2.122 Stochastic Systems

Themistoklis P. Sapsis

July 2, 2025

Contents

Introduction

Deterministic versus stochastic systems:

1. In both cases, we may have a well-defined set of equations.
2. In deterministic cases, we have repeatability.

Stochasticity in dynamical systems may be caused by

- **External factors**: these can include stochastic forcing, boundary conditions, and/or parameters. Typical systems in this class of problems include random linear vibrations, advection of tracers, porous media, and others.
- **Internal factors**: Intrinsic instabilities which leads to chaotic behavior, i.e. sensitive dependence to small perturbations. Although the system is deterministic, its chaotic behavior leads to the adoption of a stochastic perspective.

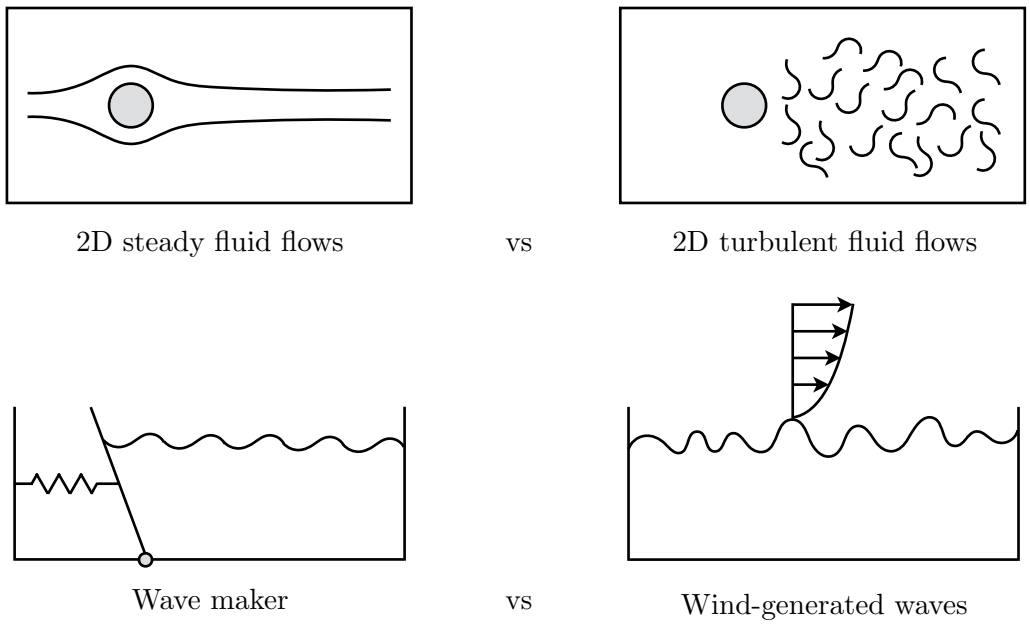


Figure 1 – Examples of deterministic versus stochastic systems.

Chapter 1

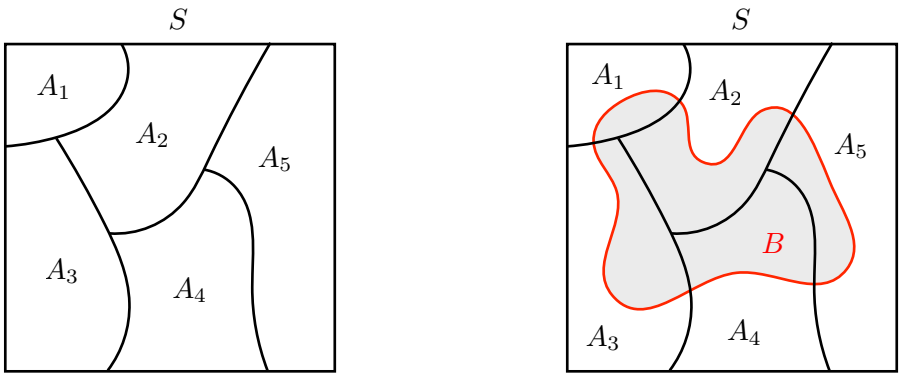
Probability Theory and Random Variables

1.1 Definition of probability

Let us begin with an empirical definition of probability,

$$P = \frac{\text{number of favorable outcomes}}{\text{total number of outcomes}}.$$

Formally, we define the event space S which contains all the possible outcomes A_i (events). Figure 1.1 shows an exemple of event space with simple and composite events. The event space includes the empty set $\emptyset$. Venn diagrams show all possible logical relations between a collection of different events and are illustrated in Figure 1.2.



$A_1, \ldots, A_5$ non-overlapping (simple) events

B overlapping (composite) event

Figure 1.1 – Event space.

Properties of logical operators:

1. Union and intersection operators, $\cup, \cap$, are commutative and associative.
2. Distribution property:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C),$$

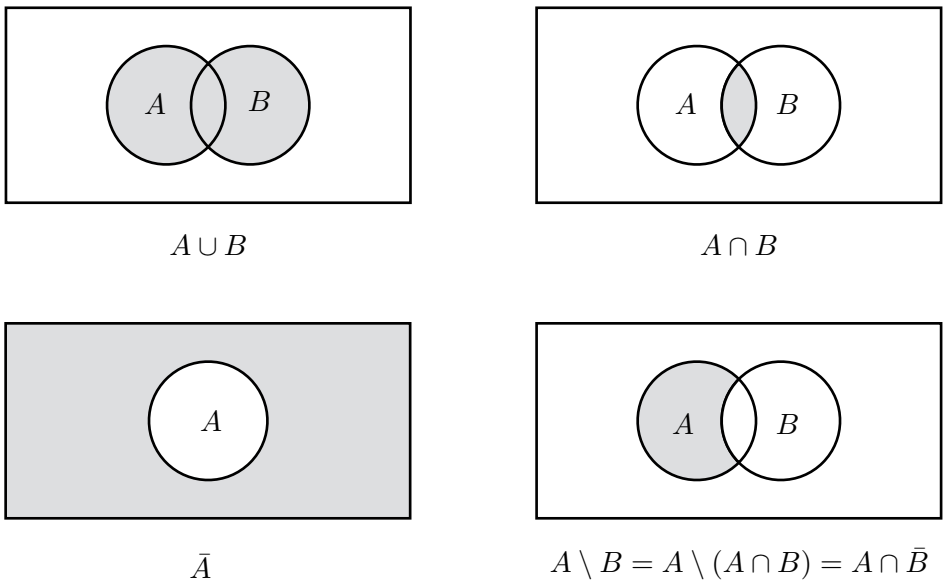


Figure 1.2 – Venn diagrams.

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

3. De Morgan's law's:

$$\overline{A \cup B} = \bar{A} \cap \bar{B},$$
$$\overline{A \cap B} = \bar{A} \cup \bar{B}.$$

A probability is then formally defined by numbers, chosen for each event, such that it satisfies the following axioms:

1. $P_i = P(A_i) \geq 0$ (Nonnegativity),
2. $P(S) = 1$ (Normalization),
3. If $A_i \cap A_j = \emptyset$, then $P(A_i \cup A_j) = P(A_i) + P(A_j)$ (Additivity).

Thus, the probability can be seen as the "area" of event A_i. The following properties can follow from the axioms of probability

1. $P(\emptyset) = 0$,
2. $P(A_i) = 1 - P(\bar{A}_i) \leq 1$,
3. If $A_i \cap A_j \neq \emptyset$, then $P(A_i \cup A_j) = P(A_i) + P(A_j) - P(A_i \cap A_j)$.

Example. *Toss a fair coin.*

The event space consists of events $A_1 = H$ (head) and $A_2 = T$ (tail), with respective probabilities $P(H) = 1/2$ and $P(T) = 1/2$. See Figure 1.3 for a graphical depiction of the corresponding event space.

Example. *Roll a dice.*

The event space consists of events $A_1, \ldots, A_6$ with probabilities $P(A_i) = 1/6$, $i = 1, \ldots, 6$. We can define the composite event $M = \{A_2, A_4, A_6\}$, representing the fact that the dice outcome is an even number.

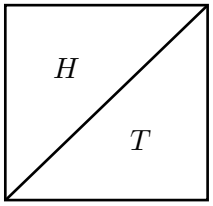


Figure 1.3 – Event space when tossing a fair coin.

1.2 Conditional probability

Given that a composite event M happened, what is the probability that A_j also happened? We denote this probability as $P(A_j|M)$.

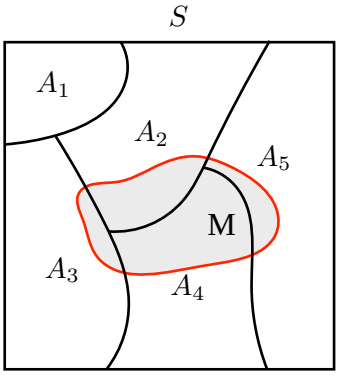


Figure 1.4 – Simple events conditioned on a composite event.

If M has occurred we must re-define probabilities because M is now the new event space:

$$P(M) = 1.$$

Based on this property we define the conditional probability of the event A_j conditioned on M as

$$P(A_j|M) = \frac{P(A_j \cap M)}{P(M)}.$$

Note that we have normalized with $P(M)$ to ensure that all axioms of probability are satisfied by the new probability $P(A_j|M)$. From this definition it immediately follows

$$P(A_j \cap M) = P(A_j|M)P(M) = P(M|A_j)P(A_j).$$

Example (previous example continued)**.** *What is the probability of A_2 given that M has happened?*

To answer this question, note that our new event space is now $M = \{A_2, A_4, A_6\}$. The conditional probability is therefore given by

$$P(A_2|M) = \frac{P(A_2 \cap M)}{P(M)} = \frac{P(A_2)}{P(A_2 \cup A_4 \cup A_6)} = \frac{1}{3}.$$

More generally, the conditional probability of A given that a composite event M happened is given by

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

Note that this implies

$$P(A \cap M) = P(A|M)P(M) = P(M|A)P(A).$$

Based on this property we can prove the *Total Probability Theorem*. If all A_j, $j = 1, \ldots, N$, are independent (simple) events and M is composite, then

$$P(M) = P(M|A_1)P(A_1) + P(M|A_2)P(A_2) + \cdots + P(M|A_N)P(A_N).$$

Proof:

$$\begin{aligned} S \cap M = M \quad &\Rightarrow \quad (A_1 \cup A_2 \cup \cdots \cup A_N) \cap M = M \\ &\Rightarrow \quad (A_1 \cap M) \cup (A_2 \cap M) \cup \cdots \cup (A_N \cap M) = M \\ &\Rightarrow \quad P(M) = P(A_1 \cap M) + P(A_2 \cap M) + \cdots + P(A_N \cap M) \\ &\Rightarrow \quad P(M) = P(M|A_1)P(A_1) + P(M|A_2)P(A_2) + \cdots + P(M|A_N)P(A_N). \end{aligned}$$

Example. *Box A contains 2000 components, of which 5% are defective. Box B contains 500 components, of which 40% are defective. Box C and D contain 1000 components each, with 10% defective. Pick one box at random and choose one component. What is the probability that it is defective?*

We define simple events A_i, $i = 1, \ldots, 4$ representing each box, and a composite event $M = \{\text{pick a defective component}\}$. We know $P(M|A_i)$ and $P(A_i)$, thus

$$\begin{aligned} P(M) &= P(M|A_1)P(A_1) + P(M|A_2)P(A_2) + P(M|A_3)P(A_3) + P(M|A_4)P(A_4) \\ &= 0.05 \cdot \frac{1}{4} + 0.40 \cdot \frac{1}{4} + 0.10 \cdot \frac{1}{4} + 0.10 \cdot \frac{1}{4} \\ &= 0.1625. \end{aligned}$$

The total probability theorem is used together with the **Bayes' Rule**, which allows to express probabilities of the form $P(A_i|M)$ in terms of $P(M|A_i)$:

$$P(A_i|M) = \frac{P(M|A_i)P(A_i)}{P(M|A_1)P(A_1) + \cdots + P(M|A_N)P(A_N)}.$$

Example. *Person X is searching for a treasure with equipment that finds the treasure with 70% success rate. Person Y has different equipment with 60% success rate. The treasure is in area I with probability 80%, or in area II with probability 20%. Person X learns that Y searched area I and did not succeed. Which area should he search first?*

Let us define the events

$$\begin{aligned} A_I &= \{\text{treasure is in I}\}, \\ A_{II} &= \{\text{treasure is in II}\}, \\ B &= \{\text{person } Y \text{ did not find it in I}\}. \end{aligned}$$

Then,

$$P(A_1|B) = \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2)} = \frac{0.4 \cdot 0.8}{0.4 \cdot 0.8 + 1 \cdot 0.2} = 0.62.$$

Person Y now finds out that X searched area I (after he did) and did not succeed. Where is the treasure most likely located? We define

$$C = \{X \text{ and } Y \text{ did not find it in I}\},$$

and, assuming independence (see below) of the failure rate of X and Y, we have

$$P(A_1|C) = \frac{P(C|A_1)P(A_1)}{P(C|A_1)P(A_1) + P(C|A_2)P(A_2)} = \frac{0.3 \cdot 0.4 \cdot 0.8}{0.3 \cdot 0.4 \cdot 0.8 + 1 \cdot 0.2} = 0.324.$$

1.3 Independence

Building on the notion of conditional probability we can now define the notion of *independence* between events. Two events A and B are called independent if and only if $P(A|B) = P(A)$ or $P(B|A) = P(B)$. This is equivalent with $P(A \cap B) = P(A)P(B)$, since

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

1.4 Random variables

A random variable X assigns a value, X_i, to each event A_i; see Figure 1.5. The input is an event, the output is a number, vector, etc.

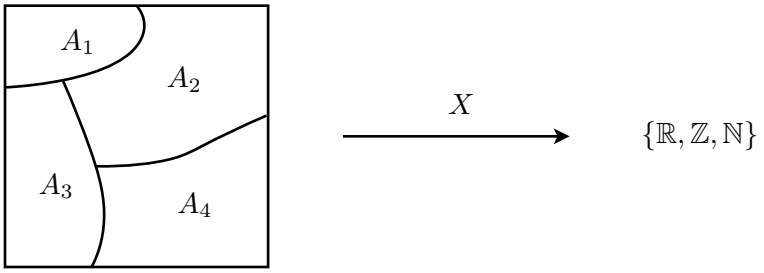


Figure 1.5 – A random variable maps the event space into a (discrete or continuous) set of values.

Example. Toss a coin:

$$\begin{aligned} \text{H} &\rightarrow +\$1, \\ \text{T} &\rightarrow -\$1. \end{aligned}$$

Example. Throw a dice:

$$\begin{aligned} 1 &\rightarrow +\$2, \\ 5 &\rightarrow -\$2, \\ 6 &\rightarrow -\$100, \\ 2,3,4 &\rightarrow \$0. \end{aligned}$$

Symbolically, $X(\zeta)$ is a variable that depends on the random outcome ζ:

$$X(\zeta) \quad \longrightarrow \quad \begin{cases} X_1 & \leftrightarrow & A_1 & \leftrightarrow & P_1 \\ X_2 & \leftrightarrow & A_2 & \leftrightarrow & P_2 \\ \vdots & & & & \\ X_n & \leftrightarrow & A_n & \leftrightarrow & P_n \end{cases}$$

We can now fully describe the random variable in terms of a table $\{X_i,\}$ of values and associated probabilities (the event no longer needs to be explicitly mentioned).

We define the *expected value or mean operator*:

$$\bar{X} = E[X(\zeta)] = \sum_{i=1}^{n} P_i X_i.$$

Statistical interpretation: After many experiments (m), we count $i_1, X_1; i_2, X_2; \ldots; i_n, X_n$. The total number of experiments can be expressed as

$$m = i_1 + i_2 + \cdots + i_n,$$

the sum of values of X over all experiments (total income) is

$$X_T = i_1 X_1 + i_2 X_2 + \cdots + i_n X_n.$$

Then, the expected value of X (average income per play) follows as

$$\bar{X} = \frac{X_T}{m} = \sum_{j=1}^{n} \frac{i_j}{m} X_j = \sum_{j=1}^{n} P_j X_j.$$

Properties of the expectation:

1. $E[X(\zeta) + Y(\zeta)] = E[X(\zeta)] + E[Y(\zeta)]$,
2. $E[\alpha X(\zeta)] = \alpha E[X(\zeta)], \quad \alpha \in \mathbb{R}$,
3. Consider a function of $X(\zeta)$, $f(X(\zeta))$. Then,

$$F_T = i_1 f(X_1) + i_2 f(X_2) + \cdots + i_n f(X_n),$$

hence

$$\overline{f(X)} = E[f(X(\zeta))] = \frac{F_T}{m} = \sum_{i=1}^{n} P_i f(X_i).$$

Next, we define the *variance* of a random variable:

$$\begin{aligned} V &= E[(X(\zeta) - \bar{X})^2] \\ &= E[X(\zeta)^2 - 2\bar{X}X + \bar{X}^2] \\ &= E[X^2] - 2\bar{X}E[X] + E[\bar{X}^2] \\ &= E[X^2] - \bar{X}^2 \geq 0, \end{aligned}$$

from which follows the definition of the *standard deviation*

$$\sigma = \sqrt{V},$$

which possesses the same units as X.

1.5 Continuous random variables

1.5.1 Probability density function

A continuous random variable is a random variable with an infinite number of outcomes. Take, for example, a random number in the continuous interval $[0, 1]$. The probability of getting any specific number is zero (improbable, but not impossible). On the other hand, the probability of getting a number contained in an infinitesimal interval of length dx is nonzero, and described by the probability density function (pdf):

$$P(X \in [x_0, x_0 + dx]) = f_X(x_0)dx,$$

where $P(\cdot)$ describes the actual probability, and $f_X(\cdot)$ is the pdf. Therefore, the pdf $f_X(x)$ is such that its area between any two points x_a, x_b gives the probability that $X \in [x_a, x_b]$; see Figure 1.6. As a result, the pdf has the following properties

1. $f_X(x) \geq 0$ for all x, (nonnegativity)
2. $\int_{-\infty}^{\infty} f_X(x)dx = 1$, (normalization)

Note that the pdf itself is *not* a probability. It is probability per unit length.

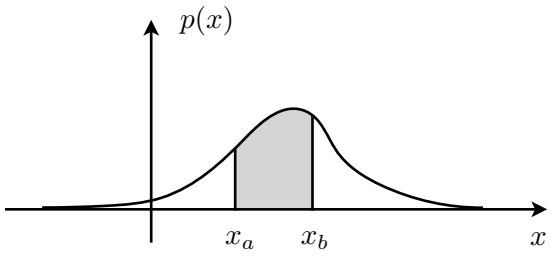


Figure 1.6 – Probability density function.

The expected value of a continuous random variable is

$$\bar{X} = E[X] = \int_{-\infty}^{\infty} x\, f_X(x)dx,$$

and that of a function of a continuous random variable is

$$\overline{G(X)} = E[G(X)] = \int_{-\infty}^{\infty} G(x)\, f_X(x)dx.$$

Example. *Consider a random variable X_u uniformly distributed in the interval $[0, 2\pi]$ and described by the pdf $\mathcal{U}(0, 2\pi)$; see Figure 1.7.*

We have

$$\bar{X} = E[X_u] = \int_{-\infty}^{\infty} x p_u(x)dx = \int_0^{2\pi} x\frac{1}{2\pi}dx = \pi,$$
$$V = E[(X - \bar{X})^2] = \int_{-\infty}^{\infty} (x - \bar{X})^2 p_u(x)dx = \int_0^{2\pi} (x - \pi)^2\frac{1}{2\pi}dx = \frac{\pi^2}{3},$$
$$\sigma = \sqrt{V} = \frac{\pi}{\sqrt{3}},$$

$$E[\cos X] = \int_{-\infty}^{\infty} \cos x \, p_u(x) dx = \int_0^{2\pi} \cos x \frac{1}{2\pi} dx = 0,$$
$$E[\cos^2 X] = \int_{-\infty}^{\infty} \cos^2 x \, p_u(x) dx = \int_0^{2\pi} \cos^2 x \frac{1}{2\pi} dx = \frac{1}{2}.$$

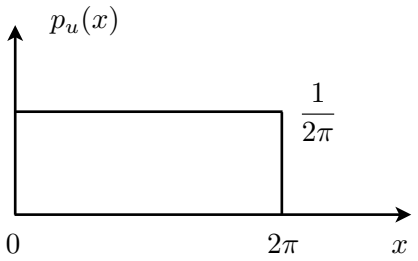


Figure 1.7 – Uniform pdf $\mathcal{U}(0, 2\pi)$.

1.5.2 Cumulative distribution function

We define the cumulative probability (or distribution) function (cdf) as

$$F_X(x_0) = P(X \leq x_0) = \int_{-\infty}^{x_0} f_X(x) dx, \quad \text{for all } x_0,$$

see Figure 1.8(a). As a result, the pdf can be derived from the cdf as

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

Moreover, the probability of any finite interval (Figure 1.8(b)) is given by

$$P(x_a \leq X \leq x_b) = \int_{x_a}^{x_b} f_X(x) dx = F_X(x_b) - F_X(x_a).$$

We can prove the following properties for the cdf

1. $\lim_{x \to -\infty} F_X(x) = 0$ (probability of the empty set),
2. $\lim_{x \to \infty} F_X(x) = 1$ (probability of the full event space),
3. $F_X(x)$ is a non-decreasing function (probability of a subset).

1.5.3 Conditional expectation

Using conditional probabilities we can define the corresponding notion for conditional probability density functions. Let a random variable $X \sim \mathcal{U}(0, 2\pi)$. Then, the conditional pdf of $\mathcal{U}$ given that $x > \pi$ is given by (see Figure 1.9 for a graphical illustration)

$$f_U(x|x > \pi) = \frac{f_U(x \cap x > \pi)}{P(x > \pi)}$$

Using this pdf we can compute the conditional expectation of X given that $X > \pi$ is given by (see Figure 1.9)

$$E[X|X > \pi] = \int_0^{2\pi} x f_U(x|x > \pi) dx = \int_0^{2\pi} x \frac{f_U(x \cap x > \pi)}{P(x > \pi)} dx = \int_\pi^{2\pi} x \frac{1/(2\pi)}{1/2} dx = \frac{3\pi}{2}.$$

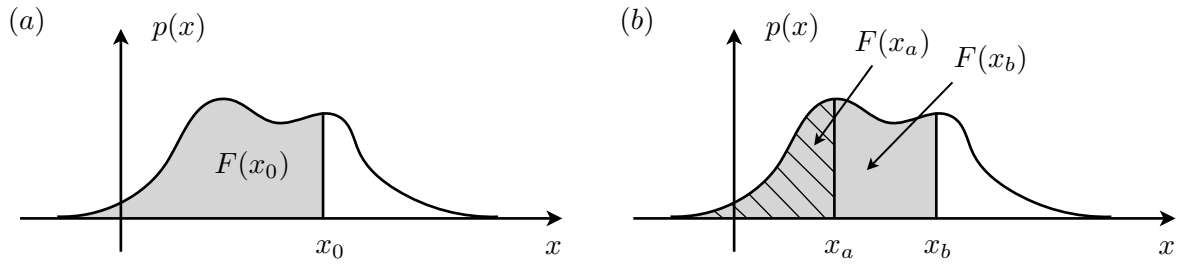


Figure 1.8 – Cumulative distribution function.

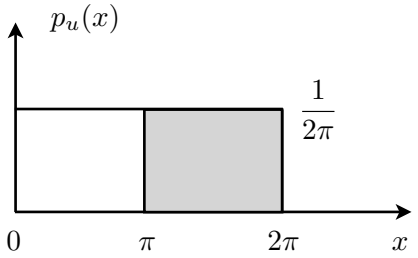


Figure 1.9 – Uniform pdf $\mathcal{U}(0, 2\pi)$; gray area represents $p_u(x \cap x > \pi)$.

1.6 Chebyshev's inequality

Chebyschev's inequality is particularly important to bound the probability for large deviations around the mean of a random variable with arbitrary probability distribution. For any random variable with finite mean μ and nonzero, finite, standard deviation σ we have:

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}, \quad k > 1.$$

1.7 Two dimensional random variables

In analogy with the scalar case we define the joint cumulative distribution function and probability density function for a two dimensional random vector (X, Y) (the extension can be applied for higher dimensional vectors). Specifically, using the joint probability for the variables X, Y we define the *joint cumulative distribution function*

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

The cumulative distribution function for the random vector (X, Y) has the same properties with the scalar case:

1. $\lim_{x\to-\infty} F_{XY}(x, y) = 0$ and $\lim_{y\to-\infty} F_{XY}(x, y) = 0$ (probability of the empty set),
2. $\lim_{x,y\to\infty} F_{XY}(x, y) = 1$ (probability of the full event space),
3. $F_{XY}(x, y)$ is a non-decreasing function with respect to x and with respect to y (probability of a subset).

Note that the probabilistic information for each scalar random variable X or Y is also contained in the joint cumulative distribution function. In particular, we have

$$\lim_{y\to\infty} F_{XY}(x, y) = \lim_{y\to\infty} P(X \leq x, Y \leq y) = P(X \leq x) = F_X(x),$$

$$\lim_{x\to\infty} F_{XY}(x,y) = \lim_{x\to\infty} P(X \le x, Y \le y) = P(X \le y) = F_Y(y).$$

In analogy with the scalar case we define the *joint probability density function* for the random vector (X, Y)

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

The joint probability density function is quantifying the probability that the random vector belongs in an infinitesimally small element $dxdy$:

$$P((X,Y) \in [x_0, x_0 + dx] \times [y_0, y_0 + dy]) = f_{XY}(x_0, y_0) dxdy.$$

We integrate over any subset $\mathcal{D} \subseteq \mathbb{R}^2$ to obtain its probability

$$P((X,Y) \in \mathcal{D}) = \iint_{\mathcal{D}} f_{XY}(x,y) dxdy.$$

Knowing the joint probability density function, we directly obtain, through integration, the probability density functions for each of the vector components X or Y, also known as *marginals*:

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y) dy \quad \text{and} \quad f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y) dx.$$

1.8 Correlation and covariance of two random variables

For applications it is often hard to quantify the dependence of two random variables through the conditional pdf of one on the other. For this reason, we use simpler measures such as second-order statistics.

Correlation

Let X be a random variable with mean $\bar{X}$ and variance $V_X = \text{var}(X) = E[(X - \bar{X})^2]$. Let Y be a random variable with mean $\bar{Y}$ and variance $V_Y = \text{var}(Y) = E[(Y - \bar{Y})^2]$. The correlation of X and Y is defined as

$$R_{XY} = E[XY] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{XY}(x,y) dxdy,$$

The correlation has the following important properties:

1. $R_{XY} = R_{YX}$ (symmetry),
2. for any vector of random variables $[X_1, X_2, ..., X_N]$ the matrix with elements $R_{X_i X_j}$ is positive semi-definite, i.e.

$$\sum_{i=1}^{N} \sum_{j=1}^{N} R_{X_i X_j} a_i a_j \ge 0, \text{ for all } a_i, a_j.$$

A direct consequence of these properties is that the matrix $R_{X_i X_j}$ is symmetric, and it has real, non-negative eigenvalues.

Covariance

Covariance is a similar measure (based on second-order statistics) that focuses on measuring how the fluctuations of two random variables around their means relate. The covariance of X and Y is defined as

$$C_{XY} = \text{cov}(X, Y) = E[(X - \bar{X})(Y - \bar{Y})] = E[XY] - \bar{X}\bar{Y} = R_{XY} - \bar{X}\bar{Y}.$$

Similarly with the correlation, covariance is both symmetric and positive semi-definite.

To characterize just the mutual dependence between the two random variables, without taking into account their magnitude, we define the correlation coefficient, which is a normalized version of the covariance. It is defined as

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)}\sqrt{\text{var}(Y)}}.$$

If $\rho(X, Y) = 0$, then X and Y are uncorrelated. Note that if X and Y are independent, then

$$\begin{aligned}
\text{cov}(X, Y) &= E[XY] - \bar{X}\bar{Y} \\
&= \sum_i \sum_j X_i Y_j P(X = X_i, Y = Y_j) - \bar{X}\bar{Y} \\
&= \sum_i X_i P(X = X_i) \sum_j Y_j P(Y = Y_j) - \bar{X}\bar{Y} \\
&= \bar{X}\bar{Y} - \bar{X}\bar{Y} = 0,
\end{aligned}$$

that is, X and Y are uncorrelated. We have used independence to write $P(X = X_i, Y = Y_j) = P(X = X_i)P(Y = Y_j)$. On the other hand, X and Y being uncorrelated does not imply that they are independent. For instance, we could have (with $\bar{X} = \bar{Y} = 0$) $E[XY] = 0$ and $E[X^3Y] \neq 0$, showing that X and Y are not independent.

By the Cauchy-Schwartz inequality we have $|\rho(X, Y)| \leq 1$. The equality $|\rho(X, Y)| = 1$ is achieved if and only if $Y = aX + b$ with a, b constants.

1.9 Gaussian random variable

A continuous random variable X is said to be normal or Gaussian if it has a pdf of the form

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}.$$

See Figure 1.10 for a graphical representation of the pdf. The mean X is

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx = \mu,$$

while its variance is

$$\begin{aligned}
\text{var}(X) = E[(X - \mu)^2] &= \int_{-\infty}^{\infty} (x - \mu)^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} dx \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 e^{-y^2/2} dy
\end{aligned}$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \left(-ye^{-y^2/2}\right)\Big|_{-\infty}^{\infty} + \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy$$
$$= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy$$
$$= \sigma^2.$$

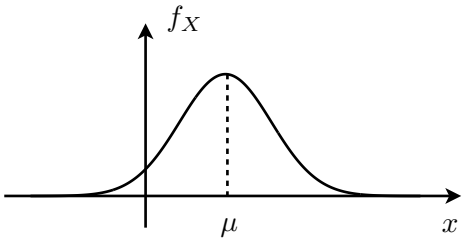


Figure 1.10 – Gaussian pdf.

Normality is preserved by linear transformations. If X is a Gaussian random variable, $Y = aX + b$, a, $b \in \mathbb{R}$ is also a Gaussian random variable with $E[Y] = aE[X] + b$ and $\text{var}(Y) = a^2\text{var}(X)$. Therefore, suppose you have a Gaussian random variable with mean μ and standard deviation σ. Then,

$$P(X \leq x) = P\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) = P\left(Y \leq \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right),$$

where Y is a standard Gaussian random variable, that is, it has zero mean and unit variance, and its associated cdf is

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

In this way, we can compute probabilities for any Gaussian random variable (with any μ, σ) using the cdf of the standard Gaussian random variable ($\mu = 0$, $\sigma = 1$).

1.10 Central limit theorem

We consider sums of the form

$$S_n = X_1 + X_2 + \cdots + X_n,$$

where $X_1, \ldots, X_n$ are iid (independent, identically distributed) random variables. If X_i, $i = 1, \ldots, n$ have mean μ and variance σ^2, then in the limit $n \to \infty$, by the central limit theorem (CLT), S_n will have a cdf given by a Gaussian random variable with mean $n\mu$ and variance $n\sigma^2$. We make the following comments:

1. The CLT is very general (imposes no requirements on X_i).
2. X_i can be continuous or discrete random variables.
3. The sum of a large number of random variables is a Gaussian random variable.
4. The CLT eliminates the need to model X_i.
5. The independence of X_i is crucial but may be hard to prove or justify.

Example. *A machine processes parts, one at a time. Processing time for each part is a random variable* $X \sim \mathcal{U}(1,5)$. *Assume that there is independence of the processing times. What is the probability that* $S_{100} \leq 320$, *where* $S_100 = X_1 + \cdots + X_{100}$*?*

For each random variable X_i, we have $\mu = E[X] = 3$, $\sigma^2 = \text{var}(X) = 4/3$. Thus, the cdf of S_{100} will have a mean $\mu_{S_{100}} = 100\mu$ and variance $\sigma^2_{S_{100}} = 100\sigma^2$, and hence

$$P(S_{100} \leq 320) = \Phi\left(\frac{320 - 100\mu}{\sqrt{100}\sigma}\right) = \Phi(1.73) = 0.958.$$

1.11 Derived distributions

1.11.1 Functions of one random variable

Example. *Let* X *be a random variable with given pdf* f_X *and cdf* F_X. *Define* $Y = aX + b$ *where* $a > 0$, b *are deterministic real numbers. What are the pdf* f_Y *and cdf* F_Y *of* Y*?*

From Figure 1.11, we see that the cdf of Y is given by

$$F_Y(y) = P(Y \leq y) = P(aX + b \leq y) = P\left(X \leq \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right).$$

From there, to get the pdf of Y, one just needs to differentiate the cdf

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{d}{dy}F_X\left(\frac{y-b}{a}\right) = f_X\left(\frac{y-b}{a}\right)\frac{1}{a}.$$

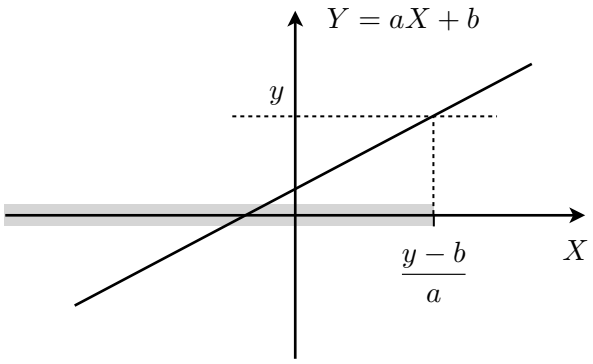


Figure 1.11 – $Y = aX + b$.

Example. *Now consider the case* $Y = 1/X$.

This time, the pdf of Y is given by

$$F_Y(y) = P(Y \leq y) = P\left(\frac{1}{X} \leq y\right).$$

As shown in Figure 1.12, there are two cases:

- For $y > 0$, we have

$$P\left(\frac{1}{X} \leq y\right) = P(X \leq 0) + P\left(X > \frac{1}{y}\right) = F_X(0) + 1 - F_X\left(\frac{1}{y}\right).$$

- For $y < 0$, we have

$$P\left(\frac{1}{X} \leq y\right) = P\left(\frac{1}{y} < X \leq 0\right) = F_X(0) - F_X\left(\frac{1}{y}\right).$$

Putting the pieces together,

$$F_Y(y) = \begin{cases} F_X(0) + 1 - F_X(1/y), & y > 0, \\ F_X(0) - F_X(1/y), & y < 0. \end{cases}$$

Taking the derivative of the cdf, we finally obtain the pdf of Y

$$f_Y(y) = \begin{cases} f_X(1/y)1/y^2, & y > 0, \\ f_X(1/y)1/y^2, & y < 0. \end{cases}$$

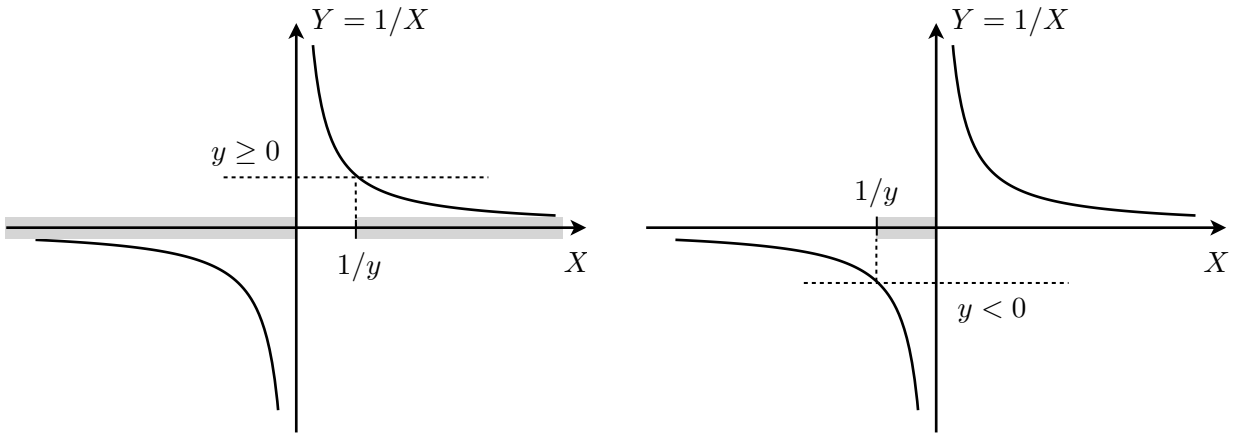


Figure 1.12 – $Y = 1/X$.

Example. *Next, consider the case* $Y = |X|$.

As before, the pdf of Y is given by

$$F_Y(y) = P(Y \leq y) = P\left(|X| \leq y\right).$$

As shown in Figure 1.13, there are two cases:

- For $y < 0$,

$$P(|X| \leq y) = 0.$$

- For $y \geq 0$,

$$P(|X| \leq y) = P(-y \leq X \leq y) = F_X(y) - F_X(-y).$$

Putting the piece back together and taking the derivative, we get

$$f_Y(y) = \begin{cases} f_X(y) + f_X(-y), & y > 0, \\ 0, & y < 0. \end{cases}$$

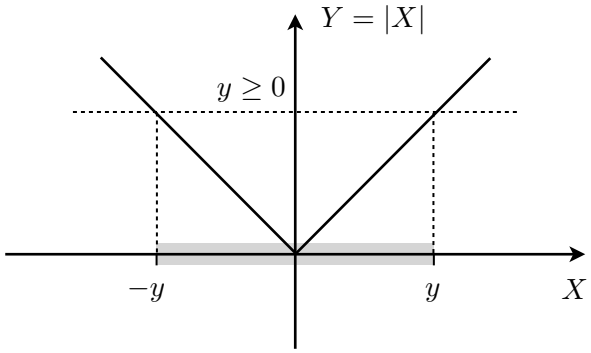


Figure 1.13 – $Y = |X|$.

For a general monotonic function, $Y = g(X)$, $g \in C^1$, g increasing as pictured in Figure 1.14,

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y)).$$

The pdf is then

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{d}{dy}P(Y \leq y) = \frac{d}{dy}F_X(g^{-1}(y)) = \left.\frac{1}{g'(x)}\right|_{x=g^{-1}(y)} f_X(g^{-1}(y)).$$

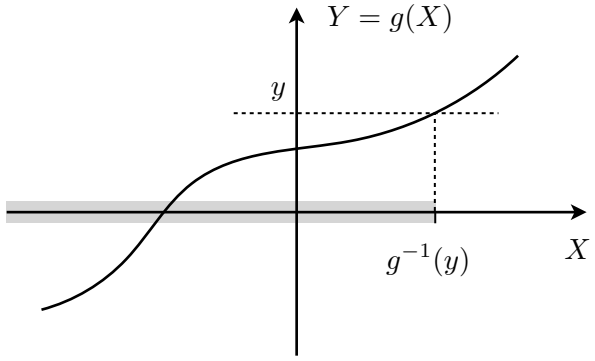


Figure 1.14 – General monotonic function $Y = g(X)$.

Finally, for an arbitrary, possibly non-monotonic function $Y = g(X)$, one needs to consider the number of solutions of $x = g^{-1}(y)$. If y is such that $x = g^{-1}(y)$ has multiple solutions $x_1(y)$, $x_2(y)$, $x_3(y)$, say, as pictured in Figure 1.15, then

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \leq x_1(y)) + P(x_2(y) \leq X \leq x_3(y)).$$

The pdf for this particular value of y follows as

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \left.\frac{1}{g'(x)}\right|_{x_1(y)} f_X(x_1(y)) - \left.\frac{1}{g'(x)}\right|_{x_2(y)} f_X(x_2(y)) + \left.\frac{1}{g'(x)}\right|_{x_3(y)} f_X(x_3(y)).$$

Another way is to cut the function $Y = g(X)$ into monotonic components and analyze them separately.

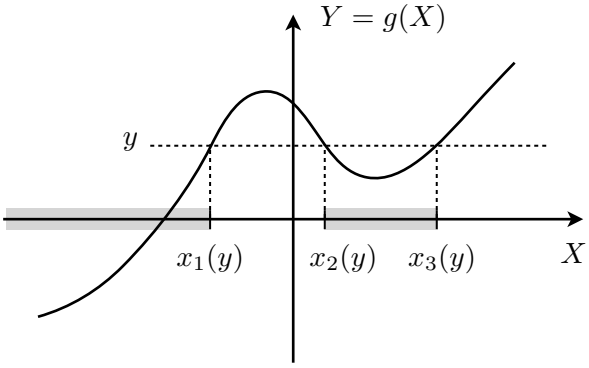


Figure 1.15 – Arbitrary non-monotonic function $Y = g(X)$.

