum of the AR process by computing the DTFT of the time autocorrelation.

$$S_X(\omega) = \frac{\sigma_C^2}{|1 - H(\omega)|^2} \ . \tag{1.10}$$

Example 1.1: Consider the P^{th} order AR random process $X_1, \cdots, X_N$ with prediction variance of σ_C^2 and prediction errors given by

$$\mathcal{E}_n \ = \ X_n - \sum_{i=1}^{P} h_i X_{n-i} \ . \tag{1.11}$$

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

Our task in this example is to compute the joint ML estimate of the prediction filter, h_n, and the prediction variance, σ_C^2. To do this, we first must compute the probability density of X. Using the PDF of equation (1.6), we can write the density for the AR process as

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

We can further simplify the expression by defining the parameter vectors

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

Then the log likelihood of the observations, X, can be written as

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

Using this, we can express the log likelihood as

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

where

$$\begin{aligned}\hat{\sigma}_x^2 &= \frac{1}{N}\sum_{n=1}^{N} X_n^2 \\ \hat{b} &= \frac{1}{N}\sum_{n=1}^{N} Z_n X_n \\ \hat{R} &= \frac{1}{N}\sum_{n=1}^{N} Z_n Z_n^t \ .\end{aligned}$$

Notice that $\hat{\sigma}_x^2$, $\hat{b}$, and $\hat{R}$ are sample statistics taken from the data, X. The three values represent the variance of X_n, the cross-correlation between Z_n and X_n, and the covariance of Z_n, respectively.

First, we compute the ML estimate of the prediction filter by taking the gradient with respect to the filter vector h.

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

where we use the convention that the gradient is represented as a row vector. Setting the gradient of the log likelihood to zero, results in the ML estimate of the prediction filter, h.

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

We can next compute the ML estimate of σ_C^2 by plugging in the expression for the ML estimate of h, and differentiating with respect to the parameter 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 of the log likelihood to zero, results in the expression

$$1 - \frac{1}{\hat{\sigma}_C^2}\left(\hat{\sigma}_x^2 - \hat{b}^t\hat{R}^{-1}\hat{b}\right) = 0$$

Which yields the ML estimate of the causal prediction variance given by

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

Perhaps a more intuitive representation for $\hat{\sigma}_C^2$ is as the average of the squared prediction errors.

$$\begin{aligned}\hat{\sigma}_C^2 &= \frac{1}{N}\sum_{n=1}^{N}\left(X_n - \hat{h}^t Z_n\right)^2 \\ &= \frac{1}{N}\sum_{n=1}^{N}\mathcal{E}_n^2\end{aligned}$$

1.4 2-D Gaussian AR Models

In fact, the analysis of 1-D AR models from the previous sections is easily generalized to a regular grid in 2 or more dimensions. To do this, each lattice point is represented by $s = (s_1, s_2)$, where s is a vector index with each coordinate taking on values in the range 1 to N.

The key issue in generalizing the AR model to 2-D is the ordering of the points in the plane. Of course, there is no truly natural ordering of the points, but a common choice is **raster ordering**, going left to right and top to bottom in much the same way that one reads a page of English text. Using this ordering, the pixels of the image, X_s, may be formed into a vector as

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

and the causal prediction errors, $\mathcal{E}$, may be similarly ordered. In this case, the 2-D causal prediction error is again given by

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

where W is a window of past pixels in 2-D. Typically, this set is given by

$$W = \{r = (r_1, r_2) : 1 \le r_2 \le P \text{ or } (r_2 = 0 \text{ and } 1 \le r_1 \le P)\} \ .$$

Notice that this definition of W only contains previous pixels in raster order. The resulting asymmetric window shown in Figure 1.3 contains a total of $2P(P+1)$ pixels. The window is not symmetric because it is constrained by the raster ordering.

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

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

1-D AR order $P = 2$ 2-D AR order $P = 2$

Figure 1.3: Structure of 1-D and 2-D AR prediction window for an order $P = 2$ models. The pixel denoted by the symbol $\otimes$ is predicted using the past values denoted by the symbol $\bullet$. For the 1-D case, $P = 2$ past values are used by the predictor, and for the 2-D case, $2P(P + 1) = 12$ past values are used by the predictor. Notice that the asymmetric shape of the 2-D prediction window results from raster ordering of the pixels.

$$H = \begin{bmatrix} \boxed{H_0} & \boxed{H_1} & \cdots & \boxed{H_{N-1}} \\ \boxed{H_{-1}} & \boxed{H_0} & \cdots & \boxed{H_{N-2}} \\ \vdots & \vdots & \ddots & \vdots \\ \boxed{H_{1-N}} & \boxed{H_{2-N}} & \cdots & \boxed{H_0} \end{bmatrix} \qquad \boxed{H_k} = \begin{bmatrix} h_{k,0} & h_{k,1} & \cdots & h_{k,N-1} \\ h_{k,-1} & h_{k,0} & \cdots & h_{k,N-2} \\ \vdots & \vdots & \ddots & \vdots \\ h_{k,1-N} & h_{k,2-N} & \cdots & h_{k,0} \end{bmatrix}$$

Figure 1.4: Diagram illustrating the structure of a Toeplitz block Toeplitz matrix. The $N^2 \times N^2$ matrix is made up of blocks which are each of size $N \times N$. Notice that the blocks are organized in a Toeplitz structure, and each block, $\boxed{H_k}$, is itself Toeplitz.

where H is the 2-D prediction matrix.

