The Geometry of Signal Detection with Applications to Radar Signal Processing

Yongqiang Cheng *, Xiaoqiang Hua, Hongqiang Wang, Yuliang Qin and Xiang Li

School of Electronic Science and Engineering, National University of Defense Technology, Changsha 410073, China; hxq712@yeah.net (X.H.); oliverwhq@tom.com (H.W.); yuliang.qin@gmail.com (Y.Q.); lixiang01@vip.sina.com (X.L.)

* Correspondence: nudtyqcheng@gmail.com; Tel.: +86-731-8457-4452

Abstract: The problem of hypothesis testing in the Neyman–Pearson formulation is considered from a geometric viewpoint. In particular, a concise geometric interpretation of deterministic and random signal detection in the philosophy of information geometry is presented. In such a framework, both hypotheses and detectors can be treated as geometrical objects on the statistical manifold of a parameterized family of probability distributions. Both the detector and detection performance are geometrically elucidated in terms of the Kullback–Leibler divergence. Compared to the likelihood ratio test, the geometric interpretation provides a consistent but more comprehensive means to understand and deal with signal detection problems in a rather convenient manner. Example of the geometry based detector in radar constant false alarm rate (CFAR) detection is presented, which shows its advantage over the classical processing method.

Keywords: hypothesis testing; signal detection; information geometry; likelihood ratio test; Neyman–Pearson detection; matrix CFAR detector

1. Introduction

Signal detection in the presence of noise is a basic issue in statistical signal processing and is fundamental to target detection applications in modern radar signal processing systems. Classical approaches to binary hypothesis testing are usually in the form of a likelihood ratio test. It has been proved that the optimal detector for a known deterministic signal in Gaussian noise is termed the matched filter, while for a random signal with known statistics, it is named the estimator–correlator in which the signal estimate is correlated with the data [1]. The essence of a basic detection problem can be regarded as decision or discrimination between two probability distributions with different models or parameters. While the likelihood ratio test provides an optimal solution to hypothesis testing problems, its meaning is not obvious in terms of the discrimination nature of a detection problem. The term "discrimination" is usually related to a decision or comparison in terms of distance between two items. There is no doubt that a better understanding of the detection problem will benefit from interpretations of its discrimination nature. Furthermore, it is promising to develop new approaches revealed by such interpretations to deal with the detection problem itself.

In statistics and information theory, a parameterized family of identically distributed probability distributions is usually regarded as a statistical manifold [2]. The statistical manifold of probability distributions has a natural geometrical structure that characterizes the intrinsic statistical properties of the family of distributions. Information geometry is the study of intrinsic properties of manifolds of probability distributions [3], where the ability of the data to discriminate those distributions is translated into a Riemannian metric [4]. The main tenet of information geometry is that many important notions in probability theory, information theory and statistics can be treated as structures in differential

geometry by regarding a space of probabilities as a differentiable manifold endowed with a Riemannian metric and a family of affine connections [5]. By providing a means to analyze the Riemannian geometric properties of various families of probability density functions, information geometry offers comprehensive results about statistical models simply by considering them as geometrical objects.

The geometric theory of statistics, pioneered in the 1940s by Rao [4] and further developed by Chentsov [6], Efron [7,8] and Amari [5,9], has found a wide range of applications in the theory of statistical inference [10,11], the study of Boltzmann machine [12], the learning of neural networks [13], the Expectation–Maximization (EM) algorithm [14], all with certain degrees of success. In the last two decades, its application has spanned several discipline areas such as information theory [15,16], systems theory [17,18], mathematical programming [19], and statistical physics [20]. As this new general theory reveals the capability of defining a new perspective on existing questions, many researchers are extending their work on this geometric theory of information to new areas of application and interpretation. For example, in the area of communication coding/decoding and channel analysis, the most important milestone is the work of Richardson where the geometric perspective clearly indicates the relationship between turbo decoding and maximum-likelihood decoding [21], while a geometric characterization of multiuser detection for synchronous direct-sequence code division multiple access (DS/CDMA) channels is presented in [22]; In the area of estimation and filtering, Smith [23] studied the intrinsic Cramér–Rao bounds on estimation accuracy for the estimation problems on arbitrary manifolds, where the set of intrinsic coordinates are not apparent. Srivastava et al. [24,25] addressed the geometric subspace estimation and target tracking problems under a Bayesian framework. In particular, in the area of hypothesis testing and signal detection, the geometrical interpretation of multiple hypothesis testing problems in the asymptotic limit was developed by Westover [26], while the interpretation of Bayes risk error in Bayesian hypothesis testing was demonstrated by Varshney [27] and it has been proved that the Bayes risk error is a member of the class of Bregman divergences. Furthermore, Barbaresco et al. [28–30] developed a new Matrix constant false alarm rate (CFAR) detection method based on Bruhat–Tits complete metric space and upper-half Siegel space in symplectic geometry, which improves the detection performance with respect to the classical CFAR detection approach based on Doppler Filter Bank or fast Fourier transform (FFT).

