Lecture 9: Gaussian Random Vectors

Random Vectors

The idea of scalar random variables can be extended to a collection of multiple random variables called a *random vector*. Consider a continuous random vector $\boldsymbol{x} : \Omega \to \mathbb{R}^2$ that maps from the sample space Ω to the two-dimensional plane. The vector has two components $\boldsymbol{x}(\omega) = [x_1(\omega), x_2(\omega)]^{\mathrm{T}}$. Each component may be thought of as an r.v. $x_1 : \Omega \to \mathbb{R}$ and $x_2 : \Omega \to \mathbb{R}$. Moreover, each random variable has an expected value $E[x_1] = \mu_{x_1}$ and $E[x_2] = \mu_{x_2}$ so that the expected value of the random vector $\boldsymbol{x}$ is

$$\boldsymbol{\mu}_x = E[\boldsymbol{x}] = \begin{pmatrix} \mu_{x_1} \\ \mu_{x_2} \end{pmatrix} . \tag{1}$$

If the r.v.s x_1 and x_2 have variances $\sigma^2_{x_1}$ and $\sigma^2_{x_2}$, respectively, then what is the variance of $\boldsymbol{x}$? To answer this question we must introduce some additional concepts.

Joint, cumulative, and marginal probabilities. Consider the event $A = \{\omega \in \Omega : [x_1(\omega), x_2(\omega)] \in B \subseteq \mathbb{R}^2\}$. That is, the event A is the set of samples in the sample space that map through the r.v.'s x_1 and x_2 to a subset of the plane B. The probability of event A is $P(A)$ and is denoted

$$P(A) = P[(x_1, x_2) \in B] = P_{x_1,x_2}(B) \tag{2}$$

where $P_{x_1,x_2}(B)$ is called the *joint probability distribution function* of the two r.v.s x_1 and x_2. The function $P_{x_1,x_2}(B)$ assigns a probability to each set B. However, for the particular choice of the rectangular set

$$B_r(x_1, x_2) = \{(s_1, s_2) \in \mathbb{R}^2 \ : \ s_1 \leq x_1, s_2 \leq x_2\} . \tag{3}$$

$$= (-\infty, x_1] \times (-\infty, x_2] \tag{4}$$

we call

$$P_{x_1,x_2}(B_r(x_1, x_2)) = F_{x_1,x_2}(x_1, x_2) \tag{5}$$

the *joint cumulative distribution function* (joint c.d.f.). If the set $B_p(x_1, x_2) = (x_1, x_2) \in \mathbb{R}^2$ is a point in the plane then we call the probability

$$P_{x_1,x_2}(B_p(x_1, x_2)) = f_{x_1,x_2}(x_1, x_2) \tag{6}$$

the *joint probability density function* (joint p.d.f.). The function (6) can be thought of as a surface defined over the plane $\mathbb{R}^2$. The volume underneath this surface integrates to one

$$\int_{\mathbb{R}^2} f_{x_1,x_2}(x_1, x_2) = 1 . \tag{7}$$

The joint c.d.f. (5) and the joint p.d.f. (6) are related by the integral

$$F_{x_1,x_2}(x_1, x_2) = \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f_{x_1,x_2}(s_1, s_2) ds_1 ds_2 . \tag{8}$$

If we choose $x_2 = \infty$, the set $B_r(x_1, \infty)$ encompasses all possible of x_2 values and only the subset $(-\infty, x_1]$ of possible x_1 values. Evaluating (8) effectively integrates out the r.v. x_2 and describes only the distribution of x_1, as in the single r.v. case. We call these *marginal c.d.f.s*:

$$F_{x_1}(x_1) = F_{x_1,x_2}(x_1, \infty) \tag{9}$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{x_1} f_{x_1,x_2}(s_1, s_2) ds_1 ds_2 \tag{10}$$
$$F_{x_2}(x_2) = F_{x_1,x_2}(\infty, x_2) \tag{11}$$
$$= \int_{-\infty}^{x_2} \int_{-\infty}^{\infty} f_{x_1,x_2}(s_1, s_2) ds_1 ds_2 \ . \tag{12}$$