Since H represents application of a linear space-invariant 2-D filter, it has a special structure and is referred to as a **Toeplitz block Toeplitz** matrix. Figure 1.4 illustrates the structure graphically. Notice that the matrix H is formed by a set of $N \times N$ blocks, H_k, organized in a Toeplitz structure. Each individual block is itself a Toeplitz matrix, which explains the terminology. Formally, the Toeplitz block Toeplitz structure means that the entries of the matrix, $H_{i,j}$, must obey the constraint

$$H_{mN+k,nN+l} = h_{m-n,k-l}$$

where in this case, $h_{i,j}$ is the 2-D prediction filters impulse response. Intuitively, the Toeplitz block Toeplitz results whenever space-invariant filters are represented as matrix operators and the pixels in the image are organized in raster ordering.

The 2-D AR model also has properties quite similar to the 1-D case. In fact, the results are formally identical, only with 1-D convolution being replace by 2-D convolution and the 1-D DTFT being replaced by the 2-D

discrete-space Fourier Transform (DSFT). More specifically,

$$R_X(s) * (\delta_s - h_s) * (\delta_s - h_{-s}) = \sigma_C^2 \delta_s \ , \tag{1.12}$$

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

$$S_X(\mu, \nu) = \frac{\sigma_C^2}{|1 - H(\mu, \nu)|^2} \ . \tag{1.13}$$

where $H(\mu, \nu)$ is the DSFT of h_s, and $S_X(\mu, \nu)$ is the 2-D power spectral density of the AR process.

- Show figure (part b) illustrating example 2-D AR process.
  - Point out anisotropic behavior.
  - Motivate next sections theme of removing causal prediction.

Chapter 1 Problems

1. Let $\{X_n\}_{n=1}^N$ be a 1-D Gaussian random process such that

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

results in $\mathcal{E}_n$ being i.i.d. $N(0, \sigma^2)$ random variables for $n = 1, \cdots, N$, and assume that $X_n = 0$ for $n \leq 0$. Compute the ML estimates of the prediction filter h_n and the prediction variance σ^2.

2. Let X_n be samples of an AR process with order P and parameters (σ^2, h). Also make the assumption that $X_n = 0$ for $n \leq 0$.

a) Use matlab to generate 100 samples of Y. Experiment with a variety of values for P and (σ^2, h). Plot your output for each experiment.

b) Use your sample values of X generated in part a) to compute the ML estimates of the (σ^2, h), and compare them to the true values.

3. Let X be a 1-D AR process with $h_n = \rho\delta_{n-1}$ and prediction variance σ^2.

a) Analytically calculate $S_x(\omega)$, the power spectrum of X, and $R_x(n)$, the autocorrelation function for X.

a) Plot $S_x(\omega)$ and $R_x(n)$ for $\rho = 0.5$ and $\rho = 0.95$.

4. Let X_n be a zero-mean wide sense stationary random process, and define

$$Z_n = \begin{bmatrix} X_{n-1} \\ \vdots \\ X_{n-P} \end{bmatrix}$$

for some fixed order P, and let

$$R = E\left[\frac{1}{N}\sum_{n=0}^{N-1} Z_n Z_n^t\right]$$

a) Show that R is a Toeplitz matrix.

b) Show that R is a positive semi-definite matrix.

5. Consider a LTI system with input x_n, output, y_n, and impulse response, h_n, so that $y_n = h_n * x_n$, where $*$ denotes convolution. Also define the

vectors, $y = [y_0, \cdots, y_{N-1}]^t$, and $x = [x_0, \cdots, x_{N-1}]^t$. Show that if $x_n = 0$ for $n \leq 0$ and $n > N$, then

$$y = Ax$$

where A is a Toeplitz matrix.

6. Consider a linear system with input $\{x_n\}_{n=0}^{N-1}$, output, $\{y_n\}_{n=0}^{N-1}$, and impulse response, $\{h_n\}_{n=0}^{N-1}$, so that $y_n = h_n * x_n$, where $*$ denotes circular convolution. Also define the vectors, $y = [y_0, \cdots, y_{N-1}]^t$, and $x = [x_0, \cdots, x_{N-1}]^t$. Show that if $x_n = 0$ for $n \leq 0$ and $n > N$, then

$$y = Ax$$

where A is a circulant matrix.

7. Let A be an $N \times N$ circulant matrix, so that $A_{i,j} = h_{n \, mod \, N}$, for some real-valued stable function h_n. Furthermore, let

$$T_{m,n} = e^{j\frac{2\pi mn}{N}}$$

be the N dimensional DFT.

a) Show that inverse transformation is given by

$$\left[T^{-1}\right]_{m,n} = \frac{1}{N} e^{j\frac{-2\pi mn}{N}}$$

where T^{-1} is the inverse DFT.

b) Show that $\Lambda = TAT^{-1}$ is a diagonal matrix with entrees given by the DFT of the function h_n. That is $\Lambda = \text{diag}\{\lambda_1, \cdots, \lambda_N\}$ where $\lambda_m = \sum_{n=1}^{N} T_{m,n} h_n$.

c) Show that the logarithm of the determinant of the matrix A is given by

$$\log|A| = \sum_{n=1}^{N} \log|\lambda_n| \ ,$$

where λ_n the N-point DFT of h_n.

d) Show that in the limit at $N \to \infty$,

$$\lim_{N\to\infty} \frac{1}{N} \log|A| = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log|H(\omega)| d\omega \ .$$

where $H(\omega) = \sum_{n=0}^{\infty} h_n e^{j\omega n}$ is the DTFT of h_n.

