On the Distribution of the Distance Between Two Multivariate Normally Distributed Points

Houssain Kettani[1] and George Ostrouchov[2]

[1]Department of Electrical and Computer Engineering and Computer Science
Polytechnic University of Puerto Rico
San Juan, PR 00919
hkettani@pupr.edu

[2]Statistics and Data Sciences Group
Computer Science and Mathematics Division
Oak Ridge National Laboratory
Oak Ridge, TN 37831
ostrouchovg@ornl.gov

Abstract

Motivated by the problem of cluster identification, we consider two multivariate normally distributed points. We seek to find the distribution of the squared Euclidean distance between these two point. Consequently, we find the corresponding distribution in the general case. We then reduce this distribution for special cases based on the mean and covariance.

1 Introduction and Formulation

Suppose we have two clusters of points C_i and C_j, in n-dimensional space. Each point in a cluster is drawn from the same multivariate normal distribution with known mean and covariance. An n-dimensional random variable $\mathbf{x}$ has a multivariate normal (or Gaussian) distribution with mean $\mathbf{m}$ and covariance matrix $\mathbf{R}$ if it has the following probability density function (pdf):

$$f(\mathbf{x}) = (2\pi)^{-n/2}|\mathbf{R}|^{-1/2}exp\{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T\mathbf{R}^{-1}(\mathbf{x}-\mathbf{m})\}.$$

Such random variable is denoted by $\mathbf{x} \sim N(\mathbf{m}, \mathbf{R})$.

Consequently, we pose the following question: given a point x_i in cluster C_i, and point x_j in cluster C_j, what is the distribution of the squared distance between the two points. To this end, the squared

Euclidean distance between two n-dimensional points $\mathbf{x_i}$ and $\mathbf{x_j}$ is given by

$$d_{ij}^2 \doteq (\mathbf{x_i} - \mathbf{x_j})^T(\mathbf{x_i} - \mathbf{x_j}),$$

where "^{T}" denotes the transpose operator.

Let us now write

$$\mathbf{z} \doteq (\mathbf{x_i} - \mathbf{x_j}), \tag{1}$$

then,

$$d_{ij}^2 = \mathbf{z}^T\mathbf{z} = \sum_{k=1}^{n} z_k^2,$$

were z_k are the entries of $\mathbf{z}$.

The distance d_{ij} is also referred to as the amplitude of the random variable $\mathbf{z}$, and its distribution is referred to as the amplitude probability distribution function (APDF). This notion is of particular interest in wave, antenna, and signal analysis [9, Appendix E], and Electromyography (EMG) [12]. The results in this paper is also useful in the areas of cluster identification [13] and dimension reduction [4].

In the following section, we present the distribution of the squared distance. We then consider special cases in Section 3, where the distribution is simplified to more familiar distributions. We end this paper by a concluding remarks in Section 4. An index of notations is presented at the end of this paper to facilitate understanding and look up of notations used in this paper.

2 Distance Distribution

Let $\mathbf{x_i}$ and $\mathbf{x_i}$ be two $n \times 1$ random vectors with $\mathbf{x_i} \sim N(\mathbf{m_i}, \mathbf{R_i})$ and $\mathbf{x_j} \sim N(\mathbf{m_j}, \mathbf{R_j})$, respectively. Then from (1), $\mathbf{z} \sim N(\mu, \mathbf{\Sigma})$, where $\mu = \mathbf{m_i} - \mathbf{m_j}$. The following theorem finds the characteristic function of $\mathbf{z}$. In this theorem, we use the function $etr(\cdot)$ to denote the exponential function of the trace of a matrix.

2.1 Theorem

Let $\mathbf{z} \sim N(\mu, \mathbf{\Sigma})$. Then the distribution of $\mathbf{z}^T\mathbf{z}$ has the following characteristic function:

$$\phi(\omega) = |\mathbf{I} + 2j\omega\mathbf{\Sigma}|^{-\frac{1}{2}} etr\{-j\omega(\mathbf{I} + 2j\omega\mathbf{\Sigma})^{-1}\mu\mu^T\}. \tag{2}$$