Similarly, the *marginal p.d.f.s* are defined by integrating out one of the variables in (6) and leaving the other one free

$$f_{x_1}(x_1) = \int_{-\infty}^{\infty} f_{x_1,x_2}(s_1, s_2) ds_2 \tag{13}$$
$$f_{x_2}(x_2) = \int_{-\infty}^{\infty} f_{x_1,x_2}(s_1, s_2) ds_1 \ . \tag{14}$$

In terms of these joint and marginal p.d.f.s, one can also define the *conditional p.d.f.*

$$f_{x_1|x_2=x_2^*}(x_1) = \frac{f_{x_1,x_2}(x_1, x_2^*)}{f_{x_2}(x_2^*)} \tag{15}$$

where x_1 can vary and x_2^* is fixed. We can interpret the numerator $f_{x_1,x_2}(x_1, x_2^*)$ in (15) as a slice of the joint p.d.f $f_{x_1,x_2}(x_1, x_2)$ cut at the the location $x_2 = x_2^*$. The numerator $f_{x_2}(x_2^*)$ is the integral of this slice (less than one). By dividing the curve $f_{x_1,x_2}(x_1, x_2^*)$ by $f_{x_2}(x_2^*)$ we obtain a new normalized curve that integrates to one. Hence, $f_{x_1|x_2=x_2^*}(x_1)$ is a valid p.d.f. Recall that when describing conditional p.d.f.s we interpret the vertical line "|" as meaning "given". So $f_{x_1|x_2=x_2^*}(x_1)$ is the p.d.f of x_1 given x_2 is equal to x_2^*. Similarly, holding x_1^* fixed and allowing x_2 to vary gives the conditional p.d.f. of x_2 given x_1

$$f_{x_2|x_1=x_1^*}(x_2) = \frac{f_{x_1,x_2}(x_1^*, x_2)}{f_{x_1}(x_1^*)} \ . \tag{16}$$

The *conditional expectation* is the expected value of one r.v. given the value of another r.v.

$$E[x_1|x_2 = x_2^*] = \int_{-\infty}^{\infty} s_1 f_{x_1|x_2=x_2^*}(s_1) ds_1 \tag{17}$$
$$E[x_2|x_1 = x_1^*] = \int_{-\infty}^{\infty} s_2 f_{x_2|x_1=x_1^*}(s_2) ds_2 \tag{18}$$

Suppose $g : \mathbb{R}^2 \to \mathbb{R}$ is a function that maps two scalars x_1 and x_2 into a real number $g(x_1, x_2)$. If we replace the scalars with r.v.s x_1 and x_2 then $g(x_1, x_2) : \Omega \to \mathbb{R}$ can be considered a r.v. itself. The expectation of $g(x_1, x_2)$ is

$$E[g(x_1, x_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(s_1, s_2) f_{x_1,x_2}(s_1, s_2) ds_1 ds_2 \ . \tag{19}$$

Now consider the particular case of $g(x_1, x_2) = (x_1 - E[x_1])(x_2 - E[x_2])$. The expectation of this function is termed the *covariance*

$$\text{Cov}(x_1, x_2) = \sigma_{x_1,x_2} = E[(x_1 - E[x_1])(x_2 - E[x_2])] \tag{20}$$

Introducing the constants $\mu_{x_1} = E[x_1]$ and $\mu_{x_2} = E[x_2]$ and using (19) with (20)

