

The Geometry of Generalized Likelihood Ratio Test

Yongqiang Cheng ^{*}, Hongqiang Wang and Xiang Li

College of Electronic Science and Engineering, National University of Defense Technology, Changsha 410073, China

^{*} Correspondence: yqcheng@nudt.edu.cn

Abstract: The generalized likelihood ratio test (GLRT) for composite hypothesis testing problems is studied from a geometric perspective. An information-geometrical interpretation of the GLRT is proposed based on the geometry of curved exponential families. Two geometric pictures of the GLRT are presented for the cases where unknown parameters are and are not the same under the null and alternative hypotheses, respectively. A demonstration of one-dimensional curved Gaussian distribution is introduced to elucidate the geometric realization of the GLRT. The asymptotic performance of the GLRT is discussed based on the proposed geometric representation of the GLRT. The study provides an alternative perspective for understanding the problems of statistical inference in the theoretical sense.

Keywords: composite hypothesis testing; generalized likelihood ratio test; maximum likelihood estimation; information geometry; statistical inference; information loss



Citation: Cheng, Y.; Wang, H.; Li, X. The Geometry of Generalized Likelihood Ratio Test. *Entropy* **2022**, *24*, 1785. <https://doi.org/10.3390/e24121785>

Academic Editor: Geert Verdoolaege

Received: 13 October 2022
Accepted: 29 November 2022
Published: 6 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The problem of hypothesis testing under statistical uncertainty arises naturally in many practical contexts. In these cases, the probability density functions (PDFs) under either or both hypotheses need not be completely specified, resulting in the inclusion of unknown parameters in the PDFs to express the statistical uncertainty in the model. The class of hypothesis testing problems with unknown parameters in the PDFs is commonly referred to as composite hypothesis testing [1]. The generalized likelihood ratio test (GLRT) is one of the most widely used approaches in composite hypothesis testing [2]. It involves estimating the unknown parameters via the maximum likelihood estimation (MLE) to implement a likelihood ratio test. In practice, the GLRT appears to be asymptotically optimal in the sense of the Neyman–Pearson criterion and usually gives satisfactory results [3]. As the GLRT combines both estimation and detection to deal with the composite hypothesis testing problem, its performance, in general, will depend on the statistical inference performance of these two aspects. However, in the literature, there is no general analytical result associated with the performance of the GLRT [1].

In recent years, the development of new theories in statistical inference has been characterized by the emerging trend of geometric approaches and their powerful capabilities, which allows one to analyze statistical problems in a unified perspective. It is important to link the GLRT to the geometrical nature of estimation and detection, which provides a new viewpoint on the GLRT. The general problem of composite hypothesis testing involves a decision between two hypotheses where the PDFs are themselves functions of unknown parameters. One approach to the understanding of performance limitations of statistical inference is via the theory of information geometry. In this context, the family of probability distributions with a natural geometrical structure is defined as a statistical manifold [4]. Information geometry studies the intrinsic properties of statistical manifolds which are endowed with a Riemannian metric and a family of affine connections derived from the log-likelihood functions of probability distributions [5]. It provides a way of analyzing the geometrical properties of statistical models by regarding them as geometric objects.

The geometric theory of statistics was firstly introduced in the 1940s by Rao [6], where the Fisher information matrix was regarded as a Riemannian metric on the manifold of probability distributions. Then, in 1972, a one-parameter family of affine connections was introduced by Chentsov in [7]. Meanwhile, Efron [8] defined the concept of *statistical curvature* and discussed its basic role in the high-order asymptotic theory of statistical inference. In 1982, Amari [5,9] developed a duality structure theory that unified all of these theories in a differential-geometric framework, leading to a large number of applications.

In the area of hypothesis testing, the geometric perspectives have acquired relevance in the analysis and development of new approaches to various testing and detection contexts. For example, Kass and Vos [10] provided a detailed introduction to the geometrical foundations of asymptotic inference of curved exponential families. Garderen [11] presented a global analysis of the effects of curvature on hypothesis testing. Dabak [12] induced a geometric structure on the manifold of probability distributions and enforced a detection theoretic specific geometry on it, while Westover [13] discussed the asymptotic limit in the problems of multiple hypothesis testing from the geometrical perspective. For the development of new approaches to hypothesis testing, Hoeffding [14] proposed an asymptotically optimal test for multinomial distributions in which the testing can be denoted in terms of the Kullback–Leibler divergence (KLD) between the empirical distribution of the measurements and the null hypothesis, where the alternate distribution is unrestricted. In the aspect of signal detection, Barbaresco et al. [15–17] studied the geometry of Bruhat–Tits complete metric space and upper-half Siegel space and introduced a matrix constant false alarm rate (CFAR) detector which improves the detection performance of the classical CFAR detection.

As more and more new analyses and new approaches have benefited from the geometric and information-theoretic perspectives of statistics, it appears to be important to clarify the geometry of existing problems that is promising to gain new ways to deal with the statistical problems. In this paper, a geometric interpretation of the GLRT is sought from the perspective of information geometry. Two pictures of the GLRT are presented for the cases where unknown parameters are and are not the same under each hypothesis, respectively. Under such an interpretation, both detection and estimation associated with the GLRT are regarded as geometric operations on the statistical manifold. As a general consideration, curved exponential families [9], which include a large number of the most common used distributions, are taken into account as the statistical model of hypothesis testing problems. A demonstration of one-dimensional curved Gaussian distribution is introduced to elucidate the geometric realization of the GLRT. The geometric structure of the curved exponential families developed by Efron [8] in 1975 and Amari [9] in 1982 provides a theoretical foundation for the analysis. The geometric formulation of the GLRT presented in this paper makes it possible for several advanced notions and conclusions in the information geometry theory to be transferred and applied to the performance analysis of the GLRT.

