Chapter 5

STOCHASTIC PROCESSES

A stochastic process refers to a family of random variables indexed by a parameter set. This parameter set can be continuous or discrete. Since we are interested in discrete systems, we will limit our discussion to processes with a discrete parameter set. Hence, a stochastic process in our context is a time sequence of random variables.

5.1 BASIC PROBABILITY CONCEPTS

5.1.1 DISTRIBUTION FUNCTION

Let $x(k)$ be a sequence. Then, $(x(k_1), \cdots, x(k_\ell))$ form an ℓ-dimensional random variable. Then, one can define the finite dimensional distribution function and the density function as before. For instance, the distribution function $F(\lambda_1, \cdots, \lambda_\ell; x(k_1), \cdots, x(k_\ell))$, is defined as:

$$F(\lambda_1, \cdots, \lambda_\ell; x(k_1), \cdots, x(k_\ell)) = \Pr\{x(k_1) \leq \lambda_1, \cdots, x(k_\ell) \leq \lambda_\ell\} \qquad (5.1)$$

The density function is also defined similarly as before.

We note that the above definitions also apply to *vector* time sequences if

$x(k_i)$ and λ_i's are taken as vectors and each integral is defined over the space that λ_i occupies.

5.1.2 MEAN AND COVARIANCE

Mean value of the stochastic variable $x(k)$ is

$$\bar{x}(k) = E\{x(k)\} = \int_{-\infty}^{\infty} \lambda dF(\lambda; x(k)) \tag{5.2}$$

Its covariance is defined as

$$\begin{aligned} R_x(k_1,k_2) &= E\{[x(k_1)-\bar{x}(k_1)][x(k_2)-\bar{x}(k_2)]^T\} \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[\lambda_1-\bar{x}(k_1)][\lambda_2-\bar{x}(k_2)]^T dF(\lambda_1,\lambda_2;x(k_1),x(k_2)) \end{aligned} \tag{5.3}$$

The cross-covariance of two stochastic processes $x(k)$ and $y(k)$ are defined as

$$\begin{aligned} R_{xy}(k_1,k_2) &= E\{[x(k_1)-\bar{x}(k_1)][y(k_2)-\bar{y}(k_2)]^T\} \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[\lambda_1-\bar{x}(k_1)][\lambda_2-\bar{y}(k_2)]^T dF(\lambda_1,\lambda_2;x(k_1),y(k_2)) \end{aligned} \tag{5.4}$$

Gaussian processes refer to the processes of which any finite-dimensional distribution function is normal. Gaussian processes are completely characterized by the mean and covariance.

5.1.3 STATIONARY STOCHASTIC PROCESSES

Throughout this book we will define *stationary* stochastic processes as those with time-invariant distribution function. *Weakly stationary* (or stationary in a wide sense) processes are processes whose first two moments are time-invariant. Hence, for a weakly stationary process $x(k)$,

$$\begin{aligned} E\{x(k)\} &= \bar{x} \quad \forall k \\ E\{[x(k)-\bar{x}][x(k-\tau)-\bar{x}]^T\} &= R_x(\tau) \quad \forall k \end{aligned} \tag{5.5}$$

In other words, if $x(k)$ is stationary, it has a constant mean value and its covariance depends only on the time difference τ. For Gaussian processes, weakly stationary processes are also stationary.

For scalar $x(k)$, $R(0)$ can be interpreted as the variance of the signal and $\frac{R(\tau)}{R(0)}$ reveals its time correlation. The normalized covariance $\frac{R(\tau)}{R(0)}$ ranges from 0 to 1 and indicates the time correlation of the signal. The value of 1 indicates a complete correlation and the value of 0 indicates no correlation.

Note that many signals have both deterministic and stochastic components. In some applications, it is very useful to treat these signals in the same framework. One can do this by defining

$$\begin{aligned} \bar{x} &= \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} x(k) \\ R_x(\tau) &= \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} [x(k)-\bar{x}][x(k-\tau)-\bar{x}]^T \end{aligned} \tag{5.6}$$

Note that in the above, both deterministic and stochastic parts are averaged out. The signals for which the above limits converge are called "quasi-stationary" signals. The above definitions are consistent with the previous definitions since,in the purely stochastic case, a particular realization of a stationary stochastic process with given mean ($\bar{x}$) and covariance ($R_x(\tau)$) should satisfy the above relationships.