$$\begin{aligned}
\text{Cov}(x_1, x_2) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[(s_1 - \mu_{x_1})(s_2 - \mu_{x_2})f_{x_1,x_2}(s_1, s_2)ds_1 ds_2 && (21)\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(s_1 s_2 - s_1\mu_{x_2} - \mu_{x_1}s_2 + \mu_{x_1}\mu_{x_2})f_{x_1,x_2}(s_1, s_2)ds_1 ds_2 && (22)\\
&= \underbrace{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(s_1 s_2)f_{x_1,x_2}(s_1, s_2)ds_1 ds_2}_{E[x_1 x_2]} - \mu_{x_2}\underbrace{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}s_1 f_{x_1,x_2}(s_1, s_2)ds_1 ds_2}_{\mu_{x_1}}\\
&\quad - \mu_{x_1}\underbrace{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}s_2 f_{x_1,x_2}(s_1, s_2)ds_1 ds_2}_{\mu_{x_2}} + \mu_{x_1}\mu_{x_2}\underbrace{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f_{x_1,x_2}(s_1, s_2)ds_1 ds_2}_{=1} && (23)\\
&= E[x_1 x_2] - \mu_{x_1}\mu_{x_2} && (24)
\end{aligned}$$

Note the result above can also be obtained by using the linearity of the expectation operator to distribute $E[\cdot]$ in (20).

$$\begin{aligned}
\text{Cov}(x_1, x_2)\sigma_{x_1,x_2} &= E[(x_1 - E[x_1])(x_2 - E[x_2])] && (25)\\
&= E[(x_1 x_2 - x_1 E[x_2] - E[x_1]x_2 + E[x_1]E[x_2]) && (26)\\
&= E[x_1 x_2 - x_1\mu_{x_2} - \mu_{x_1}x_2 + \mu_{x_1}\mu_{x_2}] && (27)\\
&= E[x_1 x_2] - E[x_1]\mu_{x_2} - \mu_{x_1}E[x_2] + \mu_{x_1}\mu_{x_2} && (28)\\
&= E[x_1 x_2] - \mu_{x_1}\mu_{x_2} - \mu_{x_1}\mu_{x_2} + \mu_{x_1}\mu_{x_2} && (29)\\
&= E[x_1 x_2] - \mu_{x_1}\mu_{x_2} && (30)
\end{aligned}$$

It can also be shown that $\text{Cov}(x_1, x_2) = \sigma_{x_1,x_2} = \text{Cov}(x_2, x_1) = \sigma_{x_2,x_1}$.

If knowledge of the value x_1 gives no information about x_2's value the elements are *uncorrelated* and $\sigma_{x_1,x_2} = 0$. From (24), $\sigma_{x_1,x_2} = 0$ implies that $E[x_1 x_2] = \mu_{x_1}\mu_{x_2}$ which is only true if the joint p.d.f is equal to the product of the individual p.d.f.s $f_{x_1,x_2}(x_1, x_2) = f_{x_1}(x_1)f_{x_2}(x_2)$ in the case of independent r.v.s. In general, if the two r.v.s are correlate, and if a larger x_1 tends to yield a larger x_2, then $\sigma_{x_1,x_2} > 0$ and the r.v.s are positively correlated. Conversely, if smaller x_1 tends to yield a larger x_2, then $\sigma_{x_1,x_2} < 0$ and the r.v.s are negatively correlated. The magnitude of σ_{x_1,x_2} indicates how strongly these two random variables are correlated.

Note that the definitions presented in this section can be extended to more than two r.v.s in a straightforward manner, and we have made no assumptions thusfar on the nature of the p.d.f.s themselves. In the following section we consider arbitrary sized collections of random variables that are restricted to be Gaussian. The approach is facilitated by using vector/matrix notation.

Gaussian Random Vectors

Much like the p.d.f. of a Gaussian r.v. x is defined by its mean μ and variance σ^2, the p.d.f of a Gaussian random vector $\boldsymbol{x}$ is defined by the vector mean $\boldsymbol{\mu}$ and a *covariance matrix* $\boldsymbol{P}_x$. For, example if the random vector $\boldsymbol{x} = [x_1, x_2]^{\text{T}}$ then the mean is

$$\boldsymbol{\mu}_x = E[\boldsymbol{x}] = \begin{pmatrix} E[x_1] \\ E[x_2] \end{pmatrix} = \begin{pmatrix} \mu_{x_1} \\ \mu_{x_2} \end{pmatrix}. \quad (31)$$

