Chapter 4

Nonparametric Techniques

"Which experiments reveal structural properties of the studied system?"

Now let us look at a converse problem. Here we do not look at the properties of the assumed model class, but we compute such properties based on experiments carried out on the studied system. In general, such methods are not tied to a specific (parameterized) model, but nevertheless embody a description of the system. Such method have the denominator 'non-parametric' or 'distribution-free'. They often come in the forms graphs, curves, tables or other intuitive representations, and give as such structural information of the system. Their use is often found in

(Preprocess) Indicate important effects present in the studied system.

(Model class) Suggest a suitable class of parametric models which can be used to capture such effects.

(Validate) Check whether the identified model behaves similarly than the actual system.

4.1 Transient Analysis

A first approach is to inject the studied system with a simple input as a pulse or a step, and to record the subsequent output of the system. This gives then an impression of the impulse response of the studied system. Let us look further into the pros and cons of this strategy. Formally, let the following input signal $\{u_t\}_t$ be injected to the system H

$$u_t = \begin{cases} K & t = 0 \\ 0 & \text{else.} \end{cases} \tag{4.1}$$

Then, if the system could be described exactly (i.e. without any effect of unknown disturbances) as $H(q^{-1}) = h_0 + \cdots + h_\tau q^{-\tau} + \ldots$, then the output of the system becomes

$$y_t = H(q)u_t = K \begin{cases} h_t & t \geq 0 \\ 0 & \text{else.} \end{cases} \tag{4.2}$$

So conversely, if one knows that the system follows very closely an LTI description $H(q^{-1}) = h_0 + \cdots + h_\tau q^{-\tau} + \ldots$, the different unknowns $\{h_\tau\}_\tau$ can be observed directly when injecting the

studied system with a pulse signal as in eq. (4.1). The pros of this approach are that (i) it is simple to understand or to (ii) implement, while the model need not be specified further except for the LTI property. The downsides are of course that (i) this method breaks down when the LTI model fits not exactly the studied system. Since models serve merely as mathematical convenient approximations of the actual system, this is why this approach is in practice not often used. (ii) It cannot handle random effects very well. (iii) such experiment is not feasible in the practical setting at hand. As for this reason it is merely useful in practice to determine some structural properties of the system. For example consider again the first order system as in the previous example, then a graph of the impulse response indicates the applicable time-constants and gain of the system.

Similarly, consider the step input signal $\{u_t\}_t$ defined as

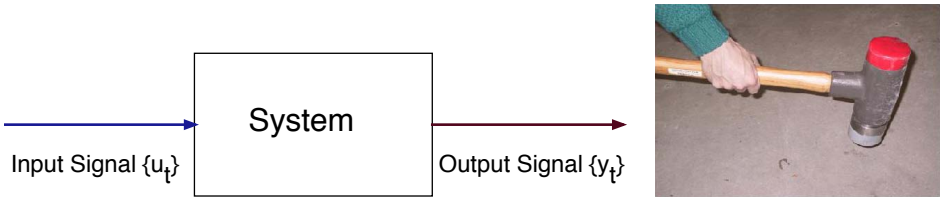
$$u_t = \begin{cases} K & t \geq 0 \\ 0 & \text{else.} \end{cases} \tag{4.3}$$

Then, if the system could be described exactly (i.e. without any effect of unknown disturbances) as $H(q^{-1}) = h_0 + \cdots + h_\tau q^{-\tau} + \ldots$, then the output of the system becomes

$$y_t = H(q^{-1})u_t = K \begin{cases} \sum_{\tau=0}^{t} h_\tau & t \geq 0 \\ 0 & \text{else.} \end{cases} \tag{4.4}$$

or equivalently

$$y_t - y_{t-1} = K \begin{cases} h_\tau & t \geq 1 \\ 0 & \text{else.} \end{cases} \tag{4.5}$$



(a) (b)

Figure 4.1: (a) A Block Representation of a system. (b) An impact hammer used for modal analysis of bridges and other large constructions.

