COMPSCI.773ST Vision-guided Control

Random Processes and Kalman Filtering

Extended Lecture Notes – 2006

A/Prof. Georgy Gimel'farb
CITR, Tamaki Campus
Department of Computer Science
University of Auckland

Room 731.320
Phone: (09) 373 7599 ext. 86609
E-mail: g.gimelfarb@auckland.ac.nz

Preface

These lecture notes cover in brief the following topics:

- Chapter 1: Random Processes
- Chapter 2: Kalman Filtering

More detail can be found in many available books and journal articles on control theory and signal processing.

Chapter1is based on the texts in [1, 2, 4, 7, 9]. Chapter 2 is based on the lecture notes prepared in 1999 by Dr. Ram Kakarala and on the texts in [2, 3, 5, 6, 8].

The above-mentioned books and articles use different notation for the same quantities, so that it is little wonder that the notation below may differ from the one more familiar to you. Sometimes, when this creates no difficulties, the same character may denote different quantities, but in any case the notation involved is explicitly explained in each section.

Contents

Chapter 1

Random Signals and Processes

Basic goal of (digital) signal processing is to solve the following two problems:

1. to determine actual parameters of a signal distorted by noise or transmission channel and

2. to find how characteristics of a channel effect signal parameters.

A temporal *continuous signal* is a real- or complex-valued oscillation in time $f(t)$, defined as a certain function of the continuous real time variable t. A spatial continuous signal is a real- or complex-valued oscillation $f(v)$ in space, defined as a certain function of the real spatial variable v. Continuous signals are frequently referred to as *analog* signals if they have a continuum of values for each variable t or v. Unless otherwise specified, we assume that signals are complex-valued variables because complex-valued data are more and more extensively used in modern signal processing.

A *discrete signal* is an arbitrary function $f[n]$ that is a sequence of real- or complex-valued numbers defined for all integer values n. A continuous time function $f(t)$, digitised at uniform intervals of T seconds, or a continuous spatial function $g(v)$, digitised at uniform linear space of V meters (or other spatial units), produce the discrete sequences $f[n] = f(nT)$ and $g[n] = g(nS)$, respectively.

We will denote continuous and discrete functions by parentheses () and brackets [] to better discriminate between them and avoid ambiguities for continuous and discrete functions denoted by the same letters.

A continuous or discrete signal that can take only a finite number of values (instead of the continuum) for each time value t or space value s is called a *digital* signal. In practice, the discrete signals are also the digital ones (we use analog-digital converters, or ADCs, to form the digital signals and then store them in a digital memory

with a fixed word length).

Both useful signals and nuisance noise in signal processing are most frequently described by probabilistic models of *random* (or *stochastic*) processes in time or space. This is the only really fruitful way to describe noise or noiselike signals, and it begins with a probabilistic description and then proceeds to derive the associated signal parameters from the probabilistic model.

1.1 Probabilistic models: basic definitions

This section overviews in brief basic components of probabilistic description of random signals. More in detail these topics are discussed in [2, 4, 9].

Let e be a discrete event that belongs to a finite set of possible outcomes of a random experiment. *Probability of the event*, $\Pr(e)$, can be intuitively considered as the limit of ratio of the number of the events e that actually happen to the total number of experiments. It is evident that $0 \leq \Pr(e) \leq 1$.

Random variable X is a variable that randomly takes values x from a continuum of possible values. Probabilities of different values are quantitatively described by a *probability distribution function* $F(x) = \Pr(X \leq x)$. It allows to define the *probability density function*, or simply *probability density*,

$$p(x) = \frac{dF(x)}{dx} \ .$$

In the discrete case, a discrete random variable x takes values from a finite or countable set of possible values, $\mathbf{X} = \{x_1, x_2, \ldots\}$. Each value $x_i$is taken with a probability $p(x_i)$, and the probability distribution is represented by the *cumulative probability function*,

$$\Pr(X \leq x) = \sum_{x_i \leq x} p(x_i). \tag{1.1}$$

Mathematical expectation, $\mathcal{E}\{X\}$, of a random variable X is:

$$\mathcal{E}\{X\} = \int_{-\infty}^{\infty} xp(x)dx. \tag{1.2}$$

It is called also the average, or mean value of the random variable X, or the first moment of X. If the number of observations is growing, the average value of x in the number of observations converges in a probabilistic sense to the mathematical

expectation.

The mathematical expectation of a function of the random variable, e.g., the function $g(x)$, can be found using the probability density $p(x)$ of a continuous variable:

$$\mathcal{E}\left\{g(X)\right\} = \int_{-\infty}^{\infty} g(x)p(x)dx \tag{1.3}$$

or the probabilities $p(x_i)$ of a discrete variable:

$$\mathcal{E}\left\{g(X)\right\} = \sum_{x\in\mathbf{X}} g(x)p(x). \tag{1.4}$$

The mathematical expectation of the squared modulo of X,

$$\mathcal{E}\{|x|^2\} = \int_{-\infty}^{\infty} |x|^2 p(x)dx, \tag{1.5}$$

is called the *mean square*, or the second moment, of the random variable x.

Variance σ^2 of the random variable x is the mean square deviation of this variable from its mean value:

$$\mathrm{var}\{x\} = \int_{-\infty}^{\infty} |x - \mathcal{E}\left\{x\right\}|^2 p(x)dx = \mathcal{E}\left\{|x|^2\right\} - |\mathcal{E}\left\{x\right\}|^2. \tag{1.6}$$

The mean square and variance are the same only for a random variable with the zero mean value.

An interrelation between two random variables is described by their *covariance*:

$$\begin{aligned} \mathrm{cov}\{xy\} &= \mathcal{E}\left\{(x - \mathcal{E}\left\{x\right\})(y^* - \mathcal{E}\left\{y\right\}^*)\right\} \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - \mathcal{E}\left\{x\right\})(y^* - \mathcal{E}\left\{y\right\}^*)p(x,y)dxdy \\ &= \mathcal{E}\left\{xy^*\right\} - \mathcal{E}\left\{x\right\}\mathcal{E}\left\{y\right\}^* \end{aligned} \tag{1.7}$$

where $p(x, y)$ is the joint probability density of the random variables x and y, and $\mathcal{E}\left\{xy\right\}$ denotes the product second moment of these variables.

For simplicity, we will not use below the different notation X and x for the random variable and its value, respectively.

1.2 Uniform and Gaussian probability distributions

Most popular probability densities are the uniform and the Gaussian (normal) ones. The *uniform distribution* of a real random variable x in a finite real interval $a \leq x \leq b$ is described by the uniform (constant) probability density

$$p(x) = \frac{1}{b-a} \; . \tag{1.8}$$

The *Gaussian distribution* of a real random variable x having the mean value $\bar{x} \equiv \mathcal{E}\{x\}$ and the variance $\sigma^2 \equiv \mathrm{var}\{x\}$ is described by the probability density

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \; \exp\left(- \frac{1}{2\sigma^2} \; (x-\bar{x})^2 \right) \tag{1.9}$$

for $-\infty < x < \infty$.

The *multivariate* Gaussian probability density for a vector of K random variables $\mathbf{x} = [x_1, x_2, \ldots, x_K]^T$ is as follows:

$$p(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^K \, |\mathbf{C}_K|}} \; \exp\left(- \frac{1}{2} \; (\mathbf{x}-\bar{\mathbf{x}})^T \mathbf{C}_K^{-1} (\mathbf{x}-\bar{\mathbf{x}}) \right) . \tag{1.10}$$

Here, $\bar{\mathbf{x}}$ is the K-element mean vector

$$\bar{\mathbf{x}} = \mathcal{E}\{\mathbf{x}\} \tag{1.11}$$

and $\mathbf{C}_K$ is the $K \times K$ covariance matrix

$$\begin{aligned} \mathbf{C}_K &= \mathcal{E}\left\{ (\mathbf{x}-\bar{\mathbf{x}})(\mathbf{x}-\bar{\mathbf{x}})^T \right\} \\ &= \begin{pmatrix} \mathrm{var}\{x_1\} & \mathrm{cov}\{x_1 x_2\} & \ldots & \mathrm{cov}\{x_1 x_K\} \\ \mathrm{cov}\{x_2 x_1\} & \mathrm{var}\{x_2\} & \ldots & \mathrm{cov}\{x_2 x_K\} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ \mathrm{cov}\{x_K x_1\} & \mathrm{cov}\{x_K x_2\} & \ldots & \mathrm{var}\{x_K\} \end{pmatrix} . \end{aligned} \tag{1.12}$$

The Gaussian probability density can also be defined for complex-valued random variables. Let $x = x_\mathrm{r} + j x_\mathrm{i}$ be a complex-valued random variable having the real, x_r, and the imaginary, x_i, parts which are real random variables with the mean values $\bar{x}_\mathrm{r}$ and $\bar{x}_\mathrm{i}$, respectively, and the same variance, $\frac{\sigma_x^2}{2}$. The random variables x_r and x_i are assumed to be statistically independent, that is, $\mathrm{cov}\{x_\mathrm{r} x_\mathrm{i}\} = 0$. Then the Gaussian

probability density of the complex-valued random variable can be represented as a particular case of the *bivariate* Gaussian probability density of the two real random variables:

$$\mathbf{x} = \begin{pmatrix} x_\mathrm{r} \\ x_\mathrm{i} \end{pmatrix}; \quad \bar{\mathbf{x}} = \begin{pmatrix} \bar{x_\mathrm{r}} \\ \bar{x_\mathrm{i}} \end{pmatrix}; \quad \mathbf{C}_2 = \begin{pmatrix} \frac{\sigma_x^2}{2} & 0 \\ 0 & \frac{\sigma_x^2}{2} \end{pmatrix}. \tag{1.13}$$

By using these values in Eq. (1.10) for $K = 2$, one obtains after some simplification that

$$p(x) = p(\mathbf{x}) = \frac{1}{\pi\sigma_x^2} \exp\left(-\frac{1}{\sigma_x^2} |x - \bar{x}|^2\right) \tag{1.14}$$

where $\bar{x} = \mathcal{E}\{x\} = \bar{x}_\mathrm{r} + j\bar{x}_\mathrm{i}$ and $\mathrm{var}\{x\} = \mathrm{var}\{x_\mathrm{r}\} + \mathrm{var}\{x_\mathrm{i}\} = \sigma_x^2$. The multivariate Gaussian probability density for a vector $\mathbf{x}$ of K complex-valued random variables can be represented as the multivariate Gaussian probability density for $2K$ real random variables as follows:

$$p(\mathbf{x}) = \frac{1}{\pi^K |\mathbf{C}_K|} \exp\left(-(\mathbf{x} - \bar{\mathbf{x}})^H \mathbf{C}_K^{-1} (\mathbf{x} - \bar{\mathbf{x}})\right) \tag{1.15}$$

where H denotes the Hermitian transposition (see Appendix B) and $\bar{\mathbf{x}}$ and $\mathbf{C}_K$ are the complex mean value and the Hermitian covariance matrix, respectively:

$$\begin{aligned} \bar{\mathbf{x}} &= \mathcal{E}\{\mathbf{x}\}; \\ \mathbf{C}_K &= \mathcal{E}\left\{(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^H\right\}. \end{aligned} \tag{1.16}$$

Eq. (1.15) is derived under the assumption that real and imaginary components of all complex-valued random variables x_k in the complex-valued random vector $\mathbf{x}$ are independent and have the same variances.

1.3 Deterministic and random processes

A signal is said to be *deterministic* if it is exactly predictable for the time of space span of interest. Examples would be:

- $x(t) = 10\sin(2\pi t)$ – Sine wave
- $x(t) = \begin{cases} 1 & \text{if} \quad t \geq 0 \\ 0 & \text{if} \quad t < 0 \end{cases}$ – Unit step
- $x(t) = \begin{cases} 1 - e^{-t} & \text{if} \quad t \geq 0 \\ 0 & \text{if} \quad t < 0 \end{cases}$ – Exponential response

Such signal are functions in the usual mathematical sense: for each specified numerical value t, the corresponding value of x is exactly determined.

In contrast with a deterministic signal, a *random signal* is not predictable in a deterministic sense. Examples of random signals are

- $X(t) = 10\sin(2\pi t + \theta)$, where θ is a random variable uniformly distributed between 0 and 2π.
- $X(t) = A\sin(2\pi t + \theta)$, where θ and A are independent random variables with known distributions.

Such signals are formally known as *random* or *stochastic* processes (both terms are equivalent and may be used interchangeably).