Information geometry opens a new prospective to study the intrinsic geometrical nature of information theory and provides a new way to deal with existing statistical problems. For signal detection in statistical signal processing, information geometry can also provide new viewpoints and better understandings in the analysis of the detection process where both hypotheses and detector can be regarded as geometrical objects on the statistical manifold. In this paper, geometrical interpretations of deterministic and random signal detection are presented from the viewpoint of information geometry. Under such interpretations, the optimal detector is a "generalized minimum distance detector" where the decision is made by comparing distances from the signal distribution estimates to two hypotheses in the sense of the Kullback–Leibler divergence (KLD). Results, presented in this paper, illustrate that the generalized minimum distance detector provides a consistent but more comprehensive means to understand signal detection problems compared to the likelihood ratio test. An example of the geometry based detector in radar CFAR detection is presented, which shows its advantage over the classical processing method.

In the next section, the equivalence between likelihood ratio test and the Kullback–Leibler divergence is discussed, followed by the geometric interpretation of classical likelihood ratio test from the viewpoint of information geometry. In Section 3 and Section 4, the geometry of deterministic and random signal detection as well as the detection performance is elucidated geometrically, respectively. An example of the geometry based detector in radar CFAR detection is presented in Section 5. Finally, conclusions are made in Section 6.

2. Geometric Interpretation of the Classical Likelihood Ratio Test

2.1. Equivalence between Likelihood Ratio Test and Kullback–Leibler Divergence

In statistics, the likelihood ratio test is a basic form for hypothesis testing. Suppose $x_1, x_2, \ldots, x_N$ are N independent and identically distributed (i.i.d.) observations from a statistical model whose probability density function is $q(x)$, i.e., $x_1, x_2, \ldots, x_N \overset{i.i.d.}{\sim} q(x)$. Consider two hypotheses for $q(x)$, denoted by $p_0(x)$ and $p_1(x)$, which are referred to as the null hypothesis and alternative hypothesis, respectively. Then, the binary hypothesis testing problem is to decide between two possible hypotheses based on the observed data $x_1, x_2, \ldots, x_N$ [1]. The likelihood ratio for the underlying hypothesis test problem is

$$L = \prod_{i=1}^{N} \frac{p_1(x_i)}{p_0(x_i)}. \tag{1}$$

The normalized log likelihood ratio can be represented by

$$\bar{l} = \frac{1}{N} \ln L = \frac{1}{N} \sum_{i=1}^{N} \ln \frac{p_1(x_i)}{p_0(x_i)}. \tag{2}$$

The normalized log likelihood ratio $\bar{l}$ is itself a random variable, and is an arithmetic average of N i.i.d. random variables $l_i = \ln \frac{p_1(x_i)}{p_0(x_i)}$. According to the strong law of large numbers [31] that for large N, as

$$\bar{l} \overset{P}{\longrightarrow} E[l_i], \tag{3}$$

where P denotes the convergence in probability, E is the expectation of random variables.

As $l_i, i = 1, \ldots, N$ are i.i.d. random variables and asymptotically

$$\begin{aligned} \bar{l} = E[l_i] &= \int q(x) \ln \frac{p_1(x)}{p_0(x)} dx = \int q(x) \ln \left[\frac{p_1(x)}{p_0(x)} \frac{q(x)}{q(x)} \right] dx \\ &= \int q(x) \left[\ln \frac{q(x)}{p_0(x)} - \ln \frac{q(x)}{p_1(x)} \right] dx \\ &= \int q(x) \ln \frac{q(x)}{p_0(x)} dx - \int q(x) \ln \frac{q(x)}{p_1(x)} dx. \end{aligned} \tag{4}$$

The quantity $\int q(x) \ln \frac{q(x)}{p(x)} dx$ is known as the Kullback–Leibler divergence (KLD) [32] from q to p, which can be denoted by

$$D(q\|p) = \int q(x) \ln \frac{q(x)}{p(x)} dx. \tag{5}$$

The Kullback–Leibler divergence is widely used to measure the similarity (distance) between two distributions and is also termed the relative entropy [33] in information theory. It should be noted that the Kullback–Leibler divergence is not a genuine metric of distance as it fails to be symmetric or satisfy the triangle inequality, which is justified by the Pinsker–Csiszar Inequality [34]. However, it has been proved that the KLD plays a central role in the theory of statistical inference [8,9].

The Kullback–Leibler divergence can be regarded as the expected log-likelihood ratio. The equivalence between the likelihood ratio and Kullback–Leibler divergence is indicated by Equation (4) as

$$\bar{l} = D(q\|p_0) - D(q\|p_1). \tag{6}$$

Therefore, for large N, the classical log likelihood ratio test $\ln L \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma$ can be performed by

$$D(q\|p_0) - D(q\|p_1) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \frac{1}{N}\gamma \triangleq \gamma'. \tag{7}$$

In particular, for $\gamma = 0$ where the prior probabilities of two hypotheses are equal, the decision becomes

$$D(q\|p_0) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} D(q\|p_1). \tag{8}$$

In such a case, the likelihood ratio test is equivalent to choosing the model that is "closer" to q in the sense of the Kullback–Leibler divergence. This is referred to as a minimum distance detector. For the general case given by Equation (7), the test can be regarded as a *generalized minimum distance detector* with a threshold γ'. Similar results can also be found in [26,33], where the analysis is performed on a probability simplex. In practical applications, typically in radar and sonar systems, the Neyman–Pearson criterion [35] is employed so that the detector is designed to maximize the probability of detection P_D for a given probability of false alarm P_F. In such a case, the probability of detection, i.e., the miss probability P_M of the detector, is especially concerned as a performance index of the detector. Generally speaking, the miss probability of the Neyman–Pearson detector decays exponentially with a fixed level as the sample size increases, and the *error exponent K* is a widely used measure as the rate of exponential decay [36], i.e.,

$$K \triangleq \lim_{n\to\infty} -\frac{1}{N}\log P_M \tag{9}$$

under the given false-alarm constraint.