4.2 Frequency Analysis

As seen in the previous chapter, an LTI is often characterized in terms of its reaction to signals with a certain frequency and phase. It is hence only natural to try to learn some properties of the studied system by injecting it with a signal having such a form. Specifically, let $\{u_t\}$ be defined for $t = \ldots, -1, 0, 1, \ldots$ as

$$u_t = a \sin(\omega t). \tag{4.6}$$

where $a > 0$ is the gain of the signal. Then as seen in the previous chapter the corresponding output of the system H is given as

$$y_t = K \sin(\omega t + \phi), \tag{4.7}$$

where

$$\begin{cases} K = a|H(e^{i\omega})| \\ \phi = \arg G(e^{i\omega}). \end{cases} \tag{4.8}$$

Note that normally the phase ϕ will be negative. By measuring the amplitude a, K and the phase ϕ for for given ω, one can find the complex variable $H(e^{i\omega})$ from (4.8). If repeating this procedure for a range of frequencies ω, one can obtain a graphical representation of $H(e^{i\omega})$. Such Bode plots (or Nyquist or related plots) are well suited for the design and analysis of automatic control systems. The procedure described above is rather sensitive to disturbances. This is not difficult to understand. If one has disturbance terms with Laplace transform $E(s)$, one gets

$$Y(s) = H(s)U(s) + E(s). \tag{4.9}$$

Then when injecting the system with a signal $\{u_t\}$ as in eq. (4.6) one gets the output signal $\{y_t\}$ where

$$y_t = K\sin(\omega t + \phi) + e_t, \tag{4.10}$$

and due to the presence of noise it will be difficult to extract good estimates of K and ϕ from those signals.

4.3 A Correlation Analysis

The above ideas are taken a step further into a correlation analysis. But instead of using simple input signals, the system is injected with a random signal $\{u_t\}_t$ which has zero mean or

$$\mathbb{E}[u_t] = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} u_t, \tag{4.11}$$

which has finite values. A formal definition of such white noise sequence is given in Chapter 4, but for now it is sufficient to let the expectation $\mathbb{E}[\cdot]$ denote an limit of an average, or $\mathbb{E}[u_t] = \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} u_t$. Then the output which is recorded at the output of the system

$$y_t = \sum_{\tau=0}^{\infty} h_\tau u_{t-\tau}. \tag{4.12}$$

When taking multiplying both sides by $u_{t'}$ for any t', and taking expectations one gets

$$\mathbb{E}[u_{t'} y_t] = \mathbb{E}\left[\sum_{\tau=0}^{\infty} h_\tau u_{t-\tau} u_{t'}\right] = \sum_{\tau=0}^{\infty} h_\tau \mathbb{E}\left[u_{t-\tau} u_{t'}\right]. \tag{4.13}$$

Summarizing this for all t, t' and canceling the cross-terms gives the linear system

$$\begin{bmatrix} r_{uy}(0) \\ r_{uy}(1) \\ \vdots \\ r_{uy}(\tau) \\ \vdots \end{bmatrix} = \begin{bmatrix} r_{uu}(0) & r_{uu}(1) & r_{uu}(2) & \dots & r_{uu}(\tau) & \dots \\ r_{uu}(1) & r_{uu}(0) & & & r_{uu}(\tau-1) & \dots \\ r_{uu}(2) & & & & & \\ \vdots & & & \ddots & & \\ r_{uu}(\tau) & r_{uu}(\tau-1) & & & r_{uu}(0) & \dots \\ \vdots & & & & & \end{bmatrix} \begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_\tau \\ \vdots \end{bmatrix} \tag{4.14}$$

where

$$r_{uu}(\tau) = \mathbb{E}[u_t u_{t-\tau}] = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} u_t u_{t-\tau}. \tag{4.15}$$

and

$$r_{uy}(\tau) = \mathbb{E}[y_t u_{t-\tau}] = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} y_t u_{t-\tau}. \tag{4.16}$$

