

Maxwell Normal Distribution in a Manifold and Mahalanobis Metric

Yukihiko Yamashita, Mariko Numakami, and Naoya Inoue

Graduate School of Science and Engineering, Tokyo Institute of Technology,
2-12-1-S6-19, O-okayama, Meguro-ku, Tokyo 152-8553, Japan
{yamasita, numakami, n708i}@ide.titech.ac.jp

Abstract. The normal distribution in Euclidean space is used widely for statistical models. However, for pattern recognition, since pattern vectors are often normalized by their norm, they are on a hyper-spherical surface. Therefore, we have to study a normal distribution in a non-Euclidean space. Here, we provide the new concept of geometrically local isotropic independence and define the Maxwell normal distribution in a manifold. We also define the Mahalanobis metric, which is an extension of the Mahalanobis distance in Euclidean space. We provide the Mahalanobis metric equation, which is covariant with coordinate transformation. Furthermore, we show its experimental results.

1 Introduction

In many fields, such as pattern recognition and data analysis, a normal distribution in Euclidean space is used for a statistical model. However, for pattern recognition, pattern vectors are often normalized by their norm and they are not in Euclidean space but on a hyper-spherical surface. Therefore, we have to study a normal distribution in a non-Euclidean space.

The normal distribution is characterized by the equality between sample mean and maximum likelihood, isotropic independence, maximum entropy or maximum number of cases, and limits such as those provided by the central limit theorem. In this paper, we extend the second characterization. We propose the concept of geometrically local isotropic independence (GLII) and define the Maxwell normal distribution in a manifold (MNDM). From the definition, we give MNDM on the n -dimensional hyper-spherical surface S^n and show that in the case of S^2 , it coincides with the Fisher distribution [1].

The Mahalanobis distance [2], which is a distance normalized by a variance-covariance matrix, is also used for pattern recognition and data analysis. By using it, we can know that a distance between two vectors that is not large in terms of Euclidean distance may be large from the viewpoint of probability (Fig. 1(a)).

In the n -dimensional Euclidean space E^n , the normal distribution with average μ and variance-covariance matrix Σ is expressed by the following probability

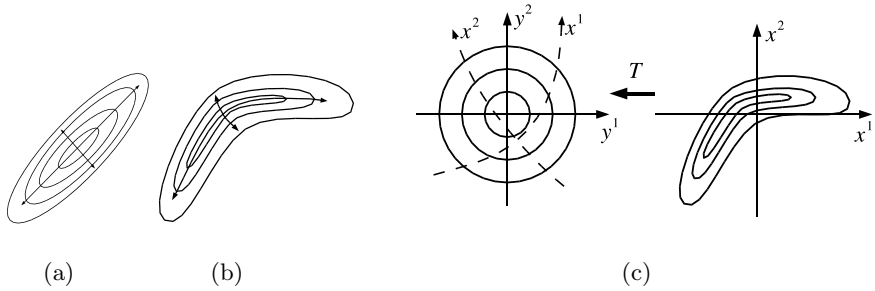


Fig. 1. Mahalanobis metric

distribution function (p.d.f.):

$$p(x) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \exp(-\langle \Sigma^{-1}(x - \mu), x - \mu \rangle / 2), \tag{1}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product and Σ is assumed to be nonsingular. When x is mapped to y by a linear transformation $y = \Sigma^{-1/2}x$, p is transformed to the p.d.f. of the standard normal distribution of y . The inner product that expresses the Mahalanobis distance is given as the pull back of the inner product as $\langle y, y' \rangle = \langle \Sigma^{-1}x, x' \rangle$.

Suppose that the counter of a p.d.f. is given as in Fig. 1(b). It is natural that the distance is evaluated not by a straight edge but by a curve. By extending the linear transform to a diffeomorphism, we obtain the Mahalanobis metric. It is very interesting that its differential equation does not depend on the coordinate system or the original metric. Furthermore, the diffeomorphism will disappear in it.

Information geometry [3] uses a manifold for probability distributions. However, since it uses a manifold as the set of distributions, there is no direct relationship to our theorem.