Stein's lemma [33,37] shows that the best error exponent under a fixed false-alarm constraint is given by the Kullback–Leibler divergence $D(p_0\|p_1)$ from the distribution of hypothesis p_0 to p_1, i.e.,

$$K = D(p_0\|p_1) \tag{10}$$

and

$$P_M \doteq 2^{-NK}, \quad \text{given } P_F = \alpha, \tag{11}$$

where $\doteq$ denotes equality to first order in the exponent, i.e., $a_n \doteq b_n$ means $\lim_{n\to\infty}\frac{1}{n}\log\frac{a_n}{b_n} = 0$.

The error exponent provides a good performance index for Neyman–Pearson detectors in the large sample regime. Furthermore, discussions of the error exponent and its properties are addressed in [36]. In the following sections of this paper, the above notions will be demonstrated through discussions on the geometry of deterministic and random signal detection.

2.2. Geometric Interpretation of the Classical Likelihood Ratio Test

The previous analysis of the geometry of hypothesis testing is performed on a probability simplex [38], which is a generalization of the notion of a triangle or tetrahedron to arbitrary dimension. The vertex labels of a probability simplex correspond to a single probability set to 1, while edge labels correspond to a sum of two probabilities set to 1 and triangular faces correspond to a sum of three probabilities set to 1. The probability simplex provides a simple structure to analyze testing problems of probability distributions by mapping the type (empirical histogram) of data to a point on the simplex. However, it is not accomplished in dealing with continuous probability distributions as well as a complete structure to perform higher level of analysis. In a sense, information geometry provides a more powerful mathematical tool to deal with statistical problems in a geometrical manner, in which a parameterized family of probability distributions can be conveniently referred to as a statistical manifold with precise geometrical structure.

As mentioned earlier, the essence of a basic hypothesis testing problem can be regarded as discrimination between two identically distributed distributions with different parameters. From the viewpoint of information geometry, probability distributions of hypotheses can be conveniently regarded as two elements of a statistical manifold. Consider the parameterized family of probability distributions $S = \{ p(x|\theta) \}$, where x is a random variable and $\theta = (\theta_1, \ldots, \theta_n)$ is a parameter vector specifying the distribution. The family S is regarded as a statistical manifold with θ as its (possibly local) coordinate system [2].

Figure 1 illustrates the definition of a statistical manifold. For a given state of interest θ in the parameter space $\Theta \in \mathbb{R}^n$, the measurement x in the sample space $X \in \mathbb{R}^m$ is a realisation of to a probability distribution $p(x|\theta)$. Each probability distribution $p(x|\theta)$ is labelled by a point $s(\theta)$ on the manifold S. The parameterized family of probability distributions $S = \{ p(x|\theta) \}$ forms an n-dimensional statistical manifold, where θ plays the role of a coordinate system of S.
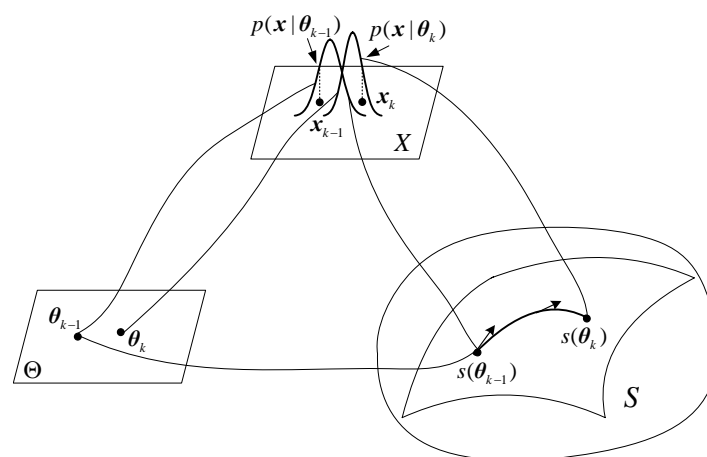
Figure 1. Definition of a statistical manifold.

