

---

---

# Networked Sensing, Estimation and Control Systems

---

Vijay Gupta

University of Notre Dame

Richard M. Murray

California Institute of Technology

Ling Shi

Hong Kong University of  
Science and Technology

Bruno Sinopoli

Carnegie Mellon University

DRAFT v1.1b, 11 January 2010

© V. Gupta, R. M. Murray, L. Shi and B. Sinopoli  
All rights reserved.

This manuscript is for review purposes only and may not be reproduced, in whole or in part, without written consent from the author. Permission is granted for individual, non-commercial use.

---

## Chapter 2

### State Estimation and Sensor Fusion

In this chapter, we provide an overview of Kalman filter and linear quadratic Gaussian (LQG) control. We will first provide a quick summary of basic theories of probability and stochastic process, which will be used to derive the Kalman filter equations. Some properties of Kalman filter and its steady-state error covariance matrix will also be provided. After that, we will introduce LQG control and derive the optimal control law using a dynamic programming approach.

The material of this chapter will be the foundation of most subsequent chapters including Chapters 4, 6 and 8. In particular, Chapter 6 considers the effect of data packet drops on Kalman filtering and LQG control and Chapter 8 considers distributed Kalman filtering.

#### 2.1 Review of Probability and Random Process

We assume the readers have some exposure to the theory of probability and random process. The material presented in this section only serves as a quick review of some basic concepts and tools from probability and random process that will be helpful to understand and derive some important results in subsequent sections and chapters. Good introductory books on probability and random process are [GS01] and [LG93].

##### Random Variables

Consider an experiment with many (possibly infinite) outcomes. All these outcomes form the *sample space*  $\Omega$ . A subset  $A \subset \Omega$  is called an *event*. Two events  $A_1, A_2$  are called *mutually disjoint* if  $A_1 \cap A_2 = \emptyset$ . The *complement* of an event  $A$  is defined as  $\bar{A} = \Omega \setminus A$ . A *probability measure*  $P(\cdot)$  is a mapping from  $\Omega$  into the interval  $[0, 1]$  such that the following axioms of probability are satisfied:

1.  $P(A) \geq 0$  for all  $A \subset \Omega$ .
2.  $P(\Omega) = 1$ .
3. If  $\{A_i, i = 1, 2, \dots\}$  is a collection of disjoint members of  $\mathcal{F}$ , i.e.,  $A_i \cap A_j = \emptyset$  for all  $i, j$ , then  $P(\cup A_i) = \sum_i P(A_i)$ .

From the axioms of probability, it follows that

$$P(A) \leq 1, \quad P(\emptyset) = 0, \quad P(\bar{A}) = 1 - P(A), \quad P(\cup A_i) \leq \sum_i P(A_i).$$

The joint probability of two events  $A$  and  $B$  is  $P(A \cap B)$  which is often written as  $P(AB)$  for simplicity. The conditional probability of  $A$  given  $B$  i.e., the probability

that  $A$  occurs if  $B$  occurs in an experiment is

$$P(A|B) = \frac{P(AB)}{P(B)}, \quad \text{assuming } P(B) \neq 0.$$

$A$  and  $B$  are *mutually independent* if

$$P(AB) = P(A)P(B).$$

If  $P(B) \neq 0$ , the conditional probability  $P(A|B)$  can be calculated from *Bayes' Rule* as

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

If  $A_i, i = 1, 2, \dots$  are mutually disjoint and  $\cup A_i = \Omega$ , then

$$P(B) = \sum_i P(B|A_i)P(A_i)$$

and

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_i P(B|A_i)P(A_i)}.$$

A *random variable* is a function  $X : \Omega \rightarrow \mathbb{R}$ . The *cumulative distribution function* of a random variable  $X$  is a function  $F_X : \mathbb{R} \rightarrow [0, 1]$  given by

$$F_X(x) = P(X \leq x).$$

The cumulative distribution function  $F$  has the following properties

1.  $\lim_{x \rightarrow -\infty} F_X(x) = 0$  and  $\lim_{x \rightarrow \infty} F_X(x) = 1$ .
2. If  $x \leq y$ , then  $F_X(x) \leq F_X(y)$ .
3.  $F_X$  is right-continuous.

When  $F_X$  is differentiable, we can define the associated *probability density function*  $p_X(x)$  as

$$p_X(x) = \frac{dF_X(x)}{dx}.$$

The *joint cumulative distribution function* of two random variables  $X$  and  $Y$ , denoted as  $F_{XY}(x, y)$ , is given by

$$F_{XY}(x, y) = P(X \leq x) \cap P(Y \leq y).$$

If its derivative exists, the associated joint probability density function is given by

$$p_{XY}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{XY}(x, y).$$

Given  $F_{XY}(x, y)$ , the *marginal distribution functions* of  $X$  and  $Y$  can be calculated as

$$F_X(x) = P(X \leq x) = F_{XY}(x, \infty), \quad F_Y(y) = P(Y \leq y) = F_{XY}(\infty, y).$$

It follows that the *marginal density functions* of  $X$  and  $Y$  are

$$p_X(x) = \int_{-\infty}^{\infty} F_{XY}(x, y) dy, \quad p_Y(y) = \int_{-\infty}^{\infty} F_{XY}(x, y) dx.$$

The *conditional density function* of  $X$  given  $Y$  is given by

$$p_{X|Y}(x|y) = \frac{p_{XY}(x, y)}{p_Y(y)}.$$

The density function of  $X$  can also be calculated as

$$p_X(x) = \int_{-\infty}^{\infty} p_{X|Y}(x|y) p_Y(y) dy.$$

If  $X$  and  $Y$  are independent random variables, then the following statements holds and are equivalent to each other:

1.  $F_{XY}(x, y) = F_X(x)F_Y(y)$ .
2.  $p_{XY}(x, y) = p_X(x)p_Y(y)$ .
3.  $p_{X|Y}(x|y) = p_X(x)$ .

