Article type: Focus Article

Transfer Functions Article ID

Stephen Pollock

University of Leicester

Keywords

Impulse response, Frequency response, Spectral density

Abstract

In statistical time-series analysis, signal processing and control engineering, a transfer function is a mathematical relationship between a numerical input to a dynamic system and the resulting output. The theory of transfer functions describes how the input/output relationship is affected by the structure of the transfer function.

The theory of the transfer functions of linear time-invariant (LTI) systems has been available for many years. It was developed originally in connection with electrical and mechanical systems described in continuous time. The basic theory can be attributed largely to Oliver Heaviside (1850–1925) [3] [4].

With the advent of digital signal processing, the emphasis has shifted to discrete-time representations. These are also appropriate to problems in statistical time-series analysis, where the data are in the form of sequences of stochastic values sampled at regular intervals.

REPRESENTATIONS OF THE TRANSFER FUNCTION

In the discrete case, a univariate and causal transfer function mapping from an input sequence $\{x_t\}$ to an output sequence $\{y_t\}$ can be represented by the equation

$$\sum_{j=0}^{p} a_j y_{t-j} = \sum_{j=0}^{q} b_j x_{t-j}, \quad \text{with} \quad a_0 = 1. \tag{1}$$

Here, the condition that $a_0 = 1$ serves to identify y_t as the current output and the elements $y_{t-1}, \ldots, y_{t-p}$ as feedback or as lagged dependent variables. The sum of the input variables on RHS of the equation, weighted by their coefficients, is described as a distributed lag scheme. (See Dhrymes [2] for a treatment of distributed lags in the context of econometric estimation.) The condition of causality implies that x_{t+j} and y_{t+j}, which are ahead of time t, are excluded from the equation.

Consider T realisations of the equation (1), with the successive outputs indexed by $t = 0, \ldots, T-1$. By associating each equation with the corresponding power z^t of an indeterminate algebraic quantity z and by adding them together, a polynomial equation is derived that can be denoted by

$$a(z)y(z) = b(z)x(z) \quad \text{or, equivalently,} \quad y(z) = \frac{b(z)}{a(z)}x(z) = \psi(z)x(z). \qquad (2)$$

Here,

$$\begin{aligned} y(z) &= y_{-p}z^{-p} + \cdots + y_0 + y_1 z + \cdots + y_{T-1}z^{T-1}, \\ x(z) &= x_{-q}z^{-q} + \cdots + x_0 + x_1 z + \cdots + x_{T-1}z^{T-1}, \\ a(z) &= 1 + a_1 z + \cdots + a_p z^p \quad \text{and} \\ b(z) &= b_0 + b_1 z + \cdots + b_q z^q \end{aligned} \qquad (3)$$

are described as the z-transforms of the corresponding sequences. This representation allows the algebra of polynomials to be deployed in analysing the dynamic system. An extensive account of the z-transform has been provided by Jury [5].

Equation (2) comprises the pre-sample elements $y_{-p}, \ldots, y_{-1}$ and $x_{-q}, \ldots, y_{-1}$ of the input and output sequences, which provide initial conditions for the system. However, if the transfer function is stable in the sense that a bounded input will lead to a bounded output—described as the BIBO condition—then it is permissible to extend the two sequences backwards in time indefinitely, as well as forwards in time. Then, $y(z)$ and $x(z)$ become infinite series; and the matter of the initial conditions can be ignored.

THE IMPULSE RESPONSE

The transfer function can be characterised by its effect on certain elementary reference signals. The simplest of these is the impulse sequence, which is defined by

$$\delta_t = \begin{cases} 1, & \text{if } t = 0; \\ 0, & \text{if } t \neq 0. \end{cases} \qquad (4)$$

The corresponding z-transform is $\delta(z) = 1$. The output generated by the impulse is described as the impulse response function. For an ordinary causal transfer function, which responds only to present and previous values of the input and output sequences, the zero-valued inputs and responses at times $t < 0$ can be ignored.

On substituting $x(z) = \delta(z) = 1$ into equation (2), it can be seen that calculating the impulse response is a matter of finding coefficients of the series expansion of the rational function $\psi(z) = b(z)/a(z)$.