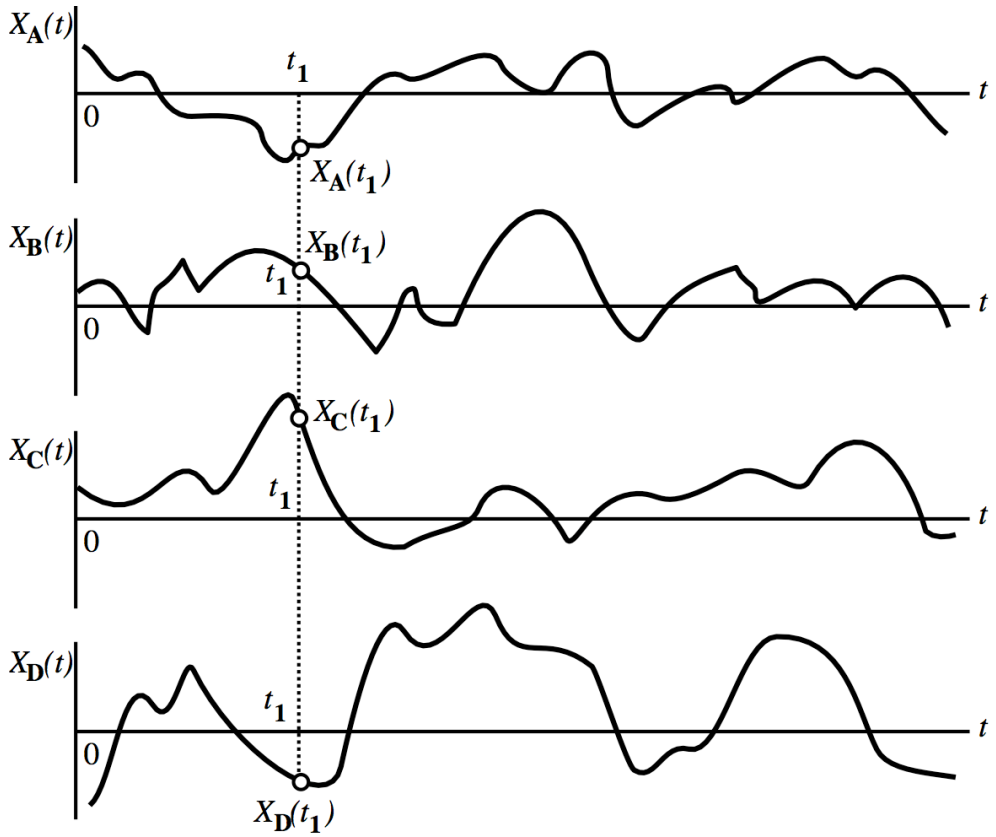


Figure 1.1: Ensemble of sample realizations of a stochastic process.

An *ensemble* of similar random signals is obtained by sampling a stochastic process at a particular point in time, say t_1 (see Figure 1.1). It can be seen that a stochastic

process is a set of random variabless that unfold in time (or space) in accordance with some conceptual chance experiment. Each of the time signals so generated is called a *sample realization* of the process. Samples of the individual signals at a particular time t_1 would then be sample realizations of the random variable $x(t_1)$. Four of these are illustrated in Figure 1.1 as $x_A(t_1)$, $x_B(t_1)$, $x_C(t_1)$, and $x_D(t_1)$. By sampling at a different time, say t_2, we would obtain samples of a different random variable $x(t_2)$ and so forth. Thus in this example an infinite set of random variables is generated by the random process $x(t)$.

In this example, we consider a *continuous-time random process* in that time evolves in a continuous manner and the probability density function describing the amplitude variations is also assumed to be continuous. However, stochastic processes may also be discrete in either time (space) or amplitude.

One way to formally specify a stochastic process is to describe in detail its probabilistic model, or the conceptual chance experiment giving rise to the process. Let $x_i = x(t_i)$; $i = 1, \ldots, k$, be a shorthand notation for the sample values taken at the successive times $t_1 < t_2 < \ldots < t_k$. Obviously, the first-order probability density functions $p_{x_i}(x)$, $i = 1, \ldots, k$, are important in describing the process because they characterize somehow the process amplitude distribution. More detailed features of the process are given by the joint densities relating any pair of random variables, for example, $p_{x_1x_2}(x_1, x_2)$, $p_{x_1x_3}(x_1, x_3)$, and so forth. It is these density functions that tell us something about how rapidly the signal changes with time, and these eventually tell us something about the signal's spectral content. Continuing on, the third, fourth, and subsequent higher-order density functions provide even more detailed information about the process in probabilistic terms. However, this leads to a formidable description of the process, to say the least, because a k-variate density function is required where k can be any positive integer.

Obviously, usually it is impossible to explicitly specify these k-order density functions. Rather, this is done more subtly by providing enough information about the process to enable one to write out any desired higher-order density function; but the actual "writing it out" is usually not done. If so, the description is complete; if not, it is incomplete to some extent, and radically different processes may fit the same incomplete description.

1.3.1 Discrete stochastic processes

A discrete *stochastic process* is a set, or *ensemble*, of real or complex-valued discrete sequences in time (or space) such that each sequence $x[n]$ can be observed as a result of a certain chance experiment. For each fixed time instant, n, the value

$x[n]$ over all the sequences in the set represents a random variable. Probability that the values of $x[n]$ are in a certain interval is given by the distribution function $F(\alpha; n) = \Pr(x[n] \leq \alpha)$ that depends on the observation time, n. The corresponding probability density is $p(\alpha; n) = \frac{\partial F(\alpha;n)}{\partial \alpha}$.

The *mean*, or *expected*, *value* of a stochastic process $x[n]$ at the time instant n is given by the relation

$$\bar{x}[n] = \mathcal{E}\left\{x[n]\right\}. \tag{1.17}$$

The *autocorrelation* of a stochastic process at two different time instants n_1 and n_2 is

$$\gamma_{xx}[n_1, n_2] = \mathcal{E}\left\{x[n_1]x^*[n_2]\right\}. \tag{1.18}$$

It is the so called "engineering" definition of autocorrelation proposed first by N. Wiener. In statistics the autocorrelation is normalised so that its absolute value lies between 0 and 1. The autocorrelation of a centered stochastic process $x[n]$ with the excluded mean value is called the *autocovariance*:

$$c_{xx}[n_1, n_2] = \mathcal{E}\left\{(x[n_1] - \bar{x}[n_1])(x^*[n_2] - \bar{x}^*[n_2])\right\}. \tag{1.19}$$

It can easily be shown that

$$c_{xx}[n_1, n_2] = \gamma_{xx}[n_1, n_2] - \bar{x}[n_1]\bar{x}^*[n_2]. \tag{1.20}$$

If the mean value of a stochastic process is equal to zero for all n, then the autocorrelation and autocovariance of the process coincide, that is,

$$c_{xx}[n_1, n_2] = \gamma_{xx}[n_1, n_2].$$

Although both the terms "autocorrelation" and "autocovariance" are sometimes used as synonyms, they are strictly identical only for the processes with zero mean value.

Two different stochastic processes $x[n]$ and $y[n]$ are described by *cross-correlation*

$$\gamma_{xy}[n_1, n_2] = \mathcal{E}\left\{x[n_1]y^*[n_2]\right\} \tag{1.21}$$

and *cross-covariance*

$$\begin{aligned} c_{xy}[n_1, n_2] &= \mathcal{E}\left\{(x[n_1] - \bar{x}[n_1])(y^*[n_2] - \bar{y}^*[n_2])\right\} \\ &= \gamma_{xy}[n_1, n_2] - \bar{x}[n_1]\bar{y}^*[n_2]. \end{aligned} \tag{1.22}$$

Two stochastic processes are referred to as the *uncorrelated* ones if $c_{xy}[n_1, n_2] = 0$ for all values n_1 and n_2.

The above definitions involve an explicit dependence of time index. A stochastic process is *wide-sense stationary* if its mean value is constant for all the time indices, that is, is independent of time, and its autocorrelation depends only on the difference of time indices $m = n_2 - n_1$. Two stochastic processes are mutually wide-sense stationary if each process is wide-sense stationary and their cross-correlation depends only on the difference of time indices.

Notice that the wide-sense stationary state is defined *only* in terms of the first and second moments of stochastic processes and the higher-order moments are not considered. Therefore the theory of the wide-sense stationary stochastic processes is actually the theory of the second-order processes.

Gaussian stochastic signals are specified as the signals such that all their joint probability distributions are the Gaussian distributions. Therefore in processing stochastic signals that are specified only by their moments up to the second order (mean values and covariances) we can restrict our consideration only to Gaussian processes because each given process can be replaced by the Gaussian process with the same moments.

A stationary discrete stochastic process $x[n]$ is specified statistically by the constant mean value

$$\bar{x}[n] = \bar{x}, \tag{1.23}$$

the *autocorrelation sequence* which is a function of the difference m between the time indices:

$$\gamma_{xx}[m] = \mathcal{E}\left\{x[n]x^*[n+m]\right\}, \tag{1.24}$$

and the *autocovariance sequence*

$$c_{xx}[m] = \mathcal{E}\left\{(x[n] - \bar{x})(x^*[n+m] - \bar{x}^*\right\} = \gamma_{xx}[m] - |\bar{x}|^2. \tag{1.25}$$

Mutually stationary discrete stochastic second-order processes $x[n]$ and $y[n]$ are specified statistically by the *cross-correlation sequence* which is a function of the difference m between the time indices:

$$\gamma_{xy}[m] = \mathcal{E}\left\{x[n]y^*[n+m]\right\}, \tag{1.26}$$

and the *cross-covariance sequence*

$$c_{xy}[m] = \mathcal{E}\{(x[n] - \bar{x})(y^*[n+m] - \bar{y}^*\} = \gamma_{xy}[m] - |\bar{x}||\bar{y}^*|. \tag{1.27}$$

The auto- and cross-correlation sequences have the following useful features:

$$\begin{array}{lcl} \gamma_{xx}[0] & \geq & |\gamma_{xx}[m]|; \\ \gamma_{xx}[-m] & = & \gamma^*_{xx}[m]; \\ \gamma_{xx}[0]\gamma_{yy}[0] & \geq & |\gamma_{xy}[m]|^2; \\ \gamma_{xy}[-m] & = & \gamma^*_{xy}[m], \end{array} \tag{1.28}$$

which hold for all integer values m.

The *power density spectrum* or *spectral density function* is defined as the discrete Fourier transform (DFT) of the autocorrelation sequence:

$$\Gamma_{xx}(\omega) = T \sum_{m=-\infty}^{\infty} \gamma_{xx}[m] \exp(-j\omega mT) \tag{1.29}$$

where ω has the usual meanung of (2π)(frequency in hertz). This function describes how the power of a stochastic process is distributed with frequency. The power density spectrum has the limited bandwidth $\pm\frac{1}{2T}$ Hz and is a periodic function of frequency with the period of $\frac{1}{T}$ Hz. The inverse DFT

$$\gamma_{xx}[m] = \int_{-\frac{1}{2T}}^{\frac{1}{2T}} \Gamma_{xx}(\omega) \exp(j\omega mT) \frac{d\omega}{2\pi} \tag{1.30}$$

allows to show that the autocorrelation with zero time shift $m = 0$ is as follows

$$\gamma_{xx}[0] = \int_{-\frac{1}{2T}}^{\frac{1}{2T}} \Gamma_{xx}(\omega) \frac{d\omega}{2\pi}. \tag{1.31}$$

It can be shown that the autocorrelation of Eq. (1.31) gives the *average power* of the stochastic process. As follows from Eq. (1.31), the area under the curve $\Gamma_{xx}(\omega)$ gives also the average power. Therefore, $\Gamma_{xx}(\omega)$ is the distribution of power as a function of frequency. For this reason, $\Gamma_{xx}(\omega)$ is called the *power density spectrum* (PDS) of the stochastic process.

The two Fourier transforms in Eqs. (1.29) and (1.30) are usually called the discrete-time *Wiener–Khinchine Theorem*. Because $\gamma_{xx}[-m] = \gamma^*_{xx}[m]$, the PDS has to be a strictly real positive function. If the autocorrelation sequence is a strictly real function, then $\gamma_{xx}[-m] = \gamma[m]$, and the PDS can be rewritten using the cosine Fourier transform:

$$\Gamma_{xx}(\omega) = 2T \sum_{m=0}^{\infty} \gamma_{xx}[m] \cos(\omega mT)$$

so that $\Gamma_{xx}(\omega) = \Gamma_{xx}(-\omega)$, and the PDS is a symmetric function.

The *cross-PDS* of two mutually stationary processes $x[n]$ and $y[n]$ is defined as the discrete-time Fourier transform of the cross-correlation sequence

$$\Gamma_{xy}(\omega) = T \sum_{m=-\infty}^{\infty} \gamma_{xy}[m] \exp(-j\omega mT). \tag{1.32}$$

Because $\gamma_{xy}[-m] \neq \gamma^*_{xy}[m]$, the cross-PDS is generally a complex-valued function. But it holds that $\Gamma_{xy}(\omega) = \Gamma^*_{yx}(\omega)$.

1.3.2 White noise and white sequence

White noise is defined to be a stationary continuous random process having a constant PDS:

$$\Gamma_{ww}(\omega) = C \tag{1.33}$$

where C is the white-noise spectral amplitude. The term "white" is a carryover from optics where white light contains all visible frequencies. The corresponding autocorrelation function for white noise is then

$$\gamma_{ww}(\tau) = C\delta(\tau) \tag{1.34}$$

where $\delta(\tau)$ is the delta-function. Because the PDS is constant for all frequencies the white noise has infinite variance. Qualitatively, white noise is jumping around infinitely far, infinitely fast! Obviously this is physically impossible, but it is a useful abstraction: all physical systems are bandlimited to some extent, and a bandlimited system driven by white noise yields a process that has finite variance; that is, the end result makes sense.

