Gaussian Probability Density Functions: Properties and Error Characterization

Maria Isabel Ribeiro
Institute for Systems and Robotics
Instituto Superior Tcnico
Av. Rovisco Pais, 1
1049-001 Lisboa PORTUGAL
{mir@isr.ist.utl.pt}

February 2004

Contents

Chapter 1

Normal random variables

A random variable X is said to be normally distributed with mean μ and variance σ^2 if its probability density function (pdf) is

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \quad -\infty < x < \infty. \tag{1.1}$$

Whenever there is no possible confusion between the random variable X and the real argument, x, of the pdf this is simply represented by $f(x)$ omitting the explicit reference to the random variable X in the subscript. The Normal or Gaussian distribution of X is usually represented by,

$$X \sim \mathcal{N}(\mu, \sigma^2),$$

or also,

$$X \sim \mathcal{N}(x-\mu, \sigma^2).$$

The Normal or Gaussian pdf (1.1) is a bell-shaped curve that is symmetric about the mean μ and that attains its maximum value of $\frac{1}{\sqrt{2\pi}\sigma} \simeq \frac{0.399}{\sigma}$ at $x = \mu$ as represented in Figure 1.1 for $\mu = 2$ and $\sigma^2 = 1.5^2$.

The Gaussian pdf $\mathcal{N}(\mu, \sigma^2)$ is completely characterized by the two parameters μ and σ^2, the first and second order moments, respectively, obtainable from the pdf as

$$\mu \quad = E[X] = \int_{-\infty}^{\infty} x f(x) dx, \tag{1.2}$$

$$\sigma^2 \quad = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx \tag{1.3}$$

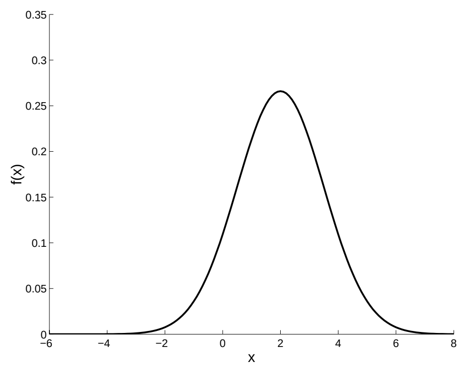

Figure 1.1: Gaussian or Normal pdf, N$(2, 1.5^2)$

The mean, or the expected value of the variable, is the centroid of the pdf. In this particular case of Gaussian pdf, the mean is also the point at which the pdf is maximum. The variance σ^2 is a measure of the dispersion of the random variable around the mean.

The fact that (1.1) is completely characterized by two parameters, the first and second order moments of the pdf, renders its use very common in characterizing the uncertainty in various domains of application. For example, in robotics, it is common to use Gaussian pdf to statistically characterize sensor measurements, robot locations, map representations.

The pdfs represented in Figure 1.2 have the same mean, $\mu = 2$, and $\sigma_1^2 > \sigma_2^2 > \sigma_3^2$ showing that the larger the variance the greater the dispersion around the mean.

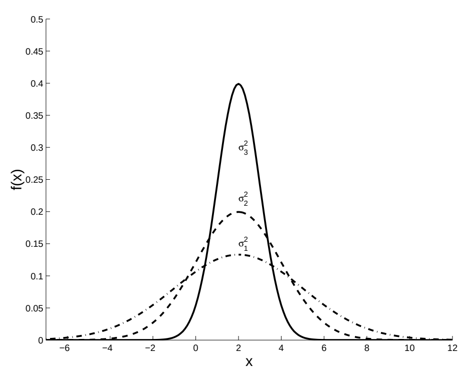

Figure 1.2: Gaussian pdf with different variances $(\sigma_1^2 = 3^2, \sigma_2^2 = 2^2, \sigma_3^2 = 1)$

Definition 1.1 *The square-root of the variance, σ, is usually known as* **standard deviation**.

Given a real number $x_a \in \mathcal{R}$, the probability that the random variable $X \sim \mathcal{N}(\mu, \sigma^2)$ takes values less or equal x_a is given by

$$Pr\{X \leq x_a\} = \int_{-\infty}^{x_a} f(x)dx = \int_{-\infty}^{x_a} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx, \qquad (1.4)$$

represented by the shaded area in Figure 1.3.

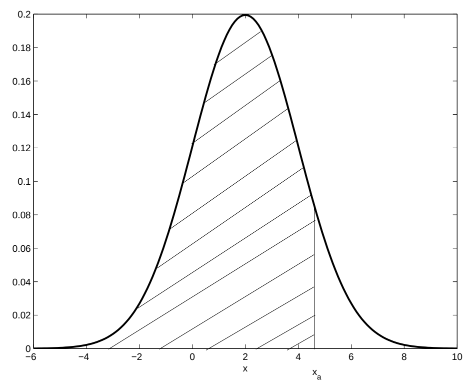

Figure 1.3: Probability evaluation using pdf

To evaluate the probability in (1.4) the *error function*, $erf(x)$, which is related with $\mathcal{N}(0,1)$,

$$erf(x) = \frac{1}{\sqrt{2\pi}} \int_0^x \exp^{-y^2/2} dy \qquad (1.5)$$

plays a key role. In fact, with a change of variables, (1.4) may be rewritten as

$$Pr\{X \leq x_a\} = \begin{cases} 0.5 - erf(\frac{\mu - x_a}{\sigma}) & for \quad x_a \leq \mu \\ \\ 0.5 + erf(\frac{x_a - \mu}{\sigma}) & for \quad x_a \geq \mu \end{cases}$$

stating the importance of the error function, whose values for various x are displayed in Table 1.1

In various aspects of robotics, in particular when dealing with uncertainty in mobile robot localization, it is common the evaluation of the probability that a

x	erf x	x	erf x	x	erf x	x	erf x
0.05	0.01994	0.80	0.28814	1.55	0.43943	2.30	0.48928
0.10	0.03983	0.85	0.30234	1.60	0.44520	2.35	0.49061
0.15	0.05962	0.90	0.31594	1.65	0.45053	2.40	0.49180
0.20	0.07926	0.95	0.32894	1.70	0.45543	2.45	0.49286
0.25	0.09871	1.00	0.34134	1.75	0.45994	2.50	0.49379
0.30	0.11791	1.05	0.35314	1.80	0.46407	2.55	0.49461
0.35	0.13683	1.10	0.36433	1.85	0.46784	2.60	0.49534
0.40	0.15542	1.15	0.37493	1.90	0.47128	2.65	0.49597
0.45	0.17365	1.20	0.38493	1.95	0.47441	2.70	0.49653
0.50	0.19146	1.25	0.39435	2.00	0.47726	2.75	0.49702
0.55	0.20884	1.30	0.40320	2.05	0.47982	2.80	0.49744
0.60	0.22575	1.35	0.41149	2.10	0.48214	2.85	0.49781
0.65	0.24215	1.40	0.41924	2.15	0.48422	2.90	0.49813
0.70	0.25804	1.45	0.42647	2.20	0.48610	2.95	0.49841
0.75	0.27337	1.50	0.43319	2.25	0.48778	3.00	0.49865

Table 1.1: erf - Error function

random variable Y (more generally a random vector representing the robot location) lies in an interval around the mean value μ. This interval is usually defined in terms of the standard deviation, σ, or its multiples.