When $a(z) = 1$, the impulse response is just the sequence of the coefficients of $b(z)$. Then, there is a finite impulse response (FIR) transfer function. When $a(z) \neq 1$, the

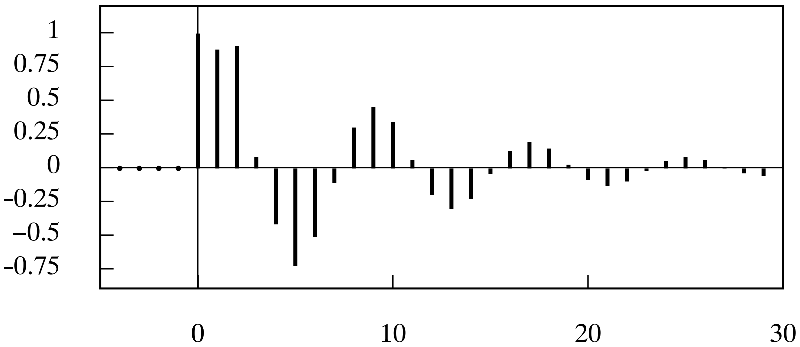

Figure 1: The impulse response of the transfer function $b(z)/a(z)$ with $a(z) = 1.0 - 0.673z + 0.463z^2 + 0.486z^3$ and $b(z) = 1.0 + 0.208z + 0.360z^2$.

impulse response is liable to continue indefinitely, albeit that, in a stable system, it will converge to zero. Then, there is an infinite impulse response (IIR) transfer function.

When $a(z) \neq 1$, the impulse response can be found via the equation $b(z) = a(z)\psi(z)$. An example is provided by the case where $p = 2$ and $q = 1$. Then,

$$b_0 + b_1 z = \{1 + a_1 z + a_2 z^2\}\{\psi_0 + \psi_1 z + \psi_2 z^2 + \cdots\}. \tag{5}$$

By performing the multiplication on the RHS, and by equating the coefficients of the same powers of z on the two sides of the equation, it is found that

$$\begin{aligned} b_0 &= \psi_0, & \psi_0 &= b_0, \\ b_1 &= \psi_1 + a_1\psi_0, & \psi_1 &= b_1 - a_1\psi_0, \\ 0 &= \psi_2 + a_1\psi_1 + a_2\psi_0, & \psi_2 &= -a_1\psi_1 - a_2\psi_0, \\ &\vdots & &\vdots \\ 0 &= \psi_n + a_1\psi_{n-1} + a_2\psi_{n-2}, & \psi_n &= -a_1\psi_{n-1} - a_2\psi_{n-2}. \end{aligned} \tag{6}$$

The resulting sequence is just the recursive solution of the homogenous difference equation $y_t + a_1 y_{t-1} + a_2 y_{t-2} = 0$, subject to the initial conditions $y_0 = b_0$ and $y_1 = b_1 - a_1 y_0$.

The transfer function is fully characterised by its response to an impulse. One is reminded that all of the harmonics of a bell are revealed when it is stuck by a single blow of the clapper, which constitutes an impulse. Figure 1 provides an example of an impulse response.

STABILITY

The stability of a rational transfer function $b(z)/a(z)$ can be investigated using its partial-fraction decomposition, which gives rise to a sum of simpler transfer functions that can be analysed readily.

If the degree of the numerator of $b(z)/a(z)$ exceeds that of the denominator, then long division can be used to obtain a quotient polynomial and a remainder that is a proper rational function. The quotient polynomial will correspond to a stable transfer function; and the remainder will be the subject of the decomposition.

Assume that $b(z)/a(z)$ is a proper rational function in which the denominator is factorised as

$$a(z) = \prod_{j=1}^{r}(1 - z/\lambda_j)^{n_j}, \tag{7}$$

where n_j is the multiplicity of the root λ_j, and where $\sum_j n_j = p$ is the degree of the polynomial. Then, the so-called Heaviside partial-fraction decomposition is

$$\frac{b(z)}{a(z)} = \sum_{j=1}^{r}\sum_{k=1}^{n_j}\frac{c_{jk}}{(1 - z/\lambda_j)^k}; \tag{8}$$

and the task is to find the series expansions of the partial fractions. (See Arfken [1] and Taylor and Mellott [10] for references to the Heaviside expansion.)