Bandlimited white noise is a random process whose spectral amplitude is constant over a finite range of frequencies, and zero outside that range. If the bandwidth includes the origin (sometimes called baseband), then

$$\Gamma_{ww}(\omega) = \begin{cases} C & \text{if} \quad |\omega| \leq 2\pi W \\ 0 & \text{if} \quad |\omega| > 2\pi W \end{cases} \tag{1.35}$$

where W is the physical bandwidth in hertz (Hz). The corresponding autocorrelation function is

$$\gamma_{ww}(\tau) = 2WC \frac{\sin(2\pi W\tau)}{2\pi W\tau} \tag{1.36}$$

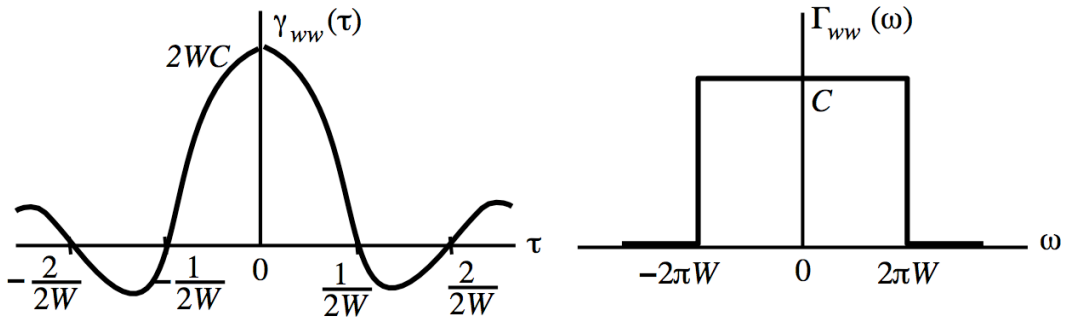


Figure 1.2: Baseband bandlimited white noise: autocorrelation function $\gamma_{ww}(\tau)$ and PDS $\Gamma_{ww}(j\omega)$.

Both the autocorrelation and spectral density functions for baseband bandlimited white noise are sketched in Figure 1.2.

It is of interest to note that the autocorrelation function for baseband bandlimited noise is zero for $\tau = \frac{1}{2W}, \frac{2}{2W}, \frac{3}{2W}$, etc. Therefore if the process is sampled at a rate of $2W$ samples per second (sometimes called the *Nyquist rate*), the resulting set of random variables are uncorrelated. Since this usually simplifies the analysis, the white bandlimited assumption is frequently made in bandlimited situations.

The frequency band for bandlimited white noise is sometimes offset from the origin and centered about some center frequency W_0. It is easily verified that the autocorrelation – PDS function pair is as follows:

$$\begin{aligned} \Gamma_{ww}(\omega) &= \begin{cases} C & \text{if} \quad 2\pi W_1 \le |\omega| \le 2\pi W_2 \\ 0 & \text{otherwise} \end{cases} \\ \gamma_{ww}(\tau) &= C\left(2W_2\,\frac{\sin(2\pi W_2\tau)}{2\pi W_2\tau} - 2W_1\,\frac{\sin(2\pi W_1\tau)}{2\pi W_1\tau}\right) \\ &= 2C\Delta_W\,\frac{\sin(\pi\Delta_W\tau)}{\pi\Delta_W\tau}\,\cos(2\pi W_0\tau) \end{aligned} \tag{1.37}$$

where $\Delta_W = W_2 - W_1$ Hz and $W_0 = \frac{W_1+W_2}{2}$ Hz. The bandlimited white noise has a finite mean-square value, and thus it is physically plausible, whereas pure white noise is not. However, the autocorrelation and spectral density functions in the bandlimited case are more complicated than for pure white noise.

The analogous discrete-time process is referred to as a white sequence. A *white sequence* is the stochastic process with zero mean value $w[n] = 0$ which is self-uncorrelated under any time shift except for $m = 0$ when its variance is equal to σ_w^2.

The autocorrelation sequence of the white noise is

$$\gamma_{ww}[m] = \sigma_w^2 \delta[m] \tag{1.38}$$

where $\delta[m]$ is the discrete delta-sequence (see Appendix **??**). Therefore the PDS of the autocorrelation sequence of the white noise satisfies the condition:

$$\Gamma_{ww}(\omega) = T\sigma_w^2, \tag{1.39}$$

that is, it is constant for all the frequencies. If the random variables are also normal, then the sequence is a *Gaussian white sequence*.

1.3.3 Gaussian random process

In the special case of the *Gaussian* or *normal* process an explicit probability density description is both feasible and appropriate because all the density functions describing the process are normal in form. The random variables $\mathbf{x} = (x_1, x_2, \ldots, x_N)^T$ are said to be *jointly normal* or *jointly Gaussian* if their joint probability density function is given by

$$p_X(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\mathbf{C}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} (\mathbf{x} - \bar{\mathbf{x}})^T \mathbf{C}^{-1} (\mathbf{x} - \bar{\mathbf{x}}) \right) \tag{1.40}$$

where $\bar{\mathbf{x}}$ and $\mathbf{C}$ are the mean vector and covariance matrix for $\mathbf{x}$, respectively. The superscripts T and -1 denote matrix transpose and inverse, respectively.

All we have to do is specify the vector random-variable mean and covariance matrix, and the density function is specified. In the case of a Gaussian random process the "variates" are the random variables $x_{t_1}, x_{t_2}, \ldots, x_{t_k}$, where points in time may be chosen arbitrarily. Thus enough information must be supplied to specify the mean and covariance matrix regardless of the choice of $t_1, t_2, \ldots, t_k$.

A *stationary Gaussian process* has for each time index n the Gaussian probability density of Eqs. (1.9) or (1.14) depending on whether it is a real-valued or complex-valued process, respectively. The successive samples $x[n], x[n+1], \ldots, x[n+M]$ will have the joint probability density of Eqs. (1.10) or (1.15) depending on whether these samples are real-valued or complex-valued, respectively. Let us consider, for example, a complex-valued process with zero mean value. In this case a vector $\mathbf{x} = (x[n]\ x[n+1] \ldots\ x[n+M-1])^T$ of the sequential samples has the following joint probability density:

$$p(\mathbf{x}) = \frac{1}{\pi^M |\mathbf{C}_M|} \exp\left(-\mathbf{x}^H \mathbf{C}_M^{-1} \mathbf{x} \right) \tag{1.41}$$

where $\mathbf{C}_M$ denotes the $M \times M$ covariance matrix:

$$\mathbf{C}_M = \gamma_M = \begin{pmatrix} \gamma_{xx}[0] & \gamma^*_{xx}[1] & \dots & \gamma^*_{xx}[M-1] \\ \gamma_{xx}[1] & \gamma_{xx}[0] & \dots & \gamma^*_{xx}[M-2] \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \gamma_{xx}[M-1] & \gamma_{xx}[M-2] & \dots & \gamma_{xx}[0] \end{pmatrix} \tag{1.42}$$

which is the Hermitian autocorrelation one.

1.4 Discrete-time state model

Discrete-time processes may arise either naturally when a sequence of events takes place in discrete steps (e.g., random walk) or from sampling a continuous process at discrete times. Irrespective of how the discretization arises in a physical problem, the discrete state model fits all situations into the following format:

$$\begin{array}{rcl} \mathbf{x}_{k+1} & = & \mathbf{F}_k \mathbf{x}_k + \mathbf{w}_k; \\ \mathbf{y}_k & = & \mathbf{B}_k \mathbf{x}_k \end{array} \tag{1.43}$$

where the following notation is used:

$\mathbf{x}_k$ is a vector state of the process at time t_k, that is, $\mathbf{x}_k = \mathbf{x}(t_k)$;

$\mathbf{F}_k$ is a matrix that relates $\mathbf{x}_k$ to $\mathbf{x}_{k+1}$ in the absense of a forcing function (in the sampled version of a continuous process this is the state transition matrix);

$\mathbf{w}_k$ is a vector whose components are white sequences, and

$\mathbf{B}_k$ is a linear connection matrix between output $\mathbf{y}_k$ and state $\mathbf{x}_k$.

A white sequence is a sequence of zero-mean random variables that are uncorrelated timewise. However, the elements of $\mathbf{w}_k$ may have a mutual nontrivial correlation at any point in time t_k. The covariance matrix associated with $\mathbf{w}_k$ is assumed to be known, and it will be denoted as $\mathbf{Q}_k$. Thus we have

$$\mathcal{E}\left\{\mathbf{w}_k \mathbf{w}_i^T\right\} = \begin{cases} \mathbf{Q}_k & \text{if} \quad i = k; \\ 0 & \text{if} \quad i \neq k \end{cases} \tag{1.44}$$

where the superscript T denotes transpose.

Chapter 2

Kalman Filtering

In most signal-processing applications, the input-output relationship illustrated in Figure 2 is important, e.g., filtering. In some situations, the inputs and outputs are either not obvious or not of interest. For example, in tracking a moving object by radar, we may be interested in monitoring how the object's position and velocity change over time, and not what causes the object to move. The approach usually taken in tracking and navigation problems is the *state-space* approach, and the solutions usually involve recursive filters known as the *Kalman filters*.

2.1 State-space modelling

Suppose that we are observing a process that is changing over time. At any given instant, the process can be described by a vector of quantities that are collectively called the *state* of the process. These quantities could be positions, velocities, temperatures, etc. The evolution of the process over time can be represented as a trajectory in the space of all possible states, as illustrated in Figure 2.1. Let us

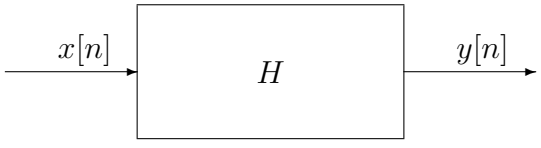


Figure 2.1: Input-output model

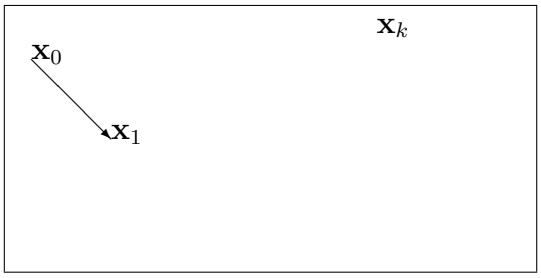


Figure 2.2: Example of trajectory in state-space

denote the state of the process at any time k by an $n \times 1$ vector $\mathbf{x}_k$, where

$$\mathbf{x}_k = \begin{bmatrix} x_k(1) \\ x_k(2) \\ \vdots \\ x_k(n) \end{bmatrix}$$

The components of $\mathbf{x}_k$ are the quantities that collectively describe the process at time k.

The observations that we are making of the process can be considered to be another sequence of vectors, denoted $\mathbf{y}_k$. It may happen that the dimension of each $\mathbf{y}_k$ is $m \times 1$, where $m < n$, the dimension of the state vectors. Generally, the problems we consider in this chapter are those in which we are to estimate the states $\mathbf{x}_k$ from the observations $\mathbf{y}_k$.

If we know enough about the process, we could theoretically describe the evolution of states:

$$\mathbf{x}_{k+1} = a_k(\mathbf{x}_k), \quad \text{for } k = 0, 1, 2, \ldots$$

Here the functions a_k change one state into the next. If we understood the process completely, then the functions $a_k(\cdot)$ would be known, and we could determine the entire trajectory starting from just $\mathbf{x}_0$. In most practical situations, we are not entirely sure how one state changes into the next; the uncertainty may be modelled by a random term $\mathbf{u}_k$ as follows:

$$\mathbf{x}_{k+1} = a_k(\mathbf{x}_k) + \mathbf{u}_k. \tag{2.1}$$

Here $\mathbf{u}_k$ is a $n \times 1$ vector of random components $u_k(1)$, $u_k(2)$, ..., $u_k(n)$.

Suppose now that we represent the relationship between the observations $\mathbf{y}_k$ and the states as follows:

$$\mathbf{y}_k = \mathbf{c}_k(\mathbf{x}_k), \quad \text{for } k = 0, 1, 2, \ldots$$

In most situations, we do not know what the measurement functions $\mathbf{c}_k(\cdot)$ are exactly, or there may be noise affecting the accuracy the measurement. We can represent either possibility with another random term:

$$\mathbf{y}_k = \mathbf{c}_k(\mathbf{x}_k) + \mathbf{v}_k. \tag{2.2}$$

Here, $\mathbf{v}_k$ is a $m \times 1$ vector of random components.

An important special case of the model represented by Eqs. (2.1) and (2.2) is when the functions $a_k(\cdot)$ and $\mathbf{c}_k(\cdot)$ are linear. In this case, they may be represented by matrices, producing the following model:

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{A}_k\mathbf{x}_k + \mathbf{u}_k, \\ \mathbf{y}_k &= \mathbf{C}_k\mathbf{x}_k + \mathbf{v}_k. \end{aligned} \tag{2.3}$$

Each matrix $\mathbf{A}_k$ is $n \times n$, and each $\mathbf{C}_k$ is $m \times n$. These matrices may represent, for example, linear approximations to the corresponding functions in (2.1) and (2.2).