8. Let $X_{m,n}$ be a zero-mean 2-D AR Gaussian random process with $h_{m,n} = \rho\delta_{m-1} + \rho\delta_{n-1} - \rho^2\delta_{m-1}\delta_{n-1}$ and prediction variance σ^2.

   a) Calculate and expression for $\mathrm{E}\big[|X_{m,n}|^2\big]$, and determine the value of σ^2 so that $\mathrm{E}\big[|X_{m,n}|^2\big] = 1$.

   b) Analytically calculate $S_x(\mu, \nu)$, the power spectrum of X.

   c) Use matlab to generate a 512×512 sample of Y. Use the value of σ^2 from a) so that $\mathrm{E}\big[|X_{m,n}|^2\big] = 1$, and use $\rho = 0.9$.

   c) Repeat part b) for $\rho = 0.5$ and $\rho = 0.98$.

   d) Plot $S_x(\mu, \nu)$ for $\rho = 0.9$.

9. Consider a 2-D LTI system with input $x_{m,n}$, output, $y_{m,n}$, and impulse response, $h_{m,n}$, so that $y_{m,n} = h_{m,n} * x_{m,n}$, where $*$ denotes 2-D convolution. Also define the vectors, $y = [y_{1,1}, \cdots, y_{1,N}, y_{2,1}, \cdots, y_{2,N}, \cdots, y_{N,1}, \cdots, y_{N,N}]^t$, and $x = [x_{1,1}, \cdots, x_{1,N}, x_{2,1}, \cdots, x_{2,N}, \cdots, x_{N,1}, \cdots, x_{N,N}]^t$. Show that if $x_{m,n} = 0$ for $m < 1$ or $m > N$ or $n < 1$ or $m > N$, then

$$y = Ax$$

   where A is a Toeplitz block Toeplitz matrix.

10. **Add problem on 2-D AR processes.** Show that number of autocorrelation lags is greater than number of prediction coefficients +1 (for prediction variance). Show that this implies that the covariance corresponding to a set of ML parameters is not unique.

Chapter 2

Non-Causal Gaussian Models

One disadvantage of AR processes is that their construction depends on a causal ordering of points in time. For many applications, it is completely reasonable to order points into the future, present, and past. For example, in real-time processing of audio signals, this is a very natural organization of the data. But sometimes, our measurements have no natural ordering. For example, the temperatures measured along a road, are a 1-D signal, but the direction of causality is not well defined. Which end of the road should represent the past and which the future?

While this example may seem a bit contrived, the problem of ordering points becomes much more sever for pixels in an image. In practical imaging applications, such as video communications, it is often necessary to impose a raster ordering on pixels in an image, but subsequent 1-D processing of the ordered pixels is likely to produce artifacts aligned with the raster pattern. The objective of this section is to introduce the basic tools that we will need to remove causality from the modeling of images, and thereby avoid the introduction of related artifacts.

2.1 Non-Causal Prediction in Gaussian Models

In order to introduce the concepts of modeling with non-causal prediction, we will start with the case of 1-D signals. Let $X_1, \cdots, X_N$ again be a zero-mean discrete-time Gaussian random process. Rather than use causal prediction, we will attempt to model X_n using predictions based on a combination of

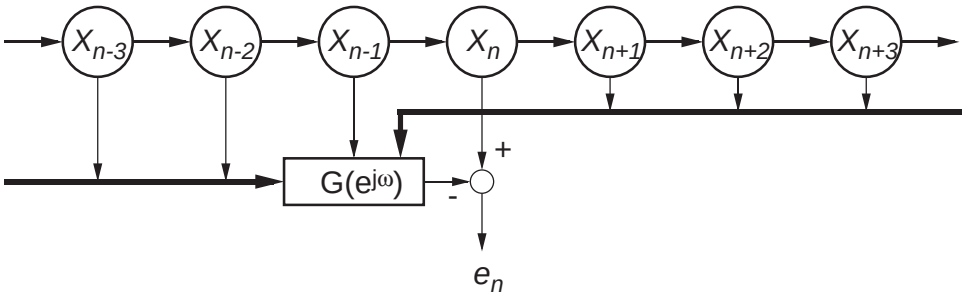

Figure 2.1: Diagram of a non-causal predictor for a P^{th} order 1-D Gaussian Markov random field. The linear time-invariant prediction filter, g_n, is symmetric and has frequency response, $G(\omega)$. In this case, the resulting prediction errors, $\mathcal{E}_n$, are not white when the predictor is optimal.
1.1

past and future information. In this case, the MMSE predictor is

$$\hat{X}_n = \mathrm{E}[X_n | X_i \text{ for } i \neq n]$$

As with the causal predictor, the non-causal predictor is a linear function of the data when the random process is zero-mean and Gaussian. So the non-causal prediction error can be written as

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

where $g_{n,n} = 0$ for all $1 \leq n \leq N$. This condition that $g_{n,n} = 0$ is very important. Otherwise, the value X_n could be used to predict X_n perfectly! In addition, we define σ_n^2 to be the non-causal prediction variance given by

$$\sigma_n^2 = \mathrm{E}\big[\mathcal{E}_n^2 | X_i \text{ for } i \neq n\big] \ .$$

Notice that because X_n is jointly Gaussian, we know from Example **??**.1 that the prediction variance is not a function of X_i for $i \neq n$; however, it may depend on the time index, n.

