Chapter 4
Spectral Estimation

Contents

4.1 Introduction

This chapter studies a very important estimation problem, which is that of estimating the power spectral density (PSD) of a stationary process. We will consider two families of estimators: 1) classical (or non-parametric) and 2) parametric estimators, which are based on a model for the PSD.

Computing the estimate of $S_x(e^{j\omega})$, which we will denote by $\hat{S}_x(e^{j\omega})$, from an arbitrarily large number of realizations of a stationary process (see Figure 4.1) would be a (relatively) easy task. Of course, this is an idealized scenario as we do not have access to all realizations and, even more, we also do not have access to all time samples of the same realization. Thus, the objective in this chapter is to compute $\hat{S}_x(e^{j\omega})$ from N samples of a single realization of the process $x[n]$.

The spectral estimation problem is defined only for wide-sense stationary (WSS) processes for which the mean function is time-independent, that is, $\mu_x = \mu_x[n] = \mathbb{E}[x[n]]$, and the auto-correlation function depends only on the time difference, i.e., $r_x[m] = r_x[n, n-m] = \mathbb{E}[x[n]x^*[n-m]]$. For non-stationary processes, the usual practice is to apply the estimators to small windows, since on a local scale we can assume that non-stationary processes are WSS. For instance, this is typically done when analyzing speech signals, which are usually described using non-stationary processes. Moreover, since only one realization is available, the process must be ergodic such that expectations can be substituted by time averages.

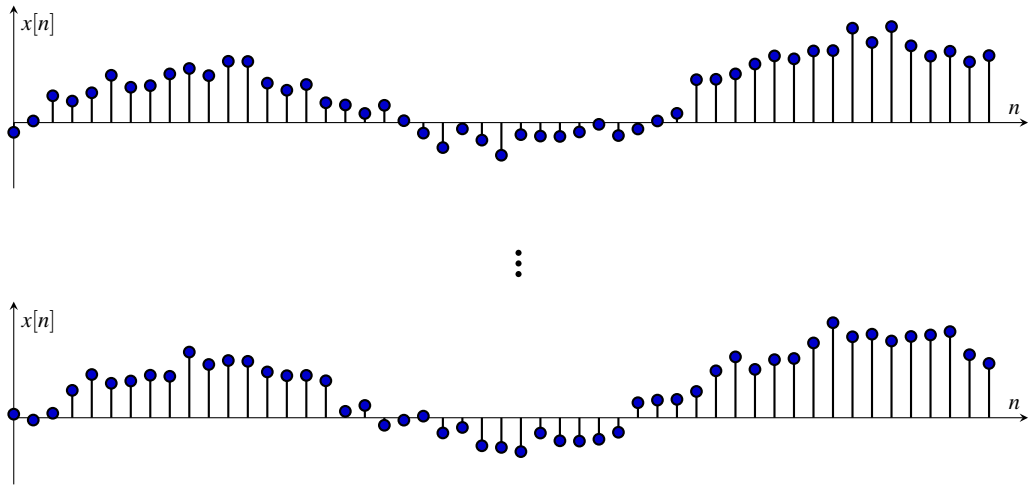

Fig. 4.1 Realizations of a discrete stochastic process

4.2 Preliminaries: Spectral analysis of deterministic signals

Before going into the spectral analysis of stochastic processes, it is convenient to study the case of deterministic signals, which will help us to understand the concept of spectral

resolution. Thus, the problem is to compute the Fourier transform of the deterministic signal $x[n]$. However, this relatively "simple" task has two problems. First, we do not have access to the whole signal $x[n]$, but only to a finite record thereof.

$$x_w[n] = \begin{cases} x[n], & n = 0, \ldots, N-1, \\ 0, & \text{otherwise.} \end{cases}$$

Defining now the window

$$w_{R,N}[n] = \begin{cases} 1, & n = 0, \ldots, N-1, \\ 0, & \text{otherwise,} \end{cases}$$

we may rewrite $x_w[n] = w_{R,N}[n]x[n]$, which allows us to compute the Fourier transform of $x_w[n]$ as[1]

$$X_w(e^{j\omega}) = \mathscr{F}(x_w[n]) = \sum_{n=0}^{N-1} x_w[n]e^{-j\omega N} = \frac{1}{2\pi} W_{R,N}(e^{j\omega}) \circledast X(e^{j\omega}), \tag{4.1}$$

where $\circledast$ denotes the circular convolution. So, the Fourier transform of the windowed signal, $x_w[n]$, is related to that of $x[n]$ through the Fourier transform of the window $w_{R,N}[n]$, which is given by

$$W_{R,N}(e^{j\omega}) = e^{-j\omega(N-1)/2} \frac{\sin\left(\frac{\omega N}{2}\right)}{\sin\left(\frac{\omega}{2}\right)} = e^{-j\omega(N-1)/2} P_N(e^{j\omega}),$$

and its amplitude $|P_N(e^{j\omega})|$ is depicted in Figure 4.2. As this figure shows, the width of the main lobe is $4\pi/N$.

