1 Paradigm of Signal Processing

1.1 Introduction

Signal processing is a branch of applied mathematics. The primary aim of signal processing is to extract information from a given-set of data, a procedure known as *signal analysis*, or conversely, to use information in generating data with some desirable property, a procedure known as *signal synthesis*. Mathematical methods are developed based on concepts from mathematics with the purpose of signal processing as an area of application. Various mathematical concepts find use in signal processing, and a systematic introduction of those mathematical concepts to the students who will become researchers and professionals in this area, has become an essential part of the curricula worldwide.

When the mathematical ideas are introduced in a course of mathematics, they are presented as abstractions. Theories are built in structured manner, based on the present state of knowledge and the aspiration of the mathematician to extend knowledge in some desired direction. A mathematician is often not concerned of the areas of application of the developed theories. In contrast, a researcher engaged in signal processing often considers an established theory and identifies ways how it can be adopted and applied in extracting information inherent in data or for generating data with some desirable property.

It is the purpose of the present book to introduce the mathematical concepts and their interpretations related to signal processing. It is illustrated how basic concepts mould into mathematical methods, and mathematical methods perform signal processing tasks. We take up the case study of spectral estimation of a signal in the next section. The spectral estimation, by definition, refers to computing the energy spectral density of a signal with finite energy or the power spectral density of a signal with finite power. In the case study, however, we generalize our tasks by stating that primarily we are interested in finding the frequency-content of a signal, and computation of the energy-level or power-level at each frequency is our secondary purpose. This case study will illustrate various connections of signal processing with mathematics.

1.2 A Case Study: Spectral Estimation

Suppose that we want to know what frequencies are present in a signal $g(t)$ which is observed over some duration of time. If the signal is periodic with period T, then it satisfies the relation $g(t) = g(t+T)$, for the minimum of such non-zero positive values of T.

Using the Fourier series expansion of a periodic signal, we can express the real-valued signal $g(t)$ as

$$g(t) = a_0 + \sum_{k=1}^{\infty}\left[a_k \cos\left(2\pi k f_0 t\right) + b_k \sin\left(2\pi k f_0 t\right)\right] \tag{1.1}$$

where $f_0 = \frac{1}{T}$ is the fundamental frequency, and $k f_0$ is the kth harmonic frequency. The Fourier series coefficients a_0, a_k and b_k are given by

$$\begin{aligned} a_0 &= \frac{1}{T}\int_{-T/2}^{T/2} g(t)dt \\ a_k &= \frac{2}{T}\int_{-T/2}^{T/2} g(t)\cos\left(2\pi k f_0 t\right)dt, \quad k = 1, 2, \ldots, \infty \\ b_k &= \frac{2}{T}\int_{-T/2}^{T/2} g(t)\sin\left(2\pi k f_0 t\right)dt, \quad k = 1, 2, \ldots, \infty \end{aligned} \tag{1.2}$$

The upper and lower bounds of the integrals are not important for a periodic signal so long the time duration of integration remains as one period T.

The Fourier series expansion of a periodic complex-valued signal $g(t)$ with period T is expressed as

$$g(t) = \sum_{k=-\infty}^{\infty} c_k \exp\left(j2\pi k f_0 t\right) \tag{1.3}$$

where $f_0 = \frac{1}{T}$ is the fundamental frequency, and $k f_0$ is the kth harmonic frequency. The complex exponential Fourier series coefficients c_k are given by

$$c_k = \frac{1}{T}\int_T g(t)\exp\left(-j2\pi k f_0 t\right) \tag{1.4}$$

To ensure that the coefficients c_k are finite, and the infinite series of (1.3) converges to $g(t)$ at every point t (in the average sense at the point of discontinuity), the signal $g(t)$ must satisfy the three Dirichlet conditions and the main among those conditions is that the signal must be absolutely integrable over one period of the signal, i.e.,

$$\int_T |g(t)|dt < \infty \tag{1.5}$$

For a real-valued signal $g(t)$, (1.3) gets a special form with $c_k^* = c_{-k}$, $k = 1, 2, \ldots, \infty$, and c_0 being real. Writing $c_k = |c_k| \exp\left(j\theta_k\right)$, the signal of (1.3) gets the form

$$g(t) = d_0 + \sum_{k=1}^{\infty} d_k \cos\left(2\pi k f_0 t + \theta_k\right) \tag{1.6}$$

where $d_0 = c_0 = |c_0|$ and $d_k = 2|c_k|$, $k = 1, 2, \ldots, \infty$. Substituting $d_0 = a_0$, $d_k \cos\theta_k = a_k$ and $-d_k \sin\theta_k = b_k$ in (1.6), we get the same signal of (1.1).