Using the error function, (1.5),the probability that the random variable X lies in an interval whose width is related with the standard deviation, is

$$Pr\{|X-\mu| \leq \sigma\} = 2.erf(1) = 0.68268 \tag{1.6}$$

$$Pr\{|X-\mu| \leq 2\sigma\} = 2.erf(2) = 0.95452 \tag{1.7}$$

$$Pr\{|X-\mu| \leq 3\sigma\} = 2.erf(3) = 0.9973 \tag{1.8}$$

In other words, the probability that a Gaussian random variable lies in the interval $[\mu - 3\sigma,\ \mu + 3\sigma]$ is equal to 0.9973. Figure 1.4 represents the situation (1.6)corresponding to the probability of X lying in the interval $[\mu - \sigma,\ \mu + \sigma]$.

Another useful evaluation is the locus of values of the random variable X where the pdf is greater or equal a given pre-specified value K_1, i.e.,

$$\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \geq K_1 \iff \frac{(x-\mu)^2}{2\sigma^2} \leq K \tag{1.9}$$

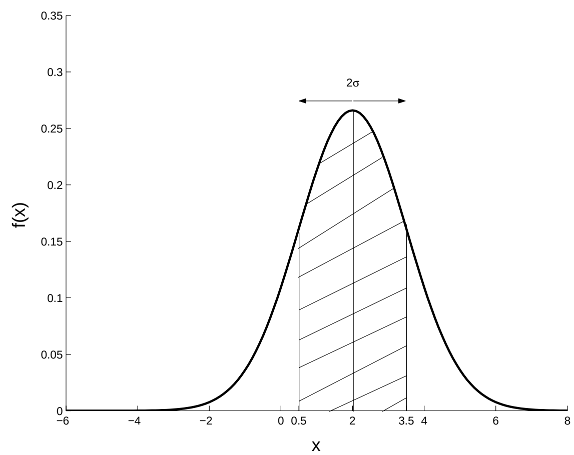

Figure 1.4: Probability of X taking values in the interval $[\mu - \sigma,\ \mu + \sigma]$, $\mu = 2, \sigma = 1.5$

with $K = -\ln(\sqrt{2\pi}\sigma K_1)$. This locus is the line segment

$$\mu - \sigma\sqrt{K} \leq x \leq \mu + \sigma\sqrt{K}$$

as represented in Figure 1.5.

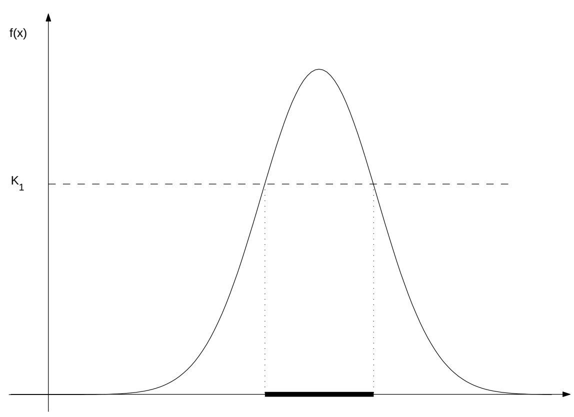

Figure 1.5: Locus of x where the pdf is greater or equal than K_1

Chapter 2

Normal random vectors

A random vector $X = [X_1,\ X_2,\ \ldots\ X_n]^T \in \mathcal{R}^n$ is Gaussian if its pdf is

$$f_X(x) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x - m_X)^T\Sigma^{-1}(x - m_X)\right\} \qquad (2.1)$$

where

- $m_X = E(X)$ is the mean vector of the random vector X,
- $\Sigma_X = E[(X - m_X)(X - m_X)^T]$ is the covariance matrix,
- $n = dimX$ is the dimension of the random vector,

also represented as

$$X \sim \mathcal{N}(m_X, \Sigma_X).$$

In (2.1), it is assumed that x is a vector of dimension n and that Σ^{-1} exists. If Σ is simply non-negative definite, then one defines a Gaussian vector through the characteristic function, [2].

The mean vector m_X is the collection of the mean values of each of the random variables X_i,

$$m_X = E\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} = \begin{bmatrix} m_{X_1} \\ m_{X_2} \\ \vdots \\ m_{X_n} \end{bmatrix}.$$

The covariance matrix is symmetric with elements,

$$\Sigma_X = \Sigma_X^T =$$
$$= \begin{bmatrix} E(X_1 - m_{X_1})^2 & E(X_1 - m_{X_1})(X_2 - m_{X_2}) & \dots & E(X_1 - m_{X_1})(X_n - m_{X_n}) \\ E(X_2 - m_{X_2})(X_1 - m_{X_1}) & E(X_2 - m_{X_2})^2 & \dots & E(X_2 - m_{X_2})(X_n - m_{X_n}) \\ \vdots & & & \vdots \\ E(X_n - m_{X_n})(X_1 - m_{X_1}) & \dots & \dots & E(X_n - m_{X_n})^2 \end{bmatrix}.$$

The diagonal elements of Σ are the variance of the random variables X_i and the generic element $\Sigma_{ij} = E(X_i - m_{X_i})(X_j - m_{X_j})$ represents the covariance of the two random variables X_i and X_j.

Similarly to the scalar case, the pdf of a Gaussian random vector is completely characterized by its first and second moments, the mean vector and the covariance matrix, respectively. This yields interesting properties, some of which are listed in Chapter 3.

When studying the localization of autonomous robots, the random vector X plays the role of the robot's location. Depending on the robot characteristics and on the operating environment, the location may be expressed as:

- a two-dimensional vector with the position in a 2D environment,
- a three-dimensional vector (2d-position and orientation) representing a mobile robot's location in an horizontal environment,
- a six-dimensional vector (3 positions and 3 orientations) in an underwater vehicle

When characterizing a 2D-laser scanner in a statistical framework, each range measurement is associated with a given pan angle corresponding to the scanning mechanism. Therefore the pair (distance, angle) may be considered as a random vector whose statistical characterization depends on the physical principle of the sensor device.

The above examples refer quantities, (e.g., robot position, sensor measurements) that are not deterministic. To account for the associated uncertainties, we consider them as random vectors. Moreover, we know how to deal with Gaussian random vectors that show a number of nice properties; this (but not only) pushes us to consider these random variables as been governed by a Gaussian distribution.

In many cases, we have to deal with low dimension Gaussian random vectors (second or third dimension), and therefore it is useful that we particularize the n-dimensional general case to second order and present and illustrate some properties.