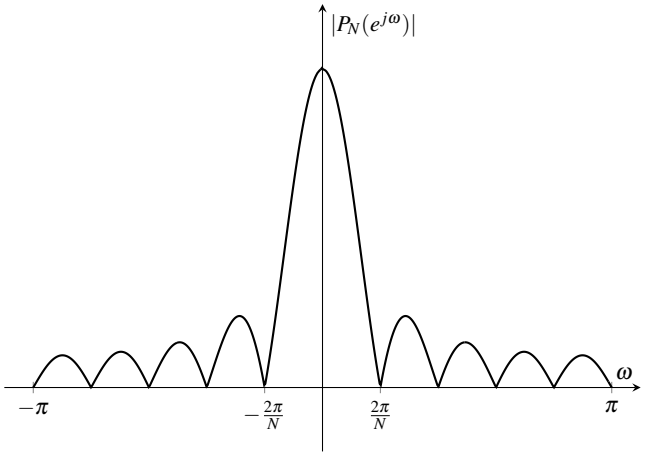

Fig. 4.2 Module of the Fourier transform of the rectangular window

[1] In the "Signals and Systems" parlance, this Fourier transform is named Discrete Time Fourier Transform (DTFT).

The second issue is that the DTFT in (4.1) is a function of a continuous variable. Hence it cannot be computed nor stored in a computer. The solution is simple and consists in discretizing the spectrum, which yields the Discrete Fourier Transform (DFT). Thus, we are only able to compute $X_w(e^{j\omega_k})$, with $\omega_k = 2\pi k/N$ and $k = 0,\ldots,N-1$. The DFT is typically computed using the fast Fourier transform (FFT) algorithm.

The aforementioned procedure based on the DFT/FFT gets only N samples of the spectrum for length-N signals, but we can get more samples by zero-padding the signals, i.e., by simply adding $N_{\text{fft}} - N$ zeros after the N samples. This procedure increases the number of frequencies but it does not increase the resolution as it does not modify the window.

Example 4.1 (Spectral analysis of a complex exponential)

This example considers the spectral analysis of a finite record of a complex exponential, i.e., $x[n] = e^{j\omega_0 n}, n = 0,\ldots,N-1$. Using the DTFT of a complex exponential, given by

$$X(e^{j\omega}) = 2\pi\delta(\omega - \omega_0),$$

and $W_N(e^{j\omega})$, $X_w(e^{j\omega})$ becomes

$$X_w(e^{j\omega}) = e^{-j(\omega-\omega_0)(N-1)/2}\frac{\sin\left(\frac{(\omega-\omega_0)N}{2}\right)}{\sin\left(\frac{\omega-\omega_0}{2}\right)} = e^{-j(\omega-\omega_0)(N-1)/2}P_N\left(\omega-\omega_0\right),$$

and its magnitude squared is

$$|X_w(e^{j\omega})|^2 = \left|P_N\left(e^{j(\omega-\omega_0)}\right)\right|^2.$$

Figure 4.3 plots, in logarithmic scale, $|X_w(e^{j\omega})|^2$ and $|X_w(e^{j\omega_k})|^2$ for two different values of N_{fft}.

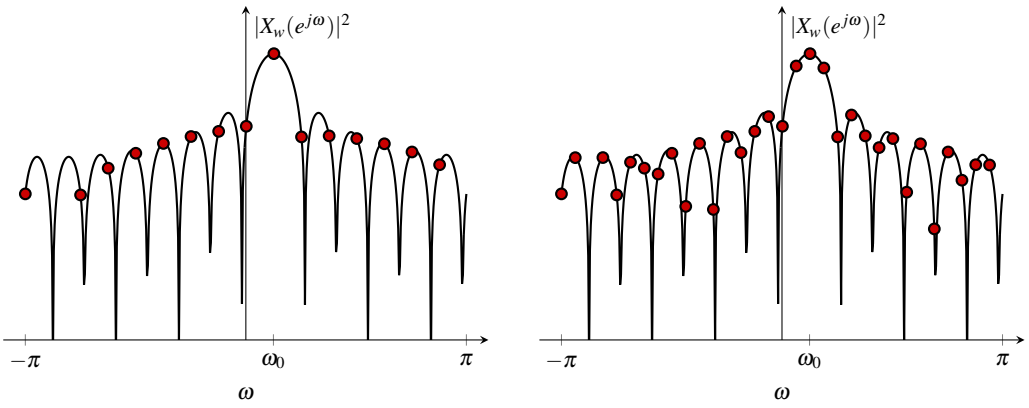

Fig. 4.3 Fourier transform (in logarithmic scale) of a windowed complex exponential

As we have seen in this example, the spectral analysis of deterministic signals depends on two factors. First, the number of available samples, which determines the window and,

therefore, the shape of the windowed spectrum. As we have seen in Figure 4.3, the rectangular window has a narrow main lobe at the expense of high secondary lobes. This effect could be reduced by pre-multiplying $x_w[n]$ by a different window, which would reduce the height of the secondary lobes, but it would widen the main lobe. Second, the number of frequencies where the estimated PSD is evaluated.