The main contributions of this paper are summarized as follows:

- A geometric interpretation of the GLRT is proposed based on the differential geometry of curved exponential families and duality structure theory developed by Amari [9]. Two geometric pictures of the GLRT are presented in the theoretical sense, which provides an alternative perspective for understanding the problems of statistical inference.
- The asymptotic performance of the GLRT is discussed based on the proposed geometric representation of the GLRT. The information loss when performing the MLE using a finite number of samples is related to the flatness of the submanifolds determined by the GLRT model.

In the next section, alternative viewpoints on the likelihood ratio test and the maximum likelihood estimation are introduced from the perspective of information theory. The equivalences between the Kullback–Leibler divergence, likelihood ratio test, and the MLE are highlighted. The principles of information geometry are briefly introduced in Section 3. In Section 4, the geometric interpretation of the GLRT is presented in consideration of

the geometry of curved exponential families. We present an example of the GLRT where a curved Gaussian distribution with one unknown parameter is involved, and a further discussion on the geometry of the GLRT. Finally, conclusions are obtained in Section 5.

2. Information-Theoretic Viewpoints on Likelihood Ratio Test and Maximum Likelihood Estimation

In statistics, the likelihood ratio test and maximum likelihood estimation are two fundamental concepts related to the GLRT. The likelihood ratio test is a very general form of testing model assumptions, while the maximum likelihood estimation is one of the most common approaches to parameter estimation. Both of them are associated with the Kullback–Leibler divergence [18], which is equivalent to the relative entropy [19] in information theory.

For a sequence of observations $\mathbf{x} = (x_1, x_2, \dots, x_N)^T \in \mathbb{R}^N$ which is independently and identically distributed (i.i.d.), the binary hypothesis testing problem is used to decide whether this sequence \mathbf{x} originates from the null hypothesis \mathcal{H}_0 or the alternative hypothesis \mathcal{H}_1 with probability distributions $p_0(\mathbf{x})$ and $p_1(\mathbf{x})$, respectively. The likelihood ratio is given by

$$L = \frac{p_1(\mathbf{x})}{p_0(\mathbf{x})} = \prod_{i=1}^N \frac{p_1(x_i)}{p_0(x_i)} \quad (1)$$

Assume $q(x)$ is the empirical distribution (frequency histogram acquired via Monte Carlo tests) of observed data. For large N , in accordance with the strong law of large numbers [20], the log likelihood ratio test in the Neyman–Pearson formulation $\ln L \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \gamma$ is equivalent to

$$D(q\|p_0) - D(q\|p_1) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \frac{1}{N} \gamma \triangleq \gamma' \quad (2)$$

where $\underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}}$ denotes that the test is to decide \mathcal{H}_1 if “ $>$ ” is satisfied, or to decide \mathcal{H}_0 , and vice versa. The quantity

$$D(q\|p) = \int_{\mathbb{R}^N} q(x) \ln \frac{q(x)}{p(x)} dx \quad (3)$$

is the KLD from $q(x)$ to $p(x)$. Note that x is dropped from the notion D for simplifying the KLD expression without confusion.

Equation (2) indicates that the likelihood ratio test is equivalent to choosing the hypothesis that is “closer” to the empirical distribution in the sense of the KLD. The test can be referred to as a generalized minimum dissimilarity detector in a geometric viewpoint.

Now, consider another, slightly different problem where the observations \mathbf{x} are from a statistical model represented by $p(\mathbf{x}|\boldsymbol{\theta})$ with unknown parameters $\boldsymbol{\theta}$. The problem is to estimate the unknown parameters $\boldsymbol{\theta}$ based on observations \mathbf{x} . The likelihood function for the underlying estimation problem is

$$p(\mathbf{x}|\boldsymbol{\theta}) = \prod_{i=1}^N p(x_i|\boldsymbol{\theta}) \quad (4)$$

In a similar way, for large N , maximizing the likelihood (4) to find the maximum likelihood estimate of $\boldsymbol{\theta}$ is equivalent to finding $\boldsymbol{\theta}$, which minimizes the KLD $D(q\|p_{\boldsymbol{\theta}})$, i.e.,

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} D(q\|p_{\boldsymbol{\theta}}) \quad (5)$$

where $p_{\boldsymbol{\theta}}$ is used as a surrogate for $p(\mathbf{x}|\boldsymbol{\theta})$.

The above results provide an information-theoretic view to the problem of hypothesis testing and maximum likelihood estimation in statistics. From the perspective of informa-

tion difference, these results have profound geometric meanings and can be geometrically analyzed and viewed in the framework of information geometry theory, from which additional insights into the analysis of these statistical problems, as well as their geometric interpretations, are obtained.

3. Principles of Information Geometry

3.1. Statistical Manifold

Information geometry studies the natural geometric structure of the parameterized family of probability distributions $S = \{p(x|\theta)\}$ specifying by a parameter vector $\theta := [\theta_1, \dots, \theta_n]$, in which x is the samples of a random variable X . When the probability measure on the sample space is continuous and differentiable and the mapping $\theta \mapsto p(x|\theta)$ is injective [5], the family S is considered as a statistical manifold with θ as its coordinate system [4].