5.1.4 SPECTRA OF STATIONARY STOCHASTIC PROCESSES

Throughout this chapter, continuous time is rescaled so that each discrete time interval represents one continuous time unit. If the sample interval T_s

is not one continuous time unit, the frequency in discrete time needs to be scaled with the factor of $\frac{1}{T_s}$.

Spectral density of a stationary process $x(k)$ is defined as the Fourier transform of its covariance function:

$$\Phi_x(\omega) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} R_x(\tau) e^{-j\tau\omega} \tag{5.7}$$

Area under the curve represents the power of the signal for the particular frequency range. For example, the power of $x(k)$ in the frequency range (ω_1, ω_2) is calculated by the integral

$$2 \cdot \int_{\omega=\omega_1}^{\omega=\omega_2} \Phi_x(\omega) d\omega$$

Peaks in the signal spectrum indicate the presence of periodic components in the signal at the respective frequency.

The inverse Fourier transform can be used to calculate $R_x(\tau)$ from the spectrum $\Phi_x(\omega)$ as well

$$R_x(\tau) = \int_{-\pi}^{\pi} \Phi_x(\omega) e^{j\tau\omega} d\omega \tag{5.8}$$

With $\tau = 0$, the above becomes

$$E\{x(k)x(k)^T\} = R_x(0) = \int_{-\pi}^{\pi} \Phi_x(\omega) d\omega \tag{5.9}$$

which indicates that the total area under the spectral density is equal to the variance of the signal. This is known as the Parseval's relationship.

Example: Show plots of various covariances, spectra and realizations!

****Exercise:** Plot the spectra of (1) white noise, (2) sinusoids, and (3)white noise filtered through a low-pass filter.

5.1.5 DISCRETE-TIME WHITE NOISE

A particular type of a stochastic process called *white noise* will be used extensively throughout this book. $x(k)$ is called a white noise (or white sequence) if

$$\mathcal{P}(x(k)|x(\ell)) = \mathcal{P}(x(k)) \text{ for } \ell < k \tag{5.10}$$

for all k. In other words, the sequence has no time correlation and hence all the elements are mutually independent. In such a situation, knowing the realization of $x(\ell)$ in no way helps in estimating $x(k)$.

A *stationary* white noise sequence has the following properties:

$$\begin{aligned} E\{x(k)\} &= \bar{x} \qquad \forall k \\ E\{(x(k)-\bar{x})(x(k-\tau)-\bar{x})^T\} &= \begin{cases} R_x & \text{if } \tau = 0 \\ 0 & \text{if } \tau \neq 0 \end{cases} \end{aligned} \tag{5.11}$$

Hence, the covariance of a white noise is defined by a single matrix.

The spectrum of white noise $x(k)$ is constant for the entire frequency range since from (5.7)

$$\Phi_x(\omega) = \frac{1}{2\pi} R_x \tag{5.12}$$

The name "white noise" actually originated from its similarity with white light in spectral properties.

5.1.6 COLORED NOISE

A stochastic process generated by filtering white noise through a dynamic system is called "colored noise."

Important:

> A stationary stochastic process with any given mean and covariance function can be generated by passing a white noise through an appropriate dynamical system.

To see this, consider

$$d(k) = H(q)\varepsilon(k) + \bar{d} \tag{5.13}$$

where $\varepsilon(k)$ is a white noise of identity covariance and $H(q)$ is a stable / stably invertible transfer function (matrix). Using simple algebra (Ljung -REFERENCE), one can show that

$$\Phi_d(\omega) = H(e^{j\omega})H^T(e^{-j\omega}) \tag{5.14}$$

The spectral factorization theorem (REFERENCE - Åström and Wittenmark, 1984) says that one can always find $H(q)$ that satisfies (5.14) for an arbitrary Φ_d and has no pole or zero outside the unit disk. In other words, the first and second order moments of any stationary signal can be matched by the above model.