A random variable  $X$  is completely specified by its distribution function  $F_X(x)$  or density function  $p_X(x)$ . In many situations,  $F_X(x)$  or  $p_X(x)$  are difficult to obtain. It turns out the *mean*  $\mu_X$  and *variance*  $\sigma_X^2$  may provide us enough (useful) information about  $X$ . The mean and variance of a random variable  $X$  are defined as follows:

$$\begin{aligned} \mu_X &= \mathbb{E}[X] = \int_{-\infty}^{\infty} xp_X(x) dx, \\ \sigma_X^2 &= \mathbb{E}[(X - \mathbb{E}[X])^2] = \int_{-\infty}^{\infty} (X - \mathbb{E}[X])^2 p_X(x) dx. \end{aligned}$$

We denote  $\mathbb{E}[\cdot]$  as the *expectation operator*. Since  $\mathbb{E}[\cdot]$  is a linear operator,  $\sigma_X^2$  can also be calculated as

$$\sigma_X^2 = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

If  $X$  is a zero-mean random variable, i.e.,  $\mathbb{E}[X] = 0$ , then  $\sigma_X = \mathbb{E}[X^2]$ . The  $k$ th *moment* of  $X$  is  $m_k = \mathbb{E}[X^k]$  and the  $k$ th *central moment* is  $\mu_k = \mathbb{E}[(X - \mathbb{E}[X])^k]$ .

The *covariance* of two random variables  $X$  and  $Y$  is defined as  $\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$ .  $X$  and  $Y$  are *uncorrelated* if  $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ . If  $X$  and  $Y$  are uncorrelated, it is easy to verify that the covariance of  $X$  and  $Y$  is equal to zero. Clearly if  $X$  and  $Y$  are independent, then they are uncorrelated. However the converse does not hold in general.

The conditional expectation of  $X$  given  $Y = y$  is

$$\mathbb{E}[X|Y = y] = \int_{-\infty}^{\infty} xp_{X|Y}(x|y) dx$$

which is a number that depends on the value of  $y$ . Similarly, the conditional expectation of  $X$  given  $Y$  is

$$\mathbb{E}[X|Y] = \int_{-\infty}^{\infty} xp_{X|Y}(x|Y)dx$$

which is also a random variable that depends on  $Y$ , i.e., it is a *function of the random variable*  $Y$ . The following property is very important and has great practical value in evaluating  $\mathbb{E}[X]$ :

$$\mathbb{E}[X] = \mathbb{E}_Y[\mathbb{E}_X[X|Y]],$$

i.e., we first find the conditional expectation of  $X$  (conditioned on  $Y$ ), and then remove the condition by taking the expectation with respect to  $Y$ . From this property, one can easily verify that if  $X$  and  $Y$  are independent, then

$$\mathbb{E}[X|Y] = \mathbb{E}[X].$$

Furthermore if  $X$  and  $Y$  are jointly independent of  $Z$ , then

$$\mathbb{E}[XY|Z] = \mathbb{E}[X|Z]\mathbb{E}[Y|Z].$$

### Random Processes

A random process  $X(t)$  is a generalization of a random variable. For a random variable, each experiment leads to a number (or a vector), while for a random process, each experiment leads to a function. For a fixed outcome  $\omega \in \Omega$ , one obtains the function  $X(t, \omega)$ , which is also called the *sample path* or *sample function* of the process. For a fixed  $t$ ,  $X(t, \omega)$  is a random variable with the underlying probability space  $\Omega$ . The *mean process* of  $X(t)$  is the time function  $\mathbb{E}[X(t)]$ . The autocorrelation of  $X(t)$  is  $\mathbb{E}[X(t_1)X(t_2)^T]$  and the autocovariance of  $X(t)$  is  $\mathbb{E}[(X(t_1) - m(t_1))(X(t_2) - m(t_2))^T]$ .

A random process  $X(t)$  is called a *Gaussian random process* if for any finite set  $\{t_1, t_2, \dots, t_N\}$ , the random variables  $\{X(t_1), X(t_2), \dots, X(t_N)\}$  have a joint Gaussian distribution, i.e., their joint probability density function is given by

$$p_X(x) = \frac{1}{(2\pi)^{N/2} \sqrt{\det[\mathcal{C}_X]}} \exp\left[-\frac{1}{2}(x - m_X)^T \mathcal{C}_X^{-1}(x - m_X)\right] \quad (2.1)$$

where  $m_X = [m_X(t_1) \ m_X(t_2) \ \dots \ m_X(t_N)]^T$  is the mean vector and  $\mathcal{C}_X = [\text{cov}(X(t_i), X(t_j))]$  is the covariance matrix. Gaussian processes have the following properties.

**Theorem 2.1.** *Let  $X(t)$  be a Gaussian process. Then  $X(t)$  is completely determined by  $m_X$  and  $\mathcal{C}_X$ .*

**Theorem 2.2.** *Let  $X$  and  $Y$  have a joint Gaussian distribution with mean and covariance given by*

$$\mu = \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} \Sigma_x & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_y \end{bmatrix}.$$

*Then  $X$  conditioned on  $Y = y$  is Gaussian with mean and covariance given by*

$$\mu_{X|Y=y} = \bar{x} + \Sigma_{xy}\Sigma_y^{-1}(y - \bar{y}) \quad \text{and} \quad \Sigma_{X|Y=y} = \Sigma_x - \Sigma_{xy}\Sigma_y^{-1}\Sigma_{yx}.$$

In other words,

$$\mathbb{E}[X|Y = y] = \bar{x} + \Sigma_{xy}\Sigma_y^{-1}(y - \bar{y}). \quad (2.2)$$

The proof can be found in Anderson and Moore [AM90].

### Stability of stochastic systems

Consider the following system dynamics:

$$x_{k+1} = f(x_k, w_k), \quad (2.3)$$

where  $x_0$  and  $w_k$  are random vectors. System (2.3) is said to be

1. *second moment stable* if

$$\lim_{k \rightarrow \infty} \mathbb{E}[|x_k|^2] = 0,$$

2. *almost sure stable* if

$$P(\lim_{k \rightarrow \infty} \|x_k\| = 0) = 1,$$

where the expectation is taken with respect to  $x_0$  and  $w_i, i = 0, \dots, k$ .

For a convex function  $f$ ,  $x_1, \dots, x_n$  in its domain, and positive weights  $\alpha_i$ , Jensen's inequality can be stated as:

$$f\left(\frac{\sum \alpha_i x_i}{\sum \alpha_i}\right) \leq \frac{\sum \alpha_i f(x_i)}{\sum \alpha_i}. \quad (2.4)$$

Jensen's inequality can also be stated in probabilistic form. Let  $X$  be a random variable and  $f$  be a convex function. Then

$$f(\mathbb{E}[X]) \leq \mathbb{E}(f(X)). \quad (2.5)$$

The above two inequalities are reversed if  $f$  is concave.