For a sequence of observations $x = (x_1, x_2, \ldots, x_n)^T$, the binary hypothesis testing problem is that of deciding whether this sequence is originated from a source (hypothesis $\mathcal{H}_1$) with a probability distribution $p(x|\theta_1)$ or from a source (hypothesis $\mathcal{H}_0$) with a probability distribution $p(x|\theta_0)$, based on observations x. The probability distribution $p(x|\theta_i)$, $i = 1, 2$, associated with the hypothesis $\mathcal{H}_i$, is a element of a parametric family of probability density functions $S = \{ p(x|\theta), \theta \in \Theta \}$, where $\Theta \in \mathbb{R}^n$ is the parameter set. This family of distributions $S = \{ p(x|\theta) \}$ makes up of a statistical manifold, where the parameter θ is the coordinate of the measurement distribution on the statistical manifold. Distributions of the two model (hypotheses) $p(x|\theta_0)$ and $p(x|\theta_1)$ can be regarded as two "datum marks" on the manifold, while the distribution $p(x|\theta)$ summarized from observations x is an estimated distribution on the manifold. Then, the above hypothesis testing problem becomes a discrimination problem where the decision is made by comparing distances from the signal distribution estimates to two hypotheses in the sense of the Kullback–Leibler divergence, i.e., selecting the model that is "closer" to the signal distribution estimates. The geometric interpretation of classical likelihood ratio test from the viewpoint of information geometry can be illustrated by Figure 2. The above analysis can be extended to the complex parameter θ and/or complex measurement x.
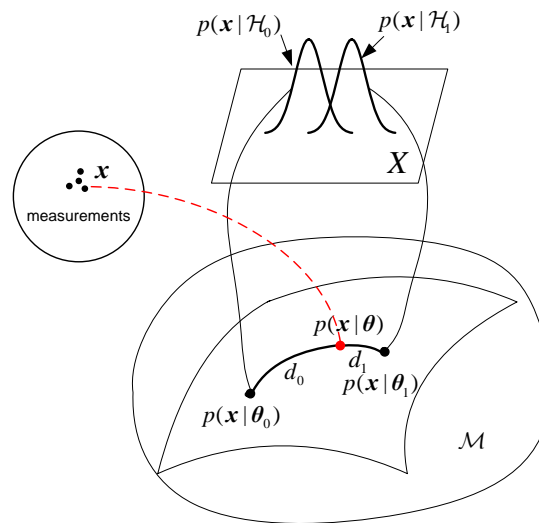
Figure 2. Geometry of the classical likelihood ratio test.

3. Geometry of Deterministic Signal Detection

3.1. Signal Model and Likelihood Ratio Test

The detection of a known deterministic signal in Gaussian noise is a basic detection problem encountered in statistical signal processing. Meanwhile, in radar applications, it usually involves the complex data. Consider the detection of a known deterministic signal in complex Gaussian noise. The hypothesis testing problem is [1]

$$\begin{cases} \mathcal{H}_0 : x(n) = w(n), & n = 0, 1, \ldots, N-1, \\ \mathcal{H}_1 : x(n) = s(n) + w(n), & n = 0, 1, \ldots, N-1, \end{cases} \tag{12}$$

where $s(n)$ is a known complex signal and $w(n)$ is a complex correlated Gaussian noise with zero mean. The above discrete quantities can be written as the vector forms,

$$\boldsymbol{x} = \big[x(0), x(1), \ldots, x(N-1)\big]^T, \quad \boldsymbol{s} = \big[s(0), s(1), \ldots, s(N-1)\big]^T, \quad \boldsymbol{w} = \big[w(0), w(1), \ldots, w(N-1)\big]^T.$$

Then, under the two hypotheses, the distributions of measurement error can be represented by

$$\begin{cases} \mathcal{H}_0 : \boldsymbol{x} \sim \mathcal{CN}\big(\boldsymbol{0}, \Sigma\big), \\ \mathcal{H}_1 : \boldsymbol{x} \sim \mathcal{CN}\big(\boldsymbol{s}, \Sigma\big), \end{cases} \tag{13}$$

where Σ is the covariance of noise vector $\boldsymbol{w}$ and $\mathcal{CN}$ denotes the complex Gaussian distribution.

Consequently, the underlying detection problem for a known deterministic signal can be regarded as discrimination between two distributions with the same covariance but different means. According to the Neyman–Pearson criterion, the likelihood ratio test reduces to deciding $\mathcal{H}_1$ if

$$T(\boldsymbol{x}) = \mathrm{Re}\left(\boldsymbol{s}^H \Sigma^{-1} \boldsymbol{x}\right) > \gamma, \tag{14}$$

where $T(\boldsymbol{x})$ is the test statistic and H denotes the complex conjugate transpose. The detector described by Equation (14) is referred to as a generalized replica–correlator or generalized matched filter [1].

3.2. Geometry of Deterministic Signal Detection

As the vectorial measurement x is from either of two hypotheses described in Equation (13), its covariance is Σ for either hypothesis. However, the mean of the measurements x is unknown and should be estimated in order to decide which hypothesis is true. The measurement distribution can be summarized (estimated) by

$$p(x|\boldsymbol{\theta}) \sim \mathcal{CN}(x, \Sigma), \tag{15}$$

where the mean is estimated by the mean of measurements, e.g., x.

The two datum marks on the manifold can be respectively denoted by

$$p(x|\boldsymbol{\theta}_1) \sim \mathcal{CN}(s, \Sigma), \;\; p(x|\boldsymbol{\theta}_0) \sim \mathcal{CN}(0, \Sigma). \tag{16}$$

For two multivariate Gaussian distributions $p_0 = \mathcal{CN}_0(\boldsymbol{\mu}_0, \Sigma_0)$ and $p_1 = \mathcal{CN}_1(\boldsymbol{\mu}_1, \Sigma_1)$, the Kullback–Leibler divergence from p_0 to p_1 can be calculated via the closed form as [32]

$$D(p_0\|p_1) = \frac{1}{2}\left[\ln\left(\frac{\det\Sigma_1}{\det\Sigma_0}\right) + \mathrm{tr}\left(\Sigma_1^{-1}\Sigma_0\right) + (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^H\Sigma_1^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) - N\right], \tag{17}$$

where "tr" and "det" denote the trace and determinant of a matrix, respectively.