Figure 1 demonstrates the diagram of a statistical manifold. For a given parameter vector $\theta \in \Theta \subset \mathbb{R}^n$, the measurement x in the sample space X is an instantiation of a probability distribution $p(x|\theta)$. Each $p(x|\theta)$ in the family of distributions is specified by a point $s(\theta)$ on the manifold S . The n -dimensional statistical manifold is composed of the parameterized family of probability distributions $S = \{p(x|\theta)\}$ with θ as a coordinate system of S .

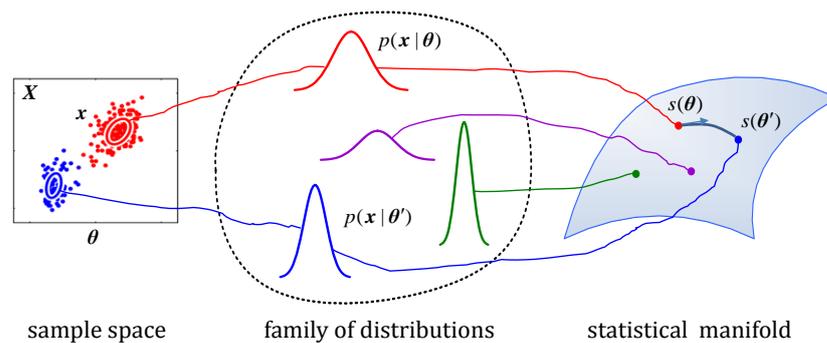


Figure 1. Diagram of a statistical manifold. θ and $s(\theta')$ denote parameters of the family of distributions from different samples X . The connection on the statistical manifold S represents a geodesic (the shortest line) between points $s(\theta)$ and $s(\theta')$. The length of the geodesic serves as a distance measure between two points on the manifold. The arrow on the geodesic starting from the point $s(\theta)$ denotes the tangent vector, which gives the direction of the geodesic.

Various families of probability distributions correspond to specific structures of the statistical manifold. Information geometry takes the statistical properties of samples as the geometric structure of a statistical manifold, and utilizes differential geometry methods to measure the variation of information contained in the samples.

3.2. Fisher Information Metric and Affine Connections

The metric and connections associated with a manifold are two important concepts in information geometry. For a statistical manifold consisting of a parameterized family of probability distributions, the Fisher information matrix (FIM) is usually adopted as a Riemannian metric tensor of the manifold [6], which is defined by the inner product between tangent vectors at a point on the manifold. It is denoted by $G(\theta) = [g_{ij}(\theta)]$, where

$$g_{ij}(\theta) = E \left\{ \frac{\partial \log p(x|\theta)}{\partial \theta_i} \cdot \frac{\partial \log p(x|\theta)}{\partial \theta_j} \right\} \tag{6}$$

where $\{\partial \log(\cdot) / \partial \theta_i\}$ is considered as a basis for the vector space of random variable X . The tangent space of S at θ , denoted as $\mathcal{T}_\theta(S)$, is identified as the vector space. Based on

the above definition, the FIM metric determines how the information distance is measured on the statistical manifold.

When considering the relationships between two tangent spaces \mathcal{T}_θ and $\mathcal{T}_{\theta+d\theta}$ at two neighboring points θ and $\theta + d\theta$ (d is the differential operator), an affine connection is defined by which the two tangent spaces become feasible for comparison. When the connection coefficients are all identically 0, then S is flat manifold that “locally looks like” a Euclidean space with zero curvatures everywhere. The most commonly used connection is called α -connections [9],

$$\Gamma_{jim}^\alpha(\theta) = E_\theta \left\{ \partial_j \partial_i l(x, \theta) \partial_m l(x, \theta) \right\} + \frac{1-\alpha}{2} E_\theta \left\{ \partial_j l(x, \theta) \partial_i l(x, \theta) \partial_m l(x, \theta) \right\} \quad (7)$$

where Γ_{jim}^α denotes the connection coefficients with $i, j, m = 1, \dots, n$, and

$$\partial_i = \frac{\partial}{\partial \theta_i} \quad \text{and} \quad l(\theta, x) = \log p(x|\theta). \quad (8)$$

In (7), $\alpha = 0$ corresponds to the Levi–Civita connection, while $\alpha = 1$ defines the e -connection and $\alpha = -1$ defines the m -connection. Under the e -connection and m -connection, an exponential family with natural parameter θ coordinate and a mixture family with expectation parameter η coordinate are both flat manifolds [9]. Statistical inference with respect to the exponential family greatly benefits from the geometric properties of the flat manifold. By using the methods of differential geometry, many additional insights into the intrinsic structure of probability distributions can be obtained, which opens a new perspective on the analysis of statistical problems. In the next section, a geometric interpretation of the GLRT and further discussions are sought based on the principles of information geometry.

4. Geometry of the Generalized Likelihood Ratio Test

As a general treatment, the curved exponential families, which encapsulate many important distributions for real-world problems, are considered as the statistical model for the hypothesis testing problems discussed in this paper. In this section, the MLE solution to parameter estimation for curved exponential families is derived. We then present two pictures of the GLRT, which are sketched based on the geometric structure of the curved exponential families developed by Efron [8] in 1975 and Amari [9] in 1982, to illustrate the information geometry of the GLRT. An example of the GLRT for a curved Gaussian distribution with a single unknown parameter is given, which is followed by a further discussion on the geometric formulation of the GLRT.