Example 4.2 (Spectral analysis of two complex exponentials)

This example considers the spectral analysis of a finite record of the sum of two complex exponential, i.e., $x[n] = e^{j\omega_0 n} + e^{j\omega_1 n}, n = 0, \ldots, N-1$, which will help us to understand the concept of resolution. Using (4.1), we have

$$X_w(e^{j\omega}) = e^{-j(\omega-\omega_0)(N-1)/2} P_N(\omega - \omega_0) + e^{-j(\omega-\omega_1)(N-1)/2} P_N(\omega - \omega_1).$$

and its magnitude squared is

$$|X_w(e^{j\omega})|^2 = \left|P_N\left(e^{j(\omega-\omega_0)}\right)\right|^2 + \left|P_N\left(e^{j(\omega-\omega_1)}\right)\right|^2 + 2\cos\left(\frac{(\omega_0-\omega_1)(N-1)}{2}\right) P_N\left(e^{j(\omega-\omega_0)}\right) P_N\left(e^{j(\omega-\omega_1)}\right).$$

Figure 4.4 plots, in logarithmic scale, $|X_w(e^{j\omega})|^2$ for two different separations between the frequencies of the exponential components. As we can see in this figure, for small frequency separations, it is impossible to identify in the spectrum the two complex exponentials.

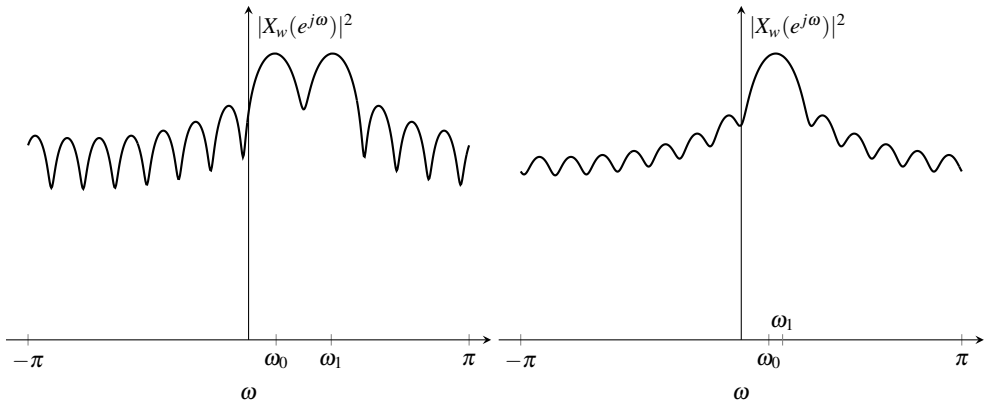

Fig. 4.4 Fourier transform (in logarithmic scale) of the sum of two complex exponentials

4.3 Non-parametric methods in spectral estimation

In this section, we turn our attention to the case of stochastic signals and, in particular, to the development of non-parametric spectral estimation methods. We will therefore study the periodogram and variations thereof.

Before proceeding, let us note that throughout this section, we will only consider DTFTs. However, we have to keep in mind that, in practice, we can only compute the DFT (using the FFT algorithm), as we have seen in Section 4.2.

Remind that the power spectral density, or power spectrum, of a stochastic process $x[n]$, is defined as

$$S_x(e^{j\omega}) = \lim_{N\to\infty} \frac{1}{2N-1} \mathbb{E}\left[\left| \sum_{n=-N+1}^{N-1} x[n] e^{-j\omega n} \right|^2 \right]. \tag{4.2}$$

If the process is WSS and the autocorrelation is absolutely summable (the usual case in practice), this definition is equivalent to the Fourier transform of the autocorrelation, i.e.,

$$S_x(e^{j\omega}) = \mathscr{F}(r_x[m]), \tag{4.3}$$

where

$$r_x[m] = \mathbb{E}[x[n]x^*[n-m]], \tag{4.4}$$

These alternative expressions for the power spectrum motivate two different strategies for spectral estimation:

1. Drop the limit in (4.2) and estimate the expectation from an finite sample.
2. Estimate the autocorrelation function and compute its Fourier transform.

Both strategies are closely related, as we will see in the following sections.

4.3.1 The periodogram and the correlogram

The **periodogram**, which is a term coined by Arthur Schuster in 1898, is obtained from on (4.2) by simply dropping the expectation and considering a finite number of samples, i.e.,

$$\hat{S}_x^p(e^{j\omega}) = \frac{1}{N} \left| \sum_{n=0}^{N-1} x[n] e^{-j\omega n} \right|^2 = \frac{1}{N} \left| X_N(e^{j\omega}) \right|^2, \tag{4.5}$$

where $X_N(e^{j\omega}) = \mathscr{F}(x_N[n])$ is the Fourier transform of the truncated version of $x[n]$

$$x_N[n] = \begin{cases} x[n], & 0 \le n \le N-1 \\ 0, & \text{otherwise} \end{cases} \tag{4.6}$$

The **correlogram**, based on (4.3), is defined as

$$\hat{S}_x^c(e^{j\omega}) = \mathscr{F}(\hat{r}_x[m]),$$

where

$$\hat{r}_x[m] = \frac{1}{N} \sum_{n=m}^{N-1} x[n]x^*[n-m], \quad m = 0,\ldots,N-1, \tag{4.7}$$

Despite the differences in names and definitions, the correlogram and the periodogram are identical estimators: rewriting $\hat{r}_x[m]$ as

$$\hat{r}_x[m] = \frac{1}{N}\sum_{n=m}^{N-1} x[n]x^*[n-m] = \frac{1}{N}\sum_{n=-\infty}^{\infty} x_N[n]x_N^*[n-m] = \frac{1}{N}x_N[n] * x_N^*[-n],$$

and taking its Fourier transform yields

$$\hat{S}_x^c(e^{j\omega}) = \mathscr{F}(\hat{r}_x[m]) = \frac{1}{N}\mathscr{F}\left(x_N[n] * x_N^*[-n]\right).$$