In Section 2, we describe the characterization of the normal distribution in Euclidean space. In Section 3, we define GLII and provide its differential equation. In Section 4, we define MNDM as a solution of the equation and provide a solution on S^n . In Section 5, we define the Mahalanobis distance and provide its differential equation. In Section 6, we present experimental results of a simulation.

In this paper, we restrict manifolds to the Riemannian manifold and use the Levi-Chivita connection. We assume that every probability distribution function (p.d.f.) is infinitely continuously differentiable and does not become 0.

2 Characterization of Normal Distribution

We explain the characterization of a normal distribution in Euclidean space. In some of the following characterizations we assume that the ensemble average is zero without loss of generality.

The normal distribution is often called the Gaussian distribution. The following characterization in E^1 was given by C.F. Gauss [4]. Let μ and x be the true and observed values, respectively. Assume that the p.d.f. depends only on the absolute error $|x - \mu|$ and the samples x_i ($i = 1, 2, \dots, m$) are extracted independently. If the maximum likelihood estimator of μ is always given by the sample mean $(\sum_{i=1}^m x_i)/n$, the distribution should be a normal distribution whose average is μ .

The velocity distribution of the ideal gas is given as a 3-dimensional normal distribution (Maxwell distribution) [5]. Below, we briefly describe its characterization in E^2 . Let (x^1, x^2) and (y^1, y^2) be two 2-dimensional stochastic variables such that x^1 and x^2 are independent of each other and so are y^1 and y^2 . We have

$$\begin{cases} y^1 = x^1 \cos \omega + x^2 \sin \omega, \\ y^2 = -x^1 \sin \omega + x^2 \cos \omega \end{cases} \quad (2)$$

for some $\omega \neq n\pi/2$, where $n = 0, \pm 1, \pm 2, \dots$. Then, the distribution of (x^1, x^2) is a 2-dimensional normal distribution whose variance-covariance matrix is given by $\sigma^2 I$, where σ is a positive number and I is a unit matrix.

The characterization by the entropy in E^1 is given as follows. Let $p(x)$ be a p.d.f. defined in $(-\infty, \infty)$. We assume that $p(x)$ maximizes the entropy

$$- \int_{-\infty}^{\infty} p(x) \log p(x) dx \quad (3)$$

under the condition that

$$\int_{-\infty}^{\infty} x^2 p(x) dx = \sigma^2. \quad (4)$$

Thus, $p(x)$ is given by the normal distribution whose average and variance are 0 and σ^2 , respectively. Similarly we can derive the normal distribution from the Maxwell-Boltzmann distribution [6], [7].

The central limit theorem is given as follows [5]. For simplicity, we neglect several conditions for convergence. Let x_i ($i = 1, \dots, m$) be independently extracted samples of the same arbitrary distribution. Then, $\sum_{i=1}^m x_i/\sqrt{n}$ converges to a normal distribution whose average and variance coincide with those of the original distribution, respectively. From the viewpoint of convergence, the normal distribution is also characterized as the limit of the Poisson distribution.

In a manifold, it is very difficult to define average, variance, and dilation uniquely. The facts show the difficulty of defining a normal distribution in a manifold by extending the characterization by Gauss or entropy or the central limit theorem. Furthermore, in physics, since the variance, which is given by a square term, means the energy, the condition of its preservation is justified. However, in general, the reason why we fix the variance is not clear.

Therefore, in this paper, we extend the characterization by Maxwell. It is based on the independence. The assumption of independence of a distribution with respect to orthogonal coordinates is reasonable in many cases. In the following discussion, we assume that the dimension of manifolds is not less than 2.

3 Geometrically Local Isotropic Independence (GLII)

In E^2 , we can describe the independence as $p(x, y) = p(x)p(y)$. However, in a manifold we cannot construct a global orthogonal coordinate system based on geodesics that correspond to the orthogonal coordinate system in E^n . Thus, we cannot use the characterization by Maxwell in E^n directly. We propose the concept of *geometrically local independence* (GLI). Since we cannot define the global direction in a manifold either and only the isotropic independence is necessary to extend the characterization by Maxwell, we define the *geometrically local isotropic independence* (GLII).

