Part III

BACKGROUND FOR ADVANCED ISSUES

Contents

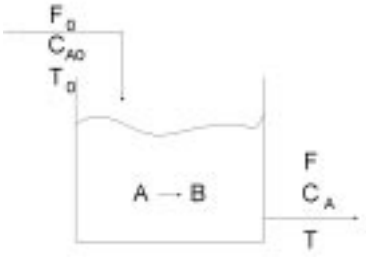
Chapter 1

BASICS OF LINEAR ALGEBRA

1.1 VECTORS

Definition of Vector

Consider a CSTR where a simple exothermic reaction occurs:



A neat way to represent process variables, F, C_A, T, is to stack them in a column.

$$\begin{bmatrix} F \\ C_A \\ T \end{bmatrix}$$

Definition of Vector (Continued)

In general, n tuples of numbers stacked in a column is called vector.

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Transpose of a Vector x:

$$x^T = [x_1 \; x_2 \; \cdots \; x_n]$$

Basic Operations of Vectors

a: a scalar, x, y: vectors

Addition:

$$x + y = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}$$

Scalar Multiplication:

$$ax = a \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} ax_1 \\ ax_2 \\ \vdots \\ ax_n \end{bmatrix}$$

Vector Norms

Norm is the measure of vector size.

p norms:

$$\|x\|_p = (|x_1|^p + \cdots + |x_n|^p)^{\frac{1}{p}} \qquad 1 \leq p < \infty$$

$$\|x\|_\infty = \max_i |x_i|$$

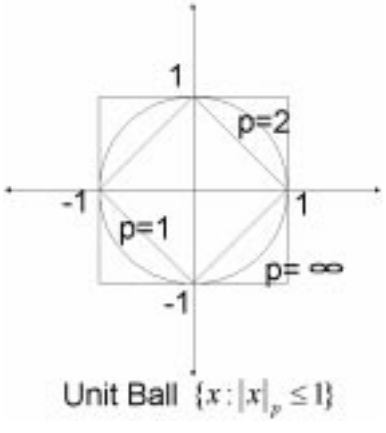
Example:

$$\|x\|_1 = |x_1| + \cdots + |x_n|$$

$$\|x\|_2 = \sqrt{|x_1|^2 + \cdots + |x_n|^2}$$

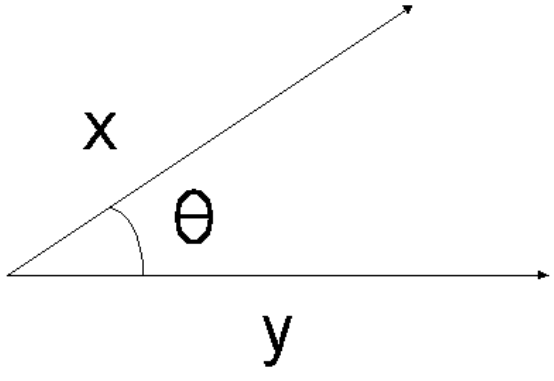
$$\|x\|_\infty = \max\{|x_1|, \cdots, |x_n|\}$$

$\|x\|_2$ coincides with the length in Euclidean sense and, thus, is called Eclidean norm. Throughout the lecture, $\|\cdot\|$ denotes $\|\cdot\|_2$.



Inner Product

Inner Product:



$$x \cdot y = x^T y = \|x\|\|y\| \cos\theta$$

$$\Downarrow$$

$$x \cdot y \begin{cases} > 0 & \text{if } \theta \text{ is acute} \\ = 0 & \text{if } \theta \text{ is right} \\ < 0 & \text{if } \theta \text{ is obtuse} \end{cases}$$

Note that two vectors x, y are orthogonal if $x^T y = 0$

Linear Independence and Basis

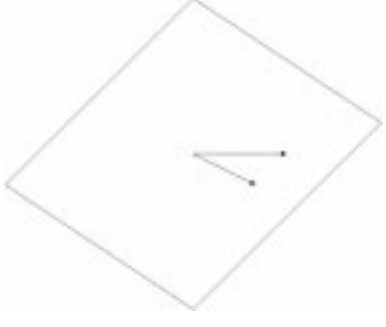
$a_1, \cdots, a_m$: scalars, $x_1, \cdots, x_m$: vectors

Linear Combination:

$$a_1 x_1 + a_2 x_2 + \cdots + a_m x_m$$

Span: Span of $x_1, \cdots, x_m$ is the set of all linear combination of them, which is a plane in $\mathbf{R}^n$.

$$span\{x_1, x_2, \cdots, x_m\} = \{x = a_1 x_1 + a_2 x_2 + \cdots + a_m x_m\}$$



Linear Independence: $\{x_1, \cdots, x_m\}$ is called linearly independent if no one of them is in the span of others.

Basis of a Space (S): A set of linearly independent vectors $\{x_1, x_2, \cdots, x_m\}$ such that $S = span\{x_1, x_2, \cdots, x_m\}$

1.2 MATRICES

Definition of Matrices

Let A be the linear mapping from a vector x to another vector y. Then A is represented by a rectangular array of numbers:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

that is called $m \times n$-matrix.

Transpose of a Matrix A:

$$A^T = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \vdots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}$$

Conjugate Transpose of a Matrix A:

$$A^* = \begin{bmatrix} \bar{a}_{11} & \bar{a}_{21} & \cdots & \bar{a}_{m1} \\ \bar{a}_{12} & \bar{a}_{22} & \cdots & \bar{a}_{m2} \\ \vdots & \vdots & \vdots & \vdots \\ \bar{a}_{1n} & \bar{a}_{2n} & \cdots & \bar{a}_{mn} \end{bmatrix}$$

Notice that $A^T = A^*$ for real matrices.

Basic Operation of Matrices

a: a scalar, A, B: matrices

Addition:

$$A + B = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2n} + b_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \cdots & a_{mn} + b_{mn} \end{bmatrix}$$

Scalar Multiplication:

$$aA = a \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} aa_{11} & aa_{12} & \cdots & aa_{1n} \\ aa_{21} & aa_{22} & \cdots & aa_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ aa_{m1} & aa_{m2} & \cdots & aa_{mn} \end{bmatrix}$$

Basic Operation of Matrices (Continued)

Matrix Multiplication:

$$AB = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1l} \\ b_{21} & b_{22} & \cdots & b_{2l} \\ \vdots & \vdots & \vdots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nl} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{n} a_{1i}b_{i1} & \sum_{i=1}^{n} a_{1i}b_{i2} & \cdots & \sum_{i=1}^{n} a_{1i}b_{il} \\ \sum_{i=1}^{n} a_{2i}b_{i1} & \sum_{i=1}^{n} a_{2i}b_{i2} & \cdots & \sum_{i=1}^{n} a_{2i}b_{il} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{i=1}^{n} a_{mi}b_{i1} & \sum_{i=1}^{n} a_{mi}b_{i2} & \cdots & \sum_{i=1}^{n} a_{mi}b_{il} \end{bmatrix}$$

Inverse of Square Matrices

Inverse of an $n \times n$ matrix A is an $n \times n$ matrix such that

$$AA^{-1} = A^{-1}A = I$$

Theorem: An $n \times n$ matrix A has its inverse iff the columns of A are linearly independent.

Suppose A defines a linear transformation:

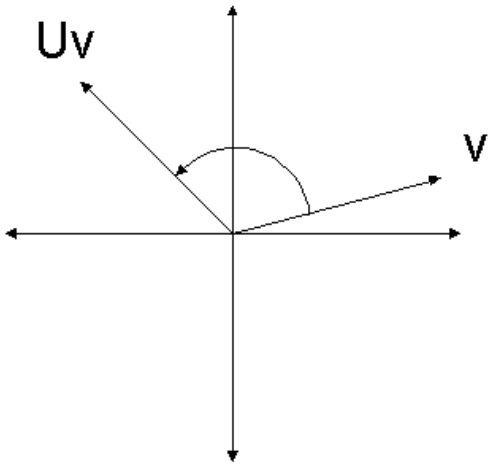
$$y = Ax$$

Then the inverse of A defines the inverse transformation:

$$x = A^{-1}y$$

Unitary Matrices

Matrix that rotates a vector without change of size is called unitary matrix

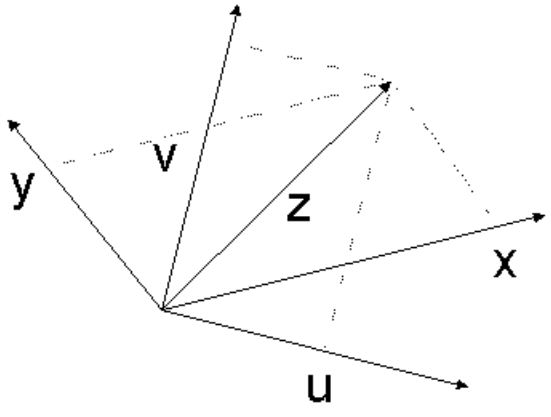


U: unitary matrix

Properties of Unitary Matrix:

$$U^*U = I = UU^*$$

Coordinate Transformation



z: a vector, $\{u, v\}$, $\{x, y\}$: coordinate systems

$$z = \bar{w}_1 u + \bar{w}_2 v = w_1 x + w_2 y \quad \Rightarrow \quad [u\ v]\bar{w} = [x\ y]w$$

$$\Downarrow$$

$$\bar{w} = Tw, \quad T = [u\ v]^{-1}[x\ y]$$

In general, the representations of a vector in two different coordinate systems are related by an invertible matrix T:

$$\bar{w} = Tw$$

$$w = T^{-1}\bar{w}$$

Coordinate Transformation (Continued)

Representations of matrix in different coordinates:

Suppose $\alpha = T\bar{\alpha}$ and $\beta = T\bar{\beta}$. Then

$$\alpha = A\beta \quad \Rightarrow \quad T\bar{\alpha} = AT\bar{\beta} \quad \Rightarrow \quad \bar{\alpha} = T^{-1}AT\bar{\beta}$$

$$\Downarrow$$

$$\bar{\alpha} = \bar{A}\bar{\beta}$$

where

$$\bar{A} = T^{-1}AT$$

that is called the similarity transformation of A.

Eigenvalues and Eigenvectors

The eigenvalues of $n \times n$ matrix A are n roots of $det(\lambda I - A)$.

If λ is an eigenvalue of A, $\exists$ nonzero v such that

$$Av = \lambda v$$

where v is called eigenvector.

Eigenvalue Decomposition

Let $A \in \mathbf{R}^{n \times n}$. Suppose λ_i be eigenvalues of A such that

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$$

Let

$$T = [v_1, v_2, \cdots, v_n] \in \mathbf{R}^{n \times n}$$

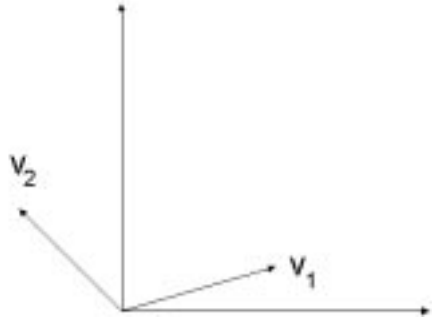
where v_i denotes eigenvector of A associated with λ_i. If A has n linearly independent eigenvectors,

$$A = T \Lambda T^{-1}$$

where

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix}$$

Notice that Λ is simply the representation of A in the coordinate system consists of eigenvectors.



Symmetric Matrices

A matrix A is called symmetric if

$$A = A^T$$

Symmetric matrix is useful when we consider a quadratic form.

Indeed, given a matrix A,

$$x^T A x = x^T S x$$

where S is the symmetric matrix defined by

$$S = \frac{1}{2}(A + A^T)$$

Positive Definiteness: A symmetric matrix A is positive definite if

$$x^T A x > 0 \quad \forall x \neq 0,\ x \in \mathbf{C}^n$$

Positive Semi-Definiteness: A symmetric matrix A is positive semi-definite if

$$x^T A x \geq 0$$

Theorem: A symmetric matrix A is positive definite iff all the eigenvalues of A are positive.

Matrix Norms

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \in \mathbf{C}^{m \times n}$$

p norms:

$$|||A|||_p = (\sum_{i,j} |a_{i,j}|^p)^{\frac{1}{p}} \qquad 1 \leq p < \infty$$

$$|||A|||_\infty = \max_{i,j} |a_{i,j}|$$

$|||\cdot|||_2$ is called Euclidean or Frobenius norm.

What is the difference between $\mathbf{C}^{m \times n}$ and $\mathbf{C}^{mn}$?

A matrix in $\mathbf{C}^{m \times n}$ defines a linear operator from $\mathbf{C}^n$ to $\mathbf{C}^m$;
$y = Ax$.

Matrix Norms (Continued)

Induced (or operator) p norms:

$$\|A\|_p = \sup_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p} = \max_{\|x\|=1} \|Ax\|_p \qquad 1 \leq p \leq \infty$$

$$\Downarrow$$

$$\|y\|_p = \|Ax\|_p \leq \|A\|_p \|x\|_p \qquad \forall x \in \mathbf{C}^n$$

Examples:

$p = 1$:

$$\|A\|_1 = \max_j \sum_{i=1}^{m} |a_{i,j}|$$

$p = 2$: spectral norm

$$\|A\|_2 = [\lambda_{max}(A^T A)]^{\frac{1}{2}}$$

$p = \infty$:

$$\|A\|_\infty = \max_i \sum_{j=1}^{m} |a_{i,j}|$$

1.3 SINGULAR VALUE DECOMPOSITION

Singular Values and Singular Vectors

Singular values of an $m \times n$ matrix A are the square roots of $\min\{m, n\}$ eigenvalues of A^*A.

$$\sigma(A) = \sqrt{\lambda(A^*A)}$$

Right singular vectors of a matrix A are the eigenvectors of A^*A.

$$\sigma(A)^2 v - A^*Av = 0$$

Left singular vectors of a matrix A are the eigenvectors of AA^*.

$$\sigma(A)^2 u - AA^*u = 0$$

$$\bar{\sigma}(A) = \text{ the largest singular value of } A = \max_{\|x\|=1} \|Ax\| = \|A\|_2$$

The largest possible size change of a vector by A.

$$\underline{\sigma}(A) = \text{ the smallest singular value of } A = \min_{\|x\|=1} \|Ax\|$$

The smallest possible size change of a vector by A.

Singular Values and Singular Vectors (Continued)

Condition number: $c(A) = \frac{\bar{\sigma}(A)}{\underline{\sigma}(A)}$

$$A\bar{v} = \bar{\sigma}\bar{u}$$

$$A\underline{v} = \underline{\sigma}\ \underline{u}$$

$$\Downarrow$$

$\bar{v}$ ($\underline{v}$): highest (lowest) gain input direction

$\bar{u}$ ($\underline{u}$): highest (lowest) gain observing direction

Singular Value Decomposition

Let $A \in \mathbf{R}^{m \times n}$. Suppose σ_i be singular values of A such that

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0, \quad p = \min\{m, n\}$$

Let

$$U = [u_1, u_2, \cdots, u_m] \in \mathbf{R}^{m \times m} \qquad V = [v_1, v_2, \cdots, v_n] \in \mathbf{R}^{n \times n}$$

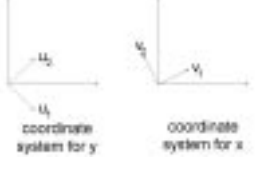
where u_i, v_j denote left and right orthonormal singular vectors of A, respectively. Then

$$A = U\Sigma V^*, \quad \Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} = \sum_{i=1}^{p} \sigma_i(A) u_i v_i^*$$

where

$$\Sigma_1 = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \sigma_p \end{bmatrix}$$

Consider $y = Ax$. Then Σ is simply the representation of A when x and y are represented in the coordinate systems consisting of right and left singular vectors, respectively.



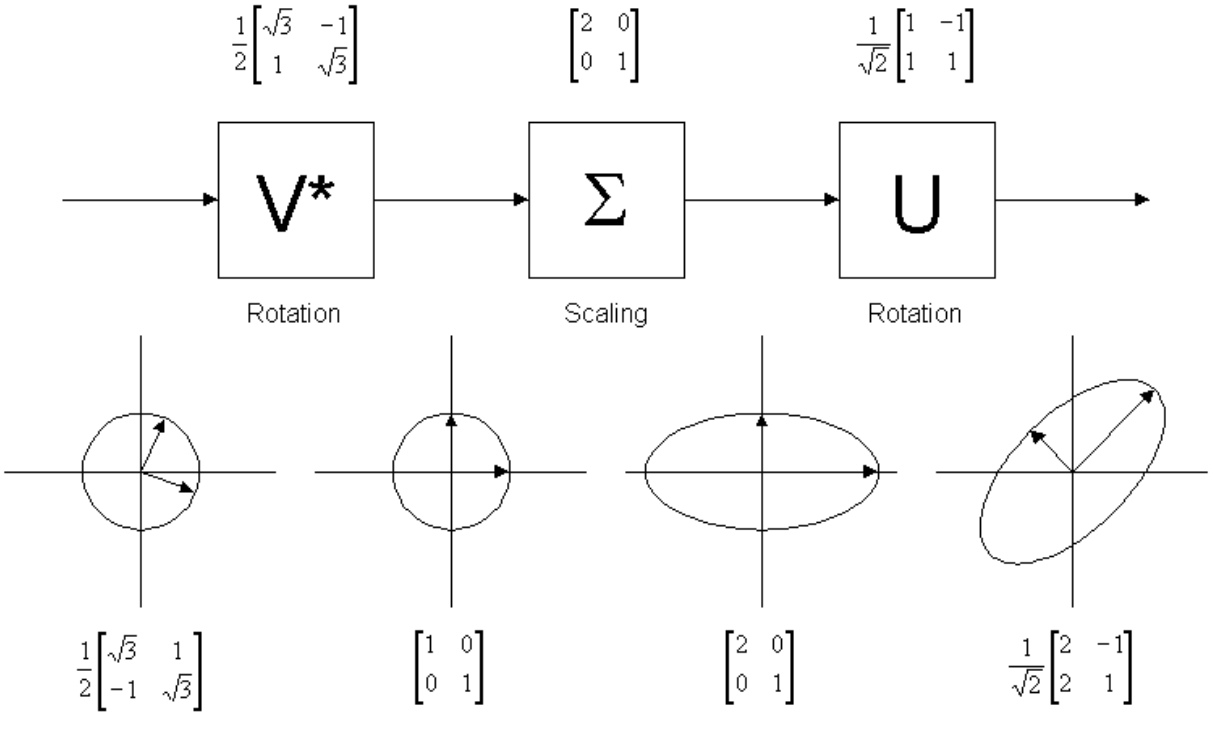
Singular Value Decomposition (Continued)

Example:

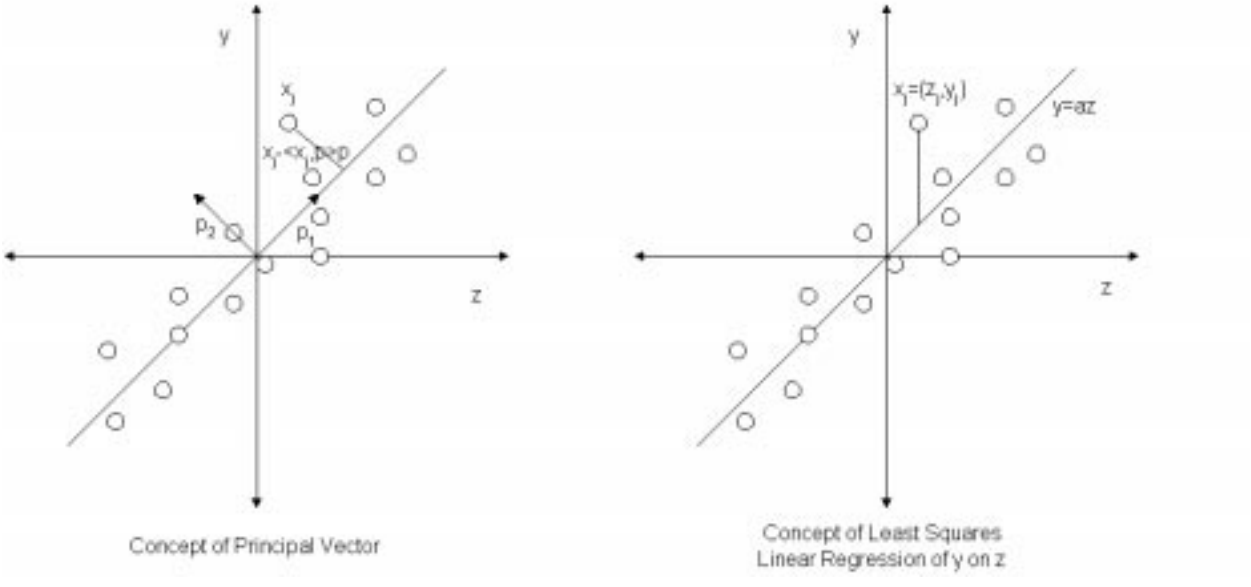
$$A = \begin{bmatrix} 0.8712 & -1.3195 \\ 1.5783 & -0.0947 \end{bmatrix}$$

$$\Downarrow$$

$$U = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}, \quad V = \frac{1}{2} \begin{bmatrix} \sqrt{3} & 1 \\ -1 & \sqrt{3} \end{bmatrix}$$



Principal Component Analysis



Given N n-dimensional vectors $\{x_1, x_2, \cdots, x_N\}$, the principal vector p is

$$p = \arg\min_{\|p\|=1} \sum_{i=1}^{N} \|x_i - \langle x_i, p\rangle p\|^2$$

$$= \arg\min_{\|p\|=1} \sum_{i=1}^{N} \left[\langle x_i, x_i\rangle - 2\langle x_i, p\rangle^2 + \langle x_i, p\rangle^2 \langle p, p\rangle\right]$$

$$= \arg\min_{\|p\|=1} \sum_{i=1}^{N} -\frac{\langle x_i, p\rangle^2}{\langle p, p\rangle} = \arg\max \sum_{i=1}^{N} \frac{\langle x_i, p\rangle^2}{\langle p, p\rangle} = \arg\max \alpha(p)$$

where

$$\alpha(p) = \sum_{i=1}^{N} \frac{x_i^T p p^T x_i}{p^T p}$$

Principal Component Analysis (Continued)

At the extremum,

$$0 = \frac{1}{2}\frac{d\alpha}{dp} = \sum_{i=1}^{N} \frac{x_i x_i^T p}{p^T p} - \sum_{i=1}^{N} \frac{x_i^T p p^T x_i p}{(p^T p)^2}$$

$$\Downarrow$$

$0 = \sum_{i=1}^{N} x_i x_i^T p - \sum_{i=1}^{N} \frac{x_i^T p p^T x_i}{p^T p} p = XX^T p - \lambda p$ Singular Value Problem for X

where

$$X = [x_1\ x_2\ \cdots\ x_N], \quad \lambda = \sum_{i=1}^{N} \frac{x_i^T p p^T x_i}{(p^T p)^2}$$

The SVD of X is

$$X = P\Lambda^{\frac{1}{2}} V^T = p_1 \lambda_1^{\frac{1}{2}} u_1^T + \cdots + p_n \lambda_n^{\frac{1}{2}} u_n^T$$

where

$$P = [p_1\ p_2\ \cdots\ p_n], \quad V = [v_1\ v_2\ \cdots\ v_N],$$
$$\Lambda = [diag[\lambda_i^{\frac{1}{2}}]\ 0] \quad 0 = X^T X v - \lambda v$$
$$\lambda_1^{\frac{1}{2}} \geq \cdots \geq \lambda_n^{\frac{1}{2}}$$

The approximation of X using first m significant principal vectors:

$$X \approx \bar{X} = \bar{P}\bar{\Lambda}^{\frac{1}{2}}\bar{U}^T = p_1 \lambda_1^{\frac{1}{2}} u_1^T + \cdots + p_m \lambda_m^{\frac{1}{2}} u_m^T$$

where

$$\bar{P} = [p_1\ p_2\ \cdots\ p_m], \quad \Lambda = diag[\lambda_i^{\frac{1}{2}}]_{i=1}^{m} \quad \bar{V} = [v_1\ v_2\ \cdots\ v_m]$$

Principal Component Analysis (Continued)

$$p_i^T X = p_i^T (p_1 \lambda_1^{\frac{1}{2}} u_1^T + \cdots + p_n \lambda_n^{\frac{1}{2}} u_n^T) = \lambda_i^{\frac{1}{2}} u_i^T$$

$$\Downarrow$$

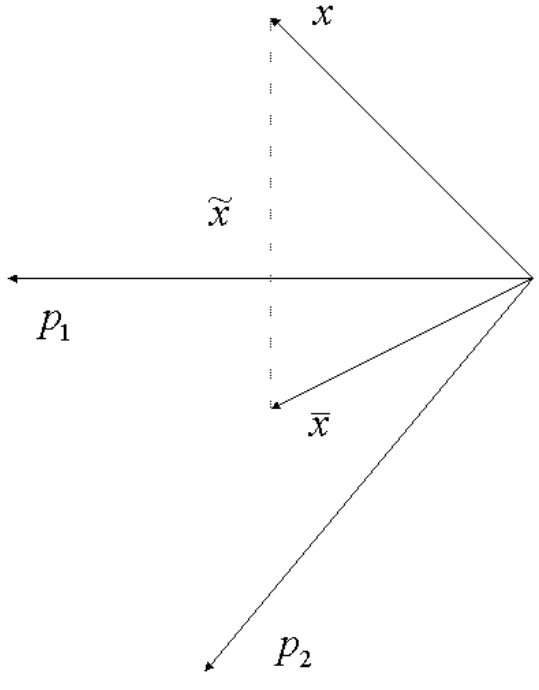
$$\bar{P}^T X = \bar{U}^T$$

$$\Downarrow$$

$$\bar{X} = \bar{P}\bar{U}^T = \bar{P}\bar{P}^T X$$

and the residual is

$$\tilde{X} = X - \bar{X} = (I - \bar{P}\bar{P}^T)X$$



Chapter 2

BASICS OF LINEAR SYSTEMS

2.1 STATE SPACE DESCRIPTION

State Space Model Development

Consider fundamental ODE model:

$$\begin{aligned} \frac{dx_f}{dt} &= f(x_f, u_f) \\ y_f &= g(x_f) \end{aligned}$$

x_f: state vector,

u_f: input vector

y_f: output vector

State Space Model Development (Continued)

$\Downarrow$ linearization w.r.t. an equilibrium $(\bar{x}, \bar{u})$

$$\begin{aligned} \frac{dx}{dt} &= \left(\frac{df}{\partial x_f}\right)_{ss} x + \left(\frac{\partial f}{\partial u_f}\right)_{ss} u \\ y &= \left(\frac{\partial g}{\partial x_f}\right)_{ss} x \end{aligned}$$

where $x = x_f - \bar{x}$, $u = u_f - \bar{u}$.