and the covariance matrix is

$$\begin{aligned}
\boldsymbol{P}_x &= E[(\boldsymbol{x} - E[\boldsymbol{x}])(\boldsymbol{x} - E[\boldsymbol{x}])^{\mathrm{T}}] && (32)\\
&= E\left[\begin{bmatrix} x_1 - \mu_{x_1} \\ x_2 - \mu_{x_2} \end{bmatrix} \begin{bmatrix} x_1 - \mu_{x_1} & x_2 - \mu_{x_2} \end{bmatrix}\right] && (33)\\
&= \begin{bmatrix} E[(x_1 - \mu_{x_1})^2] & E[(x_1 - \mu_{x_1})(x_2 - \mu_{x_2})] \\ E[(x_2 - \mu_{x_2})(x_1 - \mu_{x_1})] & E[(x_2 - \mu_{x_2})^2] \end{bmatrix} && (34)\\
&= \begin{bmatrix} \sigma_{x_1}^2 & \sigma_{x_1,x_2} \\ \sigma_{x_2,x_1} & \sigma_{x_2}^2 \end{bmatrix}. && (35)
\end{aligned}$$

where the last line follows from our definition of variance and covariance (20). This expression is equivalently obtained in vector form:

$$\begin{aligned}
\boldsymbol{P}_x &= E[(\boldsymbol{x} - E[\boldsymbol{x}])(\boldsymbol{x} - E[\boldsymbol{x}])^{\mathrm{T}}] \\
&= E[(\boldsymbol{x} - E[\boldsymbol{x}])(\boldsymbol{x}^{\mathrm{T}} - E[\boldsymbol{x}^{\mathrm{T}}])] \\
&= E[(\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}} - \boldsymbol{x}E[\boldsymbol{x}^{\mathrm{T}}] - E[\boldsymbol{x}]\boldsymbol{x}^{\mathrm{T}} + E[\boldsymbol{x}]E[\boldsymbol{x}^{\mathrm{T}}]] \\
&= E[\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}] - E[\boldsymbol{x}\boldsymbol{\mu}_x^{\mathrm{T}}] - E[\boldsymbol{\mu}_x\boldsymbol{x}^{\mathrm{T}}] + E[\boldsymbol{\mu}_x\boldsymbol{\mu}_x^{\mathrm{T}}] \\
&= E[\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}] - E[\boldsymbol{x}]\boldsymbol{\mu}_x^{\mathrm{T}} - \boldsymbol{\mu}_x E[\boldsymbol{x}^{\mathrm{T}}] + \boldsymbol{\mu}_x\boldsymbol{\mu}_x^{\mathrm{T}} \\
&= E[\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}] - \boldsymbol{\mu}_x\boldsymbol{\mu}_x^{\mathrm{T}} - \boldsymbol{\mu}_x\boldsymbol{\mu}_x^{\mathrm{T}} + \boldsymbol{\mu}_x\boldsymbol{\mu}_x^{\mathrm{T}} \\
&= E[\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}] - \boldsymbol{\mu}_x\boldsymbol{\mu}_x^{\mathrm{T}}
\end{aligned}$$

The matrix (35) includes the variances of the individual random variables $\sigma_{x_1}^2$ and $\sigma_{x_2}^2$ on the diagonal, and the covariances $\sigma_{x_1,x_2} = \sigma_{x_2,x_1}$ from (20) on the off-diagonal. Given $\boldsymbol{\mu}$ and $\boldsymbol{P}$ the p.d.f of the Gaussian random vector $\boldsymbol{x}$ is

$$f_x(\boldsymbol{x}) = \frac{\exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_x)^{\mathrm{T}}\boldsymbol{P}_x^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_x)\right)}{(2\pi)^{n/2}\sqrt{|\boldsymbol{P}_x|}} \tag{36}$$

where the $|\cdot|$ operator denotes the determinant and n is the dimension of the vector $\boldsymbol{x}$. A Gaussian random vector is denoted by $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{P}_x)$.

Log of a Gaussian p.d.f.