2.1.1 Examples of state-space models

1. Suppose that we are observing a periodic signal in a noisy background. Let us assume that the observations y_k can be represented using a finite Fourier series plus a noise term, as follows:

   $$y_k = c_1 e^{j2\pi f_1 k} + c_2 e^{j2\pi f_2 k} + \cdots + c_n e^{j2\pi f_n k} + u_k.$$

   Here the coefficients c_i are complex numbers. The evolution of the periodic function represented by the complex-exponentials can be described by letting each frequency be the component of a state. Let us define

   $$\mathbf{x}_k = \begin{bmatrix} e^{j2\pi f_1 k} \\ e^{j2\pi f_2 k} \\ \vdots \\ e^{j2\pi f_n k} \end{bmatrix}.$$

   The evolution of the states is then derived by noticing that

   $$x_{k+1}(i) = e^{j2\pi f_i (k+1)} = e^{j2\pi f_i} e^{j2\pi f_i k} = e^{j2\pi f_i} x_k(i).$$

   Therefore, we can write

   $$\mathbf{x}_{k+1} = \mathbf{A}_k \mathbf{x}_k$$

where $\mathbf{A}_k$ is the diagonal matrix

$$\mathbf{A}_k = \begin{bmatrix} e^{j2\pi f_1} & 0 & \dots & 0 \\ 0 & e^{j2\pi f_2} & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \vdots & 0 & e^{j2\pi f_n} \end{bmatrix}.$$

The observation can now be written as

$$\mathbf{y}_k = \mathbf{C}_k \mathbf{x}_k + \mathbf{v}_k$$

where $\mathbf{C}_k$ is the $1 \times n$ vector

$$\mathbf{C}_k = [c_1 \dots c_n].$$

In this problem there is no uncertainty in the state evolution, so $\mathbf{u}_k \equiv 0$.

2. Suppose that an airplane is flying directly away from an airport. The distance d from the airport to the plane is measured by radar. The pilot aims to keep the plane flying at a constant velocity V, but a slight and unpredictable headwind makes this difficult. If d_k denotes the distance at time k, and v_k the velocity, then clearly
$$d_{k+1} = d_k + v_k \Delta t,$$
where Δt is the difference in time between states (in seconds). The velocity v_k changes by an unpredictable amount due to the headwind, and therefore the change may be represented by a random term w_k. If we let the state $\mathbf{x}_k$ of this process be described by a 2-dimensional vector whose entries are $x_k(1) = d_k$ and $x_k(2) = v_k$, then the transition of states can be described as follows:
$$\begin{bmatrix} x_{k+1}(1) \\ x_{k+1}(2) \end{bmatrix} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_k(1) \\ x_k(2) \end{bmatrix} + \begin{bmatrix} 0 \\ w_k \end{bmatrix}.$$
Here, the random term w_k affects only the velocity v_k. We may suppose $v_0 = V$, the desired velocity, and that the probability density of the changes w_k due to the headwind to be centered at zero. The measurements z_k coming from the radar indicate the time delay between the emission of the radar pulse and its return. Since the time delay is obtained as the total distance travelled by the pulse $2d_k$ divided by the speed of propagation c, we may write
$$y_k = \mathbf{C}_k \mathbf{x}_k + v_k,$$
where v_k represents noise due to ground clutter or other factors, and
$$\mathbf{C}_k = \begin{bmatrix} \frac{2}{c} & 0 \end{bmatrix}.$$

2.2 Determining states from observations

The general problem that we now consider, which arises in tracking, navigation, and related applications, is how to estimate the states $\mathbf{x}_k$ from the observations $\mathbf{y}_k$. We may suppose that we have a computer with memory, so all past observations $\mathbf{y}_0$, $\mathbf{y}_1$, ..., $\mathbf{y}_k$ are retained, and can be accessed to produce the estimate of $\mathbf{x}_k$. The estimate is henceforth denoted $\hat{\mathbf{x}}_k$. The goal is to produce estimates $\hat{\mathbf{x}}_k$ that are as close as possible to the states in the sense of minimizing the average value of the error:

$$\sum_{i=1}^{n} |x_k(i) - \hat{x}_k(i)|^2 . \tag{2.4}$$

This problem was considered by Rudolf E. Kalman in 1960, and his elegant solution using a recursive filter has come to be known as the Kalman filter [5, 6]. Today, the Kalman filter is used in hundreds of applications ranging from automobile suspension systems to automatic speech recognition.

Matrix and vector notation

Some basic notation (see also Appendix B) is important to describe Kalman's solution. An n-dimensional vector $\mathbf{x}$ is represented as a column of n components $x(1)$, ..., $x(n)$, which we take to be generally complex-valued. The conjugate or Hermitian transpose of x, denoted x^H, is the $1 \times n$ row vector of conjugate entries $[x^*(1), \ldots, x^*(n)]$, where the asterisk $(*)$ means complex-conjugate. The inner product between two vectors $\mathbf{x}$ and $\mathbf{y}$ of the same dimension is $x^H y$, or in terms of their components,

$$\mathbf{x}^H \mathbf{y} = \sum_{i=1}^{n} x^*(i) y(i).$$

Two vectors are perpendicular if $x^H y = 0$. The length of a vector, denoted $\|x\|$, is computed by the formula

$$\|\mathbf{x}\| = \sqrt{\mathbf{x}^H \mathbf{x}}.$$

The measure of error in Eq. (2.4) can therefore be written as

$$\|\mathbf{x} - \hat{\mathbf{x}}\|^2.$$

The entries in an $m \times n$ matrix $\mathbf{A}$ are written $A(k, \ell)$, where $1 \leq k < m$ denote the row number and $1 \leq \ell \leq n$ denotes the column number. If the conjugate transpose operation H is applied to an $m \times n$ matrix $\mathbf{A}$, then the result, denoted $\mathbf{A}^H$ is an $n \times m$ matrix, whose rows are the columns of $\mathbf{A}$ conjugated:

$$\mathbf{A}^H(k, \ell) = A^*(\ell, k).$$

Note the law of composition for H: for matrices $\mathbf{A}$ and $\mathbf{B}$, we have

$$(\mathbf{AB})^H = \mathbf{B}^H \mathbf{A}^H.$$

Given a $n \times 1$ vector $\mathbf{x}$, and a $m \times 1$ vector $\mathbf{y}$, the outer product xy^H is the $n \times m$ matrix whose entries are as follows:

$$\begin{aligned} \mathbf{x}\mathbf{y}^H &= \begin{bmatrix} x(1) \\ x(2) \\ \vdots \\ x(n) \end{bmatrix} \begin{bmatrix} y^*(1) & y^*(2) & \cdots & y^*(m) \end{bmatrix} \\ &= \begin{bmatrix} x(1)y^*(1) & x(1)y^*(2) & \cdots & x(1)y^*(m) \\ x(2)y^*(1) & x(2)y^*(2) & \cdots & x(2)y^*(m) \\ \vdots & \vdots & \vdots & \vdots \\ x(n)y^*(1) & x(n)y^*(2) & \cdots & x(n)y^*(m) \end{bmatrix}. \end{aligned}$$

Probability concepts

The average or expected value of a random variable x is the integral

$$\mathcal{E}\{x\} = \int_{-\infty}^{\infty} x\ p(x)\ dx$$

where $p(\cdot)$ is the probability density function (p.d.f.) of x. The correlation between two variables x and y is obtained from their joint p.d.f. $p(x, y)$:

$$\mathcal{E}\{xy^*\} = \int_{-\infty}^{\infty} xy^*\ p(x, y)\ dxdy$$

The expected value of a vector $\mathbf{x}$ of random variables is another vector obtained by applying the expectation to each of its components $x(1)$, ..., $x(n)$. The expected value of a matrix $\mathbf{A}$ is the matrix of entries $\mathcal{E}\{A(k, \ell)\}$.

The expected value of the sum $\mathbf{x}+\mathbf{y}$ of two vectors is the sum of the expected values:

$$\mathcal{E}\{\mathbf{x} + \mathbf{y}\} = \mathcal{E}\{\mathbf{x}\} + \mathcal{E}\{\mathbf{y}\}.$$

The expected value of the outer product matrix $\mathbf{x}\mathbf{y}^H$ of two vectors $\mathbf{x}$ and $\mathbf{y}$ is the matrix of entries $\mathcal{E}\{x(k)y^*(\ell)\}$. This matrix is called the *correlation matrix* of $\mathbf{x}$ and $\mathbf{y}$. The correlation matrix of the error $\mathbf{x}_k - \hat{\mathbf{x}}_k$ in Eq. (2.4) is the matrix

$$\mathcal{E}\left\{(\mathbf{x}_k - \hat{\mathbf{x}}_k)(\mathbf{x}_k - \hat{\mathbf{x}}_k)^H\right\}.$$

A pair of vectors $\mathbf{x}$ and $\mathbf{y}$ are called *uncorrelated* if $\mathcal{E}\left\{xy^H\right\} = \mathbf{0}$, where $\mathbf{0}$ is the matrix of appropriate dimensions whose entries are all zero.

2.2.1 Kalman's solution

Kalman's solution [5, 6] for estimating the states $\mathbf{x}_k$ starts with the linear recursive model given by Eq. (2.3). Suitable values are chosen for $\hat{\mathbf{x}}_0$ and $\mathbf{P}_0$ (this correlation matrix will be specified later); these are often guesses when no information is available. In addition, certain assumptions are made about the random elements in the model. First, it is assumed that the uncertainty u is completely uncorrelated with measurement noise v. Second, it is assumed that the uncertainty u_k is not correlated with u_ℓ for different instants k and ℓ. The same assumption is also made on measurement noise v_k. These assumptions are listed in mathematical form below:

$$\begin{aligned} \mathcal{E}\left\{\mathbf{u}_k\mathbf{u}_\ell^H\right\} &= \begin{cases} \mathbf{U}_k, & k=\ell \\ \mathbf{0}, & k\neq\ell \end{cases} \\ \mathcal{E}\left\{\mathbf{v}_k\mathbf{v}_\ell^H\right\} &= \begin{cases} \mathbf{V}_k, & k=\ell \\ \mathbf{0}, & k\neq\ell \end{cases} \\ \mathcal{E}\left\{\mathbf{u}_k\mathbf{v}_\ell^H\right\} &= \mathbf{0}, \text{ for all } k,\ell. \end{aligned} \tag{2.5}$$

It is assumed that the correlation matrices $\mathbf{U}_k$ and $\mathbf{V}_k$ are known for all k. With these assumptions, the problem is to provide an estimate $\hat{\mathbf{x}}_k$ from the observations $\mathbf{y}_0, \ldots, \mathbf{y}_k$ to minimize the mean error

$$\mathcal{E}\left\{\|\mathbf{x}_k - \hat{\mathbf{x}}_k\|^2\right\}. \tag{2.6}$$

To make the solution easy to compute, the estimates should be linearly dependent on the observations, i.e., for all k,

$$\hat{\mathbf{x}}_k = \sum_{i=0}^{k} \mathbf{G}_i\mathbf{y}_i.$$

Here the terms $\mathbf{G}_i$ are $n \times m$ matrices. The goal is then to determine the matrices to minimize Eq. (2.6).

Kalman made a brilliant observation about the solution: **Since the states are evolving recursively by Eq. (2.3), so should their estimates.** Suppose we have constructed an optimal linear estimate $\hat{\mathbf{x}}_{k-1}$, based on observations $\mathbf{y}_0, \ldots, \mathbf{y}_{k-1}$. Then we may apply the state-transition matrix $\mathbf{A}_{k-1}$ in Eq. (2.3) to construct the vector

$$\hat{\mathbf{x}}_k^i \stackrel{\text{def}}{=} \mathbf{A}_{k-1}\hat{x}_{k-1}. \tag{2.7}$$

The lefthand side represents the natural evolution of the estimate $\hat{\mathbf{x}}_{k-1}$ by the system dynamics of Eq. (2.3). The superscript i is given to suggest that this is an "intermediate" estimate before constructing $\hat{\mathbf{x}}_k$. In other words, $\hat{\mathbf{x}}_k^i$ is our best guess

of $\hat{\mathbf{x}}_k$ from just the observations $\mathbf{y}_0, \ldots, \mathbf{y}_{k-1}$, i.e., before actually making the observation $\mathbf{y}_k$ at time k. From this vector, we can also make a prediction of $\mathbf{y}_k$ before we actually measure it; by Eq. (2.3) this is the vector

$$\mathbf{C}_k \hat{\mathbf{x}}_k^i.$$

Kalman noted that the optimal solution for $\hat{\mathbf{x}}_k$ should then be a linear combination of $\hat{\mathbf{x}}_k^i$ and the difference between the actual value of $\mathbf{y}_k$ and our prediction of it. Mathematically this gives the formula

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^i + \mathbf{G}_k \left(\mathbf{y}_k - \mathbf{C}_k \hat{\mathbf{x}}_k^i \right). \tag{2.8}$$

To understand this formula, notice that if the term in parentheses $(\cdots)$ is zero, then

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^i = \mathbf{A}_{k-1} \hat{\mathbf{x}}_{k-1},$$

which means that the estimate is evolving purely by what we know about the process.

The $n \times m$ matrix $\mathbf{G}_k$ in Eq. (2.8) is called the *gain matrix* at time k. To find the optimal gain matrix, we substitute for $\hat{\mathbf{x}}_k$ in Eq. (2.6). The problem then becomes one of finding $\mathbf{G}_k$ to minimize the error

$$\mathcal{E} \left\{ \| (\mathbf{x}_k - \hat{\mathbf{x}}_k^i) - \mathbf{G}_k \left(\mathbf{y}_k - \mathbf{C}_k \hat{\mathbf{x}}_k^i \right) \|^2 \right\}. \tag{2.9}$$

Essentially, this means taking the derivative of Eq. (2.9) with respect to the entries of the matrix $\mathbf{G}_k$ and setting the resulting nm equations to zero to find the solution. The solution is described by the following theorem.