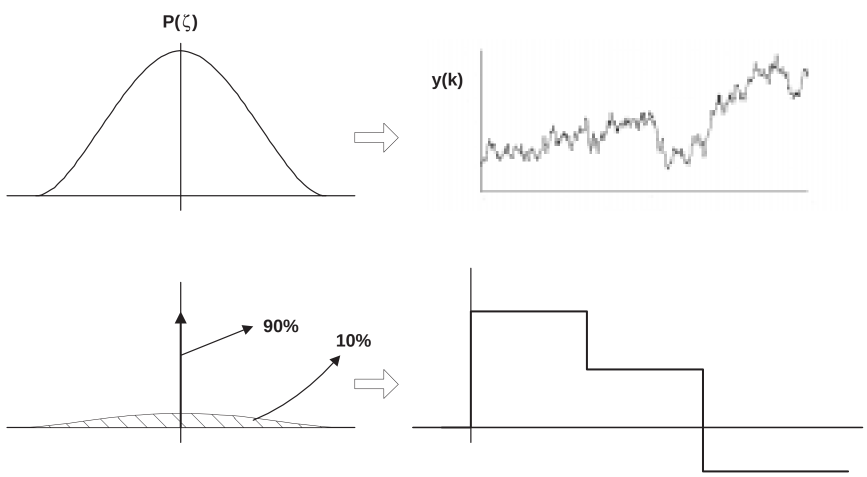
This result is very useful in modeling disturbances whose covariance functions are known or fixed. Note that a stationary Gaussian process is completely specified by its mean and covariance. Such a process can be modelled by filtering a zero-mean Gaussian white sequence through appropriate dynamics determined by its spectrum (plus adding a bias at the output if the mean is not zero).

5.1.7 INTEGRATED WHITE NOISE AND NONSTATIONARY PROCESSES

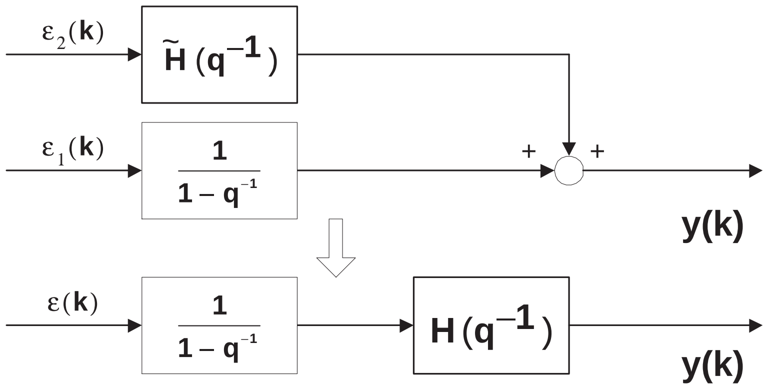
Some processes exhibit mean-shifts (whose magnitude and occurence are random). Consider the following model:

$$y(k) = y(k-1) + \varepsilon(k)$$

where $\varepsilon(k)$ is a white sequence. Such a sequence is called *integrated white noise* or sometimes random walk. Particular realizations under different distribution of $\varepsilon(k)$ are shown below:

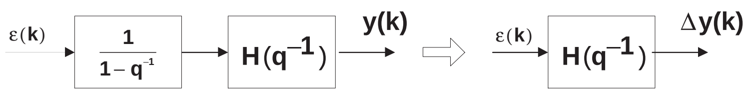


More generally, many interesting signals will exhibit stationary behavior combined with randomly occuring mean-shifts. Such signals can be modeled as



As shown above, the combined effects can be expressed as an integrated white noise colored with a filter $H(q^{-1})$.

Note that while $y(k)$ is nonstationary, the differenced signal $\Delta y(k)$ is stationary.



5.1.8 STOCHASTIC DIFFERENCE EQUATION

Generally, a stochastic process can be modeled through the following stochastic difference equation.

$$\begin{aligned} x(k+1) &= Ax(k) + B\varepsilon(k) \\ y(k) &= Cx(k) + D\varepsilon(k) \end{aligned} \tag{5.15}$$

where $\varepsilon(k)$ is a white vector sequence of zero mean and covariance R_ε.