The following section particularizes some results for a second order Gaussian pdf.

2.1 Particularization for second order

In the first two above referred cases, the Gaussian random vector is of order two or three. In this section we illustrate the case when n=2.

Let

$$Z = \begin{bmatrix} X \\ Y \end{bmatrix} \in {}^2,$$

be a second-order Gaussian random vector, with mean,

$$E[Z] = E \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} m_X \\ m_Y \end{bmatrix} \tag{2.2}$$

and covariance matrix,

$$\Sigma = \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix} \tag{2.3}$$

where σ_X^2 and σ_Y^2 are the variances of the random variables X and Y and σ_{XY} is the covariance of X and Y, defined below.

Definition 2.1 *The covariance σ_{XY} of the two random variables X and Y is the number*

$$\sigma_{XY} = E[(X - m_X)(Y - m_Y)] \tag{2.4}$$

where $m_X = E(X)$ and $m_Y = E(Y)$.

Expanding the product (2.4), yields,

$$\begin{aligned} \sigma_{XY} &= E(XY) - m_X E(Y) - m_Y E(X) + m_X m_Y & (2.5) \\ &= E(XY) - E(X)E(Y) & (2.6) \\ &= E(X) - m_X m_Y. & (2.7) \end{aligned}$$

Definition 2.2 *The* **correlation coefficient** *of the variables X and Y is defined as*

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \tag{2.8}$$

Result 2.1 *The correlation coefficient and the covariance of the variables X and Y satisfy the following inequalities ,*

$$|\rho| \leq 1, \qquad |\sigma_{XY}| \leq \sigma_X \sigma_Y. \tag{2.9}$$

Proof: [2] Consider the mean value of

$$E[a(X - m_X) + (Y - m_Y)]^2 = a^2\sigma_X^2 + 2a\sigma_{XY} + \sigma_Y^2$$

which is a positive quadratic for any a, and hence, the discriminant is negative, i.e.,

$$\sigma_{XY} - \sigma_X^2\sigma_Y^2 \leq 0$$

from where (2.9) results.

According to the previous definitions, the covariance matrix (2.3) is rewritten as

$$\Sigma = \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}. \tag{2.10}$$

For this second-order case, the Gaussian pdf particularizes as, with $z = [x \;\; y]^T \in \mathcal{R}^2$,

$$\begin{aligned} f(z) &= \frac{1}{2\pi\sqrt{det\Sigma}} \exp\left[-\frac{1}{2}[x - m_X \; y - m_Y]\Sigma^{-1}[x - m_X \; y - m_Y]^T\right] \qquad (2.11) \\ &= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{(x-m_X)^2}{\sigma_X^2} - \frac{2\rho(x-m_X)(y-m_Y)}{\sigma_X\sigma_Y} + \frac{(y-m_Y)^2}{\sigma_Y^2}\right)\right] \end{aligned}$$

where the role played by the correlation coefficient ρ is evident.

At this stage we present a set of definitions and properties that, even though being valid for any two random variables, X and Y, also apply to the case when the random variables (rv) are Gaussian.

Definition 2.3 Independence

Two random variables X and Y are called independent if the joint pdf, $f(x, y)$ equals the product of the pdf of each random variable, $f(x)$, $f(y)$, i.e.,

$$f(x, y) = f(x)f(y)$$

In the case of Gaussian random variables, clearly X and Y are independent when $\rho = 0$. This issue will be further explored later.

Definition 2.4 Uncorrelatedness

Two random varibles X and Y are called uncorrelated if their covariance is zero, i.e.,

$$\sigma_{XY} = E[(X - m_X)(Y - m_Y)] = 0,$$

which can be written in the following equivalent forms:

$$\rho = 0, \quad E(XY) = E(X)E(Y).$$

Note that

$$E(X + Y) = E(X) + E(Y)$$

but, in general, $E(XY) \neq E(X)E(Y)$. However, when X and Y are uncorrelated, $E(XY) = E(X)E(Y)$ according to Definition 2.4.

Definition 2.5 Orhthogonality

Two random variables X and Y are called orthognal if

$$E(XY) = 0,$$

which is represented as

$$X \perp Y$$

Property 2.1 *If X and Y are uncorrelated, then $X - m_X \perp Y - m_Y$.*

Property 2.2 *If X and Y are uncorrelated and $m_X = 0$ and $m_Y = 0$, then $X \perp Y$.*

Property 2.3 *If two random variables X and Y are independent, then they are uncorrelated, i.e.,*

$$f(x, y) = f(x)f(y) \Rightarrow E(XY) = E(X)E(Y)$$

but the converse is not, in general, true.

Proof: From the definition of mean value,

$$\begin{aligned} E(XY) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(xy)dxdy \\ &= \int_{-\infty}^{\infty} xf(x)dx \int_{-\infty}^{\infty} yf(y)dy = E(X)E(Y). \end{aligned}$$

Property 2.4 *If two Gaussian random variables X and Y are uncorrelated, they are also independent, i.e., for Normal or Gaussian random variables,independency is equivalent to uncorrelatedness. If $X \sim \mathcal{N}(\mu_X, \Sigma_X)$ and $Y \sim \mathcal{N}(\mu_Y, \Sigma_Y)$*

$$f(xy) = f(x)f(y) \quad \Leftrightarrow \quad E(XY) = E(X)E(Y) \Leftrightarrow \rho = 0.$$

Result 2.2 Variance of the sum of two random variables *Let X and Y be two random variables, jointly distributed, with mean m_X and m_Y and correlation coefficient ρ and let*

$$Z = X + Y.$$

Then,

$$E(Z) = m_Z = E(X) + E(Y) = m_X + m_Y \tag{2.12}$$

$$\sigma_Z^2 = E[(Z - m_Z)^2)] = \sigma_X^2 + 2\rho\sigma_X\sigma_Y + \sigma_Y^2. \tag{2.13}$$

Proof: Evaluating the second term in (2.13) yields:

$$\begin{aligned} \sigma_Z^2 &= E[(X - m_X) + (Y - m_Y)]^2 \\ &= E[(X - m_X)^2] + 2E[(X - m_X)(Y - m_Y)] + E[(Y - m_Y)^2] \end{aligned}$$

from where the result immediately holds.

Result 2.3 Variance of the sum of two uncorrelated random variables

Let X and Y be two uncorrelated random variables, jointly distributed, with mean m_X and m_Y and let

$$Z = X + Y.$$

Then,

$$\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2 \tag{2.14}$$

i.e., if two random variables are uncorrelated, then the variance of their sum equals the sum of their variances.

We regain the case of two jointly Gaussian random varaibles, X and Y, with pdf represented by (2.11) to analyze, in the plots of Gaussian pdfs, the influence of the correlation coefficient ρ in the bell-shaped pdf.