For aperiodic signals of finite duration, the signal representation consists of frequency components over a continuous domain, and the Fourier sum becomes the Fourier integral on the frequency-axis. The Fourier analysis is extended from periodic to aperiodic signals with the argument that an aperiodic signal can be thought of as a periodic signal with infinite period. As the period becomes larger and larger, the fundamental frequency becomes smaller and smaller, and in the limit, the harmonic frequencies form a continuum.

For the frequency domain representation of signals, we define the Fourier transform pair as

$$G(f) = \int_{-\infty}^{\infty} g(t) \exp(-j2\pi ft) dt \tag{1.7}$$

$$g(t) = \int_{-\infty}^{\infty} G(f) \exp(j2\pi ft) df \tag{1.8}$$

While the Fourier transform of $g(t)$, $F[g(t)] = G(f)$ shows the complex spectrum of the signal $g(t)$ at any frequency f, the inverse Fourier transform $F^{-1}[G(f)] = g(t)$ provides a means for reconstructing the signal. Once again, to guarantee that $G(f)$ is finite at every value of f and the integral of (1.8), which is over infinite duration, converges to $g(t)$ at every point of t (in the average sense at the point of discontinuity), the signal $g(t)$ must satisfy the three Dirichlet conditions. The main among those conditions is that the signal must be absolutely integrable, i.e.,

$$\int_{-\infty}^{\infty} |g(t)| dt < \infty \tag{1.9}$$

Although the periodic signals are neither absolutely integrable nor square integrable, i.e.,

$$\int_{-\infty}^{\infty} |g(t)|^2 dt < \infty \tag{1.10}$$

which is the condition for a signal being of finite energy, we can still compute the Fourier transform for periodic signals provided we permit inclusion of the impulse function(s) in the Fourier transform. The area under an impulse function $\delta(t)$ is unity, and the inverse Fourier transform of a train of impulses converges to a periodic signal:

$$\int_{-\infty}^{\infty} \left[\sum_{k=-\infty}^{\infty} c_k \delta\left(f - kf_0\right) \right] \exp\left(j2\pi ft\right) df = \sum_{k=-\infty}^{\infty} c_k \exp\left(j2\pi k f_0 t\right) \tag{1.11}$$

A periodic signal considered in (1.11) has infinite energy, but it has finite power.

One important property of the Fourier transform is that the transform preserves energy of a signal in time and frequency domains. The energy of the signal $g(t)$ is given by

$$\int_{-\infty}^{\infty} |g(t)|^2 \, dt = \int_{-\infty}^{\infty} |G(f)|^2 \, df \tag{1.12}$$

The energy spectral density $E_g(f) = |G(f)|^2$ is the measure of the energy of the signal contained within a narrow band of frequency around f divided by the width of the band. The autocorrelation function $r_g(\tau)$ of the signal defined by

$$r_g(\tau) = \int_{-\infty}^{\infty} g^*(t) g(t+\tau) dt \tag{1.13}$$

and the energy spectral density $E_g(f)$ of the signal form the Fourier transform pair, i.e.,

$$E_g(f) = \int_{-\infty}^{\infty} r_g(\tau) \exp(-j2\pi f \tau) d\tau \tag{1.14}$$

For a signal $g(t)$ with finite power, we define the autocorrelation function $r_g(\tau)$ as

$$r_g(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} g^*(t) g(t+\tau) dt \tag{1.15}$$

and the power spectral density $P_g(f)$ is given by the Fourier transform of $r_g(\tau)$, i.e.,

$$P_g(f) = \int_{-\infty}^{\infty} r_g(\tau) \exp(-j2\pi f \tau) d\tau \tag{1.16}$$

which is the measure of the power of the signal contained in the frequency range between f and $f + \Delta f$, divided by Δf, and setting $\Delta f \to 0$.