Finally, applying the properties of the Fourier transform, $\hat{S}_x^c(e^{j\omega})$ simplifies to

$$\hat{S}_x^c(e^{j\omega}) = \frac{1}{N}\mathscr{F}\left(x_N[n]\right)\mathscr{F}\left(x_N^*[-n]\right) = \frac{1}{N}X_N(e^{j\omega})X_N^*(e^{j\omega}) = \frac{1}{N}\left|X_N(e^{j\omega})\right|^2 = \hat{S}_x^p(e^{j\omega}),$$

which is the periodogram in (4.5).

4.3.1.1 Bias and variance of the periodogram

To understand why we need more refined estimators of the power spectral density, now we shall perform the statistical analysis of the periodogram (or correlogram), i.e., we will compute its bias and variance as we would do with any other estimator.

First, note that the autocorrelation estimate in (4.3.1) is biased:

$$\begin{aligned}\mathbb{E}\left[\hat{r}_x[m]\right] &= \frac{1}{N}\sum_{n=m}^{N-1}\mathbb{E}\left[x[n]x^*[n-m]\right] = \frac{1}{N}\sum_{n=m}^{N-1} r_x[m] \\ &= \frac{N-|m|}{N}r_x[m], \end{aligned} \tag{4.8}$$

Note that, for $m = 0$ or any m such that $r_x[m] = 0$, the autocorrelation estimate is unbiased. However,for any other values, $\mathbb{E}\left[\hat{r}_x[m]\right] \neq r_x[m]$, and the relative value of the bias increases with m.

Seemingly, we can easily remove this bias by replacing the factor $1/N$ in (4.7) by $1/(N-|m|)$. In this way, we would get unbiased estimates of the autocorrelation values for any m. Unfortunately, it can be show that the resulting function is, in general, not a feasible autocorrelation function (i.e. its Fourier transform may take negative values at some frequencies).

As a consequence of the autocorrelation bias, the periodogram is also biased. To see it, it is useful to define the triangular, or Barlett window, as

$$w_{T,N}[m] = \begin{cases} \frac{N-|m|}{N}, & |m| \leq N-1, \\ 0, & \text{otherwise}, \end{cases}$$

which is depicted in Figure 4.5.

The mean of the autocorrelation in (4.8) can then be written as

$$\mathbb{E}\left[\hat{r}_x^b[m]\right] = \frac{N-|m|}{N}r_x[m] = w_{T,N}[m]r_x[m], \tag{4.9}$$

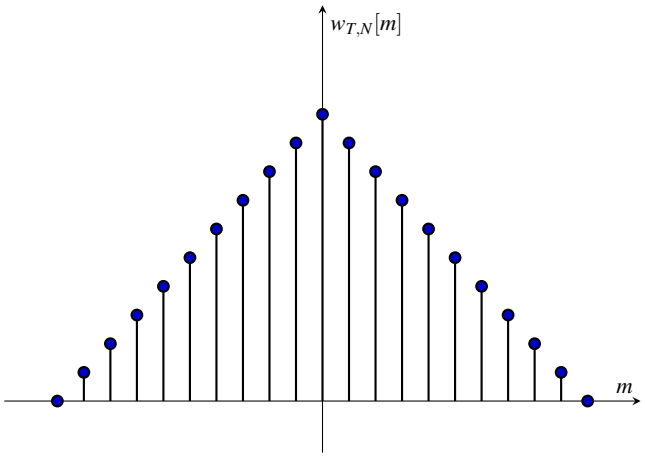

Fig. 4.5 Triangular window

Using (4.9), the bias of the periodogram becomes

$$\mathbb{E}\left[\hat{S}_x^p(e^{j\omega})\right] = \mathscr{F}\left(w_{T,N}[m]r_x[m]\right) = \frac{1}{2\pi}W_{T,N}(e^{j\omega}) \circledast S_x(e^{j\omega}), \tag{4.10}$$

where

$$\begin{aligned} W_{T,N}(e^{j\omega}) &= \mathscr{F}\left(w_{T,N}[m]\right) = \frac{1}{N}\mathscr{F}\left(w_{R,N}[m] * w_{R,N}[-m]\right) = |W_{R,N}(e^{j\omega})|^2 \\ &= \frac{1}{N}\frac{\sin^2\left(\frac{\omega N}{2}\right)}{\sin^2\left(\frac{\omega}{2}\right)}, \end{aligned}$$

is the Fourier transform of the triangular window and is depicted in Figure 4.6. Comparing Figures 4.2 and 4.6, it can be seen that the level of secondary lobes is smaller for the triangular window. By analogy with Example 4.2, we can say that the bias of the periodogram is related with its resolution.

Note that the width of the main lobe is inversely proportional to N. Therefore, by increasing the sample size we can narrow the main lobe and reduce the bias. This is because the periodogram is asymptotically unbiased: the bias reduces to zero as N grows.

$$\lim_{N\to\infty} \hat{S}_x^p(e^{j\omega}) = S_x(e^{j\omega}).$$

As a special case, note that the autocorrelation function of a white process is zero for all $m \neq 0$, therefore, the estimate of its autocorrelation is unbiased, and so the periodogram.