2.2 Proof

We start by the basic definition of the characteristic function

$$\begin{aligned}
\phi(\omega) &= \mathbf{E}[\exp(-j\omega\mathbf{z}^T\mathbf{z})] \\
&= \int (2\pi)^{-\frac{n}{2}}|\mathbf{\Sigma}|^{-\frac{1}{2}} \exp(-j\omega\mathbf{z}^T\mathbf{z}) \exp\{-\frac{1}{2}(\mathbf{z}-\mu)^T\mathbf{\Sigma}^{-1}(\mathbf{z}-\mu)\}d\mathbf{z} \\
&= \int (2\pi)^{-\frac{n}{2}}|\mathbf{\Sigma}|^{-\frac{1}{2}} \exp(-\frac{1}{2}s)d\mathbf{z},
\end{aligned}$$

where

$$\begin{aligned} s &\doteq 2j\omega\mathbf{z}^T\mathbf{z} + (\mathbf{z}-\mu)^T\mathbf{\Sigma}^{-1}(\mathbf{z}-\mu) && (3) \\ &= 2j\omega\mathbf{z}^T\mathbf{z} + \mathbf{z}^T\mathbf{\Sigma}^{-1}\mathbf{z} - 2\mu^T\mathbf{\Sigma}^{-1}\mathbf{z} + \mu^T\mathbf{\Sigma}^{-1}\mu && (4) \end{aligned}$$

2.2.1 Claim

$$s = (\mathbf{z}-\mu+\mathbf{q})^T(\mathbf{\Sigma}^{-1}+2j\omega\mathbf{I})(\mathbf{z}-\mu+\mathbf{q}) + 2j\omega\mu^T(\mu-\mathbf{q}), \quad (5)$$

where

$$\mathbf{q} = 2j\omega(\mathbf{\Sigma}^{-1}+2j\omega\mathbf{I})^{-1}\mu$$

2.2.2 Proof of Claim

We need to complete the square in s to get rid of the indefinite integral. So let us consider the following quadratic form:

$$\begin{aligned} t &\doteq (\mathbf{z}-\mu+\mathbf{q})^T(\mathbf{\Sigma}^{-1}+2j\omega\mathbf{I})(\mathbf{z}-\mu+\mathbf{q}) \\ &= s+u+v, \end{aligned}$$

where s as in (4),

$$u = 2\mathbf{z}^T\mathbf{\Sigma}^{-1}\mathbf{q} - 4j\omega\mu^T\mathbf{z} + 4j\omega\mathbf{z}^T\mathbf{q} \quad (6)$$

and

$$v = -2\mu^T\mathbf{\Sigma}^{-1}\mathbf{q} + 2j\omega\mu^T\mu - 4j\omega\mu^T\mathbf{q} + \mathbf{q}^T\mathbf{\Sigma}^{-1}\mathbf{q} + 2j\omega\mathbf{q}^T\mathbf{q}, \quad (7)$$

Now we pick $\mathbf{q}$ such that $u = 0$. Thus, we have

$$\begin{aligned} & \mathbf{z}^T(\mathbf{\Sigma}^{-1}\mathbf{q} - 2j\omega\mu + 2j\omega\mathbf{q}) = 0 \\ \Leftarrow\ & \mathbf{\Sigma}^{-1}\mathbf{q} - 2j\omega\mu + 2j\omega\mathbf{q} = \mathbf{0} && (8) \\ \Leftrightarrow\ & (\mathbf{\Sigma}^{-1}+2j\omega\mathbf{I})\mathbf{q} = 2j\omega\mu. \\ \Leftrightarrow\ & \mathbf{q} = 2j\omega(\mathbf{\Sigma}^{-1}+2j\omega\mathbf{I})^{-1}\mu && (9) \end{aligned}$$

Now, we have

$$\begin{aligned} v &= -2\mu^T\mathbf{\Sigma}^{-1}\mathbf{q} + 2j\omega\mu^T\mu - 4j\omega\mu^T\mathbf{q} + \mathbf{q}^T\mathbf{\Sigma}^{-1}\mathbf{q} + 2j\omega\mathbf{q}^T\mathbf{q}) \\ &= -2\mu^T(\mathbf{\Sigma}^{-1}\mathbf{q} - 2j\omega\mu + 2j\omega\mathbf{q}) - 2j\omega\mu^T\mu + \mathbf{q}^T\mathbf{\Sigma}^{-1}\mathbf{q} + 2j\omega\mathbf{q}^T\mathbf{q} && (10) \\ &= -2j\omega\mu^T\mu + \mathbf{q}^T\mathbf{\Sigma}^{-1}\mathbf{q} + 2j\omega\mathbf{q}^T\mathbf{q} \\ &= -2j\omega\mu^T\mu + \mathbf{q}^T(\mathbf{\Sigma}^{-1}\mathbf{q} + 2j\omega\mathbf{q} - 2j\omega\mu) + 2j\omega\mathbf{q}^T\mu && (11) \\ &= -2j\omega\mu^T\mu + 2j\omega\mathbf{q}^T\mu && (12) \\ &= -2j\omega\mu^T(\mu-\mathbf{q}) && (13) \\ & && (14) \end{aligned}$$