Theorem 1 *Let* $\mathbf{a}$, $\mathbf{b}$ *be random vectors. Then the optimal choice of matrix* $\mathbf{G}$ *to minimize*

$$\mathcal{E} \left\{ \| \mathbf{a} - \mathbf{G}\mathbf{b} \|^2 \right\} \tag{2.10}$$

is given by the formula

$$\mathbf{G} = \mathcal{E} \left\{ a b^H \right\} \left(\mathcal{E} \left\{ b b^H \right\} \right)^{-1}, \tag{2.11}$$

where it is assumed that the correlation matrix $\mathcal{E} \left\{ b b^H \right\}$ *is invertible.*

To understand this result, note that if $\mathbf{a} = \mathbf{b}$, then $\mathbf{G} = \mathbf{I}$ (the identity matrix), which makes sense. The proof of this theorem is given in Section 2.2.3.

To apply Theorem 1, let $\mathbf{a} = \mathbf{x}_k - \hat{\mathbf{x}}_k^i$, and $\mathbf{b} = \mathbf{y}_k - \mathbf{C}_k \hat{\mathbf{x}}_k^i$. Then

$$\mathcal{E} \left\{ \mathbf{a}\mathbf{b}^H \right\} = \mathcal{E} \left\{ (\mathbf{x}_k - \hat{\mathbf{x}}_k^i)(\mathbf{y}_k - \mathbf{C}_k \hat{\mathbf{x}}_k^i)^H \right\}.$$

Substituting for $\mathbf{y}_k$ from Eq. (2.3), one obtains

$$\mathcal{E}\left\{\mathbf{a}\mathbf{b}^H\right\} = \mathcal{E}\left\{(\mathbf{x}_k - \hat{\mathbf{x}}_k^i)(\mathbf{C}_k\mathbf{x}_k + \mathbf{v}_k - \mathbf{C}_k\hat{\mathbf{x}}_k^i)^H\right\}.$$

Rearranging terms, we obtain

$$\begin{aligned}\mathcal{E}\left\{\mathbf{a}\mathbf{b}^H\right\} = \mathcal{E}\left\{(\mathbf{x}_k - \hat{\mathbf{x}}_k^i)(\mathbf{x}_k - \hat{\mathbf{x}}_k^i)^H\mathbf{C}_k^H\right\} \\ +\mathcal{E}\left\{(\mathbf{x}_k - \hat{\mathbf{x}}_k^i)\mathbf{v}_k^H\right\}.\end{aligned}$$

Because $\hat{\mathbf{x}}_k^i$ depends only on $\mathbf{y}_0$, ..., $\mathbf{y}_{k-1}$, and therefore includes only the noise terms $\mathbf{v}_i$ for $i < k$, and uncertainties $\mathbf{u}_i$ for $i \leq k$, we can apply the assumptions of Eq. (2.5) to see that the second expectation on the right is zero. (We are taking advantage of the fact that the "new" noise $\mathbf{v}_k$ is uncorrelated with the "old" noise $\mathbf{v}_i$, $i < k$, and also uncorrelated with the system's uncertainty $\mathbf{u}_i$). Therefore, we obtain

$$\mathcal{E}\left\{\mathbf{a}\mathbf{b}^H\right\} = \mathcal{E}\left\{(\mathbf{x}_k - \hat{\mathbf{x}}_k^i)(\mathbf{x}_k - \hat{\mathbf{x}}_k^i)^H\right\}\mathbf{C}_k^H,$$

where we pull $\mathbf{C}_k^H$ out of the expectation because it is not random. We introduce the term $\mathbf{P}_k^i$ to denote the correlation matrix on the right side, and write the equation above as simply

$$\mathcal{E}\left\{\mathbf{a}\mathbf{b}^H\right\} = \mathbf{P}_k^i\mathbf{C}_k^H. \tag{2.12}$$

The matrix $\mathbf{P}_k^i$ is the correlation matrix for the error $\mathbf{x}_k - \hat{\mathbf{x}}_k^i$. Similar considerations lead to a simple form for $\mathcal{E}\left\{\mathbf{b}\mathbf{b}^H\right\}$; with $\mathbf{P}_k^i$ as in the previous paragraph, we obtain that

$$\mathcal{E}\left\{\mathbf{b}\mathbf{b}^H\right\} = \mathbf{C}_k\mathbf{P}_k^i\mathbf{C}_k^H + \mathbf{V}_k. \tag{2.13}$$

Here, $\mathbf{V}_k$ is measurement noise correlation matrix defined in Eq. (2.5).

Exercise 1 *Prove Eq. (2.13) by following the method used for* $\mathcal{E}\left\{\mathbf{a}\mathbf{b}^H\right\}$.

Now applying Theorem 1, we find that by Eqs. (2.12) and (2.13) the optimal choice for the gain matrix $\mathbf{G}_k$ is

$$\mathbf{G}_k = \mathbf{P}_k^i\mathbf{C}_k^H\left(\mathbf{C}_k\mathbf{P}_k^i\mathbf{C}_k^H + \mathbf{V}_k\right)^{-1}, \tag{2.14}$$

where it is assumed that the inverse on the right hand side exists. Every matrix in this equation is known except the correlation matrix $\mathbf{P}_k^i$. Is there some way to compute this recursively, starting from the matrix $\mathbf{P}_0$, which is assumed to be known? Kalman showed that there is, as follows since by definition

$$\mathbf{P}_k^i = \mathcal{E}\left\{(\mathbf{x}_k - \hat{\mathbf{x}}_k^i)(\mathbf{x}_k - \hat{\mathbf{x}}_k^i)^H\right\} \tag{2.15}$$

By Eq. (2.3), we can substitute $\mathbf{x}_k = \mathbf{A}_{k-1}\mathbf{x}_{k-1} + \mathbf{u}_{k-1}$ and get that

$$\mathbf{P}_k^i = \mathcal{E}\left\{(\mathbf{A}_{k-1}\mathbf{x}_{k-1} + \mathbf{u}_{k-1} - \hat{\mathbf{x}}_k^i)(\mathbf{A}_{k-1}\mathbf{x}_{k-1} + \mathbf{u}_{k-1} - \hat{\mathbf{x}}_k^i)^H\right\}.$$

Now substitute from Eq. (2.7) for $\hat{\mathbf{x}}_k^i$ to get that

$$P_k^i = \mathcal{E}\left\{(\mathbf{A}_{k-1}\mathbf{x}_{k-1} + \mathbf{u}_{k-1} - \mathbf{A}_{k-1}\hat{\mathbf{x}}_{k-1})(\mathbf{A}_{k-1}\mathbf{x}_{k-1} + \mathbf{u}_{k-1} - \mathbf{A}_{k-1}\hat{\mathbf{x}}_{k-1})\right\}.$$

Some rearrangement gives

$$\begin{aligned} P_k^i &= \mathbf{A}_{k-1}\mathcal{E}\left\{(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1})(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1})^H\right\}\mathbf{A}_{k-1}^H \\ &+ \mathcal{E}\left\{\mathbf{u}_{k-1}(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1})^H\right\}\mathbf{A}_{k-1}^H \\ &+ \mathbf{A}_{k-1}\mathcal{E}\left\{(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1})\mathbf{u}_{k-1}^H\right\} \\ &+ \mathcal{E}\left\{\mathbf{u}_{k-1}\mathbf{u}_{k-1}^H\right\}. \end{aligned}$$

It is possible to see that the middle two expectations are zero by the assumptions made in Eq. (2.5). Defining $\mathbf{P}_{k-1}$ to be the correlation matrix in the first term, we obtain that

$$\mathbf{P}_k^i = \mathbf{A}_{k-1}\mathbf{P}_{k-1}\mathbf{A}_{k-1}^H + \mathbf{U}_{k-1}, \tag{2.16}$$

where $\mathbf{U}_{k-1}$ is the correlation matrix defined in Eq. (2.5).

This derivation suggests that $\mathbf{P}_k^i$ can be obtained if $\mathbf{P}_{k-1}$ is known. Since we start with $\mathbf{P}_0$ being known, we can get $\mathbf{P}_1^i$ from Eq. (2.16). But to calculate $\mathbf{P}_2^i$, we need $\mathbf{P}_1$. In general, we need another recursive formula to calcuate $\mathbf{P}_k$ once $\mathbf{P}_k^i$ is obtained. Kalman derived this too:

$$\mathbf{P}_k = \mathbf{P}_k^i - \mathbf{G}_k\mathbf{C}_k\mathbf{P}_k^i. \tag{2.17}$$

This is derived by starting from the definition

$$\mathbf{P}_k = \mathcal{E}\left\{(\mathbf{x}_k - \hat{\mathbf{x}}_k)(\mathbf{x}_k - \hat{\mathbf{x}}_k)^H\right\},$$

substituting (2.8) with (2.14) for $\mathbf{G}_k$, and some amount of algebra!

2.2.2 Putting it all together

Now we are in a position to see how the iteration in the Kalman filter works. First, we choose suitable values for $\hat{\mathbf{x}}_0$ and $\mathbf{P}_0$. Then since $\mathbf{P}_0$ is known, we calculate $\mathbf{P}_1^i$ by Eq. (2.16). This then gives $\mathbf{G}_1$ by Eq. (2.14). Now, $\hat{\mathbf{x}}_1^i$ is obtained by Eq. (2.7), and then the complete estimate $\hat{\mathbf{x}}_1$ by Eq. (2.8). Simultaneously, we can get $\mathbf{P}_1$ from Eq. (2.17). The process of finding $\hat{\mathbf{x}}_2$ is similar, starting from $\hat{\mathbf{x}}_1$ and $\mathbf{P}_1$. To find each estimate $\hat{\mathbf{x}}_k$, we need $\hat{\mathbf{x}}_{k-1}$ and $\mathbf{P}_{k-1}$. The process is illustrated in the flowchart shown in Figure 2.2.2.

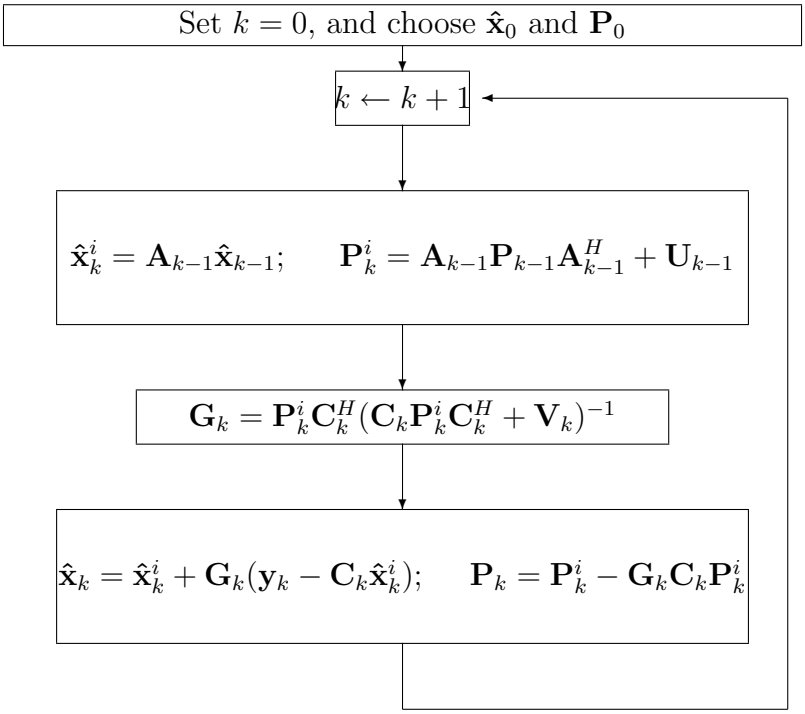


Figure 2.3: Flowchart of Kalman filter algorithm.

Examples

1. Suppose that we have an onedimensional process where the state is not changing, but the measurements are corrupted by noise.

$$\begin{aligned} x_{k+1} &= x_k \\ y_k &= x_k + v_k. \end{aligned}$$

We assume that $\mathcal{E}\{v_k\} = 0$ and $\mathcal{E}\{v_k^2\} = \sigma^2$ for all k, and furthermore $\mathcal{E}\{x_0\} = 0$ and that $\mathcal{E}\{x_0^2\} = \mu^2 = P_0$. Let us derive the Kalman filter equations for this model. The assumptions mean that $A_k = C_k = 1$ for all k, and furthermore that $U_k = 0$ and $V_k = \sigma^2$ for all k. From the flowchart we see that $\hat{x}_k^i = \hat{x}_{k-1}$, and furthermore, we obtain that $P_k^i = P_{k-1}$ for all k. In other words, the intermediate steps are not necessary here, because the state is not changing. Therefore from Eq. (2.14) we get that

$$G_k = \frac{P_{k-1}}{P_{k-1} + \sigma^2}. \tag{2.18}$$

This means from Eq. (2.17) with G_k as in Eq. (2.18) that

$$P_k = P_{k-1} - \frac{P_{k-1}^2}{P_{k-1} + \sigma^2} = \frac{P_{k-1}\sigma^2}{P_{k-1} + \sigma^2}. \tag{2.19}$$

Also, the estimate $\hat{x}_k$ is simply

$$\hat{x}_k = \hat{x}_{k-1} + \frac{P_{k-1}}{P_{k-1} + \sigma^2}\left(y_k - \hat{x}_{k-1}\right). \tag{2.20}$$

To understand this recursion, consider the special case when $\sigma = 0$, i.e., there is no measurement noise. Then $\hat{x}_0 = 0$, and by Eq. (2.20), $\hat{x}_1 = y_1$, $\hat{x}_2 = y_2$, ...; therefore, the estimates are simply the observations here, which is reasonable. Now suppose $\sigma > 0$, but $\mu = 0$, so that x_k is known to be 0 for all k. Then we would expect that all estimates equal x_0, because the states are not changing. Indeed, this is what happens, as by Eqs. (2.18) and (2.19) we get $G_1 = 0$, $P_1 = 0$, and therefore $\hat{x}_1 = 0$; continuing, we see that $P_k = 0$ for all k, and therefore by Eq. (2.20) that $\hat{x}_k = 0$ for all k. The estimates are constant here, even though the measurements z_k are randomly changing.