As in the case of causal predictors, non-causal predictors have a number of important properties. First, the non-causal prediction errors are independent of the values used in prediction.

Property 2.1: *Independence of non-causal Gaussian prediction errors from past and future* - The MMSE non-causal prediction errors for a zero-mean Gaussian random process are independent of the past and future of the random process.

$$\mathcal{E}_n \perp\!\!\!\perp \{X_i\}_{i \neq n} \ .$$

Unfortunately, one very important property is lost when using non-causal prediction. In general, the non-causal prediction errors are no longer uncorrelated and independent. This is a great loss because the independence of the causal prediction errors was essential for the computation of the probability density of the random process in Section 1.2.

2.2 Density Functions Based on Non-Causal Prediction

Although the non-causal prediction errors are not independent, we will still be able to calculate the probability density for X_n by using vector-matrix operations, but with a somewhat different strategy. Since X is a zero-mean Gaussian random vector, it must have a density function with the form

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

where B is the inverse autocorrelation matrix of X. Furthermore, we know that we can write the conditional distribution of X_n given all the remaining X_i for $i \neq n$ as

$$p(x_n | x_i \text{ for } i \neq n) = \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left\{ -\frac{1}{2\sigma_n^2} \left(x_n - \sum_{i=1}^{N} g_{n,i} x_i \right)^2 \right\} \tag{2.1}$$

where σ_n^2 is the non-causal prediction variance for X_n. Our objective is to determine the matrix B in terms of the parameters of the non-causal prediction filter, $g_{n,i}$, and variance, σ_n^2. We can do this by setting the derivatives of the log likelihoods equal for these two densities.

$$\begin{aligned} \frac{d}{dx_n} \log p(x_n | x_i \text{ for } i \neq n) &= \frac{d}{dx_n} \log p(x) \\ \frac{1}{\sigma_n^2} \left(x_n - \sum_{i=1}^{N} g_{n,i} x_i \right) &= \sum_{i=1}^{N} B_{n,i} x_i \end{aligned}$$

Since this relations must hold for all x and all n, the inverse covariance matrix, B, must be given by

$$B_{i,j} = \frac{1}{\sigma_i^2} \left(\delta_{i-j} - g_{i,j} \right) \ .$$

Alternatively, we can use this relationship to compute the MMSE non-causal predictor parameters given a specification of B.

$$\begin{aligned} \sigma_n^2 &= (B_{n,n})^{-1} \\ g_{n,i} &= \delta_{i-j} - \sigma_n^2 B_{n,i} \ . \end{aligned}$$

These relationships can be more compactly represented using matrix notation. If we define the matrix $G_{i,j} = g_{i,j}$ as the **non–causal prediction matrix,** $\Gamma = \text{diag}\left\{\sigma_1^2, \cdots, \sigma_N^2\right\}$ as the diagonal matrix of **non-causal prediction variances**, and $\mathcal{E}$ as the column vector of non-causal prediction errors, then we have that $\mathcal{E} = (I - G)X$, and

$$B = \Gamma^{-1}(I - G)$$

or alternatively

$$\begin{aligned} \Gamma &= \text{diag}(B)^{-1} \\ G &= I - \Gamma B \ . \end{aligned}$$

2.3 1-D Gaussian Markov Random Fields (GMRF)

An important special case occurs when the number of observations needed to determine the MMSE non-causal predictor is limited to a window of $n \pm P$ about the point being predicted. In order to simplify notation, we defined the window

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

so that ∂n is a set containing P neighbors on either side of n, except on the boundary, where it is truncated.

Using this new notation, a **1-D Gaussian Markov random field (GMRF)** is any Gaussian random process with the property that

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

In words, the MMSE non-causal predictor for any pixel in an GMRF is only dependent on the pixel's neighbors. Figure 2.2 illustrates the structure of the prediction window for a 1-D GMRF of order P.

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

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

We should note that the terminology "1-D random field" is clearly an oxymoron, since the term "random field" refers to a 2-D object. Later we will see that the concept of a Markov random field (MRF) grew out of the study of 2 or more dimensional objects, were they are most useful, but the concept applies perfectly well to 1-D also.

In order to reduce the number of model parameters, it is often useful to assume that the prediction coefficients are not a function of position, n. This results in the following new definition.

> *Definition:* A GMRF is said to be **homogeneous** if both the MMSE non-causal predictor and MMSE prediction variance are invariant to position.

The MMSE non-causal predictor 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 non-causal prediction variance, σ^2_{NC}, taking on a constant value. In this case, we can write the density function for the homogeneous GMRF as

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

where

$$B_{i,j} = \frac{1}{\sigma^2_{NC}} \left(\delta_{i-j} - g_{i-j} \right) \ .$$

Since we know that B must be a symmetric matrix, this implies that $g_n = g_{-n}$ must be a symmetric filter.