1.11.2 Random number generator

Let $X \sim \mathcal{U}(0,1)$, $F_X(x) = x$, $x \in [0,1]$, with samples $u_1, u_2, \ldots, u_N$. Given a cumulative distribution function $\Phi(y)$, what should be the transformation $Y = g(X)$, $X \sim \mathcal{U}(0,1)$, so that $F_Y(y) = \Phi(y)$? We utilize the fact that

$$F_Y(y) = F_X(g^{-1}(y)) = g^{-1}(y).$$

Requiring that $F_Y(y) = \Phi(y)$, this implies

$$g(x) = \Phi^{-1}(x).$$

Then the samples $\Phi^{-1}(u_1), \Phi^{-1}(u_2), \ldots, \Phi^{-1}(u_N) \sim F_Y = \Phi$.

1.11.3 Functions of two random variables

Example. *Let X, Y with joint pdf $f_{XY}(x,y)$ and cdf $F_{XY}(x,y)$. Define $Z = X + Y$. What is the pdf $f_Z(z)$ and cdf $F_Z(z)$ of Z?*

From Figure 1.16, we see that the cdf of Z for fixed z is given by

$$F_Z(z) = P(Z \leq z) = P(X + Y \leq z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f_{XY}(x,y) dy dx.$$

The pdf is

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \int_{-\infty}^{\infty} f_{XY}(x, z-x) dx,$$

where we have applied the Leibniz rule

$$\frac{d}{dx}\left[\int_{a(x)}^{b(x)} f(x,t)dt\right] = f(x,b(x))\frac{db(x)}{dx} - f(x,a(x))\frac{da(x)}{dx} + \int_{a(x)}^{b(x)} \frac{\partial f(x,t)}{\partial x} dt.$$

Note that if X and Y are independent, $f_{XY} = f_X f_Y$. As a result, the pdf of Z is given by the convolution of the pdfs of X and Y

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx = f_X * f_Y.$$

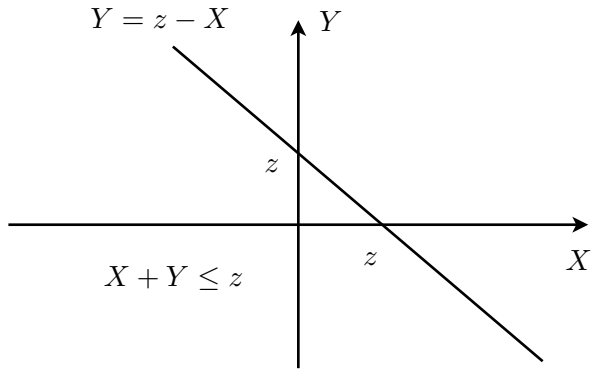


Figure 1.16 – $Z = X + Y$.

Example. *We now consider the transformation* $Z = X - Y$.

From Figure 1.17, we see that the cdf of Z for fixed z is given by

$$F_Z(z) = P(Z \leq z) = P(X - Y \leq z) = \int_{-\infty}^{\infty} \int_{x-z}^{\infty} f_{XY}(x, y) dy dx.$$

The pdf is

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \int_{-\infty}^{\infty} f_{XY}(x, x - z) dx.$$

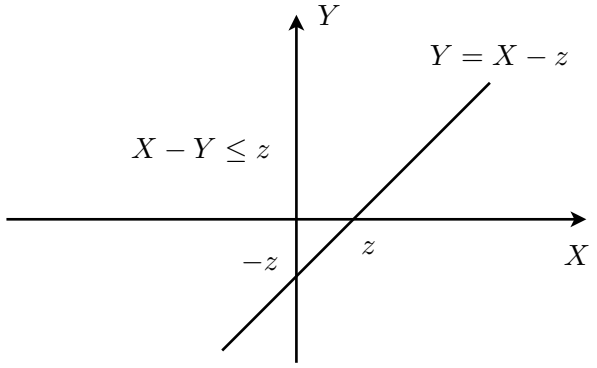


Figure 1.17 – $Z = X - Y$.

Example. *Next, we consider* $Z = XY$.

As shown in Figure 1.18, we need to consider two cases:

- For fixed $z > 0$, the cdf of Z is given by

$$F_Z(z) = P(Z \leq z) = P(XY \leq z) = \int_{-\infty}^{0} \int_{z/x}^{\infty} f_{XY}(x, y) dy dx + \int_{0}^{\infty} \int_{-\infty}^{z/x} f_{XY}(x, y) dy dx,$$

  from which the pdf follows as

$$f_Z(z) = -\int_{-\infty}^{0} \frac{1}{x} f_{XY}(x, z/x) dx + \int_{0}^{\infty} \frac{1}{x} f_{XY}(x, z/x) dx = \int_{-\infty}^{\infty} \frac{1}{|x|} f_{XY}(x, z/x) dx, \quad z > 0.$$

- For fixed $z < 0$, the cdf of Z is given by

$$F_Z(z) = P(Z \le z) = P(XY \le z) = \int_{-\infty}^{0} \int_{z/x}^{\infty} f_{XY}(x,y)dydx + \int_{0}^{\infty} \int_{-\infty}^{z/x} f_{XY}(x,y)dydx,$$

from which the pdf follows as

$$f_Z(z) = -\int_{-\infty}^{0} \frac{1}{x} f_{XY}(x, z/x)dx + \int_{0}^{\infty} \frac{1}{x} f_{XY}(x, z/x)dx = \int_{-\infty}^{\infty} \frac{1}{|x|} f_{XY}(x, z/x)dx.$$

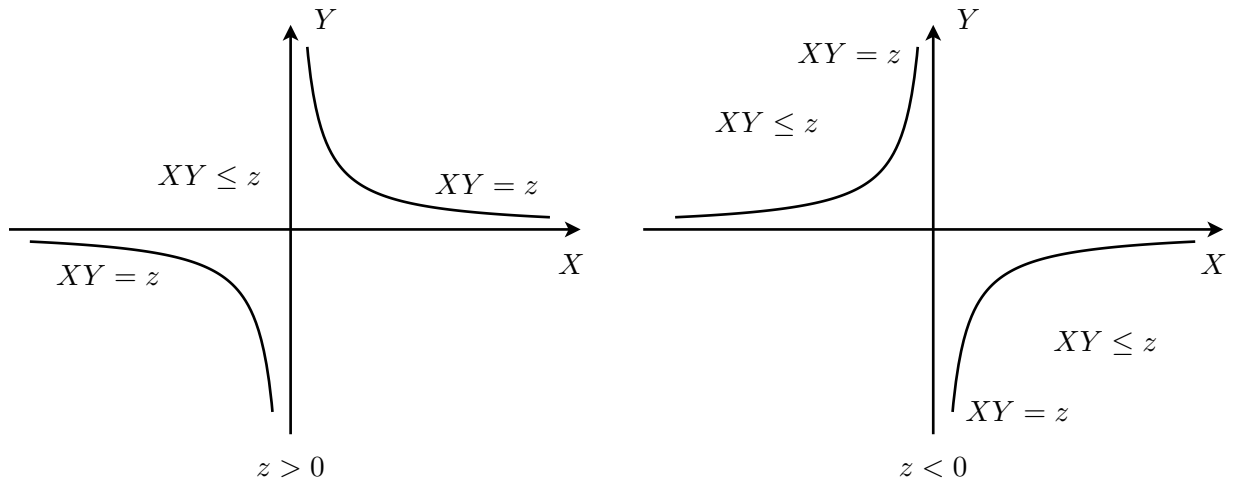


Figure 1.18 – $Z = XY$.

Example. *Next, we consider the case* $Z = \max\{X, Y\}$*, pictured in Figure 1.19.*

We note that $\max\{X, Y\} \le z \Rightarrow X \le z$ *and* $Y \le z$, hence

$$F_Z(z) = P(Z \le z) = P(\max\{X, Y\} \le z) = \int_{-\infty}^{z} \int_{-\infty}^{z} f_{XY}(x,y)dydx.$$

The pdf is

$$f_Z(z) = \int_{-\infty}^{z} f_{XY}(x,z)dx + \int_{-\infty}^{z} f_{XY}(z,y)dy.$$

Example. *Next, we consider the case* $Z = \min\{X, Y\}$*, pictured in Figure 1.20.*

We note that $\min\{X, Y\} \le z \Rightarrow X \le z$ *or* $Y \le z$. On the other hand, $\min\{X, Y\} \ge z \Rightarrow X \ge z$ *and* $Y \ge z$, hence we write

$$F_Z(z) = P(Z \le z) = P(\min\{X, Y\} \le z) = 1 - P(\min\{X, Y\} \ge z) = 1 - \int_{z}^{\infty} \int_{z}^{\infty} f_{XY}(x,y)dydx.$$

The pdf is

$$f_Z(z) = \int_{z}^{\infty} f_{XY}(x,z)dx + \int_{z}^{\infty} f_{XY}(z,y)dy.$$

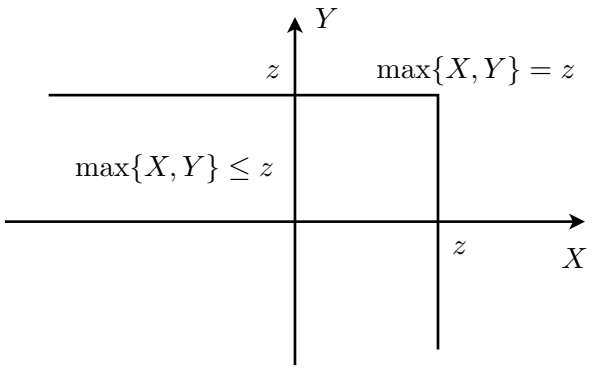


Figure 1.19 – $Z = \max\{X, Y\}$.

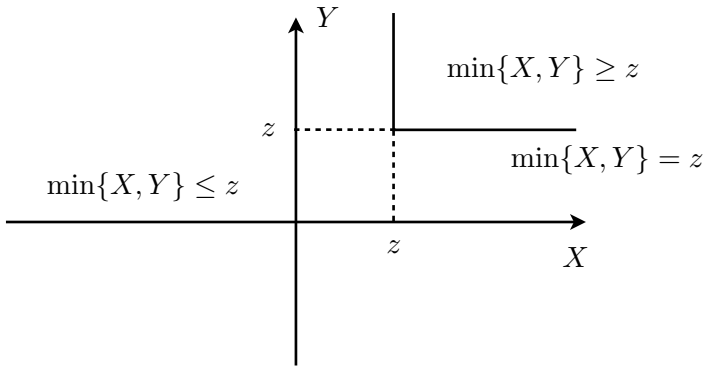


Figure 1.20 – $Z = \min\{X, Y\}$.

Finally, in the general case $Z = H(X, Y)$, we need to find the contour levels $H(X, Y) = z$, as pictured in Figure 1.21, then we have

$$F_Z(z) = P(H(X, Y) \leq z).$$

Example. *Let X, Y independent random variables that follow an exponential distribution*

$$f_X(x) = \begin{cases} \frac{1}{\theta} e^{-x/\theta}, & x \geq 0, \\ 0, & x < 0, \end{cases}$$

and likewise for Y. Compute the pdf of their difference.

Defining $Z = X - Y$, we have the pdf

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(x - z) dx.$$

It is now crucial to find the limits of integration. We note that the integrand is nonzero if and only if $f_X(x) \neq 0$ and $f_Y(x - z) \neq 0$, that is, $x \geq 0$ and $x - z \geq 0$. Depending on the sign of z, there can be two cases:

- If $z > 0$, then the range of integration is $x \geq z$, thus

$$f_Z(z) = \int_z^{\infty} \frac{1}{\theta} e^{-x/\theta} \frac{1}{\theta} e^{-(x-z)/\theta} dx = \frac{1}{\theta^2} e^{z/\theta} \frac{(-\theta)}{2} \, e^{-2x/\theta} \Big|_z^{\infty} = \frac{1}{2\theta} e^{-z/\theta}.$$

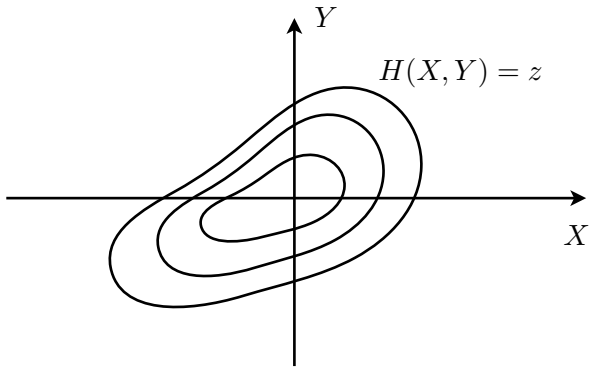


Figure 1.21 – $Z = \min\{X, Y\}$.

- If $z < 0$, then the range of integration is $x \geq 0$, thus

$$f_Z(z) = \int_0^\infty \frac{1}{\theta} e^{-x/\theta} \frac{1}{\theta} e^{-(x-z)/\theta} dx = \frac{1}{\theta^2} e^{z/\theta} \frac{(-\theta)}{2} \, e^{-2x/\theta} \Big|_0^\infty = \frac{1}{2\theta} e^{z/\theta}.$$

Combining the two pieces, we arrive at

$$f_Z(z) = \frac{1}{2\theta} e^{-|z|/\theta}, \quad z \in \mathbb{R}.$$

1.12 Random sequences

Consider a sequence of random variables $X_1, X_2, \ldots, X_N$. To see how this differs from a single random variables, consider the following questions:

1. Given the values of $X_1, X_2, \ldots, X_N$, can we say something about X_{N+1}? This involves the notion of memory.
2. What can we say about long term averages?
3. How large does N need to be in order for X_N to have a given property?

1.12.1 Bernoulli process

A Bernoulli process is a sequence X_i of independent trials, with

$$P(X_i = 1) = P(\text{success at the } i\text{th trial}) = p,$$
$$P(X_i = 0) = P(\text{failure at the } i\text{th trial}) = 1 - p.$$

This is the simplest case of a random sequence: X_{i+1} is independent from X_i (no memory). We have the following properties:

1. The probability distribution for the number S of successes in n independent trials is given by the *binomial distribution*

$$P_S(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n,$$

where the binomial coefficient is given by

$$\binom{n}{k} = \frac{n!}{(n-k)!k!}.$$

As expected, we have

$$\sum_{k=0}^{n} P_S(k) = 1.$$

The expectation and variance of S are

$$E[S] = \sum_{k=0}^{n} kP_S(k) = np, \qquad \text{var}(S) = np(1-p).$$

2. The probability distribution for the number T of trials up to (and including) the first success is given by the *geometric distribution*

$$P_T(t) = (1-p)^{t-1}p, \quad t = 1, 2, \ldots$$

The expectation and variance of T are

$$E[T] = \frac{1}{p}, \qquad \text{var}(T) = \frac{1-p}{p^2}.$$

Example. *Roll a dice 5 times. What is the probability of having 3 times the face with 6?*
We have $p = 1/6$, $q = 1 - p = 5/6$, $n = 5$, $k = 3$, therefore

$$P_S(k=3) = \binom{5}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2 = 0.03215.$$

Example. *Roll two dice 4 times. What is the probability that 7 will never be an outcome?*
The following 6 pairs correspond to the outcome being 7

$$(1,6),\ (2,5),\ (3,4),\ (4,3),\ (5,2),\ (6,1).$$

As a result, $p = 6/36 = 1/6$ and $q = 1 - p = 5/6$. Since $n = 4$ and $k = 0$, we get

$$P_S(k=0) = \binom{4}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^4 = 0.4823.$$

Let us now ask a different question. What is the probability distribution of Y_k, the total time until the kth success?

We have that

$$Y_k = T_1 + T_2 + \cdots + T_k,$$

where T_i represents the time from the $(i-1)$th to the ith success. Because the trials are independent, T_i follows the same probability distribution as T, the time to the first success, hence we have

$$E[Y_k] = \sum_{i=1}^{k} E[T_i] = \frac{k}{p}, \qquad \text{var}(Y_k) = \frac{k(1-p)}{p^2}.$$

Denoting $P_{Y_k}(t)$ the probability that $Y_k = t$, that is, the kth success happens at time t, we trivially have $P_{Y_k}(t) = 0$ for $t < k$. For $t \geq k$, we have

$$\begin{aligned} P_{Y_k}(t) &= \begin{pmatrix} \text{probability of} \\ \text{success at time } t \end{pmatrix} \cdot \begin{pmatrix} \text{probability that there are } k-1 \\ \text{successes until time } t-1 \end{pmatrix} \\ &= p\binom{t-1}{k-1} p^{k-1}(1-p)^{t-k}. \end{aligned}$$

Hence,

$$P_{Y_k}(t) = \begin{cases} \binom{t-1}{k-1} p^k (1-p)^{t-k}, & t \geq k, \\ 0, & t < k. \end{cases}$$

Example. *In a certain game, during each minute a foul occurs with probability p, and no foul with probability $1-p$. The time is discrete (counted as minutes). A foul in each minute is independent of the rest of the game. Knowing that a player has to leave the game after 6 fouls or 30 mins (whichever comes first), what is the probability of the player staying for 30 mins?*

We model the occurrence of a foul every minute using a Bernoulli process, where $P(\text{foul}) = p$. We note that the sum of $P_{Y_6}(t)$ for $t = 6, 7, \ldots, 29$ gives the probability that 6 fouls occur within the first 29 mins (or less). Therefore,

$$P(\text{6th foul occurs over 30 min}) = 1 - \sum_{t=6}^{29} P_{Y_6}(t).$$

Note that this probability is also equal to

$$P(\text{6th foul occurs over 30 min}) = \sum_{k=0}^{5} P_S(k),$$

where $P_S(k)$ is the probability that k fouls occur during the first $n = 30$ mins.

1.12.2 Poisson process

A Poisson process is the continuous time analogue of a Bernoulli process, and thus applicable to cases where there is no natural way to discretize time. Formally, an arrival process is called Poisson with rate λ if it has the following properties:

1. Time homogeneity: the probability $P(k, \tau)$ of having k arrivals in an interval τ is independent of the interval.

2. Independence: number of arrivals during a particular interval does not depend on previous times.

3. Small interval properties: $P(k, \tau), \tau \ll 1$, has the following properties:

   (a) $P(0, \tau) = 1 - \lambda\tau + O(\tau^2)$,
   (b) $P(1, \tau) = \lambda\tau + O(\tau^2)$,
   (c) $P(k, \tau) = O(\tau^2),\ k = 2, 3, \ldots$

What is the probability of having k arrivals in a finite time interval of length τ? As shown in Figure 1.22, we discretize the finite time interval into infinitesimal intervals of length δ. As a result, we end up with $n = \tau/\delta$ such intervals. In Bernoulli terms, we would have $n = \tau/\delta$ trials and we seek the probability $P_S(k)$ of having k arrivals with $P(\text{success}) = \lambda\delta$. From the binomial distribution, we have

$$\begin{aligned} P_S(k) &= \frac{n!}{(n-k)!k!}(\lambda\delta)^k(1-\lambda\delta)^{n-k} \\ &= \frac{n!}{(n-k)!k!}\frac{(\lambda\tau)^k}{n^k}\left(1-\frac{\lambda\tau}{n}\right)^{n-k} \\ &= \frac{n(n-1)(n-2)\dots(n-k+1)}{n \quad n \quad n \quad \dots \quad n}\frac{(\lambda\tau)^k}{k!}\left(1-\frac{\lambda\tau}{n}\right)^{n}\left(1-\frac{\lambda\tau}{n}\right)^{-k}. \end{aligned}$$

Taking the limit $\delta \to 0$, or $n \to \infty$, we obtain the probability of having k arrivals in an interval τ for a Poisson process

$$P(k,\tau) = \lim_{n\to\infty} P_S(k) = \frac{(\lambda\tau)^k}{k!}e^{-\lambda\tau}.$$

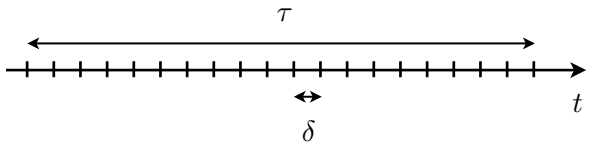


Figure 1.22 – Discretization of time interval for Poisson process.

Let us check the small interval properties for $\lambda\tau \ll 1$

(a) $P(0,\tau) = e^{-\lambda\tau} = 1 - \lambda\tau + O(\tau^2)$,
(b) $P(1,\tau) = \lambda\tau e^{-\lambda\tau} = \lambda\tau + O(\tau^2)$,
(c) $P(k,\tau) = O(\tau^2),\ k = 2,3,\dots$

Let N_τ the number of arrivals during τ. Then,

$$E[N_\tau] = \lambda\tau, \qquad \text{var}(N_\tau) = \lambda\tau.$$

Finally, define T as the time for the first arrival. Its cdf is given by

$$F_T(t) = P(T \le t) = 1 - P(T > t) = 1 - P(0,t) = 1 - e^{-\lambda t},$$

since $P(T > t) = P(\text{no arrival during } [0,t]) = P(0,t)$. The pdf of T follows as

$$f_T(t) = \frac{dF_T(t)}{dt} = \lambda e^{-\lambda t}.$$

Note that because of the independence property of a Poisson process, the inter-arrival time T between any two arrivals follows the same distribution as the time for the first arrival.

Chapter 2

Stochastic Processes and Linear Systems

2.1 Random processes

Recall the definition of a random variable. From a random event ζ, we define a random variable $X(\zeta)$ by assigning to each event A_i a number $X(A_i)$

$$\left.\begin{array}{lcl} P_1, A_1 & \rightarrow & X_1 \\ P_2, A_2 & \rightarrow & X_2 \\ \vdots & & \\ P_n, A_n & \rightarrow & X_n \end{array}\right\} \quad \longrightarrow \quad P(\zeta), X(\zeta).$$

Similarly, we define a random process, or stochastic process, $X(t, \zeta)$ by assigning for each event A_o a function $X(A_i, t)$

$$\left.\begin{array}{lcl} P_1, A_1 & \rightarrow & X_1(t) \\ P_2, A_2 & \rightarrow & X_2(t) \\ \vdots & & \\ P_n, A_n & \rightarrow & X_n(t) \end{array}\right\} \quad \longrightarrow \quad P(\zeta), X(t, \zeta).$$

A stochastic process $X(t, \zeta)$ can be interpreted in the two following ways:

1. For a given event ζ_0, the outcome function $X(t, \zeta_0)$ is a regular deterministic function of time.
2. For every $t = t_0$, the quantity $X(t_0, \zeta)$ is a random variable.

Example. Roll a dice. To each dice outcome A_i, we associate one of the functions of time shown in Figure 2.1.

Example. Roll a dice, then associate the following function to the outcome A_i

$$A_i \quad \longrightarrow \quad X(t, i) = t^i, \quad i = 1, 2, \ldots, 6.$$

2.2 Averages

There are two types of averages of a stochastic process one can perform.

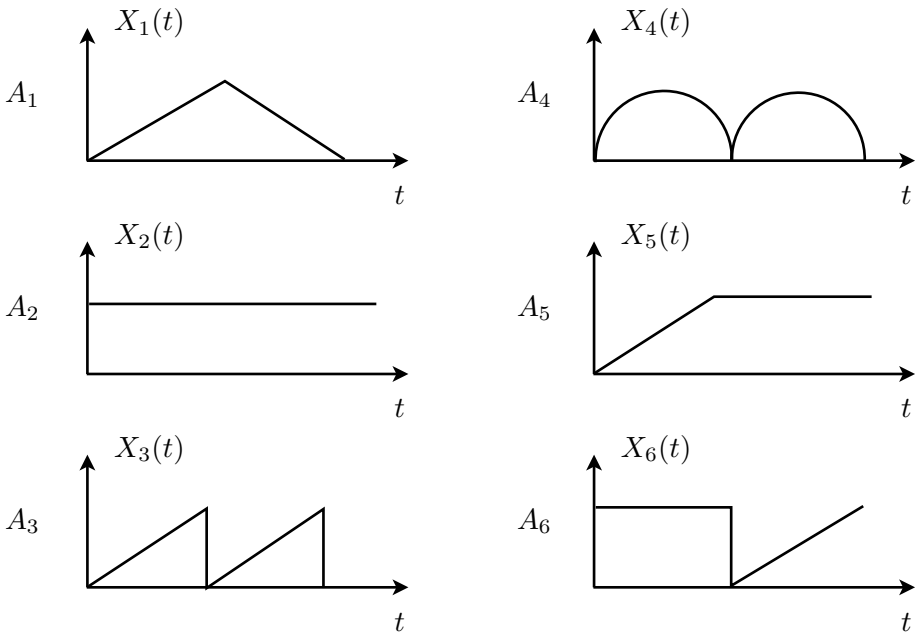


Figure 2.1 – Stochastic process example: roll a dice then pick a function of time accordingly.

2.2.1 Time averages (temporal moments)

Time averages are performed over time t for a fixed ζ, such as the following temporal mean and variance

$$m_X(\zeta) = \lim_{T\to\infty} \frac{1}{T} \int_0^T X(t,\zeta)dt,$$
$$V_X(\zeta) = \lim_{T\to\infty} \frac{1}{T} \int_0^T [X(t,\zeta) - m_X(\zeta)]^2 dt.$$

Note that these are random variables. We also define the time correlation,

$$R_X(\tau,\zeta) = \lim_{T\to\infty} \frac{1}{T} \int_0^T X(t,\zeta)X(t+\tau,\zeta)dt,$$

which measures how much, for a given ζ, the function $X(t,\zeta)$ at t is correlated with itself at $t+\tau$ (on average over time); see Figure 2.2.

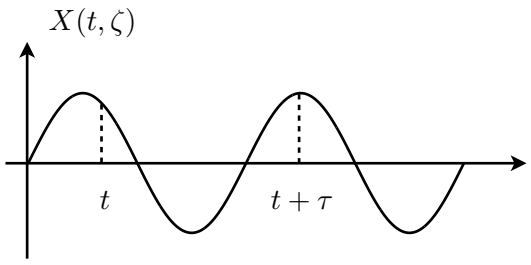


Figure 2.2 – Time correlation.

2.2.2 Ensemble averaging (statistical moments)

Ensemble averages are performed over the events ζ for a fixed t, as if the stochastic process were a random variable. For some fixed time $t = t_0$, the ensemble mean and variance are

$$\begin{aligned} m_X(t_0) &= E[X(t_0, \zeta)], \\ V_X(t_0) &= E\left[[X(t_0, \zeta) - m_X(t_0)]^2\right]. \end{aligned}$$

For two fixed times t_1, t_2, the autocorrelation and autocovariance functions are, respectively,

$$\begin{aligned} R_{XX}(t_1, t_2) &= E[X(t_1, \zeta)X(t_2, \zeta)], \\ C_{XX}(t_1, t_2) &= E\left[[X(t_1, \zeta) - m_X(t_1)][X(t_2, \zeta) - m_X(t_2)]\right]. \end{aligned}$$

Example. Roll a dice, outcome denoted as $\zeta = 1, \ldots, 6$ with probability $P(\zeta = i) = 1/6$. Define the stochastic process

$$X(t, \zeta) = a\cos(\zeta\omega_0 t),$$

where a and ω_0 are fixed. First, calculate the time averages for a fixed ζ, viewing $X(t, \zeta)$ as a function of time

$$\begin{aligned} m_X(\zeta) &= \lim_{T\to\infty} \frac{1}{T}\int_0^T a\cos(\zeta\omega_0 t)dt = 0, \\ V_X(\zeta) &= \lim_{T\to\infty} \frac{1}{T}\int_0^T a^2\cos^2(\zeta\omega_0 t)dt = \frac{a^2}{2}. \end{aligned}$$

Next, perform statistics for a fixed time t_0, viewing $X(t, \zeta)$ as a random variable

$$\begin{aligned} m_X(t_0) = E[X(t_0, \zeta)] &= \sum_{i=1}^{6} P(\zeta = i)X(t_0, i) = \sum_{i=1}^{6} \frac{1}{6} a\cos(i\omega_0 t_0), \\ V_X(t_0) = C_{XX}(t_0, t_0) &= \sum_{i=1}^{6} P(\zeta = i)[X(t_0, i) - m_X(t_0)]^2 = \sum_{i=1}^{6} \frac{1}{6}[a\cos(i\omega_0 t_0) - m_X(t_0)]^2. \end{aligned}$$

In general, time averages and ensemble averages are *not* equal.

2.2.3 Moments of derivatives and integrals

For a given stochastic process $X(t, \zeta)$, the ensemble average $m_X(t) = E[X(t, \zeta)]$ is a function of time and satisfies the following property

$$E[\alpha X(t, \zeta) + \beta] = \alpha E[X(t, \zeta)] + \beta,$$

where α and β are deterministic numbers or functions. Defining

$$Y(t, \zeta) = X'(t, \zeta) = \frac{dX(t, \zeta)}{dt},$$

the ensemble average of $Y(t, \zeta)$ is therefore

$$m_Y(t) = E[Y(t, \zeta)] = E\left[\frac{dX(t, \zeta)}{dt}\right] = \frac{d}{dt}E[X(t, \zeta)] = m_X'(t),$$

and its autocorrelation function is

$$\begin{aligned} R_{YY}(t_1,t_2) &= E\left[\left.\frac{dX(t,\zeta)}{dt}\right|_{t=t_1}\left.\frac{dX(s,\zeta)}{ds}\right|_{s=t_2}\right] \\ &= \left.\frac{\partial^2}{\partial t\partial s}E[X(t,\zeta)X(s,\zeta)]\right|_{t=t_1,s=t_2} \\ &= \left.\frac{\partial^2 R_{XX}(t,s)}{\partial t\partial s}\right|_{t=t_1,s=t_2}. \end{aligned}$$

Similar properties hold for the integral operator. Specifically, let $Z(t,\zeta)=\int_0^t X(s,\zeta)ds$. Then we will have:

$$m_Z(t) = E[\int_0^t X(s,\zeta)ds] = \int_0^t E[X(s,\zeta)]ds = \int_0^t m_X(s)ds.$$

Moreover,

$$\begin{aligned} R_{ZZ}(t_1,t_2) &= E[\int_0^{t_1} X(s_1,\zeta)ds_1 \int_0^{t_2} X(s_2,\zeta)ds_2] \\ &= \int_0^{t_1}\int_0^{t_2} E[X(s_1,\zeta)X(s_2,\zeta)]ds_1ds_2 \\ &= \int_0^{t_1}\int_0^{t_2} R_{XX}(s_1,s_2)ds_1ds_2. \end{aligned}$$

Note that $R_{YY}(t_1,t_2)$ is an example of two-point/time statistics. More generally, single time statistics such as $m_X(t), V_X(t), \ldots$ are described by the cdf

$$F_{X_t}(x) = P(X(t,\zeta)\le x),$$

and two-time statistics such as $R_{XX}(t_1,t_2), C_{XX}(t_1,t_2), \ldots$ are described by the joint cdf

$$F_{X_{t_1}X_{t_2}}(x_1,x_2) = P(X(t_1,\zeta)\le x_1, X(t_2,\zeta)\le x_2).$$

One can then differentiate the cdfs to obtain the pdfs, and hence the statistics $m_X(t),\ldots$

2.3 Stationary stochastic process

A stationary stochastic process $X(t,\zeta)$ is one whose statistics do not depend on time (origin of time). As a result, the ensemble average of such a process is constant,

$$m_X(t) = E[X(t,\zeta)] = m_X,$$

and the autocorrelation function does not dependent on t, just the difference $\tau = t_2 - t_1$,

$$R_{XX}(t,t+\tau) = R_{XX}(\tau).$$

2.3.1 Strongly stationary stochastic process

A strongly stationary stochastic process satisfies the following. Consider the families of joint random variables

$$\begin{gathered} X(t_1,\zeta), X(t_2,\zeta),\ldots,X(t_N,\zeta), \\ X(t_1+h,\zeta), X(t_2+h,\zeta),\ldots,X(t_N+h,\zeta), \end{gathered}$$

then the joint statistics (pdf) over $t_1,t_2,\ldots,t_N$ remain the same for all $t_1,\ldots,t_N$, for any N and any h.

2.3.2 Weakly or 2nd order stationary stochastic process

A weakly or 2nd order stationary stochastic process satisfies the following less stringent conditions

$$m_X(t) = m_X = \text{constant},$$
$$R_{XX}(t_1, t_2) = R_{XX}(\tau), \quad \tau = t_1 - t_2,$$

for all t, t_1, t_2. Note that strong stationarity implies weak stationarity, since

$$m_X(t) = E[X(t,\zeta)] = E[X(t+h,\zeta)] = m_X,$$
$$R_{XX}(t, t+\tau) = E[X(t,\zeta)X(t+\tau,\zeta)] = E[X(t+h,\zeta)X(t+h+\tau,\zeta)] = R_{XX}(\tau),$$

where we have chosen $h = -t$ to obtain the last equality.

Example. Consider the stochastic process

$$Y(t,\zeta) = a\cos(\omega t + \theta(\zeta)),$$

where $\theta \sim (0, 2\pi)$; see its pdf in Figure 2.3. Is $Y(t,\zeta)$ weakly stationary? We calculate the ensemble mean

$$m_Y(t) = E[Y(t,\zeta)] = E[a\cos(\omega t + \theta(\zeta))] = a\int_0^{2\pi} \cos(\omega t + \theta)\frac{1}{2\pi}d\theta = 0,$$

where we have used the formula $E[g(X)] = \int g(x) f_X(x)dx$. We also calculate the autocorrelation function

$$\begin{aligned} R_{YY}(t, t+\tau) &= E[Y(t,\zeta)Y(t+\tau,\zeta)] \\ &= a^2 E[\cos(\omega t + \theta(\zeta))\cos(\omega t + \omega\tau + \theta(\zeta))] \\ &= a^2 \int_0^{2\pi} \cos(\omega t + \theta)\cos(\omega t + \omega\tau + \theta)\frac{1}{2\pi}d\theta \\ &= a^2 \int_0^{2\pi} \frac{1}{2}\cos(\omega\tau)\frac{1}{2\pi}d\theta + a^2\int_0^{2\pi}\frac{1}{2}\cos(2\omega t + \omega\tau + 2\theta)\frac{1}{2\pi}d\theta \\ &= \frac{1}{2}a^2\cos\omega\tau, \end{aligned}$$

where we have used the formula $\cos A\cos B = 1/2[\cos(A-B) + \cos(A+B)]$. Since $m_Y(t)$ does not depend on time and $R_{YY}(t, t+\tau) = R_{YY}(\tau)$, the process $Y(t,\zeta)$ is weakly stationary. Figure 2.4 shows the autocorrelation function $R_{YY}(\tau)$.

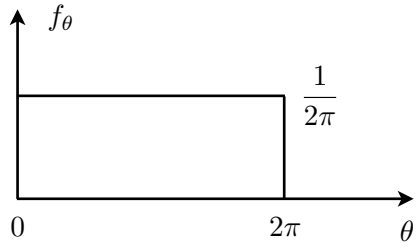


Figure 2.3 – Uniform pdf $\theta \sim (0, 2\pi)$.

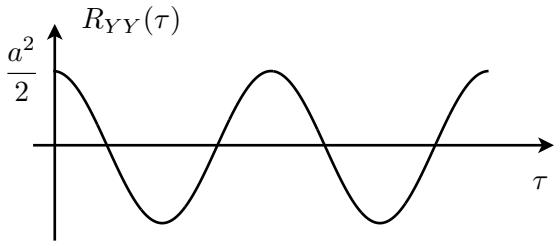


Figure 2.4 – Autocorrelation function of $Y(t,\zeta) = a\cos(\omega t + \theta(\zeta))$.

2.3.3 Properties of the autocorrelation function for stationary processes

Let $Y(t,\zeta)$ be a stationary stochastic process with $E[Y(t,\zeta)] = 0$. (If $E[Y(t,\zeta)] \neq 0$, one can always define $Y' = Y - m_Y$). Clearly all the below properties hold for the covariance function, C_{YY}.

The autocorrelation function $R_{YY}(\tau) = E[Y(t,\zeta)Y(t+\tau,\zeta)]$ satisfies the following properties:

1. $R_{YY}(0) = E[Y(t,\zeta)^2] = \sigma_Y^2 \geq 0$ does not depend on time.
2. $R_{YY}(\tau) = R_{YY}(-\tau)$ since $E[Y(t,\zeta)Y(t+\tau,\zeta)] = E[Y(t_1,\zeta)Y(t_1-\tau,\zeta)]$ and set $t_1 = t - \tau$.
3. $R_{YY}(0) \geq |R_{YY}(\tau)|$. Indeed,

$$\begin{aligned} & E\left[(Y(t,\zeta) \pm Y(t+\tau,\zeta))^2\right] \geq 0 \\ & \Rightarrow E\left[Y(t,\zeta)^2 + Y(t+\tau,\zeta)^2 \pm 2Y(t,\zeta)Y(t+\tau,\zeta)\right] \geq 0 \\ & \Rightarrow 2R_{YY}(0) \pm 2R_{YY}(\tau) \geq 0, \quad \text{using stationarity} \\ & \Rightarrow R_{YY}(0) \geq R_{YY}(\tau) \geq -R_{YY}(0). \end{aligned}$$

2.4 Ergodicity

It is not always convenient or practical to calculate ensemble statistics of a stochastic process since this requires many realizations of the stochastic process. For stationary stochastic processes, the notion of ergodicity relates to whether it is possible to substitute ensemble averages over many realizations to time averages over a single realization. More precisely, a stochastic process $X(t,\zeta)$ which is stationary and its time averages equal the corresponding statistics is called ergodic.

Example. Consider the following stochastic process. We flip a coin, then

$$\begin{aligned} H \quad &\longrightarrow \quad X_1(t) = 1 \text{ for all } t, \\ T \quad &\longrightarrow \quad X_2(t) = 2 \text{ for all } t. \end{aligned}$$

Clearly, $X(t,\zeta)$ is a stationary random process. Is it ergodic? We calculate the time average

$$m_X(\zeta) = \lim_{T\to\infty} \int_0^T \frac{1}{T} X(t,\zeta)dt = \begin{cases} 1, & \text{if } H, \\ 2, & \text{if } T, \end{cases}$$

and the ensemble average

$$m_X(t) = \frac{1}{2}X_1(t) + \frac{1}{2}X_2(t) = 1.5.$$

Thus, this stochastic process is not ergodic. In general, the statistics of a stationary stochastic process are not necessarily equal to the time averages and ergodicity is not guaranteed.

Example. Consider the stochastic process

$$Y(t,\zeta) = a\cos(\omega t + \theta(\zeta)),$$

where $\theta \sim \mathcal{U}(0, 2\pi)$ as before. We saw previously that $Y(t,\zeta)$ is stationary, with ensemble mean and autocorrelation function given by, respectively,

$$\begin{aligned} m_Y(t) &= 0, \\ R_{YY}(\tau) &= \frac{1}{2}a^2\cos\omega\tau. \end{aligned}$$

Let us now calculate the corresponding time averages. The time mean is

$$m_Y(\zeta) = \lim_{T\to\infty}\frac{1}{T}\int_0^T a\cos(\omega t + \theta(\zeta))dt = \lim_{T\to\infty}\frac{a}{T\omega}\sin(\omega T + \theta) = 0,$$

and the time correlation is

$$\begin{aligned} R_Y(\tau,\zeta) &= \lim_{T\to\infty}\frac{1}{T}\int_0^T a^2\cos(\omega t + \theta(\zeta))\cos(\omega t + \omega\tau + \theta(\zeta))dt \\ &= \lim_{T\to\infty}\frac{1}{T}\frac{a^2}{2}\int_0^T \left[\cos\omega\tau + \cos(2\omega t + \omega\tau + 2\theta)\right]dt \\ &= \lim_{T\to\infty}\frac{a^2}{2T}\left[T\cos\omega\tau + \frac{1}{2\omega}\sin(2\omega t + \omega\tau + 2\theta)\right] \\ &= \frac{a^2}{2}\cos\omega\tau. \end{aligned}$$

Since the time averages equal the ensemble averages, this stochastic process is ergodic.