$\Downarrow$ discretization

$$\begin{aligned} x(k+1) &= Ax(k) + B_u u(k) \\ y(k) &= Cx(k) \end{aligned}$$

State Space Description of Linear Systems

Consider the linear system described by the state equation:

$$x(k+1) = Ax(k) + Bu(k)$$

$$y(k) = Cx(k)$$

Take z-Transformation

$$zX(z) = AX(z) + BU(z)$$

$$Y(z) = CX(z)$$

$$\Downarrow$$

$$Y(z) = C(zI - A)^{-1}BU(z)$$

Solution to Linear System:

$$x(k) = A^n x(0) + \sum_{i=0}^{n-1} A^{n-i-1} Bu(i)$$

Transfer Function

Consider the system described by transfer function:

$$\frac{Y(z)}{U(z)} = \frac{b_1 z^{n-1} + b_2 z^{n-2} + \cdots + b_n}{z^n + a_1 z^{n-1} + \cdots + a_n}$$

Then a state space description of the system is

$$x(k+1) = Ax(k) + Bu(k)$$

$$y(k) = Cx(k)$$

where

$$A = \begin{bmatrix} -a_1 & -a_2 & \cdots & -a_{n-1} & -a_n \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \quad B = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}$$

$$C = [b_1 \; b_2 \; \cdots \; b_{n-1} \; b_n]$$

Transfer Function (Continued)

Example: Consider the transfer function:

$$\frac{b_1 z + b_2}{z^2 + a_1 z + a_2}$$

Then

$$A = \begin{bmatrix} -a_1 & -a_2 \\ 1 & 0 \end{bmatrix} \quad B = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$C = [0\ 1]$$

Then

$$\frac{Y(z)}{U(z)} = C(zI - A)^{-1}B = [b_1\ b_2] \begin{bmatrix} z + a_1 & +a_2 \\ -1 & z \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$= [b_1\ b_2] \frac{\begin{bmatrix} z & -a_2 \\ 1 & z + a_1 \end{bmatrix}^{-1}}{z^2 + a_1 z + a_2} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{b_1 z + b_2}{z^2 + a_1 z + a_2}$$

Nonuniqueness of State Space Representation

Consider a transfer function $G(z)$. Suppose the state space description of $G(z)$ is

$$x(k+1) = Ax(k) + Bu(k)$$

$$y(k) = Cx(k)$$

Consider a different coordinate system for the state space defined by

$$w(k) = T^{-1}x(k)$$

$$\Downarrow$$

$$w(k+1) = T^{-1}ATw(k) + T^{-1}Bu(k)$$

$$y(k) = CTw(k)$$

Then the transfer function of this system is

$$\frac{Y(z)}{U(z)} = CT(zI - T^{-1}AT)^{-1}T^{-1}B = CT[T^{-1}(zI - A)T]^{-1}T^{-1}B$$

$$= CTT^{-1}(zI - A)^{-1}TT^{-1}B = C(zI - A)^{-1}B = G(z)$$

There exist a multitude of state space representations of a system because there is a multiple infinity coordinate systems of the state space.

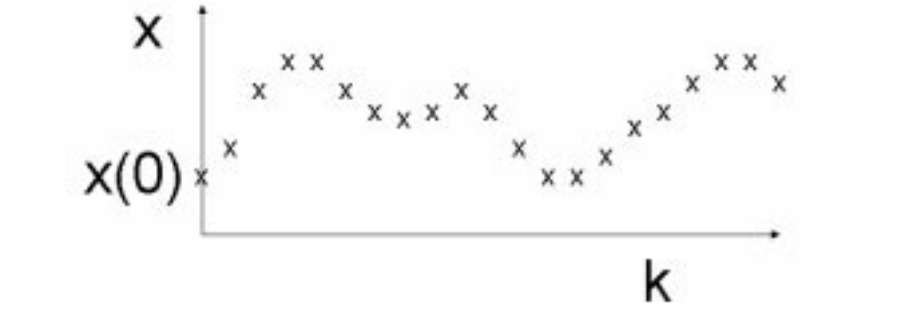
Definition of States

Given a time instant k, the state of the system is the minimal information that are necessary to calculate the future response.

For difference equations, the concept of the state is the same as that of the initial condition.

$$\Downarrow$$

$$\text{State} = x(k)$$



Stability of Linear Systems

A state x is stable if

$$\lim_{n\to\infty} A^n x = 0$$

A linear system

$$x(k+1) = Ax(k) + Bu(k)$$

$$y(k) = Cx(k)$$

is said to be stable if, for all $x \in \mathbf{R}^n$,

$$\lim_{n\to\infty} A^n x = 0$$

$$\Updownarrow$$

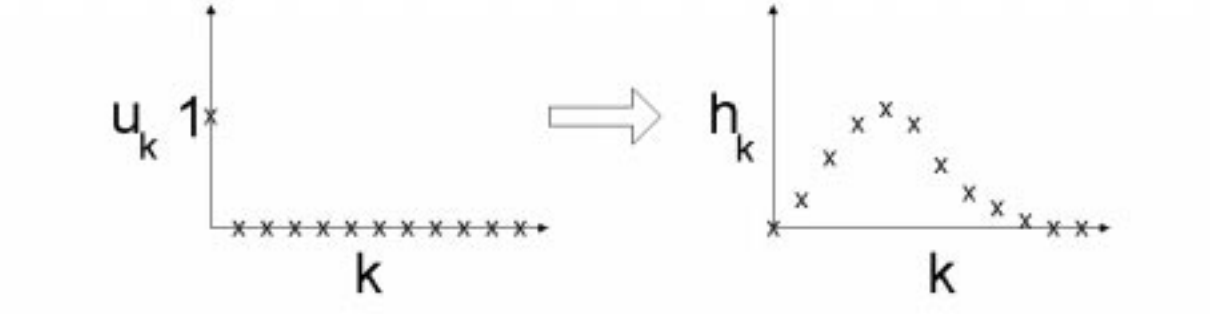
$$\max_i |\lambda_i(A)| < 1$$

2.2 FINITE IMPULSE RESPONSE MODEL

Impulse Responses of Linear Systems

$$y(k) = CA^k x(0) + \sum_{i=0}^{k-1} CA^{k-i-1} Bu(i)$$

Impulse Response Sequence $\{h(k)\}$: $\{y(k)\}$ when $x(0) = 0$ and $u(i) = \begin{cases} 1 & \text{if } i = 0 \\ 0 & \text{if } i \neq 0 \end{cases}$.



$$\{h(i) = CA^i B\}_{i=0}^{\infty}$$

$$\Downarrow$$

$$y(k) = h(k)x(0) + \sum_{i=0}^{k-1} h(k-i-1)u(i)$$

Impulse Responses of Linear Systems (Continued)

If linear system is stable,

$$\sum_{i=0}^{\infty} \|h(i)\| = \sum_{i=0}^{\infty} \|CA^iB\| < \infty$$

$$\Updownarrow$$

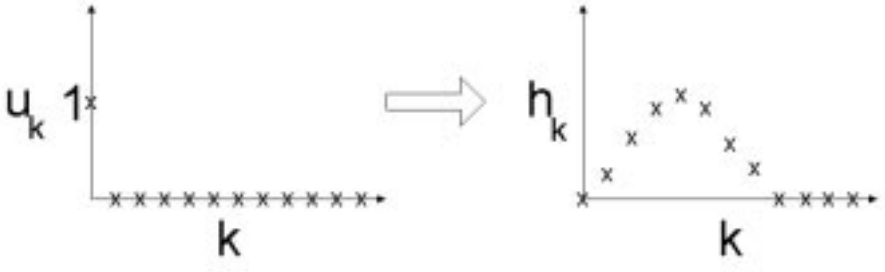
$$\{h(i)\}_{i=0}^{\infty} = \{CA^iB\}_{i=0}^{\infty} \in \ell_1$$

where ℓ_1 is the set of all absolutely summable sequences

$$\Downarrow$$

$$\lim_{i\to\infty} \|h(i)\| = \lim_{i\to\infty} \|CA^iB\| = 0$$

Finite Impulse Response Models



Finite Impulse Response (FIR) Model: Model for which there exists N such that

$$h(i) = 0 \quad \forall i \geq N$$

$$\Downarrow$$

$$y(k) = \sum_{i=1}^{N} h(i)u(k-i)$$

$$\Downarrow$$

FIR model is also called moving average model.

$$\Downarrow$$

Need to store n past inputs: $(u(i-1), \cdots, u(i-N))$

For stable linear systems, $h(i) \to 0$ as $i \to \infty$.

$$\Downarrow$$

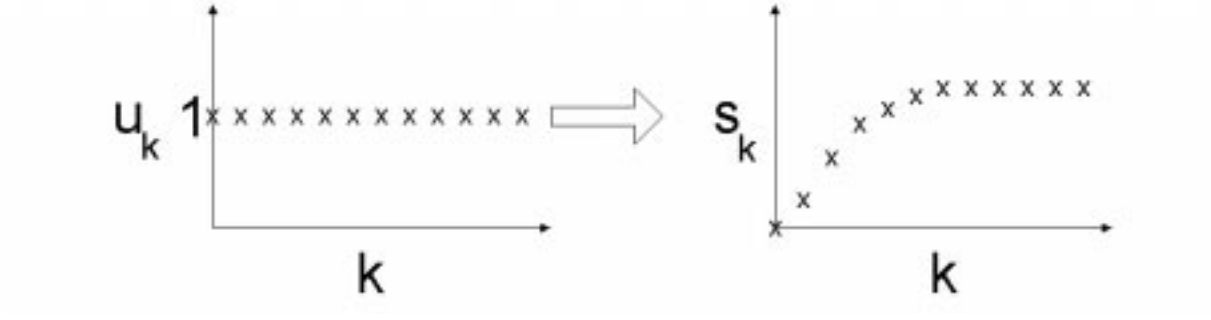
FIR model is a good approximation of a stable linear system for large enough N.

2.3 TRUNCATED STEP RESPONSE MODEL

Step Responses of Linear Systems

$$y(k) = h(k)x(0) + \sum_{i=0}^{k-1} h(k-i-1)u(i)$$

Step Response Sequence $\{s(k)\}$: $\{y(k)\}$ when $x(0) = 0$ and $u(i) = 1$, $i = 0, 1, 2, \cdots$.



Relationship between impulse and step responses:

$$s(k) = \sum_{i=1}^{k} h(i)$$

$$\Updownarrow$$

$$h(k) = s(k) - s(k-1)$$

Truncated Step Response Models

Truncated Step Response (TSR) Model: FIR model represented by its step responses.

$$y(k) = \sum_{i=1}^{N} h(i)u(k-i) = \sum_{i=1}^{N} s(i) - s(i-1)u(k-i)$$

$$= \sum_{i=1}^{N} s(i)u(k-i) - \sum_{i=1}^{N-1} s(i)u(k-i-1)$$

$$= \sum_{i=1}^{N-1} s(i)\Delta u(k-i) + s(N)u(k-N)$$

Truncated Step Response Models (Continued)

Let

$$\tilde{Y}(k) := \begin{bmatrix} y(k) \\ y(k+1) \\ \vdots \\ y(k+n-1) \end{bmatrix}$$

when $\Delta u(k) = \Delta u(k+1) = \cdots = 0$. Then

$$\tilde{Y}(k) := \begin{bmatrix} \Sigma_{i=1}^{N-1} s(i)\Delta u(k-i) + s(N)u(k-N) \\ \Sigma_{i=2}^{N-1} s(i)\Delta u(k+1-i) + s(N)u(k-N+1) \\ \Sigma_{i=3}^{N-1} s(i)\Delta u(k+2-i) + s(N)u(k-N+2) \\ \vdots \\ s(N-1)\Delta u(k-1) + s(N)u(k-2) \\ s(N)u(k-1) \end{bmatrix}$$

$$\tilde{Y}(k+1) := \begin{bmatrix} \Sigma_{i=1}^{N-1} s(i)\Delta u(k+1-i) + s(N)u(k-N+1) \\ \Sigma_{i=2}^{N-1} s(i)\Delta u(k+2-i) + s(N)u(k-N+2) \\ \Sigma_{i=3}^{N-1} s(i)\Delta u(k+3-i) + s(N)u(k-N+3) \\ \vdots \\ s(N-1)\Delta u(k) + s(N)u(k-1) \\ s(N)u(k) \end{bmatrix}$$

$$\Downarrow$$

$$\tilde{Y}(k+1) = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \tilde{Y}(k) + \begin{bmatrix} s(1) \\ s(2) \\ \vdots \\ s(N-1) \\ s(N) \end{bmatrix} \Delta u(k)$$

2.4 REACHABILITY AND OBSERVABILITY

Reachability

A state x is reachable if it can be reached from the zero state in some finite number of times by an appropriate input.

$$\Updownarrow$$

For some n and some $\{u(i)\}$,

$$x(0) = 0$$

$$x(k+1) = Ax(k) + Bu(k), \qquad 0 \le k \le n-1$$

$$x(n) = x$$

or

$$x = \sum_{i=0}^{n-1} A^{n-i-1} Bu(i) = Bu(n-1) + \cdots + A^{n-1} Bu(0)$$

A linear system

$$x(k+1) = Ax(k) + Bu(k)$$

$$y(k) = Cx(k)$$

is said to be reachable if any state in the state space is reachable.

$$\Updownarrow$$

$W_c := [B \; AB \; \cdots \; A^{n-1}B]$ has n linearly independent columns

Observability

Question: Given A, B, C, D and $\{u(i), y(i)\}_{i=1}^{n}$, can we determine the state $x(1)$ from this data?

$$y(i) = CA^{i-1}x(1) + \sum_{k=1}^{i-1} A^{i-k-1}Bu(k)$$

Define

$$\tilde{y}(i) = y(i) - \sum_{k=1}^{i-1} A^{i-k-1}Bu(k) = CA^{i-1}x(1)$$

$$\Downarrow$$

$$\begin{bmatrix} \tilde{y}(1) \\ \tilde{y}(2) \\ \vdots \\ \tilde{y}(n) \end{bmatrix} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix} x(1)$$

Observability (Continued)

A state x is observable if it is a unique solution of

$$\begin{bmatrix} \tilde{y}(1) \\ \tilde{y}(2) \\ \vdots \\ \tilde{y}(n) \end{bmatrix} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix} x$$

such that

$$y(i) = CA^{i-1}x + \sum_{k=1}^{i-1} A^{i-k-1}Bu(k)$$

A linear system

$$x(k+1) = Ax(k) + Bu(k)$$

$$y(k) = Cx(k)$$

is said to be observable if any state in the state space is observable.

$$\Updownarrow$$

$$W_o := \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix} \text{ has } n \text{ linearly independent rows}$$

2.5 STATIC STATE FEEDBACK CONTROLLER AND STATE ESTIMATOR

Linear Static State Feedback (Pole Placement)

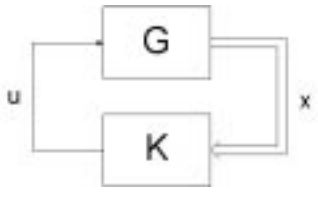
Consider a linear system

$$x(k+1) = Ax(k) + Bu(k)$$

$$y(k) = Cx(k)$$

Let $\{s_i\}_{i=1}^n$ be the set of desired closed loop poles and

$$P(z) = (z - s_1)(z - s_2)\cdots(z - s_n) = z^n + p_1 z^{n-1} + \cdots + p_n$$



Question (Pole Placement Problem): Does there exist linear static state feedback controller $u = Kx$ such that the characteristic polynomial for the closed loop system

$$x(k+1) = (A + BK)x(k)$$

is $P(z)$?

Linear Static State Feedback (Continued)

Suppose there exists T such that $z = Tx$ leads to controllable canonical form:

$$z(k+1) = \begin{bmatrix} -a_1 & -a_2 & \cdots & -a_{n-1} & -a_n \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} z(k) + \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} u(k)$$

$$\Downarrow$$

Characteristic polynomial:

$$z^n + a_1 z^{n-1} + \cdots + a_n = 0$$

If

$$u = -\bar{L}z$$

where

$$\bar{L} = [p_1 - a_1 \; p_2 - a_2 \; \cdots \; p_n - a_n]$$

$$\Downarrow$$

$$z(k+1) = \begin{bmatrix} -p_1 & -p_2 & \cdots & -p_{n-1} & -p_n \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} z(k)$$

$$\Downarrow$$

Closed loop characteristic polynomial:

$$z^n + p_1 z^{n-1} + \cdots + p_n = 0$$

Linear Static State Feedback (Continued)

Question: When does there exist such T?

Let

$$W_c := [B \; AB \; \cdots \; A^{n-1}B]$$

Then

$$\bar{W}_c := [TB \; (TAT^{-1})TB \; \cdots \; (TAT^{-1})^{n-1}TB]$$

$$= [TB \; TAB \; \cdots \; TA^{n-1}B] = TW_c$$

$$\Downarrow$$

If W_c is invertible,

$$T = \bar{W}_c W_c^{-1}$$

Theorem: Pole placement is possible iff the system is reachable.
The pole placing contoller is

$$u = -\bar{L}\bar{W}_c W_c^{-1} x$$

Linear Observer

Consider a linear system

$$x(k+1) = Ax(k) + Bu(k)$$

$$y(k) = Cx(k)$$

Suppose the states are not all measurable and y is only available.

Question: Can we design the state estimator such that the state estimate converges to the actual state?

Given the state estimate of $x(k)$ at $k-1$, $\hat{x}(k|k-1)$,

$$y(k) \neq C\hat{x}(k|k-1)$$

due to the estimation error.

$$\Downarrow$$

$$\hat{x}(k+1|k) = \underbrace{A\hat{x}(k|k-1) + Bu(k)}_{\text{prediction based on the model}} + \underbrace{K[y(k) - C\hat{x}(k|k-1)]}_{\text{correction based on the error}}$$

Define the estimation error as

$$\tilde{x} := x - \hat{x}$$

$$\Downarrow$$

$$\tilde{x}(k+1|k) = A\tilde{x}(k|k-1) - K[y(k) - C\hat{x}(k|k-1)]$$

$$= A\tilde{x}(k|k-1) - K[Cx(k) - C\hat{x}(k|k-1)] = [A - KC]\tilde{x}(k|k-1)$$

Linear Observer (Continued)

Question: Does there exist K such that the characteristic polynomial of $x(k+1) = (A + KC)x(k)$ is the desired polynomial $P(z)$?

$$\Updownarrow$$

Does there exist linear static state feedback controller $v = K^T z$ for the system

$$z(k+1) = A^T z(k) + C^T v(k)$$

such that the characteristic polynomial for the closed loop system

$$z(k+1) = (A^T + C^T K^T) z(k)$$

is the desired polynomial $P(z)$?

From pole placement, we know that this is possible iff

$$[C^T \; A^T C^T \; \cdots \; (A^T)^{n-1} C^T] =: W_o^T$$

is invertible and

$$K = -W_o^{-1} \bar{W}_o \bar{K}$$

where

$$\bar{W}_o = W_o T^T$$

$$\bar{K} = \begin{bmatrix} p_1 - a_1 \\ p_2 - a_2 \\ \vdots \\ p_n - a_n \end{bmatrix}$$

Chapter 3

BASICS OF OPTIMIZATION

3.1 INTRODUCTION

Ingredients of Optimization

- Decision variables ($x \in \mathbf{R}^n$): undetermined parameters
- Cost function ($f : \mathbf{R}^n \to \mathbf{R}$): the measure of preference
- Constraints ($h(x) = 0$, $g(x) \leq 0$): equalities and inequalities that the decision variables must satisfy

$$\min_{x \in \mathbf{R}^n} f(x)$$
$$h(x) = 0$$
$$g(x) \leq 0$$

Example

Consider control problem associated with the linear system

$$x_{k+1} = Ax_k + Bu_k$$

Decision variables: x_k, u_k, $k = 0, 1, \cdots, N$

Cost function:

- x_k is preferred to be close to the origin, the desired steady state.
- Large control action is not desirable.

$$\Downarrow$$

One possible measure of good control is

$$\sum_{i=1}^{N} x_i^T x_i + \sum_{i=0}^{N-1} u_i^T u_i$$

Constraints: decision variables, x_{k+1}, u_k, $k = 0, 1, \cdots, N$, must satisfy the dynamic constraints

$$x_{k+1} = Ax_k + Bu_k$$

$$\Downarrow$$

$$\min_{u_k, x_k} \sum_{i=1}^{N} x_i^T x_i + \sum_{i=0}^{N-1} u_i^T u_i$$

subject to

$$x_{k+1} = Ax_k + Bu_k$$

Terminologies

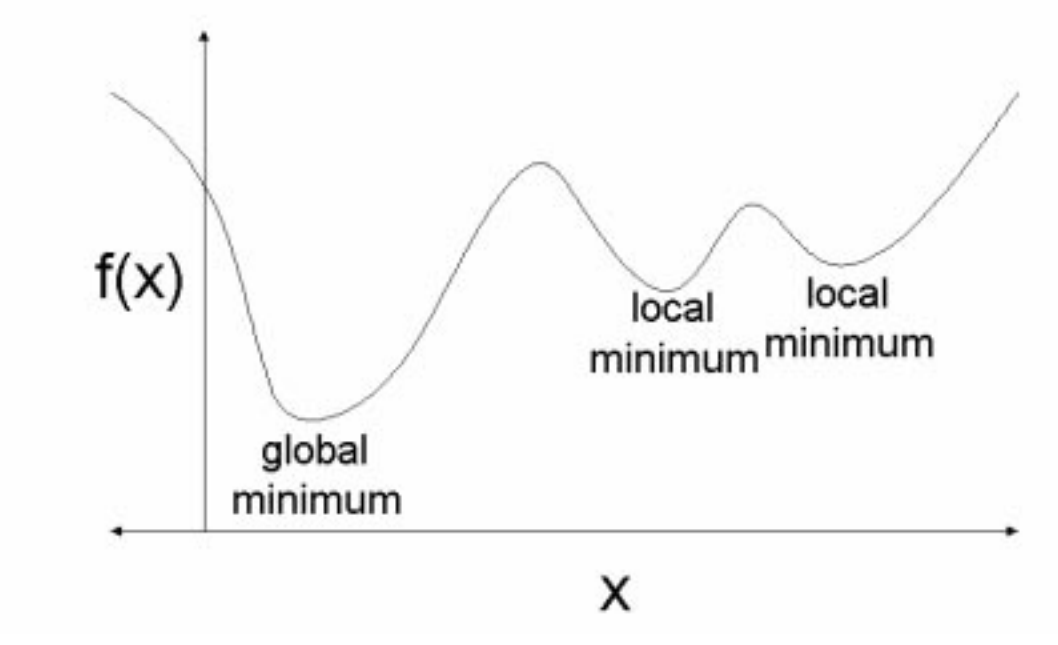
Let

$$\Omega = \{x \in \mathbf{R}^n : h(x) = 0,\ g(x) \leq 0\}$$

Feasible point: any $x \in \Omega$

Local minimum: $x^* \in \Omega$ such that $\exists \epsilon > 0$ for which $f(x^*) \leq f(x)$ for all $x \in \Omega \cap \{x \in \mathbf{R}^n : \|x - x^*\| < \epsilon\}$.

Global minimum: $x^* \in \Omega$ such that $f(x^*) \leq f(x)$ for all $x \in \Omega$.

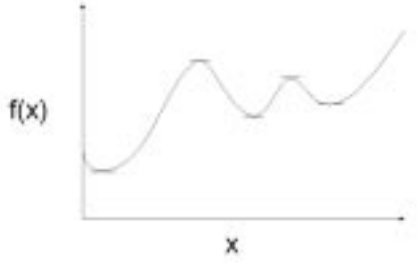


3.2 UNCONSTRAINED OPTIMIZATION PROBLEMS

Necessary Condition of Optimality for Unconstrained Optimization Problems

From calculus, the extrema x^* of a function f from $\mathbf{R}$ to $\mathbf{R}$ must satisfy

$$\frac{df}{dx}(x^*) = 0$$



$$\Downarrow$$

The minima for 1-D unconstrained problem:

$$\min_{x \in \mathbf{R}} f(x)$$

must satisfy

$$\frac{df}{dx}(x^*) = 0$$

that is only necessary.

Necessary Condition of Optimality for Unconstrained Optimization Problems (Continued)

In general, the optima for n-D unconstrained problem:

$$\min_{x \in \mathbf{R}^n} f(x)$$

satisfy the following necessary condition of optimality

$$\nabla f(x^*) = 0$$

(n equations and n unknowns)

Example: Consider

$$\min_{x \in \mathbf{R}^n} \frac{1}{2} x^T H x + g^T x$$

The necessary condition of optimality for this problem is

$$[\nabla f(x^*)]^T = Hx^* + g = 0$$

If H is invertible,

$$x^* = -H^{-1} g$$

Steepest Descent Methods

for Unconstrained Nonlinear Programs

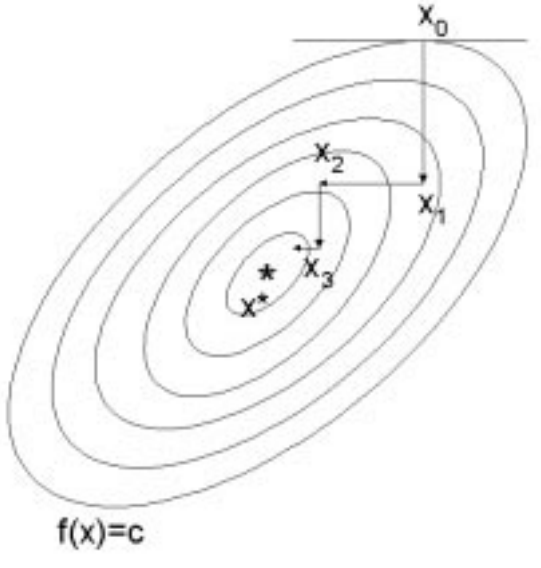
The meaning of gradient $\nabla f(x)$: the steepest ascent direction at the given point.