4.1. The MLE Solution to Statistical Estimation for Curved Exponential Families

Exponential families contain lots of the most commonly used distributions, including the normal, exponential, Gamma, Beta, Poisson, Bernoulli, and so on [21]. The curved exponential families are the distributions whose natural parameters are nonlinear functions of “local” parameters. The canonical form of a curved exponential family [9] is expressed as

$$p(x|u) = \exp \left\{ C(x) + \theta^T(u) F(x) - \varphi(\theta(u)) \right\} = p(x|\theta(u)) \quad (9)$$

where $x \in X$ is a vector of samples, $\theta := [\theta_1, \dots, \theta_n]$ are the natural parameters, $u \in \mathbb{R}^m$ ($m < n$) are local parameters standing for the parameters of interest to be estimated, which is specified by (9), while $F(x) := [F_1(x), \dots, F_n(x)]^T$ denote sufficient statistics with respect to $\theta = (\theta_1, \dots, \theta_n)$, which take values from the sample space X . φ is the potential

function of the curved exponential family and it is found from the normalization condition $\int_X p(x|\theta)dx = 1$, i.e.,

$$\varphi(\theta) = \log \int_X \exp \left\{ C(x) + \sum_i^n (\theta_i F_i(x)) \right\} dx \tag{10}$$

The term “curved” is due to the fact that the curved exponential family in (9) is a submanifold of the canonical exponential family $p(x|\theta)$ by the embedding $u \rightarrow \theta(u)$.

Let $l(\theta, x) = \log p(x|\theta)$ be the log-likelihood and $\nabla_u \theta$ be the Jacobian matrix of the natural parameter θ . According to (9),

$$\begin{aligned} \nabla l(\theta, x) &= \nabla \left\{ \theta^T(u)F(x) - \varphi(\theta(u)) \right\} \\ &= \nabla \theta^T(u)[F(x) - \eta(u)] \end{aligned} \tag{11}$$

where $\eta(u)$ is the expectation of the sufficient statistics $F(x)$, i.e.,

$$\eta(u) := E_{p(\cdot, u)} \{ F(x) \} \tag{12}$$

and is called the expectation parameter, which defines a distribution of mixture family [4]. The natural parameter $\theta(u)$ and expectation parameter $\eta(u)$ are connected by the Legendre transformation [9], as

$$\eta = \nabla_\theta \varphi(\theta), \quad \theta = \nabla_\eta \phi(\eta) \tag{13}$$

where $\phi(\eta)$ is defined by

$$\phi(\eta) = \max_\theta \{ \eta^T \theta - \varphi(\theta) \} \tag{14}$$

Therefore,

$$\nabla_u \varphi(u) = \nabla_u \theta^T(u) \nabla_\theta \varphi(\theta) = \nabla_u \theta^T(u) \eta(u) \tag{15}$$

Thus, the maximum likelihood estimator \hat{u} of the local parameter in (9) can be obtained by the following likelihood equation:

$$\nabla_u l(\hat{u}) = \nabla_u \log p(x|u) = \nabla_u \theta^T(\hat{u}) [F(x) - \eta(\hat{u})] = 0 \tag{16}$$

Equation (16) indicates that the solution to the MLE can be found by mapping the data $F(x)$ onto $\mathcal{F}_B := \{ \eta(u) : u \in \mathbb{R}^m \}$ orthogonally to the tangent of $\mathcal{F}_A := \{ \theta(u) : u \in \mathbb{R}^m \}$. As $\theta(u)$ and $\eta(u)$ live in two different spaces \mathcal{F}_A and \mathcal{F}_B , the inner product between dual spaces is defined as $\langle \theta(u), \eta(u) \rangle_\Gamma := \theta(u)^T \cdot \Gamma \cdot \eta(u)$ with a metric Γ . For the flat manifold, the identity matrix serves as the metric Γ . By analogy with the MLE for the universal distribution given by (5), Hoeffding [14] presented another interpretation for the MLE of the curved exponential family. In the interpretation, $\eta(\hat{u})$ represents a point in \mathcal{F}_B which is located closest to the data point in the sense of the Kullback–Leibler divergence, i.e.,

$$\hat{u} = \arg \min_{\eta(u) \in \mathcal{F}_B} D[F(x) || \eta_u] \tag{17}$$

where $D[F(x) || \eta_u]$ denotes the Kullback–Leibler divergence from the multivariate joint distributions of $F(x)$ to η_u .

Based on the above analysis, there are two important spaces related to a curved exponential family. One is called the natural parameter space, denoted by $\{ \theta \} \subset \mathcal{A}^n$, which denotes the enveloping space including all the distributions of exponential families, and the other is called the expectation parameter space, denoted by $\{ \eta \} \subset \mathcal{B}^n$, denoting the dual space of \mathcal{A}^n . The two spaces are “dual” with each other and flat under the e -connection and m -connection, respectively. The curved exponential family (9) is regarded as submanifolds embedded in the two spaces, and the data can also be immersed in these spaces in the form of sufficient statistics $F(x)$. Consequently, the estimators, such as the MLE given by (16),

associated with the curved exponential families can be geometrically performed in the two spaces.

4.2. Geometric Demonstration of the Generalized Likelihood Ratio Test

As mentioned earlier, the GLRT is one of the most widely used approaches in composite hypothesis testing problems with unknown parameters in the PDFs. The data x have the PDF $p(x|u_0; \mathcal{H}_0)$ under hypothesis \mathcal{H}_0 and $p(x|u_1; \mathcal{H}_1)$ under hypothesis \mathcal{H}_1 , where u_0 and u_1 are unknown parameters under each hypothesis. The GLRT enables a decision by means of replacement of the unknown parameters by their maximum likelihood estimates (MLEs) to implement a likelihood ratio test. The GLRT decides \mathcal{H}_1 if

$$L_G(x) = \frac{p(x|\hat{u}_1; \mathcal{H}_1)}{p(x|\hat{u}_0; \mathcal{H}_0)} > \gamma \tag{18}$$

where \hat{u}_i is the MLE of u_i (by maximizing $p(x|u_i; \mathcal{H}_i)$).