According to Equation (17), the KLDs from the measurement distribution $p(x|\boldsymbol{\theta})$ to hypotheses $p(x|\boldsymbol{\theta}_1)$ and $p(x|\boldsymbol{\theta}_0)$ are given by

$$D(p\|p_1) = \frac{1}{2}(x - s)^H\Sigma^{-1}(x - s), \tag{18}$$

$$D(p\|p_0) = \frac{1}{2}x^H\Sigma^{-1}x. \tag{19}$$

Thus, the distance difference is

$$D(p\|p_0) - D(p\|p_1) = \frac{1}{2}\left[x^H\Sigma^{-1}x - (x - s)^H\Sigma^{-1}(x - s)\right] = \mathrm{Re}\left(s^H\Sigma^{-1}x\right) - \frac{1}{2}s^H\Sigma^{-1}s. \tag{20}$$

By incorporating the non-data-dependent term into the threshold we decide $\mathcal{H}_1$ if

$$\mathcal{D}(x) = \mathrm{Re}\left(s^H\Sigma^{-1}x\right) > \gamma, \tag{21}$$

where $\mathcal{D}(x)$ is the test statistic of the generalized minimum distance detector.

By comparing the results given by Equation (14) and Equation (21), the detector defined by the generalized minimum distance detector coincides with the result given by the likelihood ratio test. However, the former illustrates a clear geometric interpretation of the detection problem in terms of its discrimination nature.

For above detectors, the test statistic is a complex Gaussian random variable under either hypothesis, being a linear transformation of x. Its statistics can be represented by

$$T(x) \sim \begin{cases} \mathcal{N}\left(0, s^H\Sigma^{-1}s/2\right) & \text{under } \mathcal{H}_0, \\ \mathcal{N}\left(s^H\Sigma^{-1}s, s^H\Sigma^{-1}s/2\right) & \text{under } \mathcal{H}_1, \end{cases} \tag{22}$$

where we have noted that $s^H\Sigma^{-1}s$ is real.

For a given probability of false alarm P_F, the threshold is given by

$$\gamma = \sqrt{s^H\Sigma^{-1}s/2}\, Q^{-1}(P_F), \tag{23}$$

where $Q(\cdot)$ is the right tail cumulative distribution function of the standard Gaussian distribution.

The corresponding probability of detection is given by [1]

$$P_D = Q\left[Q^{-1}(P_F) - \sqrt{d^2}\right],\tag{24}$$

where the deflection coefficient is defined by $d^2 = 2s^H\Sigma^{-1}s$, which can be regarded as the signal-to-noise ratio (SNR) of the underlying detection problem. Equation (24) indicates that the probability of detection increases monotonically with $s^H\Sigma^{-1}s$.

From the geometrical viewpoint, as shown in Equations (10) and (11), the performance of the Neyman–Pearson detector under a given false-alarm constraint can be measured by the error exponent, which is associated with the Kullback–Leibler divergence between two hypothesis distributions. In fact, in the deterministic signal detection example, the deflection coefficient d^2 is equivalent to the KLD $D(p_0\|p_1)$ from $p(x|\theta_0)$ to $p(x|\theta_1)$, which denotes distributions of the two hypotheses, as

$$D(p_0\|p_1) = \frac{1}{2}s^H\Sigma^{-1}s.\tag{25}$$

The result in Equation (25) indicates that the detection performance depends on the distance between two hypothesis distributions, which gives a clear geometric meaning of the detection performance by viewing the detection process as a discrimination between two hypotheses.

4. Geometry of Random Signal Detection

In the last section, we discussed the deterministic signal detection problem as well as its geometry. It is able to detect signals in the presence of noise by detecting the change in the mean of a test statistic. This is because the presence of a deterministic signal alters the mean of the received data. However, in some cases, the signal is more appropriately to be modeled as a random process with a known covariance structure. In this section, we consider the detection from the random signal model as well as the geometry of random signal detection.

4.1. Signal Model and Likelihood Ratio Test

The signal model of a random signal with known statistics is similar to the deterministic case given by Equation (12). The only difference is that the signal $s(n)$ is assumed as a zero-mean complex Gaussian random process, and under the two hypotheses, the distributions of measurement error can be represented by

$$\begin{cases} \mathcal{H}_0 : x \sim \mathcal{CN}(\mathbf{0}, \Sigma_w), \\ \mathcal{H}_1 : x \sim \mathcal{CN}(\mathbf{0}, \Sigma_s + \Sigma_w), \end{cases}\tag{26}$$

where Σ_s and Σ_w denote the covariance matrices of the random signal and noise, respectively.

Consequently, the underlying detection problem for a random signal with known statistics in the presence of noise can be regarded as discrimination between two distributions with the same mean but different covariance matrices. The Neyman–Pearson detector decides $\mathcal{H}_1$ if the likelihood ratio exceeds a threshold or if

$$T(x) = x^H\Sigma_w^{-1}\Sigma_s(\Sigma_s + \Sigma_w)^{-1}x > \gamma,\tag{27}$$

where $T(x)$ is the test statistic.