where we applied (8) in (10) and (11). Hence, we have

$$s = (\mathbf{z} - \mu + \mathbf{q})^T(\mathbf{\Sigma}^{-1} + 2j\omega\mathbf{I})(\mathbf{z} - \mu + \mathbf{q}) + 2j\omega\mu^T(\mu - \mathbf{q}), \quad (15)$$

where $\mathbf{q}$ as in (9). And this proves the claim.

2.2.3 Proof of Theorem

Now we have

$$\begin{aligned}
\phi(\omega) &= \int (2\pi)^{-\frac{n}{2}}|\mathbf{\Sigma}|^{-\frac{1}{2}} \exp\{-\frac{1}{2}(\mathbf{z} - \mu + \mathbf{q})^T(\mathbf{\Sigma}^{-1} + 2j\omega\mathbf{I})(\mathbf{z} - \mu + \mathbf{q})\} \exp\{-j\omega\mu^T(\mu - \mathbf{q})\}d\mathbf{z} \\
&= (2\pi)^{-\frac{n}{2}}|\mathbf{\Sigma}|^{-\frac{1}{2}} \exp\{-j\omega\mu^T(\mu - \mathbf{q})\} \int \exp\{-\frac{1}{2}(\mathbf{z} - \mu + \mathbf{q})^T(\mathbf{\Sigma}^{-1} + 2j\omega\mathbf{I})(\mathbf{z} - \mu + \mathbf{q})\}d\mathbf{z} \\
&= (2\pi)^{\frac{n}{2}}|\mathbf{\Sigma}^{-1} + 2j\omega\mathbf{I}|^{-\frac{1}{2}}(2\pi)^{-\frac{n}{2}}|\mathbf{\Sigma}|^{-\frac{1}{2}} \exp\{-j\omega\mu^T(\mu - \mathbf{q})\} \\
&= |\mathbf{I} + 2j\omega\mathbf{\Sigma}|^{-\frac{1}{2}} \exp\{-j\omega\mu^T(\mu - \mathbf{q})\} \\
&= |\mathbf{I} + 2j\omega\mathbf{\Sigma}|^{-\frac{1}{2}} \exp\{-j\omega\mu^T(\mathbf{I} - 2j\omega(\mathbf{\Sigma}^{-1} + 2j\omega\mathbf{I})^{-1})\mu\} \\
&= |\mathbf{I} + 2j\omega\mathbf{\Sigma}|^{-\frac{1}{2}} \exp\{-j\omega\mu^T(\mathbf{I} - 2j\omega\mathbf{\Sigma}(\mathbf{I} + 2j\omega\mathbf{\Sigma})^{-1})\mu\}.
\end{aligned}$$

Now note that Kailath variant matrix identity [1, p.153], states that for matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$, we have

$$(\mathbf{A} + \mathbf{BC})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{I} + \mathbf{CA}^{-1}\mathbf{B})^{-1}\mathbf{CA}^{-1}.$$

Thus, substituting $\mathbf{A} = \mathbf{I}$, $\mathbf{B} = 2j\omega\mathbf{\Sigma}$, and $\mathbf{C} = \mathbf{I}$, we get

$$\mathbf{I} - 2j\omega\mathbf{\Sigma}(\mathbf{I} + 2j\omega\mathbf{\Sigma})^{-1} = (\mathbf{I} + 2j\omega\mathbf{\Sigma})^{-1}.$$

Therefore, we have

$$\begin{aligned}
\phi(\omega) &= |\mathbf{I} + 2j\omega\mathbf{\Sigma}|^{-\frac{1}{2}} \exp\{-j\omega\mu^T(\mathbf{I} + 2j\omega\mathbf{\Sigma})^{-1}\mu\} \\
&= |\mathbf{I} + 2j\omega\mathbf{\Sigma}|^{-\frac{1}{2}} \exp\{-j\omega tr[(\mathbf{I} + 2j\omega\mathbf{\Sigma})^{-1}\mu\mu^T]\},
\end{aligned}$$

where $tr(\cdot)$ denotes the trace of a matrix. Hence the the proof of Theorem 2.1 is completed.