Consider, first, the case of a partial fraction that contains a distinct (unrepeated) real-valued root λ. The expansion is

$$\frac{c}{1 - z/\lambda} = c\{1 + z/\lambda + (z/\lambda)^2 + \cdots\}. \tag{9}$$

For this to converge for all $|z| \leq 1$, it is necessary and sufficient that $|\lambda| > 1$; and this is necessary and sufficient for the satisfaction of the BIBO condition which is that the impulse response function must be absolutely summable:

$$\sum_j |\psi_j| < \infty. \tag{10}$$

Here, we are setting $\psi_j = (1/\lambda)^j$. For a proof of that (10) is the BIBO condition for an LTI system, see, for example, Kac [6] or Pollock [7].

Next, consider the case where $a(z)$ has a distinct pair of conjugate complex roots λ and λ^*. These will come from a partial fraction with a quadratic denominator:

$$\frac{cz + d}{(z - \lambda)(z - \lambda^*)} = \frac{\kappa}{z - \lambda} + \frac{\kappa^*}{z - \lambda^*}. \tag{11}$$

It can be seen that $\kappa = (c\lambda + d)/(\lambda - \lambda^*)$ and $\kappa^* = (c\lambda^* + d)/(\lambda^* - \lambda)$ are also conjugate complex numbers.

The expansion of (9) applies to complex roots as well as to real roots:

$$\begin{aligned}\frac{c}{1 - z/\lambda} + \frac{c^*}{1 - z/\lambda^*} \quad = \quad & c\left\{1 + z/\lambda + (z/\lambda)^2 + \cdots\right\} \\ & + c^*\left\{1 + z/\lambda^* + (z/\lambda^*)^2 + \cdots\right\}\end{aligned}$$

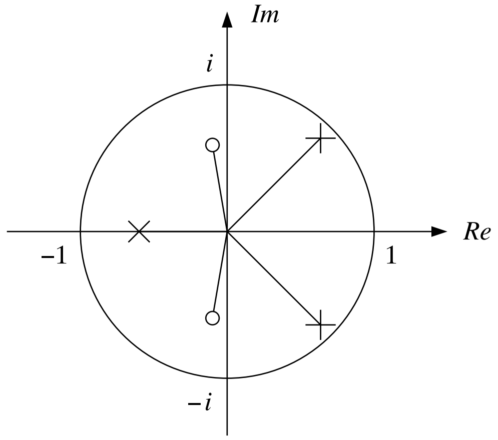

Figure 2: The pole–zero diagram for the transfer function $b(z^{-1})/a(z^{-1})$ corresponding to the impulse response function of Figure 1. The poles are marked by crosses and the zeros by circles.

$$= \sum_{t=0}^{\infty} z^t(c\lambda^{-t} + c^*\lambda^{*-t}). \tag{12}$$

The various complex quantities can be represented in terms of exponentials:

$$\begin{aligned} \lambda &= \kappa^{-1}e^{-\mathrm{i}\omega}, & \lambda^* &= \kappa^{-1}e^{\mathrm{i}\omega}, \\ c &= \rho e^{-\mathrm{i}\theta}, & c^* &= \rho e^{\mathrm{i}\theta}. \end{aligned} \tag{13}$$

Then, the generic term of the expansion becomes

$$\begin{aligned} z^t(c\lambda^{-t} + c^*\lambda^{*-t}) &= z^t\rho\kappa^t\big\{e^{\mathrm{i}(\omega t-\theta)} + e^{-\mathrm{i}(\omega t-\theta)}\big\} \\ &= z^t 2\rho\kappa^t\cos(\omega t - \theta). \end{aligned} \tag{14}$$

The expansion converges for all $|z| \leq 1$ if and only if $|\kappa| < 1$, which is a condition on the modulus of the complex number λ. But, $|\kappa| = |\lambda^{-1}| = |\lambda|^{-1}$; so it is confirmed that the necessary and sufficient condition for convergence is that $|\lambda| > 1$.