The analysis of the variance of the periodogram is cumbersome and can only be done in particular cases. For white noise, it can be shown that

$$Var\left(\hat{S}_x^p(e^{j\omega})\right) = S_x^2(e^{j\omega}),$$

and in general we can say that

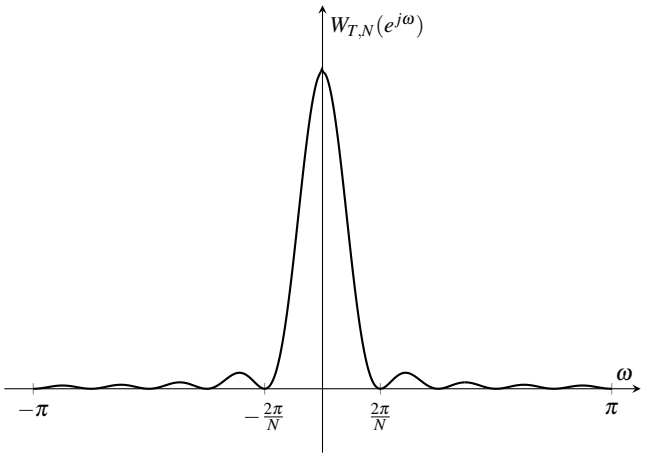

Fig. 4.6 Fourier transform of the triangular window

$$Var\left(\hat{S}_x^p(e^{j\omega})\right) \lessapprox S_x^2(e^{j\omega}),$$

where $\lessapprox$ denotes approximately proportional to. This expression tells us that the variance does not decrease for larger data records. That is, the periodogram is not a consistent estimate of the PSD.

4.3.2 The Blackman-Tukey estimator

One of the reasons for the behavior of the periodogram variance is the poor quality of the estimate $\hat{r}_x[m]$ for values of m close to N. This problem is what the Blackman-Tukey (BT) estimator tries to improve. The idea is to ignore or weight the samples of $\hat{r}_x[m]$ for m close to N. Thus, the BT estimator is

$$\hat{S}_x^{BT}(e^{j\omega}) = \mathscr{F}(w_M[m]\hat{r}_x[m]) = \sum_{m=-N+1}^{N-1} w_M[m]\hat{r}_x[m]e^{-j\omega m}, \tag{4.11}$$

where $w[m]$ is a window that must fulfill

$$w_M[m] = \begin{cases} f(|m|), & |m| \leq M-1, \\ 0, & \text{otherwise}, \end{cases}$$

where $f(|m|)$ is a monotonically decreasing function of $|m|$ and $M \leq N$. This window ignores the lags of the estimated auto-correlation for $|m| > M-1$ and weights the lags for large m. The choice of the window is critical to achieve good performance, but, in any case, it must guarantee that $\hat{S}_x^{BT}(e^{j\omega}) \geq 0$.

Using the properties of the Fourier transform, we may rewrite $\hat{S}_x^{BT}(e^{j\omega})$ as

$$\hat{S}_x^{BT}(e^{j\omega}) = \frac{1}{2\pi} W_M(e^{j\omega}) \circledast \hat{S}_x^p(e^{j\omega}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} W_M(e^{j\psi}) \hat{S}_x^p(e^{j(\omega-\psi)}) d\psi,$$

where $W_M(e^{j\omega}) = \mathscr{F}(w_M[n])$. Then, the Blackman-Tukey estimator is locally smoothing the periodogram, which reduces its variance. However, there is no free lunch and we will show that this variance reduction translates into lower resolution (or larger bias). Concretely, the bias of the BT estimator is

$$\mathbb{E}\left[\hat{S}_x^{BT}(e^{j\omega})\right] = \frac{1}{2\pi} W_M(e^{j\omega}) \circledast \mathbb{E}\left[\hat{S}_x^p(e^{j\omega})\right] = \frac{1}{2\pi} W_M(e^{j\omega}) \circledast W_{T,N}(e^{j\omega}) \circledast S_x(e^{j\omega}).$$

Finally, since $w_M[n]$ is shorter than $w_{T,N}[n]$, it can be shown that $W_M(e^{j\omega})$ is wider than $W_{T,M}(e^{j\omega})$, which translates into a lower resolution. This behavior is depicted in Figure 4.7 for $w_M[n] = w_{T,M}[n]$. Note that the y-axis is in logarithmic scale.