## Markov Chains

### 2.2 Optimal Estimation

#### Minimum mean square error estimator

Suppose we wish to know some quantity  $X$ , and we are not able to make a direct and accurate measurement of  $X$ . However we can make some indirect measurement  $Y$  that is related to  $X$ . Our task is to get an "optimal" estimate of  $X$  from  $Y$ .

One question that immediately arises before we attempt to solve the estimation problem is: what is a good estimate and when an estimate is "optimal"?

Intuitively a "good" estimate should make the estimation error  $\hat{X} - X$  "small" since we wish to reconstruct  $X$  as perfectly as possible. An "optimal" estimate should make  $\hat{X} - X$  the "smallest" among all other estimates. Many metrics can be used to define the size of the error  $\hat{X} - X$  (hence we are able to say if it is "small" or not). Since  $\hat{X} - X$  is typically a random variable, the metric that we shall use throughout the book is the following *mean squared error* (MSE)

$$\mathbb{E}[(\hat{X} - X)^T(\hat{X} - X)].$$

Therefore given  $Y = y$  (i.e., the measurement that we take), our task is to construct the optimal estimate  $\hat{X}$  that minimizes

$$\mathbb{E}[(\hat{X} - X)^T(\hat{X} - X)|Y = y].$$

It turns out that the optimal  $\hat{X}$  has a very simple form, given in the following theorem.

**Theorem 2.3.** *The optimal estimate  $\hat{X}^*$  that minimizes*

$$\mathbb{E}[(\hat{X} - X)^T(\hat{X} - X)|Y = y]$$

*is given by the following conditional expectation of  $X$*

$$\hat{X}^* = \mathbb{E}[X|Y = y].$$

*Proof.* We can rewrite  $\mathbb{E}[(\hat{X} - X)^T(\hat{X} - X)|Y = y]$  as follows

$$\begin{aligned} & \mathbb{E}[(\hat{X} - X)^T(\hat{X} - X)|Y = y] \\ &= \mathbb{E}[X^T X|Y = y] - 2\hat{X}^T \mathbb{E}[X|Y = y] + \hat{X}^T \hat{X} \\ &= (\hat{X} - \mathbb{E}[X|Y = y])^T (\hat{X} - \mathbb{E}[X|Y = y]) + \mathbb{E}[X^T X - \mathbb{E}[X]^T \mathbb{E}[X^T]|Y = y]. \end{aligned}$$

Since  $\mathbb{E}[X^T X - \mathbb{E}[X]^T \mathbb{E}[X^T]|Y = y]$  is independent of  $\hat{X}$ , we conclude that

$$\hat{X}^* = \mathbb{E}[X|Y = y].$$

□

$\hat{X}^* = \mathbb{E}[X|Y = y]$  is also called the *minimum mean squared error* (MMSE) estimate of  $X$ .

### Example 2.1 Estimate a Gaussian random variable

Consider the following equation

$$Y = X + N \tag{2.6}$$

where  $X$  and  $N$  are both scalar zero-mean Gaussian random variables with covariances  $\sigma_x$  and  $\sigma_n$  respectively. Further assume  $X$  and  $N$  are uncorrelated. Suppose we make a measurement of  $X$  and get  $y$ . The MMSE estimate of  $X$  is then given by

$$\hat{X} = \mathbb{E}[X|Y = y] = \frac{\sigma_x}{\sigma_x + \sigma_n} y.$$

▽

### Sampling of a continuous-time system

A wide variety of physical systems are modeled in the continuous-time domain. In this book, we focus on continuous-time systems with dynamics of the form

$$\frac{dx}{dt} = A_c x + B_c u + w, \quad y = C_c x + v, \tag{2.7}$$

where  $x(t) \in \mathbb{R}^n$  is the state vector with unknown initial value  $x(0)$ ,  $u(t) \in \mathbb{R}^p$  is the input vector,  $y(t) \in \mathbb{R}^m$  is the observation vector, and  $w(t)$  and  $v(t)$  are process disturbance and measurement noise. We assume  $w(t)$  and  $v(t)$  are mutually uncorrelated zero-mean Gaussian processes with autocovariances

$$\mathbb{E}[w(s)w(t)^T] = \delta_{st}\Sigma_{wc}, \quad \mathbb{E}[v(s)v(t)^T] = \delta_{st}\Sigma_{vc},$$

where  $\delta_{st} = 1$  if  $s = t$  and  $\delta_{st} = 0$  otherwise.

As more and more controllers are implemented digitally, we need a procedure to convert the continuous-time system (2.7) into an equivalent discrete-time system. This procedure is called *sampling* or *discretization*. A frequently seen approach to implement the control law on a digital computer is to use a digital to analogue converter that holds the analog signal until the next time step, called *zero-order-hold* control.

Consider the following periodic sampling scheme: we sample the system (2.7) at time instances  $t = k\tau$ ,  $k = 0, 1, \dots$ , where  $\tau > 0$  is the *sampling period*. It can be shown (see Astrom-Wittenmark) that the equivalent discrete-time system of (2.7) is given by

$$x_{k+1} = Ax_k + Bu_k + w_k, \quad y_k = Cx_k + v_k, \quad (2.8)$$

where  $x_k$  and  $y_k$  correspond to  $x(t)$  and  $y(t)$  at time  $t = k\tau$ , and  $A, B$  and  $C$  are given by

$$A = e^{A_c\tau}, \quad B = \int_0^\tau e^{A_c t} dt B_c, \quad C = C_c. \quad (2.9)$$

In the discrete-time setting, the process and measurement noises are also uncorrelated zero-mean Gaussian random processes with covariance

$$\mathbb{E}[w_s w_k] = \delta_{sk} \Sigma_w, \quad \mathbb{E}[v_s v_k^T] = \delta_{sk} \Sigma_v,$$

where

$$\Sigma_w = \int_0^\tau e^{A_c t} \Sigma_{wc} e^{A_c^T t} dt, \quad \Sigma_v = \Sigma_{vc}.$$

Computing  $\Sigma_w$  directly from the above formula is sometimes difficult due to the integral of matrix exponentials. An easier approach to compute it is given as follows. Define  $M$  and  $N$  as

$$M = \begin{bmatrix} -A_c & \Sigma_{wc} \\ 0 & A_c^T \end{bmatrix} \tau, \quad N = e^M.$$

Then can be shown that

$$N = \begin{bmatrix} * & X^{-1} \Sigma_w \\ 0 & X^T \end{bmatrix}.$$

Therefore  $\Sigma_w$  can be computed from

$$\Sigma_w = (X^T)^T X^{-1} \Sigma_w,$$

i.e.,  $\Sigma_w$  is obtained by multiplying the transpose of the lower-right submatrix of  $N$  with the upper-right submatrix of  $N$ .