Since this limit cannot be computed explicitly in practice, one settles for working with the estimates

$$\hat{r}_{uu}(\tau) = \frac{1}{n} \sum_{i=1}^{n} u_t u_{t-\tau}. \tag{4.17}$$

and

$$\hat{r}_{uy}(\tau) = \frac{1}{n} \sum_{i=1}^{n} y_t u_{t-\tau}. \tag{4.18}$$

Secondly, rather than solving the infinite system (4.19), one solves the corresponding finite linear system for appropriate $m > 0$ given as

$$\begin{bmatrix} \hat{r}_{uy}(0) \\ \hat{r}_{uy}(1) \\ \vdots \\ \hat{r}_{uy}(m-1) \end{bmatrix} = \begin{bmatrix} \hat{r}_{uu}(0) & r_{uu}(1) & \hat{r}_{uu}(2) & \dots & \hat{r}_{uu}(m-1) \\ \hat{r}_{uu}(1) & r_{uu}(0) & & & \hat{r}_{uu}(m-2) \\ \hat{r}_{uu}(2) & & & & \\ \vdots & & & \ddots & \\ \hat{r}_{uu}(m-1) & r_{uu}(m-2) & & & \hat{r}_{uu}(0) \end{bmatrix} \begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_M \end{bmatrix} \tag{4.19}$$

in order to get an idea about $\{h_\tau\}_{\tau=0}^{n}$. Those equations are known as Wiener-Hopf type of equations. This technique is related to the Least Squares estimate and the Prediction Error Method in Chapter 5.

4.4 Spectral Analysis

Now both the correlation technique and the frequency analysis method can be combined into a signal nonparametric approach as follows. The idea is to take the Discrete Fourier Transforms (DFT) of the involved signals, and find the transfer function relating them.

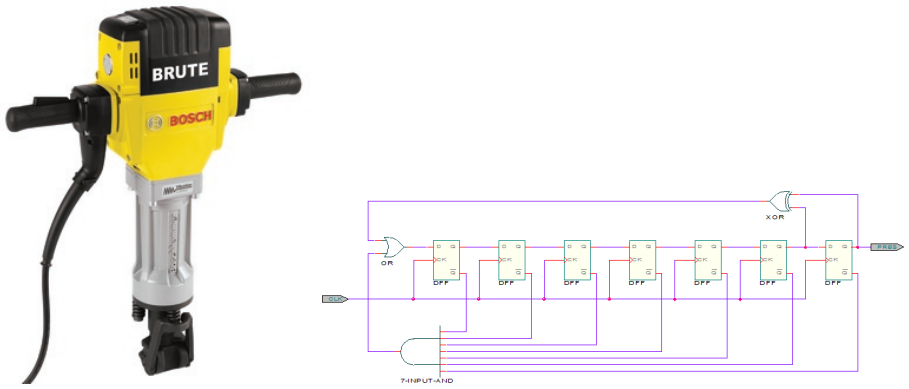
$$\begin{cases} \phi_{uu}(\omega) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} r_{uu}(\tau) e^{-i\omega\tau} \\ \phi_{uy}(\omega) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} r_{uy}(\tau) e^{-i\omega\tau}. \end{cases} \tag{4.20}$$

Then the previous chapter learns us that we have that for all ω one has

$$\phi_{uy}(\omega) = H(e^{-i\omega})\phi_{uu}(\omega), \tag{4.21}$$

where

$$H(e^{-i\omega}) = \sum_{\tau=0}^{\infty} h_\tau e^{-i\tau\omega}. \tag{4.22}$$



(a) (b)

Figure 4.2: (a) A vibration Hammer (b) A circuit scheme for realizing a Pseudo Random Binary Sequence.

Consequently, a reasonable estimate for H would be

$$\hat{H}(e^{-i\omega}) = \frac{\hat{\phi}_{uy}(\omega)}{\hat{\phi}_{uu}(\omega)}, \tag{4.23}$$

where $\hat{\phi}$ are reasonable estimates of ϕ. A straightforward estimate would be

$$\hat{\phi}_{uy} = \frac{1}{2\pi} \sum_{\tau=-n}^{n} \hat{r}_{uy} e^{-i\tau\omega}, \tag{4.24}$$

and similarly for $\hat{\phi}_{uu}$. Working out $\hat{r}_{uy}$ gives

$$\hat{\phi}_{uy} = \frac{1}{2\pi n} \sum_{\tau=-n}^{n} \sum_{t=1-\min(\tau,0)}^{\max(\tau,0)} y_{t+\tau} u_t e^{-i\tau\omega}. \tag{4.25}$$