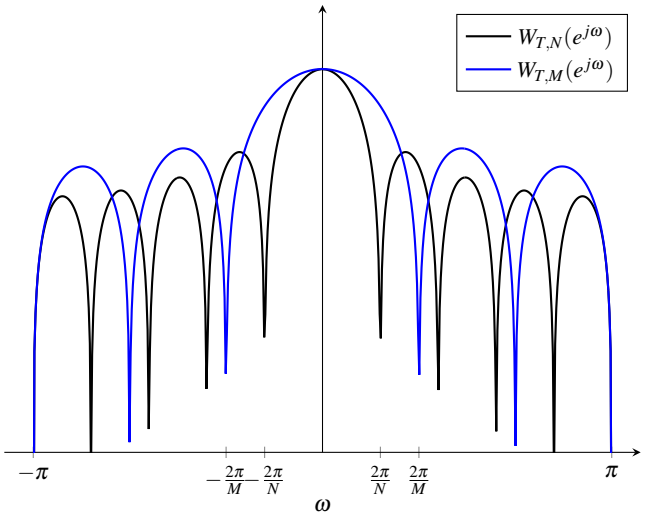

Fig. 4.7 Fourier transform (in logarithmic scale) of two triangular windows of different lengths

4.3.3 Estimators based on the averaged periodogram

The Blackman-Tukey estimator yields a smaller variance than that of the periodogram because, as we have seen, it smooths the periodogram. An alternative to reduce the variance is to average several periodograms. However, the question is: How do we obtain such periodograms? The answer is easy and consists in dividing the N observations into windows of length $M < N$.

The Barlett method is one of the possible estimators based on the averaged periodogram. First, it divides the N observations into L non-overlapping windows of length M as

$$x_l[n] = x[(l-1)M + n], \tag{4.12}$$

where $n = 0, \ldots, M-1$, and $l = 1, \ldots, L$, and computes the periodogram of each window, that is,

$$\hat{S}_{x,l}^{p}(e^{j\omega}) = \frac{1}{M}\left|\sum_{n=0}^{M-1} x_l[n]e^{-j\omega n}\right|^2 = \frac{1}{M}\left|X_l(e^{j\omega})\right|^2. \tag{4.13}$$

Then, the Barlett estimator is given by simply averaging the individual periodograms

$$\hat{S}_{x}^{B}(e^{j\omega}) = \frac{1}{L}\sum_{l=1}^{L} \hat{S}_{x,l}^{p}(e^{j\omega}). \tag{4.14}$$

Although it is out of the scope of these notes, we must point out that $\hat{S}_x^B(e^{j\omega})$ is somehow related to $\hat{S}_x^{BT}(e^{j\omega})$.

There are two further improvements of the periodogram. The first one is based on the Barlett estimator but substituting the individual periodograms by Blackman-Tukey estimates. The second one is based on dividing the N observations into L overlapping windows. The combination of both improvements is known as the Welch method.

One final question remains: What happens to the bias and variance of these methods. Regarding the bias, it is going to be higher (lower resolution) than that of the periodogram since $M < N$, as also happened to the Blackman-Tukey estimate. As for the variance, it is going to be reduced by a factor of L, the number of windows. That is,

$$Var\left(\hat{S}_x^{ap}(e^{j\omega})\right) \approx \frac{1}{L} Var\left(\hat{S}_x^{p}(e^{j\omega})\right),$$

where $\hat{S}_x^{ap}(e^{j\omega})$ is any averaged periodogram (either Barlett or Welch methods) and $\approx$ is due to the non-independence between the windows. It would be an equality when the windows are independent, i.e., the Barlett method.

4.4 Parametric methods in spectral estimation

The problem of non-parametric methods is that they estimate an infinite number of parameters (the PSD at each frequency) from a sequence of N observations. Clearly, this is an ill-posed problem since there are (many) more parameters to estimate than observations. To overcome this issue, we could postulate a parametric model for the PSD and estimate only the parameters of such model using the N observations. For instance, the model could be $S_x(e^{j\omega}) = a + b\cos^2(\omega)$ and, hence, we only have to estimate a and b.

Parametric approaches, as described above, can provide a significant performance boost if the signal fits the postulated model, otherwise the performance could be even worse than that of non-parametric methods. It is therefore of the utmost importance to select the proper model.

In this chapter we will analyze autorregresive (AR) models, which are particularly amenable for parametric estimation.

4.4.1 Auto-Regressive (AR) models

We say that the stochastic process $x[n]$ follows an auto-regressive model of order p (or, simply, an AR(p) model, if it is the output of a causal, linear and time-invariant filter driven by the recursive relation

$$x[n] = u[n] - \sum_{k=1}^{p} a_k x[n-k], \tag{4.15}$$

when the input, $u[n]$ is an IID Gaussian process with zero mean and variance $\sigma^2 > 0$.

The impulse response of the filter can be represented by means of the recursive relation (4.15), replacing $u[n]$ by $\delta[n]$

$$h[n] = \delta[n] - \sum_{k=1}^{p} a_k h[n-k],$$

Applying the Fourier transform to this equation, we get,

$$H(e^{j\omega}) = 1 - \sum_{k=1}^{p} a_k e^{-j\omega k} H(e^{j\omega}),$$

therefore

$$H(e^{j\omega}) = \frac{1}{1 + \sum_{k=1}^{p} a_k e^{-j\omega k}},$$

4.4.1.1 Power Spectum

Since $x[n]$ is the output of filter $h[n]$ for input $u[n]$,

$$x[n] = u[n] * h[n] \tag{4.16}$$

therefore, the power spectrum is

$$S_x(e^{j\omega}) = S_u(e^{j\omega}) |H(e^{j\omega})|^2 = \frac{\sigma^2}{\left| 1 + \sum_{k=1}^{p} a_k e^{-j\omega k} \right|^2}. \tag{4.17}$$

According to Weierstrass theorem, for large values of p, the PSD model in (4.17) can approximate arbitrarily close any continuous PSD. Hence, there is a strong interest in this kind of models.