Example. We now generalize the above example to multiple frequencies with the following stochastic process

$$Z(t,\zeta) = \sum_{i=1}^N a_i\cos(\omega_i t + \theta_i(\zeta)),$$

where a_i and ω_i are deterministic, and $\theta_i \sim \mathcal{U}(0, 2\pi)$ for $i = 1, \ldots, N$ are independent random variables. Because each term in the sum is independent and uniformly distributed, the central limit theorem applies and tells us that the pdf f_Z of $Z(t,\zeta)$ for fixed time t is Gaussian. We now calculate the ensemble average

$$m_Z(t) = E[Z(t,\zeta)] = E\left[\sum_{i=1}^N a_i\cos(\omega_i t + \theta_i(\zeta))\right] = \sum_{i=1}^N a_i E\left[\cos(\omega_i t + \theta_i(\zeta))\right] = 0,$$

and the autocorrelation function

$$\begin{aligned} R_{ZZ}(t, t+\tau) &= E\left[\left(\sum_{i=1}^N a_i\cos(\omega_i t + \theta_i(\zeta))\right)\left(\sum_{k=1}^N a_k\cos(\omega_k t + \omega_k\tau + \theta_k(\zeta))\right)\right] \\ &= E\left[\sum_{i=1}^N\sum_{k=1}^N a_i a_k\cos(\omega_i t + \theta_i(\zeta))\cos(\omega_k t + \omega_k\tau + \theta_k(\zeta))\right] \end{aligned}$$

$$= \sum_{i=1}^{N} \sum_{k=1}^{N} a_i a_k E\left[\cos(\omega_i t + \theta_i(\zeta)) \cos(\omega_k t + \omega_k \tau + \theta_k(\zeta))\right].$$

Let us denote

$$Q_{ik} = \cos(\omega_i t + \theta_i(\zeta)) \cos(\omega_k t + \omega_k \tau + \theta_k(\zeta)).$$

If $i = k$, we have seen that

$$E^{\zeta}[Q_{ii}] = \frac{1}{2} \cos \omega_i \tau.$$

If $i \neq k$, we have

$$\begin{aligned} E[Q_{ik}] &= \int_0^{2\pi} \int_0^{2\pi} \cos(\omega_i t + \theta_i) \cos(\omega_k t + \omega_k \tau + \theta_k) \frac{1}{(2\pi)^2} d\theta_i d\theta_k \\ &= \frac{1}{(2\pi)^2} \int_0^{2\pi} \cos(\omega_i t + \theta_i) d\theta_i \int_0^{2\pi} \cos(\omega_k t + \omega_k \tau + \theta_k) d\theta_k \\ &= 0, \end{aligned}$$

where we have used independence to write $f_{\theta_i \theta_k} = f_{\theta_i} f_{\theta_k}$. Thus, we finally obtain

$$R_{ZZ}(t, t+\tau) = \sum_{i=1}^{N} \frac{a_i^2}{2} \cos \omega_i \tau.$$

The stochastic process $Z(t, \zeta)$ is therefore stationary. It is left as an exercise to the reader to show that the corresponding time averages are equal to the ensemble averages calculated above; hence $Z(t, \zeta)$ is also ergodic.

2.5 Linear Time-Invariant (LTI) systems

A system is a set of physical objects that interact. Modelling is the process of representing the behavior of the system in terms of equations. A system usually has

- inputs, which represent the external actions/signals influencing the system, and
- outputs, which represent the quantities of interest

As shown in Figure 2.5, we define the map F from the input $u(t)$ to the output $y(t)$, that is,

$$y(t) = F[u(t)].$$

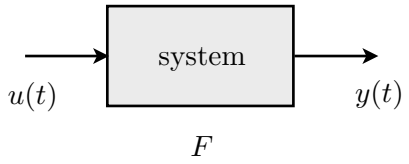


Figure 2.5 – An LTI system with input $u(t)$, output $y(t)$, and map F such that $y(t) = F[u(t)]$.

2.5.1 Dynamical systems

Dynamical systems are systems for which the output depends on present and past values of the input.

Examples. The following are dynamical systems:

1. $y(t) = \int_{t-3}^{t} u^3(\tau)d\tau$
2. $y(t) = u(t) + \sum_{n=1}^{N} u(t - n\delta)$, δ fixed.

2.5.2 Time invariant systems

Time invariant systems represents a special class of dynamical systems. Let a system with map F,

$$u(t) \xrightarrow{F} y(t).$$

We say that the system F is time-invariant if and only if shifting the input by a given time lag τ merely shifts the output by the same lag τ, that is,

$$u(t+\tau) \xrightarrow{F} y(t+\tau).$$

Examples. We consider the following systems:

1. $y(t) = [u(t)]^{3/2}$.
   Step 1: replace $u(t)$ by $u(t+\tau) \to [u(t+\tau)]^{3/2}$.
   Step 2: replace $y(t)$ by $y(t+\tau) \to y(t+\tau) = [u(t+\tau)]^{3/2}$.
   Step 3: are 1,2 equal? If yes, as is the case here, then the system is time-invariant.
2. $y(t) = \int_0^t \sqrt{u(t_1)}dt_1$.
   Step 1: $\int_0^t \sqrt{u(t_1+\tau)}dt_1 = \int_\tau^{t+\tau} \sqrt{u(s)}ds$, with $t_1 + \tau = s$.
   Step 2: $y(t+\tau) = \int_0^{t+\tau} \sqrt{u(t_1)}dt_1$.
   Step 3: clearly, 1 and 2 are not equal hence this system is not time invariant.
3. $y(t) = \int_{t-s}^{t} u^3(t_1)dt_1$.
   Step 1: $\int_{t-s}^{t} u^3(t_1+\tau)dt_1 = \int_{t-s+\tau}^{t+\tau} u^3(\xi)d\xi$, with $t_1 + \tau = \xi$.
   Step 2: $y(t+\tau) = \int_{t+\tau-s}^{t+\tau} u^3(t_1)dt_1$.
   Step 3: this time, 1 and 2 are equal hence this system is time invariant.

2.5.3 Linear systems

A system with map L is linear if, denoting

$$u_1(t) \xrightarrow{L} y_1(t),$$
$$u_2(t) \xrightarrow{L} y_2(t),$$

then, for any fixed numbers $\alpha_1, \alpha_2 \in \mathbb{R}$, we have

$$\alpha_1 u_1(t) + \alpha_2 u_2(t) \xrightarrow{L} \alpha_1 y_1(t) + \alpha_2 y_2(t).$$

Examples. We consider the following systems:

1. $y(t) = k\dfrac{du}{dt}$ is linear and time invariant (LTI).
2. $y(t) = \displaystyle\int_0^t u(t_1)dt_1$ is linear but not time invariant.
3. $y(t) = au^2(t)$ is nonlinear, time invariant.
4. $y(t) = a(t)u(t)$, where $a(t)$ is a given function, is linear but not time invariant.
5. The canonical damped harmonic oscillator, pictured in Figure 2.6(a), is LTI.
6. Small amplitude water waves excited by a wave maker in a water tank, as shown in Figure 2.6(b), are an LTI system.

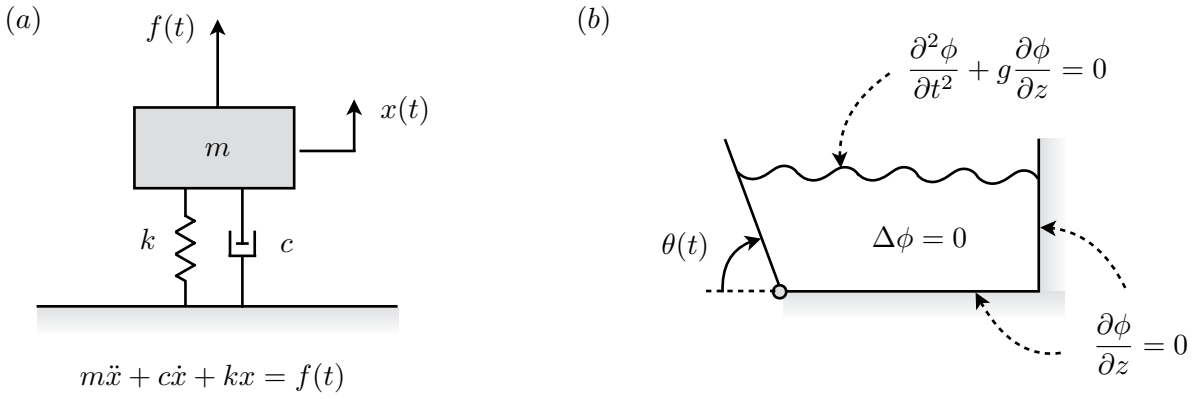


Figure 2.6 – LTI systems: (a) damped harmonic oscillator, (b) small amplitude water waves.

2.5.4 Convolution

Convolution is a basic property of LTI systems, which allows to relate the output to the input. Consider the system pictured in Figure 2.7 with input $u(t)$, output $y(t)$ and 'impulse response function' (IRF) $h(t)$, which characterizes the response of the system to an impulse input, and is unique for each system. Then, the convolution property allows to express the response of the system to an arbitrary input as

$$y(t) = \int_{-\infty}^{\infty} u(\tau)h(t-\tau)d\tau.$$

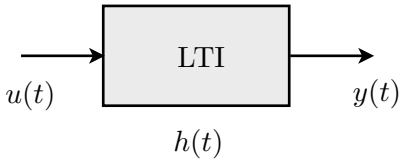


Figure 2.7 – LTI system.

To prove this formula, we will decompose the input into a series of impulses, as shown in Figure 2.8. We will then use the LTI property to express the output as a superposition of individual responses to each of these impulses, shifted in time (hence the τ variable) according to when the impulse occurred.

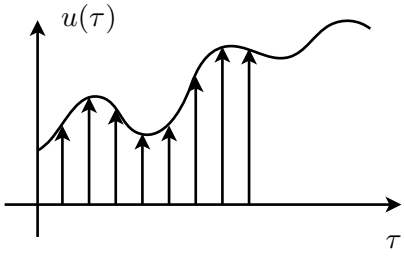


Figure 2.8 – Decomposition of the input into series of impulses.

First, we need to define what is an impulse function, or delta function. Consider for small ϵ the function

$$\delta_\epsilon(t) = \begin{cases} 1/\epsilon, & -\epsilon/2 \le t \le \epsilon/2, \\ 0, & \text{otherwise}, \end{cases}$$

which has unit area, and is represented in Figure 2.9. The delta function $\delta(t)$ is defined as

$$\delta(t) = \lim_{\epsilon \to 0} \delta_\epsilon(t).$$

Note that although the delta function has negligible support, it still integrates to one. Some further important properties of the delta function are as follows:

1. $\displaystyle\int_{-\infty}^{\infty} f(t)\delta(t)dt = \lim_{\epsilon\to 0}\int_{-\epsilon/2}^{\epsilon/2} f(t)\frac{1}{\epsilon}dt = f(0)\lim_{\epsilon\to 0}\int_{-\epsilon/2}^{\epsilon/2}\frac{1}{\epsilon}dt = f(0).$

2. $\displaystyle\int_{-\infty}^{\infty} f(t)\delta(t-\xi)dt = \lim_{\epsilon\to 0}\int_{\xi-\epsilon/2}^{\xi+\epsilon/2} f(t)\frac{1}{\epsilon}dt = f(\xi)\lim_{\epsilon\to 0}\int_{\xi-\epsilon/2}^{\xi+\epsilon/2}\frac{1}{\epsilon}dt = f(\xi).$

We are now ready to prove the convolution formula.

1. First, consider an LTI system and let $u(t) = \delta(t)$. By definition, and as pictured in Figure 2.10, the output is the impulse response function (IRF)

$$y(t) = L[\delta(t)] = h(t).$$

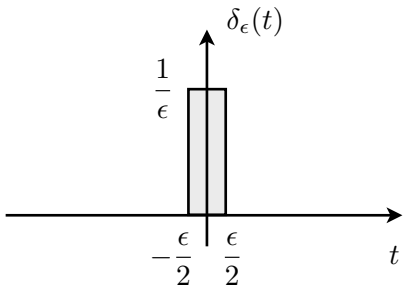


Figure 2.9 – The delta function $\delta(t)$ is the limit of $\delta_\epsilon(t)$ as $\epsilon \to 0$.

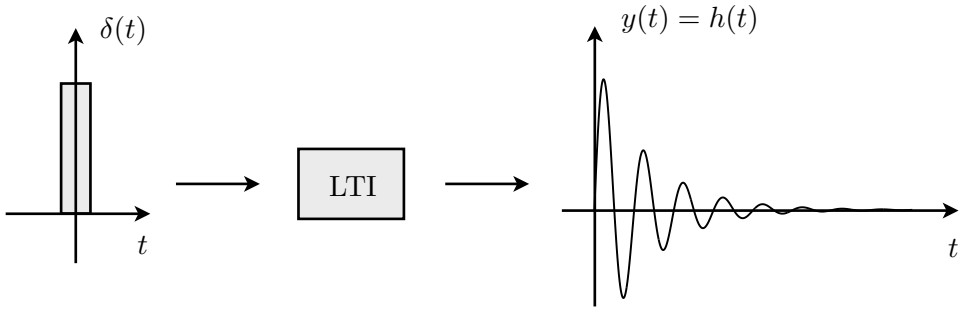


Figure 2.10 – Impulse response function.

2. Then, consider an arbitrary input $u(t)$ and express it as a superposition of impulses

$$u(t) = \int_{-\infty}^{\infty} u(\xi)\delta(t-\xi)d\xi.$$

3. Finally, utilize linearity and time-invariance to express the output as

$$\begin{aligned} y(t) = L[u(t)] &= L\left[\int_{-\infty}^{\infty} u(\xi)\delta(t-\xi)d\xi\right] \\ &= \int_{-\infty}^{\infty} u(\xi)L[\delta(t-\xi)]d\xi, \quad \text{using linearity} \\ &= \int_{-\infty}^{\infty} u(\xi)h(t-\xi)d\xi, \quad \text{using time-invariance} \\ &= \int_{-\infty}^{\infty} u(t-s)h(s)ds \\ &= u(t) * h(t), \end{aligned}$$

where the last two integrals are by definition equal to the convolution of the input $u(t)$ with the IRF $h(t)$.

The convolution has the following properties. If $y(t) = h(t) * x(t)$, then

1. $\dfrac{dy}{dt} = \dfrac{dh}{dt} * x(t) = h(t) * \dfrac{dx}{dt}$,
2. $\displaystyle\int_{-\infty}^{t} y(s)ds = \int_{-\infty}^{t} h(s)ds * x(t) = h(t) * \int_{-\infty}^{t} x(s)ds.$

Impulse Response Function

A causal system is one that responds only after being excited (note that all physical systems are causal). In this case, the impulse response function satisfies $h(t) = 0$ for $t < 0$. Suppose now that the input is such that $u(t) = 0$ for $t < 0$. Then, the convolution formula for the response becomes

$$y(t) = \int_{-\infty}^{\infty} u(\tau)h(t-\tau)d\tau = \int_{0}^{t} u(\tau)h(t-\tau)d\tau.$$

How to find the impulse response function for $t > 0$? Let us consider the example of a forced damped oscillator. By definition, the IRF is the response of the system to a delta function input and zero initial conditions, hence we need to solve

$$m\ddot{x} + b\dot{x} + kx = \delta(t), \quad x(0) = \dot{x}(0) = 0.$$

We take the integral of this equation from $-\epsilon/2$ to $\epsilon/2$ and let $\epsilon \to 0$

$$\lim_{\epsilon\to 0}\int_{-\epsilon/2}^{\epsilon/2}(m\ddot{x} + b\dot{x} + kx)dt = \lim_{\epsilon\to 0}\int_{-\epsilon/2}^{\epsilon/2}\delta(t)dt.$$

In general, only the highest order derivative on the left-hand side, here $\ddot{x}$, is expected to have an impulse behavior and thus to give a nonzero contribution to the integral as $\epsilon \to 0$. The other terms on the left-hand side, $\dot{x}$ and x, remain finite hence their integral vanishes as $\epsilon \to 0$. Note also that by definition of the delta function, the integral on the right-hand side is simply equal to one, hence

$$m\dot{x}(0^+) - m\dot{x}(0^-) = 1.$$

Therefore, the impulse imparts new initial conditions to the system at $t = 0^+$

$$x(0^+) = 0, \quad \dot{x}(0^+) = \frac{1}{m},$$

and the problem for $t > 0$ becomes

$$m\ddot{x} + b\dot{x} + kx = 0, \quad x(0^+) = 0, \quad \dot{x}(0^+) = \frac{1}{m}.$$

To solve this equation, we consider the ansatz

$$x(t) = c_1 e^{s_1 t} + c_2 e^{s_2 t},$$

which gives the quadratic equation

$$ms^2 + bs + k = 0,$$

and hence the roots are

$$s_{1,2} = -\frac{b}{2m} \pm \frac{\sqrt{b^2 - 4km}}{2m} = -c \pm i\omega_d,$$

assuming that $b^2 \ll 4k$. The solution finally gives the impulse response function, $h(t) = x(t)$.

2.5.5 Fourier series and Fourier transforms

Consider a periodic function $f(t)$ with period T, that is, $f(t+T) = f(t)$, as pictured in Figure 2.11. Then, this function admits the following Fourier series expansion

$$f(t) = A_0 + \sum_{n=1}^{\infty} \{A_n \cos(n\omega_0 t) + B_n \sin(n\omega_0 t)\},$$

where $\omega_0 = 2\pi/T$ is the fundamental frequency, and the Fourier coefficients A_0, A_n, B_n are defined as

$$\begin{aligned} A_0 &= \frac{1}{T}\int_0^T f(t)dt, \\ A_n &= \frac{2}{T}\int_0^T f(t)\cos(n\omega_0 t)dt, \\ B_n &= \frac{2}{T}\int_0^T f(t)\sin(n\omega_0 t)dt. \end{aligned}$$

A more compact equivalent representation is

$$f(t) = \sum_{n=-\infty}^{\infty} C_n e^{in\omega_0 t}$$

where the Fourier coefficients C_n are

$$C_n = \frac{1}{T}\int_0^T f(t)e^{-in\omega_0 t}dt.$$

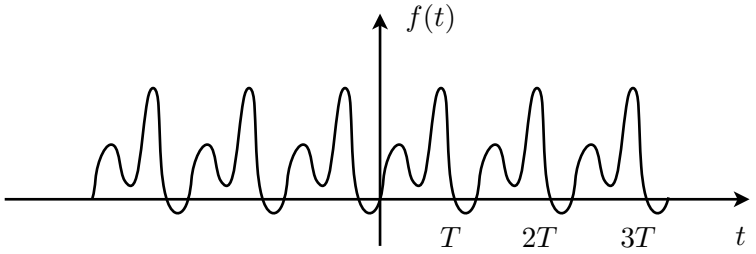


Figure 2.11 – Periodic function with period T.

The Fourier transform is an extension of the Fourier series to aperiodic functions that are absolutely integrable, that is,

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

as picture in Figure 2.12. For such functions, the direct and inverse Fourier transforms are defined, respectively, as

$$\begin{aligned} \tilde{f}(\omega) &= \mathcal{F}[f(t)] = \int_{-\infty}^{\infty} f(t)e^{-i\omega t}dt, \\ f(t) &= \mathcal{F}^{-1}[\tilde{f}(\omega)] = \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{f}(\omega)e^{i\omega t}d\omega. \end{aligned}$$

Some useful properties of the Fourier transform are as follows:

1. $\mathcal{F}\left[\dfrac{d^n f}{dt^n}\right] = (i\omega)^n \tilde{f}$.

2. $\mathcal{F}\left[\displaystyle\int_{-\infty}^{t} f dt\right] = \frac{1}{i\omega}\tilde{f}(\omega) + \pi\tilde{f}(0)\delta(\omega)$.

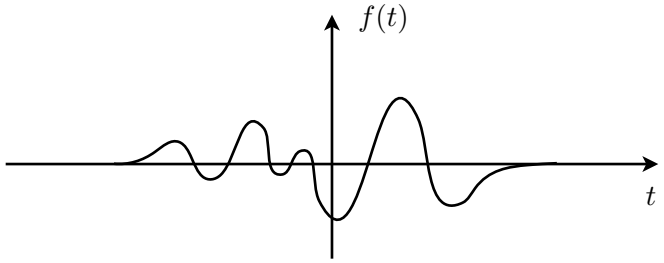


Figure 2.12 – Aperiodic function with compact support.

The Fourier transform is a very useful tool for solving linear ODEs with constant coefficients (as long as the response vanishes as time goes to infinity). Consider, for example, the forced damped oscillator

$$m\ddot{x} + b\dot{x} + kx = f(t).$$

Taking the Fourier transform on both sides of the equation leads to the algebraic equation

$$m(i\omega)^2\tilde{x} + b(i\omega)\tilde{x} + k\tilde{x} = \tilde{f},$$

hence the response can be directly expressed as

$$\tilde{x}(\omega) = \frac{\tilde{f}(\omega)}{-m\omega^2 + k + ib\omega}.$$

2.5.6 Transfer function

For LTI systems, recall that we have

$$y(t) = h(t) * u(t).$$

Taking the Fourier transform of this expression, the convolution is replaced by a standard product

$$\begin{aligned}
\tilde{y}(\omega) = \mathcal{F}[y(t)] &= \mathcal{F}\left[\int_{-\infty}^{\infty} u(\tau)h(t-\tau)d\tau\right] \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u(\tau)h(t-\tau)e^{-i\omega t}d\tau dt \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u(\tau)h(t_1)e^{-i\omega(t_1+\tau)}dt_1 d\tau, \quad t_1 = t-\tau \\
&= \int_{-\infty}^{\infty} h(t_1)e^{-i\omega t_1}dt_1 \int_{-\infty}^{\infty} u(\tau)e^{-i\omega\tau}d\tau \\
&= \tilde{h}(\omega)\tilde{u}(\omega).
\end{aligned}$$

The function $\tilde{h}(\omega)$ is the Fourier transform of $h(t)$, and is called the transfer function $H(\omega)$. Thus, the output of an LTI system can equivalently be expressed in terms of the convolution of $h(t)$ and $u(t)$, or in terms of the product of $H(\omega)$ and $\tilde{u}(\omega)$

$$y(t) = h(t) * u(t),$$
$$\tilde{y}(\omega) = \tilde{h}(\omega)\tilde{u}(\omega).$$

Example. Consider the forced damped oscillator defined by

$$m\ddot{x} + b\dot{x} + kx = f(t).$$

Since the response to a delta function input is the impulse response, one can set $f(t) = \delta(t)$

$$m\ddot{x} + b\dot{x} + kx = \delta(t),$$

so that $x(t) = h(t)$ is the impulse response function. Then, the Fourier transform of $x(t)$ is, by definition, the transfer function $H(\omega)$

$$(i\omega)^2 mH(\omega) + i\omega bH(\omega) + kH(\omega) = 1 \quad \Rightarrow \quad H(\omega) = \frac{1}{k - \omega^2 m + ib\omega}.$$

Thus, the Fourier transform gives a convenient way of calculating $H(\omega)$ without explicitly calculating the IRF $h(t)$.

2.6 Spectrum of a stochastic process

2.6.1 Definition

Let $Y(t, \zeta)$ a stationary stochastic process with $E[Y(t, \zeta)] = 0$ and autocorrelation function $R_{YY}(\tau)$. We define the power spectral density (in short, the spectrum) as the Fourier transform of the autocorrelation function, that is,

$$S_{YY}(\omega) = \int_{-\infty}^{\infty} R_{YY}(\tau) e^{-i\omega\tau} d\tau,$$
$$R_{YY}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{YY}(\omega) e^{i\omega\tau} d\omega.$$

Note that for the case of a non-zero mean stochastic process the spectrum is defined using the covariance function, $C_{YY}(\tau)$. The spectrum has the following properties:

1. $S_{YY}(\omega)$ is real and even since $R_{YY}(\tau)$ is real and even

$$\begin{aligned} S_{YY}(\omega) &= \int_{-\infty}^{\infty} R_{YY}(\tau) e^{-i\omega\tau} d\tau \\ &= \int_{-\infty}^{\infty} R_{YY}(\tau)(\cos\omega\tau - i\sin\omega\tau) d\tau \\ &= \int_{-\infty}^{\infty} R_{YY}(\tau)\cos\omega\tau d\tau. \end{aligned}$$

2. The area under S_{YY} is related to the energy (variance) of the stochastic process. Indeed,

$$\sigma_Y^2 = R_{YY}(\tau)|_{\tau=0} = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{YY}(\omega)\, e^{i\omega\tau}\big|_{\tau=0} d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{YY}(\omega) d\omega.$$

3. $S_{YY}(\omega) \geq 0$, to be proven later in Section 2.7.3.

2.6.2 Relationship with the Fourier transform

Why do we bother defining the spectrum instead of simply working with the usual Fourier transform? To answer this question, consider a stationary stochastic process $Y(t,\zeta)$. By definition, its variance is independent of time,

$$\sigma_Y^2 = R_{YY}(0),$$

which means that $Y(t,\zeta)$ cannot vanish as time goes to infinity. As a result, $Y(t,\zeta)$ is not integrable (does not have compact support) and its Fourier transform does not exist. On the other hand, it is usually the case that $R_{YY}(\tau)$ goes to zero as $|\tau|$ goes to infinity. The Fourier transform of R_{YY} is thus well-behaved, and is precisely the definition of the spectrum

$$S_{YY}(\omega) = \int_{-\infty}^{\infty} R_{YY}(\tau) e^{-i\omega\tau} d\tau.$$

To understand how the spectrum relates to the frequency content of $Y(t,\zeta)$, assume that $Y(t,\zeta)$ is Gaussian and consider the decomposition

$$Y(t,\zeta) = \sum_{i=1}^{N} a_i \cos(\omega_i t + \theta_i(\zeta)),$$

where a_i and ω_i are deterministic amplitudes and frequencies, and $\theta_i \sim \mathcal{U}(0, 2\pi)$ are i.i.d. random phases. We have previously calculated the autocorrelation function of this process,

$$R_{YY}(\tau) = \sum_{i=1}^{N} \frac{1}{2} a_i^2 \cos\omega_i\tau.$$

The spectrum is then given by the Fourier transform of R_{YY},

$$S_{YY}(\omega) = \mathcal{F}[R_{YY}(\tau)] = \sum_{i=1}^{N} \frac{1}{2} a_i^2 \mathcal{F}[\cos\omega_i\tau] = \sum_{i=1}^{N} \frac{1}{2} a_i^2 \pi[\delta(\omega-\omega_n) + \delta(\omega+\omega_n)].$$

As illustrated in Figure 2.13, the spectrum consists of delta functions of amplitude $\pi a_i^2/2$ at the discrete frequencies ω_i. Note that increasing the number of frequencies, one can recover the continuous spectrum in the limit $N \to \infty$ and $\omega_{i+1} - \omega_i \to 0$. Thus, the spectrum reflects the distribution of energy in $Y(t,\zeta)$ over different frequencies, much as we would have expected from a Fourier transform of Y (were it well-defined). There are however important differences:

1. The spectrum represents the energy (and not amplitude) content of Y, hence the amplitudes $a_1, a_2, \ldots$ appear squared in the spectrum.

2. The spectrum does not carry any information on the phases $\theta_1, \theta_2, \ldots$ of Y, and as such it does not say anything about whether the phases are correlated or not. Hence, the spectrum by itself is not sufficient to describe the shape of the pdf of Y at a given time.

Remark. In real-world applications, one cannot measure a stochastic process for an infinite amount of time, making the computation of the autocorrelation function a tricky task (especially for large values of the time lag τ). On the other hand, one can show that the autocorrelation function can also be obtained by

$$S_{YY}(\omega) = \lim_{T\to\infty} \frac{1}{2T} E\left[|\tilde{Y}_T(\omega)|^2\right],$$

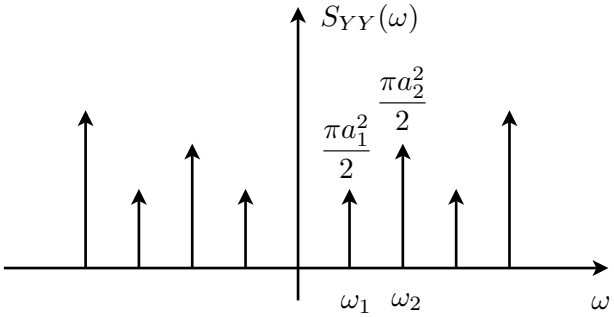


Figure 2.13 – Spectrum of a stochastic process with discrete frequencies.

where $\tilde{Y}_T(\omega) = \mathcal{F}[Y_T(t)]$, and Y_T is a truncation of Y in the interval $t \in [-T, T]$. Rather than going through the autocorrelation function, it is thus customary to approximate the spectrum by computing

$$S_{YY}(\omega) \simeq \frac{1}{2T} E\left[|\tilde{Y}_T(\omega)|^2\right].$$

2.6.3 Simulating realizations of a given spectrum

Suppose we know that a stochastic process $Y(t, \zeta)$ has spectrum $S_{YY}(\omega)$ and is Gaussian. How to compute realizations of the process $Y(t, \zeta)$? First, we need to chose a fine subdivision $0 \leq \omega_1 < \omega_2 < \cdots < \omega_{N+1}$ of the frequency axis, as shown in Figure 2.14. Then, in each frequency interval $[\omega_k, \omega_{k+1}]$, we choose a unique $\hat{\omega}_k$ by sampling uniformly a value in the interval $[\omega_k, \omega_{k+1}]$. Then, the stochastic process $Y(t, \zeta)$ can be approximated and simulated with its discrete counterpart

$$Y_d(t, \zeta) = \sum_{k=1}^{N} a_k \cos(\hat{\omega}_k t + \theta_k(\zeta)),$$

where the deterministic amplitudes a_k remain to be found, the deterministic frequencies $\hat{\omega}_k$ are known and the phases $\theta_k(\zeta)$ are i.i.d. random variables. The discrete spectrum corresponding to $Y_d(t, \zeta)$ has been calculated previously

$$S_{Y_d Y_d}(\omega) = \sum_{k=1}^{N} \frac{\pi}{2} a_k^2 \left[\delta(\omega - \hat{\omega}_k) + \delta(\omega + \hat{\omega}_k)\right].$$

We now equate the distribution of energy over different frequencies for S_{YY} and $S_{Y_d Y_d}$

$$\frac{1}{\pi} \int_{\omega_k}^{\omega_{k+1}} S_{YY}(\omega) d\omega = \frac{1}{\pi} \int_{\omega_k}^{\omega_{k+1}} S_{Y_d Y_d}(\omega) d\omega, \quad k = 1, \ldots, N.$$

Approximating the integral of S_{YY} with a midpoint rule, and recalling that delta functions in $S_{Y_d Y_d}$ integrate to one, we finally obtain a relation relating the discrete amplitudes a_k with the initial continuous spectrum S_{YY}

$$a_k = \sqrt{\frac{2}{\pi} S_{YY}(\omega) \Delta\omega}, \quad k = 1, \ldots, N.$$

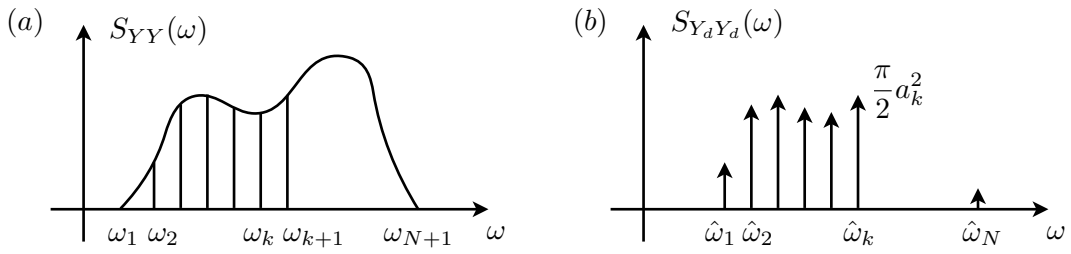


Figure 2.14 – Discretization of a continuous spectrum (only the positive frequencies are shown).

2.6.4 Wave spectrum

The wave spectrum is the spectrum of a Gaussian, stationary stochastic process representing the wave elevation

$$\eta(t,\zeta) = \sum_{i=1}^{N} A_n \cos(\omega_n t + \theta_n(\zeta)),$$

where A_n and ω_n are, respectively, deterministic amplitudes and frequencies, and $\theta_n(\zeta)$ are i.i.d. random variables in $[0, 2\pi]$. Thus, a wave field consists of many harmonics with random phases. From the central limit theorem, we know that the surface elevation η in the above representation is Gaussian. This representation usually follows from the discretization of a continuous spectrum, as described in the previous section. For water waves, a commonly used continuous spectrum is the Pierson-Moskowitz spectrum,

$$S(\omega) = \frac{A}{\omega^5} e^{-B/\omega^4},$$

where A and N are the following constants

$$A = ag^2 = 0.0081g^2,$$
$$B = 1.25\omega_p^4,$$

with ω_p a function of the wind speed at a height of 19.5 m. Other spectral forms may be used, for instance the JONSWAP spectrum which has more free parameters, hence a better description. Since the spectra are obtained through direct measurements, it is important to assume ergodicity.

2.6.5 Wind shear and turbulence

Another use of stochastic processes is in modeling wind shear and turbulence. Consider the situation picture in Figure 2.15, where the longitudinal component $\tilde{U}(z,t,\zeta)$ of the turbulent velocity field is decomposed as

$$\tilde{U}(z,t,\zeta) = \bar{U}(z,\tau) + u(t,\zeta),$$

where $\bar{U}(t,\tau)$ is a short-term mean with averaging time τ and $u(t,\zeta)$ is a fluctuating component representing turbulent eddies. The mean profile is expressed as

$$\bar{U}(z,\tau) = \bar{U}(z_r,\tau_r)\left[1 + C\ln\frac{z}{z_r}\right]\left[1 - 0.4I_u(z)\ln\frac{\tau}{\tau_r}\right],$$

where $z_r = 10$ m, $\tau_r = 3600$ s, C is a constant and $I_u(z)$ is the turbulence intensity, which goes as follows

$$I_u(z) = 0.06\left[1 + 0.043\bar{U}(z_r,\tau_r)\right]\left(\frac{z}{z_r}\right)^{-0.22}.$$

The fluctuations $u(t,\zeta)$ are modelled as a stochastic process with the following Davenport spectrum

$$S_{uu}(\omega) = \frac{\bar{U}_{10}^2 \delta}{\omega} \frac{4\theta^2}{(1+\theta^2)^{4/3}}, \quad \theta = \frac{\omega L_u}{2\pi \bar{U}_{10}},$$

where $\delta = 10^{-3}$ and $L_u = 1200$ m.

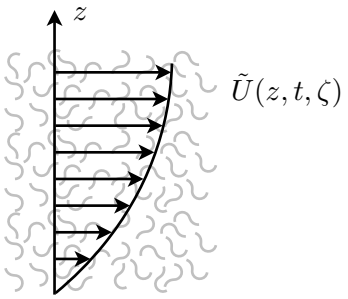


Figure 2.15 – Wind shear and turbulence.

2.7 Wiener-Khinchine relations

2.7.1 Theorem

Consider the LTI system shown in Figure 2.16 with stochastic input $u(t,\zeta)$, impulse response function $h(t)$, transfer function $H(\omega)$ and stochastic output $y(t,\zeta)$. If $u(t,\zeta)$ is a stationary and ergodic stochastic process, then the output $y(t,\zeta)$ is also a stationary and ergodic stochastic process. Moreover, if $u(t,\zeta)$ is Gaussian, then $y(t,\zeta)$ is also Gaussian.

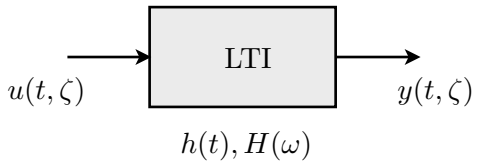


Figure 2.16 – LTI system with stochastic input and output.

We would now like to relate the statistics of $y(t,\zeta)$ to those of $u(t,\zeta)$. First, note that if $u(t,\zeta)$ is zero mean, then $y(t,\zeta)$ is also zero mean. Indeed,

$$E[y(t,\zeta)] = E\left[\int_{-\infty}^{\infty} u(\tau,\zeta)h(t-\tau)d\tau\right] = \int_{-\infty}^{\infty} E\left[u(\tau,\zeta)\right] h(t-\tau)d\tau = 0.$$

We now move to the second-order statistics of the response. The spectrum of the output is related to that of the input by the following Wiener-Khinchine relations:

$$S_{yy}(\omega) = |H(\omega)|^2 S_{uu}(\omega).$$

Proof. To show this, we start from the convolution expression of the response

$$y(t,\zeta) = \int_{-\infty}^{\infty} h(\tau)u(t-\tau,\zeta)d\tau.$$

The autocorrelation function of the output is then

$$\begin{aligned}
R_{yy}(t,\tau) &= E[y(t,\zeta)y(t+\tau,\zeta)] \\
&= E\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\tau_1)h(\tau_2)u(t-\tau_1,\zeta)u(t+\tau-\tau_2,\zeta)d\tau_2 d\tau_1\right] \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\tau_1)h(\tau_2)E[u(t-\tau_1,\zeta)u(t+\tau-\tau_2,\zeta)]d\tau_2 d\tau_1 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\tau_1)h(\tau_2)R_{uu}(\tau+\tau_1-\tau_2)d\tau_2 d\tau_1.
\end{aligned}$$

This shows that $R_{yy}(t,\tau) = R_{yy}(\tau)$, hence y is a stationary stochastic process. Its spectrum is

$$\begin{aligned}
S_{yy}(\omega) &= \int_{-\infty}^{\infty} R_{YY}(\tau)e^{-i\omega\tau}d\tau \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-i\omega\tau}h(\tau_1)h(\tau_2)R_{uu}(\tau+\tau_1-\tau_2)d\tau_2 d\tau_1 d\tau \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-i\omega(\xi+\tau_2-\tau_1)}h(\tau_1)h(\tau_2)R_{uu}(\xi)d\xi d\tau_1 d\tau_2, \quad \xi = \tau+\tau_1-\tau_2 \\
&= \int_{-\infty}^{\infty} e^{-i\omega\xi}R_{uu}(\xi)d\xi \int_{-\infty}^{\infty} e^{i\omega\tau_1}h(\tau_1)d\tau_1 \int_{-\infty}^{\infty} e^{-i\omega\tau_2}h(\tau_2)d\tau_2 \\
&= S_{uu}(\omega)H^*(\omega)H(\omega), \quad \text{since } h \text{ is real} \\
&= |H(\omega)|^2 S_{uu}(\omega).
\end{aligned}$$

2.7.2 Applications

Example. Consider a stochastic process $Y(t,\zeta)$ governed by the first-order system

$$\frac{dY}{dt} + \rho Y = X(t,\zeta),$$

where $X(t,\zeta)$ is a stationary stochastic process with zero mean and spectral density

$$S_{XX}(\omega) = \frac{a\sigma_X^2}{a^2+\omega^2}.$$

The transfer function of the system can be calculated by taking the Fourier transform of the governing equation, setting the input $X(t,\zeta) = \delta(t)$,

$$i\omega H(\omega) + \rho H(\omega) = 1 \quad \Rightarrow \quad H(\omega) = \frac{1}{i\omega+\rho}.$$

Then, from the Wiener-Khinchine relations,

$$S_{YY}(\omega) = |H(\omega)|^2 S_{XX}(\omega) = \frac{1}{|i\omega+\rho|^2}\frac{a\sigma_X^2}{a^2+\omega^2} = \frac{1}{\rho^2+\omega^2}\frac{a\sigma_X^2}{a^2+\omega^2}.$$

The variance of $Y(t,\zeta)$ is finally given by

$$\sigma_Y^2 = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{YY}(\omega)d\omega = \frac{1}{\pi}\int_0^{\infty} S_{YY}(\omega)d\omega = \frac{a\sigma_X^2}{\pi}\int_0^{\infty}\frac{d\omega}{(a^2+\omega^2)(\rho^2+\omega^2)} = \frac{\sigma_X^2}{\pi(a+\rho)\rho}.$$

Example. Consider a vibratory system,

$$\frac{dY^2}{dt^2} + b\frac{dY}{dt} + kY = X(t,\zeta),$$

where $X(t,\zeta)$ is a stationary stochastic process with zero mean and spectral density

$$S_{XX}(\omega) = \frac{a\sigma_X^2}{a^2+\omega^2}.$$

As in the previous example, the transfer function of the system can be calculated by taking the Fourier transform of the governing equation, setting the input $X(t,\zeta) = \delta(t)$,

$$-\omega^2 H(\omega) + bi\omega H(\omega) + kH(\omega) = 1 \quad \Rightarrow \quad H(\omega) = \frac{1}{k-\omega^2+ib\omega}.$$

Then, from the Wiener-Khinchine relations,

$$S_{YY}(\omega) = |H(\omega)|^2 S_{XX}(\omega) = \frac{1}{|k-\omega^2+ib\omega|^2}\frac{a\sigma_X^2}{a^2+\omega^2} = \frac{1}{(k-\omega^2)^2+b^2\omega^2}\frac{a\sigma_X^2}{a^2+\omega^2}.$$

The variance of $Y(t,\zeta)$ is finally given by

$$\sigma_Y^2 = \frac{1}{\pi}\int_0^\infty S_{YY}(\omega)d\omega = \frac{a\sigma_X^2}{\pi}\int_0^\infty \frac{d\omega}{(a^2+\omega^2)((k-\omega^2)^2+b^2\omega^2)} = \frac{(b+a)\sigma_X^2}{bk(a^2+ba+k)}.$$

2.7.3 Positivity of the spectrum

We can now give an indirect proof of the property, stated in Section 2.6.1, that the spectrum $S_{UU}(\omega)$ of a stochastic process $U(t,\zeta)$ is always non-negative. Assume that $S_{UU}(\omega) < 0$ for $\omega_1 < \omega < \omega_2$. Now, define a new stochastic process $Y(t,\zeta)$ as the output of an LTI system with input $U(t,\zeta)$ and transfer function

$$|H(\omega)| = \begin{cases} 1, & \omega_1 < \omega < \omega_2, \\ 0, & \text{otherwise.} \end{cases}$$

The spectrum of $Y(t,\zeta)$ can then be calculated with the Wiener-Khinchine relations

$$S_{YY}(\omega) = |H(\omega)|^2 S_{UU}(\omega) = \begin{cases} S_{UU}(\omega), & \omega_1 < \omega < \omega_2, \\ 0, & \text{otherwise.} \end{cases}$$

However, this implies that the variance of $Y(t,\zeta)$ is negative since

$$\sigma_Y^2 = \frac{1}{2\pi}\int_{-\infty}^\infty S_{YY}(\omega)d\omega = \frac{1}{2\pi}\int_{\omega_1}^{\omega_2} S_{UU}(\omega)d\omega < 0,$$

which is impossible. Thus, we must have $S_{UU}(\omega) \geq 0$ for all ω.