Then change of indices τ and t gives

$$\hat{\phi}_{uy} = \frac{1}{2\pi n} \sum_{s=1}^{n} \sum_{t=1}^{n} y_s u_t e^{-i(s-t)\omega} = \frac{1}{2\pi n} Y_n(\omega) U_n(-\omega), \tag{4.26}$$

where

$$\begin{cases} U_n(\omega) = \sum_{s=1}^{n} u_s e^{-is\omega} \\ Y_n(\omega) = \sum_{s=1}^{n} y_s e^{-is\omega}. \end{cases} \tag{4.27}$$

Those are the Discrete Fourier Transforms of the signals $\{u_t\}$ and $\{y_t\}$ (padded with zeros). For $\omega = 0, \frac{2\pi}{n}, \frac{4\pi}{n}, \ldots, \pi$ those can be estimated efficiently using the Fast Fourier Transform (FFT) algorithms. In a similar fashion one has

$$\hat{\phi}_{uu} = \frac{1}{2\pi n} U_n(\omega) U_n(-\omega) = \frac{1}{2\pi n} |U_n(\omega)|. \tag{4.28}$$

This estimate is called the *periodogram*. From the derivations above it follows that

$$\hat{H}(e^{-i\omega}) = \frac{Y_n(\omega)}{U_n(\omega)}. \tag{4.29}$$

This estimate is sometimes called the *empirical transfer function* estimate.

However the above estimate to the spectral densities and the transfer function will give poor results. For example, if u_t is a stochastic process, then the estimates eq. (4.28) and (4.26) do not converge in (the mean square sense) to the true spectrum as n, the number of datapoints tends to infinity. In particular, the estimate $\hat{\phi}_{uu}$ will on average behave as ϕ_{uu}, but its variance does not tend to zero as $n \to \infty$. One of the reasons for this behavior is that $\phi\phi_{uu}(\tau)$ will be quite inaccurate for large values for τ, but all covariance elements $\hat{r}_{uy}(\tau)$ are given the same weight in eq. (4.26) regardless of their accuracy. Another more subtle reason goes as follows. In eq. (4.26) $2n+1$ terms are summed. Even if the estimation error of each term goes to zero, there is no guarantee that the global sum goes to zero. These problems may be overcome if the terms of eq. (4.26) corresponding with large τ are weighted out. Thus, instead of eq. (4.26) the following improed estimate of the cross-spectrum can be used

$$\hat{\phi}'_{uy} = \frac{1}{2\pi} \sum_{\tau=-n}^{n} \hat{r}_{uy}(\tau) w(|\tau|) e^{-i\tau\omega}, \tag{4.30}$$

where $w : \mathbb{R} \to \mathbb{R}_+$ is a socalled *lag window*. It should $w(0) = 1$, and decreasing. Several forms of the lag window have been proposed in the literature. Some simple lag windows are presented in the following example.

Example 30 (Lag Windows) *The following lag windows are often used in the literature.*

- *Rectangular window:*

$$w_1(|\tau|) = \begin{cases} 1 & |\tau| \leq M \\ 0 & |\tau| > M \end{cases} \tag{4.31}$$

- *Bartlett window:*

$$w_2(|\tau|) = \begin{cases} 1 - \frac{|\tau|}{M} & |\tau| \leq M \\ 0 & |\tau| > M \end{cases} \tag{4.32}$$

- *Hamming and Tukey*

$$w_3(|\tau|) = \begin{cases} \frac{1}{2}(1 + \cos\frac{\pi\tau}{M}) & |\tau| \leq M \\ 0 & |\tau| > M. \end{cases} \tag{4.33}$$

Note that all the windows vanish for $|\tau| > M$. If the parameters M is chosen to be sufficiently large, the periodogram will not be smoothed very much. On the other hand a small M may mean that essential parts of the spectrum are smoothed out.It is not trivial to choose the parameter M. Roughly speaking M should be chosen according to trading off the following two objectives:

- *M should be small compared to n:*
- *$|\hat{r}_{uy}(\tau)| \ll \hat{r}_{uu}(0)$ for $\tau \geq M$ so as not to smooth out the parts of interest in the true spectrum.*

The use of a lag window is necessary to obtain a reasonable accuracy. On the other hand, sharp peaks in the spectrum might be smeared out. It may therefore not be possible to separate adjacent peaks. Thus the use of a lag window will give a limited frequency resolution. The effect of a lag window is illustrated in the following example.