2.3 Remark

The random variable y can be viewed as the trace of a matrix that is non-central Wishart distributed $W_1(n, \mathbf{\Sigma}, \mu\mu^T)$ of one dimension, n degrees of freedom, covariance matrix $\mathbf{\Sigma}$, and non-centrality matrix $\mu\mu^T$. See [15] for more details on such distribution. Thus, (2) gives the characteristic function and probability density distribution of such trace, respectively. The corresponding pdf is given in [14] as an infinite weighted sum of gamma density functions. The following lemma, on the other hand, presents an a asymptotic distribution of y. A plot of the pdf from the given characteristic function can be obtained using an algorithm presented in [7].

2.4 Lemma

As $n \to \infty$, the asymptotic distribution of y is

$$N\left(ntr(\mathbf{\Sigma}), n^3/2tr(\mathbf{\Sigma}^2)\right)$$

2.5 Proof

This result follows from [5], which states that if $W \sim W_m(n, \mathbf{\Sigma}, \mu\mu^T)$, then as $n \to \infty$, the asymptotic distribution of

$$\left[\frac{n}{2tr(\mathbf{\Sigma}^2)}\right]^{1/2}(trW/n - tr\mathbf{\Sigma})$$

is $N(0,1)$. See [15, pp. 517–518] for similar results regarding other functions of W. Thus, the result of the lemma follows by observing that $y \sim W_1(n, \mathbf{\Sigma}, \mu\mu^T)$.

3 Special Cases

In this section, we consider special assumptions on the random variables $\mathbf{x_i}$ and $\mathbf{x_j}$.

3.1 Case: $\mathbf{\Sigma}$ Is Diagonal

This corresponds to having the entries x_{ik} of $\mathbf{x_i}$ independent, the entries x_{jk} of $\mathbf{x_j}$ independent, and x_{ik} and x_{jk} independent.

3.1.1 Lemma

Let $\mathbf{z} \sim N(\mu, \mathbf{\Sigma})$, where $\mathbf{\Sigma}$ is a diagonal matrix with entries σ_k^2 in row k. Then the distribution of $\mathbf{z}^T\mathbf{z}$ has the following characteristic function:

$$\phi(\omega) = \prod_{k=1}^{n}\left[(1 + 2j\omega\sigma_k^2)^{-\frac{1}{2}} \exp(-j\omega\mu_k^2/(1 + 2j\omega\sigma_k^2))\right]. \tag{16}$$

3.1.2 Proof

Since $\mathbf{\Sigma}$ is diagonal, we have

$$|\mathbf{I} + 2j\omega\mathbf{\Sigma}|^{-\frac{1}{2}} = \prod_{k=1}^{n}(1 + 2j\omega\sigma_k^2)^{-\frac{1}{2}}. \tag{17}$$

and

$$\mu^T(\mathbf{I} + 2j\omega\mathbf{\Sigma})^{-1}\mu = \sum_{k=1}^{n}\mu_k^2/(1 + 2j\omega\sigma_k^2). \tag{18}$$

Now, plugging (17) and (18) in (2), we get (16).

3.1.3 Lemma

Let $z_1 \sim N(\mu_1, \sigma_1^2)$. Then the distribution of $y \doteq z_1^2$ has the following characteristic function:

$$\phi(\omega) = (1 + 2j\omega\sigma_1^2)^{-\frac{1}{2}} \exp\left(\frac{-j\omega\mu_1^2}{1 + 2j\omega\sigma_1^2}\right), \tag{19}$$

and the following probability density function:

$$f(y) = (2\pi y\sigma_1^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\frac{y + \mu_1^2}{\sigma_1^2}\right)\right\} \cosh\left(\frac{y^{\frac{1}{2}}\mu_1}{\sigma_1^2}\right) \tag{20}$$