2.8 White noise

White noise is a special type of stationary and ergodic stochastic process X which has a flat spectrum, that is,

$$S_{XX}(\omega) = S_0 \quad \text{for all } \omega.$$

As a result, the autocorrelation function of white noise is a delta function,

$$R_{XX}(\tau) = S_0\delta(\tau).$$

S_0 is called the intensity of white noise. Note that $\sigma_X^2 = R_{XX}(0) = \infty$.

How do we construct white noise? Consider another process, the so-called Wiener process. A stochastic process $W(t,\zeta)$ is called a Wiener process or Brownian motion, if:

1. $P(W(0,\zeta) = 0) = 1$.
2. For arbitrary $0 < t_0 < t_1 < \cdots < t_n$, the increments $W(t_1,\zeta) - W(t_0,\zeta), W(t_2,\zeta) - W(t_1,\zeta), \ldots, W(t_n,\zeta) - W(t_{n-1},\zeta)$ are independent.
3. For any t and $h > 0$, the increment $W(t+h) - W(t)$ follows a Gaussian distribution with

$$E[W(t+h,\zeta) - W(t,\zeta)] = 0,$$
$$E[(W(t+h,\zeta) - W(t,\zeta))^2] = h.$$

The correlation function of a Wiener process is

$$C_{WW}(t_1,t_2) = E[W(t_1,\zeta)W(t_2,\zeta)] = E[W(t_1)[W(t_2,\zeta) - W(t_1,\zeta)]] + E[W(t_1,\zeta)^2].$$

Assuming that $t_2 > t_1$, we have

$$E[W(t_1,\zeta)[W(t_2,\zeta) - W(t_1,\zeta)]] = 0,$$
$$E[W(t_1,\zeta)^2] = E[(W(t_1,\zeta) - W(0,\zeta))^2] = t_1,$$

hence

$$C_{WW}(t_1,t_2) = t_1, \quad t_2 > t_1.$$

In general,

$$C_{WW}(t_1,t_2) = \min(t_1,t_2).$$

Figure 2.17 provides intuition for the Wiener process, whose variance scales with t. Individual realizations are everywhere continuous but nowhere differentiable.

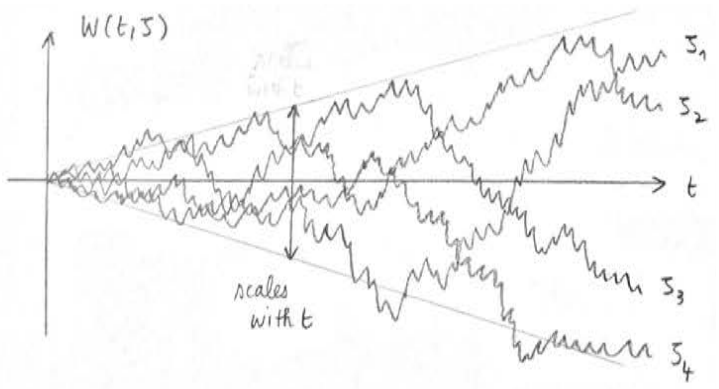


Figure 2.17 – Wiener process.

White noise is then defined as the generalized derivative of a Wiener process. The generalized derivative is defined through the correlation function as

$$\frac{\partial^2 C_{WW}(t_1,t_2)}{\partial t_1 \partial t_2} = \delta(t_1 - t_2),$$

hence white noise X is defined by the following correlation function

$$C_{XX}(\tau) = \delta(\tau).$$

The intensity of white noise is controlled by a multiplicative scalar S_0.

2.9 Direct integration

An alternative approach to working in the spectral domain with the Wiener-Khinchine relations involves direct integration in the time domain of the system governing equations. Consider, for instance, the following Langevin equation

$$\dot{Y} = -aY + \sqrt{2D}\dot{W}(t,\zeta), \quad Y(0) = Y_0,$$

where $W(t,\zeta)$ is a Wiener process, and its generalized derivative is white noise. The above equation has the following solution

$$Y(t,\zeta) = Y_0 e^{-at} + \sqrt{2D}\int_0^t e^{-a(t-s)}\frac{dW}{ds}ds.$$

In contrast to the previous spectral approach, working in the time domain allows for the analysis of initial conditions. The expected value of the solution is

$$E[Y] = Y_0 e^{-at},$$

since $E[\Delta W] = 0$. Moreover, its autocovariance function is

$$\begin{aligned}
C_{YY}(t_1,t_2) &= E\left[(Y(t_1,\zeta) - Y_0 e^{-\alpha t_1})(Y(t_2,\zeta) - Y_0 e^{-\alpha t_2})\right] \\
&= 2DE\left[\left(\int_0^{t_1}\int_0^{t_2} e^{a(s_1+s_2)}\dot{W}(s_1,\zeta)\dot{W}(s_2,\zeta)ds_1 ds_2\right)e^{-a(t_1+t_2)}\right] \\
&= 2De^{-a(t_1+t_2)}\int_0^{t_1}\int_0^{t_2} e^{a(s_1+s_2)}E\left[\dot{W}(s_1,\zeta)\dot{W}(s_2,\zeta)\right]ds_1 ds_2 \\
&= 2De^{-a(t_1+t_2)}\int_0^{t_1}\int_0^{t_2} e^{a(s_1+s_2)}\delta(s_1-s_2)ds_1 ds_2 \\
&= \sigma^2 e^{-a|\tau|}\left[1 - e^{-2a\min(t_1,t_2)}\right],
\end{aligned}$$

with $\sigma^2 = D/a$ and $\tau = t_1 - t_2$. Thus, for $\tau = 0$, we find that the variance is

$$\sigma_Y^2 = \sigma^2(1 - e^{-2at}).$$

As time goes to infinity, the expectation and covariance tend to the asymptotic limits

$$\lim_{t\to\infty} E[Y] = 0, \quad \text{and} \quad \lim_{t\to\infty} C_{YY}(t,\tau) = \sigma^2 e^{-a|\tau|}.$$

Because the expectation and covariance no longer depend on time, the system reaches a statistical steady state after a long time, characteristic of a stochastic attractor as illustrated in Figure 2.18. At the beginning of the initial transient phase, the initial condition concentrates all the probability at one point at $t = 0$. As time evolves, the system relaxes to its stable equilibrium while noise pushes the state away from this equilibrium. However, damping acts to keep the size of the attractor finite. The latter is given by the variance in the statistical steady state,

$$\sigma_Y^2 = \sigma^2 = \frac{D}{a},$$

with D indicating the noise intensity, and a the damping intensity. Note that for any stable system, the picture remains the same: stable dynamics keep the system state close to the stable equilibrium, while noise pushes the system away from it. The balance between these two competing effects results in a stochastic attractor describing the statistical steady state of the system.

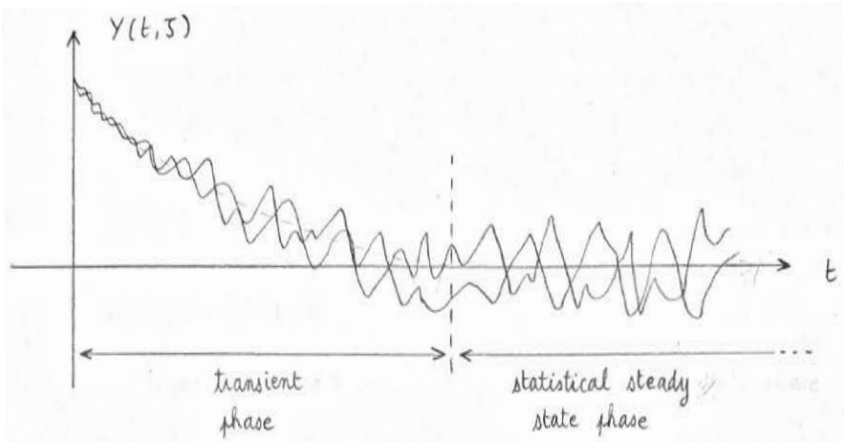


Figure 2.18 – Transient dynamics and stochastic attractor for Langevin equation.

Of course, the characteristics of the statistical steady state could have been obtained through the spectrum. The input has spectrum

$$S_{XX}(\omega) = 2D,$$

and the transfer function corresponding to the Langevin equation is

$$H(\omega) = \frac{1}{a + i\omega}.$$

Applying the Wiener-Khinchine relations, we obtain the spectrum of Y in the statistical steady-state

$$S_{YY}(\omega) = |H(\omega)|^2 S_{XX}(\omega) = \frac{2D}{a^2 + \omega^2},$$

from which we can deduce the variance in the statistical steady-state,

$$\sigma_Y^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{2D}{a^2 + \omega^2} d\omega = \frac{D}{a}.$$

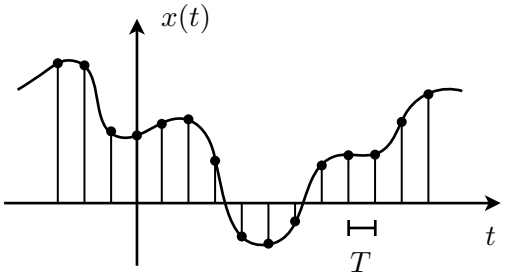


Figure 2.19 – Sampling of a continuous signal.

2.10 Nyquist sampling rate

Let $x(t)$ be a continuous signal which we want to sample every T time units, resulting in a discrete signal x_m as shown in Figure 2.19. What should T be so that we can fully recover $x(t)$ from x_m? To answer this question, we represent the sampling process as a multiplication of $x(t)$ with a train of impulses $p(t)$ defined as

$$p(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT).$$

The sampled signal $x_p(t)$ then takes the form of a train of impulses scaled by the values of $x(t)$ at the sampling locations,

$$x_p(t) = x(t)p(t) = \sum_{n=-\infty}^{\infty} x(t)\delta(t - nT) = \sum_{n=-\infty}^{\infty} x(nT)\delta(t - nT),$$

which we illustrate in Figure 2.20. What is the effect of sampling in the frequency domain? Because

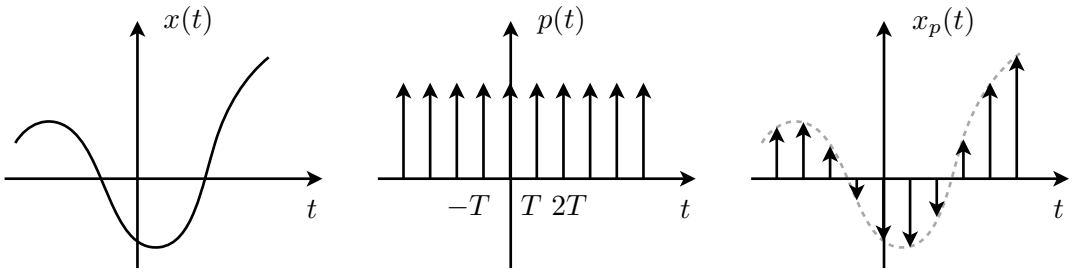


Figure 2.20 – Modeling of the sampling process.

$x_p(t) = x(t)p(t)$, the Fourier transform $\tilde{x}_p(\omega)$ of $x_p(t)$ is simply

$$\tilde{x}_p(\omega) = \tilde{x}(\omega) * \tilde{p}(\omega).$$

Let us first calculate $\tilde{p}(\omega)$, the Fourier transform of $p(t)$. Since $p(t)$ is periodic with period T, it can be expressed through a Fourier series as

$$p(t) = \sum_{k=-\infty}^{\infty} c_k e^{ik\omega_0 t}, \quad \omega_0 = \frac{2\pi}{T},$$

where the Fourier coefficients c_k are simply

$$c_k = \frac{1}{T}\int_{-T/2}^{T/2} p(t)e^{-ik\omega_0 t}dt = \frac{1}{T}.$$

Thus, we can equivalently express $p(t)$ as

$$p(t) = \frac{1}{T}\sum_{k=-\infty}^{\infty} e^{ik\omega_0 t},$$

allowing us to write its Fourier transform as a sum of delta function,

$$\tilde{p}(\omega) = \frac{2\pi}{T}\sum_{k=-\infty}^{\infty} \delta(\omega - k\omega_0),$$

since $\mathcal{F}^{-1}[\delta(\omega - k\omega_0)] = e^{ik\omega_0 t}/2\pi$. Finally, the Fourier transform of $x_p(t)$ becomes

$$\begin{aligned}
\tilde{x}_p(\omega) &= \tilde{x}(\omega) * \tilde{p}(\omega) \\
&= \int_{-\infty}^{\infty} \tilde{x}(\omega - \omega')\frac{2\pi}{T}\sum_{k=-\infty}^{\infty} \delta(\omega' - k\omega_0)d\omega' \\
&= \frac{2\pi}{T}\sum_{k=-\infty}^{\infty}\int_{-\infty}^{\infty} \tilde{x}(\omega - \omega')\delta(\omega' - k\omega_0)d\omega' \\
&= \frac{2\pi}{T}\sum_{k=-\infty}^{\infty} \tilde{x}(\omega - k\omega_0).
\end{aligned}$$

The above result shows that $\tilde{x}_p(\omega)$ is equal to a superposition of many times the function $\tilde{x}(\omega)$, each shifted by $k\omega_0$ in the frequency domain for a different integer value of k. Let us now suppose that $\tilde{x}(\omega)$ is band-limited, that is, its Fourier transform $\tilde{x}(\omega)$ is zero for all absolute frequencies above a cut-off frequency ω_M as illustrated in Figure 2.21(a). We now need to consider separately two cases:

1. When $\omega_0 > 2\omega_M$, there is no overlap of the shifted functions $\tilde{x}(\omega - k\omega_0)$ in $\tilde{x}_p(\omega)$, as shown in Figure 2.21(b). Thus, we can recover $\tilde{x}(\omega)$ from $\tilde{x}_p(\omega)$ if we filter the latter appropriately. For instance, consider the filter defined as an LTI system with $x_p(t)$ as input and the following transfer function:

   $$H(\omega) = \begin{cases} 1, & -\omega_c \le \omega \le \omega_c, \\ 0, & \text{otherwise.} \end{cases}$$

   As long as the condition $\omega_M < \omega_c < \omega_0 - \omega_M$ is satisfied, the output of this LTI system (filter) would be equal to the original continuous function $x(t)$. The action of this filter on the input in the spectral domain is sketched by the dotted box in Figure 2.21(b).

2. When $\omega_0 < 2\omega_M$, there is overlap of the shifted functions $\tilde{x}(\omega - k\omega_0)$ in $\tilde{x}_p(\omega)$, as shown in Figure 2.21(c). The resulting $\tilde{x}(\omega)$ then contains a distorted version of $\tilde{x}(\omega)$, a phenomenon that is called aliasing. In this case, it is thus not possible to recover the original function $x(t)$.

Therefore, in order to preserve all information, the continuous signal $x(t)$ must be sampled at

$$\omega_0 = \frac{2\pi}{T} > 2\omega_M,$$

where ω_M is the bandwidth of $\tilde{x}(\omega)$. This threshold sampling rate is commonly called the Nyquist rate of sampling.

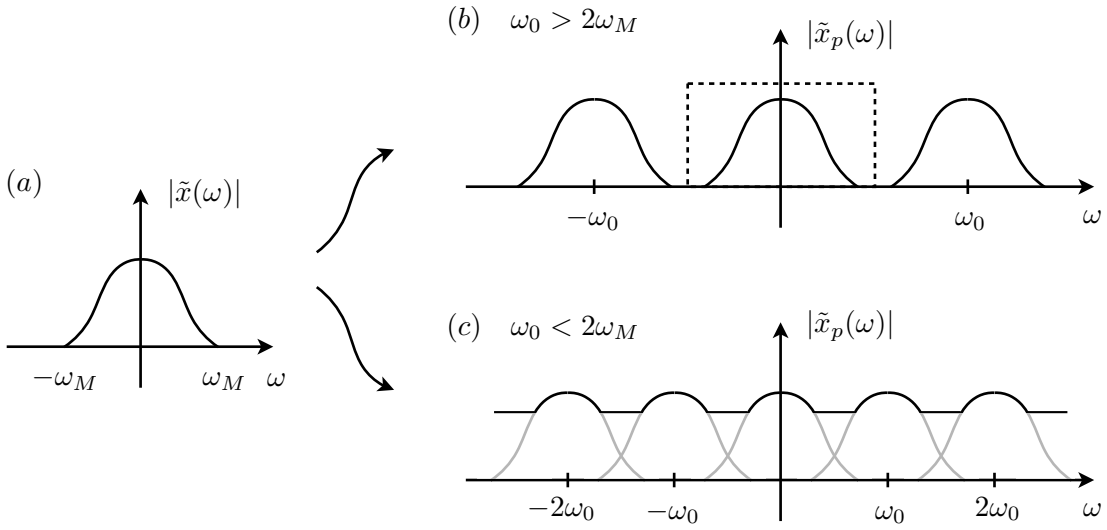


Figure 2.21 – Fourier transform of sampled signal $x_p(t)$ and Nyquist criterion.

Chapter 3

Extreme Event Statistics

3.1 Extreme value theory

Extreme value theory (EVT) is often used when e.g. building dams, where the knowledge of the statistics of extreme flooding events are required to design the specifications. To specify what EVT is about, let $\{X_1, X_2, \ldots, X_n\}$ be random variables and define

$$M_n = \max\{X_1, \ldots, X_n\}.$$

We now ask: what is the distribution of M_n? In general, it is unknown. However, in the specific case where $\{X_i\}$ are iid random variables, the answer is known and provided by EVT.

3.1.1 More on the central limit theorem

We first briefly review the central limit theorem (CLT), already covered in Section 1.10. Let $\{X_i\}$ be iid random variables with mean μ and variance σ^2, and define the following sum

$$S_n = X_1 + \cdots + X_n.$$

Then, by the central limit theorem (CLT), in the limit $n \to \infty$, the pdf of S_n will converge to that of a Gaussian random variable with mean $n\mu$ and variance $n\sigma^2$. This implies that the mean and variance of S_n grow unbounded as n increases, hence the limit $n \to \infty$ can be seen as degenerate. On the other hand, we can also define the rescaled sum,

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}},$$

which will converge to a Gaussian random variable with zero mean and unit variance as $n \to \infty$, that is,

$$P(Z_n \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz.$$

Thus, we have used a rescaling of the form $(S_n - a_n)/b_n$, where $a_n = n\mu$ and $b_n = \sigma\sqrt{n}$, to obtain a random variable that follows a normal Gaussian distribution. We will follow a similar strategy for finding a non-degenerate distribution for M_n.

3.1.2 Extremal types theorem

The below presentation follows [2]. As in the CLT, let $\{X_i\}$ be iid random variables with cdf $P(X_i \leq x) = F(x)$. Defining

$$M_n = \max\{X_1, \ldots, X_n\},$$

what is the distribution of M_n as $n \to \infty$? A preliminary calculation shows that

$$\begin{aligned} P(M_n \leq x) &= P(\max\{X_1, X_2, \ldots, X_n\} \leq x) \\ &= P(X_1 \leq x, X_2 \leq x, \ldots, X_n \leq x) \\ &= P(X_1 \leq x)P(X_2 \leq x)\ldots P(X_m \leq x) \\ &= F(x)F(x)\ldots F(x) \\ &= F(x)^n, \end{aligned}$$

where we have used independence of the $\{X_i\}$. Recall now that F is an increasing function of x, and $0 \leq F(x) \leq 1$. Thus, we define the right endpoint of the distribution as $x_{\max} = \max\{x : F(x) < 1\}$, such that

$$\begin{cases} F(x) < 1, & x < x_{\max}, \\ F(x) = 1, & x > x_{\max}. \end{cases}$$

Note that $x_{\max}$, so defined, may be infinity (for instance in the case of a Gaussian distribution). Going back to the distribution of M_n, we thus have

$$\lim_{n\to\infty} P(M_n \leq x) = \lim_{n\to\infty} F^n(x) = \begin{cases} 0, & x_{\max}, \\ 1, & x > x_{\max}. \end{cases}$$

Thus, the pdf of M_n tends to a delta function as $n \to \infty$ and the behavior of M_n can also be considered degenerate in this limit, similarly to the case of S_n in the CLT. This is pictured in Figure 3.1. On the other hand, the following theorem assures that we can define a suitable rescaling of the form $(M_n - a_n)/b_n$ which converges to a well-defined distribution in the limit $n \to \infty$.

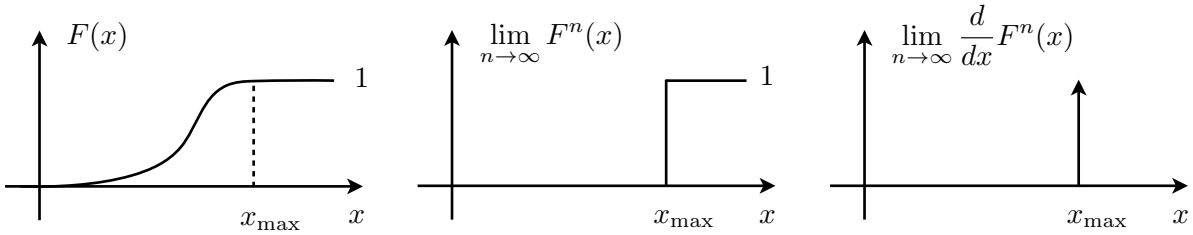


Figure 3.1 – Cdf of $\{X_i\}$ and M_n, and pdf of M_n.

Theorem (Extremal Types Theorem)**.** If $X_1, X_2, \ldots$ is a sequence of iid random variables and there exists a sequence $\{a_n\}$, $\{b_n\}$ of real numbers such that each $b_n > 0$ and

$$\lim_{n\to\infty} P\left(\frac{M_n - a_n}{b_n} \leq y\right) = G(y)$$

at all continuity points of $G(y)$, with $G(y)$ a non-degenerate function, then the limiting distribution $G(y)$ can only be one of three types:

(I) $G(y) = e^{-e^{-y}}$, (Gumbel),

(II) $G(y) = e^{-y^{-\alpha}}$, ($\alpha > 0$), for $y > 0$, (Fréchet),

(III) $G(y) = e^{-(-y)^{\alpha}}$, ($\alpha > 0$), for $y < 0$, (Weibull).

Note: The above definitions of the limiting distributions are the standard Gumbel, Frećhet, and Weibull distributions. We can define the distributions more generally by replacing y with $(y-m)\beta^{-1}$ for some location parameter m and scale parameter $\beta > 0$.

3.1.3 Domains of attraction

Given the distribution of $\{X_i\}$, and without computing M_n, do we know which of the three types does the limiting distribution of M_n belong to as $n \to \infty$? We do! The limiting distribution $G(y)$ only depends on the behavior of $F(x)$ as its tail, that is,

$$\lim_{x \to x_{\max}} F(x) = 1, \quad \text{or equivalently,} \quad \lim_{x \to x_{\max}} 1 - F(x) = 0.$$

The behavior at the tail means how fast $1 - F(x)$ converges to zero as $x \to x_{\max}$. There are three possible cases:

(I) If $x_{\max} < \infty$ and

$$\lim_{x \to x_{\max}} \frac{1 - F(x)}{(x - x_{\max})^m} = 0,$$

for all $m \in \mathbb{N}$, or $x_{\max} = \infty$ and

$$\lim_{x \to x_{\max}} \frac{1 - F(x)}{x^m} = 0,$$

for all $m \in \mathbb{N}$, then $G(y)$ is the Gumbell distribution. Note that this condition means that $1 - F(x)$ tends to zero at its tail faster than any polynomial.

(II) If $x_{\max} = \infty$ and

$$1 - F(x) \sim x^{-\alpha},$$

for some $\alpha > 0$, then $G(y)$ is the Fréchet distribution with parameter α.

(III) If $x_{\max} < \infty$ and

$$1 - F(x) \sim (x_{\max} - x)^{\alpha},$$

for some $\alpha > 0$, then $G(y)$ is the Weibull distribution with parameter α.

These three cases completely characterize the domains of attraction of the three distributions and can be illustrated in Figure 3.2. In Table 3.1, we list the type of the limiting distribution $G(y)$ that corresponds to various kinds of distributions $F(x)$ of $\{X_i\}$.

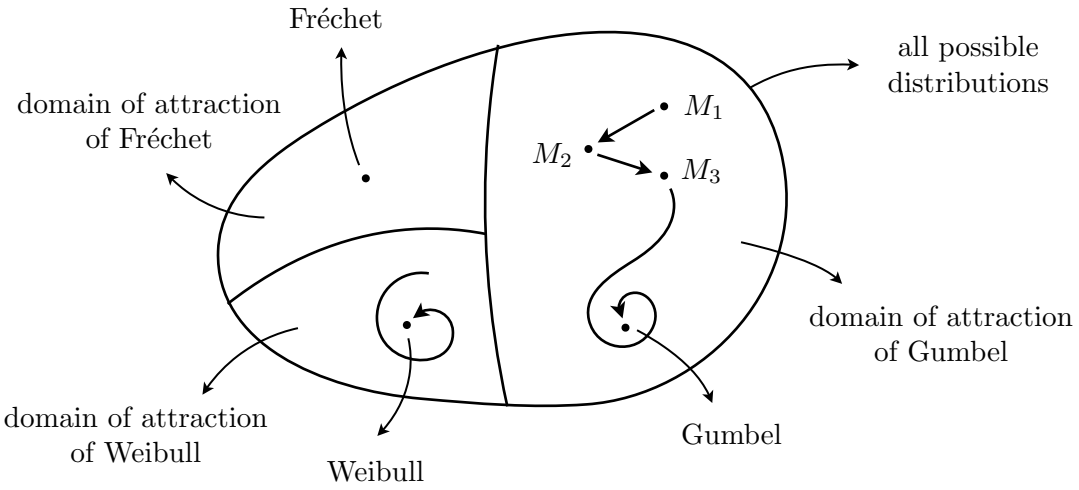


Figure 3.2 – Domains of attraction of the three possible limiting distributions.

$X_i \sim F(x)$	$M_n \sim G(y)$
Normal, Exponential, Lognormal, Gamma	Gumbel
Pareto, Cauchy, Student, Burr	Fréchet
Uniform, Beta	Weibull

Table 3.1 – Types of limiting distribution $G(y)$ corresponding to given distributions $F(x)$ of $\{X_i\}$.

3.1.4 A specific example

Consider a distribution in the second category. That is, $x_{\max} = \infty$ and

$$1 - F(x) \sim x^{-\alpha}.$$

We will find sequences a_n and b_n defined in terms of $F(x)$ such that

$$\lim_{n\to\infty} P\left(\frac{M_n - a_n}{b_n} \leq y\right) = e^{-y^{-\alpha}}.$$

We first define the quantile function as

$$Q(p) = \inf\{x : F(x) \geq p\}.$$

This returns the smallest x such that the cdf evaluated at x is greater than or equal to p. If F is smooth (there are no point masses in the pdf) and strictly increasing (there are no intervals where the pdf is zero) then this is simply the inverse of the cdf F,

$$Q(p) = F^{-1}(p), \text{ for } 0 \leq p \leq 1.$$

Now let $a_n = 0$ and $b_n = Q(1 - n^{-1})$. Using the definition of b_n and Q, we can find $F(b_n)$ as

$$\begin{aligned} F(b_n) &= F(Q(1 - n^{-1})) \\ &= F(F^{-1}(1 - n^{-1})) \\ &= 1 - n^{-1}. \end{aligned}$$

Alternatively we know from the assumed asymptotic behavior of F that

$$F(b_n) \sim 1 - b_n^{-\alpha},$$

and therefore that the asymptotic behavior of b_n is

$$b_n^{-\alpha} \sim n^{-1}.$$

Now, considering the limiting behavior of $(M_n - a_n)/b_n$ we have

$$\begin{aligned}
\lim_{n\to\infty} P\left(\frac{M_n - a_n}{b_n} \leq y\right) &= \lim_{n\to\infty} P\left(M_n \leq b_n y\right) \\
&= \lim_{n\to\infty} F^n\left(b_n y\right) \\
&\sim \lim_{n\to\infty} (1 - b_n^{-\alpha} y^{-\alpha})^n \\
&= \lim_{n\to\infty} \left(1 - \frac{y^{-\alpha}}{n}\right)^n \\
&= e^{-y^{-\alpha}}.
\end{aligned}$$

This sequence is not unique in admitting a non-degenerate function. Scaling b_n and using nonzero a_n can yield Fréchet distributions with different location and scale parameters. Take, for example,

$$\begin{aligned}
b_n &= n^{\alpha^{-1}} \beta^{-1} \\
a_n &= m b_n
\end{aligned}$$

for any m and $\beta > 0$. Then,

$$\begin{aligned}
\lim_{n\to\infty} P\left(\frac{M_n - a_n}{b_n} \leq y\right) &= \lim_{n\to\infty} P\left(M_n \leq b_n y - a_n\right) \\
&= \lim_{n\to\infty} P\left(M_n \leq n^{\alpha^{-1}} \beta^{-1}(y - m)\right) \\
&= \lim_{n\to\infty} F^n\left(n^{\alpha^{-1}} \beta^{-1}(y - m)\right) \\
&\sim \lim_{n\to\infty} \left(1 - \left(n^{\alpha^{-1}} \beta^{-1}(y - m)\right)^{-\alpha}\right)^n \\
&= \lim_{n\to\infty} \left(1 - \frac{1}{n}\left(\frac{y - m}{\beta}\right)^{-\alpha}\right)^n \\
&= e^{-\left(\frac{y-m}{\beta}\right)^{-\alpha}}
\end{aligned}$$

3.1.5 Generalized extreme value distribution

Finally, we note that the three types of limiting distributions can be unified through the following generalized extreme value (GEV) distribution,

$$H(y; \xi) = e^{-(1+\xi y)^{-1/\xi}}.$$

Indeed, the above GEV distribution reduces to one of the three types depending on the value of the parameter ξ:

(I) when $\xi \to 0$, $H(y; \xi) \to$ Gumbel,

(II) when $\xi > 0$, $H(y; \xi) \to$ Fréchet with parameter $\alpha = 1/\xi$,

(III) when $\xi < 0$, $H(y; \xi) \to$ Weibull with parameter $\alpha = -1/\xi$.

3.2 Statistics of extremes in stochastic processes

Previously, we have characterized the distribution of the maximum of an iid sequence of random variables. What about stochastic processes, which have memory (encoded in the autocorrelation function)? We ask various questions:

1. What is the frequency of upcrossings past a given level?
2. What is the frequency of local maxima larger than a given value?
3. What is the pdf describing the statistics of local maxima?

3.2.1 One-sided spectrum and bandwidth

First, we define the notions of one-sided spectrum and spectrum bandwidth. Given a (two-sided) spectrum $S(\omega)$, the one-sided spectrum is defined as

$$S^+(\omega) = \begin{cases} \frac{1}{\pi} S(\omega), & \omega \geq 0, \\ 0, & \omega < 0, \end{cases}$$

see Figure 3.3. The variance of the stochastic process corresponding to $S(\omega)$ can be expressed as

$$\sigma^2 = R(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega) d\omega = \frac{2}{2\pi} \int_0^{\infty} S(\omega) d\omega = \int_0^{\infty} S^+(\omega) d\omega.$$

Hence, the one-sided spectrum $S^+(\omega)$ is a convenient way to directly represent the distribution of energy contained in the stochastic process over different frequencies. Note that because the two-sided spectrum is even, we have not lost any information in defining $S^+(\omega)$, and $S(\omega)$ can be recovered in a straightforward manner from $S^+(\omega)$.

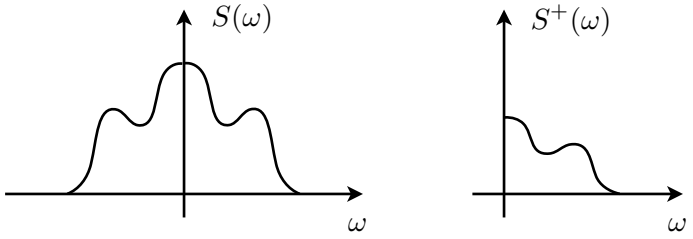


Figure 3.3 – Two-sided and one-sided spectra.

Next, we define the moments of the spectrum as

$$M_0 = \int_0^{\infty} S^+(\omega) d\omega,$$
$$M_2 = \int_0^{\infty} \omega^2 S^+(\omega) d\omega,$$
$$M_4 = \int_0^{\infty} \omega^4 S^+(\omega) d\omega,$$

which allow us to introduce a measure of the bandwidth ϵ of the spectrum,

$$\epsilon^2 = 1 - \frac{M_2^2}{M_0 M_4}.$$

The bandwidth ϵ admits the following two asymptotic limits:

1. For $\epsilon \to 0$, the spectrum is called narrow-banded, see Figure 3.4(a).
   Indeed, for $S^+(\omega) = \delta(\omega - \omega_0)$, we have $M_0 = 1$, $M_2 = \omega_0^2$, and $M_4 = \omega_0^4$, hence $\epsilon = 0$.
2. For $\epsilon \to 1$, the spectrum is called wide-banded, see Figure 3.4(b).

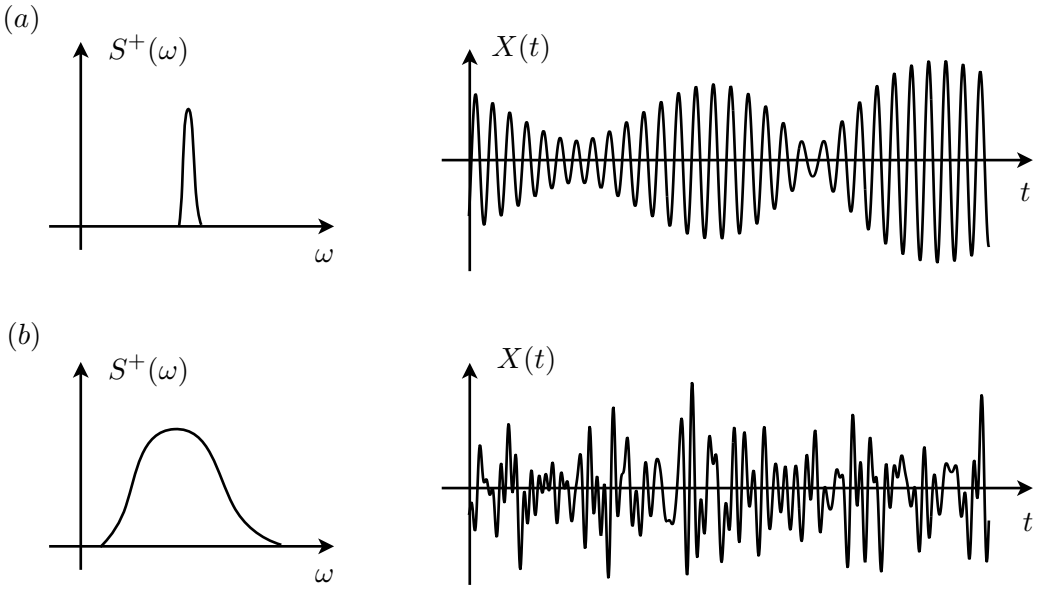


Figure 3.4 – Narrow-banded spectrum (a), wide-banded spectrum (b), and corresponding time realizations.

3.2.2 Frequency of upcrossings past a given level

Consider the stochastic process picture in Figure 3.5. We define $\bar{n}(A)$, the average frequency of upcrossings past level A, and $\bar{n}(0)$, the average frequency of upcrossings past zero level. The average period between two upcrossings of level zero is then $\bar{T} = 1/\bar{n}(0)$. As we will see below, for Gaussian, stationary and ergodic stochastic processes, we have

$$\bar{n}(A) = \frac{1}{2\pi}\sqrt{\frac{M_2}{M_0}}e^{-A^2/2M_0},$$

hence setting $A = 0$ we obtain

$$\bar{n}(0) = \frac{1}{2\pi}\sqrt{\frac{M_2}{M_0}} = \frac{1}{\bar{T}},$$

and therefore

$$\bar{n}(A) = \frac{1}{\bar{T}}e^{-A^2/2M_0}.$$

Example. Consider the platform picture in Figure. The water waves have average period $\bar{T} = 8$ sec, and standard deviation $\sigma = 2$ m. How to choose the height h of the deck above the (mean) water surface so that the deck is flooded every 10 min? We want

$$\bar{n}(h) = \frac{1}{\bar{T}}e^{-h^2/2\sigma^2} = \frac{1}{10\text{min}} \quad \Rightarrow \quad h = 5.07\text{m}.$$

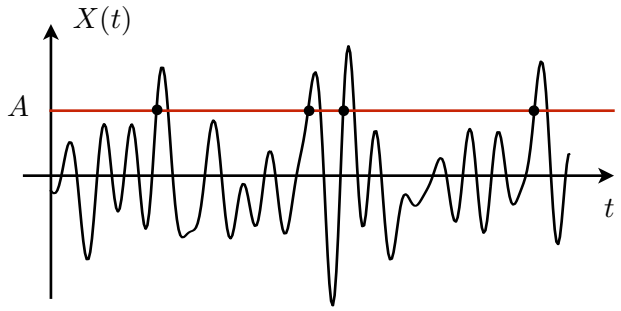


Figure 3.5 – Upcrossing rates past level A.

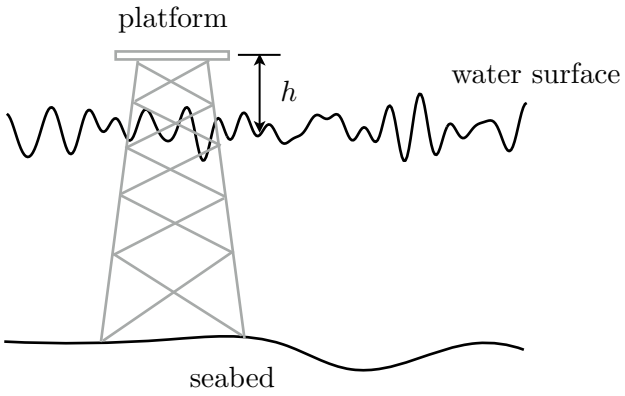


Figure 3.6 – Platform in rough sea conditions.