Most of the results developed in this book also extend to cases where the sensor measurement  $y_k$  involves a direct input term, i.e.,

$$y_k = Cx_k + Du_k + v_k. \quad (2.10)$$

For simplicity, we shall use the system model as described by (2.8) for the remainder of the book unless otherwise explicitly stated.

### Kalman filtering

Consider the following system as described by equation(2.8):

$$x_{k+1} = Ax_k + Bu_k + w_k, \quad y_k = Cx_k + v_k, \quad (2.11)$$

where  $x_k \in \mathbb{R}^n$  is the state vector with unknown initial value  $x_0$ ,  $u_k \in \mathbb{R}^p$  is the input vector,  $y_k \in \mathbb{R}^m$  is the observation vector, and  $w_k$  and  $v_k$  are process and measurement noises (or disturbances).

Clearly nothing can be said on any estimator without defining a structure on  $w_k$  and  $v_k$ . In this book, we are particularly interested in  $w_k$  and  $v_k$  that have the following properties:

- $w_k$  and  $v_k$  are zero-mean Gaussian random vectors;
- $\mathbb{E}[w_k w_j^T] = \delta_{kj} \Sigma_w$  with  $\Sigma_w \geq 0$ ;
- $\mathbb{E}[v_k v_j^T] = \delta_{kj} \Sigma_v$  with  $\Sigma_v > 0$ ;
- $\mathbb{E}[w_k v_j^T] = 0 \forall j, k$ ,

where  $\delta_{kj} = 0$  if  $k \neq j$  and  $\delta_{kj} = 1$  otherwise. We also assume the initial value  $x_0$  of system (2.11) is a zero-mean Gaussian random vector that is uncorrelated with  $w_k$  and  $v_k$  for all  $k \geq 0$ . The covariance of  $x_0$  is given by  $\Pi_0 \geq 0$ . Furthermore we assume  $(A, \sqrt{Q})$  is stabilizable.

Let  $Y_k = \{y_0, y_1, \dots, y_k\}$  be the measurements available at time  $k$  and  $U_k = \{u_0, u_1, \dots, u_k\}$  be the input applied to the system up to time  $k$ . We are interested in looking for the MMSE  $\hat{x}_k$  of  $x_k$  at each time  $k \geq 0$  given  $Y_k$  and  $U_{k-1}$ . From Theorem 2.3, we know that  $\hat{x}_k$  is given by

$$\hat{x}_k = \mathbb{E}[x_k | Y_k, U_{k-1}], \quad (2.12)$$

and the corresponding error covariance  $P_k$  is given by

$$P_k = \mathbb{E}[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T | Y_k, U_{k-1}]. \quad (2.13)$$

Calculating  $\hat{x}_k$  and  $P_k$  according to equation (2.12) and (2.13) is not trivial and is computationally intensive as  $k$  increases. The celebrated Kalman filter provides a simple and elegant way to compute  $\hat{x}_k$  and  $P_k$  recursively.

The Kalman filter [Kal60] is a well-established methodology for model-based fusion of sensor data [GA93, Gus00, May79, KSH00, AM90] that has played a central role in systems theory and has found wide applications in many fields such as control, signal processing, and communications.

Assume that  $\hat{x}_{k-1}$  and  $P_{k-1}$  defined as in equation (2.12) and (2.13) are available. Consider the one-step state prediction  $\hat{x}_{k|k-1}$  (also called the *a priori state estimate*) given by

$$\hat{x}_{k|k-1} = \mathbb{E}[x_k | Y_{k-1}, U_{k-1}]$$

and the associated estimation error covariance (also called the *a priori error covariance*)  $P_{k|k-1}$  given by

$$P_{k|k-1} = \mathbb{E}[(x_k - \hat{x}_{k|k-1})(x_k - \hat{x}_{k|k-1})^T | Y_{k-1}, U_{k-1}].$$

From (2.11), we have

$$\begin{aligned} \hat{x}_{k|k-1} &= \mathbb{E}[x_k | Y_{k-1}, U_{k-1}] \\ &= \mathbb{E}[Ax_{k-1} + Bu_{k-1} + w_{k-1} | Y_{k-1}, U_{k-1}] \\ &= A\hat{x}_{k-1} + Bu_{k-1}, \end{aligned} \tag{2.14}$$

where we use the fact that  $w_{k-1}$  is independent of any  $y_t$  ( $t \leq k-1$ ) and the expectation operator is linear. Consequently,

$$P_{k|k-1} = AP_{k-1}A^T + \Sigma_w. \tag{2.15}$$

Now consider  $y_k$  conditioned on  $Y_{k-1}$  and  $U_{k-1}$  which has mean

$$\mathbb{E}[y_k | Y_{k-1}, U_{k-1}] = \mathbb{E}[Cx_k + v_k | Y_{k-1}, U_{k-1}] = C\hat{x}_{k|k-1}$$

and covariance

$$\mathbb{E}[(y_k - \mathbb{E}[y_k])(y_k - \mathbb{E}[y_k])^T | Y_{k-1}, U_{k-1}] = CP_{k|k-1}C^T + \Sigma_v,$$

where we have used the fact that  $v_k$  is independent of  $Y_{k-1}$ . The cross covariance of  $x_k$  and  $y_k$  conditioned on  $Y_{k-1}$  and  $U_{k-1}$  is given by

$$\mathbb{E}[(x_k - \mathbb{E}[x_k])(y_k - \mathbb{E}[y_k])^T | Y_{k-1}, U_{k-1}] = P_{k|k-1}C^T.$$