Main idea: search the minimum in the steepest descnt direction

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k)$$

where

$$\alpha_k = \text{argmin}_\alpha f(x_k - \alpha \nabla f(x_k))$$



Newton's Method

for Unconstrained Nonlinear Programs

Main idea:

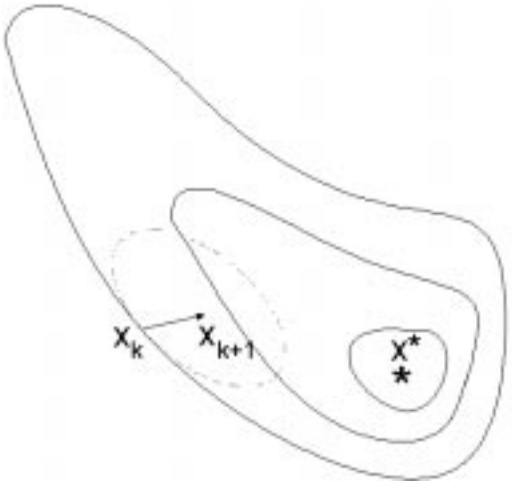
1. Approximate the object function by quadradic function
2. Solve the resulting quadratic problem

Quadratic approximation:

$$f(x) \approx f(x_k) + \nabla f(x_k)(x - x_k) + \frac{1}{2}(x - x_k)^T \nabla^2 f(x_k)(x - x_k)$$

Exact solution of the quadratic program:

$$x_{k+1} = x_k - [\nabla^2 f(x_k)]^{-1} \nabla f(x_k)^T$$



3.3 NECESSARY CONDITION OF OPTIMALITY FOR CONSTRAINED OPTIMIZATION PROBLEMS

Constrained Optimization Problems

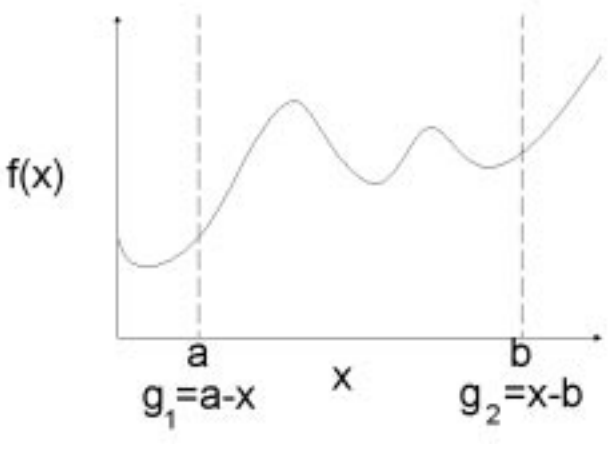
Consider

$$\min_{x \in \mathbf{R}} f(x)$$

subject to

$$g_1(x) = a - x \leq 0$$

$$g_2(x) = x - b \leq 0$$



$$\Downarrow$$

$\nabla f(x^*) = 0$ is not the necessary condition of optimality anymore.

Lagrange Multiplier

Consider

$$\min_{x \in \mathbf{R}^n} f(x)$$

subject to

$$h(x) = 0$$

At the minimum, the m constraint equations must be saisfied

$$h(x^*) = 0$$

Moreover, at the minimum,

$$df(x^*) = \frac{df}{dx}(x^*)dx = 0$$

must hold in any feasible direction.

Feasible direction, $dx^\dagger$, must satisfy

$$dh(x^*) = \frac{dh}{dx}(x^*)dx^\dagger = 0$$

$$\Updownarrow$$

For any $y = \Sigma_{i=1}^m a_i \frac{dh_i}{dx}(x^*)$,

$$y^T dx^\dagger = 0$$

Lagrange Multiplier (Continued)

$df(x^*) = \frac{df}{dx}(x^*)dx^\dagger = 0$ must hold

$$\Downarrow$$

$\frac{df}{dx}(x^*)$ is linearly dependent on $\{\frac{dh_i}{dx}(x^*)\}_{i=1}^m$

$$\Downarrow$$

$\exists\ \{\lambda_i\}_{i=1}^m$ such that

$$\frac{df}{dx}(x^*) + \sum_{i=1}^m \lambda_i \frac{dh_i}{dx}(x^*) = 0$$

$$\Downarrow$$

Necessary Condition of Optimality:

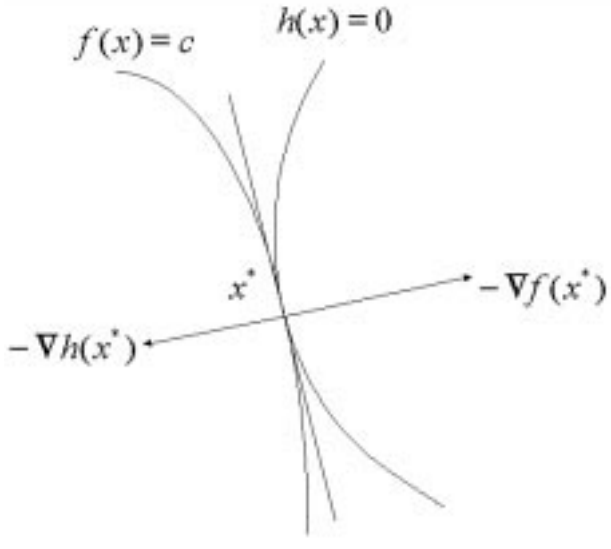
$$h(x^*) = 0 \quad m \text{ equations}$$

$$\frac{df}{dx}(x^*) + \sum_{i=1}^m \lambda_i \frac{dh_i}{dx}(x^*) = 0 \quad n \text{ equations}$$

where λ_i's are called Lagrange Multipliers.

($n + m$ equations and $n + m$ unknowns)

Lagrange Multiplier (Continued)



x* is a local minimum

Lagrange Multiplier (Continued)

Example: Consider

$$\min_{x \in \mathbf{R}^n} \frac{1}{2} x^T H x + g^T x$$

subject to

$$Ax - b = 0$$

The necessary condition of optimality for this problem is

$$[\nabla f(x^*)]^T + [\nabla h(x^*)]^T \lambda = Hx^* + g + A^T \lambda = 0$$

$$h(x^*) = Ax^* - b = 0$$

$$\Downarrow$$

$$Hx^* + A^T \lambda = -g$$

$$Ax^* = b$$

$$\Downarrow$$

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ \lambda \end{bmatrix} = \begin{bmatrix} -g \\ b \end{bmatrix}$$

If $\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix}$ is invertible,

$$\begin{bmatrix} x^* \\ \lambda \end{bmatrix} = \begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix}^{-1} \begin{bmatrix} -g \\ b \end{bmatrix}$$

Kuhn-Tucker Condition

Let x^* be a local minimum of

$$\min f(x)$$

subject to

$$h(x) = 0$$

$$g(x) \leq 0$$

and suppose x^* is a regular point for the constraints. Then $\exists\ \lambda$ and μ such that

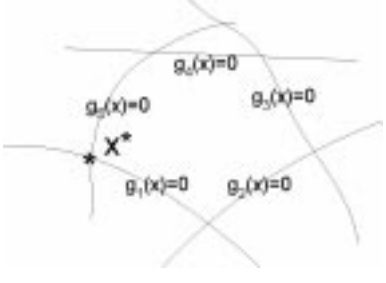
$$\nabla f(x^*) + \lambda^T \nabla h(x^*) + \mu^T \nabla g(x^*) = 0$$

$$\mu^T g(x^*) = 0$$

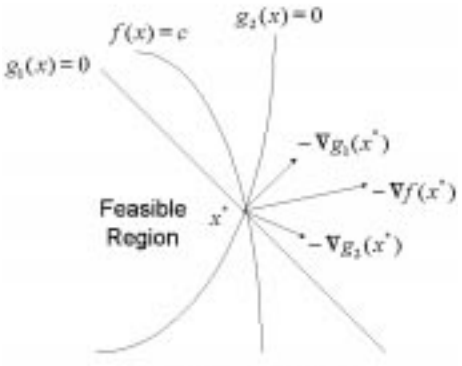
$$h(x^*) = 0$$

$$\mu \geq 0$$

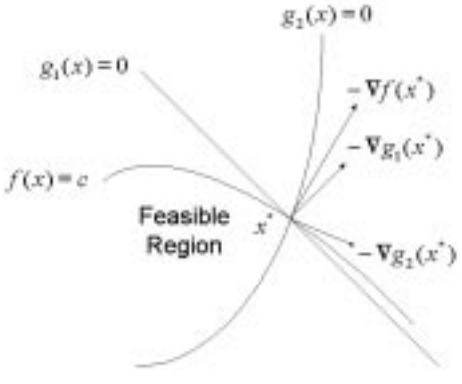
$g_i(x^*) < 0 \Rightarrow \mu_i = 0$



Kuhn-Tucker Condition(Continued)



x* is a local minimum



x* is not a local minimum

Kuhn-Tucker Condition (Continued)

Example: Consider

$$\min_{x \in \mathbf{R}^n} \frac{1}{2} x^T H x + g^T x$$

subject to

$$Ax - b = 0$$

$$Cx - d \leq 0$$

The necessary condition of optimality for this problem is

$$[\nabla f(x^*)]^T + [\nabla h(x^*)]^T \lambda + [\nabla g(x^*)]^T \mu = Hx^* + g + A^T \lambda + C^T \mu = 0$$

$$g(x^*)^T \mu = (x^{*T} C^T + d^T)\mu = 0$$

$$h(x^*) = Ax^* - b = 0$$

$$\mu \geq 0$$

$$\Downarrow$$

$$Hx^* + A^T \lambda + C^T \mu = -g$$

$$x^{*T} C^T \mu + d^T \mu = 0$$

$$Ax^* = b$$

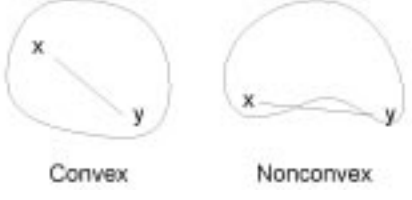
$$\mu \geq 0$$

3.4 CONVEX OPTIMIZATION

Convexity

Convex set: $C \subset \mathbf{R}^n$ is convex if

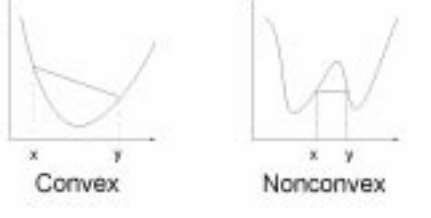
$$x, y \in C, \quad \lambda \in [0,1] \quad \Rightarrow \quad \lambda x + (1-\lambda)y \in C$$



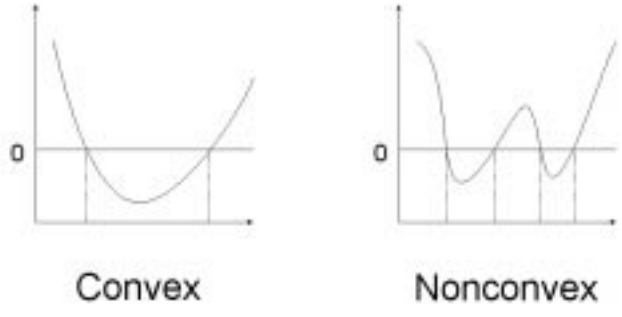
Convex Functions: $f : \mathbf{R}^n \to \mathbf{R}$ is convex if

$$x, y \in \mathbf{R}^n, \quad \lambda \in [0,1]$$

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$$



Convexity (Continued)



Notice that $\{x : g(x) \leq 0\}$ is convex if g is convex.

Theorem: If f and g are convex any local optimum is globally optimal.

Linear Programs

$$\min_{x \in \mathbf{R}^n} a^T x$$

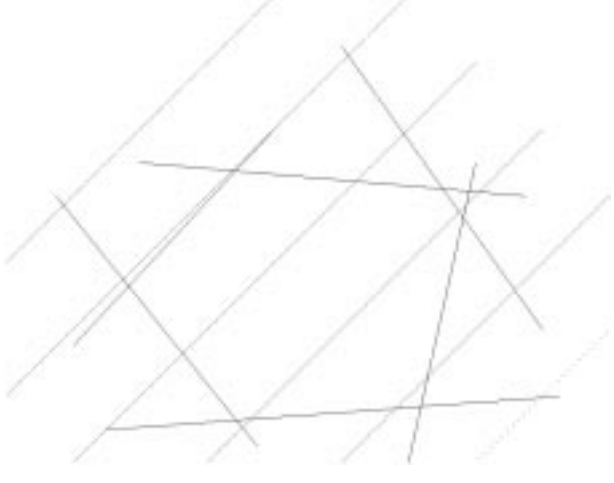
subject to

$$Bx \leq b$$

Linear program is a convex program.

Feasible basic solution: feasible solution that satisfies n of the constraints as equalities.

Fact: If an optimal solution exists, there exists a feasible basic solution that is optimal.

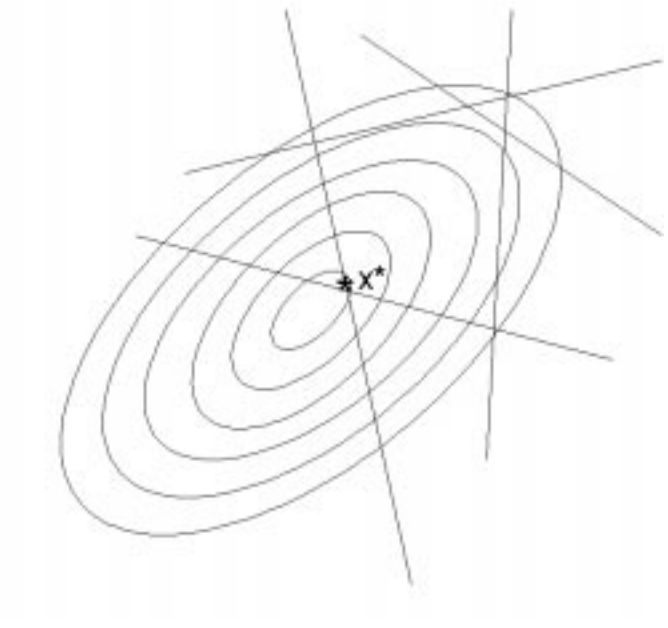


Quadratic Programs

$$\min_{x \in \mathbf{R}^n} \frac{1}{2} x^T H x + g^T x$$

subject to

$$Ax \leq b$$



Quadratic program is convex if H is positive semi-definite.

3.5 ALGORITMS FOR CONSTRAINED OPTIMIZATION PROBLEMS

Algorithms for Linear Program

Simplex Method

Motivation: There always exists a basic optimal solution.

Main Idea:

- Find a basic solution.
- Find another basic solution with lower cost function value.
- Continue until another basic solution with lower cost function value cannot be found.

Simplex algorithm always finds a basic optimal solution.

Algorithms for Linear Program (Continued)

Interior Point Method

Main Idea:

- Define barrier function:

$$B = -\sum_{i=1}^{m} \frac{1}{c_i^T x - b_i}$$

- Form the unconstrained problem:

$$\min_x a^T x + \frac{1}{K} B(x)$$

- Solve the unconstrained problem using Newton method.
- Increase K and solve the unconstrained problem again until the solution converges.

- Remarkably, problems seem to converge between 5 to 50 Newton steps regerdless of the problem size.
- Can exploit structures of the problem (e.g. sparsity) to reduce computation time per Newton step.
- Can be extended to general nonlinear convex problems such as quadratic programs.

Algorithms for Quadratic Program

Active Set Method

Main Idea:

- Determine the active constraints and set them as equality constraints.
- Solve the resulting problem.
- Check the Kuhn-Tucker condition that is also suffucuent for QP.
- If Kuhn-Tucker condition is not satisfied, try another set of active constraints.

Interior Point Method

- The main idea of interior point method for QP is the same as that for LP.

Generalized Reduced Gradient Method for Constrained Nonlinear Programs

Main idea:

1. Linearize the equality constraints that are possibly obtained adding slack variables
2. Solve the resulting linear equations for m variables
3. Apply the steepest descent method with respect to $n - m$ variables

Linearization of Constraints:

$$\nabla_y h(y,z)dy + \lambda^T \nabla_z h(y,z)dz = 0$$

$$\Downarrow$$

$$dy = -[\nabla_y h(y,z)]^{-1}\lambda^T \nabla_z h(y,z)dz$$

Generalized Reduced Gradient of Objective Function:

$$df(y,z) = \nabla_y f(y,z)dy + \lambda^T \nabla_z f(y,z)dz$$

$$= [\lambda^T \nabla_z f(y,z) - \nabla_y f(y,z)[\nabla_y h(y,z)]^{-1}\lambda^T \nabla_z h(y,z)]dz$$

$$\Downarrow$$

$$r = \frac{df}{dz} = \lambda^T \nabla_z f(y,z) - \nabla_y f(y,z)[\nabla_y h(y,z)]^{-1}\lambda^T \nabla_z h(y,z)$$

Penalty Method for Constrained Nonlinear Programs

Main idea: Instead of forcing the constraints, penalize the violation of the constraints in the objective.

$$\min_x f(x) - c_k g(x) \qquad (P_k)$$

where $c_k > 0$.

Theorem: Let x_k be the optimal solution of (P_k). Then as $c_k \to \infty$, $x_k \to x^*$.

Successive QP Method

for Constrained Nonlinear Programs

Main idea:

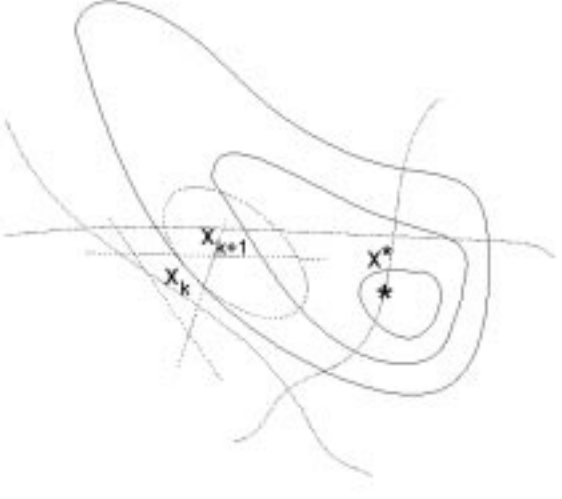
1. Approximate the object function by quadradic function and constraints linear function.
2. Solve the resulting quadratic problem

Approximate Quadratic Program:

$$\min \nabla f dx + \frac{1}{2} dx^T \nabla^2 f dx$$

subject to

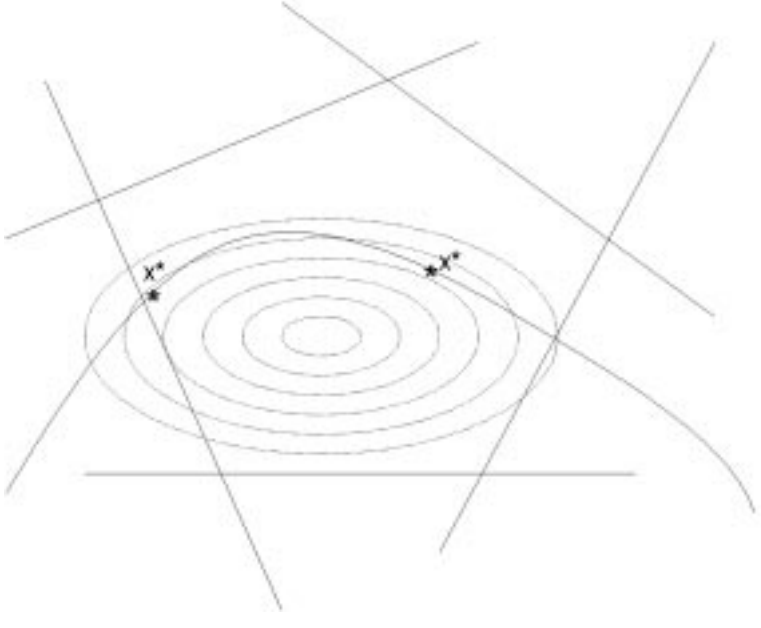
$$g(x) + \nabla g(x) dx \leq 0$$



Nonconvex Programs

The aforementioned optimization algorithms indentify only one local optimum.

However, a nonconvex optimization problem may have a number of local optima.

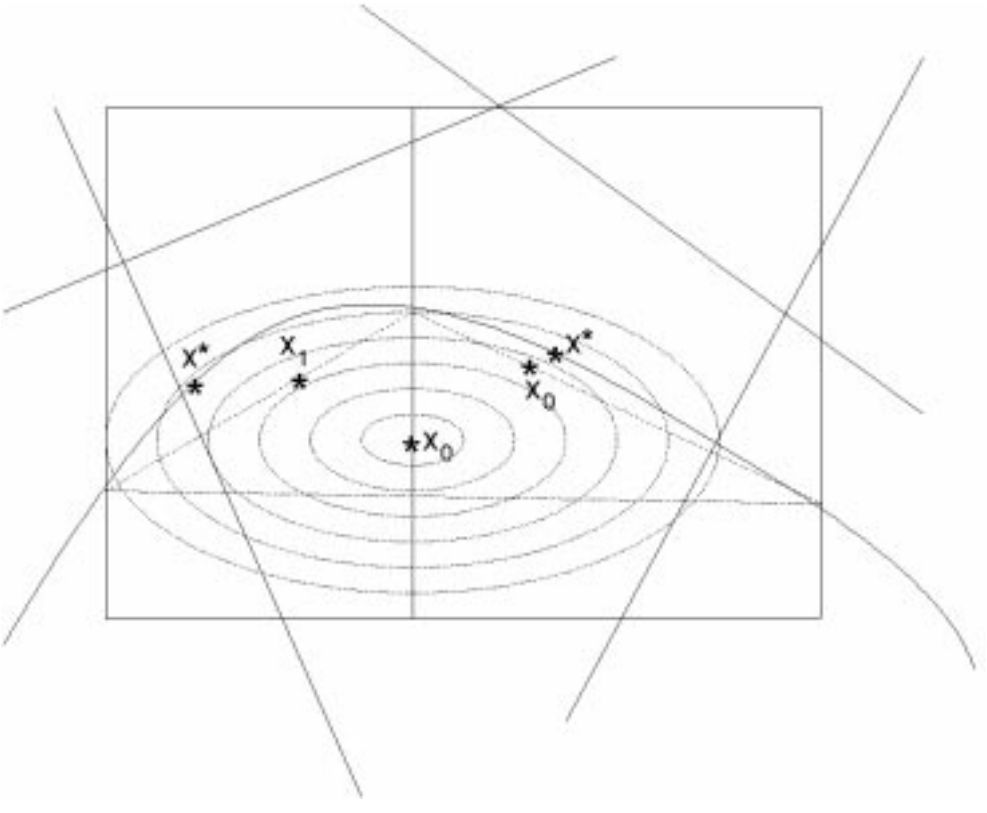


⇓

Algorithms that indentifies a global optimum are necessary

A Global Optimization Algorithm for Noconvex Programs

Branch and bound type global optimization algorithm:



- Branching Step: split the box at the optimum
- Bounding Step: find the box where the optimum is lowest

Chapter 4

RANDOM VARIABLES

4.1 INTRODUCTION

What Is Random Variable?
We are dealing with

- a physical phenomenon which exhibits randomness.
- the outcome of any one occurence (trial) cannot be predicted.
- the probability of any subset of possible outcomes is well-defined.

We ascribe the term *random variable* to such a phenomenon. Note that a random variable is not defined by a specific number; rather it is defined by the probabilities of all subsets of the possible outcomes. An outcome of a particular trial is called a *realization* of the random variable.

An example is outcome of rolling a dice. Let x represent the outcome (not of a particular trial, but in general). Then, x is not represented by a single outcome, but is defined by the set of possible outcomes ($\{1, 2, 3, 4, 5, 6\}$) and the probability of the possible outcome(s) (1/6 each). When we say x is 1 or 2 or so on, we really should say a realization of x is such.

A random variable can be discrete or continuous. If the outcome of a random variable belongs to a discrete space, the random variable is *discrete*. An example is the outcome of rolling a dice. On the other hand, if the outcome belongs to a continuous space, the random variable is *continuous*. For instance, composition or temperature of a distillation column can be viewed as continuous random variables.

What Is Statistics?

Statistics deals with the application of probability theory to real problems. There are two basic problems in statistics.

- Given a probabilistic model, predict the outcome of future trial(s). For instance one may say:

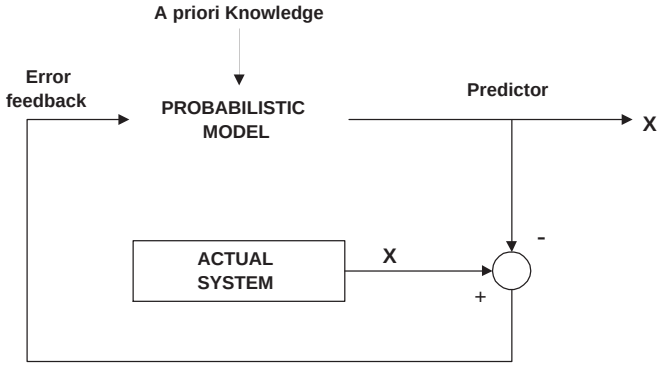
  choose the prediction $\hat{x}$ such that expected value of $(x - \hat{x})^2$ is minimized.

- Given collected data, define / improve a probabilistic model.

  For instance, there may be some unknown parameters (say θ) in the probabilistic model. Then, given data X generated from the particular probabilistic model, one should construct an estimate of θ in the form of $\hat{\theta}(X)$. For example, $\hat{\theta}(X)$ may be constructed based on the objective of minimizing expected value of $\|\theta - \hat{\theta}\|_2^2$.

  Another related topic is *hypothesis testing*, which has to do with testing whether a given hypothesis is correct (i.e, how correct defined in terms of probability), based on available data.

In fact, one does both. That is, as data come in, one may continue to improve the probabilistic model and use the updated model for further prediction.



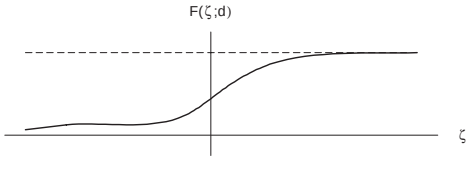
4.2 BASIC PROBABILITY CONCEPTS

4.2.1 PROBABILITY DISTRIBUTION, DENSITY: SCALAR CASE

A random variable is defined by a function describing the probability of the outcome rather than a specific value. Let d be a *continuous* random variable ($d \in \mathcal{R}$). Then one of the following functions is used to define d:

- **Probability Distribution Function**
  The probability distribution function $F(\zeta; d)$ for random variable d is defined as

$$F(\zeta; d) = \Pr\{d \leq \zeta\} \tag{4.1}$$

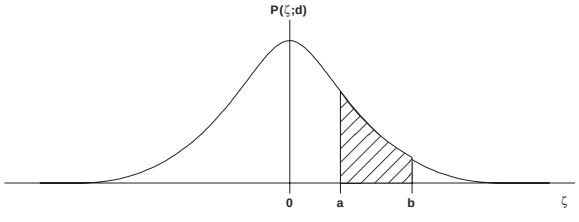


  where Pr denotes the probability. Note that $F(\zeta; d)$ is monotonically increasing with ζ and asymptotically reaches 1 as ζ approaches its upper limit.

- **Probability Density Function**

  The probability density function $\mathcal{P}(\zeta; d)$ for random variable d is defined as

$$\mathcal{P}(\zeta; d) = \frac{dF(\zeta; d)}{d\zeta} \tag{4.2}$$



  Note that

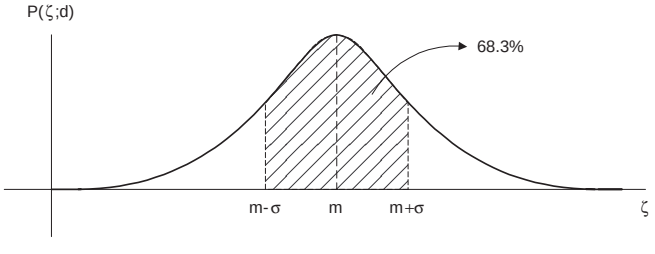
$$\int_{-\infty}^{\infty} \mathcal{P}(\zeta; d) d\zeta = \int_{-\infty}^{\infty} dF(\zeta; d) = 1 \tag{4.3}$$

  In addition,

$$\int_a^b \mathcal{P}(\zeta; d)\ d\zeta = \int_a^b dF(\zeta; d) = F(b; d) - F(a; d) = Pr\{a < d \leq b\} \tag{4.4}$$

Example: Guassian or Normally Distributed Variable

$$\mathcal{P}(\zeta; d) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\left(\frac{\zeta - m}{\sigma}\right)^2\right\} \tag{4.5}$$



Note that this distribution is determined entirely by two parameters (the mean m and standard deviation σ).