Let us now prove the formula for the average frequency of upcrossings. We build a counter function, that is, a function that counts the number of upcrossings, according to the process outlined in Figure 3.7. First, we build the function $f_1(t)$ that identifies the regions where $X(t) \geq a$,

$$f_1(t) = u(X(t) - a),$$

where u is the Heaviside function, defined as

$$u(x) = \begin{cases} 1, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Then, we take the derivative of $f_1(t)$ to obtain the function $f_2(t)$, which identifies both upcrossings and downcrossings past level a of $X(t)$,

$$f_2(t) = \frac{df_1}{dt}.$$

Finally, we only keep upcrossings by defining $f_3(t)$ as

$$f_3(t) = f_2(t)u(\dot{X}(t)).$$

Thus, we have obtained the counter function

$$f_3(t) = \delta(X(t) - a)\dot{X}(t)u(\dot{X}(t)).$$

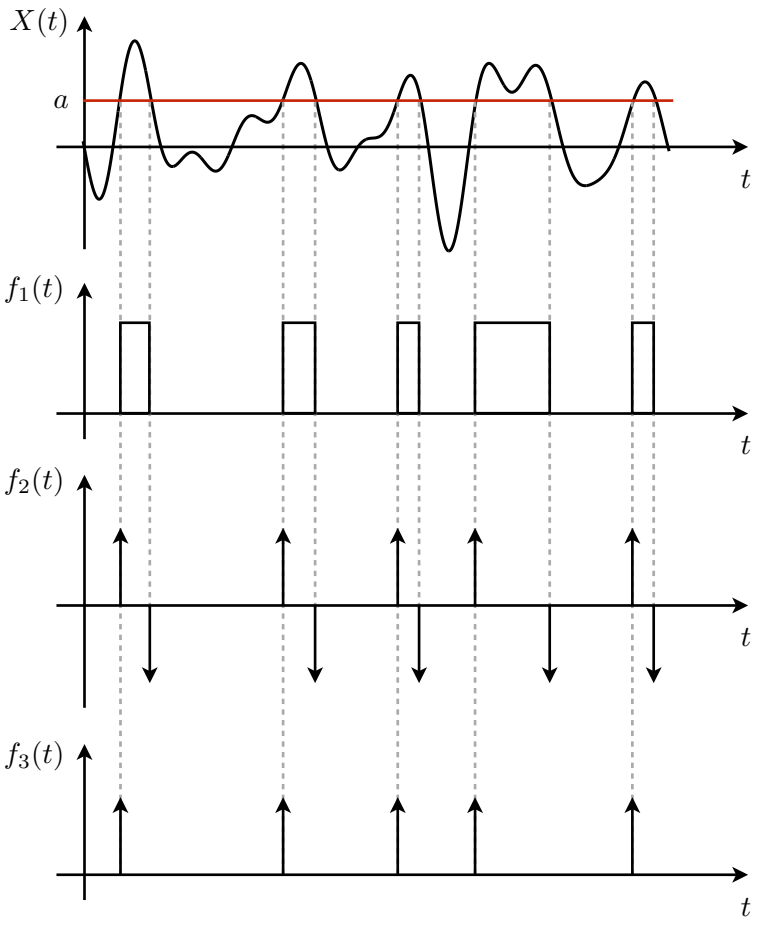


Figure 3.7 – Building a counter for the number of upcrossings above a.

The number of upcrossings past a that happen in a given time T is then simply

$$N(T) = \int_0^T \dot{X}(t)\delta(X(t) - a)u(\dot{X}(t))dt,$$

and the average frequency of upcrossings follows as

$$\bar{n}(a) = \lim_{T\to\infty} \frac{N(T)}{T} = \lim_{T\to\infty} \frac{1}{T} \int_0^T \dot{X}(t)\delta(X(t) - a)u(\dot{X}(t))dt.$$

Now, we use ergodicity of $X(t)$ to translate this time average to an ensemble average

$$\begin{aligned}
\bar{n}(a) &= E\left[\dot{X}\delta(X - a)u(\dot{X})\right] \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \dot{x}\delta(x - a)u(\dot{x})f_{X\dot{X}}(x, \dot{x})dxd\dot{x} \\
&= \int_0^{\infty} \dot{x}f_{X\dot{X}}(a, \dot{x})d\dot{x}.
\end{aligned}$$

The above is called the Rice formula. Assuming that X and $\dot{X}$ are uncorrelated Gaussian stochastic processes, the joint distribution $f_{X\dot{X}}$ is simply

$$f_{X\dot{X}}(x, \dot{x}) = f_X(x)f_{\dot{X}}(\dot{x}) = \frac{1}{\sqrt{2\pi}\sigma_X} e^{-x^2/2\sigma_X^2} \frac{1}{\sqrt{2\pi}\sigma_{\dot{X}}} e^{-\dot{x}^2/2\sigma_{\dot{X}}^2}.$$

Note that X and $\dot{X}$ are uncorrelated since

$$E[X\dot{X}] = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} X(t)\dot{X}(t)dt = \lim_{T\to\infty} \frac{1}{2T}\left.\frac{X(t)^2}{2}\right|_{-T}^{T} = 0,$$

where we have used stationarity of $X(t)$ to ensure that it remains finite as t goes to infinity. We finally obtain, after integration,

$$\bar{n}(a) = \frac{1}{2\pi}\frac{\sigma_{\dot{X}}}{\sigma_X} e^{-a^2/2\sigma_X^2},$$

where

$$\sigma_X^2 = \int_0^\infty S_X^+(\omega)d\omega = M_0,$$
$$\sigma_{\dot{X}}^2 = \int_0^\infty S_{\dot{X}}^+(\omega)d\omega = \int_0^\infty \omega^2 S_X^+(\omega)d\omega = M_2.$$

3.2.3 Frequency of local maxima past a given level

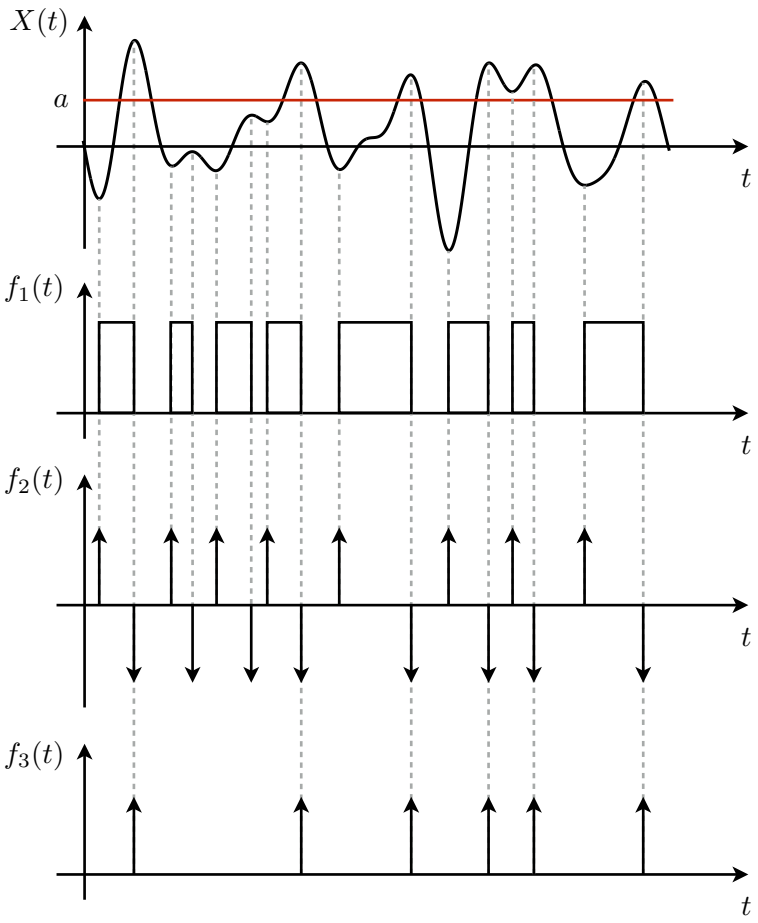


Figure 3.8 – Building a counter for the number of local maxima above a.

We are now interested in quantifying the average frequency of local maxima larger than a certain value a. As done in the previous section, we build a counter function according to the

process outlined in Figure 3.8. First, we build the function $f_1(t)$ that identifies regions of positive slope

$$f_1(t) = u(\dot{X}(t)),$$

where u is the Heaviside function defined previously. Then, we take the derivative of $f_1(t)$ to obtain the function $f_2(t)$, which identifies both local minima and local maxima of $X(t)$,

$$f_2(t) = \frac{df_1}{dt}.$$

Finally, we discard local minima and we only keep those local maxima that are above level a. This is done by defining $f_3(t)$ as

$$f_3(t) = -f_2(t)u(-\ddot{X}(t))u(X(t) - a),$$

with the minus to make the remaining delta functions become positive. Thus, we have obtained the counter function

$$f_3(t) = -\delta(\dot{X}(t))\ddot{X}(t)u(X(t) - a)u(-\ddot{X}(t)).$$

The number of local maxima above a that happen in a given time T is then simply

$$N(T) = \int_0^T -\delta(\dot{X}(t))\ddot{X}(t)u(X(t) - a)u(-\ddot{X}(t))dt,$$

and the average frequency of local maxima above a follows as

$$\bar{n}^m(a) = \lim_{T\to\infty} \frac{N(T)}{T} = \lim_{T\to\infty} \frac{1}{T} \int_0^T -\delta(\dot{X}(t))\ddot{X}(t)u(X(t) - a)u(-\ddot{X}(t))dt.$$

Now, we use ergodicity of $X(t)$ to translate this time average to an ensemble average

$$\begin{aligned}
\bar{n}^m(a) &= E\left[-\ddot{X}\delta(\dot{X})u(X - a)u(-\ddot{X})\right] \\
&= -\int_{-\infty}^{\infty} d\ddot{x} \int_{-\infty}^{\infty} d\dot{x} \int_{-\infty}^{\infty} dx\, \ddot{x}\delta(\dot{x})u(x - a)u(-\ddot{x})f_{X\dot{X}\ddot{X}}(x, \dot{x}, \ddot{x}) \\
&= -\int_{-\infty}^{0} d\ddot{x} \int_{a}^{\infty} dx\, \ddot{x} f_{X\dot{X}\ddot{X}}(x, 0, \ddot{x}).
\end{aligned}$$

For Gaussian stochastic processes, this becomes

$$\bar{n}^m(a) = -\int_{-\infty}^{0} d\ddot{x} \int_{a}^{\infty} dx\, \ddot{x} \frac{1}{\sqrt{2\pi}^3} \frac{1}{\sqrt{M_2\Delta}} e^{-Q/2},$$

where

$$Q = \frac{\dot{x}^2}{2M_2} + \frac{M_4x^2 + 2M_2x\ddot{x} + M_0\ddot{x}^2}{\Delta}, \quad \text{and} \quad \Delta = M_0M_4 - M_2^2,$$

and the average frequency of all local maxima (above any level) is

$$\bar{n}^m(a = -\infty) = \frac{1}{2\pi}\sqrt{\frac{M_4}{M_2}}.$$

3.2.4 Upcrossing rates of a transformed process

Let $X(t)$ a stationary stochastic process with average upcrossing rates $\bar{n}_X(a)$, and define

$$Y(t) = h(X(t)),$$

where h is a given function. What is the rate of upcrossings for $Y(t)$? If, for any given level b of the stochastic process $Y(t)$, the equation $h(a) = b$ has n solutions $a_1, \ldots, a_n$ (see Figure 3.9), then the rate of upcrossings past b of $Y(t)$ is given by

$$\bar{n}_Y(b) = \sum_{j=1}^{n} \bar{n}_X(a_j).$$

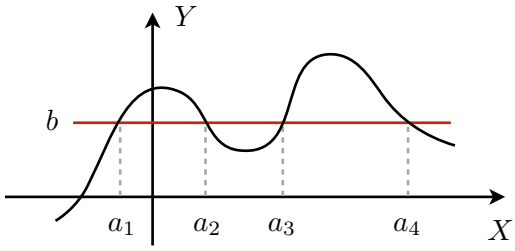


Figure 3.9 – Upcrossing rates of a transformed process.

Example. Let X be a Gaussian, stationary and ergodic stochastic process, with

$$\bar{n}_X(a) = \frac{1}{2\pi}\frac{\sigma_{\dot{X}}}{\sigma_X} e^{-a^2/2\sigma_X^2}.$$

What is $\bar{n}_Y(b)$ for $Y = X^2$? We first solve

$$b = a^2 \quad \Rightarrow \quad a_{1,2} = \pm\sqrt{b},$$

then according to the formula we have

$$\bar{n}_Y(b) = \bar{n}_X(a_1) + \bar{n}_X(a_2) = \frac{1}{2\pi}\frac{\sigma_{\dot{X}}}{\sigma_X}\left(e^{-b/2\sigma_X^2} + e^{-b/2\sigma_X^2}\right) = \frac{1}{\pi}\frac{\sigma_{\dot{X}}}{\sigma_X} e^{-b/2\sigma_X^2}.$$

3.3 Distribution of derivative at upcrossings

The previous section involves the frequency of upcrossings for a stationary stochastic process, $X(t)$. For a wide of problems it is also essential to know the probability distribution function of the upcrossing velocity at a given level a: $v_a = \dot{X}(t|X = a)$, i.e. the velocity at the moment of upcrossing. This is relevant, for example, in the ship slamming problem. Below we present a derivation of analytical expression for this case [1].

Let U describe the upcrossing event at time t:

$$U = \{X(t) < a\} \cap \{X(t + dt) > a\}$$

Obviously, in order to have an upcrossing we need positive velocity at that time. Let us define as V the event where the upcrossing velocity is positive but smaller than v:

$$V = \left\{\dot{X}(t) > 0\right\} \cap \left\{\dot{X}(t) \leq v\right\}$$

The cumulative distribution function for the velocity at the moment of the upcrossing is given by

$$F_v(v;a) = P(V|U) = \frac{P(V \cap U)}{P(U)}, \quad v \geq 0.$$

We now compute each of the terms involved. We first express the intersection of the events U and V as a system of inequalities:

$$U \cap V = \left\{ \begin{array}{l} X(t) < a \\ X(t+dt) > a \\ \dot{X}(t) < v \end{array} \right\} = \left\{ \begin{array}{l} X(t) < a \\ X(t) > a - \dot{x}(t)dt \\ \dot{X}(t) < v \end{array} \right\}$$

The probability of this event can be expressed with the following integral using the joint pdf, $f_{X\dot{X}}(x,\dot{x})$:

$$P(U \cap V) = \int_0^v \int_{a-\dot{x}dt}^a f_{X\dot{X}}(x,\dot{x}) dx d\dot{x}$$

Note that the limits of the inner integral differ only by an infinitesimal quantity, $\dot{x}dt$. Based on this observation the above integral can be written as

$$P(U \cap V) = \int_0^v \dot{x} f_{X\dot{X}}(a,\dot{x}) d\dot{x} dt$$

Using a similar argument we have the probability of the event U where we still need the requirement of a positive upcrossing velocity, $\dot{X}(t) > 0$:

$$P(U) = \int_0^\infty \dot{x} f_{X\dot{X}}(a,\dot{x}) d\dot{x} dt$$

Using the last two expression we finally obtain:

$$F_v(v;a) = \frac{\int_0^v \dot{x} f_{X\dot{X}}(a,\dot{x}) d\dot{x}}{\int_0^\infty \dot{x} f_{X\dot{X}}(a,\dot{x}) d\dot{x}}, \quad v \geq 0.$$

For the special case of independent velocity and position, $f_{X\dot{X}}(x,\dot{x}) = f_X(x) f_{\dot{X}}(\dot{x})$ we have

$$F_v(v) = \frac{\int_0^v \dot{x} f_{\dot{X}}(\dot{x}) d\dot{x}}{\int_0^\infty \dot{x} f_{\dot{X}}(\dot{x}) d\dot{x}}, \quad v \geq 0,$$

and as we observe the cumulative distribution of the upcrossing velocity does not depend on the upcrossing level a. From the last expression we can obtain the pdf of the upcrossing velocity:

$$f_v(v) = \frac{v f_{\dot{X}}(v)}{\int_0^\infty \dot{x} f_{\dot{X}}(\dot{x}) d\dot{x}}, \quad v \geq 0.$$

As it can be easily seen for the case of a normally distributed velocity, the upcrossing velocity follows a Rayleigh distribution.

3.4 Extreme value distribution over a given time interval

Let $X(t)$ a stationary stochastic process, and T a time interval. We define

$$M(T) = \max\{X(t) : 0 \leq t \leq T\},$$

$$\theta(a) = \text{first time that } X(t) \text{ crosses } a.$$

We want to find

$$P(M(T) \leq a) = P(\theta(a) \geq T),$$

for a narrow-banded stochastic process (that is, $\epsilon \simeq 0$). These events are independent, in which case we can model them through a Poisson process. Denoting the rate of hits for this Poisson process as $\bar{n}_X(a)$, the cdf of $M(T)$ is

$$F_{M(T)}(a) = P(M(T) \leq a) = 1 - P(\theta(a) \leq T) = 1 - \left(1 - e^{-\bar{n}_X(a)T}\right) = e^{-\bar{n}_X(a)T},$$

where we have used the formula for the distribution of the first arrival time of a Poisson process from Section 1.12.2.

3.5 Extreme value distribution over long time intervals

Recall that we have previously derived $\bar{n}^m(a)$, the rate of local maxima above level a,

$$\bar{n}^m(a) = -\int_a^\infty du \int_{-\infty}^0 dv\, v f_{X\dot{X}\ddot{X}}(u, 0, v).$$

From this, we can find the rate of maxima above any level as

$$\bar{n}^m(-\infty) = \lim_{a\to\infty} \bar{n}^m(a).$$

Thus, the cumulative distribution of local maxima is simply

$$P(\max X \geq a) = \frac{\text{number of } \max X \geq a}{\text{number of } \max X} = \frac{\text{average rate of } \max X \geq a}{\text{average rate of } \max X} = \frac{\bar{n}^m(a)}{\bar{n}^m(-\infty)},$$

hence

$$F_M(a) = P(\max X \leq a) = 1 - P(\max X \geq a),$$

and the pdf follows as

$$f_M(a) = \frac{dF_M(a)}{da}.$$

For a Gaussian, stationary and ergodic stochastic process, this results in the following distribution of local maxima (see [3], p.298)

$$f_M(a) = \left(\frac{\epsilon^2}{2\pi\sigma_X^2}\right)^{1/2} e^{-a^2/2\epsilon^2\sigma_X^2} + \frac{\sqrt{1-\epsilon^2}}{\sigma_X^2} a u(a) \Phi\left(\frac{\sqrt{1-\epsilon^2}a}{\sigma_X \epsilon}\right) e^{-a^2/2\sigma_X^2}, \quad a \in \mathbb{R},$$

and $u(a)$ is the step function. In the narrow-band limit, for $\epsilon = 0$, the above general formula reduces to the Rayleigh distribution,

$$f_M(a) = \frac{a}{\sigma_X^2} e^{-a^2/2\sigma_X^2}, \quad 0 \leq a < \infty,$$

while in the broad-band limit, for $\epsilon = 1$, we obtain the Gaussian distribution,

$$f_M(a) = \frac{1}{\sqrt{2\pi\sigma_X^2}} e^{-a^2/2\sigma_X^2}, \quad -\infty < a < \infty.$$

Note that the Rayleigh distribution is restricted to positive values of a because a narrow-band stochastic process cannot have local maxima below zero due to the presence of one main frequency component; see Figure 3.4. Moreover, still for the case of narrow-bad stochastic processes, the statistics of local maxima (described by the Rayleigh distribution) are related to the statistics of the envelope, since the latter goes through the local maxima.

3.5.1 The $1/N$th largest maxima

Let us call $a_1, a_2, \ldots, a_n, \ldots$ the local maxima, as pictured in Figure 3.10. We define the $1/N$th largest maxima, $a_{1/N}$, as the value that is exceeded by $1/N$ of the local maxima. For instance, $a_{1/10}$ is the value that is exceeded by 10% of the maxima. How to calculate $a_{1/N}$ from the probability distribution of the local maxima?

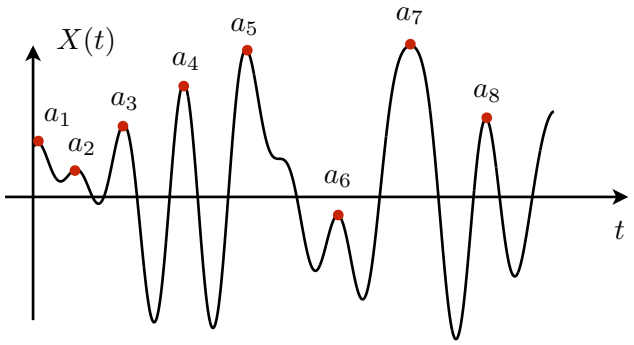


Figure 3.10 – $1/N$th largest maxima.

For simplicity, we focus on the case $\epsilon = 0$, for which the distribution of local maxima follows the Rayleigh distribution

$$f_M(a) = \frac{a}{\sigma_X^2} e^{-a^2/2\sigma_X^2}, \quad 0 \leq a < \infty.$$

Integrating the above formula, we obtain the cdf

$$F_M(\eta_0) = P(\eta \leq \eta_0) = 1 - e^{-\eta_0^2/2}, \quad 0 \leq \eta_0 < \infty,$$

where $\eta = a/\sigma_X$ is a rescaled height. We then want to find the value $\eta^{1/N}$ such that

$$\begin{aligned} P(\eta \geq \eta^{1/N}) = \frac{1}{N} \quad &\Rightarrow \quad 1 - F_M(\eta^{1/N}) = \frac{1}{N} \\ &\Rightarrow \quad e^{-\eta^{1/N\,2}/2} = \frac{1}{N} \\ &\Rightarrow \quad \eta^{1/N} = \sqrt{2\log N} = \frac{a^{1/N}}{\sigma_X}. \end{aligned}$$

In the general case $0 < \epsilon < 1$, the value above which $1/N$ of the local maxima will lie is

$$\eta^{1/N} = \sqrt{2\log\left(\frac{2\sqrt{1-\epsilon^2}}{1+\sqrt{1-\epsilon^2}} N\right)} = \frac{a^{1/N}}{\sigma_X}.$$

3.5.2 The $1/N$th largest average maxima

We now define the $1/N$th largest average maxima, $\bar{a}_{1/10}$, as the average value of all local maxima above $\eta^{1/N}$. Focusing again on the case $\epsilon = 0$, for which the distribution of local maxima follows the Rayleigh distribution,

$$f_M(\eta) = \eta e^{-\eta^2/2}, \quad 0 \leq \eta < \infty,$$

we can readily calculate

$$\begin{aligned}\bar{\eta}^{1/N} = \frac{\bar{a}^{1/N}}{\sigma_X} &= \int_{\eta^{1/N}}^{\infty} \eta\, f_M(\eta|\eta > \eta^{1/N}) d\eta \\ &= \int_{\eta^{1/N}}^{\infty} \eta\, \frac{f_M(\eta \cap \eta > \eta^{1/N})}{P(\eta > \eta^{1/N})} d\eta \\ &= \int_{\eta^{1/N}}^{\infty} \eta\, \frac{f_M(\eta)}{1/N} d\eta \\ &= N \int_{\eta^{1/N}}^{\infty} \eta^2 e^{-\eta^2/2} d\eta.\end{aligned}$$

The 1/3 highest average maxima $\bar{a}^{1/3}$ is called the significant amplitude, and the significant wave height is defined as twice the significant amplitude, $\bar{H}^{1/3} = 2\bar{a}^{1/3}$. For $\epsilon = 0$, we have $\bar{\eta}^{1/3} = 2$, hence $\bar{a}^{1/3} = 2\sigma_X$, and $\bar{H}^{1/3} = 4\sigma_X$.

3.6 Summary on short-term statistics

Let $X(t)$ be a stationary and ergodic stochastic process. We have characterized the statistics of the extreme values of $X(t)$ in the following ways:

1. Measure of bandwidth:

   $$\epsilon^2 = 1 - \frac{M_2^2}{M_0 M_4}, \quad \text{where} \quad M_{0,2,4} = \int_0^{\infty} \omega^{0,2,4} S^+(\omega) d\omega.$$

   Limiting cases: $\epsilon = 0$ (narrow-band), $\epsilon = 1$ (broad-band); see Figure 3.4.

2. Frequency of upcrossings past level a:

   $$\bar{n}(a) = \int_0^{\infty} \dot{x} f_{X\dot{X}}(a, \dot{x}) d\dot{x}.$$

   For Gaussian stochastic processes, this reduces to

   $$\bar{n}(a) = \frac{1}{2\pi} \sqrt{\frac{M_2}{M_0}} e^{-a^2/2M_0}.$$

3. Frequency of local maxima above level a:

   $$\bar{n}^m(a) = -\int_{-\infty}^{0} \int_a^{\infty} \ddot{x} f_{X\dot{X}\ddot{X}}(x, 0, \ddot{x}) dx d\ddot{x}.$$

   For Gaussian stochastic processes, this becomes

   $$\bar{n}^m(a) = -\int_{-\infty}^{0} d\ddot{x} \int_a^{\infty} dx\, \ddot{x} \frac{1}{\sqrt{2\pi}^3} \frac{1}{\sqrt{M_2 \Delta}} e^{-Q/2},$$

   where

   $$Q = \frac{\dot{x}^2}{2M_2} + \frac{M_4 x^2 + 2M_2 x\ddot{x} + M_0 \ddot{x}^2}{\Delta}, \quad \text{and} \quad \Delta = M_0 M_4 - M_2^2,$$

   and the average frequency of all local maxima (above any level) is

   $$\bar{n}^m(-\infty) = \frac{1}{2\pi} \sqrt{\frac{M_4}{M_2}}.$$

4. Frequency of upcrossings past level b for transformed stochastic process $Y = h(X)$:

$$\bar{n}_Y(b) = \sum_{j=1}^{n} \bar{n}_X(a_j),$$

where a_j is solution of $h(a_j) = b$.

5. Probability distribution of maxima over finite time T:

$$M(T) = \max\{X(t) : 0 \le t \le T\},$$
$$P(M(T) \le a) = e^{-\bar{n}(a)T}, \quad \text{for large } a.$$

6. Probability distribution of maxima over long times (for Gaussian stochastic processes):

$$f_M(a) = \frac{a}{M_0} e^{-a^2/2M_0}, \quad 0 \le a < \infty, \quad \text{for } \epsilon = 0 \text{ (narrow-band limit)},$$
$$f_M(a) = \frac{1}{\sqrt{2\pi M_0}} e^{-a^2/2M_0}, \quad -\infty < a < \infty, \quad \text{for } \epsilon = 1 \text{ (broad-band limit)}.$$

7. Average of all $1/N$th largest maxima (for Gaussian stochastic processes):

$$\bar{a}^{1/N} = \sqrt{2M_0 \log\left(\frac{2\sqrt{1-\epsilon^2}}{1+\sqrt{1-\epsilon^2}} N\right)}.$$

Significant wave height:

$$\bar{H}^{1/3} = 2\bar{a}^{1/3} \quad (= 4\sigma_X \text{ for } \epsilon = 0).$$

3.7 Long-term statistics

3.7.1 Exceedance probability

Short-term statistics are valid over a period of up to a few days, during which the spectrum characterizing the stochastic process can be considered constant. Long-term statistics can be seen as the "sum" of several short-term statistics, each with different spectra.

Example. To describe the long-term statistics of the sea surface during a series of storms, we first characterize each storm i through its spectrum $S_i^+(\omega)$, which is parameterized in terms of the significant height $\bar{H}_i^{1/3}$ and average period $\bar{T}_i$ corresponding to that storm, as done in Table 3.2.

Storm	$\bar{H}^{1/3}$	$\bar{T}$	Probability
1	2 m	4 s	0.05
2	3 m	5 s	0.08
$\vdots$			

Table 3.2 – Table of storm statistics.

During storm i, assuming Gaussian statistics, the frequency of exceeding level a_0 is

$$\lambda_i = \frac{1}{2\pi}\sqrt{\frac{M_{2,i}}{M_{0,i}}}e^{-a_0^2/2M_{0,i}} = \frac{1}{\bar{T}_i}e^{-a_0^2/2M_{0,i}} = \frac{1}{\bar{T}_i}\mu_i.$$

If T is the life of the structure, then the total duration of each storm i is

$$T_i = TP_i,$$

where P_i is the probability of storm i. Then, the total number of exceedances of a_0 in storm i is

$$N_i = \lambda_i T_i = \frac{\mu_i}{\bar{T}_i}P_i T,$$

and the total number of exceedances of level a_0 over all storms is

$$N_{a_0} = \sum_i N_i = \sum_i \frac{T}{\bar{T}_i}P_i e^{-a_0^2/2M_{0,i}}.$$

For a narrow-band stochastic process, we have seen that $\bar{H}_i^{1/3} \simeq 4\sqrt{M_{0,i}}$, hence $M_{0,i} \simeq (\bar{H}_i^{1/3}/4)^2$ and

$$N_{a_0} = \sum_i \frac{T}{\bar{T}_i}P_i e^{-2h_0^2/(\bar{H}_i^{1/3})^2},$$

where we have defined $h_0 = 2a_0$. The total number of cycles is found by setting $a_0 = 0$, that is,

$$N_0 = \sum_i \frac{T}{\bar{T}_i}P_i.$$

The long-term probability of exceeding level h_0 is thus

$$P(h > h_0) = \frac{N_{a_0}}{N_0} = \frac{\sum_i T/\bar{T}_i\, P_i e^{-2h_0^2/(\bar{H}_i^{1/3})^2}}{\sum_i T/\bar{T}_i\, P_i} = \frac{E[1/\bar{T}e^{-2h_0^2/(\bar{H}^{1/3})^2}]}{E[1/\bar{T}]}.$$

In practical calculations the effect of $\bar{T}$ is relatively small, hence its can be treated as a deterministic variable and we finally have

$$P(h > h_0) = E\left[e^{-2h_0^2/(\bar{H}^{1/3})^2}\right].$$

3.7.2 The 100-year wave h_{100}

The 100-year wave h_{100} is defined as the wave height h_{100} that will be exceeded on average once in 100 years, that is,

$$P(h > h_{100}) = E\left[e^{-2h_{100}^2/(\bar{H}^{1/3})^2}\right] = \frac{N(\text{waves above } h_{100})}{N_0} = \frac{1}{100 \text{ years}/\bar{T}}.$$

3.7.3 How safe is the design based on the 100-year wave?

A structure is designed to just barely withstand the 100-year wave. Its desired lifetime is M years. What is the probability of failure during its lifetime?

We model the waves as a sequence of independent events. We have

$$P(\text{structure fails}) = P(\text{at least one wave above } h_{100}) = 1 - P(\text{no wave above } h_{100}).$$

Denoting N_M the number of waves in M years and using a Bernoulli process, we have

$$P(\text{no wave above } h_{100}) = q^{N_M},$$

where q is the probability that any single wave is below h_{100}. To find q, note that the probability p that any particular wave is above h_{100} is by definition equal to $1/N_{100} \ll 1$, hence

$$q = 1 - p = 1 - \frac{1}{N_{100}} \simeq e^{-1/N_{100}}.$$

where we have expressed the difference of the two terms as exponential using a Taylor approximation. Finally,

$$P(\text{structure fails}) = 1 - q^{N_M} \simeq 1 - e^{-N_M/N_{100}} = 1 - e^{-M/100}.$$

Table 3.3 indicates P(structure fails) for different vales of M.

M (years)	Probability of failure
1	1%
5	4.9%
10	9.5%
20	18.1%
50	39.3%
100	63.2%

Table 3.3 – Probability of failure of design based on h_{100}.

Chapter 4

Laplace Transform and System Analysis

4.1 Laplace transform

The Fourier transform that we have seen earlier requires the function to be absolutely integrable, that is,

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

When analyzing unstable systems, however, one often encounters responses growing in time, for which the Fourier transform does not converge. The Laplace transform is a generalized version of the Fourier transform that exists for a much broader range of functions. More specifically, consider a function $y(t)$ such that $y(t) = 0$ for $t < 0$, and

$$\int_{-\infty}^{\infty} |y(t)|dt = \infty,$$

for example the function pictured in Figure 4.1. If there exists a real a_0 such that

$$\int_{-\infty}^{\infty} |y(t)e^{-a_0 t}|dt < \infty,$$

then the "weighted" function $y(t)e^{-at}$, $a \geq a_0$, is absolutely integrable and we can find its Fourier transform

$$\mathcal{F}[y(t)e^{-at}] = \int_0^{\infty} y(t)e^{-at}e^{-i\omega t}dt = \int_0^{\infty} y(t)e^{(a+i\omega)t}dt,$$

which is a function of both a and ω. We can now define the complex variable $s = a + i\omega$, and define the Laplace transform of $y(t)$ as

$$\hat{y}(s) = \mathcal{L}[y(t)] = \mathcal{F}[y(t)e^{-at}] = \int_0^{\infty} y(t)e^{-st}dt, \quad \text{Re}[s] \geq a_0.$$

By taking the inverse Fourier transform, we have

$$y(t)e^{-at} = \mathcal{F}^{-1}[\hat{y}(a + i\omega)] = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{y}(a + i\omega)e^{i\omega t}d\omega,$$

thus

$$y(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{y}(a + i\omega)e^{at}e^{i\omega t}d\omega.$$

With the change of variables $s = a + i\omega$, $ds = id\omega$, we obtain the inverse Laplace transform,

$$y(t) = \mathcal{L}^{-1}[\hat{y}(s)] = \frac{1}{2\pi i} \lim_{\omega\to\infty} \int_{a-i\omega}^{a+i\omega} \hat{y}(s)e^{st}ds, \quad \text{Re}[s] \geq a_0.$$

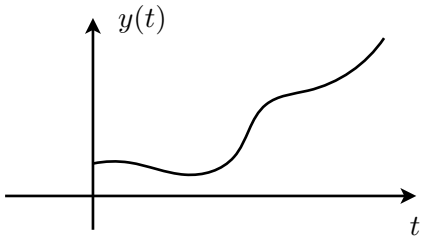


Figure 4.1 – A non-absolutely integrable function.

The evaluation of the inverse Laplace transform requires integration in the complex s-plane along a path parallel to the imaginary axis, such that $\text{Re}[s] \geq a_0$. This defines the region of convergence (ROC) of the Laplace transform; see Figure 4.2.

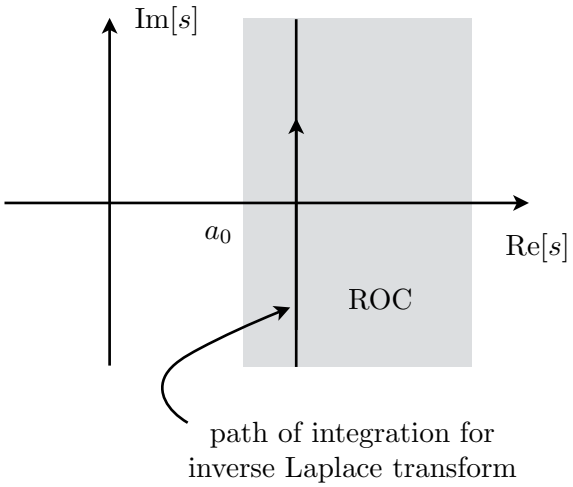


Figure 4.2 – Region of convergence (ROC) of the Laplace transform.

In summary, for a function $y(t)$ such that $y(t) = 0$, $t < 0$, the direct and inverse Laplace transforms are defined respectively as

$$\hat{y}(s) = \mathcal{L}[y(t)] = \int_0^\infty y(t)e^{-st}dt,$$
$$y(t) = \mathcal{L}^{-1}[\hat{y}(s)] = \frac{1}{2\pi i} \lim_{\omega\to\infty} \int_{a-i\omega}^{a+i\omega} \hat{y}(s)e^{st}ds.$$

These are defined only for $\text{Re}[s] \geq a_0$, the region of convergence. The Laplace transform possesses the following properties:

1. $\mathcal{L}[ag(t) + bh(t)] = a\mathcal{L}[g(t)] + b\mathcal{L}[h(t)]$.
2. $\mathcal{L}\left[\dfrac{df}{dt}\right] = \displaystyle\int_0^\infty \frac{df}{dt}e^{-st}dt = f(t)e^{-st}\Big|_0^\infty + \int_0^\infty f(t)se^{-st}dt = -f(0) + s\mathcal{L}[f(t)]$.

$$\mathcal{L}\left[\frac{d^2f}{dt^2}\right] = -\frac{df}{dt}\bigg|_{t=0} + s\mathcal{L}\left[\frac{df}{dt}\right] = -\frac{df}{dt}\bigg|_{t=0} - sf(0) + s^2\mathcal{L}[f(t)].$$

3. Final value theorem: if $\lim_{t\to\infty} f(t)$ exists, then

$$\lim_{t\to\infty} f(t) = \lim_{s\to 0} s\hat{f}(s).$$

Proof. We write

$$\mathcal{L}\left[\frac{df}{dt}\right] = \int_0^\infty \frac{df}{dt}e^{-st}dt = s\hat{f}(s) - f(0).$$

Taking the limit as $s \to 0$, we get

$$\lim_{s\to 0} s\hat{f}(s) = \lim_{s\to 0}\int_0^\infty \frac{df}{dt}e^{-st}dt + f(0) = \int_0^\infty \frac{df}{dt}\lim_{s\to 0} e^{-st}dt + f(0) = \int_0^\infty \frac{df}{dt}dt + f(0) = \lim_{t\to\infty} f(t).$$

4. $\mathcal{L}[f(t) * g(t)] = \mathcal{L}[f(t)] \cdot \mathcal{L}[g(t)]$.

Examples. Common Laplace transforms:

1. $\mathcal{L}[1] = \int_0^\infty e^{-st}dt = \frac{1}{s}, \quad \mathrm{Re}[s] > 0.$
2. $\mathcal{L}[t] = \int_0^\infty te^{-st}dt = \frac{1}{s^2}, \quad \mathrm{Re}[s] > 0.$
3. $\mathcal{L}[e^{\lambda t}] = \int_0^\infty e^{(\lambda-s)t}dt = \frac{1}{s-\lambda}, \quad \mathrm{Re}[s] > \lambda.$
4. $\mathcal{L}[t^k e^{\lambda t}] = \int_0^\infty t^k e^{(\lambda-s)t}dt = \frac{k!}{(s-\lambda)^{k+1}}, \quad \mathrm{Re}[s] > \lambda.$

4.2 Solving linear systems with the Laplace transform

4.2.1 Computing the inverse Laplace transform

To compute the inverse Laplace transform, one needs in principle to evaluate an integral in the complex s-plane, along a path parallel to the imaginary axis as shown in Figure 4.2. In practice, we can avoid computing explicitly this integral by using partial fractions expansion to expand the Laplace transform into a sum of simple fractions, which can then be converted to the time domain using Laplace transform tables or the few common Laplace transforms listed in the above section. In systems analysis, the Laplace transform is usually of the form

$$\hat{y}(s) = \frac{N(s)}{D(s)} = \frac{N(s)}{(s-\lambda_1)^{k_1}\dots(s-\lambda_n)^{k_n}},$$

where $N(s)$ and $D(s)$ are polynomials, and the degree of $N(s)$ is at most equal to that of $D(s)$. We define the roots λ_i of $D(s)$ as the poles of $\hat{y}(s)$, and the multiplicity k_i of each root as the order of the corresponding pole. Then, using partial fractions expansion we can expand $\hat{y}(s)$ as

$$\hat{y}(s) = \sum_{i=1}^{n}\sum_{j=1}^{k_i}\frac{a_{ij}}{(s-\lambda_i)^j},$$

where a_{ij} are constant coefficients. The inverse Laplace transform of each individual term can then be computed using the formula

$$\mathcal{L}^{-1}\left[\frac{k!}{(s-\lambda)^{k+1}}\right] = t^k e^{\lambda t}.$$

We note that a pole of any order $n \geq 1$ at $s = \lambda$ induces a time response that behaves as $e^{\lambda t}$ as time goes to infinity. Thus, if a function $\hat{y}(s)$ contains many poles, its region of convergence will be $\mathrm{Re}[s] \geq a_0$, where a_0 is the real part of the rightmost pole λ_i.