3.1.4 Proof

The characteristic function follows from Lemma 3.1.1 by putting $n = 1$. Now, we can find the corresponding pdf as follows. For $y > 0$, we have

$$\begin{aligned}
F_Y(y) &= P(Y \leq y) \\
&= P(Z^2 \leq y) \\
&= P(-\sqrt{y} \leq Z \leq \sqrt{y}) \\
&= (2\pi\sigma_1^2)^{-\frac{1}{2}} \int_{-\sqrt{y}}^{\sqrt{y}} \exp\left\{-\frac{1}{2}\left(\frac{(t-\mu_1)^2}{\sigma_1^2}\right)\right\} dt \\
&= (2\pi\sigma_1^2)^{-\frac{1}{2}} \left[\int_0^{\sqrt{y}} \exp\left\{-\frac{1}{2}\left(\frac{(t-\mu_1)^2}{\sigma_1^2}\right)\right\} dt - \int_0^{-\sqrt{y}} \exp\left\{-\frac{1}{2}\left(\frac{(t-\mu_1)^2}{\sigma_1^2}\right)\right\} dt\right] \\
&= \frac{1}{2}(2\pi\sigma_1^2)^{-\frac{1}{2}} \left[\int_0^y \frac{1}{\sqrt{s}} \exp\left\{-\frac{1}{2}\left(\frac{(\sqrt{s}-\mu_1)^2}{\sigma_1^2}\right)\right\} ds + \int_0^y \frac{1}{\sqrt{s}} \exp\left\{-\frac{1}{2}\left(\frac{(\sqrt{s}+\mu_1)^2}{\sigma_1^2}\right)\right\} ds\right],
\end{aligned}$$

where in the last step we applied the following change of variables, $s = t^2$. So, $dt = \frac{ds}{2\sqrt{s}}$ for $t > 0$, and $dt = \frac{ds}{-2\sqrt{s}}$ for $t < 0$.

Now, we get the pdf from the above cumulative distribution function (cdf) as follows

$$\begin{aligned}
f_Y(y) &= \frac{\partial F_Y(y)}{\partial y} \\
&= \frac{1}{2}(2\pi\sigma_1^2 y)^{-\frac{1}{2}} \left[\exp\left\{-\frac{1}{2}\left(\frac{(\sqrt{y}-\mu_1)^2}{\sigma_1^2}\right)\right\} + \exp\left\{-\frac{1}{2}\left(\frac{(\sqrt{y}+\mu_1)^2}{\sigma_1^2}\right)\right\}\right], \\
&= \frac{1}{2}(2\pi\sigma_1^2 y)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\frac{(y+\mu_1^2)}{\sigma_1^2}\right)\right\} \left[\exp\left\{-\frac{\sqrt{y}\mu_1}{\sigma_1^2}\right\} + \exp\left\{\frac{\sqrt{y}\mu_1}{\sigma_1^2}\right\}\right], \\
&= (2\pi y\sigma_1^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\frac{y+\mu_1^2}{\sigma_1^2}\right)\right\} \cosh\left(\frac{y^{\frac{1}{2}}\mu_1}{\sigma_1^2}\right)
\end{aligned}$$

3.1.5 Notes

It is noted that the characteristic function in Lemma 3.1.1 is that of the sum of n independent random variables having the distribution in Lemma 3.1.3 with parameters $(\mu_i, \sigma_i),\ i = 1, \ldots, n$. On another note, for the distribution in Lemma 3.1.3, when $\mu_1 = 0$ and $\sigma_1^2 = 1$, the distribution is chi-squared with one degree of freedom. On the other hand, when $\mu_1 \neq 0$ and $\sigma_1^2 = 1$, the distribution is non-central chi-squared with one degree of freedom and noncentrality parameter μ_1^2.

3.2 Case: $\mathbf{\Sigma} = \sigma^2 \mathbf{I}$

This corresponds to having the entries x_{ik} of $\mathbf{x_i}$ independent, the entries x_{jk} of $\mathbf{x_j}$ independent, and x_{ik} and x_{jk} independent.