4.2.2 PROBABILITY DISTRIBUTION, DENSITY: VECTOR CASE

Let $d = \begin{bmatrix} d_1 & \cdots & d_n \end{bmatrix}^T$ be a *continuous* random variable vector($d \in \mathcal{R}^n$). Now we must quantify the distribution of its individual elements as well as their correlations.

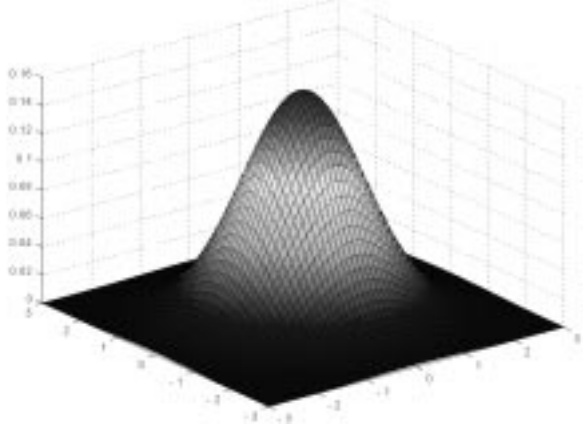
- **Joint Probability Distribution Function**
  The *joint probability distribution* function $F(\zeta_1, \cdots, \zeta_n; d_1, \cdots, d_n)$ for random variable vector d is defined as

$$F(\zeta_1, \cdots, \zeta_n; d_1, \cdots, d_n) = Pr\{d_1 \leq \zeta_1, \cdots, d_n \leq \zeta_n\} \tag{4.6}$$

  Now the domain of F is an n-dimensional space. For example, for $n = 2$, F is represented by a surface. Note that $F(\zeta_1, \cdots, \zeta_n; d_1, \cdots, d_n) \to 1$ as $\zeta_1, \cdots, \zeta_n \to \infty$.

- **Joint and Marginal Probability Density Function**
  The *joint probability density* function $\mathcal{P}(\zeta_1, \cdots, \zeta_n; d_1, \cdots, d_n)$ for random variable vector d is defined as

$$\mathcal{P}(\zeta_1, \cdots, \zeta_n; d_1, \cdots, d_n) = \frac{\partial^n F(\zeta; d)}{\partial \zeta_1, \cdots, \zeta_n} \tag{4.7}$$



For convenience, we may write $\mathcal{P}(\zeta; d)$ to denote $\mathcal{P}(\zeta_1, \cdots, \zeta_n; d_1, \cdots, d_n)$. Again,

$$\begin{aligned} & \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} \mathcal{P}(\zeta_1, \cdots, \zeta_n; d_1, \cdots, d_n) \;\; d\zeta_1 \cdots d\zeta_n \\ = \;\; & \Pr\{a_1 < d_1 \leq b_1, \cdots, a_n < d_n \leq b_n\} \end{aligned} \tag{4.8}$$

Naturally,

$$\int_{-\infty}^{\infty}, \cdots, \int_{-\infty}^{\infty} \mathcal{P}(\zeta_1, \cdots, \zeta_n; d_1, \cdots, d_n) d\zeta_1 \cdots d\zeta_n = 1 \tag{4.9}$$

We can easily derive the probability density of individual element from the joint probability density. For instance,

$$\mathcal{P}(\zeta_1; d_1) = \int_{-\infty}^{\infty}, \cdots, \int_{-\infty}^{\infty} \mathcal{P}(\zeta_1, \cdots, \zeta_n; d_1, \cdots, d_n) \;\; d\zeta_2 \cdots d\zeta_n \tag{4.10}$$

This is called marginal probability density.

While the joint probability density (or distribution) tells us the likelihood of several random variables achieving certain values simultaneously, the marginal density tells us the likelihood of one element achieving certain value when the others are not known.

Note that in general

$$\mathcal{P}(\zeta_1, \cdots, \zeta_n; d_1, \cdots, d_n) \neq \mathcal{P}(\zeta_1; d_1) \cdots \mathcal{P}(\zeta_n; d_n) \tag{4.11}$$

If

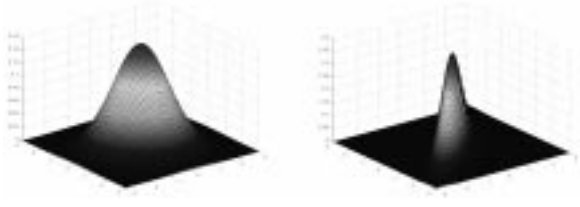
$$\mathcal{P}(\zeta_1, \cdots, \zeta_n; d_1, \cdots, d_n) = \mathcal{P}(\zeta_1; d_1) \cdots \mathcal{P}(\zeta_n; d_n) \tag{4.12}$$

$d_1, \cdots, d_n$ are called *mutually independent*.

Example: Guassian or Jointly Normally Distributed Variables

Suppose that $\mathbf{d} \triangleq [d_1 \; d_2]^T$ is a Gaussian variable. The density takes the form of

$$\begin{aligned}\mathcal{P}(\zeta_1, \zeta_2; d_1, d_2) &= \frac{1}{2\pi\sigma^1\sigma_2(1-\rho^2)^{1/2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{\zeta_1 - m_1}{\sigma_1}\right)^2\right.\right. \\ &\left.\left. -2\rho\frac{(\zeta_1 - m_1)(\zeta_2 - m_2)}{\sigma_1\sigma_2} + \left(\frac{\zeta_2 - m_2}{\sigma_2}\right)^2\right]\right\} \qquad (4.13)\end{aligned}$$



Note that this density is determined by five parameters (the means m_1, m_2, standard deviations σ_1, σ_2 and correlation parameter ρ). $\rho = 1$ represents complete correlation between d_1 and d_2, while $\rho = 0$ represents no correlation.

It is fairly straightforward to verify that

$$\begin{aligned}\mathcal{P}(\zeta_1; d_1) &= \int_{-\infty}^{\infty} \mathcal{P}(\zeta_1, \zeta_2; d_1, d_2) \; d\zeta_2 \qquad (4.14) \\ &= \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left\{-\frac{1}{2}\left(\frac{\zeta_1 - m_1}{\sigma_1}\right)^2\right\} \qquad (4.15)\end{aligned}$$

$$\begin{aligned}\mathcal{P}(\zeta_2; d_2) &= \int_{-\infty}^{\infty} \mathcal{P}(\zeta_1, \zeta_2; d_1, d_2) \; d\zeta_1 \qquad (4.16) \\ &= \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left\{-\frac{1}{2}\left(\frac{\zeta_2 - m_2}{\sigma_2}\right)^2\right\} \qquad (4.17)\end{aligned}$$

Hence, (m_1, σ_1) and (m_2, σ_2) represent parameters for the marginal density of d_1 and d_2 respectively. Note also that

$$\mathcal{P}(\zeta_1, \zeta_2; d_1, d_2) \neq \mathcal{P}(\zeta_1; d_1)\mathcal{P}(\zeta_2; d_2) \qquad (4.18)$$

except when $\rho = 0$.

General n-dimensional Gaussian random variable vector $d = [d_1, \cdots, d_n]^T$ has the density function of the following form:

$$\mathcal{P}(\zeta; d) \triangleq \mathcal{P}(\zeta_1, \cdots, \zeta_n; d_1, \cdots, d_n) \tag{4.19}$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}} |P_d|^{1/2}} \exp\left\{-\frac{1}{2}(\zeta - \bar{d})^T P_d^{-1} (\zeta - \bar{d})\right\} \tag{4.20}$$

where the parameters are $\bar{d} \in \mathcal{R}^n$ and $P_d \in \mathcal{R}^{n \times n}$. The significance of these parameters will be discussed later.

4.2.3 EXPECTATION OF RANDOM VARIABLES AND RANDOM VARIABLE FUNCTIONS: SCALAR CASE

Random variables are completely characterized by their distribution functions or density functions. However, in general, these functions are nonparametric. Hence, random variables are often characterized by their moments up to a finite order; in particular, use of the first two moments is quite common.

- **Expection of Random Variable Fnction**
  Any function of d is a random variable. Its expectation is computed as follows:
  $$E\{f(d)\} \triangleq \int_{-\infty}^{\infty} f(\zeta)\mathcal{P}(\zeta; d) \; d\zeta \tag{4.21}$$

- **Mean**

  $$\bar{d} \triangleq E\{d\} = \int_{-\infty}^{\infty} \zeta \mathcal{P}(\zeta; d) \; d\zeta \tag{4.22}$$

  The above is called mean or expectation of d.

- **Variance**

$$\text{Var}\{d\} \triangleq E\{(d-\bar{d})^2\} = \int_{-\infty}^{\infty}(\zeta-\bar{d})^2\mathcal{P}(\zeta;d)\ \ d\zeta \tag{4.23}$$

The above is the "variance" of d and quantifies the extent of d deviating from its mean.

Example: Gaussian Variable
For Gaussian variable with density

$$\mathcal{P}(\zeta;d) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2}\left(\frac{\zeta-m}{\sigma}\right)^2\right\} \tag{4.24}$$

it is easy to verify that

$$\bar{d} \triangleq E\{d\} = \int_{-\infty}^{\infty}\zeta\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2}\left(\frac{\zeta-m}{\sigma}\right)^2\right\}\ \ d\zeta = m \tag{4.25}$$

$$\text{Var}\{d\} \triangleq E\{(d-\bar{d})^2\} = \int_{-\infty}^{\infty}(\zeta-m)^2\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2}\left(\frac{\zeta-m}{\sigma}\right)^2\right\}\ \ d\zeta = \sigma^2 \tag{4.26}$$

Hence, m and σ^2 that parametrize the normal density represent the mean and the variance of the Gaussian variable.

4.2.4 EXPECTATION OF RANDOM VARIABLES AND RANDOM VARIABLE FUNCTIONS: VECTOR CASE

We can extend the concepts of mean and variance similarly to the vector case. Let $\mathbf{d}$ be a random variable vector that belongs to $\mathcal{R}^n$.

$$\begin{aligned}\bar{d}_\ell &= E\{d_\ell\} = \int_{-\infty}^{\infty}\zeta_\ell\mathcal{P}(\zeta_\ell;d_\ell)\ \ d\zeta_\ell \\ &= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\zeta_\ell\mathcal{P}(\zeta_1,\cdots,\zeta_n;d_1,\cdots,d_n)\ \ d\zeta_1,\cdots,d\zeta_n\end{aligned} \tag{4.27}$$

$$\text{Var}\{d_\ell\} = E\{(d_\ell-\bar{d}_\ell)^2\} = \int_{-\infty}^{\infty}(\zeta_\ell-\bar{d}_\ell)^2\mathcal{P}(\zeta_\ell;d_\ell)\ \ d\zeta_\ell \tag{4.28}$$

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (\zeta_\ell - \bar{d}_\ell)^2 \mathcal{P}(\zeta_1, \cdots, \zeta_n; d_1, \cdots, d_n) \;\; \zeta_1, \cdots, d\zeta_n$$

In the vector case, we also need to quantify the correlations among different elements.

$$\begin{aligned} &\mathrm{Cov}\{d_\ell, d_m\} = E\{(d_\ell - \bar{d}_\ell)(d_m - \bar{d}_m)\} \qquad (4.29) \\ = &\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (\zeta_\ell - \bar{d}_\ell)(\zeta_m - \bar{d}_m) \mathcal{P}(\zeta_1, \cdots, \zeta_n; d_1, \cdots, d_n) \;\; d\zeta_1, \cdots, d\zeta_n \end{aligned}$$

Note that

$$\mathrm{Cov}\{d_\ell, d_\ell\} = \mathrm{Var}\{d_\ell\} \qquad (4.30)$$

The ratio

$$\rho = \frac{\mathrm{Cov}\{d_\ell, d_m\}}{\sqrt{\mathrm{Var}\{d_\ell\}\mathrm{Var}\{d_m\}}} \qquad (4.31)$$

is the correlation factor. $\rho = 1$ indicates complete correlation (d_ℓ is determined uniquely by d_m and vice versa). $\rho = 0$ indicates no correlation.

It is convenient to define covariance matrix for d, which contains all variances and covariances of $d_1, \cdots, d_n$.

$$\begin{aligned} &\mathrm{Cov}\{d\} = E\{(d - \bar{d})(d - \bar{d})^T\} \qquad (4.32) \\ = &\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (\zeta - \bar{d})(\zeta - \bar{d})^T \mathcal{P}(\zeta_1, \cdots, \zeta_n; d_1, \cdots, d_n) \;\; d\zeta_1, \cdots, d\zeta_n \end{aligned}$$

The $(i,j)_{\text{th}}$ element of $\mathrm{Cov}\{d\}$ is $\mathrm{Cov}\{d_i, d_j\}$. The diagonal elements of $\mathrm{Cov}\{d\}$ are variances of elements of d. The above matrix is symmetric since

$$\mathrm{Cov}\{d_i, d_j\} = \mathrm{Cov}\{d_j, d_i\} \qquad (4.33)$$

Covariance of two different vectors $x \in \mathcal{R}^n$ and $y \in \mathcal{R}^m$ can be defined similarly.

$$\mathrm{Cov}\{x, y\} = E\{(x - \bar{x})(y - \bar{y})^T\} \qquad (4.34)$$

In this case, $\text{Cov}\{x, y\}$ is an $n \times m$ matrix. In addition,

$$\text{Cov}\{x, y\} = (\text{Cov}\{y, x\})^T \tag{4.35}$$

Example: Gaussian Variables – 2-Dimensional Case

Let $d = [d_1 \ d_2]^T$ and

$$\begin{aligned} \mathcal{P}(\zeta; d) \ &= \ \frac{1}{2\pi\sigma^1\sigma_2(1-\rho^2)^{1/2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{\zeta_1 - m_1}{\sigma_1}\right)^2\right.\right. \\ &\left.\left. -2\rho\frac{(\zeta_1 - m_1)(\zeta_2 - m_2)}{\sigma_1\sigma_2} + \left(\frac{\zeta_2 - m_2}{\sigma_2}\right)^2\right]\right\} \end{aligned} \tag{4.36}$$

Then,

$$\begin{aligned} E\{d\} \ &= \ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix} \mathcal{P}(\zeta; d) \ \ d\zeta_1 d\zeta_2 \\ &= \ \begin{bmatrix} m_2 \\ m_2 \end{bmatrix} \end{aligned} \tag{4.37}$$

Similarly, one can show that

$$\begin{aligned} \text{Cov}\{d\} \ &= \ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \begin{bmatrix} \zeta_1 - m_1 \\ \zeta_2 - m_2 \end{bmatrix} \begin{bmatrix} (\zeta_1 - m_1) & (\zeta_2 - m_2) \end{bmatrix} \mathcal{P}(\zeta; d) \ \ d\zeta_1 d\zeta_2 \\ &= \ \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{bmatrix} \end{aligned} \tag{4.38}$$

Example: Gaussian Variables – n-Dimensional Case

Let $d = [d_1 \ \cdots \ d_n]^T$ and

$$\mathcal{P}(\zeta; d) = \frac{1}{(2\pi)^{\frac{n}{2}}|P_d|^{1/2}} \exp\left\{-\frac{1}{2}(\zeta - \bar{d})^T P_d^{-1}(\zeta - \bar{d})\right\} \tag{4.39}$$

Then, one can show that

$$E\{d\} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \zeta \mathcal{P}(\zeta; d) \;\; d\zeta_1, \cdots, d\zeta_n = \bar{d} \tag{4.40}$$

$$\mathrm{Cov}\{d\} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (\zeta - \bar{d})(\zeta - \bar{d})^T \mathcal{P}(\zeta; d) \;\; d\zeta_1, \cdots, d\zeta_n = P_d \tag{4.41}$$

Hence, $\bar{d}$ and P_d that parametrize the normal density function $\mathcal{P}(\zeta; d)$ represent the mean and the covariance matrix.

Exercise: Verify that, with

$$\bar{d} = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix}; \quad P_d = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{bmatrix} \tag{4.42}$$

one obtains the expression for normal density of a 2-dimensional vector shown earlier.

NOTE: Use of SVD for Visualization of Normal Density

Covariance matrix P_d contains information about the spread (i.e., extent of deviation from the mean) for each element and their correlations. For instance,

$$\mathrm{Var}\{d_\ell\} = [\mathrm{Cov}\{d\}]_{\ell,\ell} \tag{4.43}$$

$$\rho\{d_\ell, d_m\} = \frac{[\mathrm{Cov}\{d\}]_{\ell,m}}{\sqrt{[\mathrm{Cov}\{d\}]_{\ell,\ell}\,[\mathrm{Cov}\{d\}]_{m,m}}} \tag{4.44}$$

where $[\cdot]_{i,j}$ represents the $(i,j)_{\text{th}}$ element of the matrix. However, one still has hard time understanding the correlations among all the elements and visualizing the overall shape of the density function. Here, the SVD can be useful. Because P_d is a symmetric matrix, it has the following SVD:

$$P_d \triangleq E\{(d - \bar{d})(d - \bar{d})^T\} \tag{4.45}$$

$$= V \Sigma V^T \tag{4.46}$$

$$= \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix} \begin{bmatrix} v_1^T \\ \vdots \\ v_n^T \end{bmatrix} \tag{4.47}$$

Pre-multiplying V^T and post-multiplying V to both sides, we obtain

$$E\{V^T(d-\bar{d})(d-\bar{d})^T V\} = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix} \tag{4.48}$$

Let $d^* = V^T d$. Hence, d^* is the representation of d in terms of the coordinate system defined by orthonormal basis $v_1, \cdots, v_n$. Then, we see that

$$E\{(d^*-\bar{d^*})(d^*-\bar{d^*})^T\} = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix} \tag{4.49}$$

The diagonal covariance matrix means that every element of d^* is completely independent of each other. Hence, $v_1, \cdots, v_n$ define the coordiate system with respect to which the random variable vector is independent. $\sigma_1^2, \cdots, \sigma_n^2$ are the variances of d^* with respect to axes defined by $v_1, \cdots, v_n$.

Exercise: Suppose $d \in \mathcal{R}^2$ is zero-mean Gaussian and

$$P_d = \begin{bmatrix} 20.2 & 19.8 \\ 19.8 & 20.2 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} 10 & 0 \\ 0 & 0.1 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix} \tag{4.50}$$

Then, $v_1 = [\frac{\sqrt{2}}{2} \ \frac{\sqrt{2}}{2}]^T$ and $v_2 = [\frac{\sqrt{2}}{2} \ -\frac{\sqrt{2}}{2}]^T$. Can you visualize the overall shape of the density function? What is the variance of d along the (1,1) direction? What about along the (1,-1) direction? What do you think the conditional density of d_1 given $d_2 = \beta$ looks like? Plot the densities to verify.

4.2.5 CONDITIONAL PROBABILITY DENSITY: SCALAR CASE

When two random variables are related, the probability density of a random variable changes when the other random variable takes on a particular value.

> The probability density of a random variable when one or more other random variables are fixed is called *conditional probability density.*

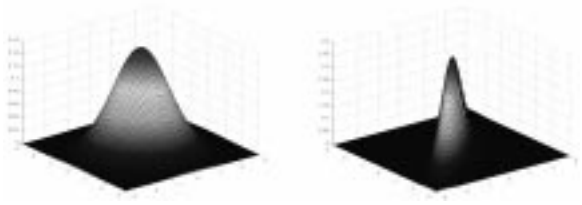
This concept is important in stochastic estimation as it can be used to develop estimates of unknown variables based on readings of other related variables.

Let x and y be random variables. Suppose $x$and y have joint probability density $\mathcal{P}(\zeta, \beta; x, y)$. One may then ask what the probability density of x is given a particular value of y (say $y = \beta$). Formally, this is called "conditional density function" of x *given* y and denoted as $\mathcal{P}(\zeta|\beta; x|y)$. $\mathcal{P}(\zeta|\beta; x|y)$ is computed as

$$\mathcal{P}(\zeta|\beta; x|y) = \frac{\lim_{\epsilon\to 0}\int_{\beta-\epsilon}^{\beta+\epsilon}\mathcal{P}(\zeta,\beta^*;x,y)d\beta^*}{\underbrace{\int_{-\infty}^{\infty}\int_{\beta-\epsilon}^{\beta+\epsilon}\mathcal{P}(\zeta,\beta^*;x,y)d\beta^* d\zeta}_{\text{normalization factor}}} \tag{4.51}$$

$$= \frac{\mathcal{P}(\zeta,\beta;x,y)}{\int_{-\infty}^{\infty}\mathcal{P}(\zeta,\beta;x,y)d\zeta} \tag{4.52}$$

$$= \frac{\mathcal{P}(\zeta,\beta;x,y)}{\mathcal{P}(\beta,y)} \tag{4.53}$$



Note:

- The above means

$$\left(\begin{array}{l} \text{Conditional Density} \\ \text{of } x \text{ given } y \end{array} \right) = \frac{\text{Joint Density of } x \text{ and } y}{\text{Marginal Density of } y} \tag{4.54}$$

  This should be quite intuitive.

- Due to the normalization,

$$\int_{-\infty}^{\infty} \mathcal{P}(\zeta|\beta; x|y) \;\; d\zeta = 1 \tag{4.55}$$

  which is what we want for a density function.

- $$\mathcal{P}(\zeta|\beta; x|y) = \mathcal{P}(\zeta, x) \tag{4.56}$$

  if and only if

$$\mathcal{P}(\zeta, \beta; x, y) = \mathcal{P}(\zeta, x)\mathcal{P}(\beta, y) \tag{4.57}$$

  This means that the conditional density is same as the marginal density when and only when x and y are independent.

We are interested in the conditional density, because often some of the random variables are measured while others are not. For a particular trial, if x is not measurable, but y is, we are intersted in knowing $\mathcal{P}(\zeta|\beta; x|y)$ for estimation of x.

Finally, note the distinctions among different density functions:

- $\mathcal{P}(\zeta, \beta; x, y)$: Joint Probability Density of x and y
  represents the probability density of $x = \zeta$ and $y = \beta$ simultaneously.

$$\int_{a_2}^{b_2} \int_{a_1}^{b_1} \mathcal{P}(\zeta, \beta; x, y) d\zeta d\beta = \Pr\{a_1 < x \leq b_1 \text{ and } a_2 < y \leq b_2\} \quad (4.58)$$

- $\mathcal{P}(\zeta; x)$: Marginal Probability Density of x
  represents the probability density of $x = \zeta$ NOT knowing what y is.

$$\mathcal{P}(\zeta, x) = \int_{-\infty}^{\infty} \mathcal{P}(\zeta, \beta; x, y) d\beta \quad (4.59)$$

- $\mathcal{P}(\beta; y)$: Marginal Probability Density of y
  represents the probability density of $y = \beta$ NOT knowing what x is.

$$\mathcal{P}(\beta, y) = \int_{-\infty}^{\infty} \mathcal{P}(\zeta, \beta; x, y) d\zeta \quad (4.60)$$

- $\mathcal{P}(\zeta|\beta; x|y)$: Conditional Probability Density of x given y
  represents the probability density of x when $y = \beta$.

$$\mathcal{P}(\zeta|\beta; x|y) = \frac{\mathcal{P}(\zeta, \beta; x, y)}{\mathcal{P}(\beta, y)} \quad (4.61)$$

- $\mathcal{P}(\beta|\zeta; y|x)$: Conditional Probability Density of y given x
  represents the probability density of y when $x = \zeta$.

$$\mathcal{P}(\beta|\zeta; y|x) = \frac{\mathcal{P}(\zeta, \beta; x, y)}{\mathcal{P}(\zeta, x)} \quad (4.62)$$

Baye's Rule:

Note that

$$\mathcal{P}(\zeta|\beta; x|y) = \frac{\mathcal{P}(\zeta, \beta; x, y)}{\mathcal{P}(\beta, y)} \quad (4.63)$$

$$\mathcal{P}(\beta|\zeta; y|x) = \frac{\mathcal{P}(\zeta, \beta; x, y)}{\mathcal{P}(\zeta, x)} \quad (4.64)$$

Hence, we arrive at

$$\mathcal{P}(\zeta|\beta;x|y) = \frac{\mathcal{P}(\beta|\zeta;y|x)\mathcal{P}(\zeta,x)}{\mathcal{P}(\beta,y)} \tag{4.65}$$

The above is known as the *Baye's Rule.* It essentially says

$$\begin{aligned} & (\text{Cond. Prob. of } x \text{ given } y) \times (\text{Marg. Prob. of } y) & (4.66)\\ = \; & (\text{Cond. Prob. of } y \text{ given } x) \times (\text{Marg. Prob. of } x) & (4.67) \end{aligned}$$

Baye's Rule is useful, since in many cases, we are trying to compute $\mathcal{P}(\zeta|\beta;x|y)$ and it's difficult to obtain the expression for it directly, while it may be easy to write down the expression for $\mathcal{P}(\beta|\zeta;y|x)$.