Example. The following Laplace transform

$$\hat{f}(s) = \frac{s+2}{s(s^2+4)},$$

has first-order poles at $s = 0$, $-2i$, $2i$, thus it can be expanded as

$$\hat{f}(s) = \frac{s+2}{s(s+2i)(s-2i)} = \frac{a_1}{s} + \frac{a_2}{s+2i} + \frac{a_3}{s-2i},$$

where the coefficients a_1, a_2, and a_3 can be found by taking the limits

$$\begin{aligned}
a_1 &= \lim_{s\to 0} s\hat{f}(s) = \lim_{s\to 0} \frac{s+2}{(s+2i)(s-2i)} = \frac{1}{2}, \\
a_2 &= \lim_{s\to -2i} (s+2i)\hat{f}(s) = \lim_{s\to -2i} \frac{s+2}{s(s-2i)} = -\frac{1-i}{4}, \\
a_3 &= \lim_{s\to 2i} (s-2i)\hat{f}(s) = \lim_{s\to 2i} \frac{s+2}{s(s+2i)} = -\frac{1+i}{4}.
\end{aligned}$$

Therefore,

$$\hat{f}(s) = \frac{1}{2s} - \frac{1-i}{4}\frac{1}{s+2i} - \frac{1+i}{4}\frac{1}{s-2i},$$

and we obtain, by identification,

$$f(t) = \frac{1}{2} - \frac{1-i}{4}e^{-2it} - \frac{1+i}{4}e^{2it} = \frac{1}{2}(1 - \cos 2t - \sin 2t), \quad t \geq 0.$$

Example. Consider the Laplace transform

$$\hat{f}(s) = \frac{5s^2+3s+1}{(s+2)^2(s+1)},$$

with a first-order pole at $s = -2$ and a second-order pole at $s = -2$. We express the function as:

$$\frac{a_1}{s+2} + \frac{a_2}{(s+2)^2} + \frac{a_3}{s+1} = \frac{5s^2+3s+1}{(s+2)^2(s+1)}.$$

Multiplying with $s+1$ and taking the limit $s \to -1$, then with $(s+2)^2$ and taking the limit $s \to -2$ and finally with $s+2$ and taking the same limit, results in the partial fractions expansion:

$$\hat{f}(s) = \frac{a_1}{s+2} + \frac{a_2}{(s+2)^2} + \frac{a_3}{s+1} = \frac{2}{s+2} - \frac{15}{(s+2)^2} + \frac{3}{s+1},$$

and the time domain response is

$$f(t) = 2e^{-2t} - 15te^{-2t} + 3e^{-t}, \quad t \geq 0.$$

Notice the te^{-2t} term arising from the second-order pole.

4.2.2 Response to forcing

Example. Consider the system shown in Figure 4.3 and governed by the following equation

$$m\frac{d^2y}{dt^2} + ky = f(t), \quad y(0) = \dot{y}(0) = 0, \quad f(t) = \delta(t).$$

Taking the Laplace transform on both sides, we obtain

$$ms^2\hat{y}(s) + k\hat{y}(s) = 1,$$

or equivalently,

$$\hat{y}(s) = \frac{1}{ms^2 + k} = \frac{1/m}{(s + i\sqrt{k/m})(s - i\sqrt{k/m})}.$$

By applying partial fractions and considering the inverse Laplace transform we have

$$y(t) = \frac{1}{\sqrt{km}} \sin \sqrt{\frac{k}{m}}t, \quad t \geq 0$$

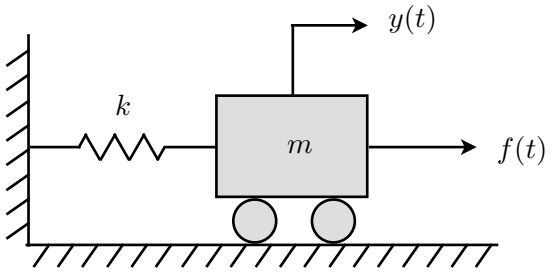


Figure 4.3 – Spring-mass system.

The above example shows that the Laplace transform is useful in calculating the response of stable or unstable systems with given initial conditions to external excitation. Another area of application of the Laplace transform is in determining the nature of the response of a system under stochastic excitation with given spectrum. Recall that the response of an LTI system with input $u(t)$ and transfer function $h(t)$ can be written as the convolution

$$y(t) = h(t) * u(t),$$

for zero initial conditions (we will explore the effect of the latter in the next section). Taking the Laplace transform, we have

$$\hat{y}(s) = \hat{h}(s)\hat{u}(s),$$

which shows that $\hat{y}(s)$ will inherit the poles contained in both $\hat{h}(s)$ and $\hat{u}(s)$. This has important consequences for the behavior of $y(t)$, as we illustrate in the example below.

Example. Let the system

$$m\frac{d^2y}{dt^2} + ky = f(t, \zeta), \quad y(0) = \dot{y}(0) = 0,$$

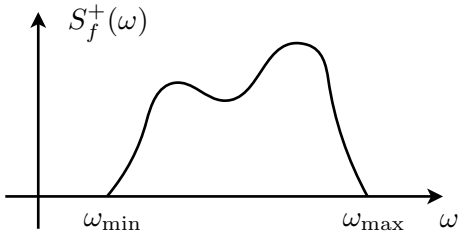


Figure 4.4 – Spectrum of forcing.

where $f(t,\zeta)$ is a Gaussian, stationary and ergodic stochastic process with one-sided spectrum shown in Figure 4.4. As time tends to infinity, will the response $y(t,\zeta)$ tend to a stationary and ergodic random process?

First, we take the Laplace transform of the governing equation and we get

$$\hat{y}(s,\zeta) = \frac{\hat{f}(s,\zeta)}{ms^2+k} = \hat{h}(s)\hat{f}(s,\zeta),$$

which shows that $\hat{h}(s)$ has first-order poles at $s = \pm i\sqrt{k/m}$. We then recall the decomposition

$$f(t,\zeta) = \sum_n a_n \cos(\omega_n t + \theta_n(\zeta)) = \sum_n a_n \frac{e^{i(\omega_n t+\theta_n)} + e^{-i(\omega_n t+\theta_n)}}{2},$$

where a_n are deterministic amplitudes, ω_n are deterministic frequencies, $\theta_n \sim \mathcal{U}(0,2\pi)$ are iid random phases, and $\Delta = \omega_{n+1} - \omega_n \to 0$ in the continuous spectrum limit. By linearity of the Laplace transform,

$$\hat{f}(s,\zeta) = \frac{1}{2}\sum_n a_n e^{i\theta_n}\mathcal{L}[e^{i\omega_n t}] + \text{c.c.} = \frac{1}{2}\sum_n a_n \frac{e^{i\theta_n}}{s - i\omega_n} + \text{c.c.},$$

where c.c. denotes complex conjugate. Thus, $\hat{f}(s,\zeta)$ has first-order poles on the imaginary axis at $s = \pm i\omega_n$, $n = 1, 2, \ldots$ with $\omega_{\min} \le \omega_n \le \omega_{\max}$. The Laplace transform of the response is

$$\begin{aligned}\hat{y}(s,\zeta) &= \frac{1/m}{(s+i\sqrt{k/m})(s-i\sqrt{k/m})}\hat{f}(s,\zeta) \\ &= \frac{1}{2m}\sum_n a_n \frac{e^{i\theta_n}}{(s-i\omega_n)(s+i\sqrt{k/m})(s-i\sqrt{k/m})} + \text{c.c.}\end{aligned}$$

As illustrated in Figure 4.5, two cases are now possible:

1. If $\sqrt{k/m} < \omega_{\min}$ or $\sqrt{k/m} > \omega_{\max}$, as shown in Figure 4.5(a), then $\hat{y}(s)$ will only contain a continuous distribution of first-order poles situated at $s = \pm i\omega_n$, $n = 1, 2, \ldots$ and a pair of poles at $s = \pm i\sqrt{k/m}$. Since

$$\mathcal{L}^{-1}\left[\frac{e^{i\theta_n}}{s \pm i\omega_n}\right] = e^{\pm i(\omega_n t+\theta_n)},$$

the response $y(t,\zeta)$ will be composed of sinusoidal frequency components with iid random phases, hence it will be stationary and ergodic and we can use the Wiener-Khinchine relations to characterize its spectrum.

2. If $\omega_{\min} \leq \sqrt{k/m} \leq \omega_{\max}$, in which case $\omega_n = \sqrt{k/m}$ for some n as shown in Figure 4.5(b), then $\hat{y}(s)$ will contain a pair of second-order poles at $s = \pm i\sqrt{k/m}$. Since

$$\mathcal{L}^{-1}\left[\frac{e^{i\theta_n}}{(s \pm i\sqrt{k/m})^2}\right] = te^{\pm i(\sqrt{k/m}\,t+\theta_n)},$$

the response $y(t,\zeta)$ will grow algebraically with time! It is thus not stationary and we cannot use the Wiener-Khinchine relations since it does not have a spectrum.

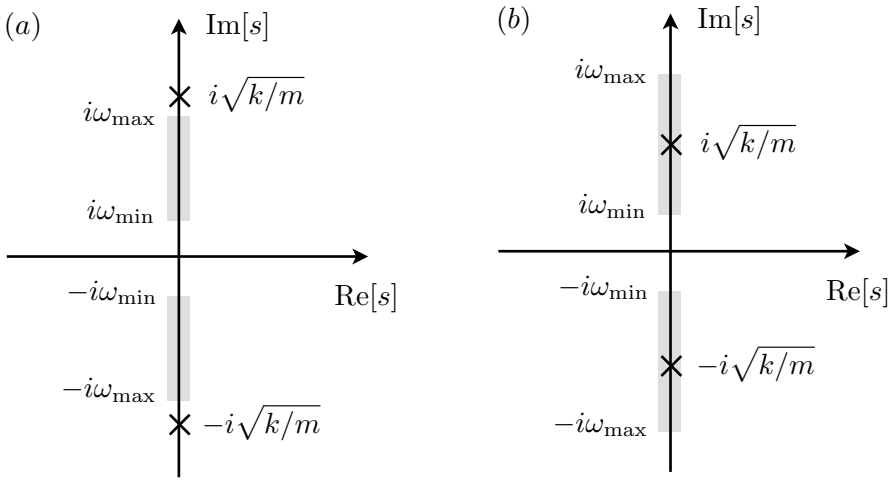


Figure 4.5 – Interplay between poles of the system and poles of the forcing.

4.3 Finite-dimensional LTI systems

Consider the finite-dimensional LTI system, written in state-space form,

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u},$$
$$\mathbf{y} = \mathbf{C}\mathbf{x},$$

where $\mathbf{x} \in \mathbb{R}^n$ is the state, $\mathbf{u} \in \mathbb{R}^m$ is the input, $\mathbf{y} \in \mathbb{R}^l$ is the output, and the matrices $\mathbf{A} \in \mathbb{R}^{n\times n}$, $\mathbf{B} \in \mathbb{R}^{n\times m}$, $\mathbf{C} \in \mathbb{R}^{l\times n}$. Note that the state variables are not unique. Indeed, we can always choose a non-singular matrix $\mathbf{M} \in \mathbb{R}^{n\times n}$ such that

$$\mathbf{x} = \mathbf{M}\mathbf{z} \quad \Leftrightarrow \quad \mathbf{z} = \mathbf{M}^{-1}\mathbf{x},$$

where $\mathbf{z}$ is the state in the new coordinate system. In the new coordinates $\mathbf{z}$, the dynamics become

$$\dot{\mathbf{z}} = (\mathbf{M}^{-1}\mathbf{A}\mathbf{M})\mathbf{z} + (\mathbf{M}^{-1}\mathbf{B})\mathbf{u},$$
$$\mathbf{y} = (\mathbf{C}\mathbf{M})\mathbf{z},$$

Example. Let $\mathbf{x} = (x_1, \ldots, x_n)$ describe a scalar field in a turbulent flow, discretized on a grid with n points. Applying a Fourier transform, we have

$$\mathbf{x} = \sum_{i=1}^{n} \alpha_i \tilde{\mathbf{x}}_i,$$

where α_i are Fourier coefficients, and $\tilde{\mathbf{x}}_i$ are Fourier modes. The vector of Fourier coefficients $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$ then represents the state in a new coordinate system (the Fourier basis).

Example. Consider the driven pendulum shown in Figure (4.6). The governing equations are

$$\begin{aligned} &l\frac{d^2\phi}{dt^2} + \frac{d^2u_0}{dt^2}\cos\phi + mg\sin\phi = 0, \\ &y = l\sin\phi, \end{aligned}$$

where u_0 is the prescribed input, and y is the observed output. Assuming that ϕ, $\dot{\phi}$ are small, we can linearize by approximating $\cos\phi \simeq 1$, $\sin\phi \simeq \phi$ and retaining only first-order terms, leading to

$$\begin{aligned} &l\frac{d^2\phi}{dt^2} + mg\phi = -\frac{d^2u_0}{dt^2} = u(t), \\ &y = l\phi. \end{aligned}$$

Defining the state as $\boldsymbol{\phi} = (\phi, \dot{\phi})^\mathsf{T}$, we can then write the above equations in state space form,

$$\begin{aligned} \dot{\boldsymbol{\phi}} &= \begin{bmatrix} 0 & 1 \\ -mg/l & 0 \end{bmatrix}\boldsymbol{\phi} + \begin{bmatrix} 0 \\ 1/l \end{bmatrix}u(t), \\ y &= \begin{bmatrix} l & 0 \end{bmatrix}\boldsymbol{\phi}. \end{aligned}$$

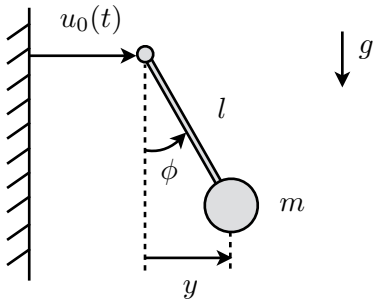


Figure 4.6 – Driven pendulum.

Example. Consider the taut string under tension T_0 depicted in Figure 4.7, driven from the top with motion $y(0,t) = u_0(t)$, and fixed at the bottom. The governing equations are

$$m\frac{\partial^2 y}{dt^2} = \left[T_0 + EA\left(\frac{\partial y}{\partial x}\right)^2\right]\frac{\partial^2 y}{\partial x^2},$$

where E is the Young's modulus, A is the cross-sectional area. The boundary conditions are $y(0,t) = u_0(t), y(l,t) = 0$ where l is the length of the string. If y and $\partial y/\partial x$ are small, we have

$$EA\left(\frac{\partial y}{\partial x}\right)^2 \ll T_0,$$

and we can linearize the governing equation to get

$$m\frac{\partial^2 y}{dt^2} = T_0\frac{\partial^2 y}{\partial x^2}.$$

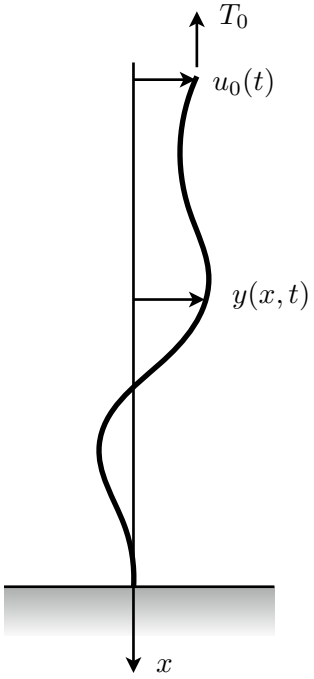


Figure 4.7 – Taut string.

To make the problem finite-dimensional, we apply the governing equation at discrete points x_k, with motion $y(x_k, t)$ discretized as

$$y_k(t) = y(x_k, t), \quad k = 1, \ldots, N.$$

Using a second-order finite difference approximation for the spatial derivative, the discretized governing equation becomes

$$m\ddot{y}_k = T_0 \frac{y_{k+1} - 2y_k + y_{k-1}}{h^2},$$

where $h = x_{k+1} - x_k$. With $N = 3$ points, as shown in Figure 4.8, and applying the boundary conditions, we get

$$\begin{aligned} m\ddot{y}_1 &= \frac{T_0}{h^2}(y_2 - 2y_1 + u_0), \\ m\ddot{y}_2 &= \frac{T_0}{h^2}(y_3 - 2y_2 + y_1), \\ m\ddot{y}_3 &= \frac{T_0}{h^2}(0 - 2y_3 + y_2). \end{aligned}$$

Thus, defining the state $\mathbf{x} = (y_1, y_2, y_3, \dot{y}_1, \dot{y}_2, \dot{y}_3)^\mathsf{T}$ and $\delta = T/h^2 m$, the state-space equations are

$$\dot{\mathbf{x}} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ -2\delta & \delta & 0 & 0 & 0 & 0 \\ \delta & -2\delta & \delta & 0 & 0 & 0 \\ \delta & \delta & -2\delta & 0 & 0 & 0 \end{bmatrix} \mathbf{x} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ \delta \\ 0 \\ 0 \end{bmatrix} u_0.$$

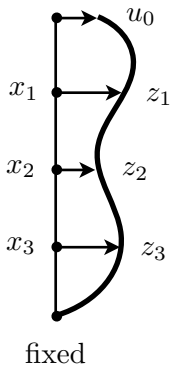


Figure 4.8 – Taut string discretization.

4.3.1 Response to forcing

The Laplace transform is useful in determining the response of stable and unstable systems to forcing. Consider the system

$$\begin{aligned}\dot{\mathbf{x}} &= \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u},\\ \mathbf{y} &= \mathbf{C}\mathbf{x},\end{aligned}$$

with initial conditions $\mathbf{x}(0) = \mathbf{x}_0$. Taking the Laplace transform of the above governing equations,

$$\begin{aligned}s\hat{\mathbf{x}}(s) - \mathbf{x}(0) &= \mathbf{A}\hat{\mathbf{x}}(s) + \mathbf{B}\hat{\mathbf{u}}(s)\\ \hat{\mathbf{y}}(s) &= \mathbf{C}\hat{\mathbf{x}}(s),\end{aligned}$$

hence, combining the two equations we get

$$\hat{\mathbf{y}}(s) = \underbrace{\mathbf{C}[s\mathbf{I} - \mathbf{A}]^{-1}\mathbf{x}_0}_{\text{effect of IC}} + \underbrace{\mathbf{C}[s\mathbf{I} - \mathbf{A}]^{-1}\mathbf{B}\hat{\mathbf{u}}(s)}_{\text{effect of forcing}}.$$

The transfer function $\mathbf{H}_L(s)$ (in the Laplace sense) is defined as

$$\mathbf{H}_L(s) = \mathbf{C}[s\mathbf{I} - \mathbf{A}]^{-1}\mathbf{B}.$$

To find the response $\mathbf{y}(t)$ in the time domain, we will decompose $\hat{\mathbf{y}}(s)$ into contributions from different modes with different frequencies, in the same spirit as the partial fractions expansion seen previously. Consider that $\mathbf{A}$ is a real $n \times n$ matrix with n distinct eigenvalues λ_i. The right eigenvectors $\mathbf{v}_i$ of $\mathbf{A}$ are defined as

$$\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i, \quad i = 1, \ldots, n,$$

and the left eigenvectors $\mathbf{w}_i$ of $\mathbf{A}$ are defined as

$$\mathbf{A}^\mathsf{T}\mathbf{w}_i = \lambda_i \mathbf{w}_i, \quad i = 1, \ldots, n,$$

where $\mathbf{A}^\mathsf{T}$ is the transpose of $\mathbf{A}$, and possesses the same eigenvalues λ_i. Note that the name 'left' eigenvectors stems from the fact that the above equation is equivalent to

$$\mathbf{w}_i^\mathsf{T}\mathbf{A} = \lambda_i \mathbf{w}_i^\mathsf{T}, \quad i = 1, \ldots, n.$$

It is straightforward to show that the right and left eigenvectors verify the biorthogonal property

$$\mathbf{w}_i^\mathsf{T}\mathbf{v}_j = \begin{cases} 1 & i = j, \\ 0 & i \neq j. \end{cases}$$

Thus, if we define $\mathbf{V} = [\mathbf{v}_1 \dots \mathbf{v}_n]$, $\mathbf{W} = [\mathbf{w}_1 \dots \mathbf{w}_n]$, $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$, we have the relation

$$\mathbf{W}^\mathsf{T}\mathbf{V} = \mathbf{V}\mathbf{W}^\mathsf{T} = \mathbf{I},$$

which implies the spectral decomposition of the matrix $\mathbf{A}$,

$$\mathbf{A}\mathbf{V} = \mathbf{V}\mathbf{\Lambda} \quad \Rightarrow \quad \mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{W}^\mathsf{T} = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i \mathbf{w}_i^\mathsf{T}.$$

Similarly, we have the spectral decompositions

$$s\mathbf{I} - \mathbf{A} = s\mathbf{V}\mathbf{W}^\mathsf{T} - \mathbf{V}\mathbf{\Lambda}\mathbf{W}^\mathsf{T} = \mathbf{V}(s\mathbf{I} - \mathbf{\Lambda})\mathbf{W}^\mathsf{T} = \sum_{i=1}^{n} \mathbf{v}_i \mathbf{w}_i^\mathsf{T}(s - \lambda_i),$$

$$(s\mathbf{I} - \mathbf{A})^{-1} = (\mathbf{W}^\mathsf{T})^{-1}(s\mathbf{I} - \mathbf{\Lambda})^{-1}\mathbf{V}^{-1} = \mathbf{V}(s\mathbf{I} - \mathbf{\Lambda})^{-1}\mathbf{W}^\mathsf{T} = \sum_{i=1}^{n} \mathbf{v}_i \mathbf{w}_i^\mathsf{T} \frac{1}{s - \lambda_i}.$$

Inserting the above decomposition into the expression for $\hat{\mathbf{y}}(s)$, we obtain

$$\hat{\mathbf{y}}(s) = \underbrace{\sum_{i=1}^{n} (\mathbf{C}\mathbf{v}_i) \frac{1}{s - \lambda_i} (\mathbf{w}_i^\mathsf{T}\mathbf{x}_0)}_{\text{effect of IC}} + \underbrace{\sum_{i=1}^{n} (\mathbf{C}\mathbf{v}_i) \frac{1}{s - \lambda_i} (\mathbf{w}_i^\mathsf{T}\mathbf{B})\hat{\mathbf{u}}(s)}_{\text{effect of forcing}}.$$

The response in the time domain can now be found by identifying the inverse Laplace transform of individual terms,

$$\mathbf{y}(t) = \sum_{i=1}^{n} (\mathbf{C}\mathbf{v}_i) e^{\lambda_i t} (\mathbf{w}_i^\mathsf{T}\mathbf{x}_0) + \sum_{i=1}^{n} (\mathbf{C}\mathbf{v}_i)(\mathbf{w}_i^\mathsf{T}\mathbf{B}) \int_0^t e^{\lambda_i (t-\tau)} \mathbf{u}(\tau) d\tau, \quad t \geq 0,$$

where the last integral comes from the fact that

$$\mathcal{L}^{-1}\left[\frac{1}{s - \lambda_i}\hat{\mathbf{u}}(s)\right] = e^{\lambda_i t} * \mathbf{u}(t) = \int_0^t e^{\lambda_i (t-\tau)} \mathbf{u}(\tau) d\tau.$$

4.3.2 Wiener-Khinchine relations

Consider the system written in state-space form,

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}(t, \zeta), \\ \mathbf{y} &= \mathbf{C}\mathbf{x}, \end{aligned}$$

with zero initial conditions, and $\mathbf{u}(t, \zeta)$ is a stationary and ergodic zero-mean random process with autocorrelation function

$$\mathbf{R}_{\mathbf{uu}}(\tau) = E[\mathbf{u}(t)\mathbf{u}^\mathsf{T}(t + \tau)] \quad \Leftrightarrow \quad \{\mathbf{R}_{\mathbf{uu}}\}_{ij}(\tau) = E[u_i(t)u_j(t + \tau)],$$

and spectrum

$$\mathbf{S_{uu}}(\omega) = \int_{-\infty}^{\infty} \mathbf{R_{uu}}(\tau) e^{-i\omega\tau} d\tau \quad \Leftrightarrow \quad \{\mathbf{S_{uu}}\}_{ij}(\omega) = \int_{-\infty}^{\infty} \{\mathbf{R_{uu}}\}_{ij}(\tau) e^{-i\omega\tau} d\tau.$$

We have seen that the transfer function (in the Laplace sense) of the system is given by

$$\mathbf{H}_L(s) = \mathbf{C}[s\mathbf{I} - \mathbf{A}]^{-1}\mathbf{B}.$$

If the system is stable (that is, the poles of $\mathbf{H}_L(s)$ are in the left-half complex plane, or equivalently, the eigenvalues λ_i of $\mathbf{A}$ have negative real part), then a Fourier transform of the governing equations leads to the transfer function $\mathbf{H}_F(\omega)$ (in the Fourier sense),

$$\mathbf{H}_F(\omega) = \mathbf{C}[i\omega\mathbf{I} - \mathbf{A}]^{-1}\mathbf{B},$$

which is the finite-dimensional generalization of the transfer function we have seen previously in Section 2.5.6. Note that the two definitions of the transfer function (in the Fourier or Laplace sense) are related as $\mathbf{H}_F(\omega) = \mathbf{H}_L(i\omega)$.

Thus, if the system is stable, we have the finite-dimensional generalization of the Wiener-Khinchine theorem,

$$\mathbf{S_{yy}}(\omega) = \mathbf{H}_F(\omega)\mathbf{S_{uu}}(\omega)\mathbf{H}_F^{\mathsf{H}}(\omega),$$

where $\mathbf{H}_F^{\mathsf{H}} = (\mathbf{H}^*)^{\mathsf{T}}$ denotes the transpose of the complex conjugate of $\mathbf{H}$. As illustrated in Figure 4.9, the Laplace transform can be applied to the initial transient regime, while the spectrum of the response in the subsequent statistical steady-state is described by the above Wiener-Khinchine theorem.

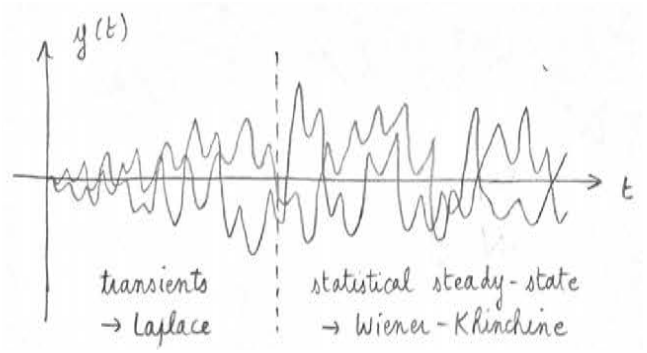


Figure 4.9 – Initial transient regime and subsequent statistical steady-state.

Example. Consider the double spring-damper-mass system pictured in Figure 4.10, and governed by the following equations

$$m_1\ddot{x}_1 - c_1\dot{x}_2 - k_1x_2 = F_1,$$
$$m_2\ddot{x}_1 + m_2\ddot{x}_2 + c_2\dot{x}_1 + (c_1 + c_2)\dot{x}_2 + k_2x_1 + (k_1 + k_2)x_2 = F_2,$$

where F_1 and F_2 are stationary, ergodic, zero-mean and Gaussian stochastic processes. First, we find the transfer function matrix between the input $\mathbf{F} = (F_1, F_2)^{\mathsf{T}}$ and the output $\mathbf{y} = (x_1, x_2)^{\mathsf{T}}$ by assuming that F_1 and F_2 are deterministic functions with well-defined Fourier transforms, and taking the Fourier transform of the governing equations

$$\begin{bmatrix} -\omega^2 m_1 & -i\omega c_1 - k_1 \\ -\omega^2 m_2 + i\omega c_2 + k_2 & -\omega^2 m_2 + i\omega(c_1 + c_2) + (k_1 + k_2) \end{bmatrix} \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{bmatrix} = \begin{bmatrix} \tilde{F}_1 \\ \tilde{F}_2 \end{bmatrix}.$$

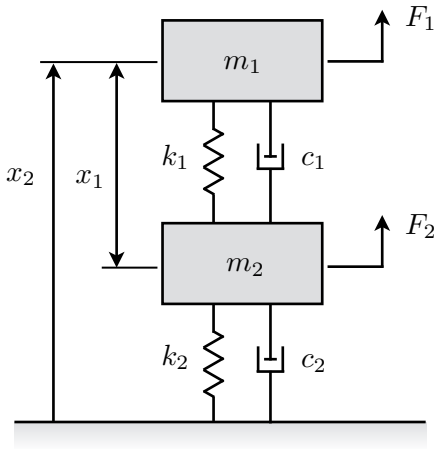


Figure 4.10 – Double spring-damper-mass system.

This can be expressed as

$$\begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{bmatrix} = \mathbf{H}_F(\omega) \begin{bmatrix} \tilde{F}_1 \\ \tilde{F}_2 \end{bmatrix},$$

where the transfer function $\mathbf{H}_F(\omega)$ is given by

$$\mathbf{H}_F(\omega) = \frac{1}{\Delta(\omega)} \begin{bmatrix} \Delta_{11}(\omega) & \Delta_{12}(\omega) \\ \Delta_{21}(\omega) & \Delta_{22}(\omega) \end{bmatrix},$$

with

$$\begin{aligned}
\Delta(\omega) &= \omega^4 m_1 m_2 - i\omega^3((m_1+m_2)c_1 + m_1 c_2) \\
&\quad - \omega^2((m_1+m_2)k_1 + m_1 k_2 + c_1 c_2) + i\omega c_2 k_1 + k_1 k_2, \\
\Delta_{11}(\omega) &= -\omega^2 m_2 + i\omega(c_1+c_2) + k_1 + k_2, \\
\Delta_{22}(\omega) &= -\omega^2 m_1, \\
\Delta_{12}(\omega) &= \omega^2 m_2 - i\omega c_2 - k_2, \\
\Delta_{21}(\omega) &= i\omega c_1 + k_1.
\end{aligned}$$

Now, given that F_1 and F_2 are stochastic processes with spectrum

$$\mathbf{S_{FF}}(\omega) = \begin{bmatrix} S_{F_1F_1}(\omega) & 0 \\ 0 & S_{F_2F_2}(\omega) \end{bmatrix},$$

and applying the Wiener-Khinchine relations, we obtain the spectrum of the output

$$\mathbf{S_{FF}}(\omega) = \begin{bmatrix} S_{x_1x_1}(\omega) & S_{x_1x_2}(\omega) \\ S_{x_2x_1}(\omega) & S_{x_2x_2}(\omega) \end{bmatrix} = \mathbf{H}_F(\omega)\mathbf{S_{FF}}(\omega)\mathbf{H}_F^{\mathsf{H}}(\omega).$$

Chapter 5

Nonlinear Systems

5.1 Deterministic analysis

5.1.1 Linearization

Consider the case of a general 2 degrees-of-freedom (DOF) first-order system,

$$\begin{aligned}\dot{x} &= f(x, y),\\ \dot{y} &= g(x, y).\end{aligned}$$

Example. Suppose we have a nonlinear oscillator governed by

$$\ddot{x} + h(x, \dot{x}) = 0.$$

Setting $y = \dot{x}$, we obtain the 2DOF first-order system

$$\begin{aligned}\dot{x} &= y,\\ \dot{y} &= -h(x, y).\end{aligned}$$

Going back to the general case, we start by looking for fixed points x^*, y^* of the system. Such fixed points are defined by

$$\begin{aligned} f(x^*, y^*) = 0 \quad &\Rightarrow \quad \dot{x}|_{x^*,y^*} = 0,\\ g(x^*, y^*) = 0 \quad &\Rightarrow \quad \dot{y}|_{x^*,y^*} = 0,\end{aligned}$$

that is, the system stays at the fixed point. Assuming that we find such pair(s) x^*, y^*, we can apply the transformation

$$\begin{aligned} u = x - x^* \quad &\Rightarrow \quad x = u + x^*, \dot{x} = \dot{u},\\ v = y - y^* \quad &\Rightarrow \quad y = v + y^*, \dot{y} = \dot{v}.\end{aligned}$$

Thus, u and v are variables that measure the distance of the system from the fixed point. Assuming that $|u|, |v| \ll 1$, we can Taylor expand the dynamics

$$\begin{aligned}\dot{u} &= f(x^* + u, y^* + v) = f(x^*, y^*) + \left.\frac{\partial f}{\partial x}\right|_{x^*,y^*} u + \left.\frac{\partial f}{\partial y}\right|_{x^*,y^*} v + \mathcal{O}(|u|^2, |v|^2),\\ \dot{v} &= g(x^* + u, y^* + v) = g(x^*, y^*) + \left.\frac{\partial g}{\partial x}\right|_{x^*,y^*} u + \left.\frac{\partial g}{\partial y}\right|_{x^*,y^*} v + \mathcal{O}(|u|^2, |v|^2).\end{aligned}$$

Since, by definition, $f(x^*, y^*) = g(x^*, y^*) = 0$, and neglecting the quadratic term, we obtain a linearized system governing the dynamics close to the fixed point x^*, y^*

$$\dot{u} \simeq \left.\frac{\partial f}{\partial x}\right|_{x^*,y^*} u + \left.\frac{\partial f}{\partial y}\right|_{x^*,y^*} v,$$
$$\dot{v} \simeq \left.\frac{\partial g}{\partial x}\right|_{x^*,y^*} u + \left.\frac{\partial g}{\partial y}\right|_{x^*,y^*} v.$$

5.1.2 Classification of linear systems

Let us first consider a few example of 2DOF linear systems, which can in general be written as

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

Example. Consider the linear oscillator

$$m\ddot{x} + kx = 0,$$
$$y = \dot{x},$$

which is equivalent with the first-order system

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -k/m & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

The solution is readily obtained as

$$x = A\cos(\omega t + \phi) \quad \Rightarrow \quad x^2 = A^2\cos^2(\omega t + \phi),$$
$$y = \dot{x} = -A\omega\sin(\omega t + \phi) \quad \Rightarrow \quad y^2 = A^2\omega^2\sin^2(\omega t + \phi),$$

hence, trajectories in the phase space (x, y) describe an ellipse,

$$x^2 + \frac{y^2}{\omega^2} = A^2,$$

as illustrated Figure 5.1.

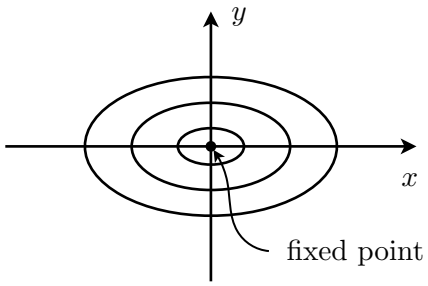


Figure 5.1 – Phase-space trajectories of the linear oscillator.

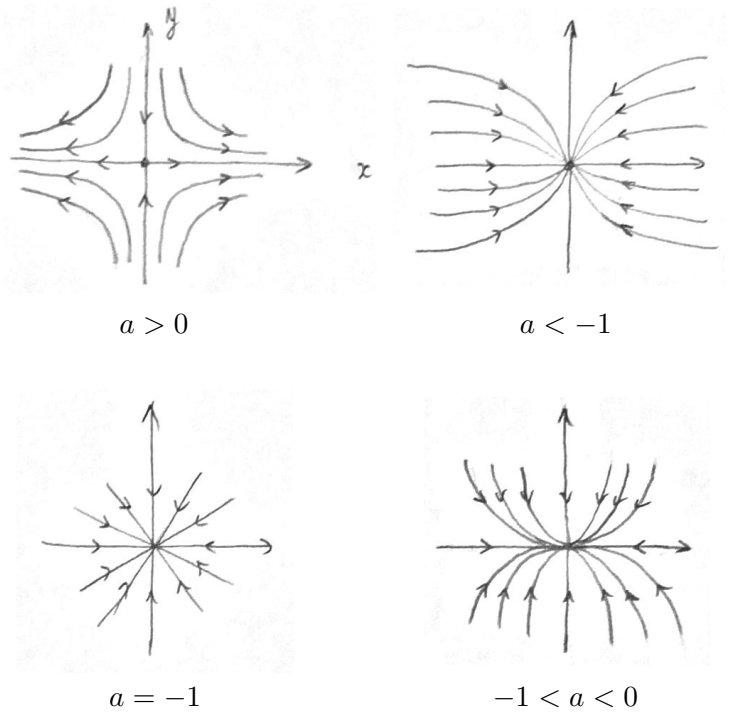
Example. Consider the diagonal system

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} a & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix},$$

which admits as solutions

$$\begin{aligned} x &= Ae^{at}, \\ y &= Be^{-t}. \end{aligned}$$

The trajectories in the phase space (x, y) of the system depend on the value of a, and the different possible cases are illustrated in Figure 5.2.



$a > 0$ $\qquad$ $a < -1$

$a = -1$ $\qquad$ $-1 < a < 0$

Figure 5.2 – Phase-space trajectories of a diagonal system.

In the general case

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix},$$

note that $x, y = 0$ is always a fixed point. Since the system is linear, we seek solutions of the form

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} e^{\lambda t} \quad \Rightarrow \quad \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \lambda e^{\lambda t} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} e^{\lambda t},$$

which leads to an eigenvalue problem $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$, where

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad \text{and} \quad \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}.$$

We thus want to find nontrivial solutions to $(\mathbf{A}-\lambda\mathbf{I})\mathbf{v}=0$, that is, solutions such that $\mathbf{v}\neq 0$. This requires $\mathbf{A}-\lambda\mathbf{I}$ to be non-invertible. We thus impose the condition $\det(\mathbf{A}-\lambda\mathbf{I})=0$, that is,

$$\det\begin{bmatrix} a-\lambda & b \\ c & d-\lambda \end{bmatrix}=0 \quad \Rightarrow \quad \lambda^2-\tau\lambda+\Delta=0 \quad \Rightarrow \quad \lambda_{1,2}=\frac{\tau\pm\sqrt{\tau^2-4\Delta}}{2},$$

where $\tau=a+d$ and $\Delta=ad-bc$. The solution can thus be written as the superposition

$$\begin{bmatrix} x \\ y \end{bmatrix}=\alpha_1\mathbf{v}_1e^{\lambda_1 t}+\alpha_2\mathbf{v}_2e^{\lambda_2 t},$$

where $\mathbf{v}_{1,2}$ are the eigenvectors associated with the eigenvalues $\lambda_{1,2}$, obtained by solving the linear system $(\mathbf{A}-\lambda_{1,2}\mathbf{I})\mathbf{v}_{1,2}=0$ once the eigenvalues are found, and $\alpha_{1,2}$ are constants that are found from the initial conditions.

Example. Consider the system

$$\begin{aligned}\dot{x}&=x+y,\\ \dot{y}&=4x-2y.\end{aligned}$$

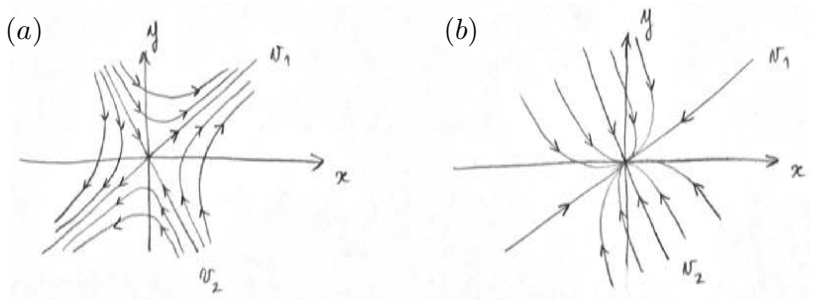
Following the procedure outlined above, we find the eigenvalues

$$\lambda^2+\lambda-6=0 \quad \Rightarrow \quad \lambda_1=2, \quad \lambda_2=-3,$$

and the associated eigenvectors

$$\mathbf{v}_1=\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2=\begin{bmatrix} 1 \\ -4 \end{bmatrix}.$$

In Figure 5.3(a), we sketch the behavior of the system in the state-space (x,y), which can be inferred from the eigenvalues and eigenvectors. The positive eigenvalue λ_1 indicates repulsion away from the fixed point along the eigendirection $\mathbf{v}_1$, while the negative eigenvalue λ_2 indicates attraction to the fixed point along the eigendirection $\mathbf{v}_2$. Should the eigenvalues be both negative, we would obtain the behavior sketched in Figure 5.3(b).



$\lambda_1>0,\lambda_2<0$ $\qquad$ $\lambda_1<0,\lambda_2<0$

Figure 5.3 – Phase-space trajectories of a system with (a) real eigenvalues of opposite sign, and (b) negative real eigenvalues.