Figure 2.1 represents four distinct situations with zero mean and null correlation between X and Y, i.e., $\rho = 0$, but with different values of the standard deviations σ_X and σ_Y. It is clear that, in all cases, the maximum of the pdf is obtained for the mean value. As $\rho = 0$, i.e., the random variables are uncorrelated,

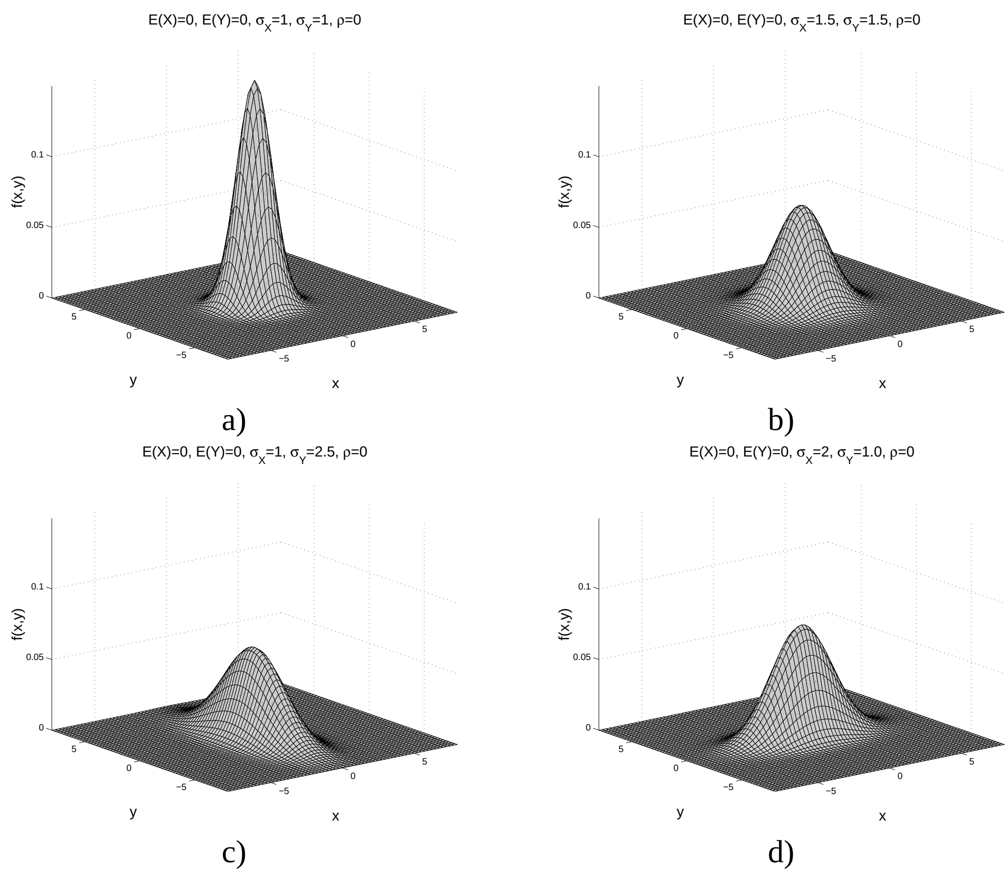

Figure 2.1: Second-order Gaussian pdfs, with $m_X = m_Y = 0$, $\rho = 0$ a) $\sigma_X = 1,\ \sigma_Y = 1$, b) $\sigma_X = 1.5,\ \sigma_Y = 1.5$, c) $\sigma_X = 1,\ \sigma_Y = 2.5$, d) $\sigma_X = 2,\ \sigma_Y = 1$

the change in the standard deviations σ_X and σ_Y has independent effects in each of the components. For example, in Figure 2.1-d) the spread around the mean is greater along the x coordinate. Moreover, the locus of constant value of the pdf is an ellipse with its axis parallel to the x and y axis. This ellipse has equal axis length, i.e, is a circumference, when both random variables, X and Y have the same standard deviation, $\sigma_X = sigma_Y$.

The examples in Figure 2.2 show the influence of the correlation coefficient on the shape of the pdf. What happens is that the axis of the ellipse referred before will no longer be parallel to the axis x and y. The greater the correlation coefficient, the larger the misalignment of these axis. When $\rho = 1$ or $\rho = -1$ the axis of the ellipse has an angle of $\pi/4$ relative to the x-axis of the pdf.

2.2 Locus of constant probability

Similarly to what was considered for a Gaussian random variable, it is also useful for a variety of applications and for a second order Gaussian random vector, to evaluate the locus (x, y) for which the pdf is greater or equal a specified constant, K_1, i.e.,

$$\left\{(x,y) : \frac{1}{2\pi\sqrt{det\Sigma}} \exp\left[-\frac{1}{2}[x - m_X \; y - m_Y]\Sigma^{-1}[x - m_X \; y - m_Y]^T\right] \geq K_1\right\} \tag{2.15}$$

which is equivalent to

$$\left\{(x,y) : [x - m_X \; y - m_Y]\Sigma^{-1} \begin{bmatrix} x - m_X \\ y - m_Y \end{bmatrix} \leq K\right\} \tag{2.16}$$

with

$$K = -2\ln(2\pi K_1 \sqrt{det\Sigma}).$$

Figures 2.3 and 2.4 represent all the pairs (x, y) for which the pdf is less or equal a given specified constant K_1. The locus of constant value is an ellipse with the axis parallel to the x and y coordinates when $\rho = 0$, i.e., when the random variables X and Y are uncorrelated. When $\rho \neq 0$ the ellipse axis are not parallel with the x and y axis. The center of the ellipse coincides in all cases with (m_X, m_Y).

The locus in (2.16) is the border and the inner points of an ellipse, centered in $(m_X,\ m_Y)$. The length of the ellipses axis and the angle they do with the axis x and y are a function of the constant K, of the eigenvalues of the covariance matrix