From the above analysis, we see that the random vector  $[x'_k \ y'_k]'$  conditioned on  $Y_{k-1}$  and  $U_{k-1}$  is Gaussian with mean and covariance

$$\begin{bmatrix} \hat{x}_{k|k-1} \\ C\hat{x}_{k|k-1} \end{bmatrix} \text{ and } \begin{bmatrix} P_{k|k-1} & P_{k|k-1}C^T \\ CP_{k|k-1} & CP_{k|k-1}C^T + \Sigma_v \end{bmatrix}.$$

Therefore from Theorem 2.2,  $x_k$  conditioned on  $y_k$  (and on  $Y_{k-1}$  and  $U_{k-1}$ , i.e., conditioned on  $Y_k$  and  $U_{k-1}$ ) has mean

$$\mathbb{E}[x_k | Y_k, U_{k-1}] = \hat{x}_{k|k-1} + K_k(y_k - C\hat{x}_{k|k-1})$$

and covariance

$$(I - K_kC)P_{k|k-1}$$

where  $K_k = P_{k|k-1}C^T[CP_{k|k-1}C^T + \Sigma_v]^{-1}$  is the so-called *Kalman gain*.

Let us summarize what we have said so far. Given the system (2.11), the MMSE estimate  $\hat{x}_k$  of  $x_k$  is given by  $\hat{x}_k = \mathbb{E}[x_k | Y_k, U_{k-1}]$ , which can be computed recursively as follows

1. *time update:*

$$\begin{aligned}\hat{x}_{k|k-1} &= A\hat{x}_{k-1} + Bu_{k-1}, \\ P_{k|k-1} &= AP_{k-1}A^T + \Sigma_w.\end{aligned}$$

2. *measurement update:*

$$\begin{aligned}K_k &= P_{k|k-1}C^T[CP_{k|k-1}C^T + \Sigma_v]^{-1}, \\ \hat{x}_k &= \hat{x}_{k|k-1} + K_k(y_k - C\hat{x}_{k|k-1}), \\ P_k &= (I - K_kC)P_{k|k-1}.\end{aligned}$$

The initial values of the recursion are set as  $\hat{x}_0 = 0$  and  $P_0 = \Pi_0$ . The Kalman filter essentially consists of the above two update steps.

**Lemma 2.1.** *The Kalman gain  $K_k$  and the error covariance  $P_k$  satisfy*

$$K_k = P_kC^T\Sigma_v^{-1}. \quad (2.16)$$

*Proof.* Since  $P_k = (I - K_kC)P_{k|k-1}$ , it suffices to show

$$(I - K_kC)P_{k|k-1}C^T\Sigma_v^{-1} = K_k$$

which is equivalent to

$$\begin{aligned}P_{k|k-1}C^T\Sigma_v^{-1} &= K_k(I + CP_{k|k-1}C^T\Sigma_v^{-1}) \\ \iff P_{k|k-1}C^T\Sigma_v^{-1} &= P_{k|k-1}C^T[CP_{k|k-1}C^T + \Sigma_v]^{-1}(I + CP_{k|k-1}C^T\Sigma_v^{-1}) \\ \iff \Sigma_v &= (I + CP_{k|k-1}C^T\Sigma_v^{-1})^{-1}(CP_{k|k-1}C^T + \Sigma_v) \\ \iff \Sigma_v &= \Sigma_v(\Sigma_v + CP_{k|k-1}C^T)^{-1}(CP_{k|k-1}C^T + \Sigma_v)\end{aligned}$$

where the last equation holds trivially.  $\square$

Let  $\mathbb{S}_+^n$  be the set of  $n$  by  $n$  positive semi-definite matrices. To simplify the analysis, define the function  $h : \mathbb{S}_+^n \rightarrow \mathbb{S}_+^n$  as

$$h(X) \triangleq AXA^T + \Sigma_w, \quad (2.17)$$

and  $\tilde{g} : \mathbb{S}_+^n \rightarrow \mathbb{S}_+^n$  as

$$\tilde{g}(X) \triangleq X - XC^T[CC^T + \Sigma_v]^{-1}CX. \quad (2.18)$$

Further define  $g : \mathbb{S}_+^n \rightarrow \mathbb{S}_+^n$  as

$$g(X) \triangleq h \circ \tilde{g} = AXA^T + \Sigma_w - AXC^T[CC^T + \Sigma_v]^{-1}CXA. \quad (2.19)$$

For functions  $f, f_1, f_2 : \mathbb{S}_+^n \rightarrow \mathbb{S}_+^n$ ,  $f_1 \circ f_2$  is defined as

$$f_1 \circ f_2(X) \triangleq f_1(f_2(X)), \quad (2.20)$$

and  $f^t$  is defined as

$$f^t(X) \triangleq \underbrace{f \circ f \circ \dots \circ f}_{t \text{ times}}(X). \quad (2.21)$$

With these definitions, it can be verified that in the Kalman filter time update and measurement update equations,  $P_{k+1|k}$  and  $P_{k+1}$  satisfy

$$\begin{aligned} P_{k+1|k} &= h(P_k), \\ P_{k+1|k} &= g(P_{k|k-1}), \\ P_{k+1} &= \tilde{g}(P_{k+1|k}), \\ P_{k+1} &= \tilde{g} \circ h(P_k). \end{aligned}$$

The equation  $g(X) = X$  or

$$AXA^T + \Sigma_w - AXC^T[CC^T + \Sigma_v]^{-1}CXA = X \quad (2.22)$$

is called the *discrete-time algebraic Riccati equation* (DARE). When  $(A, \sqrt{\Sigma_w})$  is stabilizable and  $(A, C)$  is detectable,  $P_k$  converges to a unique positive semi-definite matrix  $\bar{P}$  which satisfies  $\bar{P} = \tilde{g} \circ h(\bar{P})$ .  $\bar{P}$  is called the steady-state error covariance, and it reflects how well the estimate  $\hat{x}_k$  approximates  $x_k$  in the steady state.

### Properties of the Kalman filter

We first introduce a few well-known lemmas without proofs.

**Lemma 2.2** (Matrix Inversion Lemma). *Let  $X > 0$ . If  $X = B^{-1} + CD^{-1}C'$ , then the inverse of  $X$  can be written as*

$$X^{-1} = B - BC(D + C'BC)^{-1}C'B.$$

The second lemma is the Schur Complement lemma. It provides a set of equivalent relationships for a positive definite matrix  $M$ .