Now suppose that both $\mu > 0$ and $\sigma > 0$. From (2.19) it can be seen that $P_k < P_{k-1}$, i.e., that the error variance is always decreasing. In fact, it can be seen from Eq. (2.19) that the factor of reduction is

$$\frac{P_k}{P_{k-1}} = \frac{\sigma^2}{P_{k-1} + \sigma^2} < 1.$$

Since $P_0 > 0$, this recursion has the limit $\lim_{k\to\infty} P_k = 0$. This means that the estimate (2.20) changes by less and less each time, so that for k large enough, $\hat{x}_k$ is essentially constant. This is again reasonable, because the state x_k itself is not changing.

2. *Tracking a moving object.* Suppose we have an object moving with an unknown velocity and acceleration. The position of the object is measured by radar, but background noise prevents this measurement from being exact. We now derive the Kalman filter for tracking this object from the noisy position measurements.

   Suppose for simplicity that the object moves in one dimension only. The three-dimensional case is similar, only requiring somewhat more cumbersome notation. The position of the object at any time t, where t is now measured continuously, is not in general a linear function of t. Let us expand the position function, denoted $x(t)$, in a Taylor series. For small h, we can write

$$x(t+h) \approx x(t) + hx'(t) + \frac{h^2}{2}x''t. \tag{2.21}$$

   This equation shows the relationship between position at a future time, current position $x(t)$, velocity $x'(t)$, and acceleration $x''(t)$. We make a discrete approximation to this and construct a state-space model as follows. Let $\mathbf{x}_k$ be a 3×1 state vector whose entries are $x_k(1) = x(k)$ (position), $x_k(2) = x'(k)$ (velocity) and $x_k(3) = x''(k)$ (acceleration). We may apply Eq. (2.21) to obtain this recursion for position:

$$x_{k+1}(1) = x_k(1) + hx_k(2) + \frac{h^2}{2}x_k(3). \tag{2.22}$$

   A similiar recursion for velocity can be obtained by differentiating Eq. (2.21) with respect to h; this yeilds

$$x'(t+h) = x'(t) + hx''(t). \tag{2.23}$$

   In discrete form, this is

$$x_{k+1}(2) = x_k(2) + hx_k(3). \tag{2.24}$$

   A similar analysis gives

$$x_{k+1}(3) = x_k(3). \tag{2.25}$$

Putting Eqs. (2.22), (2.24), and (2.25) together yields the matrix equation

$$\begin{bmatrix} x_{k+1}(1) \\ x_{k+1}(2) \\ x_{k+1}(3) \end{bmatrix} = \begin{bmatrix} 1 & h & \frac{h^2}{2} \\ 0 & 1 & h \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_k(1) \\ x_k(2) \\ x_k(3) \end{bmatrix}. \tag{2.26}$$

If we let $\mathbf{A}_k$ denote the 3×3 matrix above, we obtain the process equation in Eq. (2.3), with the vector $\mathbf{u}_k$ of random terms denoting the uncertainty as to how the object's trajectory is evolving. The measurement of Eq. (2.3) is simple to derive; with $\mathbf{C}_k = [1, 0, 0]$, the observation z_k contains only the noisy measurement of position $x_k(1)$.

The Kalman filter equations are easy to write down following the recipe given in Figure 2.2.2.

2.2.3 Appendix: proof of Theorem 1

Expanding Eq. (2.10) gives

$$\begin{aligned} \mathcal{E}\left\{\|\mathbf{a}-\mathbf{G}\mathbf{b}\|^2\right\} &= \mathcal{E}\left\{(\mathbf{a}-\mathbf{G}\mathbf{b})^H(\mathbf{a}-\mathbf{G}\mathbf{b})\right\} \\ &= \mathcal{E}\left\{(\mathbf{a}^H-\mathbf{b}^H\mathbf{G}^H)(\mathbf{a}-\mathbf{G}\mathbf{b})\right\} \\ &= \mathcal{E}\left\{\mathbf{a}^H\mathbf{a}-\mathbf{b}^H\mathbf{G}^H\mathbf{a}-\mathbf{a}^H\mathbf{G}\mathbf{b}+\mathbf{b}^H\mathbf{G}^H\mathbf{G}\mathbf{b}\right\} \\ &= \mathcal{E}\left\{\mathbf{a}^H\mathbf{a}\right\}-\mathcal{E}\left\{\mathbf{b}^H\mathbf{G}^H\mathbf{a}\right\}-\mathcal{E}\left\{\mathbf{a}^H\mathbf{G}\mathbf{b}\right\}+\mathcal{E}\left\{\mathbf{b}^H\mathbf{G}^H\mathbf{G}\mathbf{b}\right\}. \end{aligned}$$

Differentiating this with respect to $\mathbf{G}$ may seen difficult because $\mathbf{G}$ is a matrix, and both $\mathbf{G}$ and $\mathbf{G}^H$ are appearing. To make it easier, we can treat the elements of $\mathbf{G}$ as independent from the elements of $\mathbf{G}^H$. They are not of course, but it is a mathematical fact (which we won't prove) that the same result is obtained by treating them independently.

The rule for differentiating with respect to matrices is as follows [2]. If $\mathbf{x}$, $\mathbf{y}$ are arbitrary vectors, and $\mathbf{A}$ is a matrix, then

$$\frac{\partial}{\partial A}\mathbf{x}^H\mathbf{A}\mathbf{y} = \mathbf{y}\mathbf{x}^H.$$

The right hand side is a matrix with the same dimension as $\mathbf{A}$, and its i,j-th entries are precisely the entries obtained by differentiating the scalar $\mathbf{x}^H\mathbf{A}\mathbf{y}$ with respect to the i,j-th entry a_{ij} of $\mathbf{A}$.

Using this rule, and differentiating the expression above with respect to $\mathbf{G}^H$ and setting the result equal to zero, we get an equation involving only $\mathbf{G}$:

$$-\mathcal{E}\left\{\mathbf{a}\mathbf{b}^H\right\} + \mathbf{G}\,\mathcal{E}\left\{\mathbf{b}\mathbf{b}^H\right\} = 0.$$

Solving, this gives

$$\mathbf{G} = \mathcal{E}\left\{\mathbf{a}\mathbf{b}^H\right\}\left(\mathcal{E}\left\{\mathbf{b}\mathbf{b}^H\right\}\right)^{-1}.$$

2.2.4 Kalman vs. Wiener filtering

Let a stochastic discrete-time linear system be given by the following equations that specify the state vector $\mathbf{x}$ and the output vector $\mathbf{y}$:

$$\mathbf{x}[n+1] = \mathbf{A}_n\mathbf{x}[n] + \mathbf{u}[n]; \quad \mathbf{y}[n] = \mathbf{C}_n\mathbf{x}[n] + \mathbf{w}[n].$$

Let the matrices $\mathbf{A}_n$ and $\mathbf{C}_n$ specifying the system be known. Let the state noise $\mathbf{u}[n]$ and the observation noise $\mathbf{v}[n]$ be the zero-mean vector white–noise processes ($\mathcal{E}\{\mathbf{u}[k]\} = \mathcal{E}\{\mathbf{v}[k]\} = 0$ with known covariance matrices:

$$\mathcal{E}\left\{\mathbf{u}[k]\mathbf{u}^H[l]\right\} = \mathbf{U}_k\delta(k-l); \;\; \mathcal{E}\left\{\mathbf{v}[k]\mathbf{v}^T[l]\right\} = \mathbf{V}_k\delta(k-l).$$

Let the initial state $\mathbf{x}[0]$ be a random zero-mean vector ($\mathcal{E}\{\mathbf{x}[0]\} = 0$) with known covariance matrix $\mathbf{P}_0$. Let the initial state, state noise, and observation noise be mutually uncorrelated.

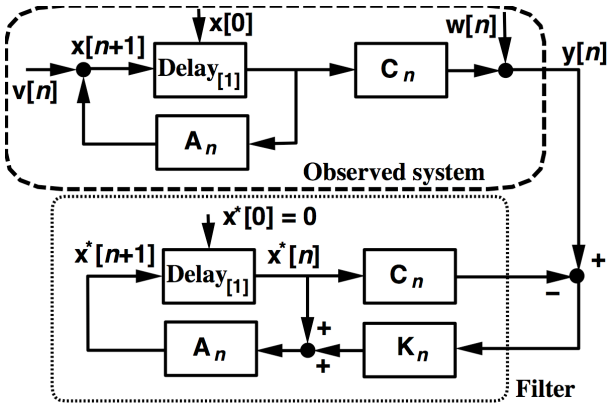


Figure 2.4: Kalman filter.

Then the linear unbiased least mean square estimate of the system is given by the following recurrent filtering algorithm:

$$\hat{\mathbf{x}}[n+1] = \mathbf{A}_n\left(\hat{\mathbf{x}}[n] + \mathbf{G}_n\left(\mathbf{y}[n] - \mathbf{C}_n\hat{\mathbf{x}}[n]\right)\right)$$

where the gain matrix $\mathbf{G}_n$ and the correlation matrices $\mathbf{P}_n$ and $\mathbf{P}_n^i$ of the errors are changing in time as follows:

$$\begin{array}{lcl} \mathbf{G}_n & = & \mathbf{P}_n^i \mathbf{C}_n^H \left(\mathbf{C}_n \mathbf{P}_n^i \mathbf{C}_n^H + \mathbf{V}_n \right)^{-1}; \\ \mathbf{P}_n & = & \mathbf{P}_n^i - \mathbf{G}_n \mathbf{C}_n \mathbf{P}_n^i; \\ \mathbf{P}_{n+1}^i & = & \mathbf{A}_n \mathbf{P}_n \mathbf{A}_n^H + \mathbf{U}_n. \end{array} \tag{2.27}$$

Just as Wiener filtering of stationary random processes, Kalman filtering minimizes the total square error, but in contrast to the former it need not simultaneously process all the input and output data samples but instead it permits to sequentially process the data samples. Kalman filtering can also be used for system identification, for example, to estimate the unknown matrices $\mathbf{B}_n$ in the more complex linear system model

$$\mathbf{x}[n+1] = \mathbf{A}_n \mathbf{x}[n] + \mathbf{B}_n \mathbf{u}[n]; \quad \mathbf{y}[n] = \mathbf{C}_n \mathbf{x}[n] + \mathbf{v}[n]$$

where $\mathbf{u}[n]$ is the known input sequence of finite energy.

Bibliography

[1] M. G. Bellanger, *Adaptive Digital Filters and Signal Analysis*. New York, Marcel Dekker, 1988.

[2] R. G. Brown and P. Y. C. Hwang: *Introduction to Random Signals and Applied Kalman Filtering*. 2nd Ed. New York, John Wiley & Sons, 1992.

[3] M. S. Grewal and A. P. Andrews, *Kalman Filtering: Theory and Practice*. Englewood Cliffs: Prentice-Hall, 1993.

[4] E. C. Ifeachor and B. W. Jervis, *Digital Signal Processing: A Practical Approach*. Workingham, England: Addison-Wesley, 1993.

[5] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems", *Trans. of the ASME – J. of Basic Engineering*, Series 82D, pp. 16–26, 1960.

[6] R. E. Kalman and R. Bucy: New Results in Linear Filtering and Prediction, *Trans. of the ASME – J. of Basic Engineering*, Series 83D, pp. 95–108, 1961.

[7] J. G. Proakis, and D. Manolakis, *Digital Signal Processing: Principles, Algorithms and Applications*. 2nd Ed. Sydney: Maxwell McMillan, 1992.

[8] H. W. Sorenson (Ed.) *Kalman Filtering: Theory and Application*. New York: IEEE Press, 1985.

[9] B. Widrow, A. Stearns, *Adaptive signal processing*. Englewood Cliffs: Prentice-Hall, 1985.