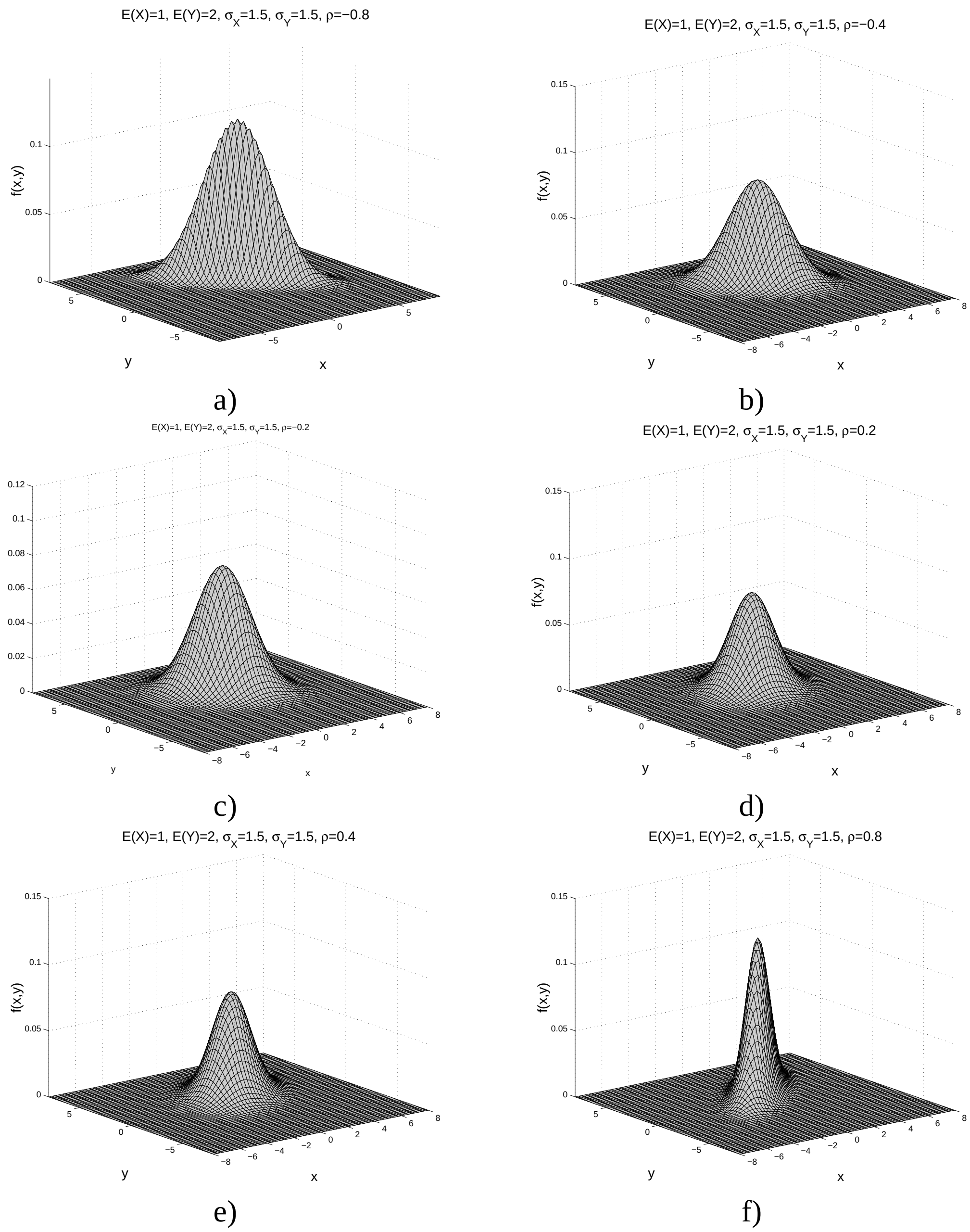

Figure 2.2: Second-order Gaussian pdfs, with $m_X = 1,\ m_Y = 2,\ \sigma_X = 1.5,\ \sigma_Y = 1.5$

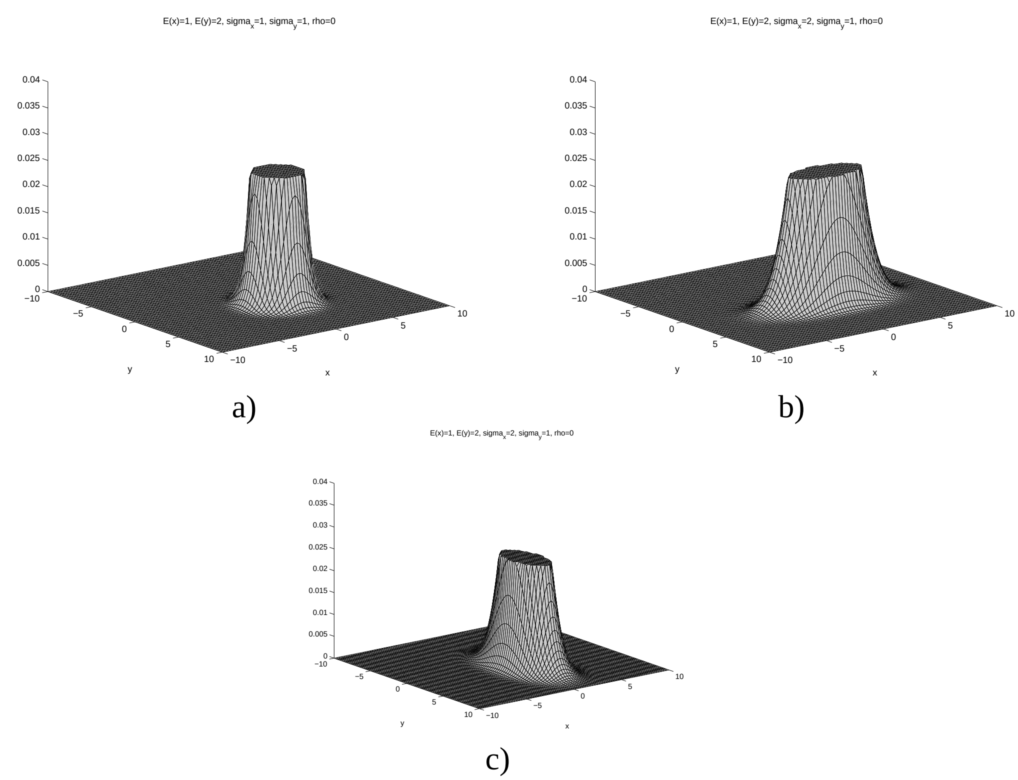

Figure 2.3: Locus of (x, y) such that the Gaussian pdf is less than a constant K for uncorrelated Gaussian random variables with $m_X = 1, \; m_Y = 2$ - a) $\sigma_X = \sigma_Y = 1$, b) $\sigma_X = 2, \; \sigma_Y = 1$, c) $\sigma_X = 1, \; \sigma_Y = 2$