Let $\{x^\mu\}$ be a local coordinate system in a manifold. Let $g_{\mu\nu}$ be a metric tensor on the coordinate system $\{x^\mu\}$. Let $g = \det(g_{\mu\nu})$. Let p be a p.d.f. on the coordinate system $\{x^\mu\}$. Let q be its p.d.f. normalized by \sqrt{g} (We call it the normalized p.d.f. for short), which is written as

$$q = \frac{p}{\sqrt{g}}. \tag{5}$$

This q is a scaler and it is invariant for the coordinate transformation. We define $\delta_{\mu\nu}$ as

$$\delta_{\mu\nu} = \begin{cases} 1 & (\mu = \nu) \\ 0 & (\text{else}) \end{cases}. \tag{6}$$

Here, let $\{x^\mu\}$ be a normal coordinate system [8]. There exists a normal coordinate system at every point in a Riemannian manifold. At the origin ($x^\alpha = 0, \alpha = 1, 2, \dots, n$) of the normal coordinate system, we have $g_{\mu\nu} = \delta_{\mu\nu}$ and all Levi-Chivita connections $\Gamma_{\nu\gamma}^\mu$ vanish. Now, we define GLI.

Definition 1. (*Geometrically local independence, GLI*) Let $\{x^\mu\}_{\mu=1}^n$ be a normal coordinate system. We define a normalized p.d.f. q as being geometrically locally independent (GLI) with respect to x^μ and x^ν ($\mu \neq \nu$) at the origin ($x^\alpha = 0, \alpha = 1, 2, \dots, n$) if and only if $\frac{1}{q} \frac{\partial q}{\partial x^\mu}$ (the changing rate of q normalized by q with respect to x^μ) does not depend on x^ν at the origin with the approximation of the first order of coordinates.

Theorem 1. Let q be a normalized p.d.f. At the origin of the normal coordinate system, the probability distribution is GLI with respect to x^μ and x^ν ($\mu \neq \nu$) if and only if at the origin we have

$$\frac{\partial^2}{\partial x^\nu \partial x^\mu} \log q = 0. \tag{7}$$

The proof of this theorem is clear. Theorem 1 yields that x^μ and x^ν are commutative in Definition 1. Thus, we can say 'GLI with respect to x^μ and x^ν '.

GLI is an extension of the independence since in E^2 GLI for any parallel translation of a coordinate system is equivalent to the independence.

Theorem 2. Let $\{x^1, x^2\}$ be the orthogonal coordinate system of E^2 . Let $\{y^1, y^2\}$ be an orthogonal coordinate system, which is given by its parallel translation. A p.d.f. is GLI with $\{y^1, y^2\}$ for every parallel translation if and only if x^1 and x^2 are independent.

The proof of this theorem is also clear since every coordinate system in E^2 is a normal coordinate system. Now, we define GLII.

Definition 2. (*Geometrically local isotropic independence, GLII*) Let $\{x^\mu\}$ be a normal coordinate system. Let $\{x'^\mu\}$ be a normal coordinate system given by a rotation at the origin. A p.d.f. is said to be geometrical isotropically locally independent (GLII) at the origin if and only if the normalized p.d.f. is GLI with respect to any pair of coordinates y^μ and y^ν ($\mu \neq \nu$) of a system which is given by an arbitrary rotation of $\{x^\mu\}$ whose center is the origin.

Theorem 3. Let q be a normalized p.d.f. At the origin, q is GLII if and only if

$$\frac{\partial^2}{\partial x^\mu \partial x^\nu} \log q = 0, \quad \frac{\partial^2}{\partial (x^\mu)^2} \log q = \frac{\partial^2}{\partial (x^\nu)^2} \log q \tag{8}$$

for any pair of x^μ and x^ν ($\mu \neq \nu$).

This theorem can be proved by the transformation of partial derivatives. We extend Theorem 3 from a normal coordinate system to a general coordinate system in the following Theorem 4. We denote the covariant differential by ∇_μ .

Theorem 4. A normalised p.d.f. q is GLII if and only if

$$\nabla_\mu \nabla_\nu \log q = f g_{\mu\nu} \tag{9}$$

with a scalar f .