3.2.1 Lemma

Let $z \sim N(\mu, \sigma^2\mathbf{I})$. Then the distribution of $\mathbf{z}^T\mathbf{z}$ has the following characteristic function:

$$\phi(\omega) = (1 + 2j\omega\sigma^2)^{-\frac{n}{2}} \exp\left(-j\omega\lambda/(1 + 2j\omega\sigma^2)\right), \tag{21}$$

where $\lambda = \sum_{k=1}^n \mu_k^2$, and the following probability density function:

$$\begin{aligned} f(y) &= 2^{-\frac{n}{2}} \exp[-\frac{(y+\lambda)}{2\sigma^2}] \sum_{j=0}^{\infty} \frac{y^{n/2+j-1}\lambda^j}{\Gamma(n/2+j)2^{2j}j!\sigma^{n+4j}} \\ &= \frac{1}{2\sigma^2}\,(y/\lambda)^{(n-2)/4} \exp(-\frac{(y+\lambda)}{2\sigma^2}) I_{(n-2)/2}(\frac{\sqrt{\lambda y}}{\sigma^2}) \end{aligned}$$

where $I_\alpha(\cdot)$ is the modified Bessel function of the first kind of degree α. This is the noncentral Gamma distribution with parameters $a = n/2$, $b = 2\sigma^2$, and $c = \lambda/(4\sigma^4)$.

3.2.2 Proof

The characteristic function follows from Lemma 3.1.1 by taking $\sigma_k^2 = \sigma^2$ and $\lambda = \sum_{k=1}^n \mu_k^2$. Now, to obtain the corresponding pdf, we proceed as follows. First note that the characteristic function and pdf of noncentral chi squared random variable with n degrees of freedom and noncentrality parameter λ is given as follows:

$$\varphi_\lambda(\omega) = (1 + 2j\omega)^{-\frac{n}{2}} \exp\left(-j\omega\lambda/(1 + 2j\omega)\right)),$$

and

$$\begin{aligned} g(y) &= 2^{-\frac{n}{2}} \exp[-\frac{1}{2}(y+\lambda)] \sum_{j=0}^{\infty} \frac{y^{(n/2)+j-1}\lambda^j}{\Gamma(\frac{n}{2}+j)2^{2j}j!} \\ &= \exp(-\frac{1}{2}(\lambda+y))\frac{1}{2}\left(\frac{y}{\lambda}\right)^{(n-2)/4} I_{(n-2)/2}(\sqrt{\lambda y}) \end{aligned}$$

where $I_\alpha(\cdot)$ is the modified Bessel function of the first kind of degree α and is given by the following infinite sum

$$I_\alpha(y) = (y/2)^\alpha \sum_{i=0}^{\infty} \frac{(y/2)^{2i}}{i!\Gamma(\alpha+i+1)}. \tag{22}$$

See [6, pp. 900 – 932] for more information on various kinds of Bessel functions and some of the identities and approximations associated with them.

Thus, from the properties of Fourier transform, since $\phi(\omega) = \varphi_{\lambda/\sigma^2}(\omega\sigma^2)$, we get $f(y) = \frac{1}{\sigma^2} g_{\lambda/\sigma^2}(\frac{y}{\sigma^2})$. And this concludes the proof.

3.2.3 Note

A non-central Gamma distribution is defined as having the following characteristic function

$$\varphi(\omega) = (1 + j\omega b)^{-a} \exp\left(-j\omega b^2 c/(1 + j\omega b)\right),$$

where a, b, and c are referred to as shape, scale, and non-centrality parameters, respectively [10].

The corresponding probability distribution function is given as

$$\begin{aligned} \gamma_{a,b,c}(x) &= b^{-a} e^{-bc} e^{-x/b} x^{a-1} \sum_{k=0}^{\infty} \frac{(cx)^k}{k!\Gamma(a+k)}, \quad x > 0, \\ &= b^{-a} e^{-bc} e^{-x/b} (x/c)^{(a-1)/2} I_{a-1}(2\sqrt{cx}), \quad x > 0. \end{aligned}$$

3.3 Case: $\boldsymbol{\Sigma}$ Is Diagonal and $\mu = \mathbf{0}$

Then from (16) we get

$$\phi(\omega) = \prod_{k=1}^{n} (1 + 2j\omega\sigma_k^2)^{-\frac{1}{2}}. \tag{23}$$

But this is the characteristic function of the sum of the n independent random variables each is distributed as $\Gamma(\frac{1}{2}, 2\sigma_k^2)$, where $\Gamma(a, b)$ is the Gamma distribution with probability density function

$$\gamma_{a,b}(x) = \frac{x^{a-1} e^{-\frac{x}{b}}}{b^a \Gamma(a)}, \quad x > 0,$$

and $\Gamma(a)$ is the Gamma function, defined as

$$\Gamma(a) = \int_0^{\infty} x^{a-1} \exp(x) dx,$$

and a characteristic function:

$$\varphi(\omega) = (1 + j\omega b)^{-a}$$

The distribution of such sums is considered in [16], from which we get the Stacy distribution with the following pdf:

$$f(x) = x^{\frac{n}{2}-1} \sum_{i=0}^{\infty} \frac{(-x)^i}{\Gamma(\frac{n}{2}+i)} \sum_{\sum_{j=1}^{n} k_j = i} \prod_{l=1}^{n} \frac{\Gamma(k_l + \frac{1}{2})}{k_l! \sqrt{\pi} (2\sigma_l^2)^{k_l+1/2}}.$$