Finally, consider the case of a repeated root with a multiplicity of n. Then, a binomial expansion is available that gives

$$\frac{1}{(1 - z/\lambda)^n} = 1 + n\frac{z}{\lambda} + \frac{n(n+1)}{2!}\left(\frac{z}{\lambda}\right)^2 + \frac{n(n+1)(n+2)}{3!}\left(\frac{z}{\lambda}\right)^3 + \cdots. \tag{15}$$

If λ is real, then $|\lambda| > 1$ is the condition for convergence. If λ is complex, then it can be combined with the conjugate root in the manner of (13) to create a trigonometric function; and, again, the condition for convergence is that $|\lambda| > 1$.

This result can be understood by regarding the LHS of (15) as a representation of n transfer functions in series, each of which fulfils the BIBO condition.

The general conclusion is that the transfer function is stable if and only if all of the roots of the denominator polynomial $a(z)$, which are described as the poles of the transfer function, lie outside the unit circle in the complex plane.

It is helpful to represent the poles of the transfer function graphically by showing their locations within the complex plane together with the locations of the roots of the numerator polynomial, which are described as the zeros of the transfer function.

It is more convenient to represent the poles and zeros of $b(z^{-1})/a(z^{-1})$, which are the reciprocals of those of $b(z)/a(z)$. For a stable and invertible transfer function, these must lie within the unit circle. This recourse has been adopted for Figure 2, which shows the pole–zero diagram for the transfer function that gives rise to Figure 1.

THE RESPONSE TO A COSINE

One must also consider the response of the transfer function to a simple sinusoidal signal. Any finite data sequence can be expressed as a sum of discretely sampled sine and cosine functions with frequencies that are integer multiples of a fundamental frequency that produces one cycle in the period spanned by the sequence. The finite sequence is to be regarded as a single cycle within a infinite sequence, which is the periodic extension of the data.

Consider, therefore, the consequences of mapping the signal sequence $\{x_t = \cos(\omega t)\}$ through the transfer function with the coefficients $\{\psi_0, \psi_1, \ldots\}$. The output is

$$y(t) = \sum_j \psi_j \cos\big(\omega[t-j]\big). \tag{16}$$

By virtue of the trigonometrical identity $\cos(A-B) = \cos A \cos B + \sin A \sin B$, this becomes

$$\begin{aligned} y(t) &= \Big\{\sum_j \psi_j \cos(\omega j)\Big\}\cos(\omega t) + \Big\{\sum_j \psi_j \sin(\omega j)\Big\}\sin(\omega t) \\ &= \alpha\cos(\omega t) + \beta\sin(\omega t) = \rho\cos(\omega t - \theta), \end{aligned} \tag{17}$$

where

$$\begin{aligned} \alpha &= \sum_j \psi_j \cos(\omega j), \qquad \beta = \sum_j \psi_j \sin(\omega j), \\ \rho^2 &= \alpha^2 + \beta^2 \qquad \text{and} \qquad \theta = \tan^{-1}\Big(\frac{\beta}{\alpha}\Big). \end{aligned} \tag{18}$$

It can be seen that the transfer function has a twofold effect upon the signal. First, there is a *gain effect* whereby the amplitude of the sinusoid is increased or diminished by the factor ρ. Then, there is a *phase effect* whereby the peak of the sinusoid is displaced by a time delay of θ/ω periods. The frequency of the output is the same as the frequency of the input, which is a fundamental feature of all linear dynamic systems.

Observe that the response of the transfer function to a sinusoid of a particular frequency is akin to the response of a bell to a tuning fork. It gives very limited information regarding the characteristics of the system. To obtain full information, it is necessary to excite the system over the whole range of frequencies.