Going back to the general case, recall that the eigenvalues are given by

$$\lambda_{1,2} = \frac{\tau \pm \sqrt{\tau^2 - 4\Delta}}{2},$$

where $\tau = a + d$ and $\Delta = ad - bc$. Depending on the signs of τ and Δ, we can distinguish different possible scenario which are illustrated in Figure 5.4:

1. If $\Delta < 0$, then
$$\lambda_{1,2} = \frac{\tau}{2} \pm \frac{1}{2}\sqrt{\tau^2 - 4\Delta} \quad \text{with} \quad \lambda_1 > 0, \lambda_2 < 0,$$
hence the fixed point is an unstable saddle.

2. If $\Delta > 0$, we further need to distinguish between three cases:
   (a) if $\tau^2 - 4\Delta = 0$, then
   $$\lambda_{1,2} = \frac{\tau}{2},$$
   hence the fixed point is an unstable node for $\tau > 0$, and a stable node for $\tau < 0$.
   (b) if $\tau^2 - 4\Delta > 0$, then
   $$\lambda_{1,2} = \frac{\tau}{2} \pm \frac{1}{2}\sqrt{\tau^2 - 4\Delta} \quad \text{with} \quad \text{sign}(\lambda_{1,2}) = \text{sign}(\tau),$$
   hence the fixed point is an unstable node for $\tau > 0$, and a stable node for $\tau < 0$.
   (c) if $\tau^2 - 4\Delta < 0$, then
   $$\lambda_{1,2} = \frac{\tau}{2} \pm \frac{1}{2} i\sqrt{|\tau^2 - 4\Delta|},$$
   hence the fixed point is an unstable spiral for $\tau > 0$, and a stable spiral for $\tau < 0$.

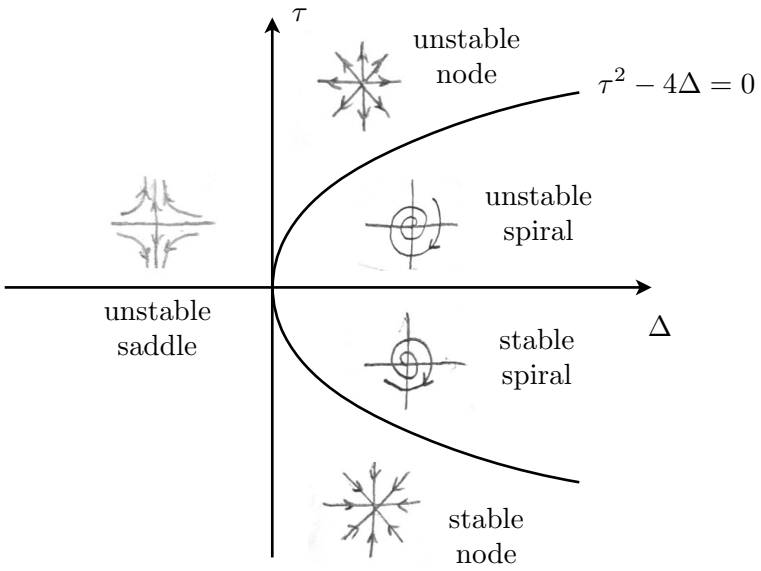


Figure 5.4 – Stability landscape for 2D linearized systems.

Example. Consider the system

$$\dot{x} = -x + x^3,$$
$$\dot{y} = -2y.$$

We first look for fixed points

$$-x^* + x^{*3} = 0 \quad \Rightarrow \quad x^* = 0, \pm 1,$$
$$-2y^* = 0 \quad \Rightarrow \quad y^* = 0.$$

Thus we have the fixed points $(x^*, y^*) = (0,0), (x^*, y^*) = (1,0), (x^*, y^*) = (-1,0)$. Linearizing the system around the fixed point $(x^*, y^*) = (0,0)$ leads to the matrix

$$\mathbf{A} = \begin{bmatrix} -1 & 0 \\ 0 & -2 \end{bmatrix},$$

thus this fixed point is stable. Linearizing the system around the fixed points $(x^*, y^*) = (\pm 1, 0)$ leads to the matrix

$$\mathbf{A} = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix},$$

thus these fixed points are unstable saddles. Since we cannot have chaotic dynamics between the fixed points for a 2D system, the knowledge of these 3 fixed points and their stability behavior enables us to draw in Figure 5.5 the behavior of the system in phase space.

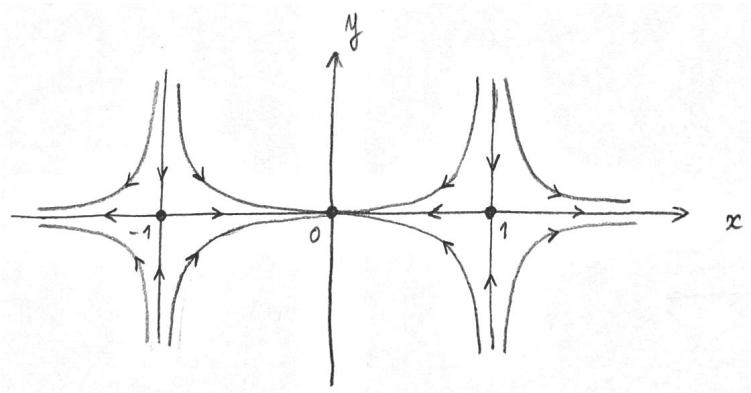


Figure 5.5 – Phase portrait of system with 3 fixed points.

5.2 Nonlinear systems with white noise excitation

5.2.1 The Fokker-Planck-Kolmogorov (FPK) equation

Consider the nonlinear system

$$\frac{d\mathbf{Y}}{dt} = \mathbf{F}[\mathbf{Y}, t] + \boldsymbol{\sigma}[\mathbf{Y}, t]\dot{\mathbf{W}}, \qquad \mathbf{Y}(t_0) = \mathbf{Y}_0,$$

where $\mathbf{Y}(t, \zeta) \in \mathbb{R}^N$ is the state, $\mathbf{F} \in \mathbb{R}^N$ is a nonlinear 'drift' term, $\boldsymbol{\sigma} \in \mathbb{R}^{N \times N}$ is a nonlinear 'diffusion' term, and $\dot{\mathbf{W}} \in \mathbb{R}^N$ is white noise. For nonlinear systems, the statistics are non-Gaussian

and hence 2nd-order moments are not sufficient to give a full description of the response. We thus have to use the Fokker-Planck-Kolmogorov (FPK) equation, which transforms the trajectory-based description of the dynamics into an evolution equation for the pdf $f_{\mathbf{Y}}$ of the state. It is formulated as follows

$$\frac{\partial f_{\mathbf{Y}}}{\partial t} + \sum_{i=1}^{N} \frac{\partial}{\partial y_i}[F_i(\mathbf{y},t) f_{\mathbf{Y}}] - \frac{1}{2} \sum_{i,j=1}^{N} \frac{\partial^2}{\partial y_i \partial y_j}[b_{ij}(\mathbf{y},t) f_{\mathbf{Y}}] = 0,$$

where $f_{\mathbf{Y}}(\mathbf{y},t) = f_{\mathbf{Y}}(y_1, y_2, \ldots, y_N, t)$ is the pdf of the stochastic process $\mathbf{Y}(t,\zeta)$ describing the state, and the diffusion coefficients b_{ij} are defined as

$$b_{ij}(\mathbf{y},t) = \sum_{r=1}^{N} \sigma_{ir}(\mathbf{y},t)\sigma_{jr}(\mathbf{y},t), \quad i,j = 1,\ldots,N.$$

The Fokker-Planck-Kolmogorov equation represents the conservation of probability in phase space. It is a linear partial differential equation (PDE); more specifically, if the system is described by an N-dimensional nonlinear ODE, then the FPK is a linear PDE defined in an N-dimensional domain. The FPK equation is hard to solve (numerically) for $N > 4$, since it must always integrate to one and be non-negative. Finally, it is valid for white noise only (similar equations exist for Poisson noise and Levy noise).

5.2.2 Application examples

Example. Consider the 1D nonlinear system

$$\frac{dY}{dt} + h(Y) = D\dot{W}, \qquad Y(0) = Y_0,$$

where the state $Y \in \mathbb{R}$, D is a constant, $h(Y)$ is a nonlinear function and the initial condition Y_0 is deterministic. The associated FPK is expressed as

$$\frac{\partial f_Y(y,t)}{\partial t} = \frac{\partial}{\partial y}[h(y) f_Y(y,t)] + \frac{1}{2} D^2 \frac{\partial^2 f_Y(y,t)}{\partial y^2},$$

with the initial and boundary conditions

$$f_Y(y,0) = \delta(y - y_0), \qquad \lim_{|y|\to\infty} f_Y(y,t) = 0.$$

When solving the FPK numerically, we need to ensure that $f_Y \geq 0$ everywhere, and that

$$\int_{-\infty}^{\infty} f_Y(y,t)dy = 1.$$

As $t \to \infty$, the long-time solution converges to the statistical steady-state

$$\lim_{t\to\infty} f_Y(y,t) = f_{Y,s}(y).$$

To find $f_{Y,s}(y)$, we set the time derivative to zero in the FPK,

$$\frac{\partial}{\partial y}[h(y) f_{Y,s}(y)] + \frac{1}{2} D^2 \frac{\partial^2 f_{Y,s}(y)}{\partial y^2} = 0.$$

Making use of the fact that $f_{Y,s}(y)$ must vanish as $|y| \to \infty$, we can integrate the above equation

$$h(y) f_{Y,s}(y) + \frac{1}{2} D^2 \frac{\partial f_{Y,s}}{\partial y} = 0.$$

Thus, we get the long-time statistics by integrating one more time

$$f_{Y,s}(y) = \frac{1}{C} \exp\left[-\frac{2}{D^2} \int_0^y h(z) dz \right],$$

where C is a normalizing constant such that

$$\int_{-\infty}^{\infty} f_{Y,s}(y) dy = 1 \quad \Rightarrow \quad C = \int_{-\infty}^{\infty} \exp\left[-\frac{2}{D^2} \int_0^y h(z) dz \right] dy.$$

Example. Consider the nonlinear oscillator,

$$\frac{d^2 Y}{dt^2} + \beta \frac{dY}{dt} + h(Y) = 2D\dot{W},$$

where $h(Y)$ is a nonlinear restoring force. Since the FPK is formulated for 1st-order equations, we set $Y = Y_1$, $\dot{Y} = Y_2$ and we rewrite the above 2nd-order equation as a set of coupled 1st-order equations (similar to a nonlinear state-space description)

$$\begin{aligned} \frac{dY_1}{dt} &= Y_2, \\ \frac{dY_2}{dt} &= -\beta Y_2 - h(Y_1) + 2D\dot{W}. \end{aligned}$$

The FPK for the pdf $f_{Y_1 Y_2}(y_1, y_2, t)$ is then expressed as

$$\begin{aligned} \frac{\partial f_{Y_1 Y_2}(y_1, y_2, t)}{\partial t} = &-\frac{\partial}{\partial y_1}[y_2 f_{Y_1 Y_2}(y_1, y_2, t)] + \frac{\partial}{\partial y_2}[(\beta y_2 + h(y_1)) f_{Y_1 Y_2}(y_1, y_2, t)] \\ &+ 2D^2 \frac{\partial^2 f_{Y_1 Y_2}(y_1, y_2, t)}{\partial y_2^2}. \end{aligned}$$

In the long-time limit, the solution converges to the statistical steady-state

$$\lim_{t \to \infty} f_{Y_1 Y_2}(y_1, y_2, t) = f_{Y_1 Y_2, s}(y_1, y_2),$$

which is described by the steady-state FPK equation

$$0 = -\frac{\partial}{\partial y_1}[y_2 f_{Y_1 Y_2, s}(y_1, y_2)] + \frac{\partial}{\partial y_2}[(\beta y_2 + h(y_1)) f_{Y_1 Y_2, s}(y_1, y_2)] + 2D^2 \frac{\partial^2 f_{Y_1 Y_2, s}(y_1, y_2)}{\partial y_2^2}.$$

Integrating the above equation, we obtain the Maxwell Boltzmann distribution,

$$f_{Y_1 Y_2, s}(y_1, y_2) = C \exp\left[-\frac{\beta}{2D^2} \left(\frac{y_2^2}{2} + \int_0^{y_1} h(z) dz \right) \right] = C \exp\left[-\frac{\beta}{2D^2} \mathcal{E}(y_1, y_2) \right],$$

where $\mathcal{E}$ represents the total energy of the system, that is, the sum of the kinetic energy and the potential energy

$$\mathcal{E}(Y, \dot{Y}) = \frac{1}{2} \dot{Y}^2 + \int_0^Y h(z) dz.$$

Note that if the system is linear, that is, $h(Y) = kY$, then $f_{Y_1Y_2,s}$ is Gaussian as expected. Consider instead the case $h(Y) = \alpha Y + kY^3$, which corresponds to the Duffing oscillator,

$$\frac{d^2Y}{dt^2} + \beta\frac{dY}{dt} + \alpha Y + kY^3 = 2D\dot{W},$$

with total energy given by

$$\mathcal{E}(Y,\dot{Y}) = \frac{1}{2}\dot{Y}^2 + \frac{1}{2}\alpha Y^2 + \frac{1}{4}kY^4.$$

In the long-time limit, the steady-state pdf is then

$$f_{Y_1Y_2,s}(y_1,y_2) = C\exp\left[-\frac{\beta}{2D^2}\left(\frac{y_2^2}{2} + \frac{1}{2}\alpha y_1^2 + \frac{1}{4}ky_1^4\right)\right].$$

We now investigate the relationship between the steady-state pdf obtained above and the stability properties of the system without taking the noise into account. Let us first focus on the case $\alpha, k > 0$. As done in the previous section, we first look for fixed points satisfying $\dot{Y}_1 = \dot{Y}_2 = 0$, which implies

$$\alpha Y_1^* + kY_1^{*3} = 0 \quad \text{and} \quad Y_2^* = 0 \quad \Rightarrow \quad Y_1^* = Y_2^* = 0.$$

To find the stability properties of the system in the neighborhood of the fixed point $Y_1^* = Y_2^* = 0$, we linearize the governing equations around Y_1^*, Y_2^*, which leads to

$$\begin{bmatrix} \dot{Y}_1 \\ \dot{Y}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\alpha & -\beta \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \mathbf{A} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}.$$

We have $\operatorname{tr}\mathbf{A} = -\beta < 0$ and $\det\mathbf{A} = \alpha > 0$, hence as we saw in the previous section, the fixed point Y_1^*, Y_2^* is stable and attracts neighboring trajectories. Therefore, as illustrated in Figure 5.6, the shape of the steady-state pdf of the system with noise present is a result of the interplay between attraction to the fixed point and diffusion due to the noise.

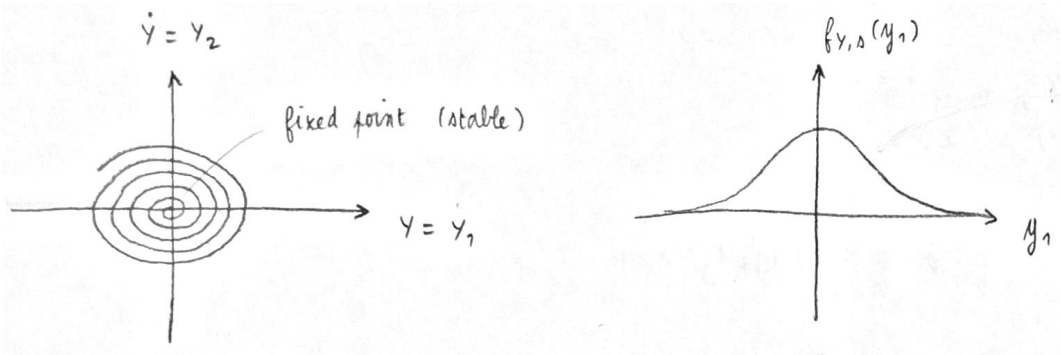


Figure 5.6 – Steady-state pdf of a system with a single stable fixed point.

Let us now investigate the case $\alpha < 0$, $k > 0$. The fixed points satisfy $\dot{Y}_1 = \dot{Y}_2 = 0$, which implies

$$\alpha Y_1^* + kY_1^{*3} = 0 \quad \text{and} \quad Y_2^* = 0 \quad \Rightarrow \quad Y_1^* = 0 \text{ or } \pm\sqrt{-\frac{\alpha}{k}}, \quad Y_2^* = 0.$$

As before, the linearized governing equations around the fixed point $Y_1^* = Y_2^* = 0$ are

$$\begin{bmatrix} \dot{Y}_1 \\ \dot{Y}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\alpha & -\beta \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \mathbf{A} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}.$$

We have $\operatorname{tr}\mathbf{A} = -\beta < 0$ and $\det\mathbf{A} = \alpha < 0$, hence the fixed point $Y_1^* = Y_2^* = 0$ is an unstable saddle. Now, the linearized governing equations around the fixed points $Y_1^* = \pm\sqrt{-\alpha/k}, Y_2^* = 0$ are

$$\begin{bmatrix} \dot{Y}_1 \\ \dot{Y}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 2\alpha & -\beta \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \mathbf{A} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}.$$

We have $\operatorname{tr}\mathbf{A} = -\beta < 0$ and $\det\mathbf{A} = -2\alpha > 0$, hence the fixed points $Y_1^* = \pm\sqrt{-\alpha/k}, Y_2^* = 0$ are stable. Thus, as illustrated in Figure 5.7, the shape of the steady-state pdf of the system with noise present is a result of the interplay between attraction to the stable fixed points at $Y_1^* = \pm\sqrt{-\alpha/k}$, attraction and repulsion away from the unstable saddle at $Y_1^* = 0$, and diffusion due to the noise.

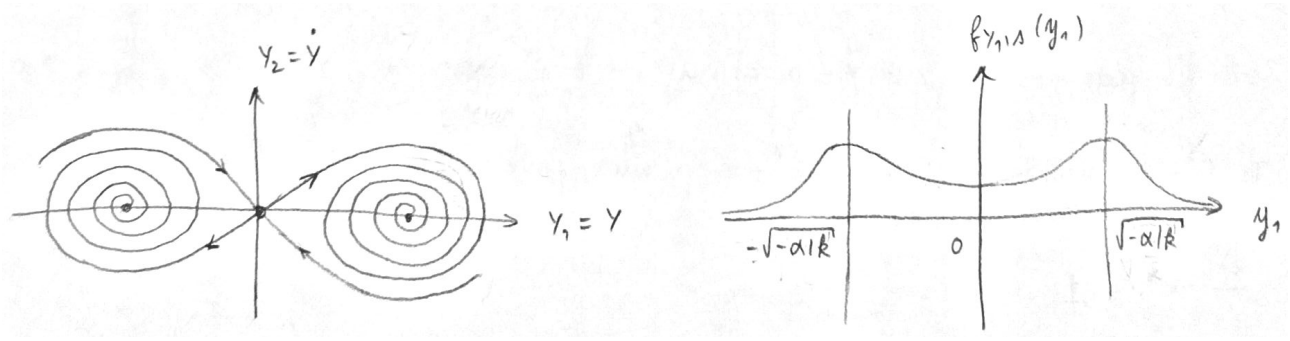


Figure 5.7 – Steady-state pdf of a system with two stable fixed points and an unstable saddle.

One can also look for the most probable state (Y_1, Y_2), which is found by setting

$$\begin{aligned} \nabla f_{Y_1Y_2,s} = 0 \quad &\Rightarrow \quad \frac{\partial f_{Y_1Y_2,s}}{\partial y_1} = 0 \quad \text{and} \quad \frac{\partial f_{Y_1Y_2,s}}{\partial y_2} = 0 \\ &\Rightarrow \quad \alpha y_1 + k y_1^3 = 0 \quad \text{and} \quad y_2 = 0, \end{aligned}$$

which coincides with the fixed points of the system. Evaluating the pdf at these local extrema then allows us to find the global maximum, which in this case will be located at the two stable fixed points $Y_1^* = \pm\sqrt{-\alpha/k}, Y_2^* = 0$.

Example. Consider the nonlinear oscillator,

$$\frac{d^2Y}{dt^2} + \beta\frac{dY}{dt} + h(Y) = \sigma\dot{W},$$

where $h(Y)$ is a nonlinear restoring force given by

$$h(Y) = \begin{cases} -ky - 2ky_0, & y \le -y_0, \\ ky, & -y_0 < y \le y_0, \\ -ky + 2ky_0, & y > y_0. \end{cases}$$

The restoring force $h(Y)$ is illustrated in Figure 5.8, and models the restoring force for a ship undergoing rolling motion due to waves.

First, we find the equilibrium points of the system by setting $\dot{Y} = \ddot{Y} = 0$ and neglecting the noise. This leads to

$$h(Y^*) = 0 \quad \Rightarrow \quad Y^* = 0 \text{ or } Y^* = \pm 2y_0.$$

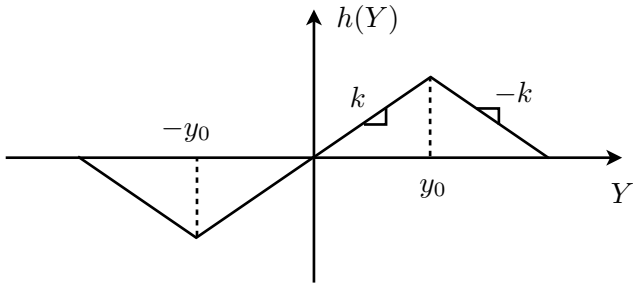
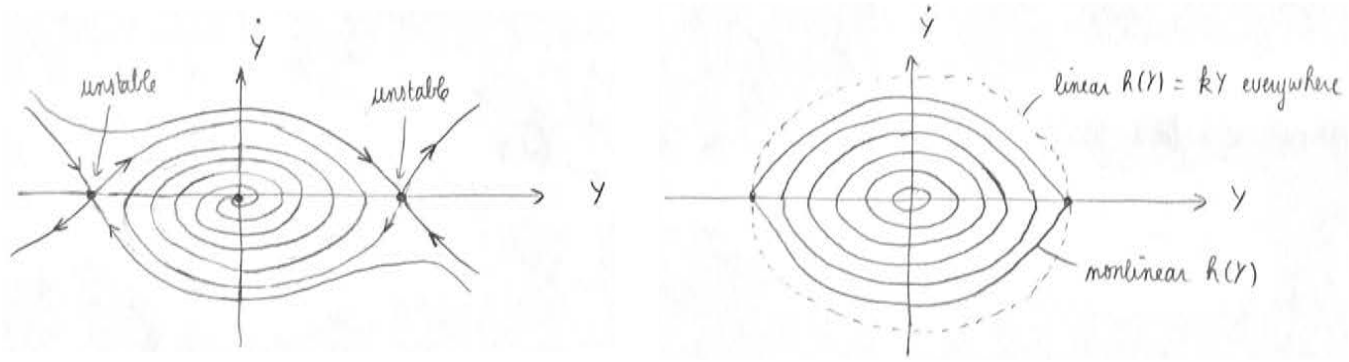


Figure 5.8 – Restoring force for the rolling motion of a ship.



$\beta > 0$ $\beta = 0$

Figure 5.9 – Phase portrait for rolling motion of a ship.

Linearizing the system around each fixed point, we find that the fixed points $Y^* = \pm 2y_0$ are unstable saddles. On the other hand, the fixed point $Y^* = 0$ is stable if the damping $\beta > 0$, and marginally stable if $\beta = 0$. The resulting phase portraits are illustrated in Figure 5.9 for the two cases $\beta > 0$ and $\beta = 0$.

To find the pdf of the statistical steady-state, we note that

$$\int_0^y h(z)dz = \begin{cases} -\frac{1}{2}ky^2 - 2ky_0y - ky_0^2, & y \le -y_0, \\ \frac{1}{2}ky^2, & -y_0 < y \le y_0, \\ \frac{1}{2}ky^2 - 2ky_0y - ky_0^2, & y > y_0. \end{cases}$$

hence we will obtain the steady-state pdf pictured in Figure 5.10. Note that the non-Gaussian character of the pdf is directly caused by the nonlinear restoring force (softening nonlinearity), inducing heavy tails to the pdf.

5.2.3 Extended phase space

Suppose that we are given a stochastic process $Y(t, \zeta)$ with given spectrum $\tilde{S}_{YY}(\omega)$, and we would like to determine the pdf of the system

$$\frac{dX}{dt} = h(X) + Y.$$

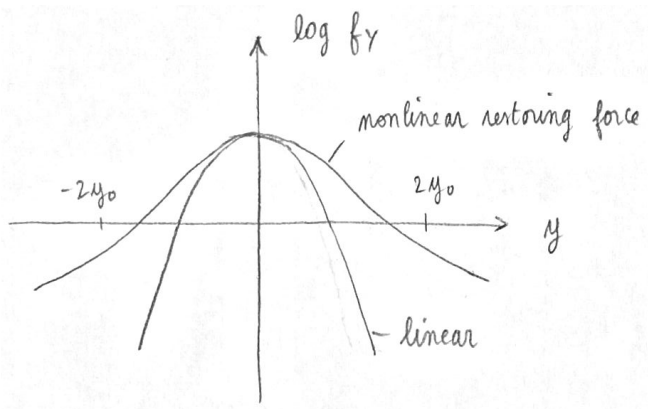


Figure 5.10 – Steady-state pdf for the rolling motion of a ship.

Since Y is not white noise, we cannot apply FPK directly. However, we can represent $Y(t,\zeta)$ as the response of a linear oscillator excited by white noise,

$$\frac{d^2Y}{dt^2} + \beta\frac{dY}{dt} + kY = \sigma\dot{W}.$$

Because of linearity of the oscillator, we can apply the Wiener-Khinchine relations

$$S_{YY}(\omega) = |H(\omega)|^2 S_{\dot{W}\dot{W}}(\omega) = \frac{1}{(k-\omega^2)^2 + \beta^2\omega^2}\sigma^2.$$

Thus, we can pick σ, k, β so that $S_{YY}(\omega)$ is as close as possible to the given spectrum $\tilde{S}_{YY}(\omega)$. The coupled system which results is then simply excited by white noise,

$$\begin{aligned}
\frac{dX}{dt} &= h(X) + Y,\\
\frac{dY}{dt} &= Z,\\
\frac{dZ}{dt} &= -\beta Z - kY + \sigma\dot{W}.
\end{aligned}$$

We can therefore formulate the FPK for the pdf $f_{XYZ}(x,y,z,t)$ as follows

$$\frac{\partial f_{XYZ}}{\partial t} + \frac{\partial}{\partial x}[(h(x)+y)f_{XYZ}] + \frac{\partial}{\partial y}[zf_{XYZ}] + \frac{\partial}{\partial z}[(-\beta z - ky)f_{XYZ}] - \frac{1}{2}\sigma^2\frac{\partial^2 f_{XYZ}}{\partial z^2} = 0.$$

5.3 Statistical linearization

We saw in the previous section that if the input spectrum is different from white noise, we need to augment the dimension of the phase space by adding additional equations to model the input. For instance,

$$\dot{X} = aX + \dot{W},$$

where $\dot{W}$ is white noise, which gives the spectrum

$$S_{XX}(\omega) = \frac{1}{a^2+\omega^2}.$$

For more complicated input spectrum, we would need to use higher-order equations, such as

$$\dddot{X} + K_2\ddot{X} + K_1\dot{X} + K_0 = \dot{W},$$

hence the dimensionality of the resulting augmented system can grow quickly, which becomes problematic when it comes to solving the FPK. Obviously, high dimensionality of the actual system would also make the PFK problematic. On the other hand, if the original nonlinear system is close enough to a linear system, could we find a linear system that mimics the behavior of the nonlinear system? Statistical linearization, which we will see in this section, tries to provide an answer. Note that this is a different tool from the usual linearization performed with Taylor expansions. Consider the nonlinear oscillator,

$$\ddot{Y} + f(Y, \dot{Y}) = X,$$

where f is a nonlinear function, and X is the excitation with given spectrum. We are trying to find a linear system

$$\ddot{Y} + \beta_e\dot{Y} + k_eY = X,$$

that is close to the original one, with β_e and k_e arbitrary coefficients to be found. To that effect, we define the difference

$$\Delta = \beta_e\dot{Y} + k_eY - f(Y, \dot{Y}).$$

We then look for β_e, k_e such that Δ is small. But since Δ is stochastic, the relevant quantity to minimize is the variance $E[\Delta^2]$. Hence, we want to find β_e, k_e such that

$$E[\Delta^2] = E[(\beta_e\dot{Y} + k_eY - f(Y, \dot{Y})^2]$$

is minimum. The associated extrema conditions are

$$\frac{\partial E[\Delta^2]}{\partial \beta_e} = 0, \quad \text{and} \quad \frac{\partial E[\Delta^2]}{\partial k_e} = 0.$$

Expanding

$$E[\Delta^2] = E[\beta_e^2\dot{Y}^2 + k_e^2Y^2 + f(Y, \dot{Y})^2 + 2\beta_e k_e Y\dot{Y} - 2\beta_e f(Y, \dot{Y})\dot{Y} - 2k_eYf(Y, \dot{Y})],$$

the two extrema conditions are thus

$$\begin{aligned}\beta_e E[\dot{Y}^2] + k_e E[Y\dot{Y}] - E[\dot{Y}f(Y, \dot{Y})] = 0,\\ k_e E[Y^2] + \beta_e E[Y\dot{Y}] - E[Yf(Y, \dot{Y})] = 0.\end{aligned}$$

This is a system of 2 equations for 2 unknowns k_e and β_e. We obtain

$$\begin{aligned}\beta_e &= \frac{E[Y^2]E[\dot{Y}f(Y, \dot{Y})] - E[Y\dot{Y}]E[Yf(Y, \dot{Y})]}{E[Y^2]E[\dot{Y}^2] - E[Y\dot{Y}]^2},\\ k_e &= \frac{E[\dot{Y}^2]E[Yf(Y, \dot{Y})] - E[Y\dot{Y}]E[\dot{Y}f(Y, \dot{Y})]}{E[Y^2]E[\dot{Y}^2] - E[Y\dot{Y}]^2}.\end{aligned}$$

The problem is that we don't know the moments of Y and $\dot{Y}$! We thus need to make the following assumptions on the form of the solution to compute the above coefficients:

1. Y and $\dot{Y}$ are Gaussian,

2. Y and $\dot{Y}$ are independent (remember that this was the case in the FPK examples), hence $E[Y\dot{Y}] = 0$,

3. X is stationary and Gaussian.

Setting $E[Y\dot{Y}] = 0$ gives

$$\beta_e = \frac{E[\dot{Y} f(Y, \dot{Y})]}{E[\dot{Y}^2]},$$
$$k_e = \frac{E[Y f(Y, \dot{Y})]}{E[Y^2]}.$$

Let us now consider a specific nonlinear term,

$$f(Y, \dot{Y}) = 2h\dot{Y} + \omega_0^2 Y + \epsilon Y^3.$$

Then, we get

$$\beta_e = \frac{E[2h\dot{Y}^2 + \omega_0^2 Y\dot{Y} + \epsilon Y^3\dot{Y}]}{E[\dot{Y}^2]} = \frac{2hE[\dot{Y}^2]}{E[\dot{Y}^2]} = 2h,$$
$$k_e = \frac{E[2hY\dot{Y} + \omega_0^2 Y^2 + \epsilon Y^4]}{E[Y^2]} = \frac{\omega_0^2 E[Y^2] + \epsilon E[Y^4]}{E[Y^2]} = \omega_0^2 + \epsilon\frac{E[Y^4]}{E[Y^2]}.$$

Note that it is expected that $\beta_e = 2h$, since f is linear in $\dot{Y}$ with coefficient $2h$. We now make use of the assumption that Y is Gaussian. For a Gaussian random variable, we indeed know that

$$E[Y^4] = 3E[Y^2]^2,$$

which is a specific case of a more general property of Gaussian random variables,

$$E[Y^p] = \begin{cases} 0, & p \text{ odd}, \\ \sigma_Y^p (p-1)!!, & p \text{ even}, \end{cases}$$

where $n!! = n(n-2)(n-4)\dots 2$. Thus, we finally have

$$k_e = \omega_0^2 + 3\epsilon E[Y^2], \quad \beta_e = 2h,$$

where, even if unknown, $E[Y^2] = \sigma_Y^2$ is a constant as $t \to \infty$. To find σ_Y^2, we note that the above parameters give the following linear system, which is closest to the original nonlinear one

$$\ddot{Y} + \beta_e \dot{Y} + k_e Y = X,$$

with transfer function

$$H(\omega) = \frac{1}{k_e - \omega^2 + i\omega\beta_e} \quad \Rightarrow \quad |H(\omega)|^2 = \frac{1}{(k_e - \omega^2)^2 + \omega^2\beta_e^2}.$$

By the Wiener-Khinchine relations,

$$S_{YY}(\omega) = \frac{S_{XX}(\omega)}{(k_e - \omega^2)^2 + \omega^2\beta_e^2} = \frac{S_{XX}(\omega)}{(\omega_0^2 + 3\epsilon\sigma_Y^2 - \omega^2)^2 + 4h^2\omega^2},$$

hence the variance σ_Y^2 is equal to

$$\sigma_Y^2 = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{YY}(\omega)d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{S_{XX}(\omega)}{(\omega_0^2 + 3\epsilon\sigma_Y^2 - \omega^2)^2 + 4h^2\omega^2}d\omega.$$

This is a nonlinear equation for σ_Y^2 that can be solved numerically to obtain σ_Y^2! The coefficients k_e and β_e of the linear system are thus entirely specified. In the special case where $S_{XX}(\omega) = c_0 =$ const, that is X is white noise, we obtain

$$\sigma_Y^2 = \frac{\pi c_0}{2h(\omega_0^2 + 3\epsilon\sigma_Y^2)} \quad \Rightarrow \quad \sigma_Y^2 = \frac{\sqrt{\omega_0^4 + 6\pi c_0\epsilon/h} - \omega_0^2}{6\epsilon}.$$

5.4 Moments equations

5.4.1 Ito's formula

Let a stochastic process $X(t,\zeta)$ governed by

$$\frac{dX}{dt} = a(t) + b(t)\dot{W}(t),$$

where $X(t) \in \mathbb{R}$, and $\dot{W}(t)$ is white noise. The differential of a function of $X(t)$, $f(t, X(t))$, is provided by Ito's formula

$$df(t,X(t)) = \left[\frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial x}adt + \frac{\partial f}{\partial x}bdW\right] + \frac{1}{2}b^2\frac{\partial^2 f}{\partial x^2}dt,$$

where the last diffusion term is a correction due to white noise. To generalize Ito's formula to higher dimensions, consider the finite-dimensional system

$$\frac{d\mathbf{X}}{dt} = \mathbf{A}(t) + \mathbf{B}(t)\dot{\mathbf{W}}(t),$$

where $\mathbf{X} \in \mathbb{R}^m$, $\mathbf{A} \in \mathbb{R}^m$, $\mathbf{B} \in \mathbb{R}^{m\times m}$, and $\dot{\mathbf{W}} \in \mathbb{R}^m$ is white noise. For a function $\mathbf{f}(t, \mathbf{X}(t)) = \mathbf{Y}(t)$, Ito's formula becomes

$$\frac{d\mathbf{Y}}{dt} = \left[\frac{\partial \mathbf{f}}{\partial t} + \sum_{i=1}^{m}\frac{\partial \mathbf{f}}{\partial x_i}A_i + \sum_{i=1}^{m}\sum_{j=1}^{m}\frac{\partial \mathbf{f}}{\partial x_i}B_{ij}\frac{d\mathbf{W}_j}{dt}\right] + \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m}\frac{\partial^2 \mathbf{f}}{\partial x_i \partial x_j}B_{ik}B_{jk}.$$

Example. Consider the system

$$\frac{dX_1}{dt} = a_1 + b_1\dot{W}, \quad \frac{dX_2}{dt} = a_2 + b_2\dot{W}.$$

The differential of the function $f(t, \mathbf{X}) = f(X_1, X_2) = X_1X_2$ is

$$\frac{d(X_1X_2)}{dt} = [X_1a_1 + X_2a_1 + b_1b_2] + [X_1b_2 + X_2b_1]\dot{W}.$$

Example. Setting $a(t) = 0$ and $b(t) = 0$, we trivially get that $X(t) = W(t)$, where $W(t)$ is a Wiener process. Then, the differential of a function $f(X) = f(W)$ of the Wiener process follows as

$$\frac{df(W)}{dt} = \frac{1}{2}\frac{\partial^2 f(W)}{\partial x^2} + \frac{\partial f(W)}{\partial x}dW.$$

Letting $f(W) = W^n$, $n = 2, 3, \ldots$, we have

$$\frac{d(W^n)}{dt} = nW^{n-1}dW + \frac{1}{2}n(n-1)W^{n-2}.$$

5.4.2 Moments equations and closure schemes

Let the system

$$\frac{d\mathbf{Y}}{dt} = \mathbf{A}(\mathbf{Y}) + \mathbf{B}(\mathbf{Y})\dot{\mathbf{W}},$$

where $\mathbf{Y}$ and $\mathbf{A}$ are vectors, $\mathbf{B}$ is a matrix, and $\dot{\mathbf{W}}$ is a white noise vector. We define

$$h(\mathbf{Y}) = Y_1^{k_1}(t)Y_2^{k_2}(t)\dots Y_n^{k_n}(t),$$

where $k_1, k_2, \dots, k_n$ are positive integers. The family of moments are then given by

$$m_{k_1 k_2 \dots k_n}(t) = E[Y_1^{k_1} Y_2^{k_2} \dots Y_n^{k_n}] = E[h(\mathbf{Y})].$$

Let us now focus on the following 1D system,

$$\frac{dY}{dt} = -[Y + \mu Y^3] + \dot{W}.$$

In this case, $h(Y) = Y^k$ and $m_k = E[Y^k]$. Applying Ito's formula then taking the expectation,

$$\frac{dm_k}{dt} = -km_k - k\mu m_{k+2} + \frac{1}{2}k(k-1)m_{k-2}, \quad k = 1, 2, \dots$$

For $k = 1$, this becomes

$$\frac{dm_1}{dt} = -m_1 - \mu m_3,$$

while for $k = 2$, we have

$$\frac{dm_2}{dt} = -2m_2 - 2\mu m_4 + 1.$$

This illustrate the closure problem: moments equations always involve higher-order moments, and hence the system can never be closed exactly. To circumvent this issue, suitable approximations can be made, resulting in different closure schemes:

1. The easiest option is to set higher-order moments to zero. The results are rather inaccurate, and often lead to negative 2nd order moments. In the above example, this is equivalent with linearizing the original system (but not true in general).

2. Gaussian closure is a more accurate option. Assuming a Gaussian pdf for the response, one can express higher-order moments in terms of lower-order moments.

3. A third closure scheme is the cumulant neglect hypothesis. Given a pdf g, we denote the Fourier transform

   $$\phi(\lambda) = \mathcal{F}[g].$$

   Then, the cumulants are defined as

   $$k_m = \frac{1}{i^m}\frac{d^m}{d\lambda^m}\ln\phi(\lambda)\bigg|_{\lambda=0}.$$

   The cumulants are connected with the moments in the following way

   $$\begin{aligned} m_1 &= k_1, \\ m_2 &= k_2 + k_1^2, \\ m_3 &= k_3 + 3k_1k_2 + k_1^3, \end{aligned}$$

   and they measure the distance from a Gaussian distribution. Indeed, for a Gaussian distribution we would have $k_3 = k_4 = \dots = 0$. We may now close the moments equations by neglecting the cumulants.

Chapter 6

Bayesian Regression

Suppose we have a training set $\mathcal{D}$ of n observations

$$\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^n,$$

where $\mathbf{x}$ is the input vector and y is a (scalar) output. Our goal is to infer an input–output relationship in the form $y = f(\mathbf{x})$ from our dataset $\mathcal{D}$. In other words, we seek the conditional distribution of the outputs given the inputs.