From the perspective of information geometry, the probability distribution $p(x|u_i; \mathcal{H}_i)$ is an element of the parameterized family of PDFs $S = \{p(x|u), u \in \Omega\}$, where $\Omega \subset \mathbb{R}^m$ is the parameter set. For the curved exponential family $S = \{p(x|u)\}$, it can be regarded as a submanifold embedding in the natural parameter space $\{\theta\} \subset \mathcal{A}^n$, which includes all the distributions of exponential families. The curved exponential family S can be represented by a curve (or surface) $\{\theta = \theta(u)\}$ embedded in the enveloping space \mathcal{A}^n by the nonlinear mapping $u \rightarrow \theta(u)$. The expectation parameter space $\{\eta\} \subset \mathcal{B}^n$ of S is a dual flat space to the natural parameter space $\{\theta\}$, while the “realizations” of sufficient statistics $F(x)$ of the distribution $p(x|\theta)$ can be immersed in this space. Consequently, the MLE is performed in the space \mathcal{B}^n by mapping the samples $F(x)$ onto the submanifold specified by $\{\eta = \eta(u)\}$ under the m -projection.

As the parameters u_0 and u_1 , as well as their dimensionalities, may or may not be the same under the null and alternative hypotheses, two pictures of the GLRT are presented for the two cases: one is with the same unknown parameters under each hypotheses and the other is with different parameters or different dimensionalities. The picture for the first case is illustrated in Figure 2a. In this case, distributions under two hypotheses share the same form and the same unknown parameter u . However, the parameter takes different value sets under different hypotheses. The family of $S = \{p(x|u)\}$ can be smoothly embedded as a surface \mathcal{F}_B specified by $\{\eta(u) : u \in \mathbb{R}^m\}$ in the space \mathcal{B}^n . The hypotheses $p(x|u_i; \mathcal{H}_i)$ with unknown u_i define two “uncertainty volumes” Ω_0 and Ω_1 on the submanifold \mathcal{F}_B . These volumes are the collections of probability distributions specified by the value sets of the unknown parameter u_i . The measurements x are immersed in \mathcal{B}^n in the form of sufficient statistics $F(x)$. Consequently, the MLE can be found by “mapping” the samples $F(x)$ onto the uncertainty volumes Ω_0 and Ω_1 on \mathcal{F}_B . The points p_0 and p_1 in Figure 2 are the corresponding projections, i.e., the MLEs of the unknown parameter under two hypotheses. As indicated in (17), the MLEs can also be obtained by finding the points on Ω_0 and Ω_1 which are located closest to the data point in the sense of KLD, i.e.,

$$\hat{u}_0 = \arg \inf_{\eta(u) \in \Omega_0} D[F(x)||\eta_u], \quad \hat{u}_1 = \arg \inf_{\eta(u) \in \Omega_1} D[F(x)||\eta_u] \tag{19}$$

and the corresponding minimum KLDs can be represented by

$$D_0 = D[F(x)||p_0], \quad D_1 = D[F(x)||p_1], \tag{20}$$

respectively.

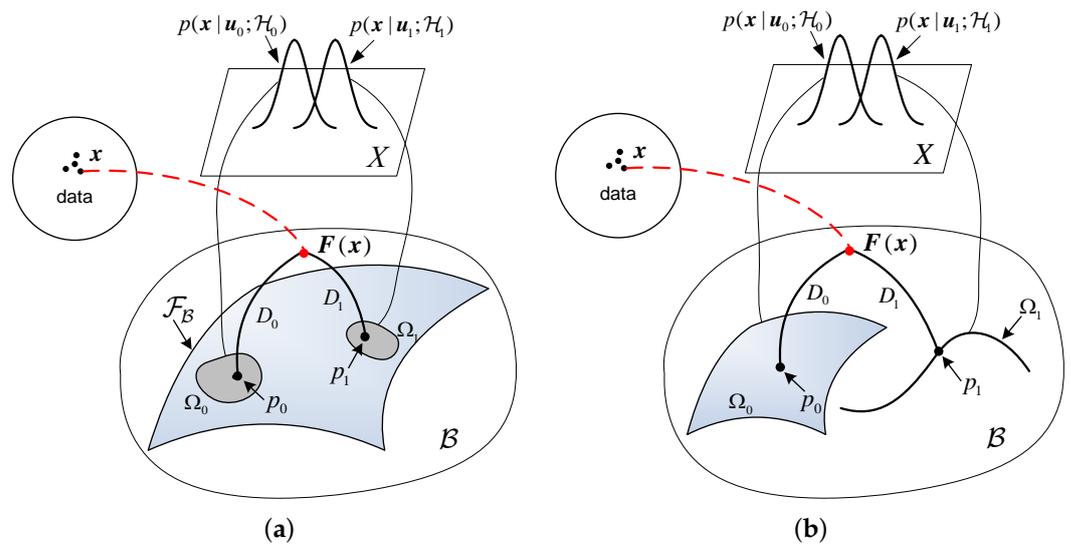


Figure 2. Geometry of the GLRT. (a) The case for the same unknown parameters under two hypotheses. (b) The case for different unknown parameters and different dimensionalities under two hypotheses.