Once again, if we again extend X_n so that $n = -\infty, \cdots, -1, 0, 1, \cdots, \infty$, then we can express the relation between X_n and the non-causal prediction errors, $\mathcal{E}_n$, using convolution.

$$\mathcal{E}_n = X_n * (\delta_n - g_n) \tag{2.2}$$

and

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

Using this expression, we can compute the cross-correlation between $\mathcal{E}_n$ and X_n as

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

By combining this result with Property 2.1, we get the following expression for the cross-correlation between the prediction error and X_n.

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

This result just indicates that the prediction errors are independent of the values used in the prediction. Using this fact, we have that

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

Rearranging terms results in

$$R_{\mathcal{E}}(n) = \sigma_{NC}^2 \left(\delta_n - g_n\right) \ , \tag{2.3}$$

where $R_{\mathcal{E}}(n)$ is the time autocorrelation of the non-causal prediction errors. So from this we see that, in general, the noncausal prediction errors are not white. From equation (2.2), we know that autocorrelation functions of $\mathcal{E}_n$ and X_n must be related by

$$R_{\mathcal{E}}(n) = R_X(n) * \left(\delta_n - g_n\right) * \left(\delta_n - g_{-n}\right) \ . \tag{2.4}$$

Equating the expressions of (2.3) and (2.4) and using the fact that g_n is a symmetric function of n then yields an important expression for the time autocorrelation of X_n.

$$R_X(n) * (\delta_n - g_n) = \sigma^2_{NC}\delta_n \ . \tag{2.5}$$

From (2.3) and (2.5), we can compute the power spectral density of the non-causal prediction errors.

$$S_{\mathcal{E}}(\omega) = \sigma^2_{NC}\left(1 - G(\omega)\right) \ ,$$

where $G(\omega)$ is the DTFT of g_n, and the power spectral density of the homogeneous GMRF process is given by

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

2.4 2-D Gaussian Markov Random Fields (GMRF)

In fact, all of the derivations of this section are easily generalized to regular grids in 2 or more dimensions. To do this, each lattice point is represented by $s = (s_1, s_2)$, where s is a vector index with each coordinate taking on values in the range 1 to N. We denote the set of all lattice points as $S = [1, \cdots, N]^2$.

To generalize the vector-matrix relationships, we can order the pixels of the vector X in raster order so that

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

and $\mathcal{E}$ is order similarly. If the GMRF is homogeneous, then the 2-D non-causal prediction error is again given by

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

where ∂s is a set of neighbors in 2-D. Typically, this set is given by

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

where P defines a $(2P+1) \times (2P+1)$ window about the point being predicted. Figure 2.2 illustrates the structure of this 2-D prediction window for an order P GMRF.

1-D GMRF order $P = 2$ 2-D GMRF order $P = 2$

Figure 2.2: Structure of 1-D and 2-D GMRF prediction window for an order $P = 2$ models. The pixel denoted by the symbol $\otimes$ is predicted using the past values denoted by the symbol $\bullet$. For the 1-D case, $2P = 4$ past and future values are used by the predictor, and for the 2-D case, $(2P+1)^2 - 1 = 24$ values are used by the predictor.

Again, the prediction errors can be expressed in matrix form as

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

where G is the 2-D prediction matrix. As in the case of the 2-D AR model, the matrix G is Toeplitz block Toeplitz.

The 2-D stationary GMRF also has properties quite similar to the 1-D case. In fact, the results are formally identical, only with 1-D convolution being replace by 2-D convolution and the 1-D DTFT being replaced by the 2-D DSFT. More specifically,

$$R_{\mathcal{E}}(s) = \sigma^2_{NC} \left(\delta_s - g_s\right) \tag{2.6}$$

$$R_X(s) * \left(\delta_s - g_s\right) = \sigma^2_{NC}\delta_s \ , \tag{2.7}$$

where $*$ denotes 2-D convolution of 2-D functions. From this we can express the 2-D power spectral density of both the homogeneous GMRF process and its associated prediction errors.

$$S_X(\mu,\nu) \ = \ \frac{\sigma^2_{NC}}{1 - G(\mu,\nu)} \ . \tag{2.8}$$

$$S_{\mathcal{E}}(\mu,\nu) \ = \ \sigma^2_{NC}\left(1 - G(\mu,\nu)\right) \tag{2.9}$$

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

2.5 Relation Between GMRF and Gaussian AR Models

An obvious question that arises at this point is which model is more general, the AR or GMRF? In other words, is an AR model a GMRF, and visa-versa?

	1-D AR Model	1-D GMRF Model
Model Parameters	σ_C^2 , h_n	σ_{NC}^2 , g_n
Time Auto-Correlation	$R_X(n) * (\delta_n - h_n) * (\delta_n - h_{-n}) = \sigma_C^2 \delta_n$	$R_X(n) * (\delta_n - g_n) = \sigma_{NC}^2 \delta_n$
Power Spectrum	$S_X(\omega) = \dfrac{\sigma_C^2}{\lvert 1 - H(\omega) \rvert^2}$	$S_X(\omega) = \dfrac{\sigma_{NC}^2}{1 - G(\omega)}$

	2-D AR Model	2-D GMRF Model
Model Parameters	σ_C^2 , h_s	σ_{NC}^2 , g_s
Space Auto-Correlation	$R_X(s) * (\delta_s - h_s) * (\delta_s - h_{-s}) = \sigma_C^2 \delta_s$	$R_X(s) * (\delta_s - g_s) = \sigma_{NC}^2 \delta_s$
Power Spectrum	$S_X(\mu, \nu) = \dfrac{\sigma_C^2}{\lvert 1 - H(\mu, \nu) \rvert^2}$	$S_X(\mu, \nu) = \dfrac{\sigma_{NC}^2}{1 - G(\mu, \nu)}$

Table 2.1: Auto-correlation and power spectrum relationships for 1-D and 2-D models.