**Lemma 2.3** (Schur Complement). *Let*

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}.$$

*Then the following three conditions are equivalent to each other.*

1.  $M > 0$ .
2.  $A > 0$  and  $S_A \triangleq D - CA^{-1}B > 0$ .
3.  $D > 0$  and  $S_D \triangleq A - BD^{-1}C > 0$ .

The last one is the Block Matrix Inversion lemma, which, as its name suggests, inverts a block matrix using the Schur complement of the matrix.

**Lemma 2.4** (Block Matrix Inversion). *Let*

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix} > 0.$$

*Then  $M^{-1}$  can be computed as*

$$M^{-1} = \begin{bmatrix} A^{-1} + A^{-1}BS_A^{-1}CA^{-1} & -A^{-1}BS_A^{-1} \\ -S_A^{-1}CA^{-1} & S_A^{-1} \end{bmatrix},$$

or it can be computed as

$$M^{-1} = \begin{bmatrix} S_D^{-1} & -S_D^{-1}BD^{-1} \\ -D^{-1}CS_D^{-1} & D^{-1} + D^{-1}CS_D^{-1}BD^{-1} \end{bmatrix}.$$

Many useful properties of the functions  $h$ ,  $\tilde{g}$  and  $g$  are presented below.

**Lemma 2.5.** For any  $X, Y \in \mathbb{S}_+^n$ , and  $X \leq Y$ ,

1.  $h(X) \leq h(Y)$ .
2.  $g(X) \leq g(Y)$ .
3.  $\tilde{g}(X) \leq \tilde{g}(Y)$ .
4.  $\tilde{g}(X) \leq X$ .
5.  $g(X) \leq h(X)$ .

When the measurement matrix  $C$  is invertible, the function  $g$  exhibits a very nice property. When we apply  $g$  to any  $X \geq 0$ , we have a bounded value. The following lemma gives this bound.

**Lemma 2.6.** Assume  $C^{-1}$  exists and let  $\bar{M} = C^{-1}RC^{-1'}$ . Then for any  $X \geq 0$ ,  $\tilde{g}(X) \leq \bar{M}$ .

*Proof.* For any  $t > 0$ , we have  $\tilde{g}(t\bar{M}) = \frac{t}{t+1}\bar{M} \leq \bar{M}$ . For all  $X \geq 0$ , since  $\bar{M} > 0$ , it is clear that there exists  $t_1 > 0$  such that  $t_1\bar{M} > X$ . Therefore  $\tilde{g}(X) \leq \tilde{g}(t_1\bar{M}) \leq \bar{M}$ .  $\square$

The steady-state error covariance  $\bar{P}$  has the following property.

**Lemma 2.7.**  $\bar{P} \leq h(\bar{P})$ .

*Proof.* Let  $P^*$  satisfy  $P^* = g(P^*)$ . Then one can verify that  $\bar{P} = \tilde{g}(P^*)$ . Since  $g = h \circ \tilde{g}$ , we have

$$\bar{P} = \tilde{g}(P^*) \leq P^* = g(P^*) = h \circ \tilde{g}(P^*) = h(\bar{P}).$$

$\square$

Let  $0 \leq \lambda \leq 1$ . Consider the following modified DARE.

$$g_\lambda(X) \triangleq AXA^T + \Sigma_w - \lambda AX C^T [CXC^T + \Sigma_v]^{-1} CXA = X. \quad (2.23)$$

The modified DARE will be studied in detail in Chapter 6 and the parameter  $\lambda$  will represent data packet arrival rate. Some preliminary results on the modified DARE are stated in the following lemmas. The proofs are omitted and can be found in the appendix of [SSS<sup>+</sup>03].

**Lemma 2.8.** Let the operator

$$\phi(K, X) = (1 - \lambda)(AXA' + \Sigma_w) + \lambda(FXF' + V) \quad (2.24)$$

where  $F = A + KC$ ,  $V = \Sigma_w + K\Sigma_vK'$ . Assume  $X \in \mathbb{S}_+^n$ ,  $\Sigma_v > 0$ ,  $\Sigma_w \geq 0$ , and  $(A, \Sigma_w^{\frac{1}{2}})$  is controllable. Then the following facts are true:

1. With  $K_X = -AXC'(CX C' + \Sigma_v)^{-1}$ ,  $g_\lambda(X) = \phi(K_X, X)$
2.  $g_\lambda(X) = \min_K \phi(K, X) \leq \phi(K, X), \forall K$
3. If  $X \leq Y$ , then  $g_\lambda(X) \leq g_\lambda(Y)$
4. If  $\lambda_1 \leq \lambda_2$  then  $g_{\lambda_1}(X) \geq g_{\lambda_2}(X)$
5. If  $\alpha \in [0, 1]$ , then  
 $g_\lambda(\alpha X + (1 - \alpha)Y) \geq \alpha g_\lambda(X) + (1 - \alpha)g_\lambda(Y)$
6.  $g_\lambda(X) \geq (1 - \lambda)AXA' + \Sigma_w$
7. If  $\bar{X} \geq g_\lambda(\bar{X})$ , then  $\bar{X} > 0$
8. If  $X$  is a random variable, then  
 $(1 - \lambda)A\mathbb{E}[X]A' + \Sigma_w \leq \mathbb{E}[g_\lambda(X)] \leq g_\lambda(\mathbb{E}[X])$

The next two Lemmas show that when the modified DARE has a solution  $\hat{P}$ , this solution is also stable, i.e., every sequence based on the difference Riccati equation  $P_{t+1} = g_\lambda(P_t)$  converges to  $\hat{P}$  for all initial positive semidefinite conditions  $P_0 \geq 0$ .

**Lemma 2.9.** *Define the linear operator*

$$\mathcal{L}(Y) = (1 - \lambda)(AY A') + \lambda(FY F')$$

Suppose there exists  $\bar{Y} > 0$  such that  $\bar{Y} > \mathcal{L}(\bar{Y})$ .

1. For all  $W \geq 0$ ,

$$\lim_{k \rightarrow \infty} \mathcal{L}^k(W) = 0$$

2. Let  $U \geq 0$  and consider the linear system

$$Y_{k+1} = \mathcal{L}(Y_k) + U \quad \text{initialized at } Y_0.$$

Then, the sequence  $Y_k$  is bounded.