Let us now investigate the effect of finite length of observation time of the signal on the estimation of frequency. When the signal $g(t)$ is available for the time duration $t \le \left|\frac{T}{2}\right|$ for processing, we compute the Fourier transform of the windowed signal as follows,

$$G_W(f) = F[g(t)w(t)] = G(f) * W(f) \tag{1.17}$$

where $*$ stands for the convolution operation, and the window function is

$$w(t) = \begin{cases} 1, & t \le \left|\frac{T}{2}\right| \\ 0, & \text{elsewhere} \end{cases} \tag{1.18}$$

and its Fourier transform is given by

$$W(f) = \frac{\sin(\pi f T)}{\pi f} = T\,\mathrm{sinc}(fT) \tag{1.19}$$

The maximum of the sinc function occurs at $f = 0$, and its first zero-crossings appear at $f = \pm\frac{1}{T}$. When the signal $g(t)$ comprises of two complex exponentials at frequencies f_1 and f_2, and its Fourier transform is

$$G(f) = \delta(f - f_1) + \delta(f - f_2) \tag{1.20}$$

we compute the Fourier transform of the windowed signal as

$$G_W(f) = W(f - f_1) + W(f - f_2) \tag{1.21}$$

Thus, the two peaks at $f = f_1$ and $f = f_2$ in the plot of the square-magnitude of the Fourier transform can be resolved provided the difference of frequencies, $f_1 \sim f_2 \geq \frac{1}{T}$. In other words, the resolution of estimation of frequency by the method based on the Fourier transform is inversely proportional to the duration of observation time of the signal. This condition on the resolution of frequency can be of concern when the signal is observable only for short duration of time.

We shall now consider discrete-time signal for processing. Even for those cases where the signal $g(t)$ is observable over continuous time, we need to sample the signal over some discrete points of the time axis before a digital computer can be employed for signal processing. When a band-limited signal is sampled at a uniform time-interval T_s, the spectrum of the discrete-time signal becomes periodic with period T_s. The discrete-time Fourier transform $G_d(f)$ of the sequence $\{g_n = g(nT_s)\}$ is related to the continuous-time Fourier transform $G(f)$ as follows,

$$\begin{aligned} G_d(f) &\triangleq \sum_{n=-\infty}^{\infty} g_n \exp(-j2\pi f n T_s) \\ &= \frac{1}{T_s}\sum_{k=-\infty}^{\infty} G(f + k f_s) \end{aligned} \tag{1.22}$$

where $f_s = \frac{1}{T_s}$ is the sampling rate.

Assuming that $\frac{f_s}{2} \geq B$ where B is the bandwidth of the signal, i.e., $|G(f)|$ is non-zero only for $|f| \leq B$, we can low-pass filter $G_d(f)$ with gain T_s to get $G(f)$, and then compute the inverse Fourier transform to obtain the signal $g(t)$. In the process, we get the formula for reconstructing the original signal from its sampled sequence as given below,

$$g(t) = \sum_{n=-\infty}^{\infty} g_n\,\mathrm{sinc}(f_s t - n), \quad -\infty \leq t \leq \infty \tag{1.23}$$

When B is comfortably less than $\frac{f_s}{2}$, some other formula for reconstructing the signal may be more desirable because in that case, a generating function can be involved which is localized on the time-axis, unlike the sinc function which stretches for all time, $-\infty < t < \infty$. Note that the sinc function is the inverse Fourier transform of the rectangular window. By choosing a smooth transition-band between pass-band and stop-band of low-pass filtering, we get an appropriate kernel function in (1.23), which is time-limited.

Using the notation $g[n] = g(nT_s)$ for the discrete-time signal, we compute the autocorrelation function $r_g[l]$ of the signal with finite power as

$$r_g[l] = \lim_{M\to\infty} \frac{1}{2M+1} \sum_{n=-M}^{M} g^*[n]g[n+l] \tag{1.24}$$

The power spectral density $P_g(f)$ is now defined as

$$P_g(f) = \sum_{l=-\infty}^{\infty} r_g[l]\exp(-j2\pi fl), \quad -\frac{1}{2} \le f \le \frac{1}{2} \tag{1.25}$$

where the frequency variable f is the discrete frequency defined as the relative value in comparison with the sampling rate f_s. Thus, if f and f_s are known, then the frequency in Hertz can be determined.