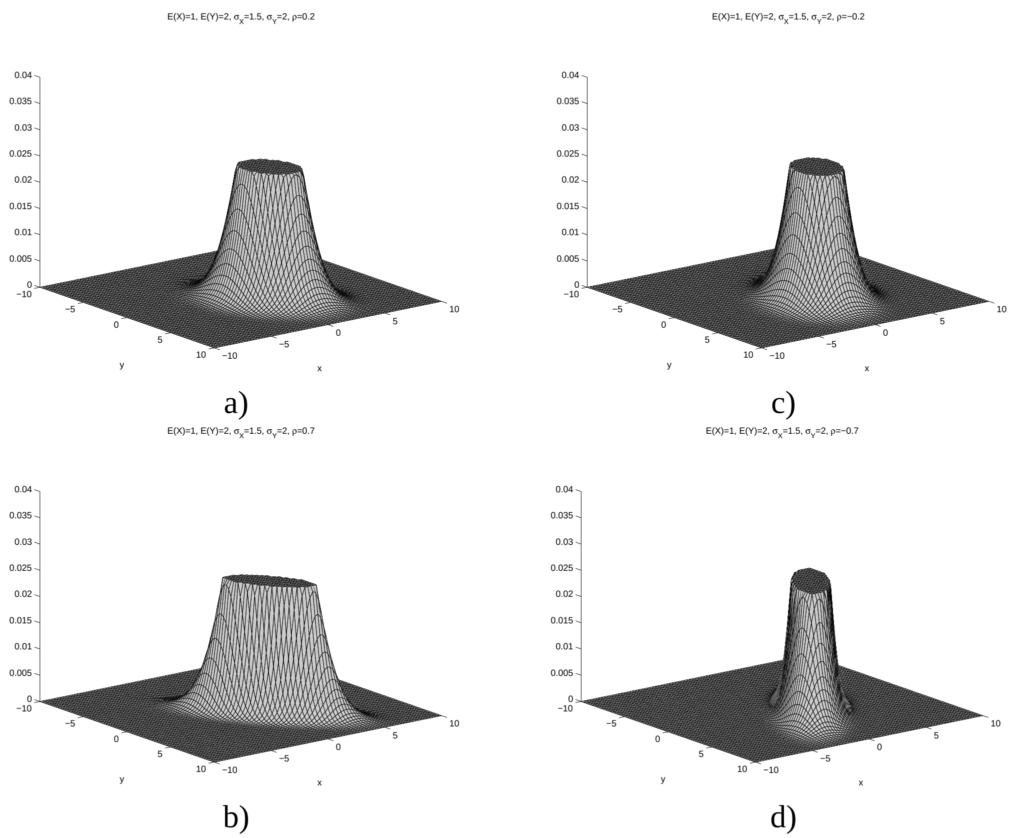

Figure 2.4: Locus of (x, y) such that the Gaussian pdf is less than a constant K for correlated variables . $m_X = 1,\ m_Y = 2,\ \sigma_X = 1.5,\ \sigma_Y = 2$ - a) $\rho = 0.2$, b) $\rho = 0.7$, c) $\rho = -0.2$, d) $\rho = -0.7$

Σ and of the correlation coefficient. We will demonstrate this statement in two different steps. We show that:

1. **Case 1** - if Σ in (2.16) is a diagonal matrix, which happens when $\rho = 0$, i.e., X and Y are uncorrelated, the ellipse axis are parallel to the frame axis.

2. **Case 2** - if Σ in (2.16) is non-diagonal, i.e, $\rho \neq 0$, the ellipse axis are not parallel to the frame axis.

In both cases, the length of the ellipse axis is related with the eigenvalues of the covariance matrix Σ in (2.3) given by:

$$\lambda_1 = \frac{1}{2}\left[\sigma_X^2 + \sigma_Y^2 + \sqrt{(\sigma_X^2 - \sigma_Y^2)^2 + 4\sigma_X^2\sigma_Y^2\rho^2}\right], \tag{2.17}$$

$$\lambda_2 = \frac{1}{2}\left[\sigma_X^2 + \sigma_Y^2 - \sqrt{(\sigma_X^2 - \sigma_Y^2)^2 + 4\sigma_X^2\sigma_Y^2\rho^2}\right]. \tag{2.18}$$

Case 1 - Diagonal covariance matrix

When $\rho = 0$, i.e., the variables X and Y are uncorrelated, the covariance matrix is diagonal,

$$\Sigma = \begin{bmatrix} \sigma_X^2 & 0 \\ 0 & \sigma_Y^2 \end{bmatrix} \tag{2.19}$$

and the eigenvalues particularize to $\lambda_1 = \sigma_X^2$ and $\lambda_2 = \sigma_Y^2$. In this particular case, illustrated in Figure 2.5, the locus (2.16) may be written as

$$\left\{(x, y) : \frac{(x - m_X)^2}{\sigma_X^2} + \frac{(y - m_Y)^2}{\sigma_Y^2} \leq K\right\} \tag{2.20}$$

or also,

$$\left\{(x, y) : \frac{(x - m_X)^2}{K\sigma_X^2} + \frac{(y - m_Y)^2}{K\sigma_Y^2} \leq 1\right\}. \tag{2.21}$$

Figure 2.5 represents the ellipse that is the border of the locus in (2.21) having:

- x-axis with length $2\sigma_X\sqrt{K}$
- y-axis with length $2\sigma_Y\sqrt{K}$.

Case 2 - Non-diagonal covariance matrix

When the covariance matrix Σ in (2.3) is non-diagonal, the ellipse that borders the locus (2.16) has center in (m_X, m_Y) but its axis are not aligned with the

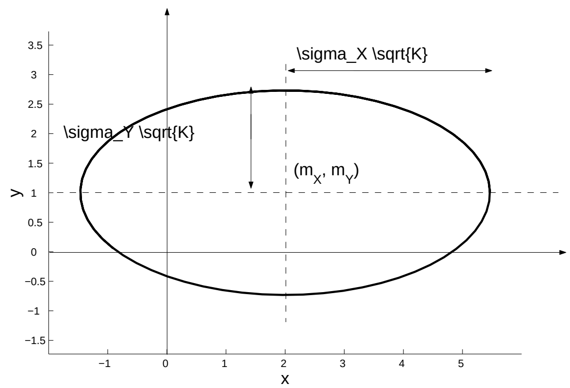

Figure 2.5: Locus of constant pdf: ellipse with axis parallel to the frame axis

coordinate frame. In the sequel we evaluate the angle between the ellipse axis and those of the coordinate frame. With no loss of generality we will consider that $m_X = m_Y = 0$, i.e., the ellipse is centered in the coordinated frame. Therefore, the locus under analysis is given by

$$\left\{(x,y) : [x \ \ y]\Sigma^{-1}\left[\begin{array}{c} x \\ y \end{array}\right] \leq K\right\} \tag{2.22}$$

where Σ is the matrix in (2.3). As it is a symmetric matrix, the eigenvectors corresponding to distinct eigenvalues are orthogonal. When

$$\sigma_X \neq \sigma_Y$$

the eigenvalues (2.17) and (2.18) are distinct, the corresponding eigenvectors are orthogonal and therefore Σ has simple structure which means that there exists a non-singular and unitary coordinate transformation T such that

$$\Sigma = T \ D \ T^{-1} \tag{2.23}$$

where

$$T = [\ v_1 \quad | \ \ v_2 \ ], \quad D = diag(\lambda_1, \ \lambda_2)$$

and v_1, v_2 are the unit-norm eigenvectors of Σ associated with λ_1 and λ_2. Replacing (2.23) in (2.22) yields

$$\left\{(x,y) : [x \quad y]TD^{-1}T^{-1}\left[\begin{array}{c} x \\ y \end{array}\right] \leq K\right\}. \tag{2.24}$$