4.2. Geometry of Random Signal Detection

As measurements x are from either of the two hypotheses described in Equation (26), its mean is zero for either hypothesis. However, the covariance of x is uncertain and should be estimated in order

to decide which hypothesis is true. Assume that both the signal and noise are zero mean wide sense stationary (WSS) random process; then, the covariance of data $\boldsymbol{x}$ is defined by

$$
\boldsymbol{R} \triangleq E\left[(\boldsymbol{x} - \boldsymbol{\mu}_x)(\boldsymbol{x} - \boldsymbol{\mu}_x)^H\right] = \begin{bmatrix} r_0 & \bar{r}_1 & \cdots & \bar{r}_{N-1} \\ r_1 & r_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \bar{r}_1 \\ r_{N-1} & \cdots & r_1 & r_0 \end{bmatrix}, \tag{28}
$$

where $r_k = E[x_n \bar{x}_{n+k}]$ is called the correlation coefficient and $\bar{x}$ denotes the complex conjugate of x. $\boldsymbol{R}$ is a Toeplitz Hermitian Positive Definite matrix with $\boldsymbol{R}^H = \boldsymbol{R}$. According to the ergodicity of a WSS random process, the correlation coefficients of the signal can be calculated by averaging over time instead of its statistical expectation, as

$$
\hat{r}_k = \frac{1}{N} \sum_{n=0}^{N-1-|k|} x(n)\bar{x}(n+k), \quad |k| \leqslant N - 1. \tag{29}
$$

Thus, the measurement error distribution can be summarized (estimated) by

$$
p(\boldsymbol{x}|\boldsymbol{\theta}) \sim \mathcal{CN}\left(\boldsymbol{0}, \boldsymbol{R}\right), \tag{30}
$$

where the covariance $\boldsymbol{R}$ is estimated from measurements $\boldsymbol{x}$.

Then, the two datum marks on the manifold can be, respectively, denoted by

$$
p(\boldsymbol{x}|\boldsymbol{\theta}_1) \sim \mathcal{CN}\left(\boldsymbol{0}, \Sigma_s + \Sigma_w\right), \quad p(\boldsymbol{x}|\boldsymbol{\theta}_0) \sim \mathcal{CN}\left(\boldsymbol{0}, \Sigma_w\right). \tag{31}
$$

Using Equation (17), the KLDs from the measurement distribution $p(\boldsymbol{x}|\boldsymbol{\theta})$ to hypotheses $p(\boldsymbol{x}|\boldsymbol{\theta}_1)$ and $p(\boldsymbol{x}|\boldsymbol{\theta}_0)$ are given by

$$
D(p\|p_1) = \frac{1}{2}\left[\mathrm{tr}\left((\Sigma_s + \Sigma_w)^{-1}\boldsymbol{R} - \boldsymbol{I}\right) - \ln\det\left((\Sigma_s + \Sigma_w)^{-1}\boldsymbol{R}\right)\right], \tag{32}
$$

$$
D(p\|p_0) = \frac{1}{2}\left[\mathrm{tr}\left(\Sigma_w^{-1}\boldsymbol{R} - \boldsymbol{I}\right) - \ln\det\left(\Sigma_w^{-1}\boldsymbol{R}\right)\right]. \tag{33}
$$

Consequently, the distance difference is

$$
\begin{aligned}
D(p\|p_0) - D(p\|p_1) &= \frac{1}{2}\left[\mathrm{tr}\left(\left(\Sigma_w^{-1} - (\Sigma_s + \Sigma_w)^{-1}\right)\boldsymbol{R}\right) - \ln\det\left((\Sigma_s + \Sigma_w)\Sigma_w^{-1}\right)\right] \\
&= \frac{1}{2}\mathrm{tr}\left[\Sigma_w^{-1}\Sigma_s(\Sigma_s + \Sigma_w)^{-1}\boldsymbol{R}\right] - \frac{1}{2}\ln\det\left(\boldsymbol{I} + \Sigma_s\Sigma_w^{-1}\right).
\end{aligned} \tag{34}
$$

By incorporating the non-data-dependent term into the threshold, we decide $\mathcal{H}_1$ if

$$
\mathcal{D}(\boldsymbol{x}) = \mathrm{tr}\left[\Sigma_w^{-1}\Sigma_s(\Sigma_s + \Sigma_w)^{-1}\boldsymbol{R}\right] > \gamma, \tag{35}
$$

where $\mathcal{D}(\boldsymbol{x})$ is the test statistic of the generalized minimum distance detector.

As the test statistic in Equation (27) is a quadratic form in the data and thus will not be a Gaussian random variable, the detection performance of the detector is quite often difficult to determine analytically [1], especially when the signal is a correlated random process with arbitrary covariance matrix. While the detector given in Equation (35) is associated with an estimated covariance $\boldsymbol{R}$ based on measurements, analysis of its detection performance is also non-trivial. Thus, here, the performance comparison of two detectors is performed via Monte Carlo simulations. In addition, 1,000,000 Monte Carlo runs are performed and the performances of two detectors are illustrated in Figure 3. In the simulation, the number of measurement samples is $N = 100$ and the probabilities