It should be emphasized that the above “mapping” is a general concept. When the parameters to be estimated are not restricted by a given “value set”, the MLE is simply obtained by maximizing the likelihood and the projections will fall onto the submanifold $\mathcal{F}_{\mathcal{B}}$. However, if the parameters to be estimated are restricted in a given “value set”, the MLE should be operated by maximizing the likelihood with respect to the given parameter space. In the case where the projections fall outside the “uncertainty volumes”, the MLE solutions are given by those points which are closest to the data point described by (19).

Let $\mathcal{R}(\eta_0, \rho)$ be a *divergence sphere* centered at η_0 with radius ρ ; that is, the submanifold of the enveloping space \mathcal{B}^n consisting of points η for which the KLD $D(\eta_0||\eta)$ is equal to ρ . Denote this divergence sphere by

$$\mathcal{R}(\eta_0, \rho) = \{\eta \in \mathcal{B}^n \mid D(\eta_0||\eta) = \rho\} \tag{21}$$

Then, the closest point in (19) may be more easily found via the divergence sphere with center $F(x)$ and radius D_i tangent to Ω_i at p_i , as illustrated in Figure 3. Consequently, according to (2), the GLRT can be geometrically performed by comparing the difference between the minimum KLDs D_0 and D_1 with a threshold γ' , i.e.,

$$D[F(x)||p_0] - D[F(x)||p_1] \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma' \tag{22}$$

In practice, the Neyman–Pearson criterion is commonly employed to determine the threshold γ' in (22) and the detector is of maximum probability of detection P_D under a given probability of false alarm P_F . As a commonly used performance index, the missingprobability P_M usually decays exponentially as the sample size increases. The rate of exponential decay can be represented by [22], as

$$K \triangleq \lim_{n \rightarrow \infty} -\frac{1}{N} \log P_M \tag{23}$$

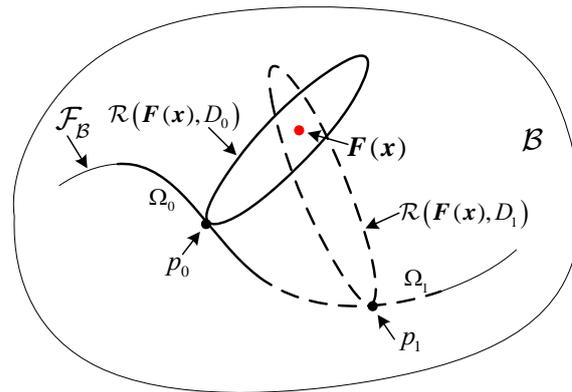


Figure 3. Illustration of the mapping via divergence spheres.

Based on Stein’s lemma, for a constant false-alarm constraint, the best error exponent is related to the Kullback–Leibler divergence $D(p_0||p_1)$ from p_0 to p_1 [23], i.e.,

$$K = D(p_0||p_1) \tag{24}$$

and

$$P_M \doteq 2^{-NK} \tag{25}$$

where \doteq denotes the first-order equivalence in the exponent. For example,

$$a_n \doteq b_n \quad \text{means} \quad \lim_{n \rightarrow \infty} \frac{1}{n} \log \frac{a_n}{b_n} = 0. \tag{26}$$

In the above sense, the KLD from p_0 to p_1 is equivalent to the signal-to-noise ratio (SNR) of the underlying detection problem. Therefore, information geometry offers an insightful geometrical explanation for the detection performance of a Neyman–Pearson detector.

In the second case, the dimensionality of the unknown parameters u_0 and u_1 is different, while the dimensionality of the enveloping spaces is common for both hypotheses due to the same measurements x . However, the hypotheses may correspond to two separated submanifolds, Ω_0 and Ω_1 , embedded in \mathcal{B}^n caused by the different dimensionality between the unknown parameters. As illustrated in Figure 2b, a surface and a curve are used to denote the submanifolds Ω_0 and Ω_1 , corresponding to the two hypotheses, respectively. Similar to the first case, the GLRT with different unknown parameters may also be geometrically interpreted.

4.3. A Demonstration of One-Dimensional Curved Gaussian Distribution

Consider the following detection problem:

$$\begin{cases} \mathcal{H}_0 : u < 0 \\ \mathcal{H}_1 : u > 0 \end{cases} \tag{27}$$

The measurement originates from a curved Gaussian distribution

$$x \sim \mathcal{N}(u, u^2 a^2) \tag{28}$$

where a is a positive constant and u is an unknown parameter.

The probability density function of the measurement is

$$p(x|u) = \frac{1}{\sqrt{2\pi a|u|}} \exp\left[-\frac{(x-u)^2}{2a^2 u^2}\right] \tag{29}$$

By reparameterization, the probability density function can be represented in the general form of a curved exponential family as

$$\begin{aligned} p(x|u) &= \exp\left\{\frac{1}{a^2u}x - \frac{1}{2a^2u^2}x^2 - \frac{1}{2a^2} - \ln(a|u|) - \frac{1}{2}\ln(2\pi)\right\} \\ &= \exp\left\{\ln a + \theta_1x + \theta_2x^2 - \left[-\frac{\theta_1^2}{4\theta_2} - \frac{1}{2}\ln(-\theta_2) + \frac{1}{2}\ln\pi\right]\right\} \\ &= \exp\{C(x) + F_1(x)\theta_1 + F_2(x)\theta_2 - \varphi(\theta_1, \theta_2)\} \end{aligned} \quad (30)$$

where $C(x) = \ln a$ and the potential function φ is

$$\varphi(\theta_1, \theta_2) = -\frac{\theta_1^2}{4\theta_2} - \frac{1}{2}\ln(-\theta_2) + \frac{1}{2}\ln\pi \quad (31)$$