Note that

$$\begin{aligned} E\{x(k)\} &= AE\{x(k-1)\} = A^k E\{x(0)\} \\ E\{x(k)x^T(k)\} &= AE\{x(k-1)x^T(k-1)\}A^T + BR_\varepsilon B^T \end{aligned} \tag{5.16}$$

If all the eigenvalues of A are strictly inside the unit disk, the above approaches a stationary process as $k \to \infty$ since

$$\begin{aligned} \lim_{k\to\infty} E\{x(k)\} &= 0 \\ \lim_{k\to\infty} E\{x(k)x^T(k)\} &= R_x \end{aligned} \tag{5.17}$$

where R_x is a solution to the Lyapunov equation

$$R_x = AR_xA^T + BR_\varepsilon B^T \tag{5.18}$$

Since $y(k) = Cx(k) + D\varepsilon(k)$,

$$\begin{aligned} E\{y(k)\} &= CE\{x(k)\} + DE\{\varepsilon(k)\} = 0 \\ E\{y(k)y^T(k)\} &= CE\{x(k)x^T(k)\}C^T + DE\{\varepsilon(k)\varepsilon^T(k)\}D^T = CR_xC^T + DR_\varepsilon D^T \end{aligned} \tag{5.19}$$

The auto-correlation function of $y(k)$ becomes

$$R_y(\tau) \triangleq E\{y(k+\tau)y^T(k)\} = \begin{cases} CR_xC^T + DR_\varepsilon D^T \text{ for } \tau = 0 \\ CA^\tau R_xC^T + CA^{\tau-1}BR_\varepsilon D^T \text{ for } \tau > 0 \end{cases} \tag{5.20}$$

The spectrum of w is obtained by taking the Fourier transform of $R_y(\tau)$) and can be shown to be

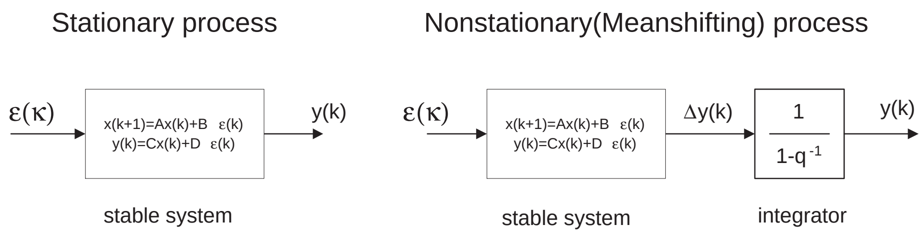
$$\Phi_y(\omega) = \left(C(e^{j\omega}I - A)^{-1}B + D\right) R_\varepsilon \left(C(e^{j\omega}I - A)^{-1}B + D\right)^T \tag{5.21}$$

In the case that A contains eigenvalues on or outside the unit circle, the process is nonstationary as its covariance keeps increasing (see Eqn. (5.16). However, it is common to include integrators in A to model *mean-shifting* (*random-walk*-like) behavior. If all the outputs exhibit this behavior, one can use

$$\begin{aligned} x(k+1) &= Ax(k) + B\varepsilon(k) \\ \Delta y(k) &= Cx(k) + D\varepsilon(k) \end{aligned} \tag{5.22}$$

Note that, with a stable A, while $\Delta y(k)$ is a stationary process, $y(k)$

includes an integrator and therefore is nonstationary.



5.2 STOCHASTIC SYSTEM MODELS

Models used for control will often include both deterministic and stochastic inputs. The deterministic inputs correspond to known signals like manipulated variables. The stochastic signals cover whatever remaining parts that cannot be predicted *a priori*. They include the effect of disturbances, other process variations and instrumentation errors.

5.2.1 STATE-SPACE MODEL

The following stochastic difference equation may be used to characterize a stochastic disturbance:

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) + \varepsilon_1(k) \\ y(k) &= Cx(k) + \varepsilon_2(k) \end{aligned} \tag{5.23}$$

$\varepsilon_1(k)$ and $\varepsilon_2(k)$ are white noise sequences that represent the effects of disturbances, measurement error, etc. They may or may not be correlated.