4.5 Nonparameteric Techniques for Timeseries

Let us study a similar technique for estimating the transfer function of a timeseries. Here, we do not have an input sequence available, but we assume that the observed sequence $\{y_t\}_t$ is driven by unobserved white noise $\{e_t\}$. The previous chapter enumerates some common models for such systems. Here we will give a main nonparametric technique useful for recovering the underlying structure. Again such approaches are based on working with the covariances observed in the system.

4.5.1 Yule-Walker Correlation Analysis

Let's consider the equivalent of the correlation approach when timeseries are concerned. At first, assume the studied timeseries follows an AR(m) model as

$$y_t - a_1 y_{t-1} - \cdots - a_m y_{t-m} = e_t, \tag{4.34}$$

where $\{e_t\}$ is zero mean white noise such that

$$\begin{cases} \mathbb{E}[e_t] = 0 \\ \mathbb{E}[e_t e_{t-\tau}] = \sigma^2 \delta_\tau \end{cases} \tag{4.35}$$

with δ_τ equal to one if $\tau = 0$, and equals zero otherwise. Note that then $\{y_t\}$ can be seen as a linear combination of past values of the noise $\{e_s\}_{s \le t}$. B multiplication of both sides of eq. (4.34) with a (delayed) value of the process $y_{t-\tau}$ for all $\tau = 0, 1, 2, \ldots$, and taking the expectation one gets

$$\mathbb{E}[y_{t-\tau}(y_t - a_1 y_{t-1} - \cdots - a_m y_{t-m})] = \mathbb{E}[y_{t-\tau} e_t], \tag{4.36}$$

or

$$r_y(\tau) - a_1 r_y(\tau - 1) - \cdots - a_m r_y \tau - m = \begin{cases} \sigma^2 & \text{if } \tau = 0 \\ 0 & \text{otherwise,} \end{cases} \tag{4.37}$$

where we defined as before

$$r_y(\tau) = \mathbb{E}[y_t y_{t-\tau}] = \lim_{n \to \infty} \frac{1}{n} \sum_{t=\tau+1}^{n} y_t y_{t-\tau}. \tag{4.38}$$

Assuming those are given for all $\tau = 0, 1, 2, \ldots$, those can be organized as a system of linear equations as follows

$$\begin{bmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} r_y(0) & r_y(1) & r_y(2) & \ldots & r_y(M) & \ldots \\ r_y(1) & r_y(0) & & & r_y(M-1) & \ldots \\ r_y(2) & & & & & \\ \vdots & & & \ddots & & \\ r_y(M) & r_y(M-1) & & & r_y(0) & \ldots \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_M \end{bmatrix} \tag{4.39}$$

Those are known as the Yule-Walker (YW) equations. Conversely, if one knows the parameters $\{a_1, \ldots, a_{m_a}\}$, the covariances $\{r_y(\tau)\}$ are given as solutions to the system

$$\begin{bmatrix} 1 & a_1 & a_2 & \ldots & a_{m_a} \\ a_1 & 1+a_2 & a_3 & & 0 \\ \vdots & & \ddots & & \vdots \\ a_{m_a} & 0 & \ldots & & 1 \end{bmatrix} \begin{bmatrix} r_y(0) \\ r_y(1) \\ \vdots \\ r_y(m_a) \end{bmatrix} = \begin{bmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{4.40}$$

Next, assume that the timeseries follows an ARMA(m_a, m_c) model given as

$$y_t - a_1 y_{t-1} - \cdots - a_{m_a} y_{t-m_a} = e_t + c_1 e_{t-1} + \cdots + c_{m_c} e_{t-m_c}, \tag{4.41}$$

and where $\{e_t\}$ is a white zero mean process satisfying the conditions (4.35). Then, again multiplying both sides of eq. (4.41) with $y_{t-\tau}$ and taking the expectation gives the relations

$$r_y(\tau) - a_1 r_y(\tau-1) - \cdots - a_{m_a} r_y \tau - m_a = r_{ey}(\tau) + c_1 r_{ey}(\tau-1) + \cdots + r_{ey}(\tau - m_c). \tag{4.42}$$