By reducing eq. (9) with $g^{\mu\nu}$, we can get f . Then, eq. (9) is equivalent to

$$\left(\nabla_\mu \nabla_\nu - \frac{1}{n} g_{\mu\nu} \Delta \right) \log q = 0, \tag{10}$$

where $\Delta = g^{\alpha\beta} \nabla_\alpha \nabla_\beta$ and n is the dimension of the manifold.

Since $\nabla_\nu \log q$ is a covariant differential of a scalar, it is equivalent to a partial differential, that is, eq. (9) is equivalent to $\nabla_\mu \partial_\nu \log q = f g_{\mu\nu}$. However, we describe it by a covariant differential in order to see the symmetry easily.

Proof. First, let $\{x^\mu\}$ be a normal coordinate system. From equations in (8), the distribution is GLII if and only if

$$\frac{\partial^2}{\partial x^\mu \partial x^\nu} \log q = f \delta_{\mu\nu}. \tag{11}$$

Since x^μ is a normal coordinate and $\log q$ is a scalar, $\frac{\partial^2}{\partial x^\mu \partial x^\nu} \log q$ and $\delta_{\mu\nu}$ are equal to $\nabla_\mu \nabla_\nu \log q$ and $g_{\mu\nu}$, respectively. Then, eq. (11) can be described as

$$\nabla_\mu \nabla_\nu \log q = f g_{\mu\nu}. \tag{12}$$

We can let f be a scalar since f is the same for any normal coordinate system at the point. Then, both sides of eq. (12) are transformed as 2nd order covariant tensors so that eq. (12) holds for any coordinate system.

Conversely, if eq. (12) holds for a scalar f in a coordinate system, then eq. (11) holds in the normal coordinate system. This completes the proof. □

4 Maxwell Normal Distribution in a Manifold (MNMDM)

For the GLII probability distribution in E^n , we have the following theorem.

Theorem 5. *If a probability distribution is GLII at every point in E^n , it is an uncorrelated normal distribution whose variables have the same variance.*

This theorem can be proved since the variables of the partial differential equation can be separated. From Theorem 5, we propose the following definition.

Definition 3. *(Maxwell normal distribution in a manifold (MNMDM)) A probability distribution is MNMDM if and only if it is GLII at every point.*

As an example of MNMDM in a manifold, we provide it in S^n .

Theorem 6. *The normalized p.d.f. of MNMDM on S^n depends only on the angle x^1 from some point on S^n , and for a constant κ it is given by and for a constant κ it is given by*

$$q = C \exp(\kappa \cos x^1), \tag{13}$$

where C is a constant given by

$$C = \frac{\kappa^{(n-1)/2}}{(2\pi)^{(n+1)/2} I_{(n-1)/2}(\kappa)}, \tag{14}$$

where $I_p(\kappa)$ is the deformed Bessel function.

The proof of this theorem is so long we provide it in another opportunity.

Remark. MNMDM in S^2 coincides with the Fisher distribution, which was originally given as a Maxwell-Boltzmann distribution of a magnetic dipole in a magnetic field, since its energy is proportional to $\cos x^1$, where x_1 is the angle between the magnetic dipole and the magnetic field. When x^1 is small, we have $\cos x^1 \simeq 1 - (x^1)^2/2$, so it is approximated by a normal distribution in Euclidean space. Although the Fisher distribution has similar characterizations by Gauss and by entropy, the distance used for the definition of the average or variance is not the ordinary distance on S^n but the distance in E^{n+1} embedding S^n . One of the advantages of the characterization by GLII is that such embedding is not needed. Another advantage is that it can provide the Mahalanobis metric, which we discuss in the next section.

5 Mahalanobis Metric

First, we describe how a p.d.f is transformed with a diffeomorphism (Fig. 1 (c)). Let M and M' be manifolds. Let $\{x^\mu\}$ and $\{y^\mu\}$ be the coordinate systems, and let $g_{\mu\nu}$ and $g'_{\mu\nu}$ be the metric tensors in M and M' , respectively. Let $g = \det(g_{\mu\nu})$ and $g' = \det(g'_{\mu\nu})$. Let $T : M \rightarrow M'$ be a diffeomorphism.