- If the above model is derived from fundamental principles, $\varepsilon(k)$ may be a signal used to generate physical disturbance states (which are included in the state x) or merely artifical signals added to represent random errors in the state equation. $\varepsilon_2(k)$ may be measurement noise or signals representing errors in the output equations.

- If the model is derived on an empirical basis, $\varepsilon_1(k)$ and $\varepsilon_2(k)$ together represent the combined effects of the process / measurement randomness. In other words, the output is viewed as a composite of two signals ($y(k) = y_d(k) + y_s(k)$), one of which is the output of the deterministic system

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) \\ y_d(k) &= Cx(k) \end{aligned} \tag{5.24}$$

and the other is the random component

$$\begin{aligned} x(k+1) &= Ax(k) + \varepsilon_1(k) \\ y_s(k) &= Cx(k) + \varepsilon_2(k) \end{aligned} \tag{5.25}$$

With such a model available, one problem treated in statistics is to predict future states, $(x(k+i), i \geq 0)$ given collected output measurements $(y(k), \cdots, y(1))$. This is called state estimation and will be discussed in the next chapter.

The other problem is building such a model. Given data $(y(i), u(i), i = 1, \cdots, N)$, the following two methods are available.

- One can use the so called subspace identification methods.
- One can build a time series model or more generally a transfer function model of the form

$$y(k) = G(q^{-1})u(k) + H(q^{-1})\varepsilon(k)$$

Then, one can perform a state-space realization of the above to obtain the state-space model.

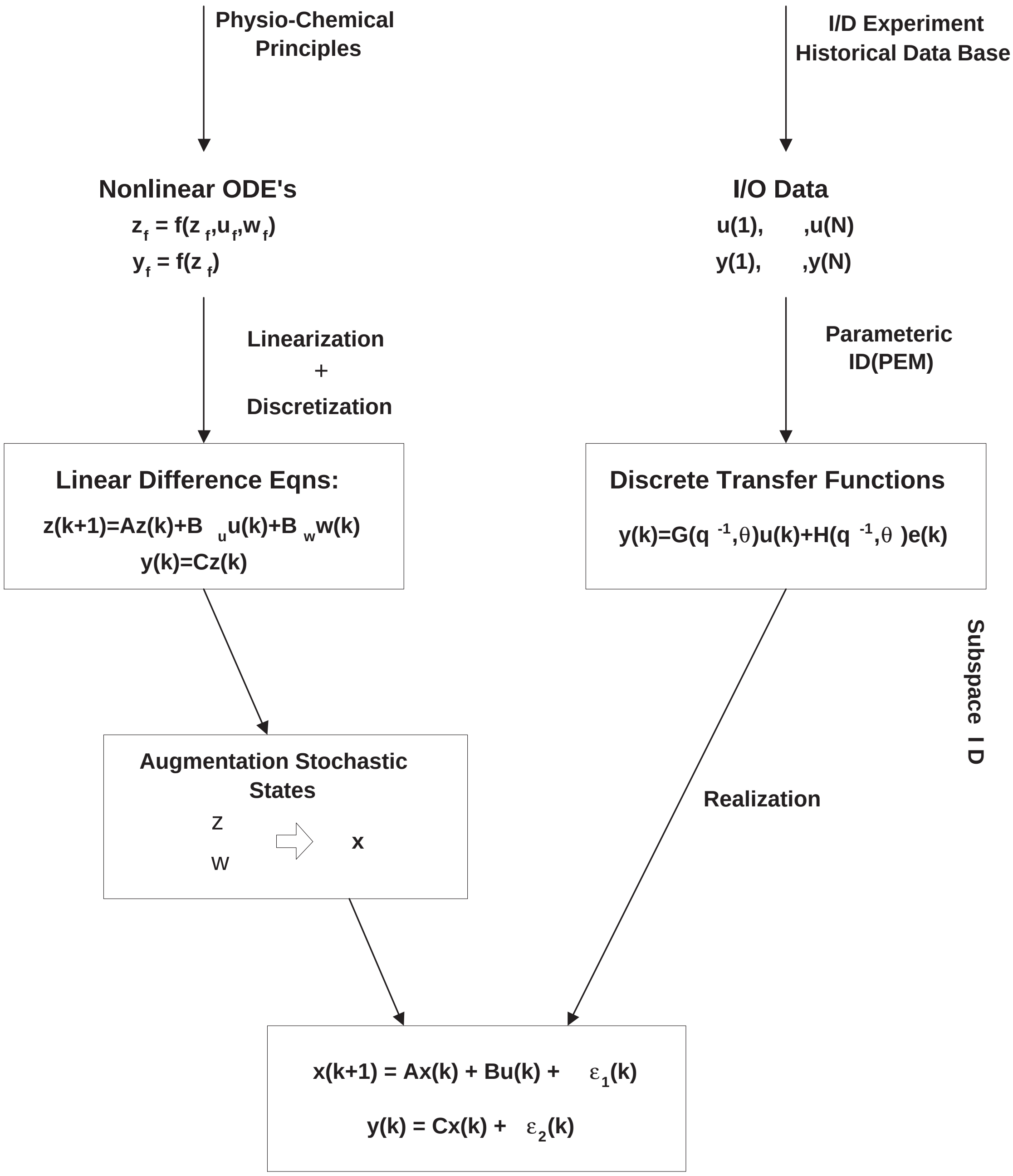
Details of these methods will be discussed in the identification chapter.