Note now that for an integer m, we have

$$\Gamma(m + \frac{1}{2}) = \frac{(2m-1)!!\sqrt{\pi}}{2^m}.$$

So, putting $m = k_l$, we get

$$f(x) = x^{\frac{n}{2}-1} \sum_{i=0}^{\infty} \frac{(-x)^i}{\Gamma(\frac{n}{2}+i)} \sum_{\sum_{j=1}^{n} k_j = i} \prod_{l=1}^{n} \frac{(2k_l-1)!!}{k_l! 2^{2k_l+1/2} \sigma_l^{2k_l+1}}.$$

Since we have

$$\sum_{j=1}^{n} k_j = i,$$

we can rewrite $f(x)$ as

$$f(x) = x^{\frac{n}{2}-1} \sum_{i=0}^{\infty} \frac{(-x)^i}{\Gamma(\frac{n}{2}+i) 2^{2i+n/2}} \sum_{\sum_{j=1}^{n} k_j = i} \prod_{l=1}^{n} \frac{(2k_l-1)!!}{k_l! \sigma_l^{2k_l+1}}.$$

Note again that for an integer m, we have

$$\Gamma(\frac{m}{2}) = \frac{(m-2)!! \sqrt{\pi}}{2^{(m-1)/2}}.$$

Thus, putting $m = n + 2i$, we get

$$f(x) = \frac{x^{\frac{n}{2}-1}}{\sqrt{2\pi}} \sum_{i=0}^{\infty} \frac{(-x/2)^i}{[n+2(i-1)]!!} \sum_{\sum_{j=1}^{n} k_j = i} \prod_{l=1}^{n} \frac{(2k_l-1)!!}{k_l! \sigma_l^{2k_l+1}}.$$

Now note that for an integer m, we have

$$(2m-1)!! = \frac{(2m)!}{2^m m!}.$$

Hence, putting $m = k_l$, and noting that $\sum_{j=1}^{n} k_j = i$, we get

$$f(x) = \frac{x^{\frac{n}{2}-1}}{\sqrt{2\pi}} \sum_{i=0}^{\infty} \frac{(-x)^i}{2^{2i}[n+2(i-1)]!!} \sum_{\sum_{j=1}^{n} k_j = i} \prod_{l=1}^{n} C_{k_l}^{2k_l} \sigma_l^{-(2k_l+1)}.$$

3.4 Case: $\mathbf{\Sigma} = \sigma^2 \mathbf{I}$ and $\mu = \mathbf{0}$

In this case, (23) simplifies to

$$\phi(\omega) = (1 + 2j\omega\sigma^2)^{-\frac{n}{2}}. \tag{24}$$

This is the characteristic function of gamma distributed random variable with parameters $a = \frac{n}{2}$ and $b = 2\sigma^2$.

3.5 Case: $\mathbf{\Sigma} = \mathbf{I}$ and $\mu = \mathbf{0}$

In this case we get

$$\phi(\omega) = (1 + 2j\omega)^{-\frac{n}{2}} \tag{25}$$

This is the characteristic function of chi-squared random variable with n degrees of freedom. The corresponding pdf is given by

$$f(x) = \frac{x^{\frac{n}{2}-1}\exp(-\frac{x}{2})}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}}; \quad x > 0.$$

In this case, the distance d_{ij} is chi distributed with n degrees of freedom having the corresponding pdf

$$f(x) = \frac{x^{n-1}\exp(-\frac{x^2}{2})}{\Gamma(\frac{n}{2})2^{\frac{n}{2}-1}}; \quad x > 0.$$

4 Concluding Remarks

We presented the distribution of the squared Euclidean distance between two multivariate normally distributed points. We then reduced this distribution for special cases based on the mean and covariance. We believe that such result may find its application in the area of clusters identification by developing an algorithm that infers the cluster to which some points belong from histogram of their distances from a particular point.