Let q' be a normalized p.d.f. in M' . Since we can unify elements in M' and M , we can consider a normalized p.d.f. q in M as corresponding to q' . We define $\det T$ as

$$\det T = \sqrt{\frac{g'}{g}} \det \left(\frac{\partial y^\mu}{\partial x^\nu} \right). \tag{15}$$

Note that $\det T$ is invariant with coordinate transformations in both M and M' . Then, we have

$$q(x) = \det T q'(T(x)). \tag{16}$$

We propose the following definition of the Mahalanobis metric.

Definition 4. (*Mahalanobis metric*) We assume that there exists a transformation $T : M \rightarrow M'$ and the normalized p.d.f. q' of MNDM in M' such that a p.d.f. p is given by transforming q' by T . The Mahalanobis metric $\tilde{g}_{\mu\nu}$ with respect to p in M is given as the pull back of $g'_{\mu\nu}$.

From the assumption, we have

$$\nabla'_\mu \nabla'_\nu \log q' = f g'_{\mu\nu} \tag{17}$$

where ∇'_μ is the covariant differential defined by the metric tensor $g'_{\mu\nu}$ in M' . Let $\tilde{g} = \det(\tilde{g}_{\mu\nu})$. Since we have

$$\det \left(\frac{\partial y^\mu}{\partial x^\nu} \right) = \sqrt{\frac{\tilde{g}}{g'}}, \tag{18}$$

eq. (15) and $p(x) = \sqrt{g} q(x)$ yield that

$$q'(T(x)) = p(x) / \sqrt{\tilde{g}}. \tag{19}$$

Furthermore, the transformation from $(x, \tilde{g}_{\mu\nu})$ to $(y, g'_{\mu\nu})$ by T is equivalent to the coordinate transformation. Then, we get the following theorem by rewriting $\tilde{g}_{\mu\nu}$ as $g_{\mu\nu}$ and calculating $\nabla_\mu \nabla_\nu \log g$.

Theorem 7. A metric $g_{\mu\nu}$ is a Mahalanobis metric with respect to p if and only if

$$\nabla_\mu \nabla_\nu \log p - \frac{\partial}{\partial x^\nu} \Gamma_{\alpha\mu}^\alpha + \Gamma_{\nu\mu}^\eta \Gamma_{\alpha\eta}^\alpha = f g_{\mu\nu}. \tag{20}$$

with a scalar f .

Remarks. Eq.(20) is covariant with the coordinate transformation. Since eq. (20) does not include a term that depends on the original metric of M , the Mahalanobis metric does not depend on the original metric either. That is an extension of Mahalanobis distance not depending on the original inner product [9]. Although we introduced a diffeomorphism T for its definition, terms with respect to T and M' disappear in eq. (20).

6 Experiment on Mahalanobis Metric

It may be difficult to solve eq. (20) directly since it includes an arbitrary scalar f . When we assume that $M' = E^n$ and the variance of the normal distribution in M' is one, we have $f = -1$. (For another example, if $M' = S^n$, we have $f = -\log q$.) Then, the equation can be solved. Since this Mahalanobis metric $g_{\mu\nu}$ is given as a metric transformed from E^n , we have $R_{\nu\alpha\beta}^\mu = 0$. We add this equation to the criterion.

For the experiment, we let $M' = E^2$ and a diffeomorphism be

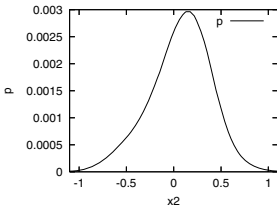
$$y^1 = \alpha x^1, \tag{21}$$

$$y^2 = \alpha(x^2 + \beta h(x^1)h(x^2)), \tag{22}$$

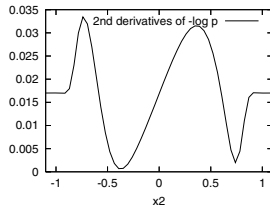
where $\alpha = 3$, $\beta = 0.3$, and

$$h(x) = \begin{cases} e^{4(1-\frac{1}{1-x^2})} & (-1 < x < 1) \\ 0 & (\text{else}) \end{cases}. \tag{23}$$