The above distributions with local parameter u correspond to a one-dimensional curved exponential family embedded in the natural parameter space \mathcal{A} . The natural coordinates are

$$\theta_1 = \frac{1}{a^2u}, \quad \theta_2 = -\frac{1}{2a^2u^2} \quad (32)$$

which defines a parabola (denoted by \mathcal{F}_A)

$$\theta_2 = -\frac{a^2}{2}\theta_1^2 \quad (33)$$

in \mathcal{A} . The underlying distribution (28) can also be represented in the expectation parameter space \mathcal{B} with expectation coordinates

$$\eta_1 = \frac{\partial\varphi}{\partial\theta_1} = u, \quad \eta_2 = \frac{\partial\varphi}{\partial\theta_2} = (a^2 + 1)u^2 \quad (34)$$

which also defines a parabola (denoted by \mathcal{F}_B)

$$\eta_2 = (a^2 + 1)\eta_1^2 \quad (35)$$

in \mathcal{B} .

The sufficient statistics $F(x)$ obtained from samples x can be represented by

$$F(x) = [x, x^2]^T \quad (36)$$

Figure 4 shows the expectation parameter space and illustrates geometric interpretation of the underlying GLRT, where the blue parabola in the figure denotes embedding of the curved Gaussian distribution with parameter u . The submanifolds associated with two hypotheses can be geometrically represented by the blue parabolas (specified by $\eta_1 < 0$ and $\eta_1 > 0$, respectively). Without loss of generality, assume that $a = 1, u = 2$. The blue dots signify $N = 100$ observations (measurements) in the expectation parameter space with the coordinates (x, x^2) . The statistical mean of the measurements are used to calculate the sufficient statistics $F(x)$ which are denoted by a red asterisk. The MLEs of parameter u under two hypotheses are obtained by finding the points on the two submanifolds which are closest to the data point in the sense of KLD. According to (22), the GLRT can be geometrically performed by comparing the difference between the minimum KLDs D_0 and D_1 with a threshold γ' .

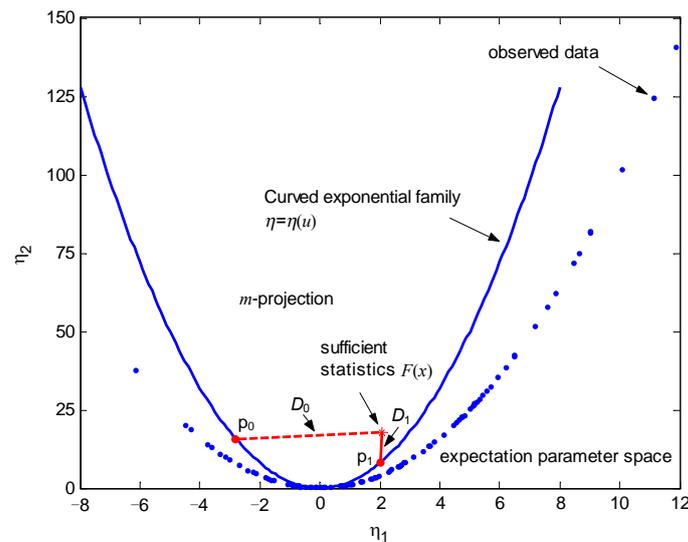


Figure 4. The geometric interpretation of the GLRT for one-dimensional curved Gaussian distribution.

4.4. Discussions

The geometric formulation of the GLRT presented above provides additional insights into the GLRT. To the best of our knowledge, there is no general analytical result associated with the performance of the GLRT in the literature [1]. The asymptotic analysis is only valid under the conditions that (1) the data sample size is large; and (2) the MLE asymptotically attains the Cramér-Rao lower bound (CRLB) of the underlying estimation problems.

It is known that the MLE with natural parameters is a sufficient statistic for an exponential family, and achieves the CRLB if a suitable measurement function is chosen for the estimation [8]. For the curved exponential families the MLE is not, in general, an efficient estimator, which means that the variance of MLE may not achieve CRLB with a finite number of samples. This indicates that when using a finite number of samples there will be a deterioration in performance for both MLE and GLRT when the underlying statistical model is a curved exponential family. There will be an inherent information loss (compared with the Fisher information) when implementing an estimation process if the statistical model is of nonlinearity. Roughly speaking, if the embedded submanifold \mathcal{F}_B in Figure 2a and Ω_0, Ω_1 in Figure 2b are curved, the MLEs will not achieve the CRLB due to the inherent information loss caused by the non-flatness of the statistical model. The information loss may be quantitatively calculated using the e -curvature of the statistical model [9].

Consequently, if the statistical model associated with a GLRT is not flat, i.e., the submanifolds shown in Figure 2 are curved, there will be a deterioration in performance for the GLRT using a finite number of samples. As sample size N increases, the sufficient statistics $F(x)$ will be better matched to the statistical model and thus closer to the submanifolds (see Figure 2), and the divergence from data to the submanifold associated with the true hypothesis \mathcal{H}_i will be shorter. Asymptotically, as $N \rightarrow \infty$, the sufficient statistics will fall onto the submanifold associated with the true hypothesis \mathcal{H}_i , so that the corresponding divergence D_i reduces to zero. By then, the GLRT achieves a perfect performance.