SPECTRAL DENSITY AND THE FREQUENCY RESPONSE

In a discrete-time system, there is a problem of aliasing whereby signal frequencies (i.e. angular velocities) in excess of π radians per sampling interval are confounded with frequencies within the interval $[0, \pi]$. To understand this, consider a cosine wave of unit amplitude and zero phase with a frequency ω in the interval $\pi < \omega < 2\pi$ that is sampled at unit intervals. Let $\omega^* = 2\pi - \omega$. Then

$$\begin{aligned}\cos(\omega t) &= \cos\big\{(2\pi - \omega^*)t\big\} \\ &= \cos(2\pi)\cos(\omega^* t) + \sin(2\pi)\sin(\omega^* t) \\ &= \cos(\omega^* t);\end{aligned} \tag{19}$$

which indicates that ω and ω^* are observationally indistinguishable. Here, $\omega^* \in [0, \pi]$ is described as the alias of $\omega > \pi$.

The maximum frequency in discrete data is π radians per sampling interval and, as the Shannon–Nyquist sampling theorem indicates, aliasing is avoided only if there are at least two observations in the time that it takes the signal component of highest frequency to complete a cycle. In that case, the discrete representation will contain all of the available information on the system. (The classical article of Shannon [9] that conveys this result is readily available in its reprinted form.)

Any stationary stochastic process defined over the doubly infinite set of positive and negative integers can be expressed as a weighted combination of the non-denumerable infinity of sines and cosines that have frequencies in Nyquist interval $[0, \pi]$. Thus, if x_t is an element of such a process, then it can be represented by

$$x_t = \int_0^\pi \Big\{\cos(\omega t)dA(\omega) + \sin(\omega t)dB(\omega)\Big\} = \int_{-\pi}^{\pi} e^{\mathrm{i}\omega t} dZ(\omega). \tag{20}$$

This is commonly described as the spectral representation of the process generating x_t. Here, $dA(\omega)$ and $dB(\omega)$ are the infinitesimal increments of stochastic functions, defined on the frequency interval, that are everywhere continuous but nowhere differentiable. Moreover, it is assumed that the increments $dA(\omega)$ and $dB(\omega)$ are uncorrelated with each other and with preceding and succeeding increments. (See Pollock [7] or Priestley [8] for fuller accounts.)

The concise expression on the RHS of (20) entails the following definitions:

$$dZ(\omega) = \frac{dA(\omega) - \mathrm{i}dB(\omega)}{2} \quad \text{and} \quad dZ(-\omega) = dZ^*(\omega) = \frac{dA(\omega) + \mathrm{i}dB(\omega)}{2}. \tag{21}$$

The expression in terms of sines and cosines can be recovered via the identities

$$e^{\mathrm{i}\omega t} = \cos(\omega t) + \mathrm{i}\sin(\omega t) \quad \text{and} \quad e^{-\mathrm{i}\omega t} = \cos(\omega) - \mathrm{i}\sin(\omega t). \tag{22}$$

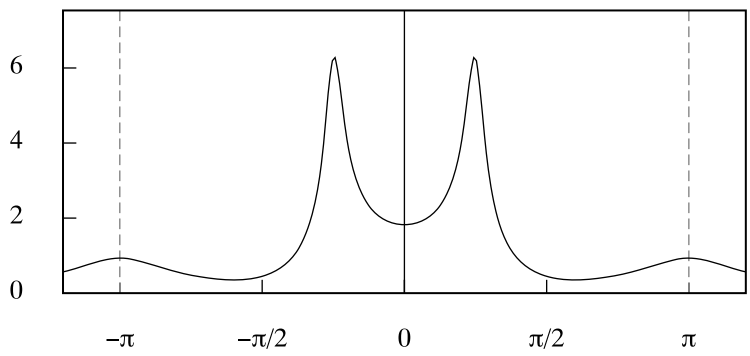

Figure 3: The gain of the transfer function depicted in Figures 1 and 2.

The spectral density function $f(\omega)$ of the process is defined by

$$E\{dZ(\omega)dZ^*(\omega)\} = E\{dA^2(\omega) + dB^2(\omega)\} = f(\omega)d\omega. \tag{23}$$

The quantity $f(\omega)d\omega$ is the power or the variance of the process that is attributable to the elements in the frequency interval $[\omega, \omega + d\omega]$; and the integral of $f(\omega)$ over the frequency range $[-\pi, \pi]$ is the overall variance of the process. A white-noise process $\{\varepsilon_t\}$ of independently and identically distributed elements, with a variance of σ_ε^2, has a uniform spectral density function $f_\varepsilon(\omega) = \sigma_\varepsilon^2/(2\pi)$.