In order to answer this question, we can relate the autocorrelation functions and power spectrums for stationary AR and GMRF models. Table 2.1 summarizes the important relationships from the previous chapter.

We know that if the autocorrelation of the stationary AR and GMRF processes are the same, then they must have the same distribution. So equating the expressions for $R_X(n)$ for the 1-D AR case and the 1-D GMRF case, we find the following important relationship.

$$\sigma_{NC}^2(\delta_n - h_n) * (\delta_n - h_{-n}) = \sigma_C^2(\delta_n - g_n) \tag{2.10}$$

If we are given a specific AR process, we can use this expressions to compute the parameters of the associated GMRF process. If we evaluated the equation for $n = 0$, we get a general relationship between the causal and non-causal prediction error.

$$\sigma_{NC}^2 = \frac{\sigma_C^2}{1 + \sum_{n=1}^{P} h_n^2}$$

Using this relationship, we have that

$$g_n = \delta_n - \frac{(\delta_n - h_n) * (\delta_n - h_{-n})}{1 + \sum_{n=1}^{P} h_n^2} \ .$$

So if we select an order P AR model, then the resulting GMRF is given by

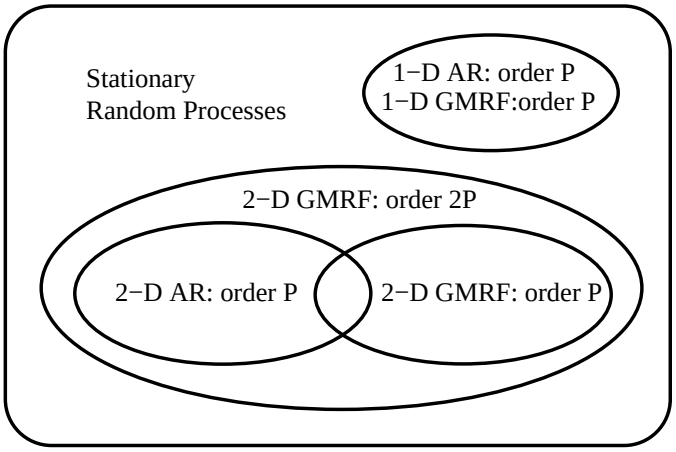

Figure 2.3: Venn diagram illustrating the relationship between AR and GMRF models in 1 and 2-D. Notice that in 1-D, AR and GMRF models are equivalent, but in 2-D they are not.

the auto-correlation of the function $\delta_n - h_n$.[1] This auto-correlation operation is shown graphically in Fig. 2.4. Notice, that in 1-D, and order P AR model results in an order P GMRF. You should be able to convince yourself of this by working out a simple case of $P = 2$.

In order to find the parameters of an AR model from the parameters of a GMRF, it is necessary to find a causal predictor, h_n, so that (2.10) holds in either 1-D or 2-D. In the 1-D case, this is a classic problem that has been solved for the case of Weiner filtering. Because g_n is a symmetric function, it is always possible to factor its rational Z-transform into a product of causal and anti-causal parts. However, this result can not be generalized to 2-D because polynomials in more than one dimension can not, in general, be factored. This leads to the following result.

Property 2.2: *Equivalence of AR and GMRF models in 1-D* - A stationary discrete time zero-mean Gaussian random process is a 1-D order P AR model if and only if it is 1-D order P GMRF.

Interesting, this equivalence relationship does not hold in 2-D. First, Fig. 2.4 shows how a 2-D AR model of order P produces a 2-D GMRF of order $2P$. This is because the resulting 2-D convolution of the prediction filter produces a asymmetric function which is $4P + 1$ wide and $2P + 1$ high. However, the converse relationship simply no longer holds.

Property 2.3: *In 2-D, an AR model of order P is a GMRF of order $2P$*

[1]The **auto-correlation** of a function f_n is defined as $f_n * f_{-n}$, i.e. the convolution with its time reverse.