We can define the concepts of conditional expectation and conditional covariance using the conditional density. For instance, the conditional expectation of x given $y=\beta$ is defined as

$$E\{x|y\} \triangleq \int_{-\infty}^{\infty} \zeta \mathcal{P}(\zeta|\beta;x|y)d\zeta \tag{4.68}$$

Conditional variance can be defined as

$$\begin{aligned} \text{Var}\{x|y\} &\triangleq E\{(\zeta - E\{x|y\})^2\} & (4.69)\\ &= \int_{-\infty}^{\infty} (\zeta - E\{x|y\})^2 \mathcal{P}(\zeta|\beta;x|y)d\zeta & (4.70) \end{aligned}$$

Example: Jointly Normally Distributed or Gaussian Variables
Suppose that x and y have the following joint normal densities parametrized by $m_1, m_2, \sigma_1, \sigma_2, \rho$:

$$\begin{aligned} \mathcal{P}(\zeta,\beta;x,y) &= \frac{1}{2\pi\sigma^x\sigma_y(1-\rho^2)^{1/2}} & (4.71)\\ &\times \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{\zeta-\bar{x}}{\sigma_x}\right)^2 - 2\rho\frac{(\zeta-\bar{x})(\beta-\bar{y})}{\sigma_x\sigma_y} + \left(\frac{\beta-\bar{y}}{\sigma_y}\right)^2\right]\right\} \end{aligned}$$

Some algebra yields

$$\mathcal{P}(\zeta,\beta;x,y) = \underbrace{\frac{1}{\sqrt{2\pi\sigma_y^2}}\exp\left\{-\frac{1}{2}\left(\frac{\beta-\bar{y}}{\sigma_y}\right)^2\right\}}_{\text{marginal density of } y} \tag{4.72}$$

$$\times \underbrace{\frac{1}{\sqrt{2\pi\sigma_x^2(1-\rho^2)}}\exp\left\{-\frac{1}{2}\left(\frac{\zeta-\bar{x}-\rho\frac{\sigma_x}{\sigma_y}(\beta-\bar{y})}{\sigma_x\sqrt{1-\rho^2}}\right)^2\right\}}_{\text{conditional density of } x}$$

$$= \underbrace{\frac{1}{\sqrt{2\pi\sigma_x^2}}\exp\left\{-\frac{1}{2}\left(\frac{\zeta-\bar{x}}{\sigma_x}\right)^2\right\}}_{\text{marginal density of } x} \tag{4.73}$$

$$\times \underbrace{\frac{1}{\sqrt{2\pi\sigma_y^2(1-\rho^2)}}\exp\left\{-\frac{1}{2}\left(\frac{\beta-\bar{y}-\rho\frac{\sigma_y}{\sigma_x}(\zeta-\bar{x})}{\sigma_y\sqrt{1-\rho^2}}\right)^2\right\}}_{\text{conditional density of } y}$$

Hence,

$$\mathcal{P}(\zeta|\beta;x|y) = \frac{1}{\sqrt{2\pi\sigma_x^2(1-\rho^2)}}\exp\left\{-\frac{1}{2}\left(\frac{\zeta-\bar{x}-\rho\frac{\sigma_x}{\sigma_y}(\beta-\bar{y})}{\sigma_x\sqrt{1-\rho^2}}\right)^2\right\} \tag{4.74}$$

$$\mathcal{P}(\beta|\zeta;y|x) = \frac{1}{\sqrt{2\pi\sigma_y^2(1-\rho^2)}}\exp\left\{-\frac{1}{2}\left(\frac{\beta-\bar{y}-\rho\frac{\sigma_y}{\sigma_x}(\zeta-\bar{x})}{\sigma_y\sqrt{1-\rho^2}}\right)^2\right\} \tag{4.75}$$

Note that the above conditional densities are normal. For instance, $\mathcal{P}(\zeta|\beta;x|y)$ is a normal density with mean of $\bar{x}+\rho\frac{\sigma_x}{\sigma_y}(\beta-\bar{y})$ and variance of $\sigma_x^2(1-\rho^2)$. So,

$$E\{x|y\} = \bar{x}+\rho\frac{\sigma_x}{\sigma_y}(\beta-\bar{y}) \tag{4.76}$$

$$= \bar{x}+\frac{\rho\sigma_x\sigma_y}{\sigma_y^2}(\beta-\bar{y}) \tag{4.77}$$

$$= E\{x\}+\mathrm{Cov}\{x,y\}\mathrm{Var}^{-1}\{y\}(\beta-\bar{y}) \tag{4.78}$$

Conditional covariance of x given $y = \beta$ is:

$$\begin{aligned} E\{(x - E\{x|y\})^2|y\} &= \sigma_x^2(1-\rho^2) && (4.79)\\ &= \sigma_x^2 - \frac{\sigma_x^2\sigma_y^2\rho^2}{\sigma_y^2} && (4.80)\\ &= \sigma_x^2 - (\sigma_x\sigma_y\rho)\frac{1}{\sigma_y^2}(\sigma_x\sigma_y\rho) && (4.81)\\ &= \mathrm{Var}\{x\} - \mathrm{Cov}\{x,y\}\mathrm{Var}^{-1}\{y\}\mathrm{Cov}\{y,x\} && (4.82) \end{aligned}$$

Notice that the conditional distribution becomes a point density as $\rho \to 1$, which should be intuitively obvious.

4.2.6 CONDITIONAL PROBABILITY DENSITY: VECTOR CASE

We can extend the concept of conditional probability distribution to the vector case similarly as before.

Let x and y be n and m dimensional random vectors respectively. Then, the conditional density of x given $y = [\beta_1, \cdots, \beta_m]^T$ is defined as

$$\begin{aligned} &\mathcal{P}(\zeta_1, \cdots, \zeta_n|\beta_1, \cdots, \beta_m; x_1, \cdots, x_n|y_1, \cdots, y_m)\\ = \;& \frac{\mathcal{P}(\zeta_1, \cdots, \zeta_n, \beta_1, \cdots, \beta_m; x_1, \cdots, x_n, y_1, \cdots, y_m)}{\mathcal{P}(\beta_1, \cdots, \beta_m; y_1, \cdots, y_m)} \end{aligned} \quad (4.83)$$

Baye's Rule can be stated as

$$\begin{aligned} &\mathcal{P}(\zeta_1, \cdots, \zeta_n|\beta_1, \cdots, \beta_m; x_1, \cdots, x_n|y_1, \cdots, y_m) && (4.84)\\ = \;& \frac{\mathcal{P}(\beta_1, \cdots, \beta_m|\zeta_1, \cdots, \zeta_n; y_1, \cdots, y_m|x_1, \cdots, x_n)\mathcal{P}(\zeta_1, \cdots, \zeta_n; x_1, \cdots, x_n)}{\mathcal{P}(\beta_1, \cdots, \beta_m; y_1, \cdots, y_m)} \end{aligned}$$

The conditional expectation and covariance matrix can be defined similarly:

$$E\{x|y\} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \begin{bmatrix} \zeta_1 \\ \vdots \\ \zeta_n \end{bmatrix} \mathcal{P}(\zeta|\beta; x|y) \;\; d\zeta_1, \cdots, d\zeta_n \tag{4.85}$$

$$\mathrm{Cov}\{x|y\} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \begin{bmatrix} \zeta_1 - E\{x_1|y\} \\ \vdots \\ \zeta_n - E\{x_n|y\} \end{bmatrix} \begin{bmatrix} \zeta_1 - E\{x_1|y\} \\ \vdots \\ \zeta_n - E\{x_n|y\} \end{bmatrix}^T \mathcal{P}(\zeta|\beta; x|y) \;\; d\zeta_1, \cdots, d\zeta_n \tag{4.86}$$

Example: Gaussian or Jointly Normally Distributed Variables

Let x and y be jointly normally distributed random variable vectors of dimension n and m respectively. Let

$$z = \begin{bmatrix} x \\ y \end{bmatrix} \tag{4.87}$$

The joint distribution takes the form of

$$\mathcal{P}(\zeta, \beta; x, y) = \frac{1}{(2\pi)^{\frac{n+m}{2}} |P_z|^{1/2}} \exp\left\{ -\frac{1}{2} (\eta - \bar{z})^T P_z^{-1} (\eta - \bar{z}) \right\} \tag{4.88}$$

where

$$\bar{z} = \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}; \quad \eta = \begin{bmatrix} \zeta \\ \beta \end{bmatrix} \tag{4.89}$$

$$P_z = \begin{bmatrix} \mathrm{Cov}(x) & \mathrm{Cov}(x,y) \\ \mathrm{Cov}(y,x) & \mathrm{Cov}(y) \end{bmatrix} \tag{4.90}$$

Then, it can be proven that (see Theorem 2.13 in [Jaz70])

$$E\{x|y\} = \bar{x} + \mathrm{Cov}(x,y)\mathrm{Cov}^{-1}(y)(\beta - \bar{y}) \tag{4.91}$$

$$E\{y|x\} = \bar{y} + \mathrm{Cov}(y,x)\mathrm{Cov}^{-1}(x)(\zeta - \bar{x}) \tag{4.92}$$

and

$$\begin{aligned}
\mathrm{Cov}\{x|y\} &\triangleq E\left\{(\zeta - E\{x|y\})(\zeta - E\{x|y\})^T\right\} && (4.93)\\
&= \mathrm{Cov}\{x\} - \mathrm{Cov}\{x,y\}\mathrm{Cov}^{-1}\{y\}\mathrm{Cov}\{y,x\} && (4.94)\\
\mathrm{Cov}\{y|x\} &\triangleq E\left\{(\beta - E\{y|x\})(\beta - E\{y|x\})^T\right\} && (4.95)\\
&= \mathrm{Cov}\{y\} - \mathrm{Cov}\{y,x\}\mathrm{Cov}^{-1}\{x\}\mathrm{Cov}\{x,y\} && (4.96)
\end{aligned}$$

4.3 STATISTICS

4.3.1 PREDICTION

The first problem of statistics is prediction of the outcome of a future trial given a probabilistic model.

> Suppose $\mathcal{P}(x)$, the probability density for random variable x, is given. Predict the outcome of x for a new trial (which is about to occur).

Note that, unless $\mathcal{P}(x)$ is a point distribution, x cannot be predicted exactly.

To do optimal estimation, one must first establish a formal criterion. For example, the most likely value of x is the one that corresponds to the highest density value:

$$\hat{x} = \arg\left[\max_x \mathcal{P}(x)\right]$$

A more commonly used criterion is the following minimum variance estimate:

$$\hat{x} = \arg\left[\min_{\hat{x}} E\left\{\|x - \hat{x}\|_2^2\right\}\right]$$

The solution to the above is $\hat{x} = E\{x\}$.
Exercise: Can you prove the above?

If a related variable y (from the same trial) is given, then one should use $\hat{x} = E\{x|y\}$ instead.

4.3.2 SAMPLE MEAN AND COVARIANCE, PROBABILISTIC MODEL

The other problem of statistics is inferring a probabilistic model from collected data. The simplest of such problems is the following:

> We are given the data for random variable x from N trials. These data are labeled as $x(1), \cdots, x(N)$. Find the probability density function for x.

Often times, a certain density shape (like normal distribution) is assumed to make it a well-posed problem. If a normal density is assumed, the following sample averages can then be used as estimates for the mean and covariance:

$$\hat{\bar{x}} = \frac{1}{N}\sum_{i=1}^{N} x(i)$$

$$\hat{Rx} = \frac{1}{N}\sum_{i=1}^{N} x(i)x^T(i)$$

Note that the above estimates are *consistent* estimates of real mean and covariance $\bar{x}$ and R_x (i.e., they converge to true values as $N \to \infty$).

A slightly more general problem is:

> A random variable vector y is produced according to
>
> $$y = f(\theta, u) + x$$
>
> In the above, x is another random variable vector, u is a *known* deterministic vector (which can change from trial to trial) and θ is

an *unknown* deterministic vector (which is invariant). Given data for y from N trials, find the probability density parameters for x (e.g., $\bar{x}$, R_x) and the unknown deterministic vector θ.

This problem will be discussed later in the regression section.

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.

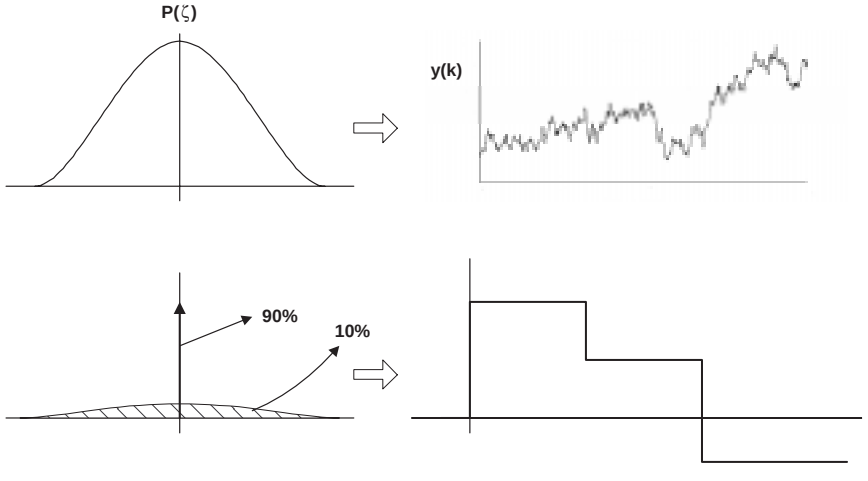
This result is very useful in modeling disturbances whose covariance functions are known or fixed. Note that a stationary Gaussian process is completely specified by its mean and covariance. Such a process can be modelled by filtering a zero-mean Gaussian white sequence through appropriate dynamics determined 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

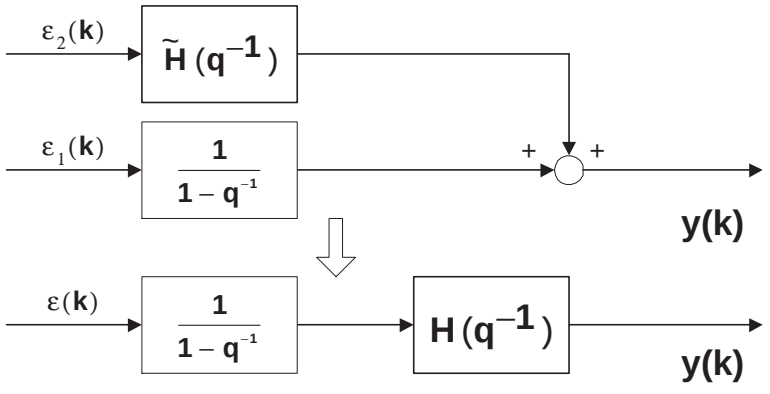
Some processes exhibit mean-shifts (whose magnitude and occurence are random). Consider the following model:

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

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:

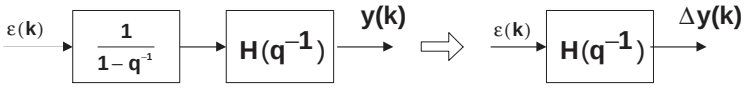


More generally, many interesting signals will exhibit stationary behavior combined with randomly occuring mean-shifts. Such signals can be modeled as



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

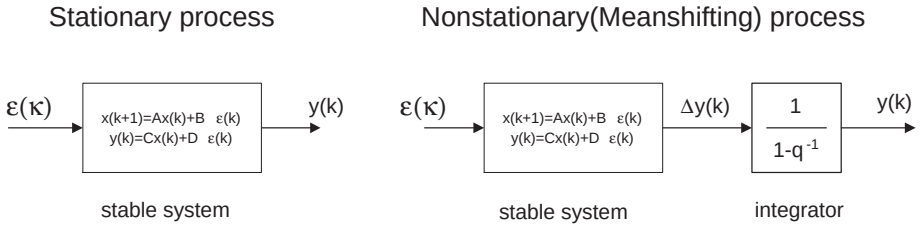
$$\Phi_y(\omega) = \left(C(e^{j\omega}I - A)^{-1}B + D\right) R_\varepsilon \left(C(e^{j\omega}I - A)^{-1}B + D\right)^T \tag{5.21}$$

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.

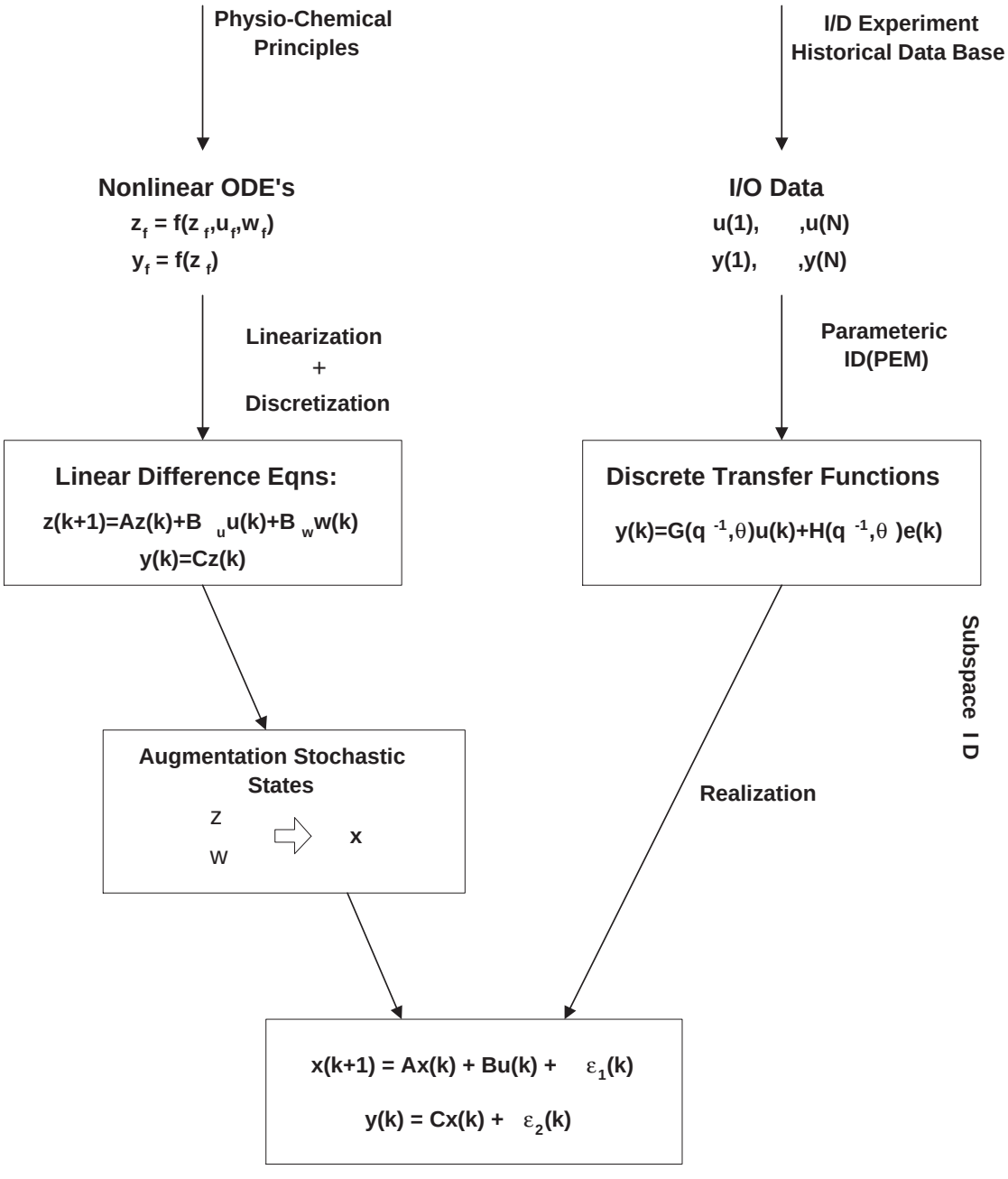
Details of these methods will be discussed in the identification chapter.

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

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

Chapter 6

STATE ESTIMATION

In practice, it is unrealistic to assume that all the disturbances and states can be measured. In general, one must *estimate* the states from the measured input / output sequences. This is called *state estimation.*

Let us assume the standard state-space system description we developed in the previous chapter:

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) + \varepsilon_1(k) \\ y(k) &= Cx(k) + \varepsilon_2(k) \end{aligned} \tag{6.1}$$

$\varepsilon_1(k)$ and $\varepsilon_2(k)$ are mutually independent white noise sequences of covariances R_1 and R_2 respectively. The problem of state estimation is to estimate $x(k+i),\ i \geq 0$, given $\{y(j), u(j), j \leq k\}$ (i.e., inputs and outputs up to the kth sample time). Estimating $x(k+i)$ for $i > 0$ is called *prediction*, while that for $i = 0$ is called *filtering.* Some applications require $x(k+i), i < 0$ to be estimated and this is referred to as *smoothing.*

There are many state estimation techniques, ranging from a simple open-loop observer to more sophisticated optimal observers like the Kalman filter. Since state estimation is an integral part of a model predictive controller, we examine some popular techniques in this chapter. These

techniques are also useful for parameter estimation problems, such as those arise in system identification discussed in the next chapter.

An extremely important, but often overlooked point is the importance of correct disturbance modelling. Simply adding white noises into the state and output equations, as often done by those who misunderstand the role of white noise in a standard system description, can result in extreme bias. In general, to obtain satisfactory results, disturbances (or their effects) must be modelled as appropriate stationary / nonstationary stochastic processes and the system equations must be augmented with their describing stochastic equations before a state estimation technique is applied.

6.1 LINEAR OBSERVER STRUCTURE

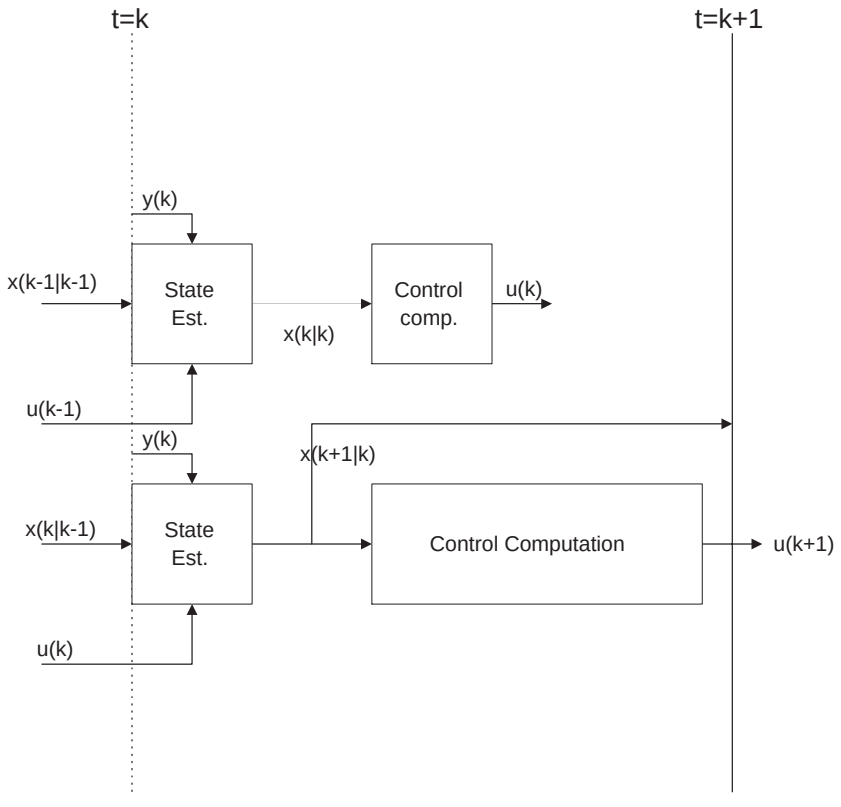
A linear observer for system (6.1) takes the form of

$$\begin{aligned} \hat{x}(k|k-1) &= A\hat{x}(k-1|k-1) + Bu(k-1) \\ \hat{x}(k|k) &= \hat{x}(k|k-1) + K(y(k) - \hat{x}(k|k-1)) \end{aligned} \tag{6.2}$$

In the above, $x(i|j)$ represents an estimate of $x(i)$ constructed using measurements up to time j. The above equations can be used to construct the filtered estimate $\hat{x}(k|k)$ recursively.

Comments:

- In some applications, one may need to compute the one-step-ahead prediction $\hat{x}(k+1|k)$ rather than the filtered estimate. For instance, in a control application, the control computation may require one sample period to complete and in this case, one may want to compute $\hat{x}(k+1|k)$ at time k in order to begin the computation for the control input $u(k+1)$.



Notice that (6.2) can be rewritten as a one-step-ahead predictor simply by switching the order of the two equations:

$$\begin{aligned} \hat{x}(k|k) &= \hat{x}(k|k-1) + K(y(k) - \hat{x}(k|k-1)) \\ \hat{x}(k+1|k) &= A\hat{x}(k|k) + Bu(k) \end{aligned} \tag{6.3}$$

- The free parameter in the above is K, which is called the *observer gain matrix*. What remains to be discussed is how to choose K. In general, it should be chosen so that the estimation error ($x_e(k) \stackrel{\Delta}{=} x(k) - \hat{x}(k|k)$ or $\hat{x}_e(k+1) \stackrel{\Delta}{=} x(k+1) - \hat{x}(k+1|k)$) is minimized in some sense.

- Equations for error dynamics can be easily derived. For instance, the equations for the filter estimation error is

$$x_e(k) = (A - KCA)x_e(k-1) + (I - KC)\varepsilon_1(k-1) + K\varepsilon_2(k) \tag{6.4}$$

The above can be derived straightforwardly by replacing $y(k)$ in the observer equation (6.2) with $CAx(k-1) + CBu(k-1) + C\varepsilon_1(k-1)$. The equation for prediction error $\hat{x}_e(k)$ can be derived similarly as

$$\hat{x}_e(k+1) = (A - AKC)\hat{x}_e(k) + \varepsilon_1(k-1) + AK\varepsilon_2(k) \tag{6.5}$$

- In some cases, it is advantageous to allow K to vary with time. This results in a time varying observer.

6.2 POLE PLACEMENT

From (6.4), it is clear that the eigenvalues of the transition matrix $A - KCA$ determine how the estimation error propagates. For instance, one must take care that all the eigenvalues lie strictly inside the unit circle in order to ensure stable error dynamics (i.e., asymptotically vanishing initialization error, finite error variance, etc.). The eigenvalues of $A - KCA$ are called observer poles and determining K on the basis of prespecified observer pole location is called *pole placement*. For instance, if (C, A) is an observable pair, the observer poles can be placed in an arbitrary manner through K.

One can also work with the one-step-ahead prediction error equation (6.5). In this case one can let $AK = \hat{K}$ and determine $\hat{K}$ so that the eigenvalues of $A - \hat{K}C$ are placed at desired locations. Again, with an observer system, the eigenvalues can be placed at arbitrary locations.

Pole placement is most coveniently carried out by first putting the system in an observer canonical form through an appropriate coordinate transformation (given by the observability matrix). For instance, consider the following observer canonical form for a single-input, single-output

system:

$$\begin{aligned} x(k+1) &= \begin{bmatrix} -a_1 & 1 & 0 & \cdots & 0 \\ -a_2 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ -a_{n-1} & 0 & 0 & \cdots & 1 \\ -a_n & 0 & 0 & \cdots & 0 \end{bmatrix} x(k) + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_{n-1} \\ b_n \end{bmatrix} u(k) \\ y(k) &= \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \end{bmatrix} x(k) \end{aligned} \tag{6.6}$$

Then, assuming $K = [k_1 \; k_2 \; \cdots \; k_{n-1} \; k_n]^T$, we have

$$A - KC = \begin{bmatrix} -(a_1 + k_1) & 1 & 0 & \cdots & 0 \\ -(a_2 + k_2) & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ -(a_{n-1} + k_{n-1}) & 0 & 0 & \cdots & 1 \\ -(a_n + k_n) & 0 & 0 & \cdots & 0 \end{bmatrix} \tag{6.7}$$

The characteristic polynomial for the above matrix is

$$z^n + (a_1 + k_1)z^{n-1} + \cdots + (a_{n-1} + k_{n-1})z + (a_n + k_n) = 0 \tag{6.8}$$

Hence, $k_1, \cdots, k_n$ can be easily determined to place the roots at desired locations.

6.3 KALMAN FILTER

An observer gain can also be determined from a stochastic optimal estimation viewpoint. For example, the observer gain for the linear observer structure can be chosen to minimize the variance of the estimation error. The resulting estimator is the celebrated *Kalman filter*, which has by far been the most popular state estimation technique. When the additional assumption is made that the disturbances are Gaussian, the Kalman filter is

indeed the optimal estimator (not just the optimal linear estimator).

6.3.1 KALMAN FILTER AS THE OPTIMAL LINEAR OBSERVER

Note that the the linear observer (6.3) can be written in the following one-step-ahead predictor form:

$$\hat{x}(k+1|k) = A\hat{x}(k|k-1) + Bu(k) + \underbrace{AK(k)}_{\hat{K}(k)}\{y(k) - C\hat{x}(k|k-1)\} \tag{6.9}$$

In the above, we allowed the observer gain to vary with time for generality.