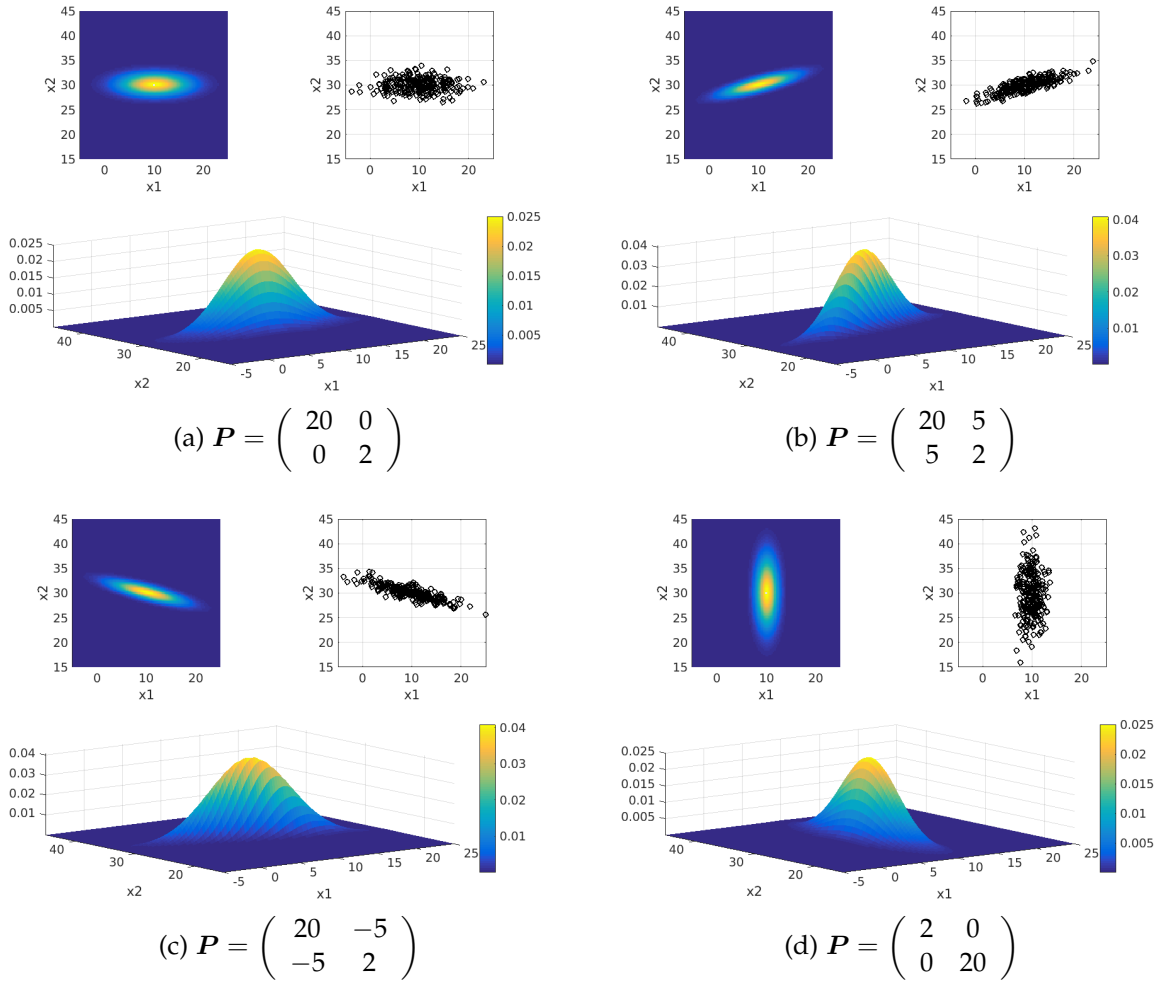
A common operation involves taking the log of (36) which can be computed as follows:

$$\begin{aligned}
\log f_x(\boldsymbol{x}) &= \log\left[\frac{\exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_x)^{\mathrm{T}}\boldsymbol{P}_x^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_x)\right)}{(2\pi)^{n/2}\sqrt{|\boldsymbol{P}_x|}}\right] \\
&= \log\left[\exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_x)^{\mathrm{T}}\boldsymbol{P}_x^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_x)\right)\right] - \log\left[(2\pi)^{n/2}\sqrt{|\boldsymbol{P}_x|}\right] \\
&= \left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_x)^{\mathrm{T}}\boldsymbol{P}_x^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_x)\right) - \frac{1}{2}\log|\boldsymbol{P}_x| - \frac{n}{2}\log 2\pi
\end{aligned}$$