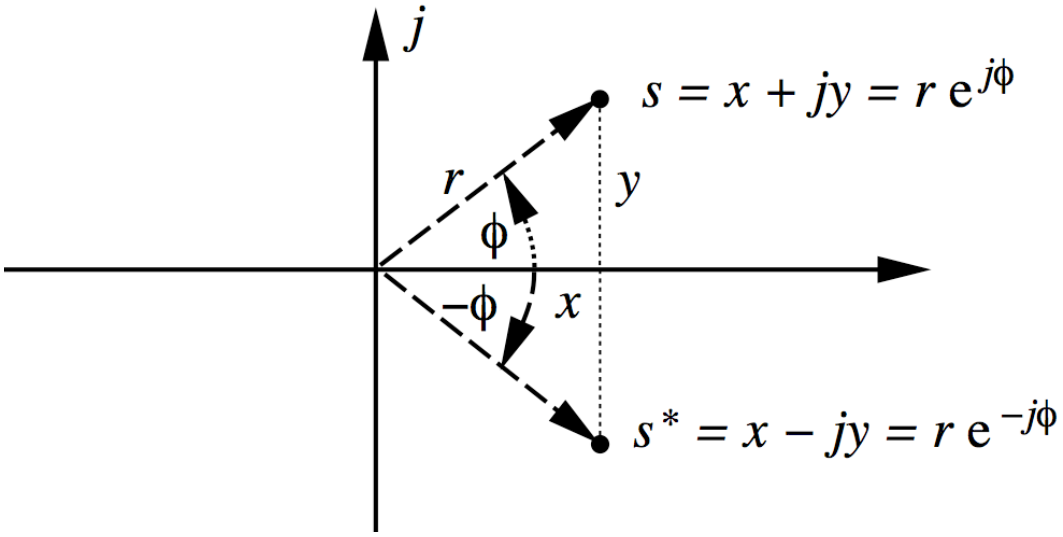
Appendix A

Complex variables and functions

A complex number has a real part and an imaginary part, both of which are constants. If the real part and/or imaginary part are variables, a complex number is called a *complex variable*. In the Laplace transform a complex variable is denoted $s = \sigma + j\omega$ where σ and ω are the real part and the imaginary part, respectively.

Basic definitions:

- $j = \sqrt{-1}$ is the imaginary unit ($j^2 = -1$);
- if $s = x + jy$ is the complex number or variable then $x = \Re(s)$ and $y = \Im(s)$ are the real and imaginary parts, respectively:



- $s^* = x - jy$ is the complex conjugate of s;
- $r \equiv |s| = \sqrt{x^2 + y^2}$ denotes the magnitude of s;

- $\phi \equiv \text{arc } s = \tan^{-1}(y/x)$ denotes the phase of s.

A complex function $F(s) = F_x + jF_y$ has a real part F_x and an imaginary part F_y where F_x and F_y are real quantities. The magnitude of $F(s)$ is $\sqrt{F_x^2 + F_y^2}$, and the angle, or phase θ of $F(s)$ is $\tan^{-1}(F_y/F_x)$. The angle is measured counterclockwise from the positive real axis. The complex conjugate of $F(s)$ is $F^*(s) = F_x - jF_y$.

Complex functions in linear control systems analysis are commonly single-valued functions of s and are uniquely determined for a given value of s.

A complex function $F(s)$ is *analytic* in a region if $F(s)$ and all its derivatives exist in that region. The derivative of an analytic function $F(s)$ is given by

$$\frac{d}{ds}F(s) = \lim_{\Delta s \to 0} \frac{F(s+\Delta s) - F(s)}{\Delta s} = \lim_{\Delta s \to 0} \frac{\Delta F}{\Delta s}$$

Since $\Delta s = \Delta\sigma + j\Delta\omega$, Δs can approach zero along an infinite number of different paths. It can be shown that if the following two Cauchy–Riemann conditions

$$\frac{\partial F_x}{\partial \sigma} = \frac{\partial F_y}{\partial \omega} \quad \text{and} \quad \frac{\partial F_y}{\partial \sigma} = -\frac{\partial F_x}{\partial \omega}$$

are satisfied, then the derivative $\frac{dF(s)}{ds}$ is uniquely determined, and the function $F(s)$ is analytic.

Example A.1 Let $F(s) = \frac{1}{s+1}$. Then

$$F(\sigma + j\omega) = \frac{1}{\sigma + 1 + j\omega} = F_x + jF_y$$

where

$$F_x = \frac{\sigma + 1}{(\sigma+1)^2 + \omega^2} \quad \text{and} \quad F_y = \frac{-\omega}{(\sigma+1)^2 + \omega^2}$$

It can be seen that, except at $s = -1$ (that is, $\sigma = -1$ and $\omega = 0$, the function $F(s)$ satisfies the Cauchy–Riemann conditions:

$$\begin{aligned} \frac{\partial F_x}{\partial \sigma} &= \frac{\partial F_y}{\partial \omega} = \frac{\omega^2 - (\sigma+1)^2}{((\sigma+1)^2 + \omega^2)^2} \\ \frac{\partial F_y}{\partial \sigma} &= -\frac{\partial F_x}{\partial \omega} = \frac{2\omega(\sigma+1)}{((\sigma+1)^2 + \omega^2)^2} \end{aligned}$$

Hence $F(s) = \frac{1}{s+1}$ is analytic in the entire s plane except at $s = -1$. The derivative $\frac{dF(s)}{ds}$, except at $s = -1$, is found to be

$$\begin{array}{rcl} \frac{d}{ds}F(s) & = & \frac{\partial F_x}{\partial \sigma} + j\frac{\partial F_y}{\partial \sigma} = \frac{\partial F_y}{\partial \omega} - j\frac{\partial F_x}{\partial \omega} \\ & = & -\frac{1}{(\sigma+1+j\omega)^2} = -\frac{1}{(s+1)^2} \end{array}$$

Note that the derivative of an analytic function can be obtained simply by differenciating $F(s)$ with respect to s. In this example,

$$\frac{d}{ds}\left(\frac{1}{s+1}\right) = -\frac{1}{(s+1)^2}$$

Poles and zeros. Points in the s plane at which the function $F(s)$ is analytic are called *ordinary* points, while points in the s plane at which the function $F(s)$ is not analytic are called *singular* points. Singular points at which the function $F(s)$ or its derivatives approach infinity are called *poles*. In the previous example, $s = -1$ is a singular point and is a pole of the function $F(s)$.

If $F(s)$ approaches infinity as s approaches $-p$ and if the function $F(s)(s+p)^n$ has a finite, nonzero value at $s = -p$ for $n = 1, 2, 3, \ldots$, then $s = -p$ is called a pole of order n. If $n = 1$, the pole is called a simple pole. If $n = 2, 3, \ldots$, the pole is called a second-order pole, a third-order pole, and so on. Points at which the function $F(s)$ equals zero are called *zeros*.

Example A.2 Let us consider the complex function

$$F(s) = \frac{K(s+2)(s+10)}{s(s+1)(s+5)(s+15)^2}$$

This function has zeros at $s = -2$, $s = -10$, simple poles at $s = 0$, $s = -1$, $s = -5$, and a second-order pole (multiple pole of order 2) at $s = -15$. Note that $F(s)$ becomes zero at $s = \infty$. Since for large values of s this function is approximately equal to

$$F(s) \approx \frac{K}{s^3}$$

$F(s)$ possesses a triple zero (multiple zero of order 3) at $s = \infty$. If points at infinity are included, this function has the same number of poles as zeros, namely, five zeros $(s = -2, -10, \infty, \infty, \infty)$ and five poles $(s = 0, -1, -5, -15, -15)$.

Appendix B

Complex-valued vectors and matrices

A *vector*-column is the $(m \times 1)$-matrix:

$$\mathbf{c} = \begin{pmatrix} c_1 \\ . \\ . \\ . \\ c_m \end{pmatrix}.$$

The vector-row is obtained by simply transposing the vector-column:

$$(c_1 \dots c_m) = \mathbf{c}^T.$$

Let the superscript $*$ denote the complex conjugate:

$$\text{if } z = x_{\mathrm{r}} + jx_{\mathrm{i}} \text{ then } z^* = x_{\mathrm{r}} - jx_{\mathrm{i}}.$$

The inner product of two complex-valued $(m \times 1)$-vectors is defined as the dot product:

$$< \mathbf{w}, \mathbf{v} >= \mathbf{w}^H \mathbf{v} = \sum_{i=1}^{m} w_i^* v_i \,.$$

The outer product $\mathbf{w}\mathbf{v}^H$ of the two $(m \times 1)$-vectors is the square $(m \times m)$-matrix:

$$\mathbf{w}\mathbf{v}^H = \begin{pmatrix} w_1 v_1^* & w_1 v_2^* & \dots & w_1 v_m^* \\ w_2 v_1^* & w_2 v_2^* & \dots & w_2 v_m^* \\ . & . & & . \\ . & . & & . \\ . & . & & . \\ w_m v_1^* & w_m v_2^* & \dots & w_m v_m^* \end{pmatrix}.$$

The scalar non-negative vector length, or norm, can be defined as $\| \mathbf{v} \| = < \mathbf{v}, \mathbf{w} >^{\frac{1}{2}}$, that is,

$$\|\mathbf{v}\| = \left(\sum_{i=1}^{m} |v_i|^2 \right)^{\frac{1}{2}}.$$

The Hermitian transposed, or adjoint, matrix that corresponds to a complex-valued $(m \times n)$-matrix

$$\mathbf{A} = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,m} \\ a_{2,1} & a_{2,2} & \dots & a_{2,m} \\ . & . & & . \\ . & . & & . \\ . & . & & . \\ a_{n,1} & a_{n,2} & \dots & a_{n,m} \end{pmatrix}$$

is the $(n \times m)$-matrix $\mathbf{A}^H$ obtained by replacing all the elements of $\mathbf{A}$ with their complex conjugate values and transposing the obtained matrix (that is, $\mathbf{A}^H = (\mathbf{A}^T)^* = (\mathbf{A}^*)^T$):

$$\mathbf{A}^H = \begin{pmatrix} a^*_{1,1} & a^*_{2,1} & \dots & a^*_{m,1} \\ a^*_{1,2} & a^*_{2,2} & \dots & a^*_{m,2} \\ . & . & & . \\ . & . & & . \\ . & . & & . \\ a^*_{1,n} & a^*_{2,n} & \dots & a^*_{m,n} \end{pmatrix}$$

Here, the Hermitian conjugation is denoted by the superscript H. In the similar way to transposition, it is easily shown that

$$\begin{aligned} (\mathbf{A} + \mathbf{B})^H &= \mathbf{A}^H + \mathbf{B}^H \\ (\mathbf{AB})^H &= \mathbf{B}^H \mathbf{A}^H \end{aligned}$$

A square $(m \times m)$-matrix $\mathbf{A}$ is *symmetric* if $\mathbf{A} = \mathbf{A}^T$ so that the elements $a_{k,l} = a_{l,k}$ are symmetric with respect to the main diagonal containing the elements $a_{k,k}$; $k = 1, \dots, m$. A simple example of the symmetric (4×4)-matrix is as follows:

$$\mathbf{S} = \begin{pmatrix} s_{1,1} & s_{1,2} & s_{1,3} & s_{1,4} \\ s_{1,2} & s_{2,2} & s_{2,3} & s_{2,4} \\ s_{1,3} & s_{2,3} & s_{3,3} & s_{3,4} \\ s_{1,4} & s_{2,4} & s_{3,4} & s_{4,4} \end{pmatrix}$$

A square complex-valued $(m \times m)$-matrix $\mathbf{A}$ is *Hermitian* if $\mathbf{A} = \mathbf{A}^H$.

Appendix C

Fourier and Laplace transforms

C.1 Fourier spectrum

Each real- or complex-valued function $f(t)$ can be expanded into the following complex-valued Fourier series:

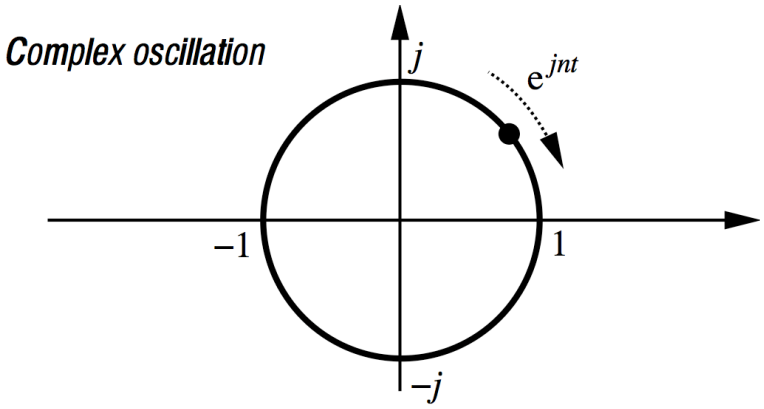
$$\begin{aligned} f(t) &= \sum_{n=-\infty}^{\infty} c_n \mathrm{e}^{jnt} \\ c_n &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) \mathrm{e}^{-jnt} dt; \quad (n = 0; \pm 1, \pm 2, \ldots) \end{aligned}$$

For the real-valued function $f(t)$, the coefficient c_0 is also real-valued, and the coefficient $c_{-n} = c_n^*$ so that

$$\begin{aligned} f(t) &= c_0 + \sum_{n=1}^{\infty} c_n \mathrm{e}^{jnt} + c_n^* \mathrm{e}^{-jnt} = c_0 + 2 \sum_{n=1}^{\infty} \Re(c_n \mathrm{e}^{jnt}) \\ &= r_0 + 2 \sum_{n=1}^{\infty} \Re(r_n \mathrm{e}^{j(nt-\phi_n)}) \\ &= r_0 + 2 \sum_{n=1}^{\infty} r_n \cos(nt - \phi_n) \end{aligned}$$

The term $c_n \mathrm{e}^{jnt}$ where $c_n = r_n \mathrm{e}^{-j\phi_n}$ means the complex oscillation with the cyclic frequency n, magnitude $r_n = |c_n|$, and initial phase $-\phi_n = \text{arc } c_n$.

A set of the successive values c_n is called the Fourier spectrum of the function $f(t)$.