Recall that the error dynamics for $\hat{x}_e(k) = x(k) - \hat{x}(k|k-1)$ are given by

$$\hat{x}_e(k+1) = (A - \hat{K}(k)C)\hat{x}_e(k) + \varepsilon_1(k) + \hat{K}(k)\varepsilon_2(k) \tag{6.10}$$

Let

$$\begin{aligned} P(k) &= \text{Cov}\{\hat{x}_e(k)\} && (6.11)\\ &= E\left\{(\hat{x}_e(k) - E\{\hat{x}_e(k)\})\,(\hat{x}_e(k) - E\{\hat{x}_e(k)\})^T\right\} && (6.12) \end{aligned}$$

Assuming that the initial guess is chosen so that $E\{\hat{x}_e(0)\} = 0$, $E\{\hat{x}_e(k)\} = 0$ for all $k \geq 0$ and

$$\begin{aligned} P(k+1) &= \left\{\hat{x}_e(k+1)\hat{x}_e^T(k+1)\right\} && (6.13)\\ &= (A - \hat{K}(k)C)P(k)(A - \hat{K}(k)C)^T + R_1 - \hat{K}(k)R_2\hat{K}^T(k) \end{aligned}$$

In the above, we used the fact that $\hat{x}_e(k), \varepsilon_1(k)$ and $\varepsilon_2(k)$ in (6.10) are mutually independent.

Now let us choose $K(k)$ such that $\alpha^T P(k+1)\alpha$ is minimized for an arbitrary choice of α. Since α is an arbitrary vector, this choice of $K(k)$ minimizes

the expected value of any norm of $\hat{x}_e$ (including the 2-norm which represents the error variance). Now, it is straightforward algebra to show that

$$\begin{aligned}\alpha^T P(k+1)\alpha &= \alpha^T \Big[AP(k)A^T + R_1 - \hat{K}(k)CP(k)A^T \\ &\quad -AP(k)C^T\hat{K}^T(k) + \hat{K}(k)(R_2 + CP(k)C^T)\hat{K}^T(k)\Big]\alpha\end{aligned} \tag{6.14}$$

Completing the square on the terms involving $\hat{K}(k)$, we obtain

$$\begin{aligned}\alpha^T P(k+1)\alpha &= \alpha^T \Big\{\Big[\hat{K}(k) - AP(k)C^T(R_2 + CP(k)C^T)^{-1}\Big]\left[R_2 + CP(k)C^T\right] \\ &\quad \times \Big[\hat{K}(k) - AP(k)C^T(R_2 + CP(k)C^T)^{-1}\Big]^T\Big\}\alpha \\ &\quad +\alpha^T\left[AP(k)A^T + R_1 - AP(k)C^T(R_2 + CP(k)C^T)^{-1}CP(k)A^T\right]\alpha\end{aligned} \tag{6.15}$$

Hence, $\hat{K}(k)$ minimizing the above is

$$\hat{K}(k) = AP(k)C^T(R_2 + CP(k)C^T)^{-1} \tag{6.16}$$

and

$$P(k+1) = AP(k)A^T + R_1 - AP(k)C^T\left(R_2 + CP(k)C^T\right)^{-1}CP(k)A^T \tag{6.17}$$

Given $x(1|0)$ and $P(1)$, the above equations can be used along with (6.9) to recursively compute $\hat{x}(k+1|k)$. They are referred to as the *time-varying* Kalman filter equations.

Note:

- For detectable systems, it can be shown that $P(k)$ converges to a constant matrix $\bar{P}$ as $K \to \infty$. Hence, for linear time-invariant systems, it is customary to implement an observer with a constant gain matrix derived from $\bar{P}$ according to (6.16). This is referred to as the *steady-state* Kalman filter.

- Also, recall the relationship between the one-step-ahead predictor gain $\hat{K}(k)$ and the filter gain $K(k)$ ($\hat{K}(k) = AK(k)$). Hence,

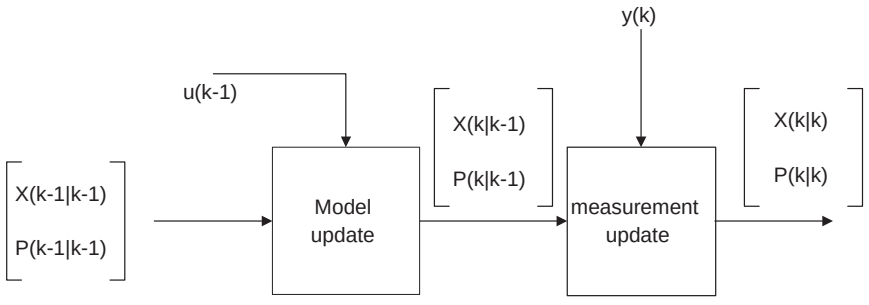
$$K(k) = P(k)C^T(R_2 + CP(k)C^T)^{-1} \tag{6.18}$$

  This gain can be used to implement a filter of the form (6.2) that recursively computes $\hat{x}(k|k)$ rather than $\hat{x}(k+1|k)$.

6.3.2 KALMAN FILTER AS THE OPTIMAL ESTIMATOR FOR GAUSSIAN SYSTEMS

In the previous section, we assumed a linear observer structure and posed the problem as a parametric optimization where the expected value of the estimation error variance is minimized with respect to the observer gain. In fact, the Kalman filter can be derived from an entirely probabilistic argument, i.e., by deriving a Bayesian estimator that recursively computes the conditional density of $x(k)$.

Assume that $\varepsilon_1(k)$ and $\varepsilon_2(k)$ are Gaussian noise sequences. Then, assuming $x(0)$ is also a Gaussian variable, $x(k)$ and $y(k)$ are jointly-Gaussian sequences. Now we can simply formulate the state estimation problem as computing the conditional expectation $E\{x(k) \mid Y(k)\}$ where $Y(k) = [y^T(0),\ y^T(1),\ \cdots, y^T(k)]^T$. Let us denote $E\{x(i) \mid Y(j)\}$ as $x(i|j)$. We divide the estimation into the following two steps.



- **Model Update:** *Compute* $E\{x(k)|Y(k-1)\}$ *given* $E\{x(k-1)|Y(k-1)\}$, $P(k-1|k-1)$, *and* $u(k-1)$.
  Since $x(k) = Ax(k-1) + Bu(k-1) + \varepsilon_1(k-1)$ and $\varepsilon_1(k-1)$ is a zero-mean variable independent of $y(0), \cdots, y(k-1)$,

$$\begin{aligned} \hat{x}(k|k-1) &= E\left\{Ax(k-1) + Bu(k-1) + e(k-1) \mid Y(k-1)\right\} \\ &= AE\{x(k-1) \mid Y(k-1)\} + Bu(k-1) \end{aligned} \tag{6.19}$$

  Hence, we obtain

$$\hat{x}(k|k-1) = A\hat{x}(k-1|k-1) + Bu(k-1) \tag{6.20}$$

  In addition, note that

$$x(k) - \hat{x}(k|k-1) = A\left(x(k) - \hat{x}(k-1|k-1)\right) + \varepsilon_1(k-1) \tag{6.21}$$

  Therefore,

$$P(k|k-1) = E\left\{\left(x(k) - \hat{x}(k|k-1)\right)\left(x(k) - \hat{x}(k|k-1)\right)^T\right\} \tag{6.22}$$

$$= AP(k-1|k-1)A^T + R_1 \tag{6.23}$$

  Since the conditional density for $x(k)$ given $Y(k-1)$ is Gaussian, it is completely specified by $\hat{x}(k|k-1)$ and $P(k|k-1)$.

- **Measurement Update:** *Compute* $E\{x(k)|Y(k)\}$ *given* $E\{x(k)|Y(k-1)\}$ $P(k|k-1)$ *and* $y(k)$.

The conditional density $\mathcal{P}\{x(k) \mid Y(k)\}$ is equivalent to the conditional density $\mathcal{P}\{x(k) \mid y(k)\}$ with the prior density of $x(k)$ given instead by $\mathcal{P}\{x(k) \mid Y(k-1)\}$. Note that $\mathcal{P}\{x(k) \mid Y(k-1)\}$ is a Gaussian density of mean $\hat{x}(k|k-1)$ and covariance $P(k|k-1)$. In other words, we view $x(k)$ as a Gaussian variable of mean $\hat{x}(k|k-1)$ and covariance $P(k|k-1)$.

In addition, $y(k) = Cx(k) + \varepsilon_2(k)$ and therefore is also Gaussian.

$$E\{y(k)\} = CE\{x(k)\} + E\{\varepsilon(k)\} = C\hat{x}(k|k-1)$$

$$E\left\{\left(y(k) - E\{y(k)\}\right)\left(y(k) - E\{y(k)\}\right)^T\right\} = CP(k|k-1)C^T + R_2$$

In fact, $x(k)$ and $y(k)$ are jointly Gaussian with the following covariance:

$$\begin{aligned} & E\left\{\begin{bmatrix} x(k) - \hat{x}(k|k-1) \\ y(k) - y(k|k-1) \end{bmatrix}\begin{bmatrix} x(k) - \hat{x}(k|k-1) \\ y(k) - y(k|k-1) \end{bmatrix}^T\right\} \\ = & \begin{bmatrix} P(k|k-1) & P(k|k-1)C^T \\ CP(k|k-1) & CP(k|k-1)C^T + R_2 \end{bmatrix} \end{aligned} \tag{6.24}$$

Recall the earlier results for jointly Gaussian variables:

$$E\{x|y\} = E\{x\} + R_{xy}R_y^{-1}(y - E\{y\}) \tag{6.25}$$

$$\mathrm{Cov}\{x|y\} = R_x - R_{xy}R_y^{-1}R_{yx} \tag{6.26}$$

Applying the above to $x(k)$ and $y(k)$,

$$\begin{aligned} \hat{x}(k|k) &= E\{x(k)|y(k)\} \\ &= \hat{x}(k|k-1) \\ &\quad + P(k|k-1)C^T\left(CP(k|k-1)C^T + R_2\right)^{-1}\left(y(k) - C\hat{x}(k|k-1)\right) \\ P(k|k) &= \mathrm{Cov}\{x(k)|y(k)\} \\ &= P(k|k-1) - P(k|k-1)C^T\left(CP(k|k-1)C^T + R_2\right)^{-1}CP(k|k-1) \end{aligned} \tag{6.27}$$

In short, for Gaussian systems, we can compute the conditional mean and covariance of $x(k)$ recursively using

$$\hat{x}(k|k-1) = A\hat{x}(k-1|k-1) + Bu(k-1)$$

$$\hat{x}(k|k) = \hat{x}(k|k-1) + \underbrace{P(k|k-1)C^T\left(CP(k|k-1)C^T + R_2\right)^{-1}}_{K(k)}(y(k) - C\hat{x}(k|k-1))$$

and

$$P(k|k-1) = AP(k-1|k-1)A^T + R_1$$

$$P(k|k) = P(k|k-1) - P(k|k-1)C^T\left(CP(k|k-1)C^T + R_2\right)^{-1}CP(k|k-1)$$

Note that this above has a linear observer structure with the observer gain given by the Kalman filter equations derived earlier ($P(k|k-1)$ in the above is $P(k)$ in Eq. (6.16)–(6.17)).

Chapter 7

SYSTEM IDENTIFICATION

Identification of process dynamics is perhaps the most time consuming step in implementing an MPC and one that requires relatively high expertise from the user. In this section, we give introduction to various identification methods and touch upon some key issues that should help an engineer obtain models on a reliable basis. Since system identification is a very broad subject that can easily take up an entire book, we will limit our objective to giving just an overview and providing a starting point for further exploration of the field. Because of this, our treatment of various methods and issues will necessarily be brief and informal. References will be given at the end for more complete, detailed treatments of the various topics presented in this chapter.

7.1 PROBLEM OVERVIEW

The goal of identification is to obtain a mathematical relation that reliably predicts the behavior of outputs, using input output data gathered from the process. For conveninence, the mathematical relation searched for is often limited to linear ones. As we saw in the previous chapter, both known and

unknown inputs affect the outputs. Since many inputs change in a random, but correlated manner, it is often desirable to identify a model that has both deterministic and stochastic components.

In terms of how input output data are translated into a mathematical relation, the field of identification can be divided broadly into two branches: parametric identification and nonparametric identification. In parameteric identification, the structure of the mathematical relation is fixed *a priori* and parameters of the structure are fitted to the data. In nonparametric identification, no (or very little) assumption is made with respect to the model structure. Frequency response identification is nonparametric. Impulse response identification is also nonparametric, but it can also be viewed as parametric identification since a impulse response of a finite length is often identified.

As a final note, it is important not to forget the end-use of the model, which is to analyze and design a feedback control system in our case. Accuracy of a model must ultimately be judged in view of how well the model predicts the output behavior with the intended feedback control system in place. This consideration must be reflected in all phases of identification including test input design, data filtering, model structure selection, and parameter estimation.

7.2 PARAMETRIC IDENTIFICATION METHODS

In parametric identification, the model structure is set in prior to the model fitting. The key objective is then to identify the model parameters, based on given input output data. Although a particular model structure is assumed for parameter estimation, one often adjusts the model structure iteratively based on the result of fitting (for example, through residual analysis).

7.2.1 MODEL STRUCTURES

A general structure for parametric identification is:

$$y(k) = G(q,\theta)u(k) + H(q,\theta)\varepsilon(k) \tag{7.1}$$

where y is the output and u is the input (most of times, this will be a manipulated input, but it can also be a measured disturbance variable). For systems with stationary disturbances, $\varepsilon(k)$ can be assumed to be white noise and $H(q,\theta)$ a stable, stably invertible and normalized (*i.e.*, $H(\infty,\theta) = 1$) transfer function, without loss of generality. In the case that the disturbance is better described by a stochastic process driven by *integrated* white noise, we can replace $y(k)$ and $u(k)$ with $\Delta y(k)$ and $\Delta u(k)$.

Within the general structure, different parametrizations exist. Let us discuss some popular ones, first in the single input, single output context.

- *ARX Model* If we represent G as a rational function and express it as a linear equation with an additive error term, we obtain

$$y(k)+a_1y(k-1)+\cdots+a_ny(k-n) = b_1u(k-1)+\cdots+b_mu(k-m)+\varepsilon(k) \tag{7.2}$$

  When the equation error $\varepsilon(k)$ is taken as a white noise sequence, the resulting model is called an ARX model (AR for Auto-Regressive and X for eXtra input u). Hence, the ARX model corresponds to the following parametrization of the transfer functions:

$$\begin{aligned} G(q,\theta) &= \frac{B(q)}{A(q)} \triangleq \frac{b_1q^{-1}+\cdots+b_mq^{-m}}{1+a_1q^{-1}+\cdots+a_nq^{-n}} \\ H(q,\theta) &= \frac{1}{A(q)} \triangleq \frac{1}{1+a_1q^{-1}+\cdots+a_nq^{-n}} \end{aligned} \tag{7.3}$$

  A high-order ARX model is a good choice when the system order is

unknown. To see this, note that (7.1) can be written as

$$H^{-1}(q,\theta)y(k) = H^{-1}(q,\theta)G(q,\theta)u(k) + \varepsilon(k) \tag{7.4}$$

Since H^{-1} is assumed stable, if G is stable,

$$\begin{aligned} H^{-1}(q,\theta) &\approx 1 + a_1 q^{-1} + \cdots + a_n q^{-n} \\ H^{-1}(q,\theta)G(q,\theta) &\approx 1 + b_1 q^{-1} + \cdots + b_m q^{-m} \end{aligned} \tag{7.5}$$

for sufficiently large n and m.

- *ARMAX Model* A natural extension to the ARX parametrization is the ARMAX model, which expresses the equation error term as a moving average of white noise:

$$\begin{aligned} & y(k) + a_1 y(k-1) + \cdots + a_n y(k-n) \\ = \; & b_1 u(k-1) + \cdots + b_m u(k-m) \\ & + \varepsilon(k) + c_1 \varepsilon(k-1) + \cdots + c_\ell \varepsilon(k-\ell) \end{aligned} \tag{7.6}$$

  For the ARMAX model, the parametrization of the noise transfer function changes to

$$H(q,\theta) = \frac{C(q)}{A(q)} \triangleq \frac{1 + c_1 q^{-1} + \cdots + c_\ell q^{-\ell}}{1 + a_1 q^{-1} + \cdots + a_n q^{-n}} \tag{7.7}$$

  Because of the moving average term, an ARMAX model can potentially represent a system with much fewer parameters when compared to an ARX model. In fact, a state-space system of order n always have an input output representation given by an n_{th} order ARMAX model. However, parameter estimation is more complicated and over-parametrization can cause loss of identifiability (i.e., the parameter values can become nonunique).

- *Output Error Model* Both ARX and ARMAX model puts common poles on G and H. In some cases, it may be more natural to model

them separately. One such parametrization is the Output Error (OE) model given below:

$$\begin{aligned} \tilde{y}(k) + a_1\tilde{y}(k-1) + \cdots + a_n\tilde{y}(k-n) &= b_1u(k-1) + \cdots + b_mu(k-m) \\ y(k) &= \tilde{y}(k) + \varepsilon(k) \end{aligned} \tag{7.8}$$

In the above $\tilde{y}(k)$ represents the disturbance-free output. Customarily, $\varepsilon(k)$ is assumed to be white noise. This means the OE structure gives

$$G(q,\theta) = \frac{A(q)}{B(q)} \quad \text{and} \quad H(q) = 1 \tag{7.9}$$

A slightly more general case is when $H(q)$ is not 1, but completely known (i.e., disturbance is a colored noise with known spectrum). In this case, we can write

$$\underbrace{H^{-1}(q)y(k)}_{y_f(k)} = G(q,\theta)\underbrace{H^{-1}u(k)}_{u_f(k)} + \varepsilon(k) \tag{7.10}$$

Note that the above is in the form of (7.8). Simple prefiltering of input and output decorrelates the noise and gives the standard OE structure. Parameter estimtion is complicated by the fact that $\tilde{y}$'s are not known, and depend on the parameters.

- *FIR and Orthogonal Expansion Model* A special kind of output error structure is obtained when $G(q,\theta)$ is parametrized linearly. For instance, when $G(q)$ is stable, it can be expanded as a power series of q^{-1}. One obtains

$$G(q) = \sum_{i=1}^{\infty} b_i q^{-i} \tag{7.11}$$

Truncating the power series after n-terms, one obtains the model

$$y(k) = \left(b_1q^{-1} + b_2q^{-2} + \cdots + b_nq^{-n}\right)u(k) + H(q)\varepsilon(k) \tag{7.12}$$

This is the Finite Impulse Response model that we used in the basic

part of this book.

A general form of an orthogonal expansion model is

$$G(q) = \sum_{i=1}^{\infty} b_i B_i(q) \tag{7.13}$$

One of the popular choices for $\{B_i(q)\}$ is the so called Laguere functions defined as

$$B_i(q) = \frac{\sqrt{1-\alpha^2}}{q-\alpha}\left(\frac{1-\alpha q}{q-\alpha}\right)^{i-1} \tag{7.14}$$

An advantage of using this function is that the knowledge of process's dominant time constant can be incorporated into the choice of α to speed up the convergence (since it helps curtail the number of parameters).

- *Box-Jenkins Model* A natural generalization of the output error model is to let the disturbance transfer function be a rational function of unknown parameters. This leads to the Box-Jenkins model which has the structure of

$$y(k) = \frac{B(q)}{A(q)}u(k) + \frac{C(q)}{D(q)}\varepsilon(k) \tag{7.15}$$

This model structure is quite general, but the parameter estimation is nonlinear and loss of identifiability can occur.

All of the above models can be generalized to the case where $H(q,\theta)$ contains an integrator. For instance, we can extend the ARMAX model to

$$y(k) = \frac{B(q)}{A(q)}u(k) + \frac{1}{1-q^{-1}}\frac{C(q)}{A(q)}\varepsilon(k) \tag{7.16}$$

The above is called ARIMAX model (I for integration). In terms of parameter estimation, the resulting problem is the same since we can

transform the above to

$$\Delta y(k) = \frac{B(q)}{A(q)}\Delta u(k) + \frac{C(q)}{A(q)}\varepsilon(k) \tag{7.17}$$

The same holds for all the other model structues.

Extensions to the multivariable case are mostly straightforward, but can involve some complications. One may choose to fit each output independently using one of the above structures (This is called "MISO identification"). In this case, the only modification to the above is that $B(q)$ is now a row vector containing n_u polynomials, where n_u is the number of inputs. The parameter estimation problem remains e the same except in the number of parameters. On the other hand, some applications require a model that capture disturbance correlations among different outputs. This requires MIMO identification where all the outputs are fitted to a single multivariable model on a simultaneous basis. In this case, $A(q), B(q)$, etc. are matrix polynomials of appropriate dimension. For instance, the ARX model becomes

$$y(k) = A^{-1}(q)B(q)u(k) + A^{-1}(q)\varepsilon(k) \tag{7.18}$$

where $A(q)$ and $B(q)$ are $n_y \times n_y$ and $n_y \times n_u$ matrix polynomials respectively. The parameterization of these matrices can be a subtle issue. For instance, if all matrix entries are assumed to be unknown, one can easily lose identifiability. In general, significant prior knowledge is needed to obtain a correct parameterization. In addition, parameter estimation can be numerically challenging due to the large number of parameters, especially when the model structure leads to a nonlinear estimation problem.

7.2.2 PARAMETER ESTIMATION VIA PREDICTION ERROR MINIMIZATION

7.2.2.1 Prediction Error Method

The optimal one-step ahead predictor based on the model (7.1) can be written as

$$y(k|k-1) = G(q,\theta)u(k) + \left(I - H^{-1}(q,\theta)\right)\left(y(k) - G(q,\theta)u(k)\right) \qquad (7.19)$$

By comparing (7.1) with (7.19), we see that the prediction error $(y(k) - y(k|k-1))$ is simply $\varepsilon(k)$, assuming that the model is perfect. Note that $\left(I - H^{-1}(q,\theta)\right)$ contains at least one delay since $I - H^{-1}(\infty,\theta) = 0$. Hence, the right hand side does not require $y(k)$ to be known.

Because the primary function of a model in control is to provide a prediction of the future output behavior, it is logical to choose θ such that the prediction error resulting from the model is minimized for the available data record. Let us denote the data record we have as $(\hat{y}(1), \cdots, \hat{y}_N)$. Then, this objective is formulated as

$$\min_{\theta} \sum_{k=1}^{N} \|\hat{e}_{pred}(k,\theta)\|_2^2 \qquad (7.20)$$

where $\hat{e}_{pred}(k,\theta) = \hat{y}(k) - y(k|k-1)$, and $\|\cdot\|_2$ denotes the Euclidean norm. Use of other norms are possible, but the 2-norm is by far the most popular choice. Using (7.19), we can write

$$\hat{e}_{pred}(k,\theta) = H^{-1}(q,\theta)\left(\hat{y}(k) - G(q,\theta)u(k)\right) \qquad (7.21)$$

For certain model structures, the 2-norm minimization of prediction error is

formulated as a linear least-squares problem. For example, for the ARX structure, $G(q,\theta) = \frac{B(q)}{A(q)}$, and $H(q,\theta) = \frac{1}{A(q)}$ and

$$\begin{aligned} \hat{e}_{pred}(k,\theta) &= A(q)\hat{y}(k) - B(q)u(k) \\ &= \hat{y}(k) + a_1\hat{y}(k-1) + \cdots + a_n\hat{y}(k-n) - b_1u(k-1) - \cdots - b_mu(k-m) \end{aligned} \tag{7.22}$$

Since $\hat{e}_{pred}(k,\theta)$ is linear with respect to the unknown parameters, the minimization of
$\sum_{k=1}^{N} \hat{e}^2_{pred}(k,\theta)$ is a linear least squares problem.

Another such example is an FIR model with known disturbance characteristics for which $G(q,\theta) = \sum_{i=1}^{n} h_i q^{-i}$ and $H(q)$ contains no unknown parameters. In this case

$$\hat{e}_{pred}(k,\theta) = \hat{y}_f(k) - h_1u_f(k-1) - \cdots - h_nu_f(k-n) \tag{7.23}$$

where $\hat{y}_f(k) = H^{-1}(q)\hat{y}(k)$ and $u_f(k) = H^{-1}(q)u(k)$. Again, the expression is linear in the unknowns and the prediction error minimization (PEM) is a linear least squares problem. If the noise model was $\frac{1}{1-q^{-1}}H(q)$, then $\hat{y}_f(k)$ and $u_f(k)$ should be redefined as $H^{-1}(q)\Delta\hat{y}(k)$ and $H^{-1}(q)\Delta u(k)$ respectively. The same observation applies to Laguerre or other orthogonal expansion models.

PEM for other model structures such as the ARMAX and Box-Jenkins structures is not a linear least squares problem and pseudo-linear regression is often used for them.

7.2.2.2 Properties of Linear Least Squares Identification

We saw that prediction error minimization for many model structures can be cast as a linear regression problem. The general linear regression problem can be written as

$$\hat{y}(k) = \phi^T(k)\theta + e(k,\theta) \tag{7.24}$$

where $\hat{y}$ is the observed output (or filtered output), ϕ is the regressor vector, θ is the parameter vector to be identified, and e the residual error (that depends on the choice of θ). $\{\cdot\}(k)$ denotes the k_{th} sample. In the least squares identification, θ is found such that the sum of squares of the residuals is minimized, i.e., $\theta_N^{LS} = \arg\left\{\min_\theta \sum_{k=1}^N e^2(k,\theta)\right\}$. We saw in the previous section that 2-norm minimization of prediction error for certain model structures can be cast in this form.

For a data set collected over N sample intervals, (7.24) can be written collectively as the following set of linear equations:

$$\hat{Y}_N = \Phi_N\theta + E_N \tag{7.25}$$

where

$$\Phi_N = \left[\begin{array}{ccc} \phi(1) & \cdots & \phi(N) \end{array}\right]^T \tag{7.26}$$

$$\hat{Y}_N = \left[\begin{array}{ccc} \hat{y}(1) & \cdots & \hat{y}(N) \end{array}\right]^T \tag{7.27}$$

$$E_N = \left[\begin{array}{ccc} e(1) & \cdots & e(N) \end{array}\right]^T \tag{7.28}$$

The least squares solution is

$$\hat{\theta}_N^{LS} = (\Phi_N^T\Phi_N)^{-1}\Phi_N^T Y_N \tag{7.29}$$

Convergence

Let us assume that the underlying system (from which the data are

generated) is represented by the model

$$y(k) = \phi^T(k)\theta_o + \varepsilon(k) \tag{7.30}$$

where θ_o is the true parameter vector (which is assumed to be well defined since we are discussing the convergence here) and $\varepsilon(k)$ is a term due to disturbance, noise, etc.