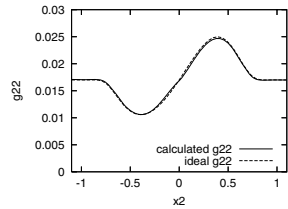
From this diffeomorphism we can calculate p as shown in Fig. 2(a). We transform the differential equation (20) and $R_{\nu\alpha\beta}^\mu$ to difference equations on a 51×51 mesh. We let the differentials of $g_{\mu\nu}$ on the boundary be zero. Let F be the squared sum of errors of the difference equations. We calculate the Mahalanobis metric by the steepest descent method.



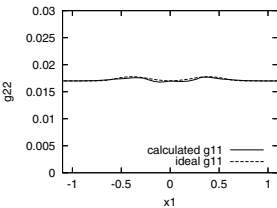
(a) $p (x^1 = 0)$



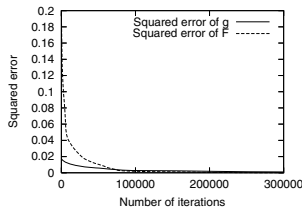
(b) $\partial_2 \partial_2 \log p (x^1 = 0)$



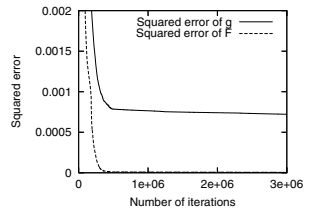
(c) $g_{22} (x^1 = 0)$



(d) $g_{11} (x^2 = 0)$



(e) Mean square errors of equation F and metric G



(f) Rescaled version of (e) (1/100 times horizontally and 10 times vertically)

Fig. 2. Experimental results

Figures 2(c) and (d) illustrate the calculated Mahalanobis metric tensor components g_{22} and g_{11} . Since we know the diffeomorphism here, we know the ideal Mahalanobis metric shown by dashed lines in the figures. From these figures, we can see that the calculated metric can approximate the ideal Mahalanobis metric. Figure 2(b) illustrates $\partial_2 \partial_2 \log p$. From $g_{\mu\nu} = -\nabla_\mu \nabla_\nu \log p$, the connection terms correspond to the difference between Figs. 2(b) and (c), which are very large. Figure 2(e) and its rescaled version (f) illustrate the convergence of F and the square error of the Mahalanobis metric denoted by G . From this figure, we can see it takes a long time to calculate it. We have to develop a more efficient method.

7 Conclusions

In this paper, we defined GLI, GLII, and MNDM. We clarified that they are extensions of independence, isotropic independence, and Maxwell distribution in Euclidean space, respectively. We provided MNDM on a hyper-spherical surface. We gave a definition of the Mahalanobis metric, which is an extension of the Mahalanobis distance. In future work, we will obtain MNDM in other manifolds, develop a method of obtaining the Mahalanobis metric from samples, and apply it to nonlinear dimensional reduction.

References

1. Mardia, K.: Statistics of directional data. Academic Press, London and New York (1972)
2. Mahalanobis, P.: Normalization of statistical variates and the use of rectangular co-ordinates in the theory of sampling distributions. *Sankhy* **3** (1937) 1–40
3. Amari, S.: Differential geometrical method in statistics. Springer-Verlag, Berlin Heidelberg (1985)
4. Maistrov, L.: Probability theory, a historical sketch. Academic Press, New York and London (1974) (Translated and Edited by Kotz, S.).
5. Feller, W.: An introduction to probability theory and its application. third edition edn. Volume 1. John Wiley & Sons, New York (1968)
6. T.C., F.: Probability and its engineering uses. second edition edn. D. Van Nostrand Company, Princeton, New Jersey (1965)
7. Tien, C., Lienhard, J.: Statistical Thermodynamics. Revised printing edn. Hemisphere Publishing Corporation, Washington (1985)
8. Spivak, M.: A comprehensive introduction to differential geometry. Volume 1–5. Publish or Perish, Inc., Houston, Texas (1979)
9. Yamashita, Y., Ogawa, H.: Optimum image restoration and topological invariance. *System and Computers in Japan* **24** (1993) 53–63