For systems with (mean-shifting) nonstationary disturbances in all output channels, it may be more convenient to express the model in terms of the differenced inputs and outputs:

$$\begin{aligned} x(k+1) &= Ax(k) + B\Delta u(k) + \varepsilon_1(k) \\ \Delta y(k) &= Cx(k) + \varepsilon_2(k) \end{aligned} \tag{5.26}$$

If the undifferenced y is desired as the output of the system, one can simply rewrite the above as

$$\begin{aligned} \begin{bmatrix} x(k+1) \\ y(k+1) \end{bmatrix} &= \begin{bmatrix} A & 0 \\ CA & I \end{bmatrix} \begin{bmatrix} x(k) \\ y(k) \end{bmatrix} + \begin{bmatrix} B \\ CB \end{bmatrix} \Delta u(k) + \begin{bmatrix} I \\ C \end{bmatrix} \varepsilon_1(k) + \begin{bmatrix} 0 \\ I \end{bmatrix} \varepsilon_2(k) \\ y(k) &= \begin{bmatrix} 0 & I \end{bmatrix} \begin{bmatrix} x(k) \\ y(k) \end{bmatrix} \end{aligned} \tag{5.27}$$



5.2.2 INPUT-OUTPUT MODELS

One can also use input-output models. A general form is

$$y(k) = G(q)u(k) + H(q)\varepsilon(k) \tag{5.28}$$

Within the above general structure, different parameterizations exist. For instance, a popular model is the following ARMAX (AR for Auto-Regressive, MA for Moving-Average and X for eXtra input) process:

$$\begin{aligned} y(k) &= (I + A_1 q^{-1} + \cdots + A_n q^{-n})^{-1}(B_1 q^{-1} + \cdots + B_m q^{-m})u(k) \\ &\quad + (I + A_1 q^{-1} + \cdots + A_n q^{-n})^{-1}(I + C_1 q^{-1} + \cdots + C_n q^{-n})\varepsilon(k) \end{aligned} \tag{5.29}$$

Note that the above is equivalent to the following linear time-series equation:

$$\begin{aligned} y(k) &= -A_1 y(k-1) - A_2 y(k-2) - \cdots - A_n y(k-n) \\ &\quad + B_1 u(k-1) + \cdots + B_m u(k-m) \\ &\quad + \varepsilon(k) + C_1 \varepsilon(k-1) + \cdots + C_n \varepsilon(k-n) \end{aligned} \tag{5.30}$$

In most practical applications, matrices A_i's and C_i's are restricted to be diagonal, which results in a MISO (rather than a MIMO) structure. In such a case, stochastic components for different output channels are restricted to be *mutually independent*.

For systems with integrating type disturbances in all output channels, a more appropriate model form is

$$y(k) = G(q)u(k) + \frac{1}{1 - q^{-1}} H(q)\varepsilon(k) \tag{5.31}$$

The above can be easily rewritten as

$$\Delta y(k) = G(q)\Delta u(k) + H(q)\varepsilon(k) \tag{5.32}$$