Examples of Gaussian Random Vector Densities

Consider the numerical examples shown in Fig. 1 in which various joint p.d.f.s are plotted (using Eq. 36). Each distribution has the same mean $\boldsymbol{\mu} = (10 \;\; 30)^{\mathrm{T}}$, but a varying covariance matrix

$\boldsymbol{P}$. In addition to plotting the (continuous) probability distribution, a set of 250 samples are drawn from each distribution (indicated by black markers). Figure 1 illustrates how the four entires of the covariance matrix $\boldsymbol{P}$ determine the shape of the joint p.d.f. and the distribution of the samples. The term $\sigma_{x_1}^2$ determines how spread out the samples are in the x_1 direction. Similarly, $\sigma_{x_1}^2$ determines how spread out the samples are in the x_2 direction. The off-diagonal terms indicate in which direction the distribution is slanted. In the case, $\sigma_{x_1,x_2} = 5$ the two variables are positively correlated (as one increases, so does the other). In the case $\sigma_{x_1,x_2} = -5$ the opposite is true.



(a) $\boldsymbol{P} = \begin{pmatrix} 20 & 0 \\ 0 & 2 \end{pmatrix}$

(b) $\boldsymbol{P} = \begin{pmatrix} 20 & 5 \\ 5 & 2 \end{pmatrix}$

(c) $\boldsymbol{P} = \begin{pmatrix} 20 & -5 \\ -5 & 2 \end{pmatrix}$

(d) $\boldsymbol{P} = \begin{pmatrix} 2 & 0 \\ 0 & 20 \end{pmatrix}$

Figure 1: Probability distributions with mean $\boldsymbol{\mu} = (\mu_{x_1} \ \mu_{x_2})^{\mathrm{T}} = (10 \ \ 30)^{\mathrm{T}}$

Confidence Ellipsoids and Mahalanobis Distance

Observe that the multivariate pdf (36) is a function of the vector $\boldsymbol{x}$ that only appears within the exponential term which we denote

$$\Delta = \sqrt{(\boldsymbol{x}-\boldsymbol{\mu}_x)^{\mathrm{T}}\boldsymbol{P}_x^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_x)}\,. \tag{37}$$

This quantity is known as the *Mahalanobis distance* from $\boldsymbol{\mu}_x$ to $\boldsymbol{x}$ and quantifies how many standard deviations away $\boldsymbol{x}$ is from the mean $\boldsymbol{\mu}_x$. If the covariance matrix is the identity matrix then $\boldsymbol{\Delta}$ reduces to Euclidean distance.

Let $\tilde{\boldsymbol{x}} = \boldsymbol{x} - \boldsymbol{\mu}_x$ denote the deviation from the mean. Then the equation

$$\Delta = \sqrt{\tilde{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{P}_x^{-1}\tilde{\boldsymbol{x}}} = c \tag{38}$$

represents contours of the Gaussian pdf that are equal to some constant probability c. We can select c to define an ellipse with a certain *containment probability* p_c by selecting $c = \sqrt{-2\ln(1-p_c)}$ [1, Sec. 14.1.4]. Often a 90% confidence (containment probability) is selected ($p_c = 0.9$) in which case the area enclosed by the ellipse represents a 90% probability of containing the random vector.

We will now show that the geometry of this contour is an n-dimensional ellipsoid. First, factor the covariance matrix as

$$\boldsymbol{P}_x = \boldsymbol{M}\boldsymbol{\Lambda}\boldsymbol{M}^{\mathrm{T}} \tag{39}$$

where $\boldsymbol{M} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n]$ is the *modal matrix* of eigenvectors and $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ is the *spectral matrix* containing diagonal matrix of eigenvalues where $\lambda_i \boldsymbol{v}_i = \boldsymbol{P}_x \boldsymbol{v}_i$ for $i = 1, 2, \ldots, n$.