When a finite number of signal samples $\{g[n]; n = 0,1,\ldots,N-1\}$ are available for processing, we compute the discrete Fourier transform (DFT) given by

$$G[k] = \sum_{n=0}^{N-1} g[n]\exp(-j2\pi kn/N), \quad k = 0,1,\ldots,N-1 \tag{1.26}$$

and the inverse DFT given by

$$g[n] = \frac{1}{N}\sum_{k=0}^{N-1} G[k]\exp(j2\pi kn/N), \quad n = 0,1,\ldots,N-1 \tag{1.27}$$

Since the total range of discrete frequency is sampled by the integer k at N points, the resolution of discrete frequency is $\frac{1}{N}$. When multiplied by $f_s = \frac{1}{T_s}$, the resolution of frequency in Hertz becomes $\frac{1}{T}$, where T is the duration of sampling in seconds. Once again, two close frequencies can be resolved by the DFT only if they are more than $\frac{1}{T}$ apart, which gives a condition similar to what we have obtained for the case of continuous-time signal.

When some finite number of signal samples $g[n]$ are available, we can compute the autocorrelation function $r_g[l]$ accurately only for some limited number of lags. In this case, we do not use (1.25) to compute the power spectral density $P_g(f)$ of the signal. For accurate estimation of $P_g(f)$ by (1.25), we need the sequence $r_g[l]$ for a large number of lags. Therefore, for accurate estimation of $P_g(f)$ with a short length of the sequence $r_g[l]$, we introduce the approach of model-based signal processing.

In the model-based signal processing approach, we first assume a model for the signal, and then we use the model for extracting information about the signal. Basically,

a model identifies a class of signals, and signal processing becomes the way of extracting features relevant to the class. Often a signal is identified as the output of a system with known input, and the model-based signal processing is done via system identification. A parametric model is the one which represents the signal in terms of a set of unknown parameters, which may be the coefficients of a set of known simple functions. Alternatively, a parametric model can be formulated by using a set of parameters related by some simple mathematical expression. For non-parametric modelling, we assume the form of the signal before extracting the signal properties.

A model-mismatch problem occurs where the signal does not fit the model used for it. For instance, if an algorithm is developed to estimate the pitch-frequency or the pitch-period of a signal assuming that the underlying process is periodic, then the algorithm cannot provide any sensible results when applied for an aperiodic signal. For parametric modelling, parsimony of parameters is an important point to ensure that parameter estimation can be done accurately. An error analysis can reveal how closely the signal is fitting the parametric model. We shall address many issues relevant to signal modelling in a later chapter when we consider various concepts of function representation.

In our approach of signal modelling, let the signal $g[n]$ be represented by an autoregressive (AR) process of order J defined by the linear difference equation

$$g[n] = -\sum_{k=1}^{J} \alpha_k g[n-k] + e[n] \tag{1.28}$$

where $\{-\alpha_k; k = 1, 2, \ldots, J\}$ are the AR coefficients and $e[n]$ is a zero-mean uncorrelated random sequence. Taking the z-transform of the sides, (1.28) can be rewritten in terms of the system function of an all-pole filter as follows,

$$\frac{G(z)}{E(z)} = \frac{1}{A(z)} = \frac{1}{\sum_{k=0}^{J} \alpha_k z^{-k}} \tag{1.29}$$

where $G(z)$ and $E(z)$ are the z-transforms of the response $g[n]$ and the excitation $e[n]$ respectively, and $\alpha_0 = 1$.

By definition of the AR process, the autocorrelation function of $e[n]$ is given by

$$r_e[l] = \sigma^2 \delta[l] \tag{1.30}$$

with σ^2 being an unknown constant, and the cross-correlation of $g[n]$ and $e[n]$ is given by

$$r_{ge}[l] = \lim_{M \to \infty} \frac{1}{2M+1} \sum_{n=-M}^{M} g^*[n-l] e[n] = \sigma^2 \delta[l] \tag{1.31}$$

Multiplying both sides of (1.28) by $g^*[n-l]$ and performing the time averaging as in (1.24) or (1.31), we get the following relation

$$r_g[l] = -\sum_{k=1}^{J} \alpha_k \, r_g[l-k] + \sigma^2 \delta[l] \tag{1.32}$$