of false alarm P_F are assumed $P_F = 10^{-4}, 10^{-5}$, respectively. Figure 3 illustrates that the detection performance of two detectors coincide with each other, i.e., the likelihood ratio-based detector is equivalent to the geometry-based detector.
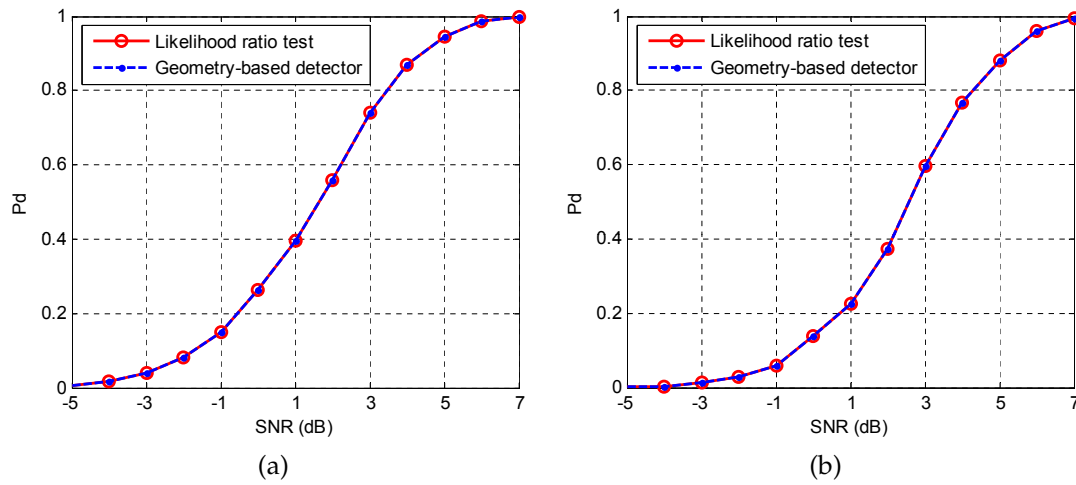
Figure 3. Performance comparison of the likelihood ratio-based detector and geometry-based detector. (**a**) $P_F = 10^{-4}$; and (**b**) $P_F = 10^{-5}$.

The error exponent of the above Neyman–Pearson detector under the fixed false-alarm constraint is given by the Kullback–Leibler divergence from p_0 to p_1, as

$$D(p_0 \| p_1) = \frac{1}{2} \left[\text{tr}\big((\Sigma_s + \Sigma_w)^{-1}\Sigma_w\big) - \ln \det\big((\Sigma_s + \Sigma_w)^{-1}\Sigma_w\big) - N \right]. \tag{36}$$

As a special case, when the signal and noise are both white random processes with variances σ_s^2 and σ_w^2, the error exponent is simplified by

$$\begin{aligned} K &= \frac{N}{2} \left(\frac{\sigma_w^2}{\sigma_s^2 + \sigma_w^2} + \ln \frac{\sigma_s^2 + \sigma_w^2}{\sigma_w^2} - 1 \right) \\ &= \frac{N}{2} \left[\frac{1}{\text{SNR} + 1} + \ln(\text{SNR} + 1) - 1 \right], \end{aligned} \tag{37}$$

where $\text{SNR} = \frac{\sigma_s^2}{\sigma_w^2}$ denotes the signal-to-noise ratio.

As

$$\frac{dK}{d\text{SNR}} = \frac{N \cdot \text{SNR}}{2(\text{SNR} + 1)^2} > 0, \tag{38}$$

the error exponent is monotonically increasing as SNR increases. Moreover, as shown in Figure 4, the error exponent K increases linearly with respect to SNR at high SNR. The result indicates that the detection performance improves as the divergence between two hypotheses increases, which gives a clear geometric interpretation of the detection performance of the Neyman–Pearson detector.
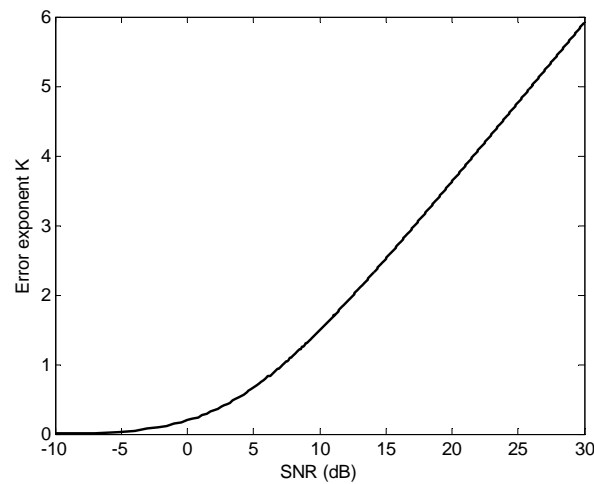
Figure 4. Error exponent versus signal-to-noise ratio ($N = 1$).

5. Application to Radar Target Detection

In this section, the application of geometry based detector to radar constant false alarm rate (CFAR) detection with pulse Doppler radar is presented. The performance between the geometry based CFAR detector and the classical cell-averaging CFAR (CA-CFAR) detector is compared. As illustrated in Figure 5, in the classical CA-CFAR detector for pulse Doppler radar, the data z_i in each range resolution cell is obtained from a square-law detector of Doppler filter banks, where the Doppler power spectral density of the sample data x is estimated by fast Fourier transform (FFT). The CA-CFAR detector performs by comparing the data z_D in the test cell to an adaptive detection threshold γ such that a CFAR is maintained. The detection threshold γ is determined by estimating the background clutter power of the test cell and multiplying a scaling factor T based on the desired probability of false alarm P_F. The power of the background clutter is estimated by the arithmetical mean of the data in the background range cells, which is given by