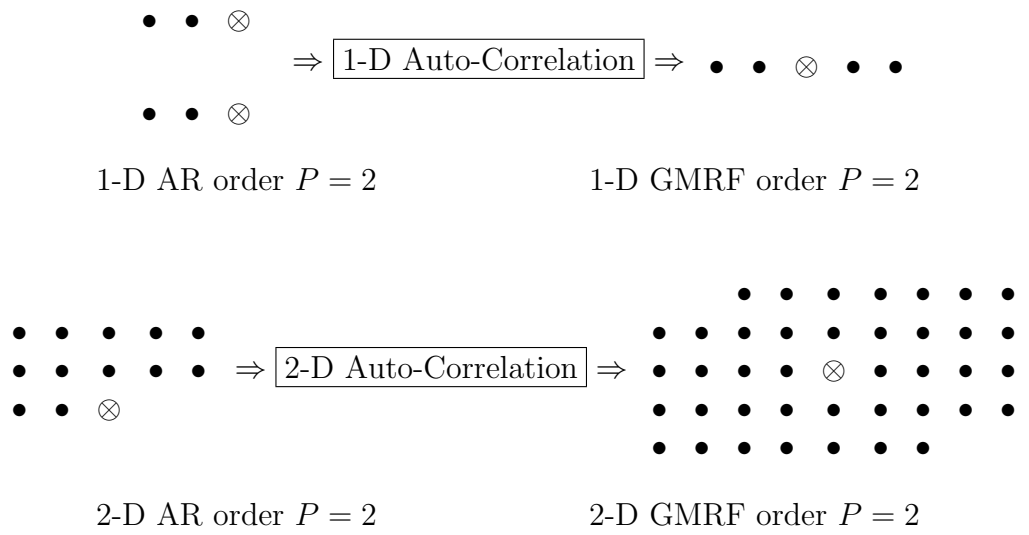

Figure 2.4: Relationship between an AR model and the corresponding GMRF in both 1-D and 2-D. In both 1-D and 2-D, the GMRF prediction window is produced by the auto-correlation the AR filter. Notice that in 1-D, an order P AR model produces an order P GMRF. However, in 2-D and order P AR model produces a GMRF which requires an order of $2P$. In each case, the pixels denoted by the symbol $\otimes$ are predicted using the values denoted by the symbol $\bullet$.

- In 2-D, a stationary discrete time zero-mean order P AR model is also an order $2P$ GMRF.

So from this we see that GMRFs are more general than AR models in 2-D, but not in 1-D. This means that 2-D GMRFs can be used to model a broader class of distributions, which is another major justification for their use.

Example 2.1: Consider a zero-mean stationary AR process with order $P = 1$, and prediction filter

$$h_n = \rho\delta_{n-1} \ .$$

where $|\rho| < 1$, and prediction variance σ_C^2.

From this AR model, we would like to calculate the parameters of an equivalent GMRF. The non-causal prediction variance is given by

$$\sigma_{NC}^2 = \frac{\sigma_C^2}{1+\rho^2} \ .$$

Notice, that this non-causal prediction variance is always smaller than the causal prediction variance. The corresponding non-causal prediction filter is

given by

$$\begin{aligned} g_n &= \delta_n - \frac{(\delta_n - h_n) * (\delta_n - h_{-n})}{1+\rho^2} \\ &= \frac{\rho}{1+\rho^2}\left(\delta_{n-1} + \delta_{n+1}\right) \ . \end{aligned}$$

From this we can also calculate the power spectrum for both the AR and GMRF processes as

$$\begin{aligned} S_X(\omega) &= \frac{\sigma_C^2}{\left|1 - H(\omega)\right|^2} = \frac{\sigma_C^2}{\left|1 - \rho e^{-j\omega}\right|^2} \\ &= \frac{\sigma_C^2}{(1+\rho^2)} \frac{1}{\left(1 - \frac{2\rho}{1+\rho^2}\cos(\omega)\right)} \ , \end{aligned}$$

or equivalently using the GMRF power spectrum

$$\begin{aligned} S_X(\omega) &= \frac{\sigma_{NC}^2}{1 - G(\omega)} = \frac{\sigma_{NC}^2}{1 - g_1 e^{-j\omega} + g_1 e^{j\omega}} \\ &= \frac{\sigma_{NC}^2}{1 - 2g_1\cos(\omega)} \\ &= \frac{\sigma_C^2}{(1+\rho^2)} \frac{1}{\left(1 - \frac{2\rho}{1+\rho^2}\cos(\omega)\right)} \ . \end{aligned}$$

This verifies that the two models result in the same power spectral density.

2.6 GMRF Models on General Lattices

Now that we have seen the GMRF in 1-D, we can take a slightly more abstract approach and develop a general formulation of the GMRF which is applicable to observations indexed on any lattice. To do this, we consider a random process, X_s, indexed on a finite set of lattice points $s \in S$.

The neighbors of a pixel, s, are again denoted by ∂s, but we have made no specific assumptions regarding the structure of S, or the neighbors of a pixel, ∂s. In fact, our specification of neighbors must only meet the two constraints stated in the following definition.

Definition: For each $s \in S$, let $\partial s \subset S$. Then we say that ∂s is a **neighborhood system** if it meets two conditions. First, for all $s \in S$, $s \notin \partial s$. Second, for all $s, r \in S$, if $s \in \partial r$, then $r \in \partial s$.

So for ∂s to be a legitimate neighborhood system, s can not be a neighbor of itself, and if s is a neighbor of r, then r must be a neighbor of s. Cliques are another very important concept that are very closely related to neighborhoods.

Definition: An unordered pair of pixels $\{s, r\}$ with $s, r \in S$ is said to be a **pair-wise clique** if $s \in \partial r$. We denote the set of all pair-wise cliques as

$$\mathcal{P} = \{\, \{s, r\} | s, r \in S \text{ and } s \in \partial r \} \ .$$

Notice that by convention, the pixel pair $\{s, r\}$ is unordered, so each unique pair only appears once in the set $\mathcal{P}$.