Assuming that the sequence of autocorrelation functions $\{r_g[l]; l = -J, \ldots, 0, \ldots, J\}$ is available for processing, from (1.32) we write the matrix equation

$$\begin{bmatrix} r_g[0] & r_g[-1] & \cdots & r_g[-(J-1)] \\ r_g[1] & r_g[0] & \cdots & r_g[-(J-2)] \\ \vdots & \vdots & \ddots & \vdots \\ r_g[J-1] & r_g[J-2] & \cdots & r_g[0] \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_J \end{bmatrix} = -\begin{bmatrix} r_g[1] \\ r_g[2] \\ \vdots \\ r_g[J] \end{bmatrix} \tag{1.33}$$

which is solved for the vector containing the AR coefficients α_k. Next we estimate the variance of excitation σ^2 as

$$\sigma^2 = r_g[0] + \sum_{k=1}^{J} \alpha_k \, r_g[-k] \tag{1.34}$$

The power spectral density $P_g(f)$ of the output of the AR process is related to the power spectral density $P_e(f)$ of the input as follows,

$$P_g(f) = \frac{1}{|A(f)|^2} P_e(f) \tag{1.35}$$

where $A(f) = A(z)\big|_{z=e^{j2\pi f}} = \sum_{k=0}^{J} \alpha_k \, e^{-j2\pi kf}$ and $P_e(f) = \sigma^2$ from (1.30). Rewriting (1.35), we get

$$P_g(f) = \frac{\sigma^2}{\left|1 + \sum_{k=1}^{J} \alpha_k e^{-j2\pi kf}\right|^2} \tag{1.36}$$

Note that although the autocorrelation sequence $r_g[l]$ is available for the lags $-J \le l \le J$, we can extrapolate the sequence $r_g[l]$ for lags $J+1, J+2, \ldots$ by using the relation (1.32) with the estimated values of α_k. This implied extension of the autocorrelation sequence beyond the window of available lags is the reason why the computation of $P_g(f)$ by (1.36) does not put any restriction on resolution of frequency. If we compare the present situation with the other option of computing $P_g(f)$ by (1.25) using the autocorrelation sequence of available lags, then we see that the advantage of the model-based approach is to remove the restriction on resolution of frequency.

In an approach of fitting an equivalent model, we propose that the signal samples $\{g[n]; n = 0, 1, \ldots, N-1\}$ be fitted into a linear predictor of order J. Then, the predicted signal samples $\hat{g}[n]$ are given by

$$\hat{g}[n] = -\sum_{k=1}^{J} \alpha_k \, g[n-k], \quad n \ge J \tag{1.37}$$

where now $\{-\alpha_k; k = 1, 2, \ldots, J\}$ are the prediction coefficients. The prediction error sequence $e[n]$ is expressed as

$$\begin{aligned} e[n] &= g[n] - \hat{g}[n] \\ &= g[n] + \sum_{k=1}^{J} \alpha_k g[n-k], \ n \geq J \end{aligned} \tag{1.38}$$

and the power of prediction error ρ is given by

$$\begin{aligned} \rho &= \frac{1}{N-J} \sum_{n=J}^{N-1} |e[n]|^2 \\ &= \frac{1}{N-J} \sum_{n=J}^{N-1} \left(g^*[n] + \sum_{l=1}^{J} \alpha_l^* g^*[n-l] \right) \left(g[n] + \sum_{k=1}^{J} \alpha_k g[n-k] \right) \end{aligned} \tag{1.39}$$

Setting $\frac{\partial \rho}{\partial \alpha_l^*} = 0$ for minimization of ρ, we get

$$\frac{1}{N-J} \sum_{n=J}^{N-1} g^*[n-l] \left(g[n] + \sum_{k=1}^{J} \alpha_k g[n-k] \right) = 0 \quad \text{for} \quad l = 1, 2, \ldots, J \tag{1.40}$$

Defining the autocorrelation functions for finite-length signal sequence appropriately, (1.40) is rewritten as

$$r_g[l] + \sum_{k=1}^{J} \alpha_k \, r_g[l-k] = 0 \quad \text{for} \quad l = 1, 2, \ldots, J \tag{1.41}$$