**Lemma 2.10.** *Consider the operator  $\phi(K, X)$  defined in equation (6.28). Suppose there exists a matrix  $\bar{K}$  and a positive definite matrix  $Z$  such that*

$$Z > 0 \quad \text{and} \quad Z > \phi(\bar{K}, Z).$$

Then, for any  $P_0$ , the sequence  $P_t = g_\lambda^t(P_0)$  is bounded, i.e., there exists  $M_{P_0} \geq 0$  dependent of  $P_0$  such that

$$P_t \leq M \quad \text{for all } t.$$

### 2.3 Linear Quadratic Optimal Control

The optimal linear quadratic regulator problem is posed as follows. Consider the process

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

with the initial condition  $x_0$ , where  $x_k \in \mathbf{R}^n$  is the state and  $u_k \in \mathbf{R}^m$  is the control input that needs to be designed to minimize the cost

$$J_K = \sum_{k=0}^K (x_k^T Q x_k + u_k^T R u_k) + x_{k+1}^T P_{k+1} x_{k+1},$$

with  $Q > 0$  and  $R \geq 0$ . If the parameter  $K$  is finite, the problem is termed the finite horizon LQR problem. The case when  $K \rightarrow \infty$  is termed the infinite horizon LQR problem. We shall assume that the pair  $(A, B)$  is controllable. In general, the results given below extend to the case when the matrices  $A$ ,  $B$ ,  $Q$  and  $R$  are time varying.

The solution to the finite horizon problem can be obtained through standard dynamic programming arguments. The following theorem summarizes the results.

**Theorem 2.4.** *Consider the finite horizon LQR problem posed above. The optimal control law is a linear function of the state*

$$u_k = - (B^T P_{k+1} B + R)^{-1} B^T P_{k+1} A x_k,$$

where the matrix  $P_k$  evolves according to the backward Riccati recursion

$$P_k = A^T P_{k+1} A + Q - A^T P_{k+1} B (B^T P_{k+1} B + R)^{-1} B^T P_{k+1} A$$

with the initial condition  $P_{k+1}$ . Moreover, the achieved cost is given by  $x_0^T P_0 x_0$ .

*Proof.* We begin by rewriting the cost function  $J_K$  to identify terms in the cost that depend on  $x_k$  and  $u_k$ :

$$J_K = \sum_{k=1}^{K-1} (x_k^T Q x_k + u_k^T R u_k) + T_k$$

$$T_k = x_k^T Q x_k + u_k^T R u_k + x_{k+1}^T P_{k+1} x_{k+1}.$$

The only term in the cost that can be affected by the choice of  $u_k$  is  $T_k$ . To choose  $u_k$ , we minimize  $T_k$  by a completion of squares argument. We obtain

$$T_k = x_k^T Q x_k + u_k^T R u_k + (Ax_k + Bu_k)^T P_{k+1} (Ax_k + Bu_k)$$

$$= x_k^T P_k x_k + (u_k + S_k^{-1} B^T P_{k+1} A x_k)^T S_k (u_k + S_k^{-1} B^T P_{k+1} A x_k),$$

where

$$S_k = B^T P_{k+1} B + R$$

$$P_k = Q + A^T P_{k+1} A - A^T P_{k+1} B S_k^{-1} B^T P_{k+1} A.$$

Thus, the optimal choice of  $u_k$  is

$$u_k = -S_k^{-1}B^T P_{k+1}x_k.$$

With the optimal choice of  $u_k$  the term  $T_k$  reduces to  $x_k^T P_k x_k$ . Thus, the cost function  $J_K$  can be rewritten as

$$J_K = \sum_{k=1}^{K-2} (x_k^T Q x_k + u_k^T R u_k) + T_{K-1}$$

$$T_{K-1} = x_{K-1}^T Q x_{K-1} + u_{K-1}^T R u_{K-1} + x_{K-1}^T P_{K-1} x_{K-1}.$$

Thus, the problem of choosing  $u_{K-1}$  is formally identical to the problem that we solved above for choosing  $u_k$ , and the same argument can be repeated at any time step recursively. At a general time  $k$ , the control input  $u_k$  given  $r_k = i$  is given by

$$u_k = - (B^T P_{k+1} B + R)^{-1} B^T P_{k+1} A x_k,$$

where the matrix  $P_k$  evolves according to the backward Riccati recursion

$$P_k = A^T P_{k+1} A + Q - A^T P_{k+1} B (B^T P_{k+1} B + R)^{-1} B^T P_{k+1} A$$

with the final condition  $P_{K-1}$ . Moreover, the cost  $J_0$  obtained through this procedure equals  $x_0^T P_0 x_0$ .  $\square$

For the infinite horizon case, we provide the solution without proof below.

**Theorem 2.5.** *Consider the infinite horizon LQR problem posed above. The optimal control law is a linear function of the state*

$$u_k = - (B^T P B + R)^{-1} B^T P A x_k,$$

where the matrix  $P$  is the unique positive semi-definite solution of the Riccati equation

$$P = A^T P A + Q - A^T P B (B^T P B + R)^{-1} B^T P A.$$

Moreover, the achieved cost is given by  $x_0^T P x_0$ .

## 2.4 LQG Problem

The finite horizon Linear Quadratic Gaussian optimal control problem is posed as follows. Consider the process

$$x_{k+1} = A x_k + B u_k + w_k$$

with the initial condition  $x_0$  as zero mean Gaussian, where  $x_k \in \mathbf{R}^n$  is the state,  $u_k \in \mathbf{R}^m$  is the control input that needs to be designed, and  $w_k$  is the process noise modeled Gaussian and white with mean zero and covariance  $\Sigma_w > 0$ . The process is observed using a sensor that generates measurements of the form

$$y_k = C x_k + v_k,$$

where the sensor noise  $v_k$  is modeled Gaussian and white with mean zero and covariance  $\Sigma_v > 0$ . The noise sequences  $\{w(j)\}$ ,  $\{v(j)\}$  and the initial condition  $x_0$  are assumed to be independent. The cost function that needs to be minimized is

$$J_K = \sum_{k=0}^K \mathbb{E}[(x_k^T Q x_k + u_k^T R u_k)] + \mathbb{E}[x_{k+1}^T P_{k+1} x_{k+1}],$$

with  $Q > 0$  and  $R \geq 0$ . The expectation is taken with respect to all the random parameters in the system. The controller at time  $k$  is allowed access to the measurements until time  $k$  and control inputs until time  $k - 1$ . If the parameter  $K$  is finite, the problem is termed the finite horizon LQG problem. As  $K \rightarrow \infty$ , the cost would necessarily diverge. Thus, the infinite horizon LQG problem considers the cost

$$J_\infty = \lim_{K \rightarrow \infty} \frac{1}{K} J_K.$$

We shall assume that the pair  $(A, B)$  is controllable and  $(A, C)$  is observable. In general, the results given below extend to the case when the matrices  $A$ ,  $B$ ,  $C$ ,  $Q$  and  $R$  are time varying.