5. Conclusions

In this paper, the generalized likelihood ratio test is addressed from a geometric viewpoint. Two pictures of the GLRT are presented in the philosophy of the information geometry theory. Both the detection and estimation associated with a GLRT are regarded as geometric operations on the manifolds of a parameterized family of probability distributions. As demonstrated in this work, the geometric interpretation of GLRT provides additional insights in the analysis of GLRT.

Potentially, more constructive analysis can be generalized based on the information geometry of GLRT. For example, the error exponent defined by (24) and (25) provides a useful performance index for the detection process associated with GLRT. When p_0 and p_1 in (24) are the estimates of an MLE (rather than the true values) of unknown parameters under each hypothesis, there may be a deterioration in performance in the estimation process. Determining how to incorporate such an “estimation loss” into the error exponent is an issue. Another open issue is the GLRT with PDFs of different forms for each hypothesis, which leads to a different distribution embedding associated with each hypothesis.

Author Contributions: Y.C. proposed the original ideas and performed the research. H.W. conceived and designed the demonstration of one-dimensional curved Gaussian distribution. X.L. reviewed the paper and provided useful comments. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China under grant No. 61871472.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kay, S.M. *Fundamentals of Statistical Signal Processing (Volume II): Detection Theory*; Prentice Hall PTR: Hoboken, NJ, USA, 1993.
2. Liu, W.; Liu, J.; Hao, C.; Gao, Y.; Wang, Y. Multichannel adaptive signal detection: Basic theory and literature review. *Sci. China Inf. Sci.* **2022**, *65*, 5–44. [[CrossRef](#)]
3. Levitan, E.; Merhav, N. A competitive Neyman-Pearson approach to universal hypothesis testing with applications. *IEEE Trans. Inform. Theory* **2002**, *48*, 2215–2229. [[CrossRef](#)]
4. Amari, S. Information geometry on hierarchy of probability distributions. *IEEE Trans. Inform. Theory* **2001**, *47*, 1701–1711. [[CrossRef](#)]
5. Amari, S.; Nagaoka, H. *Methods of Information Geometry (Translations of Mathematical Monographs)*; American Mathematical Society and Oxford University Press: New York, NY, USA, 2000.
6. Rao, C.R. Information and accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.* **1945**, *37*, 81–91.
7. Chentsov, N.N. *Statistical Decision Rules and Optimal Inference*; English Translation; American Mathematical Society: Providence, RI, USA, 1982.
8. Efron, B. Defining the curvature of a statistical problem (with applications to second order efficiency). *Ann. Stat.* **1975**, *3*, 1189–1242. [[CrossRef](#)]
9. Amari, S. Differential geometry of curved exponential families—curvatures and information loss. *Ann. Stat.* **1982**, *10*, 357–385. [[CrossRef](#)]
10. Kass, R.E.; Vos, P.W. *Geometrical Foundations of Asymptotic Inference*; John Wiley & Sons Inc: Hoboken, NJ, USA, 1997.
11. Garderen, K.J.V. An alternative comparison of classical tests: Assessing the effects of curvature. In *Applications of Differential Geometry to Econometrics*; Cambridge University Press: Cambridge, UK, 2000; pp. 230–280.
12. Dabak, A.G. A Geometry for Detection Theory. Ph.D. Thesis, Rice University, Houston, TX, USA, 1993.
13. Westover, M.B. Asymptotic geometry of multiple hypothesis testing. *IEEE Trans. Inform. Theory* **2008**, *54*, 3327–3329. [[CrossRef](#)] [[PubMed](#)]
14. Hoeffding, W. Asymptotically optimal tests for multinomial distributions. *Ann. Math. Statist.* **1965**, *36*, 369–408. [[CrossRef](#)]
15. Barbaresco, F. Innovative tools for radar signal processing based on Cartan’s geometry of SPD matrices & information geometry. In Proceedings of the 2008 IEEE Radar Conference, Rome, Italy, 26–30 May 2008.
16. Barbaresco, F. New foundation of radar Doppler signal processing based on advanced differential geometry of symmetric spaces: Doppler matrix CFAR and radar application. In Proceedings of the International Radar Conference, Bordeaux, France, 12–16 October 2009.
17. Barbaresco, F. Robust statistical radar processing in Fréchet metric space: OS-HDR-CFAR and OS-STAP processing in Siegel homogeneous bounded domains. In Proceedings of the International Radar Symposium, Leipzig, Germany, 7–9 September 2011.
18. Kullback, S. *Information Theory and Statistics*; Dover Publications: New York, NY, USA, 1968.
19. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; Wiley: New York, NY, USA, 1991.
20. Etemadi, N. An elementary proof of the strong law of large numbers. *Probab. Theory Relat. Fields* **1981**, *55*, 119–122. [[CrossRef](#)]
21. Garcia, V.; Nielsen, F. Simplification and hierarchical representations of mixtures of exponential families. *Signal Process.* **2010**, *90*, 3197–3212. [[CrossRef](#)]

-
22. Sung, Y.; Tong, L.; Poor, H.V. Neyman-Pearson detection of Gauss-Markov signals in noise: closed-form error exponent and properties. *IEEE Trans. Inform. Theory* **2006**, *52*, 1354–1365. [[CrossRef](#)]
 23. Chamberland, J.; Veeravalli, V.V. Decentralized detection in sensor networks. *IEEE Trans. Signal Process.* **2003**, *51*, 407–416. [[CrossRef](#)]