Denoting

$$\left[\begin{array}{c} w_1 \\ w_2 \end{array}\right] = T^{-1} \left[\begin{array}{c} x \\ y \end{array}\right] \tag{2.25}$$

and given that $T^T = T^{-1}$, it is immediate that (2.24) can be expressed as

$$\left\{(w_1, w_2) : [w_1 \quad w_2] \left[\begin{array}{cc} \lambda_1 & 0 \\ 0 & \lambda_2 \end{array}\right]^{-1} \left[\begin{array}{c} w_1 \\ w_2 \end{array}\right] \leq K\right\} \tag{2.26}$$

that corresponds, in the new coordinate system defined by the axis w_1 and w_2, to the locus bordered by an ellipse aligned with those axis. Given that v_1 and v_2 are unit-norm orthogonal vectors, the coordinate transformation defined by (2.25) corresponds to a rotation of the coordinate system (x, y), around its origin by an angle

$$\alpha = \frac{1}{2} \tan^{-1} \left(\frac{2\rho\sigma_X\sigma_Y}{\sigma_X^2 - \sigma_Y^2} \right), \quad -\frac{\pi}{4} \leq \alpha \leq \frac{\pi}{4}, \quad \sigma_X \neq \sigma_Y. \tag{2.27}$$

Evaluating (2.26), yields,

$$\left\{(w_1, w_2) : \frac{w_1^2}{K\lambda_1} + \frac{w_2^2}{K\lambda_2} \leq 1\right\} \tag{2.28}$$

that corresponds to an ellipse having

- w_1-axis with length $2\sqrt{K\lambda_1}$
- w_2-axis with length $2\sqrt{K\lambda_2}$

as represented in Figure 2.6.

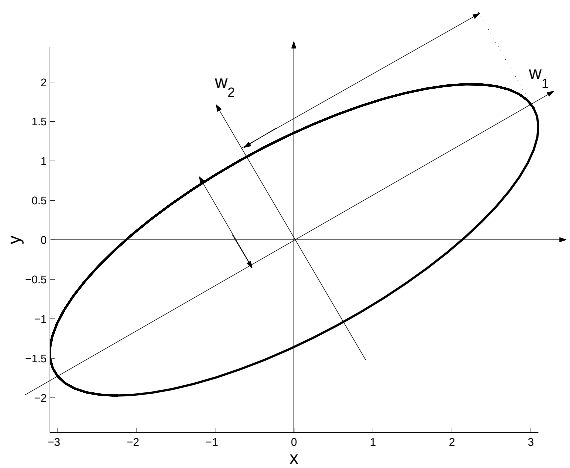

Figure 2.6: Ellipses non-aligned with the coordinate axis x and y.

Chapter 3

Properties

Let X and Y be two jointly distributed Gaussian random vectors, of dimension n and m, respectively, i.e,

$$X \sim \mathcal{N}(m_X, \Sigma_X) \quad Y \sim \mathcal{N}(m_Y, \Sigma_Y)$$

and Σ_X a square matrix of dimension n and Σ_Y a square matrix of dimension m.

Result 3.1 *The conditional pdf of X and Y is given by*

$$f(x|y) = \frac{1}{\sqrt{(2\pi)^n det\Sigma}} \exp\left[-\frac{1}{2}(x-m)^T\Sigma^{-1}(x-m)\right] \sim \mathcal{N}(m, \Sigma) \tag{3.1}$$

with

$$m = E[X|Y] = m_X + \Sigma_{XY}\Sigma_Y^{-1}(Y - m_Y) \tag{3.2}$$

$$\Sigma = \Sigma_X - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_Y X \tag{3.3}$$

The previous result states that, when X and Y are jointly Gaussian, $f(x|y)$ is also Gaussian with mean and covariance matrix given by (3.1) and (3.3), respectively.

Result 3.2 *Let $X \in \mathcal{R}^n$, $Y \in \mathcal{R}^m$ and $Z \in \mathcal{R}^r$ be jointly distributed Gaussian random vectors. If Y and Z are independent, then*

$$E[X|Y,Z] = E[X|Y] + E[X|Z] - m_X \tag{3.4}$$

where $E[X] = m_X$.

Result 3.3 *Let $X \in \mathcal{R}^n$, $Y \in \mathcal{R}^m$ and $Z \in \mathcal{R}^r$ be jointly distributed Gaussian random vectors. If Y and Z are not necessarily independent, then*

$$E[X|Y, Z] = E[X|Y, \tilde{Z}] \tag{3.5}$$

where

$$\tilde{Z} = Z - E[Z|Y]$$

yildeing

$$E[X|Y, Z] = E[X|Y] + E[X|\tilde{Z}] - m_X$$

Chapter 4

Covariance matrices and error ellipsoid

Let X be a n-dimensional Gaussian random vector, with

$$X \sim \mathcal{N}(m_X, \Sigma_X)$$

and consider a constant, $K_1 \in \mathcal{R}$. The locus for which the pdf $f(x)$ is greater or equal a specified constant K_1,i.e.,

$$\left\{ x : \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}[x - m_X]^T \Sigma_X^{-1} [x - m_X] \right] \geq K_1 \right\} \tag{4.1}$$

which is equivalent to

$$\left\{ x : [x - m_X]^T \Sigma_X^{-1} [x - m_X] \leq K \right\} \tag{4.2}$$

with $K = -2\ln((2\pi)^{n/2} K_1 |\Sigma|\ 1/2)$ is an n-dimensional ellipsoid centered at the mean m_X and whose axis are only aligned with the cartesian frame if the covariance matrix Σ is diagonal. The ellipsoid defined by (4.2) is the region of minimum volume that contains a given probability mass under the Gaussian assumption.

When in (4.2) rather than having an inequality there is an equality, (4.2), i.e.,

$$\left\{ x : [x - m_X]^T \Sigma_X^{-1} [x - m_X] = K \right\}$$

this locus may be interpreted as the contours of equal probability.

Definition 4.1 Mahalanobis distance *The scalar quantity*

$$[x - m_X]^T \Sigma_X^{-1} [x - m_X] = K \tag{4.3}$$

is known as the Mahalanobis distance of the vector x to the mean m_X.

The Mahalanobis distance, is a normalized distance where normalization is achieved through the covariance matrix. The surfaces on which K is constant are ellipsoids that are centered about the mean m_X, and whose semi-axis are $\sqrt{K}$ times the eigenvalues of Σ_X, as seen before. In the special case where the random variables that are the components of X are uncorrelated and with the same variance, i.e., the covariance matrix Σ is a diagonal matrix with all its diagonal elements equal, these surfaces are spheres, and the Mahalanobis distance becomes equivalent to the Euclidean distance.