4.4.2 Auto-correlation

Using (4.15), we can express the auto-correlation function of the AR(p) process $x[n]$ as

$$\begin{aligned} r_x[m] &= \mathbb{E}\{x[n]x^*[n-m]\} \\ &= \mathbb{E}\left\{\left(u[n] - \sum_{k=1}^{p} a_k x[n-k]\right) x^*[n-m]\right\} \\ &= \mathbb{E}\{u[n]x^*[n-m]\} - \sum_{k=1}^{p} a_k r_x[m-k] \end{aligned} \tag{4.18}$$

We can develop this expression by analyzing the cases $m > 0$, $m = 0$ and $m < 0$ separately.

- For $m > 0$, since the filter is causal, $x^*[n-m]$ does not depend on $u[n]$, therefore,

$$\mathbb{E}\{u[n]x^*[n-m]\} = \mathbb{E}\{u[n]\}\mathbb{E}\{x^*[n-m]\} = 0, \qquad m > 0 \tag{4.19}$$

  therefore, (4.18) simplifies to

$$r_x[m] = -\sum_{k=1}^{p} a_k r_x[m-k], \qquad m > 0$$

- For $m = 0$, we get,

$$\begin{aligned} \mathbb{E}\{u[n]x^*[n-m]\} &= \mathbb{E}\{u[n]x^*[n]\} \\ &= \mathbb{E}\left\{u[n]\left(u^*[n] - \sum_{k=1}^{p} a_k^* x^*[n-k]\right)\right\} \\ &= \mathbb{E}\{u[n]u^*[n]\} = \sigma^2 \end{aligned} \tag{4.20}$$

  therefore

$$r_x[0] = \sigma^2 - \sum_{k=1}^{p} a_k r_x[m-k].$$

- Finally, for $m < 0$, since the autocorrelation function is Hermitian, we have $r_x[m] = r_x^*[-m]$.

  Joining the cases $m > 0$, $m = 0$ and $m < 0$, we get

$$r_x[m] = \begin{cases} -\sum_{k=1}^{p} a_k r_x[m-k] + \sigma^2, & m = 0, \\ -\sum_{k=1}^{p} a_k r_x[m-k], & m > 0, \\ r_x^*[-m], & m < 0. \end{cases} \tag{4.21}$$

We see in (4.21) that the relationship between the model parameters (σ^2, $a_1, \ldots, a_p$) and the auto-correlation is linear. which simplifies the estimation of such parameters. The

estimation procedure consists in substituting the theoretical auto-correlation by an estimate and then solving a linear system of equations. The PSD estimate is obtained by substituting the estimated parameters in the corresponding model.

4.4.3 Parameter estimation from the autocorrelation

Since we need to obtain $p+1$ parameters, i.e., $a_1,\ldots,a_p$ and σ^2, we need $p+1$ equations from (4.21), which are known as the Yule-Walker equations:

$$\begin{aligned} r_x[0] &= -a_1 r_x[-1] - a_2 r_x[-2] + \ldots - a_p r_x[-p] + \sigma^2, \\ r_x[1] &= -a_1 r_x[0] - a_2 r_x[-1] + \ldots - a_p r_x[-p+1], \\ &\vdots \\ r_x[p] &= -a_1 r_x[p-1] - a_2 r_x[p-2] + \ldots - a_p r_x[0]. \end{aligned}$$

The last p equations depend only on $a_1,\ldots,a_p$. We can write them in matrix form by using $r_x[-m] = r_x^*[m]$ and defining

$$\mathbf{r}_x = \begin{bmatrix} r_x[1] & r_x[2] & \cdots & r_x[p] \end{bmatrix}^T,$$

and

$$\mathbf{R}_x = \begin{bmatrix} r_x[0] & r_x^*[1] & \cdots & r_x^*[p-1] \\ r_x[1] & r_x[0] & \cdots & r_x^*[p-2] \\ \vdots & \vdots & \ddots & \vdots \\ r_x[p-1] & r_x[p-2] & \cdots & r_x[0] \end{bmatrix},$$

to arrive at

$$\begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \vdots \\ \hat{a}_p \end{bmatrix} = -\mathbf{R}_x^{-1}\mathbf{r}_x,$$

The matrix $\mathbf{R}_x$ has a special structure, namely, it is constant along diagonals, and is therefore known as Toeplitz. This fact is important for solving the system of equations (computing the matrix inverse) as it reduces the complexity from $\mathcal{O}(p^3)$ to $\mathcal{O}(p^2)$. The remaining parameter to be estimated, the variance, is easily obtained as

$$\hat{\sigma}^2 = r_x[0] + \hat{a}_1 r_x^*[1] + \hat{a}_2 r_x^*[2] + \ldots + \hat{a}_p r_x^*[p].$$

Finally, as we have already seen before, in practical scenarios the auto-correlation function is not available and must therefore be replaced by an estimate. The overall procedure consist on three steps:

1. Estimate the autocorrelation function
2. Compute the model parameters solving the Yule-Walker equations
3. Estimate the power spectrum using the parameter estimates.