Substituting (1.41) in (1.39), we get the minimum prediction-error power $\rho_{\min}$ as

$$\rho_{\min} = r_g[0] + \sum_{k=1}^{J} \alpha_k \, r_g[-k] \tag{1.42}$$

Comparing (1.41) and (1.42) with (1.32) and (1.34), we find that an AR process and a linear predictor, both having the same order, essentially model a signal in the same way so far the inter-relationship of the autocorrelation functions is expressed, and hence, the power spectral density of the signal will be represented in the same form whether the signal is modelled by an AR process or a linear predictor. Now, taking the z-transform of both the sides, (1.38) can be rewritten in the form of the system function of the prediction-error filter (PEF) as follows,

$$\frac{E(z)}{G(z)} = A(z) = \sum_{k=0}^{J} \alpha_k z^{-k} \quad \text{with} \quad \alpha_0 = 1 \tag{1.43}$$

which is the inverse-system of the AR process of same order. The zeros of the PEF evaluated by extracting the roots of the polynomial equation, $A(z) = 0$, are related to the frequencies f_i present in the signal as $z_i = \exp(j2\pi f_i); i = 1, 2, \ldots, J$.

Although the estimation of frequencies of a signal by plotting the AR-PSD or by computing the zeros of the PEF polynomial equation can provide optimal resolution with

finite data-length, the resolution can be adversely affected with presence of noise. It has been shown empirically that the resolution of the AR spectral estimator decreases with the decrease of the signal-to-noise ratio (SNR). The reason for this degradation can be given as a model-mismatch problem. In other words, the noisy signal does not fit the all-pole model, unless we use a large model order. An AR process with a large model-order models the signal as well as the additive noise. However, the use of a large model-order will require the autocorrelation sequence to be computed for a large number of lags, which in turn, will require that a larger number of signal samples be available for processing. Therefore with finite data-length, the resolution of the AR spectral estimator may not be acceptable at low SNR level.

We now describe the approach of parametric modelling for spectral estimation. Let the sequence of signal y_n be represented by a weighted sum of complex exponentials with or without additive noise,

$$y_n = \sum_{i=1}^{M} b_i z_i^n + w_n \,;\, n = 0, 1, \ldots, N-1 \tag{1.44}$$

where $b_i = |b_i| \angle \theta_i$ is a complex number denoting in polar coordinate the amplitude and phase of the ith complex exponential, $z_i = \exp(j2\pi f_i)$, f_i being the frequency, and w_n is the noise which may be present in the signal samples. While estimating the frequency-parameters of the model, we set initially the noise sequence $w_n = 0$, and later, we modify the estimation procedure to take account of the presence of noise.

Letting $w_n = 0$ in (1.44), we form the polynomial equation

$$a_M \prod_{i=1}^{M} (z - z_i) = 0 \quad \Rightarrow \quad a_M z^M + a_{M-1} z^{M-1} + \cdots + a_1 z + a_0 = 0 \tag{1.45}$$

whose roots $z_i, i = 1, 2, \ldots, M$ can be extracted for frequency estimation provided the unknown coefficients $a_0, a_1, \ldots, a_M$ are determined first. Utilizing the N samples of the sequence y_n, we can write the following matrix equation according to (1.45),

$$\mathbf{Ya} = \mathbf{0} \quad \Rightarrow \quad \begin{bmatrix} y_0 & y_1 & \cdots & y_M \\ y_1 & y_2 & \cdots & y_{M+1} \\ \vdots & \vdots & & \vdots \\ y_{N-M-1} & y_{N-M} & \cdots & y_{N-1} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_M \end{bmatrix} = \mathbf{0} \tag{1.46}$$

where $N \geq 2M$ to ensure that the matrix $\mathbf{Y}$ has at least M rows. In order to solve (1.46), we assume, without loss of generality, that one of the coefficients $a_0, \ldots, a_M$ is unity. Setting $a_M = 1$, we rewrite (1.46) as follows,

$$\begin{bmatrix} y_0 & y_1 & \cdots & y_{M-1} \\ y_1 & y_2 & \cdots & y_M \\ \vdots & \vdots & & \vdots \\ y_{N-M-1} & y_{N-M} & \cdots & y_{N-2} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{M-1} \end{bmatrix} = -\begin{bmatrix} y_M \\ y_{M+1} \\ \vdots \\ y_{N-1} \end{bmatrix} \quad \Rightarrow \quad \mathbf{Y}_0 \bar{\mathbf{a}} = \mathbf{y} \tag{1.47}$$