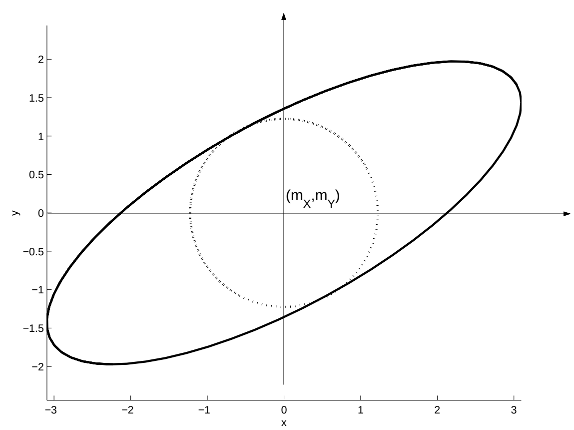

Figure 4.1: Contours of equal Mahalanobis and Euclidean distance around (m_X, m_Y) for a second order Gaussian random vector

Figure 4.1 represents the contours of equal Mahalanobis and Euclidean distance around (m_X, m_Y) for a second order Gaussian random vector. In oher words, any point (x, y) in the ellipse is at the same Mahalanobis distance to the center of the ellipses. Also, any point (x, y) in the circumference is at the same Euclidean distance to the center. This plot enhances the fact that the Mahalanobis distance is weighted by the covariance matrix.

For decision making purposes (e.g., the field-of-view, a validation gate), and given m_X and Σ_X, it is necessary to determine the probability that a given vector will lie within, say, the 90% confidence ellipse or ellipsoid given by (4.3). For a given K, the relationship between K and the probability of lying within the ellipsoid specified by K is, [3],

$$\begin{array}{ll} n=1; & Pr\{\text{x inside the ellipsoid}\} = -\frac{1}{\sqrt{2\pi}} + 2erf(\sqrt{K}) \\ n=2; & Pr\{\text{xinside the ellipsoid}\} = 1 - e^{-K/2} \\ n=3; & Pr\{\text{x inside the ellipsoid}\} = -\frac{1}{\sqrt{2\pi}} + 2erf(\sqrt{K}) - \sqrt{\frac{2}{\pi}}\sqrt{K}e^{-K/2} \end{array} \tag{4.4}$$

where n is the dimension of the random vector. Numeric values of the above expression for $n = 2$ are presented in the following table

Probability	K
50%	1.386
60%	1.832
70%	2.408
80%	3.219
90%	4.605

For a given K the ellispoid axis are fixed. The probability that a given value of the random vector X lies within the ellipsoid centered in the mean value, increases with the increase of K.

This problem can be stated the other way around. In the case where we specify a fixed probability value, the question is the value of K that yields an ellipsoid satisfying that probability. To answer the question the statistics of K has to be analyzed.

The scalar random variable (4.3) has a known random distribution, as stated in the following result.

Result 4.1 *Given the n-dimensional Gaussian random vector X, with mean m_X and covariance matrix Σ_X, the scalar random variable K defined by the quadratic form*

$$[x - m_X]^T \Sigma_X^{-1} [x - m_X] = K \tag{4.5}$$

has a chi-square distribution with n degrees of freedom.

Proof: see, p.e., in [1].

The pdf of K in (4.5), i.e., the chi-square density with n degrees of freedom is, (see, p.e., [1])

$$f(k) = \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} k^{\frac{n-2}{2}} \exp^{-\frac{k}{2}}$$

where the gamma function satisfies,

$$\Gamma(\frac{1}{2}) = \sqrt{\pi}, \quad \Gamma(1) = 1 \quad \Gamma(n+1) = \Gamma(n).$$

The probability that the scalar random variable, K in (4.5) is less or equal a given constant, χ_p^2

$$Pr\{K \leq \chi_p^2\} = Pr\{[x - m_X]^T \Sigma^{-1} [x - m_X] \leq \chi_p^2\} = p$$

is given in the following table where n is the number of degrees of freedom and the sub-indice p in χ_p^2 represents the corresponding probability under evaluation.

n	$\chi^2_{0.995}$	$\chi^2_{0.99}$	$\chi^2_{0.975}$	$\chi^2_{0.95}$	$\chi^2_{0.90}$	$\chi^2_{0.75}$	$\chi^2_{0.50}$	$\chi^2_{0.25}$	$\chi^2_{0.10}$	$\chi^2_{0.05}$
1	7.88	6.63	5.02	3.84	2.71	1.32	0.455	0.102	0.0158	0.0039
2	10.6	9.21	7.38	5.99	4.61	2.77	1.39	0.575	0.211	0.103
3	12.8	11.3	9.35	7.81	6.25	4.11	2.37	1.21	0.584	0.352
4	14.9	13.3	11.1	9.49	7.78	5.39	3.36	1.92	1.06	0.711

From this table we can conclude, for example, that for a third-order Gaussian random vector, $n = 3$,

$$Pr\{K \leq 6.25\} = Pr\{[x - m_X]^T \Sigma^{-1} [x - m_X] \leq 6.25\} = 0.9$$

Example 4.1 Mobile robot localization and the error ellipsoid

This example illustrates the use of the error ellipses and ellipsoids in a particular application, the localization of a mobile robot operating in a given environment.

Consider a mobile platform, moving in an environment and let $P \in \mathcal{R}^2$ be the position of the platform relative to a world frame. P has two components,

$$P = \begin{bmatrix} X \\ Y \end{bmatrix}.$$

The exact value of P is not known and we have to use any particular localization algorithm to evaluate P. The most common algorithms combine internal and external sensor measurements to yield an estimated value of P.

The uncertainty associated with all the quantities involved in this procedure, namely vehicle model, sensor measurements, environment map representation, leads to consider P as a random vector. Gaussianity is assumed for simplicity. Therefore, the localization algorithm provides an estimated value of P, denoted *as $\hat{P}$, which is the mean value of the Gaussian pdf, and the associated covariance matrix, i.e.,*

$$P \sim \mathcal{N}(\hat{P}, \Sigma_P)$$

At each time step of the algorithm we do not know the exact value of P, but we have an estimated value, $\hat{P}$ and a measure of the uncertainty of this estimate, given by Σ_P. The evident question is the following: "Where is the robot?", i.e., "What is the exact value of P"? It is not possible to give a direct answer to this question, but rather a probabilistic one. We may answer, for example: "Given $\hat{P}$ and Σ_P, with 90% of probability, the robot is located in an ellipse centered in $\hat{P}$ and whose border is defined according to the Mahalanobis distance". In this case the value of K in (4.5) will be $K = 4.61$.

Someone may say that, for the involved application, a probability of 90% is small and ask to have an answer with an higher probability, for example 99%. The answer will be similar but, in this case, the error ellipse, will be defined for $K = 9.21$, i.e.,the ellipse will be larger than the previous one.

Bibliography

[1] Yaakov Bar-Shalom, X. Rong Li, Thiagalingam Kirubarajan, ”Estimation with Applications to Tracking and Navigation,” John Wiley & Sons, 2001.

[2] A. Papoulis, ”Probability, Random Variables and Stochastic Processes,” McGraw-Hill, 1965.

[3] Randall C. Smith, Peter Cheeseman, ”On the Representation and Estimation of Spatial Uncertainty,” the International Journal of Robotics Research, Vol.5, No.4, Winter 1986.