Some insight can be drawn by rewriting the least squares solution in the following form:

$$\begin{aligned} \hat{\theta}_N^{LS} &= \left[\tfrac{1}{N}\Sigma_{k=1}^N \phi(k)\phi^T(k)\right]^{-1} \tfrac{1}{N}\Sigma_{k=1}^N \phi(k)\left[\phi^T(k)\theta_o + \varepsilon(k)\right] \\ &= \theta_o + \left[\tfrac{1}{N}\Sigma_{k=1}^N \phi(k)\phi^T(k)\right]^{-1} \tfrac{1}{N}\Sigma_{k=1}^N \phi(k)\varepsilon(k) \end{aligned} \tag{7.31}$$

A desirable property of $\hat{\theta}_N^{LS}$ is that under fairly mild assumptions it converges to θ_o as the number of data points becomes large ($N \to \infty$). Note that the term

$$\left[\frac{1}{N}\sum_{k=1}^N \phi(k)\phi^T(k)\right]^{-1} \frac{1}{N}\sum_{k=1}^N \phi(k)\varepsilon(k)$$

represents the error in the parameter estimate. Assume that

$$\lim_{N\to\infty}\left(\frac{1}{N}\sum_{k=1}^N \phi(k)\phi^T(k)\right)$$

exists. This is true if the input is a quasi-stationary signal. In order that

$$\lim_{N\to\infty}\left[\frac{1}{N}\sum_{k=1}^N \phi(k)\phi^T(k)\right]^{-1} \frac{1}{N}\sum_{k=1}^N \phi(k)\varepsilon(k) = 0 \tag{7.32}$$

the following two conditions must be satisfied:

1.
$$\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^N \phi(k)\varepsilon(k) = 0 \tag{7.33}$$

2.
$$\text{rank}\left\{\lim_{N\to\infty}\left[\frac{1}{N}\sum_{k=1}^{N}\phi(k)\phi^T(k)\right]\right\} = \dim\{\phi\} \tag{7.34}$$

The first condition is satisfied if the regressor vector and the residual sequences are uncorrelated. There are two scenarios under which this condition holds:

- $\varepsilon(k)$ *is a zero-mean white sequence.* Since $\phi(k)$ does not contain $\varepsilon(k)$, $E\{\phi(k)\varepsilon(k)\} = 0$ and $\frac{1}{N}\sum_{k=1}^{N}\phi(k)\varepsilon(k) \to 0$ as $N \to \infty$. In the prediction error minimization, if the model structure is unbiased, $\varepsilon(k)$ is white.
- $\phi(k)$ *and* $\varepsilon(k)$*are independent sequences and one of them is zero-mean.* For instance, in the case of an FIR model (or an orthogonal expansion model), $\phi(k)$ contains inputs only and is therefore independent of $\varepsilon(k)$ whether it is white or nonwhite. This means that the FIR parameters can be made to converge to the true values even if the disturbance transfer function $H(q)$ is not known perfectly (resulting in nonwhite prediction errors), as long as $u_f(k)$ is designed to be zero-mean and independent of $\varepsilon(k)$. The same is not true for an ARX model since $\phi(k)$ contains past outputs that are correlated with a nonwhite $\varepsilon(k)$.

In order for the second condition to be satisfied, $\lim_{N\to\infty}\left[\frac{1}{N}\sum_{k=1}^{N}\phi(k)\phi^T(k)\right]$ must exist and should be nonsingular. The rank condition on the matrix $\lim_{N\to\infty}\left[\frac{1}{N}\sum_{k=1}^{N}\phi(k)\phi^T(k)\right]$ is called the *persistent excitation* condition as it is closely related to the notion of *order of persistent excitation* (of an input signal) that we shall discuss in Section 7.2.2.3.

Statistical Properties

Let us again assume that the underlying system is represented by (7.30). We further assume that $\varepsilon(k)$ is an independent, identically distributed

(i.i.d.) random variable sequence of zero mean and variance r_ε. Then, using (7.31), we can easily see that

$$E\{\hat{\theta}_N^{LS} - \theta_0\} = E\left\{\frac{1}{N}\left(\sum_{k=1}^{N}\phi(k)\phi^T(k)\right)^{-1}\frac{1}{N}\sum_{k=1}^{N}\phi(k)\varepsilon(k)\right\} = 0 \tag{7.35}$$

and

$$\begin{aligned} & E\{(\hat{\theta}_N^{LS} - \theta_o)(\hat{\theta}_N^{LS} - \theta_o)^T\} \\ = & \left(\tfrac{1}{N}\Sigma_{k=1}^{N}\phi(k)\phi^T(k)\right)^{-1}\left(\tfrac{1}{N^2}\Sigma_{k=1}^{N}\phi(k)r_\varepsilon\phi^T(k)\right)\left(\tfrac{1}{N}\Sigma_{k=1}^{N}\phi(k)\phi^T(k)\right)^{-1}]] \\ = & \left(\tfrac{1}{N}\Sigma_{k=1}^{N}\phi(k)\phi^T(k)\right)^{-1}\tfrac{r_\varepsilon}{N} \\ = & r_\varepsilon(\Phi_N^T\Phi_N)^{-1} \end{aligned} \tag{7.36}$$

(7.35) implies that the least squares estimate is "unbiased." (7.36) defines the covariance of the parameter estimate. This information can be used to compute confidence intervals. For instance, when normal distribution is assumed, one can compute an ellipsoid corresponding to a specific confidence level.

7.2.2.3 Persistency of Excitation

In the linear least squares identification, in order for parameters to converge to true values in the presence of noise, we must have

$$\text{rank}\left\{\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}\phi(k)\phi^T(k)\right\} = \dim\{\phi\} \tag{7.37}$$

This condition is closely related to the so called *persistency of excitation.* A signal $u(k)$ is said to b *persistently exciting of order* n if the following condition is satsified:

$$\text{rank}\{C_u^n\} = n \tag{7.38}$$

where

$$C_u^n = \lim_{N\to\infty} \frac{1}{N} \Sigma_{k=1}^{N} \left\{ \begin{bmatrix} u(k-1)u(k-1) & u(k-1)u(k-2) & \cdots & u(k-1)u(k-n) \\ u(k-2)u(k-1) & u(k-2)u(k-2) & \cdots & u(k-2)u(k-n) \\ \vdots & \ddots & \ddots & \vdots \\ u(k-n)u(k-1) & u(k-n)u(k-2) & \cdots & u(k-n)u(k-n) \end{bmatrix} \right\} \tag{7.39}$$

The above is equivalent to requiring the power spectrum of $u(k)$ to be nonzero at n or more distinct frequency points between $-\pi$ and π.

Now, suppose $\phi(k)$ consists of past inputs and outputs. A necessary and sufficient condition for (7.37) to hold is that:

$u(k)$ is persistently exciting of order $\dim\{\phi\}$.

This is obvious in the case that $\phi(k)$ is made of n past inputs only (as in FIR models). In this case,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \phi(k)\phi^T(k) = C_u^n \tag{7.40}$$

The condition also holds when $\phi(k)$ contains filtered past inputs $u_f(k-1), \cdots, u_f(k-n)$ (where $u_f(k) = H^{-1}(q)u(k)$). Note that:

$$\Phi_{u_f}(\omega) = \frac{\Phi_u(\omega)}{|H(e^{j\omega})|^2} \tag{7.41}$$

Hence, if $u(k)$ is persistently exciting of order n, so is $u_f(k)$. What is not so obvious (but can be proven) is that the above holds even when $\phi(k)$ contains past outputs.

An important conclusion that we can draw from this is that, in order to assure convergence of parameter estimates to true values, we must design the input signal $u(k)$ to be persistently exciting of order $\dim\{\theta\}$. A pulse is

not persistently exciting of any order since the rank of the matrix C_u^1 for such a signal is zero. A step signal is persistently exciting of order 1. A single step test is inadequate in the presence of significant disturbance or noise since only one parameter may be identified without error using such a signal. Sinusoidal signals are persistently exciting of second order since their spectra are nonzero at two frequency points. Finally, a random signal can be persistently exciting of any order since its spectrum is nonzero over a frequency interval. It is also noteworthy that a signal periodic with period n can at most be persistently exciting of order n.

Violation of the persistent excitation condition does not mean that obtaining estimates for parameters is impossible. It implies, however, that parameters do not converge to true values no matter how many data points are taken.

7.2.2.4 Frequency-Domain Bias Distribution Under PEM

The discussion of parameter convergence is based on the assumption that there exists a "true" parameter vector. Even when the parameters converge to their "best" values, it is still possible for the model to show significant bias from the true plant model if the model structure used for identification is not rich enough. For example, an FIR model with too few coefficients may differ from the true system significantly even with the best choice of impulse response coefficients. Understanding how the choice of input signal affects distribution of model bias in the frequency domain is important, especially for developing a model for closed-loop control purposes, since accuracy of fit in certain frequency regions (*e.g.*, cross-over frequency region) can be more important than others.

In the prediction error method, parameters are fitted based on the criterion

$$\min_{\theta} \frac{1}{N} \sum_{k=1}^{N} \hat{e}_{pred}^2(k,\theta) \tag{7.42}$$

where $\hat{e}_{pred}(k,\theta) = H^{-1}(q,\theta)\{\hat{y}(k) - G(q,\theta)u(k)\}$. Suppose the true system is represented by

$$\hat{y}(k) = G_o(q)u(k) + H_o(q)\varepsilon(k) \tag{7.43}$$

Then,

$$\hat{e}_{pred}(k,\theta) = \frac{G_o(q) - G(q,\theta)}{H(q,\theta)}u(k) + \frac{H_o(q)}{H(q,\theta)}\varepsilon(k) \tag{7.44}$$

By Parseval's theorem,

$$\begin{aligned} & \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \hat{e}_{pred}^2(k,\theta) && (7.45)\\ = & \int_{-\pi}^{\pi} \Phi_{\hat{e}}(\omega) d\omega && (7.46)\\ = & \int_{-\pi}^{\pi} \left(\left| G_o(e^{j\omega}) - G(e^{j\omega},\theta) \right|^2 \frac{\Phi_u(\omega)}{\left| H(e^{j\omega},\theta) \right|^2} + \frac{\left| H_o(e^{j\omega}) \right|^2}{\left| H(e^{j\omega},\theta) \right|^2} \Phi_{\varepsilon}(\omega) \right) d\omega \end{aligned}$$

where $\Phi_{\hat{e}}(\omega)$ is the spectrum of $\hat{e}_{pred}(k)$.

Note that, in the case that the disturbance model does not contain any unknown parameter,

$$\begin{aligned} & \lim_{N\to\infty} \tfrac{1}{N} \textstyle\sum_{k=1}^{N} \hat{e}_{pred}^2(k,\theta) \\ = & \int_{-\pi}^{\pi} \left(\left| G_o(e^{j\omega}) - G(e^{j\omega},\theta) \right|^2 \frac{\Phi_u(\omega)}{\left| H(e^{j\omega}) \right|^2} + \frac{\left| H_o(e^{j\omega}) \right|^2}{\left| H(e^{j\omega}) \right|^2} \Phi_{\varepsilon}(\omega) \right) d\omega \end{aligned} \tag{7.47}$$

Since the last term of the integrand is unaffected by the choice of θ, we may conclude that PEM selects θ such that the L_2-norm of the error $G_o(q) - G(q,\theta)$ weighted by the filtered input spectrum $\Phi_{u_f}(\omega)$ (where $u_f(k) = H^{-1}(q)u(k)$) is minimized. An implication is that, in order to obtain a good frequency response estimate at a certain frequency region, the filtered input u_f must be designed so that its power is concentrated in

the region. If we want good frequency estimates throughout the entire frequency range, an input signal with a flat spectrum (*e.g.*, a sequence of independent, zero mean random variables) is the best choice.

Frequency domain bias distribtuion can be made more flexible by minimizing the filtered prediction error $\hat{e}_{fpred}$ ($\triangleq L(q)e_{pred}$). In this case,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \hat{e}^2_{fpred}(k,\theta) \tag{7.48}$$

$$= \int_{-\pi}^{\pi} \left(\left| G_o(e^{j\omega}) - G(e^{j\omega},\theta) \right|^2 \frac{\Phi_u(\omega)}{\left|L(e^{j\omega})\right|^2 \left|H(e^{j\omega})\right|^2} + \frac{\left|H_o(e^{j\omega})\right|^2}{\left|L(e^{j\omega})\right|^2 \left|H(e^{j\omega})\right|^2} \Phi_\varepsilon(\omega) \right) d\omega$$

Hence, by prefiltering the data before the parameter estimation, one can affect the bias distribution. This gives an added flexibility when the input spectrum cannot be adjusted freely.

Finally, we have based our argument on the case where the disturbance model does not contain any parameter. When the disturbance model contains some of the parameters, the noise spectrum $|H_o(e^{j\omega})|^2$ does affect the bias distribution. However, the qualitative effects of the input spectrum and prefiltering remain the same.

7.2.3 PARAMETER ESTIMATION VIA STATISTICAL METHODS

In formulating the prediction error minimization, we did not require an exact statistical description of the *underlying plant*. Prediction error minimization is a logical criterion for parametric identification regardless of the true nature of the underlying plant (*i.e.*, even if the assumed model structure does not match the real plant exactly). In stochastic identification, a specific stochastic model is assumed for the underlying plant and plant parameters are estimated in an optimal fashion based on some well-defined

criterion. Although it may be difficult to come up with an exact description of the plant in reality, studying these methods can provide some useful insights into the performance of empirical methods like the prediction error minimization. We present the two most popular methods here.

7.2.3.1 Maximum Likelihood Estimation

In system identification, one is trying to extract system information out of measurements that are inherently unreliable. In maximum likelihood estimation, this is formalized by describing each observation as a realizaton of a random variable with certain probability distribution. For instance, if we assume a model

$$y(k) = \phi^T(k)\theta + \varepsilon(k) \tag{7.49}$$

where $\varepsilon(k)$ is a Gaussian variable with zero mean and variance r_ε, then the probability density function (PDF) of $y(k)$ becomes

$$dF(\zeta; y(k)) = \frac{1}{\sqrt{2\pi r_\varepsilon}} \exp\left\{-\frac{(\zeta - \phi^T(k)\theta)^2}{2r_\varepsilon}\right\} \tag{7.50}$$

In the above, ζ represents a particular realized value for $y(k)$.

In parametric identification with N data points, we can work with a joint PDF for $Y_N \triangleq (y(1), \cdots, y(N))$. Let us denote the joint PDF as $dF(\zeta_N; Y_N)$. Again, ζ^N is a variable representing realization of Y_N. Suppose the actual observations are given as $\hat{Y}_N = (\hat{y}(1), \cdots, \hat{y}(N))$. Once we insert these values into the probability density function, $dF(\hat{Y}_N; Y_N)$ is now a deterministic function of θ called "likelihood function." We denote the likelihood function for the observation $\hat{Y}_N$ as $\ell(\theta|\hat{Y}_N)$.

The basic idea of maximum likelihood estimation is to make the observations "as likely as possible" by choosing θ such that the liklihood

function is maximized. In other words,

$$\hat{\theta}_N^{ML} = \arg\left\{\max_\theta \ell(\theta|\hat{Y}_N)\right\} \tag{7.51}$$

Often, it is generally quite difficult to derive the likelihood function from a stochastic system model. An exception is the case when the model can be put into a linear predictor form in which the observation is linear with respect to both the unknown parameters and random variables.

Let us apply the maximum likelihood method to the following linear identification problem:

$$Y_N = \Phi_N\theta + \mathcal{E}_N \tag{7.52}$$

In the above, we assume that $\mathcal{E}_N$ is a zero-mean Gaussian variable vector of covariance $R_\mathcal{E}$. Then, we have

$$\begin{aligned} dF(\hat{Y}_N; Y_N) &= dF(\hat{Y}_N - \Phi_N\theta; \mathcal{E}_N) && (7.53)\\ &= \frac{1}{\sqrt{(2\pi)^N \det(R_\mathcal{E})}} \exp\left\{-\frac{1}{2}(\hat{Y}_N - \Phi_N\theta)^T R_\mathcal{E}^{-1}(\hat{Y}_N - \Phi_N\theta)\right\} \end{aligned}$$

The maximum likelihood estimator is defined as

$$\begin{aligned} \hat{\theta}_N^{ML} &= \arg\left\{\max_\theta dF(\hat{Y}_N; Y_N)\right\} && (7.54)\\ &= \arg\left\{\max_\theta \log\left(dF(\hat{Y}_N; Y_N)\right)\right\} && (7.55)\\ &= \arg\left\{\max_\theta \left(-\frac{1}{2}(\hat{Y}_N - \Phi_N\theta)^T R_\mathcal{E}(\hat{Y}_N - \Phi_N\theta)\right)\right\} && (7.56)\\ &= \arg\left\{\min_\theta \left(\frac{1}{2}(\hat{Y}_N - \Phi_N\theta)^T R_\mathcal{E}(\hat{Y}_N - \Phi_N\theta)\right)\right\} && (7.57) \end{aligned}$$

Note that the above is a weighted least squares estimator. We see that, when the weighting matrix is chosen as the inverse of the covariance matrix for the output error term $\mathcal{E}_N$, the weighted least squares estimation is equivalent to the maximum likelihood estimation. In addition, the unweighted least squares estimator is a maximum likelihood estimator for

the case when the output error is an i.i.d. Gaussian sequence (in which case the covariance matrix for $\mathcal{E}_N$ is in the form of $r_\varepsilon I_N$).

7.2.3.2 Bayesian Estimation

Bayesian estimation is a philosophically different approach to the parameter estimation problem. In this approach, parameters themselves are viewed as random variables with a certain prior probability distribution. If the observations are described in terms of the parameter vector, the probability distribution of the parameter vector changes after the observations. The distribution after the observations is called posterior probability distribution, which is given by the conditional distribution of the parameter vector conditioned with the observation vector. The parameter value for which the posterior PDF attains its maximum is called the "maximum a posteriori (MAP) estimate." It is also possible to use the mean of the posterior distribution as an estimate, which gives the "minimum variance estimate."

One of the useful rules in computing the posterior PDF is Bayes's rule. Let us denote the conditional PDF of the parameter vector for given observations as $dF(\hat{\theta}|\zeta_N;\theta|Y_N)$. Then, Bayes's rule says

$$dF(\hat{\theta}|\zeta_N;\theta|Y_N) = \frac{dF(\zeta_N|\hat{\theta};Y_N|\theta)\cdot dF(\hat{\theta};\theta)}{dF(\zeta_N;Y_N)} \tag{7.58}$$

$dF(\zeta_N;Y_N)$ is independent of θ and therefore is constant once it is evaluated for given observation $\hat{Y}_N$. Hence, the MAP estimator becomes

$$\hat{\theta}_N^{MAP} = \arg\left\{\max_{\hat{\theta}} dF(\zeta_N|\hat{\theta};Y_N|\theta)\cdot dF(\hat{\theta};\theta)\right\} \tag{7.59}$$

Note that we end up with a parameter value that maximizes the product of the likelihood function and the prior density.

Let us again apply this concept to the linear parameter estimation problem of

$$Y_N = \Phi_N \theta + \mathcal{E}_N \tag{7.60}$$

where $\mathcal{E}_N$ is a Gaussian vector of zero mean and covariance $R_\mathcal{E}$. We also treat θ as a Gaussian vector of mean $\hat{\theta}(0)$ and covariance $P(0)$. Hence, the prior distribution is a normal distribution of the above mean and covariance.

Next, let us evaluate the posterior PDF using Bayes's rule.

$$dF(\hat{\theta}|\hat{Y}_N;\theta|Y_N) = [\text{constant}] \times dF_\mathcal{N}(\hat{Y};Y_N)_{(\Phi_N\theta,R_\mathcal{E})} \cdot dF_\mathcal{N}(\hat{\theta},\theta)_{(\hat{\theta}(0),P(0))} \tag{7.61}$$

where

$$dF_\mathcal{N}(\hat{x},x)_{(\bar{x},R)} = \frac{1}{\sqrt{(2\pi)^N \det(R)}} \exp\left\{-\frac{1}{2}(\hat{x}-\bar{x})^T R^{-1}(\hat{x}-\bar{x})\right\} \tag{7.62}$$

The MAP estimate can be obtained by maximizing the logarithm of the posterior PDF:

$$\begin{aligned}\hat{\theta}_N^{MAP} &= \arg\left\{\max_{\hat{\theta}}\left(-\frac{1}{2}(\hat{Y}_N-\Phi_N\hat{\theta})^T R_\mathcal{E}^{-1}(\hat{Y}_N-\Phi_N\hat{\theta}) - \frac{1}{2}(\hat{\theta}-\hat{\theta}(0))^T P^{-1}(0)(\hat{\theta}-\hat{\theta}(0)\right)\right\} \\ &= \arg\left\{\min_{\hat{\theta}}\frac{1}{2}\left((\hat{Y}_N-\Phi_N\hat{\theta})^T R_\mathcal{E}^{-1}(\hat{Y}_N-\Phi_N\hat{\theta})\right) + (\hat{\theta}-\hat{\theta}(0))^T P^{-1}(0)(\hat{\theta}-\hat{\theta}(0)\right\}\end{aligned} \tag{7.63}$$

Solving the above least squares problem, we obtain

$$\hat{\theta}_N^{MAP} = \left(\Phi_N^T R_\mathcal{E}^{-1}\Phi_N + P^{-1}(0)\right)^{-1}\left(\Phi_N^T R_\mathcal{E}^{-1}\hat{Y}_N + P^{-1}(0)\hat{\theta}(0)\right) \tag{7.64}$$

Using the Matrix Inversion Lemma, one can rewrite the above as

$$\hat{\theta}_N^{MAP} = \hat{\theta}(0) + P(0)\Phi_N^T\left(\Phi_N^T P(0)\Phi_N + R_\mathcal{E}\right)^{-1}\left(\hat{Y}_N - \Phi_N\hat{\theta}(0)\right) \tag{7.65}$$

We make the following observations:

- The above indicates that, as long as $P(0)$ is chosen as a nonsingular

matrix and the persistent excitation condition is satisfied, $\hat{\theta}_N^{MAP}$ converges to $\hat{\theta}_N^{LS}$ as $N \to \infty$. Hence, all the asymptotic properties of the least squares identification apply to the above method as well.

- If $P(0)$ is chosen as a singular matrix, the estimate of θ may be *biased* since the null space of $P(0)$ represents the parameter subspace corresponding to zero update gain.
- From (7.63), we see that specifying the initial parameter covariance matrix $P(0)$ to be other than ∞I is equivalent to penalizing the deviation from the initial parameter guess through weighting matrix $P^{-1}(0)$ in the least squares framework. The standard least squares solution is interpreted in the Bayesian framework as the MAP solution corresponding to a uniform initial parameter distribution (i.e., $P(0) = \infty I$).

Utilizing prior knowledge in the above framework can help us obtain a smoother and more realistic impulse response. In Section **??**, we suggested using a diagonal weighting matrix to penalize the magnitudes of the impulse response coefficients so that a smoother step response can be obtained. We now see that this is equivalent to specifying the initial parameter covariance as a diagonal matrix (*i.e.*, the inverse of the weighting matrix) in the Bayesian framework. The statistical interpretation provides a formal justification for this practice and a systematic way to choose the weighting matrix (possibly as a nondiagonal matrix).

(7.65) can be written in the following recursive form:

$$\begin{aligned} \hat{\theta}(k) &= \hat{\theta}(k-1) + K(k)\left(y(k) - \phi^T(k)\hat{\theta}(k-1)\right) \\ K(k) &= \frac{P(k-1)\phi(k)}{1+\phi^T(k)P(k-1)\phi(k)} \\ P(k) &= P(k-1) - \frac{P(k-1)\phi(k)\phi^T(k)P(k-1)}{1+\phi^T(k)P(k-1)\phi(k)} \end{aligned} \tag{7.66}$$

where $\hat{\theta}(k)$ represents $\hat{\theta}_k^{MAP}$ or $E\{\theta|Y_k\}$ and

$P(k) = E\left\{(\theta - \hat{\theta}(k))(\theta - \hat{\theta}(k))^T | Y_k\right\}$. The above formula is easily derived by formulating the problem as a special case of state estimation and applying the Kalman filtering.

One could generalize the above to the time-varying parameters by using the following system model for parameter variation:

$$\begin{aligned} \theta(k) &= \theta(k-1) + w(k) \\ y(k) &= \phi^T(k)\theta(k) + \nu(k) \end{aligned} \tag{7.67}$$

where $w(k)$ is white noise. This way, the parameter vector $\theta(k)$ can be assumed to be time-varying in a random walk fashion. One may also model $w(k)$ and $\nu(k)$ as nonwhite signals by further augmenting the state vector as described earlier

We will demonstrate an application of the Bayesian approach to the impulse response coefficient identification through the following example.

Example:
In practice, it may be more appropriate to assume (in prior to the identification) the derivatives of the impulse response as zero-mean random variables of Gaussian distribution and specify the covariance of the derivative of the impulse response coefficients. In other words, one may specify

$$E\left\{\frac{dh}{dt}_{t=i\cdot T_s}\right\} \approx E\left\{\frac{h_i - h_{i-1}}{T_s}\right\} = 0; \qquad 1 \le i \le n \tag{7.68}$$

$$E\left\{\left(\frac{dh}{dt}_{t=i\cdot T_s}\right)^2\right\} \approx E\left\{\left(\frac{h_i - h_{i-1}}{T_s}\right)^2\right\} = \frac{\sigma_i}{T_s^2} \tag{7.69}$$

In this case, $P(t_0)$ (the covariance for θ) takes the following form:

$$P(t_0) = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & -1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \sigma_1 & & & & \\ & \sigma_2 & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & \sigma_n \end{bmatrix} \left(\begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & -1 & 1 \end{bmatrix}^T \right)^{-1} \tag{7.70}$$

Note that the above is translated as penalizing the 2-norm of the difference between two successive impulse response coefficients in the least squares identification method. It is straightforward to extend the above concepts and model the second order derivatives of the impulse response as normally distributed zero-mean random variables.

(Comment: ADD NUMERICAL EXAMPLE HERE!!!)

7.2.4 OTHER METHODS

There are other methods for estimating parameters in the literature. Among them, a method that stands out is the *instrumental variable (IV) method*. The basic idea behind this method is that, in order for a model to be good, the prediction error must show little or no correlation with past data. If they show significant correlation, it implies that there is information left over in the past data not utilized by the predictor.

In the IV method, a set of variables called "instruments" (denoted by vector η hereafter) must be defined first. η contains some transformations of past data $(y(k-1), \cdots, y(0), u(k-1), \cdots, u(0))$. Then, θ is determined from the following relation:

$$\frac{1}{N} \sum_{k=1}^{N} \eta(k) e_{pred}(k, \theta) = 0 \tag{7.71}$$

$\eta(k)$ is typically chosen to be of same dimension as the parameter vector θ. This way, one obtains the same number of equations as unknowns. Sometimes, η is chosen to be of higher dimension. Then, θ can be determined by minimizing some norm of $\frac{1}{N}\sum_{k=1}^{N}\eta(k)e_{pred}(k,\theta)$. Filtered e_{pred} can be used as well in the above. The success of the method obviously depends on the choice of instruments. See Ljung (1987) for guidelines on how to choose them. If $\eta(k)$ is chosen as $\phi(k)$, one obtains the same estimate as the least squares estimate. It is also possible to choose η that contains parameters. This leads to pseudo-linear regression.