Index of Notations

- $n!!$ is the double factorial function defined as

$$n!! = \begin{cases} n(n-2)\ldots 3\cdot 1 & \text{if} \quad n > 0 \quad \text{is odd;} \\ n(n-2)\ldots 4\cdot 2 & \text{if} \quad n > 0 \quad \text{is even,} \\ 1 & \text{if} \quad n = -1, 0. \end{cases}$$

- C_i^j where i and j non-negative integers with $j \geq i$ denotes the binomial coefficient defined as

$$C_i^j = \frac{j!}{i!(j-i)!}.$$

- $\cosh_i(z)$ is the alternate hyperbolic cosine/sine function defined as

$$\begin{aligned} \cosh_i(z) &= \begin{cases} \cosh(z) & \text{if} \quad i \quad \text{is even;} \\ \sinh(z) & \text{if} \quad i \quad \text{is odd,} \end{cases} \\ &= \frac{e^z + (-1)^i e^{-z}}{2}. \end{aligned}$$

References

[1] C. Bishop (1995), "*Neural networks for pattern recognition,*" Oxford University Press.

[2] A. G. Constantine (1963), "Some non-central distribution problems in multivariate analysis," *the Annals of Mathematical Statistics,* Vol. 34, No. 4, pp. 1270–1285, December, 1963.

[3] C. Donati-Martin, Y. Doumerc, H. Matsumoto, and M. Yor (2004), "Some properties of the Wishart processes and a matrix extension of the Hartman-Watson laws," Publications of the Research Institute of Mathimatical Science, Kyoto University, Kyoto, Japan, Vol. 40, No. 4, pp. 1385–1412.

[4] I. K. Fodor (2002), "A survey of dimension reduction techniques," Lawrence Livermore National Laboratory (LLNL) Technical Report, UCRL-ID-148494, University of California, Livermore, CA, 2002.

[5] Y. Fujikoshi (1970), "Asymptotic expansions of distributions of test statistics in multivariate analysis," Journal of Science of Hiroshima University, Series A-1, Vol. 34, pp. 73–144.

[6] I. S. Gradshteyn and I. M. Ryzhik (2000), "*Table of integrals, series, and products,*" Sixth Edition, Academic Press.

[7] J. A. Gubner (1996), "Computation of shot–noise probability distributions and densities," *SIAM Journal of Scientific Computing*, Vol. 17, No. 3, pp. 750–761, May 1996.

[8] C. S. Herz (1955), "Bessel functions of matrix argument," *the Annals of Mathematics*, 2nd Scr., Vol. 61, No. 3, pp. 474–523, May, 1955.

[9] J. R. Hoffman, M. G. Cotton, R. J. Achatz, R. N. Statz, and R. A. Dalke (2001), "Measurements to determine potential interference to GPS receivers from ultrawideband transmission systems," National Telecommunications and Information Administration (NTIA) Report 01-384, Institute for Telecommunication Sciences, Boulder, Colorado, February, 2001.

[10] M. M. Islam (2003), "Family of non-central transformed chi-square distributions," *Pakistan Journal of Statistics*, Volume 19(3), pp. 325–330.

[11] N. L. Johnson, S. Kotz, and N. Balakrishnan (1994), "*Continuous univarate distributions,*" Vol. 1, Second Edition, John Wiley and Sons, Inc.

[12] B. Jonsson (1978), "Kinesiology: with special reverence to electromyographic kinesiology," *Contemporary Clinical Neurophysiology*, EEG Suppl. No. 34, pp. 417–428.

[13] L. Kaufman and P. J. Rousseeuw, (1990), "Finding groups in data: an introduction to cluster analysis," John Wiley and Sons, Inc.

[14] S. Kourouklis and P. G. Moschopoulos (1985), "On the distribution of the trace of a noncentral Wishart matrix,"Metron, Vol. XLIII, No. 1-2, pp. 85–92.

[15] R. J. Muirhead (2005), "*Aspects of multivariate statistical theory,*" John Wiley and Sons, Inc.

[16] E. W. Stacy (1962), "A generalization of the gamma distribution," *Annals of Mathematical Statistics*, Vol. 33, No. 3, pp. 1187–1192.

[17] L. L. Scharf (1991), "*Statistical signal processing: detection, estimation, and time series analysis*," Addison Wesley.