The cross-correlations $r_{ey}(\tau)$ are found as follows. Multiply both sides of eq.(4.41) with $e_{t-\tau}$ and take expectations, then we have that

$$r_{ey}(\tau) - a_1 r_{ey}(\tau-1) - \cdots - a_{m_a} r_y \tau - m_a = \sigma^2 \left(c_0 \delta_\tau + \cdots + c_{m_c} \delta_{\tau - m_c}\right), \tag{4.43}$$

where

$$r_{ey}(\tau) = \mathbb{E}[e_{t+\tau} y_t] = \lim_{n \to} \frac{1}{n} \sum_{t=1}^{n-\tau} e_{t+\tau} y_t. \tag{4.44}$$

As y_t is a linear combination of $\{e_s\}_{s \leq t}$, we have that $r_{ey}(\tau) = 0$ for $\tau > 0$. It as such follows that for all $\tau > m_c$ one has

$$r_y(\tau) - a_1 r_y(\tau-1) - \cdots - a_{m_a} r_y(\tau - m_a) = 0. \tag{4.45}$$

Note that those equations involve only the AR parameters of the ARMA process.

4.5.2 Spectral Factorization

In order to get an initial estimate of the MA parameters the following technique is often used. Rather than setting up the YW equations, one extracts the MA part from the covariance structure using the following Lemma.

Lemma 5 (Spectral Factorization) *Let $\phi : \mathbb{C} \to \mathbb{C}$ be a spectrum that can be written for all $z \in \mathbb{C}$ as*

$$\phi(z) = \frac{\sum_{k=-m_\beta}^{m_\beta} \beta_k z^k}{\sum_{k=-m_\alpha}^{m_\alpha} \alpha_k z^k}, \tag{4.46}$$

for $\beta_{-m_\beta}, \ldots, \beta_{m_\beta}, \alpha_{-m_\alpha}, \ldots, \alpha_{m_\alpha} \in \mathbb{R}$. Then there exists two functions

$$\begin{cases} A(z) = 1 + a_1 z + \cdots + a_{m_\alpha} z^{m_\alpha} \\ C(z) = 1 + c_1 z + \cdots + c_{m_\beta} z^{m_\beta}, \end{cases} \tag{4.47}$$

with $a_1, \ldots, a_{m_\alpha}, c_1, \ldots, c_{m_\beta} \in \mathbb{C}$, and a constant $\sigma > 0$ such that

1. *$A(z)$ has all zeros inside the unit circle.*
2. *$C(z)$ has all zeros inside or on the unit circle.*
3. *one has for all $z \in \mathbb{C}$ that*
$$\phi(z) = \sigma^2 \frac{C(z)}{A(z)} \frac{C^*(z^{-*})}{A^*(z^{-*})}, \tag{4.48}$$
where $A^*(z) = 1 + a_1^* z + \cdots + a_{m_\alpha}^*$ and $C^*(z^{-*}) = 1 + c_1^* z + \cdots + c_{m_\beta}^*$ and with z^* denoting the conjugate of $z \in \mathbb{C}$.

The proof of this result hinges on complex algebra, see e.g. [5].

Example 31 (MA Process) *Suppose one has given a MA(m) process such that*

$$y_t = e_t + c_1 e_{t-1} + \cdots + c_m e_{t-m}, \tag{4.49}$$

and where $\{e_t\}_t$ is zero mean, white noise with standard deviation σ. Then the covariances of this process are given as

$$r_y(\tau) = \begin{cases} \sigma^2 c_{|\tau|}^2 & |\tau| \leq m \\ 0 & \textit{elsewhere.} \end{cases} \tag{4.50}$$

The spectral density is then given for any $\omega \in]-\pi, \pi]$ as

$$\begin{aligned} \phi_y(\omega) = \frac{1}{2\pi} \sum_\tau r_y(\tau) e^{-i\omega\tau} &= \frac{1}{2\pi} \left(r_y(0) + \sum_{k=1}^m c_k^2 (e^{-i\omega k} + e^{i\omega k}) \right) \\ &= \frac{1}{2\pi} \left(r_y(0) + \sum_{k=1}^m c_k^2 \cos(\omega k) \right). \end{aligned} \tag{4.51}$$