4.4.4 Maximum Likelihood estimation

We can avoid the estimation of the autocorrelation function by applying estimation theory to compute the model parameters from the signal observations. Assume we have recorded a set of N observations from the process, $\{x[n], 0 \leq n \leq N-1\}$. Our goal is to estimate the model parameters based on these observations. To do so, the following vector notation will be useful:

$$\mathbf{a} = (a_1, a_2, \ldots, a_{p-1})^\top \tag{4.22}$$

$$\mathbf{x} = (x[p], x[p+1], \ldots, x[N-1])^\top \tag{4.23}$$

$$\mathbf{x}_n = (x[n-1], x[n-2], \ldots, x[n-p])^\top \tag{4.24}$$

Note that $\mathbf{x}_n$ contains all values from the signal record that multiply the coefficients to determine $x[n]$, in such a way that the signal model (4.15) can be expressed as

$$x[n] = u[n] - \mathbf{x}_n^\top \mathbf{a}, \tag{4.25}$$

Also, note that vector $\mathbf{x}$ contains all observations from the process starting for $n > p$. We have excluded the observations for $n < p$ because the maximization of the complete likelihood including these values (i.e. $p(\mathbf{x}, \mathbf{x}_p \mid \mathbf{a})$) becomes much more complex. To avoid these complications, we will take $\mathbf{x}_p$ as fixed data, maximizing the likelihood for the rest of observations, that is, we will compute the estimate

$$\hat{\mathbf{a}}_{\text{ML}} = \underset{\mathbf{a}}{\operatorname{argmax}}\, p(\mathbf{x} \mid \mathbf{x}_p, \mathbf{a}) \tag{4.26}$$

To do so, using (4.25), we factorize the likelihood as

$$\begin{aligned} p(\mathbf{x} \mid \mathbf{x}_p, \mathbf{a}) =& p(x[N-1], x[N-2], \ldots, x[p] \mid \mathbf{x}_p, \mathbf{a}) \\ =& p(x[N-1] \mid x[N-2], \ldots, x[p], \mathbf{x}_p, \mathbf{a}) \cdot p(x[N-2] \mid x[N-3], \ldots, x[p], \mathbf{x}_p \mathbf{a}) \\ & \cdot \ldots \cdot p(x[p] \mid \mathbf{x}_p, \mathbf{a}) \\ =& \prod_{n=p}^{N-1} p(x[n] \mid \mathbf{x}_n, \mathbf{a}) \end{aligned} \tag{4.27}$$

Since the input process is IID, zero-mean, Gaussian, we can write

$$p(x[n] \mid \mathbf{x}_n, \mathbf{a}) = \mathcal{N}\left(x[n] - \mathbf{x}_n^\top \mathbf{a}, \sigma^2\right) \tag{4.28}$$

therefore

$$\begin{aligned} p(\mathbf{x} \mid \mathbf{x}_p, \mathbf{a}) =& \prod_{n=p}^{N-1} \mathcal{N}\left(x[n] - \mathbf{x}_n^\top \mathbf{a}, \sigma^2\right) \\ =& \frac{1}{\sigma^{N-p}(2\pi)^{(N-p)/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{n=p}^{N-1} \left(x[n] - \mathbf{x}_n^\top \mathbf{a}\right)^2\right) \end{aligned} \tag{4.29}$$

Therefore, the ML estimate is the solution of a least squares problem

$$\hat{\mathbf{a}}_{\mathrm{ML}} = \underset{\mathbf{a}}{\operatorname{argmin}} \sum_{n=p}^{N-1} \left(x[n] - \mathbf{x}_n^\top \mathbf{a} \right)^2 \tag{4.30}$$

which can be expressed in matrix form by defining the observation matrix

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_p^\top \\ \mathbf{x}_{p+1}^\top \\ \vdots \\ \mathbf{x}_{N-1}^\top \end{pmatrix} \tag{4.31}$$

to arrive at

$$\hat{\mathbf{a}}_{\mathrm{ML}} = \underset{\mathbf{a}}{\operatorname{argmin}} \|\mathbf{x} - \mathbf{X}\mathbf{a}\|^2 = \left(\mathbf{X}\mathbf{X}^\top \right)^{-1} \mathbf{X}^\top \mathbf{x} \tag{4.32}$$

The same approach can be applied to the estimation of the variance, σ^2:

$$\begin{aligned} \hat{\sigma}^2_{\mathrm{ML}} &= \underset{\sigma}{\operatorname{argmax}} \frac{1}{\sigma^{N-p}(2\pi)^{(N-p)/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{n=p}^{N-1} \left(x[n] - \mathbf{x}_n^\top \mathbf{a} \right)^2 \right) \\ &= \frac{1}{N-p} \sum_{n=p}^{N-1} \left(x[n] - \mathbf{x}_n^\top \mathbf{a} \right)^2 \end{aligned} \tag{4.33}$$

4.4.5 Signal prediction

The estimation of the parameters and an AR process can be done by simply solving a system of equations. Moreover, from the expression of an AR model

$$x[n] = u[n] - \sum_{k=1}^{p} a_k x[n-k],$$

we note that they can be used to predict future samples by ignoring the input, i.e.,

$$x[n] = -\sum_{k=1}^{p} \hat{a}_k x[n-k],$$

where the coefficients of the model have been replaced by some estimates. That is, from a record of N samples, $x[0], \ldots, x[N-1]$, we can estimate the model parameters and, afterwards, we can predict $x[N], x[N+1], \ldots$.