Other variations to the least squares regression is the so called *biased regression* methods in which the regression is restricted to a subspace of the parameter space. The subspace is not chosen *a priori*, but is formed by incrementally adding on a one-dimensional space chosen to maximize the covariance of data ϕ (as in the *Principal Component Regression*) or to maximize the covariance between ϕ and y (as in the *Partial Least Squares*). These methods are designed to reduce the variance (esp. when the data do not show adequate excitation of the whole parameter space) at the expense of bias. In the Bayesian estimation setting, this can be interpreted as choosing a singular initial covariance matrix $P(0)$. However, the singular directions are determined on the basis of data rather than prior knowledge.

7.3 NONPARAMETRIC IDENTIFICATION METHODS

When one has little prior knowledge about the system, nonparametric identification which assumes very little about the underlying system is an alternative. Nonparametric model structures include frequency response models, impulse response models, *etc.*. These model structures intrinsically have no finite-dimensional parameter representations. In reality, however,

the dividing line between the parametric identification and the nonparametric identification is somewhat blurred: In nonparametric identification, some assumptions are always made about the system structure (*e.g.*, a finite length impulse response, smoothness of the frequency response) to obtain a well-posed estimation problem. In addition, in parametric identification, a proper choice of model order is often determined by examining the residuals from fitting models of various orders.

7.3.1 FREQUENCY RESPONSE IDENTIFICATION

Dynamics of a general linear system can be represented by the system's frequency response, which is defined through amplitude ratio and phase angle at each frequency. The frequency response information is conveniently represented as a complex function of ω whose modulus and argument define the amplitude ratio and the phase angle respectively. Such a function can be easily derived from the systems transfer function $G(q)$ by replacing q with $e^{j\omega}$. Hence, the amplitude ratio and phase angle of the system at each frequency is related to the transfer function parameters through the following relations:

$$A.R.(\omega) = |G(e^{j\omega})| = \sqrt{\mathrm{Re}\{G(e^{j\omega})\}^2 + \mathrm{Im}\{G(e^{j\omega})\}^2} \tag{7.72}$$

$$P.A.(\omega) = |G(e^{j\omega})| = \tan^{-1}\left[\frac{\mathrm{Im}\{G(e^{j\omega})\}}{\mathrm{Re}\{G(e^{j\omega})\}}\right] \tag{7.73}$$

Since $G(e^{j\omega})$ $(0 \leq \omega \leq \pi$ for system with sample time of 1) defines system dynamics completely, one approach to system identification is to identify $G(e^{j\omega})$ directly. This belongs to the category of nonparametric identification as frequency response is not parametrized by a finite-dimensional parameter vector (there are infinite number of frequency points).

7.3.1.1 Frequency Response Computation

The most immediate way to identify the frequency response is through a sine-wave testing, where sinusoidal perturbations are made directly to system input at different frequencies. Although conceptually straightforward, this method is of limited value in practice since (1) sinusoidal perturbations are difficult to make in practice, and (2) each experiment gives frequency response at only a single frequency.

A more practical approach is to use the results from the Fourier analysis. From the z-domain input / output relationship, it is immediate that, for system $y(k) = G(q)u(k)$,

$$G(e^{j\omega}) = \frac{Y(\omega)}{U(\omega)} \tag{7.74}$$

where

$$Y(\omega) = \sum_{k=1}^{\infty} y(k)e^{-j\omega k} \tag{7.75}$$

$$U(\omega) = \sum_{k=1}^{\infty} u(k)e^{-j\omega k} \tag{7.76}$$

Hence, by dividing the Fourier transform of the output data with that of the input data one can compute the system's frequency response. What complicates the frequency response identification in practice is that one only has finite length data. In addition, output data are corrupted by noise and disturbances.

Let us assume that the underlying system is represented by

$$y(k) = G(q)u(k) + e(k) \tag{7.77}$$

where $e(k)$ is a zero-mean stationary sequence and collectively describes the

effect of noise and disturbance. We define

$$Y_N(\omega) \triangleq \frac{1}{\sqrt{N}} \sum_{k=1}^{N} y(k) e^{-j\omega k} \tag{7.78}$$

$$U_N(\omega) \triangleq \frac{1}{\sqrt{N}} \sum_{k=1}^{N} u(k) e^{-j\omega k} \tag{7.79}$$

Then,

$$G_N(\omega) \triangleq \frac{Y_N(\omega)}{U_N(\omega)} = G(e^{j\omega}) + \frac{R_N(\omega)}{U_N(\omega)} + \frac{E_N(\omega)}{U_N(\omega)} \tag{7.80}$$

where $|R_N(\omega)| = \frac{c_1}{\sqrt{N}}$ for some c_1 (Ljung, 1987). $G_N(\omega)$ computed as above using N data points is an estimate of the true system frequency response $G(e^{j\omega})$ and will be referred to as the "Empirical Transfer Function Estimate (ETFE)."

7.3.1.2 Statistical Properties of the ETFE

Let us take expectation of (7.80):

$$E\{G_N(\omega)\} = E\left\{G(e^{j\omega}) + \frac{R_N(\omega)}{U_N(\omega)} + \frac{E_N(\omega)}{U_N(\omega)}\right\} = G(e^{j\omega}) + \frac{R_N(\omega)}{U_N(\omega)} \tag{7.81}$$

We can also compute the variance as

$$E\left\{\left(G_N(\omega) - G(e^{j\omega})\right)\left(G_N(\omega) - G(e^{-j\omega})\right)\right\} = \frac{\Phi_e + \rho_N}{|U_N(\omega)|^2} \tag{7.82}$$

where $\rho_N \leq \frac{c_2}{N}$ (Ljung, 1987).

The implications of the above are as follows:

- Since the second term of the RHS of (7.81) decays as $\frac{1}{\sqrt{N}}$, $G_N(\omega)$ is an *asymptotically unbiased estimate* of $G(e^{j\omega})$.
- If $u(k)$ is a periodic signal with period of N, $|U_N(\omega)|$ is nonzero only at

N frequency points (at $\omega = \frac{2\pi \cdot k}{N}, k = 0, \cdots, N-1$). This means that the ETFE is defined only at the N frequency points. $|U_N(\omega)|$ at these frequency points keeps growing larger as $N \to \infty$, and from (7.82), we see that the variance goes to zero.

- If $u(k)$ is a randomly generated signal, as N increases,the number of frequency points at which the ETFE can be computed also increases. However, $|U_N(\omega)|^2$ is a function that fluctuates around the spectrum of $u(k)$ and therefore does not increase with data. From (7.82), we conclude that the variance does not decay to zero. This is characteristic of any nonparameteric identification where, roughly speaking, one is trying to estimate infinite number of parameters.

A practical implication of the last comment is that the estimate can be very sensitive to noise in the data (no matter how many data points are used). Hence, some smoothing is needed. The following are some simple smoothing methods:

- Select a finite number of frequency points, $\omega_1, \cdots \omega_N$ between 0 and π. Assume that $G(e^{j\omega})$ is constant over $\omega_i - \delta\omega \leq \omega \leq \omega_i + \delta\omega$. Hence, the EFTE ($G_N(\omega)$) obtained within this window are averaged, for instance, according to the signal-to-noise ratio $\frac{\Phi_e}{|U_N(\omega)|^2}$. Since the number of frequency response parameters become finite under the assumption, the variance decays to zero as $1/N$. However, the assumption leads to bias.

- A generalization of the above is to use the weighting function $W_s(\zeta - \omega)$ for smoothing. The ETFE is smoothed according to

$$G_N^s(\omega) = \frac{\int_{-\pi}^{\pi} W_s(\zeta - \omega) G_N(\omega) \frac{|U_N(\zeta)|^2}{\Phi_e(\omega)} d\zeta}{\int_{-\pi}^{\pi} W_s(\zeta - \omega) \frac{|U_N(\zeta)|^2}{\Phi_e(\omega)} d\zeta} \tag{7.83}$$

  W_s is a function that is centered around zero and is symmetric. It usually includes a parameter that determines the width of the

smoothing window and therefore the trade-off between bias and variance. Larger window reduces variance, but increases bias and vice versa. For typical choices of W_s, see Table 6.1 of Ljung (1987). Again, the variance can be shown to decay as $1/N$ under a nonzero smoothing window.

7.3.2 IMPULSE RESPONSE IDENTIFICATION

Impulse response identification is another form of nonparametric identification, that is commonly used in practice. Suppose the underlying system is described by convolution model

$$y(k) = \sum_{i=1}^{\infty} H_i u(k-i) + e_k \tag{7.84}$$

Now post-multiply $u^T(k-\tau)$ to the above equation to obtain

$$y(k)u^T(k-\tau) = \sum_{i=1}^{\infty} H_i u(k-i)u^T(k-\tau) + e(k)u^T(k-\tau) \tag{7.85}$$

Summing up the data from $k=1$ to $k=N$,

$$\left(\frac{1}{N}\sum_{k=1}^{N} y(k)u^T(k-\tau)\right) = \sum_{i=1}^{\infty} H_i \left(\frac{1}{N}\sum_{k=1}^{N} u(k-i)u^T(k-\tau)\right) + \left(\frac{1}{N}\sum_{k=1}^{N} e(k)u^T(k-\tau)\right) \tag{7.86}$$

Assuming the input had remained at the steady-state value (i.e., $u(k)=0$ for $k \le 0$), the above can be represented by

$$R_{yu}(\tau) = \sum_{i=1}^{\infty} H_i R_{uu}(\tau - i) + R_{eu}(\tau) \tag{7.87}$$

where

$$R_{yu}(\tau) = \frac{1}{N}\sum_{k=1}^{N} y(k)u^T(k-\tau) \tag{7.88}$$

$$R_{uu}(\tau) = \frac{1}{N}\sum_{k=1}^{N} u(k)u^T(k-\tau) \tag{7.89}$$

$$R_{eu}(\tau) = \frac{1}{N}\sum_{k=1}^{N} e(k)u^T(k-\tau) \tag{7.90}$$

The above equation can also be derived from a statistical argument. More specifically, we can take expectation of (7.87) to obtain

$$E\{y(k)u^T(k-\tau)\} = \sum_{i=1}^{\infty} H_i E\{u(k-i)u^T(k-\tau)\} + E\{e(k)u^T(k-\tau)\} \tag{7.91}$$

Assuming $\{u(k)\}$ and $\{e(k)\}$ are stationary sequences, R_{uu}, R_{yu} and R_{eu} are estimates of the expectations based on N data points.

Now, let us assume that $\{u(k)\}$ is a zero-mean stationary sequence that is uncorrelated with $\{e(k)\}$, which is also stationary (or $\{e(k)\}$ is a zero-mean stationary sequence uncorrelated with $\{u(k)\}$). Then, $R_{eu}(\tau) \to 0$ as $N \to \infty$. Let us also assume that $H_i = 0$ for $i > n$. An appropriate choice of n can be determined by examining $R_{yu}(\tau)$ under a white noise perturbation. When the input perturbation signal is white, $R_{uu}(i) = 0$ except $i = 0$. From the above, it is clear that $R_{yu}(\tau) =)$ if $H_\tau = 0$. Hence, one can choose n where $R_{yu} \approx 0$ for $\tau > n$.

With these assumptions, as $N \to \infty$, we can write (7.87) as

$$\begin{bmatrix} R_{yu}(1) & R_{yu}(2) & \cdots & R_{yu}(n) \end{bmatrix} \tag{7.92}$$

$$\approx \begin{bmatrix} H_1 & H_2 & \cdots & H_n \end{bmatrix} \begin{bmatrix} R_{uu}(0) & R_{uu}(1) & \cdots & R_{uu}(n-1) \\ R_{uu}(-1) & R_{uu}(0) & \cdots & R_{uu}(n-1) \\ \vdots & \ddots & \ddots & \vdots \\ R_{uu}(-n+1) & R_{uu}(-n+2) & \cdots & R_{uu}(0) \end{bmatrix}$$

Taking transpose of the above equation and rearranging it gives

$$\begin{bmatrix} H_1^T \\ H_2^T \\ \vdots \\ H_n^T \end{bmatrix} \approx \begin{bmatrix} R_{uu}(0) & R_{uu}(1) & \cdots & R_{uu}(n-1) \\ R_{uu}(-1) & R_{uu}(0) & \cdots & R_{uu}(n-2) \\ \vdots & \ddots & \ddots & \vdots \\ R_{uu}(-n+1) & R_{uu}(-n+2) & \cdots & R_{uu}(0) \end{bmatrix}^{-1} \begin{bmatrix} R_{yu}^T(1) \\ R_{yu}^T(2) \\ \vdots \\ R_{yu}^T(n) \end{bmatrix} \tag{7.93}$$

With finite-length data, parameter variance can be significant. However, because we limited the number of impulse response coefficients to n by assuming $H_i = 0, i > n$, the variance decays as $1/N$ (assuming the matrix Φ remains nonsingular). However, some bias results because of the truncation. Again, the choice of n determines the trade-off between the variance and the bias.

Note that (7.93) gives the same estimate as the least squares identification. In the case that $\{e(k)\}$ is nonstationary due to integrating type disturbances, differenced data, $\Delta y(k)$ and $\Delta u(k)$, can be used as before.

7.3.3 SUBSPACE IDENTIFICATION

There applications where it is necessary to embed into the model disturbance correlations among different outputs. In this case, MIMO identification (rather than SISO or MISO identification) is needed. Transfer function models are difficult to work with in this context, since it gives rise to a numerically ill-conditioned, nonlinear estimation problem with possible local minima. In addition, significant prior knowledge (*e.g.*, the system order, the observability index) is needed to obtain a model parameterization. An alternative is to identify a state-space model directly, using a *subspace identification* method. Different subspace identification algorithms available in the literature share the same basic concept, which

will be presented here.

Assume that the underlying system is given by

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) + \varepsilon_1(k) \\ y(k) &= Cx(k) + \varepsilon_2(k) \end{aligned} \tag{7.94}$$

where $\begin{bmatrix} \varepsilon_1(k) \\ \varepsilon_2(k) \end{bmatrix}$ is a zero-mean, i.i.d. vector sequence.

The system is assumed to be controllable (from $[u^T \ \varepsilon_1^T]^T$) and observable. In addition, the stochastic part of the system is assumed to be stationary. The objective is to identify from input-output data a state-space model

$$\begin{aligned} \tilde{x}(k+1) &= \tilde{A}\tilde{x}(k) + \tilde{B}u(k) + \tilde{\varepsilon}_1(k) \\ y(k) &= \tilde{C}\tilde{x}(k) + \tilde{\varepsilon}_2(k) \end{aligned} \tag{7.95}$$

that is equal to (7.94) in an *input-output sense*. We will assume for the sake of simplicity that the input sequence $u(k)$ used in the identification is a white noise sequence.

Consider the following optimal multi-step prediction equation (of finite memory):

$$\begin{aligned} \begin{bmatrix} y(k+1) \\ y(k+2) \\ \vdots \\ y(k+\bar{n}) \end{bmatrix} &= L_1 \begin{bmatrix} y(k-\bar{n}+1) \\ y(k-\bar{n}+2) \\ \vdots \\ y(k) \end{bmatrix} + L_2 \begin{bmatrix} u(k-\bar{n}+1) \\ u(k-\bar{n}+2) \\ \vdots \\ u(k) \end{bmatrix} \\ &\quad + L_3 \begin{bmatrix} u(k+1) \\ u(k+2) \\ \vdots \\ u(k+\bar{n}-1) \end{bmatrix} + \begin{bmatrix} e(k+1|k) \\ e(k+2|k) \\ \vdots \\ e(k+\bar{n}|k) \end{bmatrix} \end{aligned} \tag{7.96}$$

$$= \begin{bmatrix} y(k+1|k) \\ y(k+2|k) \\ \vdots \\ y(k+\bar{n}|k) \end{bmatrix} + \begin{bmatrix} e(k+1|k) \\ e(k+2|k) \\ \vdots \\ e(k+\bar{n}|k) \end{bmatrix} \tag{7.97}$$

$\bar{n} > n$ where n is the system order. $y(k+i|k)$ represents the optimal prediction of $y(k+i)$ on the basis of data $y(k-\bar{n}+1), \cdots, y(k)$ and $u(k-\bar{n}+1), \cdots, u(k+\bar{n}-1)$. $e(k+i|k)$ denotes the respective prediction error. $L_1 \in \mathcal{R}^{n_y \cdot \bar{n} \times n_y \cdot \bar{n}}$, $L_2 \in \mathcal{R}^{n_y \cdot \bar{n} \times n_u \cdot \bar{n}}$ and $L_3 \in \mathcal{R}^{n_y \cdot \bar{n} \times n_u \cdot (\bar{n}-1)}$ are functions of system matrices.

The optimal prediction error $e(k+i|k),\ i \leq 1 \leq \bar{n}$ is zero-mean and uncorrelated with $y(k-\bar{n}+1), \cdots, y(k)$ and $u(k-\bar{n}+1), \cdots, u(k+\bar{n}-1)$. Hence, unbiased, consistent esimates of L_1, L_2 and L_3 can be obtained by applying linear least squares identification. L_1, L_2 and L_3 are related to the system matrices and covariance matrices in a complex manner, and extracting the system matrices directly from L_1, L_2 and L_3 would involve a very difficult nonlinear optimization. It also requires a special parameterization of model matrices in order to prevent a loss of identifiability. Clearly, an alternative way to generate the system matrices is desirable.

We can rewrite the optimal predictions in (7.96) in terms of a Kalman filter estimate as follows:

$$\begin{bmatrix} y(k+1|k) \\ y(k+2|k) \\ \vdots \\ y(k+\bar{n}|k) \end{bmatrix} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{\bar{n}-1} \end{bmatrix} x(k+1|k) + L_3 \begin{bmatrix} u(k+1) \\ u(k+2) \\ \vdots \\ u(k+\bar{n}-1) \end{bmatrix} \tag{7.98}$$

$x(k+1|k)$ represents an estimate of $x(k+1)$ that is obtained by running a nonsteady-state Kalman filter started with an initial estimate of

$x(k-\bar{n}+1|k-\bar{n})=0$ and initial covariance matrix corresponding to the open-loop, steady-state covariance of x.[1] Comparing (7.98) with (7.96), one can conclude that

$$\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{\bar{n}-1} \end{bmatrix} x(k+1|k) = \begin{bmatrix} L_1 & L_2 \end{bmatrix} \begin{bmatrix} y(k-\bar{n}+1) \\ \vdots \\ y(k) \\ u(k-\bar{n}+1) \\ \vdots \\ u(k) \end{bmatrix} \tag{7.99}$$

Hence, the extended observability matrix and $\begin{bmatrix} L_1 & L_2 \end{bmatrix}$ have the same image space and examining the rank of the latter gives the system order.

In constructing a state-space model from input-output data, there clearly exists some degrees-of-freedom since the basis for the state vector can be chosen arbitrarily without affecting the input-output relation. This means that the extended observability matrix for the identified model (7.95) (denoted as Γ_o from this point on) can be *any* matrix (of dimension $(\bar{n}\cdot n_y)\times n$) that has the same image space as $\begin{bmatrix} L_1 & L_2 \end{bmatrix}$. Let the SVD of $\begin{bmatrix} L_1 & L_2 \end{bmatrix}$ be represented as follows:

$$\begin{bmatrix} L_1 & L_2 \end{bmatrix} = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} \tag{7.100}$$

We choose $\Gamma_o = U_1\Sigma_1^{1/2}$. This defines the basis for the state vector. Let $\tilde{x}$ denote x written in terms of the above-defined basis. We then express the

[1] This interpretation does not hold in the case of time-correlated input sequence since future inputs can then contribute to the estimation of past outputs. However, a similar interpretation can be developed and the theory extends straightforwardly with some modifications.

system equation in terms of the new basis as follows:

$$\tilde{x}(k+1) = \tilde{A}\tilde{x}(k) + \tilde{B}u(k) + \tilde{\varepsilon}_1(k) \tag{7.101}$$

$$y(k) = \tilde{C}\tilde{x}(k) + \tilde{\varepsilon}_2(k) \tag{7.102}$$

The form of the state-space model that will actually be identified is the following Kalman filter equation for the above system:

$$\tilde{x}(k+2|k+1) = \tilde{A}\tilde{x}(k+1|k) + \tilde{B}u(k+1) + \underbrace{K(k+1)\zeta(k+1)}_{\tilde{\varepsilon}_1(k+1)} \tag{7.103}$$

$$y(k+1) = \tilde{C}\tilde{x}(k+1|k) + \underbrace{\zeta(k+1)}_{\tilde{\varepsilon}_2(k+1)} \tag{7.104}$$

$\tilde{x}(k+1|k)$ and $\tilde{x}(k+2|k+1)$ are two consecutive estimates generated from a nonsteady-state Kalman filter and $K(k+1)$ is the Kalman filter gain. ζ represents the innovation term (note $\zeta(k+1) = y(k+1) - \tilde{x}(k+1|k)$).

Now that the identification problem is well-defined, we discuss the construction of system matrices. In order to identify the system matrices using the relations in (7.103)–(7.104), we need data for the Kalman filter estimates $\tilde{x}(k+1|k)$ and $\tilde{x}(k+2|k+1)$. Let us define $\tilde{x}(k+2|k+1)$ and $\tilde{x}(k+1|k)$ as the estimates from the nonsteady-state Kalman filter for system (7.101), started with the initial estimate of $\tilde{x}(k-\bar{n}+1|k-\bar{n}) = 0$ and initial covariance given by the open-loop, steady-state covariance of $\tilde{x}$. Then, according to (7.99),

$$, \; {}_o\tilde{x}(k+1|k) = \begin{bmatrix} L_1 & L_2 \end{bmatrix} \begin{bmatrix} y(k-\bar{n}+1) \\ \vdots \\ y(k) \\ u(k-\bar{n}+1) \\ \vdots \\ u(k) \end{bmatrix} \tag{7.105}$$

Hence, the data for $\tilde{x}(k+1|k)$ can be found through the following formula:

$$\tilde{x}(k+1|k) = ,_o^{\dagger} \begin{bmatrix} L_1 & L_2 \end{bmatrix} \begin{bmatrix} y(k-\bar{n}+1) \\ \vdots \\ y(k) \\ u(k-\bar{n}+1) \\ \vdots \\ u(k-1) \end{bmatrix} \tag{7.106}$$

It is important to recognize that the data for $\tilde{x}(k+2|k+1)$ cannot be obtained by time-shifting the data for $\tilde{x}(k+1|k)$, since this will result in the Kalman filter estimate for $\tilde{x}(k+2)$ with a different starting estimate of $x(k-\bar{n}+2|k-\bar{n}+1) = 0$. Instead, one must start from the prediction equation below and follow the same procedure as before:

$$\begin{bmatrix} y(k+2|k+1) \\ y(k+3|k+1) \\ \vdots \\ y(k+\bar{n}|k+1) \end{bmatrix} = \hat{L}_1 \begin{bmatrix} y(k-\bar{n}+1) \\ y(k-\bar{n}+2) \\ \vdots \\ y(k+1) \end{bmatrix} + \hat{L}_2 \begin{bmatrix} u(k-\bar{n}+1) \\ u(k-\bar{n}+2) \\ \vdots \\ u(k+1) \end{bmatrix} + \hat{L}_3 \begin{bmatrix} u(k+2) \\ u(k+3) \\ \vdots \\ u(k+\bar{n}-1) \end{bmatrix} \tag{7.107}$$

Once the data for
$\begin{bmatrix} y^T(k+2|k+1) & y^T(k+3|k+1) & \cdots & y^T(k+\bar{n}|k+1) \end{bmatrix}^T$ are obtained by using the estimates, the data for $\tilde{x}(k+2|k+1)$ can be derived by multiplying them with the pseduo-inverse of $\hat{,}_o$ (which is $,_o$ with the last n_y rows eliminated).

Once the data for $\tilde{x}(k+1|k)$ and $\tilde{x}(k+2|k+1)$ are generated, one can find the system matrices by applying least squares identification to (7.103). Since $\zeta(k+1)$ is a zero-mean sequence that is independent of $\tilde{x}(k+1|k)$ and $u(k+1)$, the least squares method gives unbiased, consistent estimates of $\tilde{A}, \tilde{B}$ and $\tilde{C}$. The covariance matrix for $[\tilde{\varepsilon}_1^T \ \tilde{\varepsilon}_2^T]^T$ can also be computed from the residual sequence.

7.3.3.1 Properties and Issues

The subspace identification method we just described has the following properties (Comment: see Van Overschee and De Moor REF for proofs and discussions):

- The resulting model is asymptotically unbiased.
- The estimates for the covariance matrices are biased, however, due to the fact that (7.103) is a nonsteady-state Kalman filter. The approximation error diminishes as $\bar{n} \to \infty$.

Strengths of the method are that it requires only numerically stable, noniterative linear algebra operations only and that very little prior knowledge (an upper-bound on the system order) is needed to start up the algorithm. However, there are some drawbacks as well. Although the method yields an asymptotically unbiased model, very little can be said about the model quality obtained with finite data. In practice, one must always work with finite-length data sets. In addition, various nonideal factors like nonlinearity and nonstationarity make the residual sequence $e(k+i|k)$ in (7.96) become correlated with the regression data. Because of these reasons, L_1, L_2 obtained from the least squares identification (which are critical for determining the system order and generating data for the Kalman filter estimates) may have significant variance. Although expected errors in the estimates of these matrices can be quantified, it is difficult to say how these errors affect the final model quality (measured in terms of prediction error, frequency response error, etc.). One implication is that, in general, one needs a large amount of data in order to guarantee much success with these algorithms (which is only natural since these algorithms use very little prior knowledge). Another implication is that the above does not replace the traditional parametric identification, but complements it.

For instance, it has been suggested that the subspace methods be used to provide a starting estimate for the prediction error minimization.

Another related issue is that, becuase of the variance, the SVD of $\begin{bmatrix} L_1 & L_2 \end{bmatrix}$ is likely to show many more nonzero singular values than the intrinsic system order. In order not to overfit the data, one has to limit the system order by eliminating the *negligible* singular values in forming the , $_o$ matrix. In the context of model reduction, this is along the same line as the Hankel norm reduction. An alternative for deciding the system order and the basis for the states is to use the SVD of the matrix $\begin{bmatrix} L_1 & L_2 \end{bmatrix} Y$, where Y is the matrix whose columns contain the data for $\begin{bmatrix} y(k-\bar{n}+1)^T & \cdots & y(k)^T & u(k-\bar{n}+1)^T & \cdots & u(k-1)^T \end{bmatrix}^T$. In this case, the singluar values indicate how much of the output data are explained by different linear modes (in the 2-norm sense). In the context of model reduction, this corresponds to a frequency-weighting with the input spectrum (for the deterministic part). This step of determining the model order and basis is somewhat subjective, but is often critical.

Finally, the requirement that the stochastic part of the system be stationary should not be overlooked. If the system has integrating type disturbances, one can difference the input output data before applying the algorithm. Further low-pass filtering may be necessary not to over-emphasize the high frequency fitting (recall the discussion on the frequency-domain bias distribution).