Using this concept of neighbors, we may give a formal definition for GMRFs.

Definition: Let X_s be a jointly Gaussian random process indexed on $s \in S$. Then we say that X is a **Gaussian Markov random field (GMRF)** with neighborhood system ∂s, if for all $s \in S$

$$\mathrm{E}[X_s | X_r \text{ for } r \neq s] = \mathrm{E}[X_s | X_r \text{ for } r \in \partial s] \ .$$

So a GMRF is a Gaussian random process X_s such that the non-causal predictor is only dependent on neighboring values, X_r for $r \in \partial s$.

Using this new and somewhat more general formulation, we can restate the results of Section 2.2 that relate the non-causal prediction parameters of a zero-mean GMRF to the parameters of its density function. So the equations of non-causal prediction become

$$\begin{aligned} \mathcal{E}_s &= X_s - \sum_{r \in \partial s} g_{s,r} X_r \\ \sigma_s^2 &= \mathrm{E}\big[\mathcal{E}_s^2 | X_r \text{ for } r \in \partial s\big] \end{aligned}$$

or in vector-matrix form this is written as

$$\begin{aligned} \mathcal{E} &= (I - G)X \\ \Gamma &= \mathrm{diag}\left(\mathrm{E}\big[\mathcal{E}\mathcal{E}^t\big]\right) \end{aligned}$$

With this notation, we can define the following general property of zero-mean GMRFs.

Property 2.4: *Density of a zero-mean GMRF from non-causal prediction parameters* - The density function of a zero-mean Gaussian random process, X_s for $s \in S$, with a non-causal prediction matrix G and positive define non-causal prediction variance matrix, Γ, is given by

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

where the inverse covariance, B, is given by

$$B \ = \ \Gamma^{-1}(I - G) \ ,$$

or equivalently, the non-causal prediction parameters are given by

$$\begin{aligned} \Gamma \ &= \ \text{diag}(B)^{-1} \\ G \ &= \ I - \Gamma B \ . \end{aligned}$$

The relationships of Property 2.2 can also be written more explicitly as

$$\begin{aligned} B_{s,r} \ &= \ \frac{1}{\sigma_s^2} \left(\delta_{s-r} - g_{s,r} \right) \\ \sigma_s^2 \ &= \ \frac{1}{B_{s,s}} \\ g_{s,r} \ &= \ \delta_{s-r} - \sigma_s^2 B_{s,r} \ . \end{aligned}$$

Chapter 2 Problems

1. Find the error in the proof of Property 1.3 when the predictor is assumed to be non-causal, i.e. $\hat{X}_i = \mathrm{E}[X_k \; k \neq i]$.

2. Let X_n be a 1-D zero-mean Gaussian AR process with MMSE causal predition filter given by $h_n = \rho\delta_{n-1}$ and causal prediction variance σ_c^2.

   a) Calculate, $S_X(\omega)$, the power spectral density of the random process.

   b) Calculate, $R_X(n)$, the time auto-correlation of the random process.

   c) Calculate (σ_{NC}^2, g_n) the noncausal prediction variance and the non-causal prediction filter for the equivalent GMRF.

3. Let $\{X_n\}_{n=1}^5$ be a zero-mean 1-D order 2 GMRF.

   a) A friend tells you that the non-causal prediction variance for X_n is $\sigma_n^2 = n$ and non-causal prediction filter is $g_n = \frac{1}{4}(\delta_{n-1} + \delta_{n+1})$. Is this possible? If so, why? If not, why not?

   b) If you know that

$$B_{n,m} = \sqrt{nm}\left(\delta_{n-m} - \frac{1}{4}\left(\delta_{n-m-1} + \delta_{n-m+1}\right)\right) \; ,$$

   then calculate the non-causal prediction variance σ_n^2 and the non-causal prediction filter $g_{m,n}$ for X.

4. Let X_n be a zero-mean GMRF with prediction filter g_n and prediction variance σ_{NC}^2. Can g_n be any symmetric function? If not, what properties must g_n have?

5. Let X_n be a 1-D Gaussian MRF with noncausal predictor g_n. Prove that $\Sigma_n \, g_n < 1$.

6. Let $G(\omega)$ be the DTFT of the non-causal prediction filter, g_n, for a homogeneous GMRF. Prove that $G(\omega)$ is real valued, with $G(\omega) \leq 1$.

7. Let $Y_{m,n}$ be a 2-D zero-mean wide sense stationary random process with autocorrelation function $R(k,l) = E[Y_{m,n}Y_{m+k,n+l}]$.
   a) Show that $\forall k, l$, $R(k,l) = R(-k,-l)$
   b) Give an example of a random process for which it is false that $\forall k, l$, $R(k,l) = R(-k,l)$. (Hint this is equivalent to $\exists k, l$, $R(k,l) \neq R(-k,l)$.)

8. Let Y_s be a 2-D Gaussian AR process indexed by $s = (s_1, s_2)$ where s_1 is the column index and s_2 is the row index. Let the MMSE causal prediction filter be given by

$$h_s = \rho\delta_{s_1-1,s_2} + \rho\delta_{s_1,s_2-1}$$

and the causal prediction variance be given by σ_C^2. Compute (σ_{NC}^2, g_s) the noncausal prediction variance and the noncausal prediction filter.

Bibliography