Let $\{\psi_0, \psi_1, \ldots\}$ be the impulse response of the transfer function. Then the effects of the transfer function upon the spectral elements of the process defined by (20) are shown by the equation

$$\begin{aligned} y_t &= \sum_j \psi_j x(t-j) = \sum_j \psi_j \left\{ \int_\omega e^{\mathrm{i}\omega(t-j)} dZ_x(\omega) \right\} \\ &= \int_\omega e^{\mathrm{i}\omega t} \left(\sum_j \psi_j e^{-\mathrm{i}\omega j} \right) dZ_x(\omega). \end{aligned} \tag{24}$$

These effects are summarised by the complex-valued frequency-response function

$$\psi(\omega) = \sum \psi_j e^{-\mathrm{i}\omega j} = |\psi(\omega)| e^{-\mathrm{i}\theta(\omega)}, \tag{25}$$

The final expression, which is in polar form, entails the following definitions:

$$|\psi(\omega)|^2 = \left\{ \sum_{j=0}^{\infty} \psi_j \cos(\omega j) \right\}^2 + \left\{ \sum_{j=0}^{\infty} \psi_j \sin(\omega j) \right\}^2$$

$$\theta(\omega) \quad = \quad \arctan\left\{\frac{\sum \psi_j \sin(\omega j)}{\sum \psi_j \cos(\omega j)}\right\}. \tag{26}$$

The two components of the frequency response are the amplitude response or the gain $|\psi(\omega)|$ and the phase response $\theta(\omega)$. Their definitions subsume those of (18). These two functions of ω, which are circular or periodic, are plotted in Figures 3 and 4 over the interval $[-\pi, \pi]$.

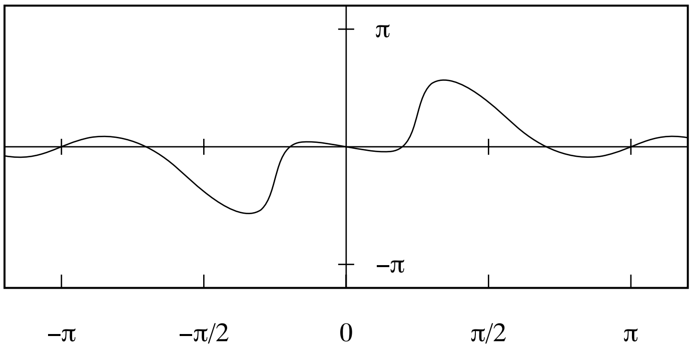

Figure 4: The phase plot to accompany Figure 3.

In general, the phase response does not relate in any simple way to the lags of the transfer function relationship that are perceptible via the impulse response function. An exception concerns the transfer function $\psi(z) = z$ that imposes a delay of one period. Then, the phase response is a line of unit slope passing through the origin, rising to π when $\omega = \pi$ and falling to $-\pi$ when $\omega = -\pi$; and the absolute time delay that is imposed on all frequencies is $\tau = \theta(\omega)/\omega = 1$.

The frequency response is just the discrete-time Fourier transform of the impulse response function. Therefore, the quantities of (25) and (26) can be expressed in terms of the z-transform of the functions with $z = e^{-\mathrm{i}\omega}$. They can be obtained in this way from the following expressions:

$$\psi(z) = \sum_j \psi_j z^j, \quad |\psi(z)|^2 = \psi(z)\psi(z^{-1}), \quad \theta(z) = \text{Arg}\{\psi(z)\}. \tag{27}$$

In that case, there is a minor abuse of notation when we write $\psi(\omega)$ in place of $\psi(e^{-\mathrm{i}\omega})$.

As ω progresses from $-\pi$ to π, or, equally, as $z = e^{-\mathrm{i}\omega}$ travels around the unit circle in the complex plane, the frequency-response function defines a trajectory that becomes a closed contour when ω reaches π.