6.1 The standard linear model

In the standard linear model, we assume a *linear* relationship between inputs and outputs:

$$f(\mathbf{x}) = \mathbf{x}^\mathsf{T}\mathbf{w}, \quad y = f(\mathbf{x}) + \varepsilon,$$

where $\mathbf{w}$ is the vector of parameters and ε is a noise term that reflects the fact that our observations may differ from the function values. We assume that the noise is Gaussian, that is,

$$\varepsilon \sim \mathcal{N}(0, \sigma_n^2).$$

We first define the **likelihood function** as the probability density of the observations given the parameters, which we write as

$$p(\mathbf{y}|X, \mathbf{w}) = \prod_{i=1}^n p(y_i|\mathbf{x}_i, \mathbf{w}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left[-\frac{(y_i - \mathbf{x}_i^\mathsf{T}\mathbf{w})^2}{2\sigma_n^2}\right] = \mathcal{N}(X^\mathsf{T}\mathbf{w}, \sigma_n^2 I).$$

In deriving the above, we have made use of the independence assumption between input–output pairs. We may also have some beliefs about the parameters before we look at the observations. This is specified in a **prior** over $\mathbf{w}$. For simplicity, we assume a zero-mean Gaussian prior with covariance matrix Σ_p:

$$\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \Sigma_p).$$

The **posterior** is the probability density of the parameters given the observations we have made and the prior. Using Bayes rule, we have

$$p(\mathbf{w}|X, \mathbf{y}) = \frac{p(\mathbf{y}|X, \mathbf{w})p(\mathbf{w})}{p(\mathbf{y}|X)} = \frac{\text{likelihood} \times \text{prior}}{\text{marginal likelihood}},$$

where the marginal likelihood is given by

$$p(\mathbf{y}|X) = \int p(\mathbf{y}|X, \mathbf{w}) p(\mathbf{w}) \, \mathrm{d}\mathbf{w}.$$

We arrive at

$$p(\mathbf{w}|X, \mathbf{y}) \propto \exp\left[-\frac{1}{2\sigma_n^2}(\mathbf{y} - X^\mathsf{T}\mathbf{w})^\mathsf{T}(\mathbf{y} - X^\mathsf{T}\mathbf{w})\right] \exp\left[-\frac{1}{2}\mathbf{w}^\mathsf{T}\Sigma_p^{-1}\mathbf{w}\right] = \mathcal{N}\left(\frac{A^{-1}X\mathbf{y}}{\sigma_n^2}, A^{-1}\right),$$

where $A = \sigma_n^{-2} X X^\mathsf{T} + \Sigma_p^{-1}$. Note the interplay between data and prior knowledge in the variance A. Also note that for Gaussian input, the output is generally non-Gaussian.

To make prediction for $\mathbf{x}_*$ not in the original dataset, we compute the predictive distribution for y_* at $\mathbf{x}_*$:

$$y_*|\mathbf{x}_*, X, \mathbf{y} \sim \mathcal{N}\left(\frac{\mathbf{x}_*^\mathsf{T} A^{-1} X \mathbf{y}}{\sigma_n^2}, \mathbf{x}_*^\mathsf{T} A^{-1} \mathbf{x}_*\right).$$

The predictive distribution is Gaussian, with a mean depending linearly on $\mathbf{x}_*$ and a variance depending quadratically on $\mathbf{x}_*$. This means that predictive uncertainties grow with the magnitude of the input.

Figure 6.1 shows an example of Bayesian linear regression for a simple two dimensional problem. Note that while input x is one dimensional we include an offset term in the regression by augmenting each measurement with a constant so that $\mathbf{x} = (x, 1)$. With few observations, the variance of the likelihood function is large, allowing the prior to have a dominant effect. As more samples are collected the likelihood becomes sharper and the prior plays a less significant role.

6.2 Nonlinear regression (projection of inputs into feature space)

Instead of using a linear model for our input–output relationship, we may decide to projet the inputs into some high-dimensional space and then apply the linear model in this "lifted" space. To this end, we introduce the function $\boldsymbol{\phi}$ which maps the input vector $\mathbf{x}$ into an N-dimensional feature space. Our model becomes

$$f(\mathbf{x}) = \boldsymbol{\phi}(\mathbf{x})^\mathsf{T}\mathbf{w}, \quad y = f(\mathbf{x}) + \varepsilon.$$

We may proceed as before, except that everywhere $\boldsymbol{\phi}(\mathbf{x})$ is substituted for $\mathbf{x}$. The predictive distribution becomes

$$y_*|\mathbf{x}_*, X, \mathbf{y} \sim= \mathcal{N}\left(\frac{\boldsymbol{\phi}(\mathbf{x}_*)^\mathsf{T} A^{-1} \Phi \mathbf{y}}{\sigma_n^2}, \boldsymbol{\phi}(\mathbf{x}_*)^\mathsf{T} A^{-1} \boldsymbol{\phi}(\mathbf{x}_*)\right),$$

where $\Phi = \Phi(X)$ is the aggregation of columns $\boldsymbol{\phi}(\mathbf{x})$ and $A = \sigma_n^{-2}\Phi\Phi^\mathsf{T} + \Sigma_p^{-1}$.

Nonlinear models have more expressiveness than linear models, but they become expensive to compute as the dimension of the feature space increases. They also have a tendency to overfit the data. That is, to identify patterns in the data which do not truly exist but are the consequence of having a finite sample size. An example of nonlinear regression with a quintic feature map is shown in the top row of Fig. 6.2. Since $\mathbf{w}$ is a six-dimensional vector we only show the prior and posterior distributions for y.

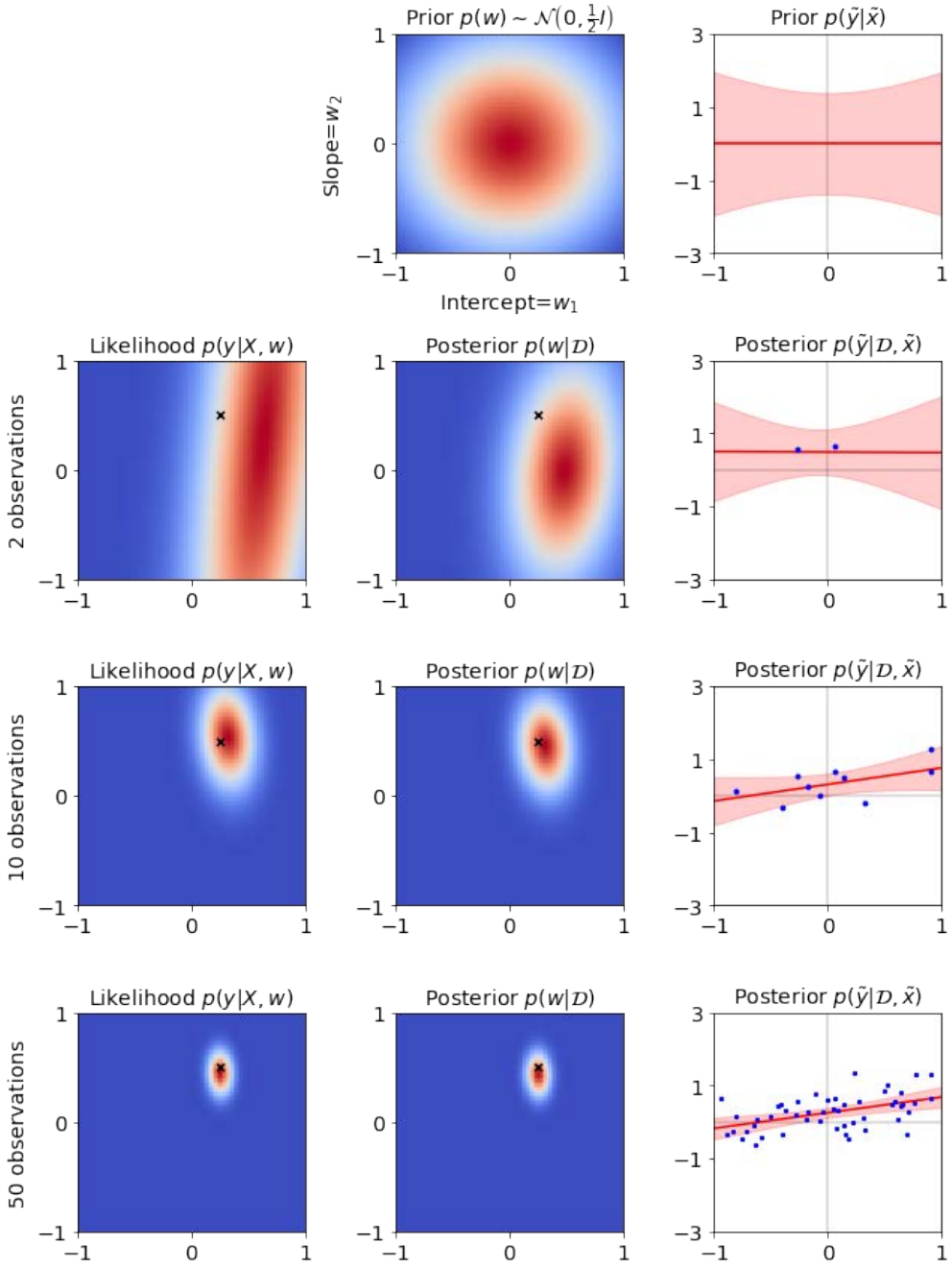


Figure 6.1 – Top row: Prior distributions for $\mathbf{w}$ and $y_*|\mathbf{x}_*$ for a two dimensional linear problem. Lower three rows: Likelihood $p(\mathcal{D}|\mathbf{w})$, $\mathbf{w}$-posterior $p(\mathbf{w}|\mathcal{D})$, and y-posterior $p(y_*|\mathcal{D}, \mathbf{x}_*)$ for datasets of size 2, 10, and 50. Black crosses at $\mathbf{w} = (0.5, 0.25)$ indicates model used to generate data with $\sigma^2 = 1/4$.

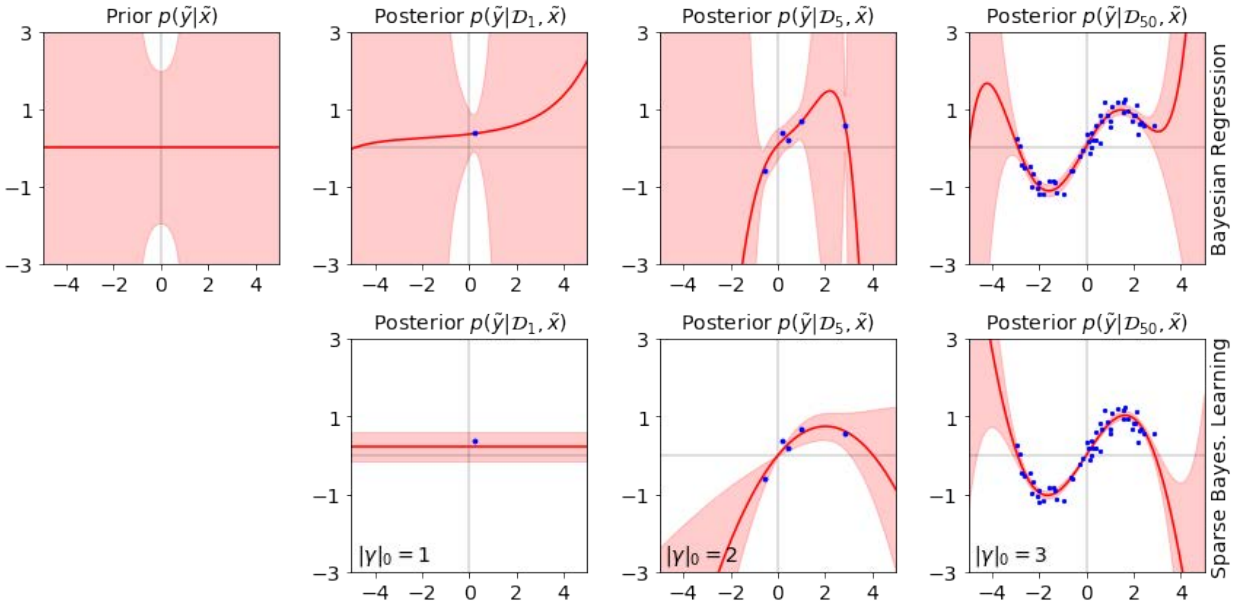


Figure 6.2 – Top row: Prior and posterior distributions for y_* for datasets of size 1, 5, and 50 using features $\phi(x) = (1, x, x^2, x^3, x^4, x^5)$. Bottom row: posterior distribution for y_* using automatic relevance determination. $|\gamma|_0$ indicates the number of non-zero elements of γ and therefore the number of features used in the model.

6.3 Selecting a prior

In order to apply the Bayesian regression framework we have assumed that we are given a prior in the form of $\mathbf{w} \sim \mathcal{N}(0, \Sigma_p)$. There are two natural questions to ask;

- Why did we assume $\mathbf{w} \sim \mathcal{N}(0, \Sigma_p)$?
- Where does Σ_p come from?

The answer to the first question is simple, if not entirely satisfying. We first define a conjugate prior. We say that $p(\mathbf{w})$ is a **conjugate prior** for likelihood $p(\mathcal{D}|\mathbf{w})$ if the posterior distribution $p(\mathbf{w}|\mathcal{D})$ is in the same family of distributions as $p(\mathbf{w})$, e.g. they are both Gaussian. Using a conjugate prior simplifies the problem algebraically and allows us to see the difference between prior and posterior clearly by looking at how parameters change. Gaussian distributions are self conjugate, so we selected $\mathbf{w} \sim \mathcal{N}(0, \Sigma_p)$ because our likelihood is a Gaussian. As a consequence, we are able to derive the posterior distribution analytically and to easily see how the likelihood acts to update the prior.

Of course, what makes sense algebraically may be arbitrary in the context of the problem we are considering and we are not required to pick a conjugate prior. In some cases domain knowledge might lead us to other choices or we may have a sufficiently complex likelihood function that we are unable to find a conjugate prior. Computation of the posterior distribution in these cases will be much more challenging.

We now consider the second question; once we have established that it is reasonable to use a Gaussian prior for $\mathbf{w}$, how do we choose an appropriate covariance, Σ_p? The Bayesian approach is

to let Σ_p itself be a random variable with its own prior distribution. That is,

$$\Sigma_p \sim p(\Sigma_p|\eta)$$

where η is a parameter characterizing the prior distribution of Σ_p. This gives the joint posterior distribution over $\mathbf{w}$ and Σ_p,

$$p(\mathbf{w}, \Sigma_p|\mathcal{D}) = p(\mathcal{D}|\mathbf{w})p(\mathbf{w}|\Sigma_p)p(\Sigma_p|\eta).$$

At this point it may seem like we have not made any progress. We no longer need to pick Σ_p, but we do need to pick an appropriate prior $p(\Sigma_p|\eta)$ and hyperparameter η. Furthermore, our posterior is now higher dimensional and more complicated than before. We solve both of the problems with a method called **empirical Bayes**, also known as type-II maximum likelihood or evidence maximization. Using empirical Bayes we set,

$$p(\Sigma_p) = \delta\left(\Sigma_p - \underset{\tilde{\Sigma}_p}{arg\,max}\, p\left(\mathcal{D}|\tilde{\Sigma}_p\right)\right) = \delta\left(\Sigma_p - \underset{\tilde{\Sigma}_p}{arg\,max} \int p\left(\mathcal{D}|\mathbf{w}\right) p\left(\mathbf{w}|\tilde{\Sigma}_p\right)\, d\mathbf{w}\right)$$

where δ is the Dirac delta distribution so that $\Sigma_p = \underset{\tilde{\Sigma}_p}{arg\,max}\, p\left(\mathcal{D}|\tilde{\Sigma}_p\right)$ with probability one. For the nonlinear Bayesian regression problem we considered in the previous section the evidence is given by,

$$\begin{aligned} p\left(\mathcal{D}|\tilde{\Sigma}_p\right) &= \int p\left(\mathcal{D}|\mathbf{w}\right) p\left(\mathbf{w}|\tilde{\Sigma}_p\right)\, d\mathbf{w} \\ &= \frac{1}{(2\pi)^{n/2}|\Sigma_y|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\mathbf{y}^T\Sigma_y^{-1}\mathbf{y}\right), \end{aligned}$$

where,

$$\Sigma_y = \sigma_n^2 I + \Phi\tilde{\Sigma}_p\Phi^T.$$

Maximizing the evidence is equivalent to minimizing the negative log of the evidence, which is often an easier computational problem. We therefor find Σ_p through solving,

$$\Sigma_p = \underset{\tilde{\Sigma}_p}{arg\,min} \log|\Sigma_y| + \mathbf{y}^T\Sigma_y^{-1}\mathbf{y}.$$

For $p(\mathbf{w}|\Sigma_p)$ to be a probability density function we need Σ_p to be symmetric positive definite, though we may also consider symmetric positive semidefinite Σ_p under the restriction that $\mathbf{w} \perp \ker(\Sigma_p)$. Optimization on the space of symmetric positive (semi)definite matrices is highly nontrivial, so we often assume a simple Anstaz for Σ_p. The most common are $\Sigma_p = \lambda^{-1}I$ with $\lambda > 0$, called **Bayesian ridge regression**, and $\Sigma_p = \text{diag}(\boldsymbol{\gamma})$ with $\gamma_i \geq 0$, called **automatic relevance determination** (ARD).

Automatic relevance determination is also often called sparse Bayesian learning (SBL) due to the tendency of the empirical Bayes estimate of $\boldsymbol{\gamma}$ to be sparse. In this case the support of $p(\mathbf{w})$ lies on a low dimensional hyperplane and any sample from $p(\mathbf{w})$ or $p(\mathbf{w}|\mathcal{D})$ will be sparse. The second row of Fig. 6.2 shows the result of applying SBL to a simple regression problem. Compared to Bayesian regression with $\Sigma_p = \lambda^{-1}I$, SBL learns models with fewer nonzero terms.

6.4 MAP Estimation

In classical statistics we assume there exists some true value of $\mathbf{w}$ and are concerned with estimating its value. A common technique that borrows from Bayesian methods is to assume a prior $p(\mathbf{w})$ and

estimate $\mathbf{w}$ as the mode of the posterior distribution,

$$\hat{\mathbf{w}}_{MAP} = \underset{\mathbf{w}}{arg\,max}\, p(\mathbf{w}|\mathcal{D}) = \underset{\mathbf{w}}{arg\,max}\, p(\mathcal{D}|\mathbf{w})p(\mathbf{w})$$

This is called **maximum a-posteriori** estimation, often abbreviated as MAP-estimation. For the nonlinear regression problem with $\Sigma_p = \lambda^{-1}I$ the MAP-estimate $\hat{\mathbf{w}}_{MAP}$ is given in closed form by,

$$\hat{\mathbf{w}}_{MAP} = \left(\Phi\Phi^T + \sigma_n^2\lambda I\right)^{-1}\Phi\mathbf{y}.$$

Taking the singular value decomposition $\Phi = USV^T$ with $S = diag(\mathbf{s})$ and noting that $\mathbf{y} = \Phi^T\mathbf{w} + \epsilon$ with $\epsilon = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)$ we can rewrite $\hat{\mathbf{w}}_{MAP}$ as,

$$\begin{aligned}
\hat{\mathbf{w}}_{MAP} &= \left(\Phi\Phi^T + \sigma_n^2\lambda I\right)^{-1}\Phi\Phi^T\mathbf{w} + \left(\Phi\Phi^T + \sigma_n^2\lambda I\right)^{-1}\Phi\epsilon \\
&= \left(U(S^2 + \lambda\sigma_n^2 I)U^T\right)^{-1}US^2U^T\mathbf{w} + \left(U(S^2 + \lambda\sigma_n^2 I)U^T\right)^{-1}USV^T\epsilon \\
&= Udiag\left(\frac{\mathbf{s}^2}{\mathbf{s}^2 + \lambda\sigma_n^2}\right)U^T\mathbf{w} + Udiag\left(\frac{\mathbf{s}}{\mathbf{s}^2 + \lambda\sigma_n^2}\right)V^T\boldsymbol{\epsilon} \\
&= \mathbf{w} - \underbrace{Udiag\left(\frac{\lambda\sigma_n^2}{\mathbf{s}^2 + \lambda\sigma_n^2}\right)U^T\mathbf{w}}_{\text{bias}} + \underbrace{Udiag\left(\frac{\mathbf{s}}{\mathbf{s}^2 + \lambda\sigma_n^2}\right)V^T\boldsymbol{\epsilon}}_{\text{variance}}.
\end{aligned}$$

We see that $\hat{\mathbf{w}}_{MAP}$ is a biased estimator of the true $\mathbf{w}$ in that $\mathbb{E}_\epsilon[\hat{\mathbf{w}}_{MAP}] \neq \mathbf{w}$, but that as λ becomes large the variance term decays. Setting $\lambda > 0$ will result in more accurate predictions in many practical scenarios. Furthermore, using MAP estimation allows one to estimate $\mathbf{w}$ even with fewer observations than features.

6.5 Gaussian Process Regression (GPR)

We begin out discussion of Gaussian processes by rewriting the posterior distribution $y_*|\mathbf{x}_*, X, \mathbf{y}$ for the nonlinear regression problem in terms of a symmetric function called a kernel. Recall that for nonlinear regression the posterior is given by,

$$y_*|\mathbf{x}_*, X, \mathbf{y} \sim \mathcal{N}\left(\frac{\boldsymbol{\phi}(\mathbf{x}_*)^\mathsf{T}A^{-1}\Phi\mathbf{y}}{\sigma_n^2}, \boldsymbol{\phi}(\mathbf{x}_*)^\mathsf{T}A^{-1}\boldsymbol{\phi}(\mathbf{x}_*)\right),$$

with,

$$A = \sigma_n^{-2}\Phi\Phi^\mathsf{T} + \Sigma_p^{-1}$$

Define $k(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x})\Sigma_p\phi(\mathbf{x}')$. Inserting $\left(k(X,X) + \sigma_n^2 I\right)\left(k(X,X) + \sigma_n^2 I\right)^{-1} = I$ into the expression for the posterior mean gives,

$$\begin{aligned}
\frac{1}{\sigma_n^2}\boldsymbol{\phi}(\mathbf{x}_*)^\mathsf{T}A^{-1}\Phi\mathbf{y} &= \frac{1}{\sigma_n^2}\boldsymbol{\phi}(\mathbf{x}_*)^\mathsf{T}A^{-1}\Phi\left(k(X,X) + \sigma_n^2 I\right)\left(k(X,X) + \sigma_n^2 I\right)^{-1}\mathbf{y} \\
&= \frac{1}{\sigma_n^2}\boldsymbol{\phi}(\mathbf{x}_*)^\mathsf{T}A^{-1}\left(\Phi\Phi^T\Sigma_p\Phi + \sigma_n^2\Phi\right)\left(k(X,X) + \sigma_n^2 I\right)^{-1}\mathbf{y} \\
&= \frac{1}{\sigma_n^2}\boldsymbol{\phi}(\mathbf{x}_*)^\mathsf{T}A^{-1}\left(\Phi\Phi^T + \sigma_n^2\Sigma_p^{-1}\right)\Sigma_p\Phi\left(k(X,X) + \sigma_n^2 I\right)^{-1}\mathbf{y} \\
&= \boldsymbol{\phi}(\mathbf{x}_*)^\mathsf{T}\Sigma_p\Phi\left(k(X,X) + \sigma_n^2 I\right)^{-1}\mathbf{y} \\
&= k(\mathbf{x}_*, X)\left(k(X,X) + \sigma_n^2 I\right)^{-1}\mathbf{y}
\end{aligned}$$

Applying the Woodbury identity to A^{-1} gives,

$$\begin{aligned}\phi(\mathbf{x}_*)^T \left(\sigma_n^{-2}\Phi\Phi^\mathsf{T} + \Sigma_p^{-1}\right)^{-1} \phi(\mathbf{x}_*) &= \phi(\mathbf{x}_*)^T\Sigma_p\phi(\mathbf{x}_*) - \phi(\mathbf{x}_*)^T\Sigma_p\Phi\left(\sigma_n^2 I + \Phi^T\Sigma_p\Phi\right)\Phi^T\Sigma_p\phi(\mathbf{x}_*) \\ &= k(\mathbf{x}_*,\mathbf{x}_*) - k(\mathbf{x}_*,X)\left(k(X,X)+\sigma_n^2 I\right)^{-1}k(X,\mathbf{x}_*)\end{aligned}$$

It follows that the posterior distribution for y_* is equivalently written as,

$$y_*|\mathbf{x}_*, X, \mathbf{y} \sim \mathcal{N}\left(k(x_*,X)\left(k(X,X)+\sigma_n^2 I\right)^{-1}\mathbf{y}, k(\mathbf{x}_*,\mathbf{x}_*) - k(\mathbf{x}_*,X)\left(k(X,X)+\sigma_n^2 I\right)^{-1}k(X,\mathbf{x}_*)\right),$$

The kernel function here is an inner product with weights given by the prior covariance matrix Σ_p, but we can define a kernel more generally. Let $\mathcal{X}$ be a closed subset of $\mathbb{R}^m$. A **kernel** is a function $k : \mathcal{X}\times\mathcal{X}\to\mathbb{R}$ such that $k(\mathbf{x},\mathbf{x}') = k(\mathbf{x}',\mathbf{x})$ and for any collection $\{\mathbf{x}_1,\dots\mathbf{x}_n\}$ the matrix defined with $K_{ij} = k(\mathbf{x}_i,\mathbf{x}_j)$ is positive semi-definite. We can construct a wide variety of functions that meet the definition for a kernel. It turns out that just about any kernel is equivalent to an inner product in a Hilbert space. We present a slightly simplified version of **Mercer's theorem**.

Theorem. (Mercer) *If k is a kernel with $\iint |k(\mathbf{x},\mathbf{x}')|^2\mu(\mathbf{x})\mu(\mathbf{x}')\,d\mathbf{x}d\mathbf{x}' < \infty$ for some $\mu > 0$ then,*

$$k(\mathbf{x},\mathbf{x}') = \sum_{j=0}^{\infty}\lambda_j\phi_j(\mathbf{x})\phi_j(\mathbf{x}')$$

where λ_j, ϕ_j are eigenvalue, eigenvector pairs of the operator T_k defined by,

$$T_k f(\mathbf{x}) = \int k(\mathbf{x},\mathbf{x}')f(\mathbf{x}')\mu(\mathbf{x}')\,d\mathbf{x}'.$$

In other words kernels are, under mild assumptions, equivalent to inner products with infinitely many features. This fact, together with the expression for the posterior distribution of y_* in terms of a kernel function motivates Gaussian processes. Rather than specifying the functional form of the input–output relationship, we may say that $f(\mathbf{x})$ is a **Gaussian process** with mean $m(\mathbf{x})$ and covariance given by kernel function $k(\mathbf{x},\mathbf{x}')$ if,

$$\begin{aligned}\mathbb{E}[f(\mathbf{x})] &= m(\mathbf{x}), \\ \operatorname{cov}(f(\mathbf{x}), f(\mathbf{x}')) &= \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))] = k(\mathbf{x},\mathbf{x}').\end{aligned}$$

For the simple Bayesian linear regression model discussed in §6.1, we have

$$\begin{aligned}m(\mathbf{x}) &= \mathbf{x}\,\mathbb{E}[\mathbf{w}] = 0, \\ k(\mathbf{x},\mathbf{x}') &= \mathbf{x}\,\mathbb{E}[\mathbf{w}\mathbf{w}^\mathsf{T}]\mathbf{x}' = \mathbf{x}^\mathsf{T}\Sigma_p\mathbf{x}'.\end{aligned}$$

In Gaussian Process Regression, we only specify the covariance between data points. In some cases the function $m(\mathbf{x})$ is also specified, generally as a constant, but here we assume $m(\mathbf{x}) = 0$. As an example, we may choose

$$k(\mathbf{x},\mathbf{x}') = \exp\left(-\frac{|\mathbf{x}-\mathbf{x}'|^2}{2}\right).$$

For this particular covariance function, the covariance is almost unity between variables whose corresponding inputs are very close, and decreases as their distance in the input space increases.

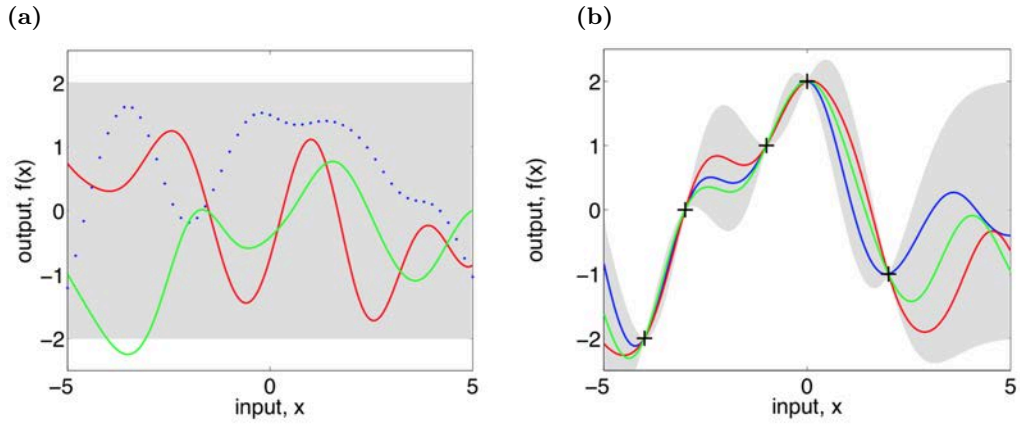


Figure 6.3 – (a) Three random functions drawn from a GP prior, and (b) three random functions drawn from the posterior. The shaded area corresponds to two standard deviations away from the pointwise mean. (Reproduced from Rasmussen & Williams, 2006).

For noise-free observations ($\sigma_n^2 = 0$), the joint distribution of the training outputs $\mathbf{y}$ and the test output y_* according to the prior is

$$\begin{bmatrix} \mathbf{y} \\ y_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} k(X,X) & k(X,\mathbf{x}_*) \\ k(\mathbf{x}_*,X) & k(\mathbf{x}_*,\mathbf{x}_*) \end{bmatrix}\right).$$

Here, we consider only one test point, so $k(X,\mathbf{x}_*)$ denotes the $n \times 1$ matrix of the covariances between the test point and all of the training points, and similarly for the other entries $k(X,X)$, $k(\mathbf{x}_*,\mathbf{x}_*)$ and $k(\mathbf{x}_*,X)$. To get the posterior distribution, we compute the conditional,

$$y_*|\mathbf{x}_*, X, \mathbf{y} \sim \mathcal{N}(k(\mathbf{x}_*,X)k(X,X)^{-1}\mathbf{y}, k(\mathbf{x}_*,\mathbf{x}_*) - k(\mathbf{x}_*,X)k(X,X)^{-1}k(X,\mathbf{x}_*)),$$

which can be thought of as a superposition of Gaussian kernels. Note that the variance is zero for the training data. For noisy observations we have,

$$\begin{bmatrix} \mathbf{y} \\ y_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} k(X,X) + \sigma_n^2 I & k(X,\mathbf{x}_*) \\ k(\mathbf{x}_*,X) & k(\mathbf{x}_*,\mathbf{x}_*) \end{bmatrix}\right).$$

so,

$$y_*|\mathbf{x}_*, X, \mathbf{y} \sim \mathcal{N}(k(\mathbf{x}_*,X)\left(k(X,X) + \sigma_n^2 I\right)^{-1}\mathbf{y}, k(\mathbf{x}_*,\mathbf{x}_*) - k(\mathbf{x}_*,X)\left(k(X,X) + \sigma_n^2 I\right)^{-1}k(X,\mathbf{x}_*)).$$

This is precisely the form we had for the posterior in the case of nonlinear regression except that now we have generalized our kernel to include infinite dimensional feature spaces. For illustration of these computations, see figure 6.3.

Common choices of covariance functions include the square exponential kernel given by,

$$k(\mathbf{x},\mathbf{x}') = \sigma_f^2 \exp\left(-\frac{\|\mathbf{x}-\mathbf{x}'\|^2}{2\sigma_l^2}\right),$$

and the Matern class of kernels given by,

$$k_\nu(\mathbf{x},\mathbf{x}') = \sigma_f^2 \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{\sqrt{2\nu}\|\mathbf{x}-\mathbf{x}'\|}{\sigma_l}\right)^\nu K_\nu\left(\frac{\sqrt{2\nu}\|\mathbf{x}-\mathbf{x}'\|}{\sigma_l}\right)$$

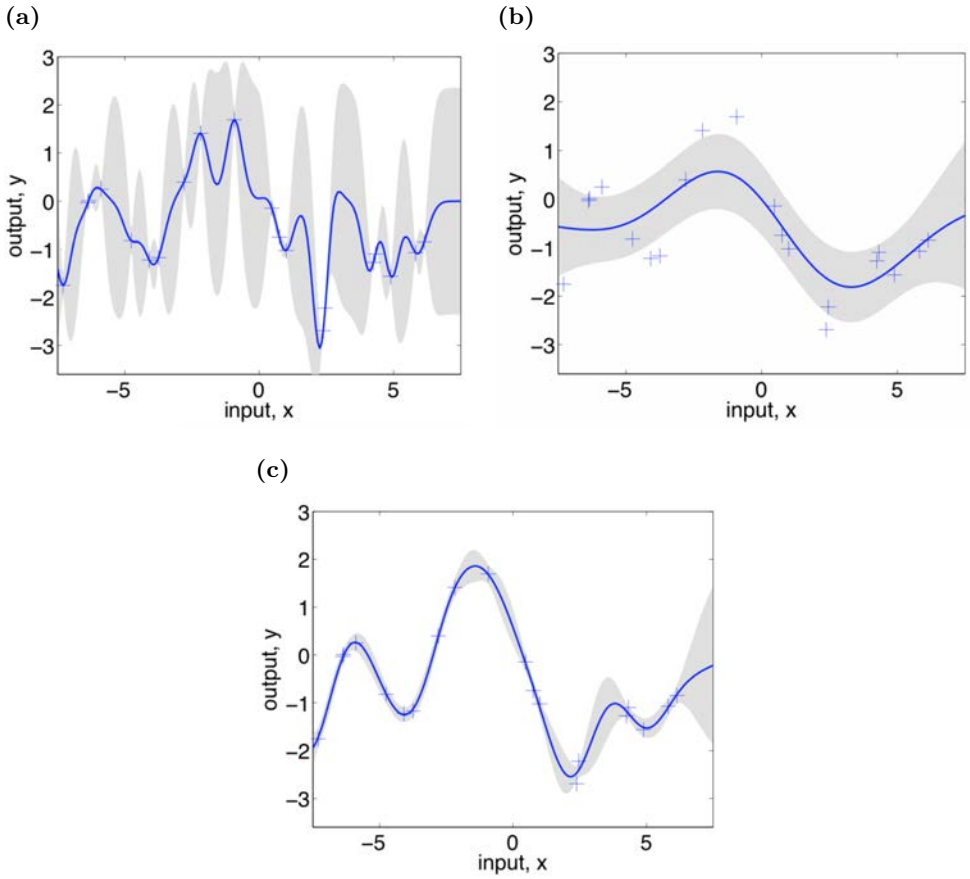


Figure 6.4 – Illustration of the effect of hyperparameters on data generated from a GP. (Reproduced from Rasmussen & Williams, 2006).

where K_ν is a modified Bessel function. ν is generally selected to be a half-integer value $3/2, 5/2, \ldots$ so that the kernel simplifies to,

$$k_{3/2}(\mathbf{x}, \mathbf{x}') = \left(1 + \sqrt{3}r\right) \exp\left(-\sqrt{3}r\right)$$
$$k_{5/2}(\mathbf{x}, \mathbf{x}') = \left(1 + \sqrt{5}r + \frac{5}{3}r^2\right) \exp\left(-\sqrt{5}r\right)$$

where $r = \|\mathbf{x} - \mathbf{x}'\|/\sigma_l$. Samples from the square exponential kernel are infinitely differentiable and those from Matern kernel with $\nu = (2n+1)/2$ are n-times differentiable.

In general, the covariance function may have free parameters, the so-called **hyperparameters**. In each of the above examples the parameter σ_l governs the "spread" of the kernel. For small values of σ_l, the covariance decays rapidly with the distance between input points; this may lead to overfitting (figure 6.4a). On the other hand, large values of σ_l lead to slow decay of the covariance as the distance between input points grows, which smoothes out asperities (figure 6.4b). The challenge is to find a trade-off between the hyperparameters to achieve reasonable predictions (figure 6.4c). The parameter σ_f is the marginal variance of the prior at a single point.

In a practical setting, we learn the coefficients in a similar manner to Empirical Bayes. From the initial assumptions we used to define a Gaussian process, the likelihood $p(\mathcal{D}|\sigma_l, \sigma_f) = p(\mathbf{y}|X, \sigma_l, \sigma_f)$

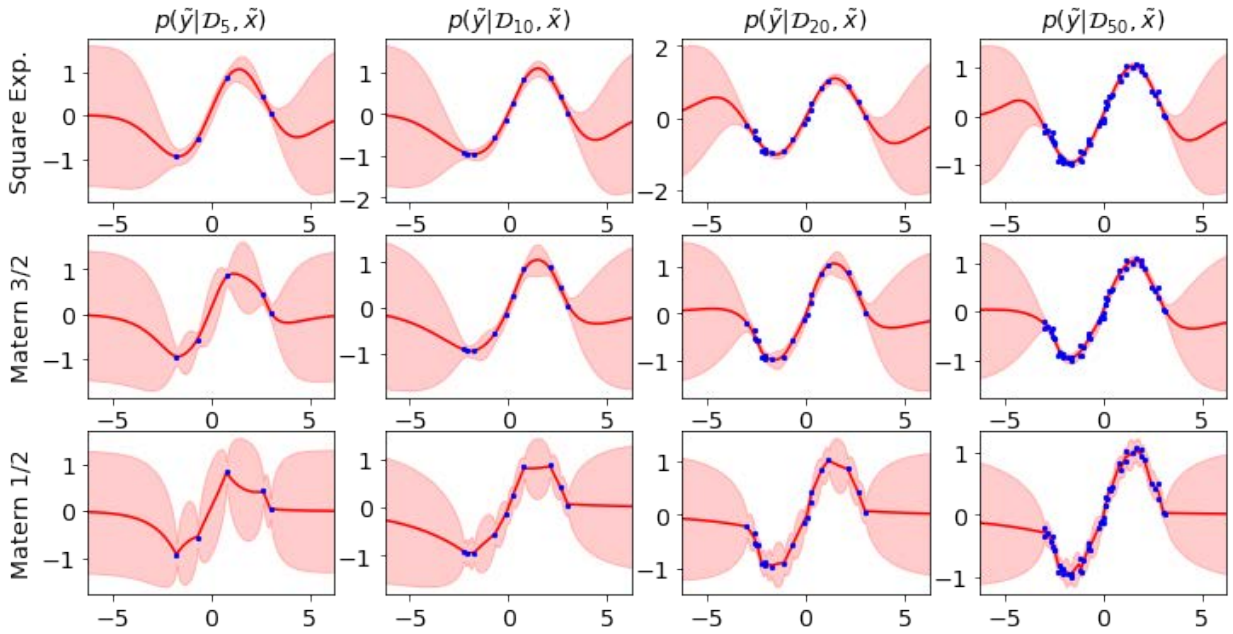


Figure 6.5 – Posterior distribution for GP regression using 3 different kernels and 4 sample sizes. The Matern 1/2 covariance function is not generally used in practice but we include it here to highlight that small values of ν result in less smooth functions.

is given by,

$$p(\mathbf{y}|X, \sigma_l, \sigma_f) = \mathcal{N}\left(\mathbf{y}|k(X,X) + \sigma_n^2 I\right).$$

To find optimal hyper-parameters we minimize the negative log-likelihood. We therefore set,

$$\sigma_l, \sigma_f = \underset{\sigma_l, \sigma_f}{arg\,min} \log\left|k(X,X) + \sigma_n^2 I\right| + \mathbf{y}^T\left(k(X,X) + \sigma_n^2 I\right)^{-1}\mathbf{y}.$$

This is a non-convex optimization problem and in general is not solvable. Gradient decent and quasi-Newton optimization algorithms often yield good results nonetheless. It is common practice to search for the optimal hyperparameters using a quasi-Newton method and from several initial conditions and select the result with lowest function value. Examples of the posterior predictive distribution using several kernels are shown in Fig. 6.5.

Bibliography

[1] V. Belenky, K. M. Weems, and W.-m. Lin. Numerical Procedure for Evaluation of Capsizing Probability with Split Time Method. In *27th Symposium on Naval Hydrodynamics, Seoul, Korea*, number October, 2008.

[2] M. Nicodemi. Extreme value statistics. *in Encyclopedia of complexity and systems science*, E:3317, 2009.

[3] T. Soong and M. Grigoriu. *Random Vibration of Mechanical and Structural Systems*. PTR Prentice Hall, 1993.