> **Aside:** For a given matrix $\boldsymbol{P}$ we can compute the modal and spectral matrices in Matlab using `[M,L] = eig(P)`. The eigenvalues in L are arranged from smallest to largest.

Introduce a new variable $\boldsymbol{z} = \boldsymbol{M}^{\mathrm{T}}\tilde{\boldsymbol{x}}$. Then,

$$\begin{aligned}
c^2 &= \tilde{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{P}_x^{-1}\tilde{\boldsymbol{x}} && (40)\\
&= \tilde{\boldsymbol{x}}^{\mathrm{T}}(\boldsymbol{M}\boldsymbol{\Lambda}\boldsymbol{M}^{\mathrm{T}})^{-1}\tilde{\boldsymbol{x}} && (41)\\
&= \tilde{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{M}^{-1}\boldsymbol{\Lambda}^{-1}(\boldsymbol{M}^{\mathrm{T}})^{-1}\tilde{\boldsymbol{x}} && (42)\\
&= \boldsymbol{z}^{\mathrm{T}}\boldsymbol{\Lambda}\boldsymbol{z} && (43)\\
&= \begin{bmatrix} z_1 & z_2 & \cdots & z_n \end{bmatrix} \begin{bmatrix} 1/\lambda_1 & 0 & \cdots & 0 \\ 0 & 1/\lambda_1 & \cdots & \vdots \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & 1/\lambda_n \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} && (44)\\
\implies c^2 &= \frac{z_1^2}{\lambda_1} + \frac{z_2^2}{\lambda_2} + \cdots + \frac{z_n^2}{\lambda_n} && (45)
\end{aligned}$$

which is the equation of an n-dimensional ellipsoid. In the above equation we made use of the fact that the inverse of a diagonal matrix (i.e., $\boldsymbol{\Lambda}^{-1}$) is simply the inverse of each diagonal element.

Aside: The generic equation for an ellipse center on the origin is

$$\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 = 1 \tag{46}$$

where $2a$ is called the major axis and $2b$ is the minor axis. This expression can be generalized to n dimensions to define an n-dimensional ellipsoid.

From (45) we see that the square roots of the eigenvalues are the lengths of the ellipsoid axes. In the z coordinates the ellipse is unrotated (i.e., aligned with the $z_1, \ldots, z_n$ axis. However, in the original x coordinates the columns of $\boldsymbol{M}$ (eigenvectors) define the directions of the principle axes of the ellipsoid. In two dimensions, we can determine the angle that the major axis makes as follows. Let $\lambda^{\max}$ denote the largest eigenvalue and $\boldsymbol{v}^{\max} = [u_1, u_2]^{\mathrm{T}}$ be the corresponding eigenvector. The angle that this eigenvector makes with the x_1 axis is $\psi = \operatorname{atan}(u_2/u_1)$. The parametric equation (x_e, y_e) describing a two-dimensional ellipse that is rotated by an angle ψ and translated to a point $(\hat{x}_1, \hat{x}_2)$ with largest eigenvalue $\lambda^{\max}$ and smallest eigenvalue $\lambda^{\min}$ is

$$x_e(\theta) = \hat{x}_1 + c\sqrt{\lambda^{\max}}\cos\psi\cos\theta - c\sqrt{\lambda^{\min}}\sin\psi\sin\theta \tag{47}$$

$$y_e(\theta) = \hat{y}_1 + c\sqrt{\lambda^{\max}}\sin\psi\cos\theta + c\sqrt{\lambda^{\min}}\cos\psi\sin\theta \tag{48}$$

where $\theta \in [0, 2\pi]$.

References

[1] Anton J Haug. *Bayesian Estimation and Tracking: a Practical Guide*. John Wiley & Sons, 2012.

[2] Bruce P Gibbs. *Advanced Kalman Filtering, Least-squares and Modeling*. John Wiley & Sons, 2011.