The points on the trajectory are characterised by their polar co-ordinates. These are the modulus $|\psi(\omega)|$, which is the length of the radius vector joining $\psi(\omega)$ to the origin, and

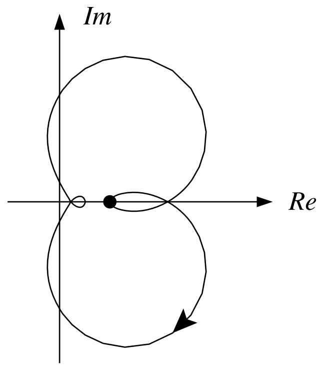

Figure 5: The path described in the complex plane by the frequency response function corresponding to the gain and phase functions of Figures 3 and 4. The trajectory originates, when $\omega = 0$, in the point on the real axis marked by a dot and it travels in the direction of the arrow, of which the tip is reached when $\omega = \pi/4$.

the argument $\text{Arg}\{\psi(\omega)\} = \theta(\omega)$ which is the (anticlockwise) angle in radians which the radius makes with the positive real axis. Figure 5 provides an illustration.

The spectral density $f_y(\omega)$ of the output process $y(t)$ is given by

$$\begin{aligned} f_y(\omega)d\omega &= E\{dZ_y(\omega)dZ_y^*(\omega)\} \\ &= \psi(\omega)\psi^*(\omega)E\{dZ_x(\omega)dZ_x^*(\omega)\} \\ &= |\psi(\omega)|^2 f_x(\omega)d\omega. \end{aligned} \tag{28}$$

When the input sequence is a white-noise process with a uniform spectral density function $f_\varepsilon(\omega) = \sigma_\varepsilon^2/(2\pi)$, this becomes the spectral density function of an autoregressive moving-average (ARMA) process. In the notation of the z-transform, the autocovariance generating function of the ARMA process is

$$\gamma(z) = \sigma_\varepsilon^2 \frac{b(z)b(z^{-1})}{a(z)a(z^{-1})} = \sigma_\varepsilon^2 \psi(z)\psi(z^{-1}). \tag{29}$$

Setting $z = e^{-\mathrm{i}\omega}$ and dividing by 2π gives the spectral density function in the form of

$$f_y(\omega) = \frac{\gamma(e^{-\mathrm{i}\omega})}{2\pi}. \tag{30}$$

In calculating this, we would evaluate the numerator and the denominator of $\gamma(e^{-\mathrm{i}\omega})$ separately.

Conclusion

The theory of linear time-invariant transfer functions is part of the basic grammar of systems analysis. Whereas it finds numerous applications in electrical and mechanical engineering, it is also used in statistical time-series analysis in connection with stationary stochastic processes.

In circumstances where the assumption of stationarity is unsustainable, linear dynamic models need to be replaced by more sophisticated models. The linear theory provides a point of departure for such developments.

References

[1] Arfken, G. *Mathematical Methods for Physicists, Third Edition*. Academic Press, San Diego, California, 1986.

[2] Dhrymes, P.J. *Distributed Lags: Problems of Estimation and Formulation*. Holden-Day, San Francisco 1971.

[3] Heaviside, O. *Electrical Papers*. American Mathematical Society, 1970.

[4] Heaviside, O. *Electromagnetic Theory*. E. Benn, London 1925, reprinted by American Mathematical Society, 1971.

[5] Jury, E.I. *Theory and Application of the z-Transform Method*. John Wiley and Sons, New York, 1964.

[6] Kac, R. *Introduction to Digital Signal Processing*. McGraw-Hill Book Co., New York, 1982.

[7] Pollock, D.S.G. *A Handbook of Time-Series Analysis, Signal Processing and Dynamics*. Academic Press, London 1999.

[8] Priestley, M.B. *Spectral Analysis and Time Series,* Academic Press, London, 1981.

[9] Shannon, C.E. Communication in the Presence of Noise, *Proceedings of the Institute of Radio Engineers,* 37, 10–21, reprinted in *Proceedings of the IEEE,* 86, 447–457, 1998.

[10] Taylor, F. and Mellott, J. *Hands-on Digital Signal Processing*. McGraw-Hill, New York, 1998.

Cross-References

Autoregressive process, Discrete Fourier transform, Filtering for time series and signals