Each exponential function e^{jnt} represents in the complex plane the movement of the complex point e^{jnt} along the unit circle with the cyclic frequency $n = 0, \pm 1, \pm 2, \ldots$. The point e^{jnt} runs around the unit circle n times for $t = 2\pi$ time units, the movement being clockwise if the frequency is positive and counterclockwise if it is negative.

Such a movement around the origin in the complex plane can be considered as a **complex oscillation** with the positive or negative cyclic frequency, depending on the direction of movement.

Let a function $f(t)$ be defined over the interval $(-\pi, \pi)$ and let it be extended periodically outside this range. Then its spectrum $(c_n : \ n = 0, \pm 1, \pm 2, \ldots)$ with respect to the set of oscillations e^{jnt} completely specifies the function. Most of physical phenomena described by the function $f(t)$ are much more obvious in terms of these oscillations.

C.2 Fourier integral

Let the physical phenomenon under consideration be non-periodical, that is, let the function $f(t)$ be defined for all t such that $-\infty < t < \infty$. Then this function can be represented by the following Fourier integral (under certain conditions yielding that this integral exists):

$$f(t) = \int_{-\infty}^{\infty} F(y) e^{jyt} dy \quad \text{where} \quad F(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t) e^{-jyt} dt$$

This integral is called a **spectral density** of the function $f(t)$.

The function $F(y)$ relates to the above Fourier coefficients c_n where the integer variable n is replaced by the continuous variable y. The value $r(y) = |F(y)|$ is called the **magnitude density**, and the initial phase is arc $F(y)$.

Example 1: the spectral density of the rectangular pulse

$$\begin{aligned} f(t) &= \begin{cases} A & \text{if} \quad |t| \le \tau \\ 0 & \text{if} \quad |t| > \tau \end{cases} \\ F(y) &= \frac{1}{2\pi} \int_{-\tau}^{\tau} A \mathrm{e}^{jyt} dt = \frac{A}{2\pi} \frac{\mathrm{e}^{jy\tau} - \mathrm{e}^{-jy\tau}}{jy} = \frac{A}{\pi} \frac{\sin \tau y}{y} \end{aligned}$$

Example 2: the spectral density of the decreasing exponential function:

$$\begin{aligned} f(t) &= \mathrm{e}^{-|t|} \\ F(y) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{e}^{-jyt} \mathrm{e}^{-|t|} dt = \frac{1}{2\pi} \int_{0}^{\infty} \left(\mathrm{e}^{jyt} + \mathrm{e}^{-jyt} \right) \mathrm{e}^{-t} dt \\ &= \frac{1}{2\pi} \int_{0}^{\infty} \mathrm{e}^{-t} \cos yt \; dt = \frac{1}{\pi(1+y^2)} \end{aligned}$$

Example 3. The spectral density of the oscillation with a cyclic frequency ω:

$$\begin{aligned} f(t) &= \mathrm{e}^{jwt} \\ F(y) &= \delta y - \omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{e}^{j(\omega - y)t} = \begin{cases} \infty & \text{if} \quad y = \omega \\ 0 & \text{if} \quad y \neq \omega \end{cases} \end{aligned}$$

where $\delta(\ldots)$ denotes the Dirac's delta function.

The Dirac's delta, or impulse function differs much from the conventional mathematical functions. For any continuous function $h(y)$ of y, it holds that

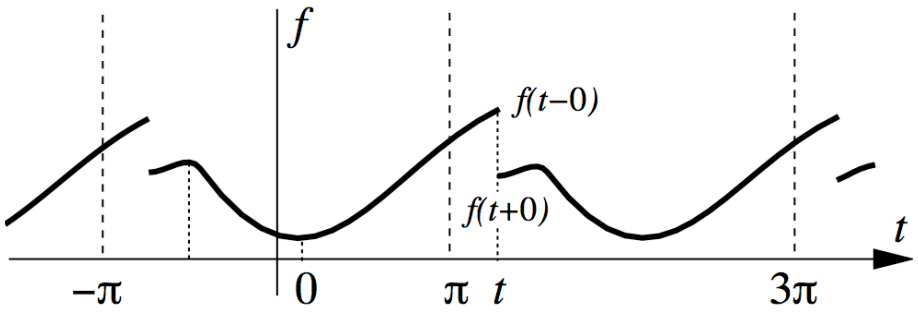
$$\int_{-\infty}^{\infty} h(y)\delta(y)dy = h(0).$$

C.3 From Fourier to Laplace integral

The spectral representation can only be obtained for the functions $f(t)$; $(-\infty < t < \infty)$ such that

$$\int_{-\infty}^{\infty} |f(t)|dt < \infty$$

and such that the Dirichlet's conditions hold, that is, the function has at each finite interval a finite number of extrema and a finite nimber of discontinuities of the first kind.



Periodic piecewise–continuous and piecewise–monotonic function

By involving the Dirac's delta-function, the spectral representation can be also obtained for the stationary oscillations $e^{j\omega t}$. But the increasing or decreasing oscillations $e^{(\alpha+j\omega)t}$ cannot be represented by the Fourier integral because the integral defining $F(y)$ diverges at $t = +\infty$ if $\alpha > 0$ or at $t = -\infty$ if $\alpha < 0$. The Laplace integral overcomes this difficulty.

In practice, no process can be observed over the interval $(-\infty, \infty)$. Usually, each process starts at the finite instant which can be considered as the origin ($t = 0$), and then is observed for a long time (theoretically, till $t = +\infty$). Assuming $f(t) = 0$ for $t < 0$, the spectral representation of $f(t)$ is:

$$\begin{aligned} F(y) &= \frac{1}{2\pi}\int_0^\infty e^{-jyt} f(t)dt \\ \int_{-\infty}^{\infty} e^{jyt} F(y)dy &= \begin{cases} f(t) & \text{if} \quad t \geq 0 \\ 0 & \text{if} \quad t < 0 \end{cases} \end{aligned}$$

To avoid the divergent Fourier integrals, $f(t)$ can be replaced by a decreasing function $e^{-xt}f(t)$ with a control parameter $x > 0$. The Fourier integral giving the spectral density of this function depends on x and converges for all bounded functions $f(t)$, even for the exponential ones such as $e^{-\alpha t}$, $\alpha > 0$ if $x > \alpha$. The spectral representation is now as follows:

$$F_x(y) = \frac{1}{2\pi}\int_0^\infty e^{-jyt}\left[e^{-xt}f(t)\right]dt = \frac{1}{2\pi}\int_0^\infty e^{-(x+jy)t}f(t)dt$$

$$\int_{-\infty}^{\infty} e^{(x+jy)t}F_x(y)dy = \begin{cases} f(t) & \text{if} \quad t \geq 0 \\ 0 & \text{if} \quad t < 0 \end{cases}$$

Because the spectral density $F_x(y)$ depends now on the complex variable $s = x + jy$, let us denote

$$F_x(y) = F(x + jy) = F(s)$$

For the integration, x is constant wheras y is changing from $-\infty$ to $+\infty$. Therefore, s is changing from $x - j\infty$ to $x + j\infty$ (in the complex plane, it is a movement along the vertical line with abscissa x).

By replacing $f(t)$ with $2\pi f(t)$ and using $ds = jdy$, we obtain the **Laplace integrals**:

$$\begin{aligned} \int_0^{\infty} \mathrm{e}^{-st} f(t)dt &= F(s) \\ \frac{1}{2\pi j} \int_{x-j\infty}^{x+j\infty} \mathrm{e}^{st} F(s)ds &= \begin{cases} f(t) & \text{if} \quad t \geq 0 \\ 0 & \text{if} \quad t < 0 \end{cases} \end{aligned}$$

The first integral is called the **Laplace integral**. and the second one is the **inverse Laplace integral**.

Physical meaning of the Laplace integral is that $F(x + jy)$ with the constant x is a spectral density of the decreasing function $\mathrm{e}^{-xt} f(t)$ which has the cyclic frequency y.

Examples of the Laplace integral:

Example 1. Unit step function

$$u(t) = \begin{cases} 1 & \text{if} \quad t > 0 \\ 0 & \text{if} \quad t < 0 \\ \varepsilon & \text{if} \quad t = 0; \quad (0 \leq \varepsilon \leq 1) \end{cases}$$

The spectral density (if $\mathrm{e}^{-st} \to 0$ when $t \to \infty$, that is, if $\Re(s) > 0$) is:

$$F(s) = \int_0^{\infty} \mathrm{e}^{-st} dt = \left. \frac{\mathrm{e}^{-st}}{-s} \right|_0^{\infty} = \frac{1}{s}$$

Example 2. Delayed unit step function

$$u_\tau(t) = \begin{cases} 1 & \text{if} \quad t > \tau > 0 \\ 0 & \text{if} \quad t < \tau \\ \varepsilon & \text{if} \quad t = \tau; \quad (0 \leq \varepsilon \leq 1) \end{cases}$$

The spectral density of $u_\tau(t) = u(t - \tau)$ is:

$$F(s) = \int_0^{\infty} \mathrm{e}^{-st} u(t - \tau)dt = \int_{\tau}^{\infty} \mathrm{e}^{-st} dt = \frac{\mathrm{e}^{-\tau s}}{s} \quad \text{if} \quad \Re(s) > 0$$

Thus, the delay τ of the function corresponds to multiplcation of the spectral density by the factor $\mathrm{e}^{-\tau s}$.

Example 3. Exponential function $f(t) = u(t)\mathrm{e}^{\alpha t}$ with $\alpha = \sigma + j\omega$:

σ	ω	Type of oscillation
0	$\neq 0$	Complex oscillation
< 0	$\neq 0$	Decreasing oscillation
> 0	$\neq 0$	Increasing oscillation
$\neq 0$	0	Aperiodic rise or decline

The spectral density is:

$$\int_0^\infty \mathrm{e}^{-st}\mathrm{e}^{\alpha t}dt = \int_0^\infty \mathrm{e}^{-(s-\alpha)t}dt = \frac{1}{s-\alpha} \quad \text{if} \quad \Re(s) > \Re(\alpha)$$

Example 4. Real-valued oscillation $u(t)\cos\omega t$ is the sum of two complex oscillations

$$\cos\ \omega t = \frac{1}{2}\left(\mathrm{e}^{j\omega t} + \mathrm{e}^{-j\omega t}\right)$$

The spectral density is

$$\int_0^\infty \mathrm{e}^{-st}\cos\ \omega t\ dt\frac{1}{2}\left(\frac{1}{s-j\omega} + \frac{1}{s+j\omega}\right) = \frac{s}{s^2+\omega^2}$$

Exercise: Find the spectral density of the real-valued oscillation $u(t)\sin\ \omega t$.

C.4 Laplace transform

The Laplace transform maps the space of initial functions to the space of transform functions:

$$F(s) = \int_0^\infty \mathrm{e}^{-st}f(t)dt \quad \leftrightarrow \quad F(s) = \mathcal{L}\{f(t)\}$$

The inverse Laplace transform inverses such a mapping:

$$f(t) = \frac{1}{2\pi j}\int_{x-j\infty}^{x+\infty} \mathrm{e}^{st}F(s)ds; \ \ t > 0 \quad \leftrightarrow \quad f(t) = \mathcal{L}^{-1}\{F(s)\}$$

Many operations with functions become very simple in the transform space:

Similarity	$f(at) \rightarrow \frac{1}{a}F\left(\frac{s}{a}\right)$ $F(as) \leftarrow \frac{1}{a}f\left(\frac{t}{a}\right)$	$a > 0$
Delay (shift)	$u(t-a)f(t-a) \rightarrow \mathrm{e}^{-as}F(s)$	$a > 0$
Decay	$\mathrm{e}^{-\alpha s}f(t) \rightarrow F(s+\alpha)$	$\alpha = \sigma + j\omega$
Differentiation	$f'(t) \rightarrow sF(s) - f(+0)$	
Integration	$\int_0^t f(\tau)d\tau \rightarrow \frac{1}{s}F(s)$ $\frac{1}{t}f(t) \leftarrow \int_s^\infty F(\sigma)d\sigma$	
Convolution	$f_1(t) * f_2(t) \rightarrow F_1(s) \cdot F_2(s)$	

C.5 Convolution

Convolution is an integral combination of functions which is frequently met in physics, signal processing, and control:

$$f_1(t) * f_2(t) = \int_0^t f_1(\tau)f_2(t-\tau)d\tau$$

The convolution is a commutative and associative operation:

$$f_1(t) * f_2(t) = f_2(t) * f_1(t);$$
$$f_1(t) * (f_2(t) * f_3(t)) = (f_1(t) * f_2(t)) * f_3(t)$$

The rules of differentiation

$$\begin{aligned}
f'(t) \equiv \frac{d}{dt}f(t) &\rightarrow sF(s) - f(+0) \\
f''(t) \equiv \frac{d^2}{dt^2}f(t) &\rightarrow s^2F(s) - f(+0)s - f'(+0) \\
&\cdots \quad \cdots \quad \cdots \\
f^{(n)}(t) \equiv \frac{d^n}{dt^n}f(t) &\rightarrow s^nF(s) - f(+0)s^{n-1} - f'(+0)s^{n-2} - \ldots \\
&\qquad\qquad -f^{(n-2)}(+0)s - f^{n-1}(+0)
\end{aligned}$$

and the rule of convolution are the most important in practice because they permit us to replace integral and differential equations by simpler algebraic ones.