The solution to the finite horizon problem is provided by the separation principle. The principle essentially states that the optimal control input is calculated as the input in the LQR problem, but with the state  $x_k$  replaced by the minimum mean squared error (MMSE) estimate of the state  $x_k$  based on the measurements until time  $k$  and control inputs until time  $k - 1$ . Note that the estimate can be calculated recursively through the Kalman filter. The following theorem summarizes the results.

**Theorem 2.6.** *Consider the finite horizon LQG problem posed above. The optimal control law is a linear function of the state*

$$u_k = - (B^T P_{k+1} B + R)^{-1} B^T P_{k+1} A \hat{x}_k,$$

where the matrix  $P_k$  evolves according to the backward Riccati recursion

$$P_k = A^T P_{k+1} A + Q - A^T P_{k+1} B (B^T P_{k+1} B + R)^{-1} B^T P_{k+1} A$$

with the final condition  $P_{k+1}$ , and  $\hat{x}_k$  is the MMSE estimate of the state  $x_k$  based on the measurements until time  $k$  and control inputs until time  $k - 1$ , calculated, e.g., using a Kalman filter.

*Proof.* The proof again follows dynamic programming arguments. We begin by rewriting the cost function  $J_K$  to identify terms in the cost that depend on  $x_k$  and  $u_k$ :

$$J_K = \mathbb{E} \left[ \sum_{k=1}^{K-1} (x_k^T Q x_k + u_k^T R u_k) \right] + T_k$$

$$T_k = \mathbb{E} [x_k^T Q x_k + u_k^T R u_k + x_{k+1}^T P_{k+1} x_{k+1}].$$

The only term in the cost that can be affected by the choice of  $u_k$  is  $T_k$ . To choose  $u_k$ , we minimize  $T_k$  by a completion of squares argument. We obtain

$$T_k = \mathbb{E} [x_k^T Q x_k + u_k^T R u_k + (A x_k + B u_k + w_k)^T P_{k+1} (A x_k + B u_k + w_k)]$$

$$= \mathbb{E} [x_k^T P_k x_k + w_k^T P_{k+1} w_k + (u_k + S_k^{-1} B^T P_{k+1} A x_k)^T S_k (u_k + S_k^{-1} B^T P_{k+1} A x_k)],$$

where we have used the fact that the process noise is white (hence  $w_k$  is independent of both  $x_k$  and  $u_k$ ) and zero mean, and have defined

$$\begin{aligned} S_k &= B^T P_{k+1} B + R \\ P_k &= Q + A^T P_{k+1} A - A^T P_{k+1} B S_k^{-1} B^T P_{k+1} A. \end{aligned}$$

Note that the controller does not have access to  $x_k$  and hence the quadratic term cannot be minimized to zero. Instead, the controller estimates (based on the measurements until time  $k$  and control inputs until time  $k-1$ ) the term  $S_k^{-1} B^T P_{k+1} A x_k$  in the MMSE sense. Thus, the optimal choice of  $u_k$  is

$$u_k = -S_k^{-1} B^T P_{k+1} \hat{x}_k.$$

Denote by  $\Lambda_{e,k}$  the error covariance thus obtained. Since the controller utilizes all control inputs until time  $K-1$  while calculating  $u_k$ ,  $\Lambda_{e,k}$  does not depend on the choice of control inputs  $u_0, \dots, u_{K-1}$ . With the optimal choice of  $u_k$ , the term  $T_k$  reduces to  $T_k = \mathbb{E}[x_k^T P_k x_k + w_k^T P_{k+1} w_k + \Lambda_{e,k}]$ . Thus, the cost function  $J_K$  can be rewritten as

$$\begin{aligned} J_K &= \mathbb{E}\left[\sum_{k=1}^{K-2} (x_k^T Q x_k + u_k^T R u_k)\right] + T_{K-1} + w_{K-1}^T P_{K-1} w_{K-1} + \Lambda_{e,K-1} \\ T_{K-1} &= \mathbb{E}[x_{K-1}^T Q x_{K-1} + u_{K-1}^T R u_{K-1} + x_{K-1}^T P_{K-1} x_{K-1}]. \end{aligned}$$

Now note that the terms  $w_k^T P_{k+1} w_k$  and  $\Lambda_{e,k}$  are not impacted by the choice of  $u[K-1]$  and hence may be dropped from the minimization problem. Thus, the problem of choosing  $u[K-1]$  is formally identical to the problem that we solved above for choosing  $u_k$ , and the same argument can be repeated at any time step recursively. At a general time  $k$ , the control input  $u_k$  given  $r[k] = i$  is given by

$$u_k = - (B^T P_{k+1} B + R)^{-1} B^T P_{k+1} A \hat{x}_k,$$

where the matrix  $P_k$  evolves according to the backward Riccati recursion

$$P_k = A^T P_{k+1} A + Q - A^T P_{k+1} B (B^T P_{k+1} B + R)^{-1} B^T P_{k+1} A$$

with the final condition  $P_{K+1}$ . □

The separation principle also holds for the infinite horizon case. In particular, under the assumptions above, both the backward Riccati recursion in the control calculation and the forward Riccati recursion in the Kalman filter are replaced by the corresponding Riccati equations.

## 2.5 Further Reading

### Exercises

**2.1** Show  $\mathbb{E}[X|Y = y] = \frac{\sigma_x}{\sigma_x + \sigma_n} y$  in Example 2.1.