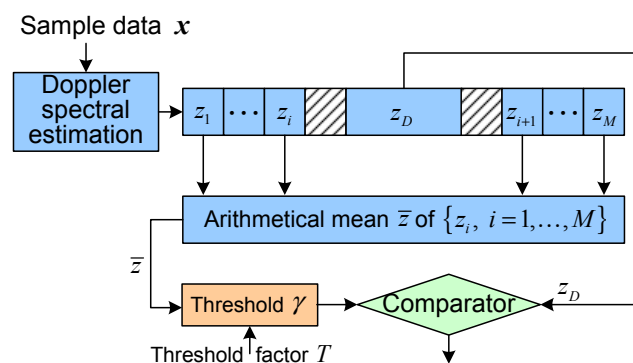$$\bar{z} = \frac{1}{M} \sum_{i=1}^{M} z_i. \tag{39}$$

Figure 5. Classical cell-averaging constant false alarm rate (CA-CFAR) detector based on Doppler spectral estimation.

Due to the poor Doppler resolution as well as the energy spread of the Doppler filter banks, the classical CA-CFAR detector based on Doppler spectral estimation suffers from a severe performance degradation. Barbaresco et al. [28] developed a matrix CFAR detector based on Cartan's geometry of

the symmetric positive-definite (SPD) matrices space, which improves the detection performance of the classical CA-CFAR detector based on Doppler filter banks. As illustrated in Figure 6, the data R_i in each resolution cell is a covariance matrix estimated by the sample data x according to Equation (28). Then, calculate the distance between the covariance matrix R_D of the cell under test and the mean matrix $\overline{R}$ of the reference cells around the cell under test. Finally, the detection is made by comparing the distance between R_D and $\overline{R}$ with an adaptive detection threshold γ, which is related to the clutter power level as well as a multiplier corresponding to the desired probability of false alarm P_F. In [28], the Riemannian distance measure is used when calculating the distance as well as the mean matrix, as the geometry of the manifold of SPD matrices is considered.
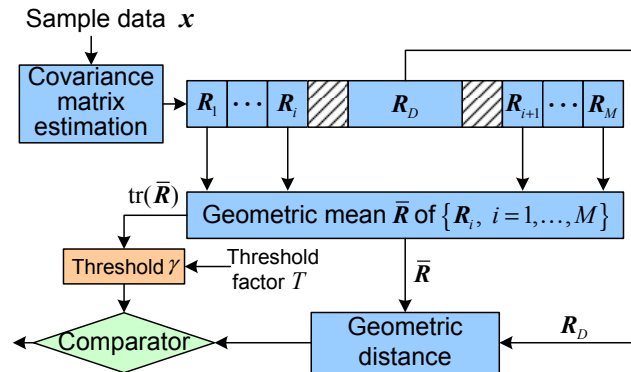

Figure 6. Geometry-based matrix CFAR detector.

The Riemannian distance between the covariance matrix of the cell under test and the mean matrix of reference cells is calculated by

$$d^2 \left(R_D, \overline{R} \right) = \| \ln \left(R_D^{-1/2} \overline{R} R_D^{-1/2} \right) \|^2 = \sum_{i=1}^{N} \ln^2 \left(\lambda_k \right), \tag{40}$$

where λ_k is the kth eigenvalue of $R_D^{-1/2} \overline{R} R_D^{-1/2}$.

The Riemannian mean of a set of SPD matrices $\{ R_1, R_2, \ldots, R_M \}$ is calculated via the minimum sum of distance between R and R_i as [39]

$$\overline{R} = \arg \min_{R} \sum_{i=1}^{M} d^2 \left(R, R_i \right), \tag{41}$$

where $d(\cdot, \cdot)$ represents the Riemannian distance between two matrices.

One approach to solve Equation (41) is to use an iterative gradient algorithm that is based on the Jacobi field and the exponential map. In summary, the Karcher barycenter based Riemannian mean is given by the following iteration [40]

$$\overline{R}_{k+1} = \overline{R}_k^{1/2} \exp \left(\eta_k \sum_{i=1}^{M} \ln \left(\overline{R}_k^{-1/2} R_i \overline{R}_k^{-1/2} \right) \right) \overline{R}_k^{1/2}, \tag{42}$$

where η_k denotes the step-length of iterations.

Then, the matrix CFAR detector based on Riemannian distance can be performed as

$$d \left(R_D, \overline{R} \right) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma. \tag{43}$$

The matrix CFAR detector proposed in [28] improves the performance of the classical CA-CFAR detector. However, as discussed in Section 2, the likelihood ratio test is equivalent to the generalized minimum distance detector in terms of the Kullback–Leibler divergence, which indicates that the Kullback–Leibler divergence is suitable for detection applications. Thus, it is necessary to compare the performance of matrix CFAR detector in terms of the Riemannian distance measure and Kullback–Leibler divergence. The Kullback–Leibler divergence between two matrix is given by Equation (17), where the mean of a set of SPD matrices $\{R_1, R_2, \ldots, R_M\}$ is calculated by

$$\overline{R}_{KL} = \arg\min_{R} \sum_{i=1}^{M} D^2(R\|R_i) = \left(\frac{1}{M} \sum_{i=1}^{M} R_i^{-1} \right)^{-1}. \tag{44}$$
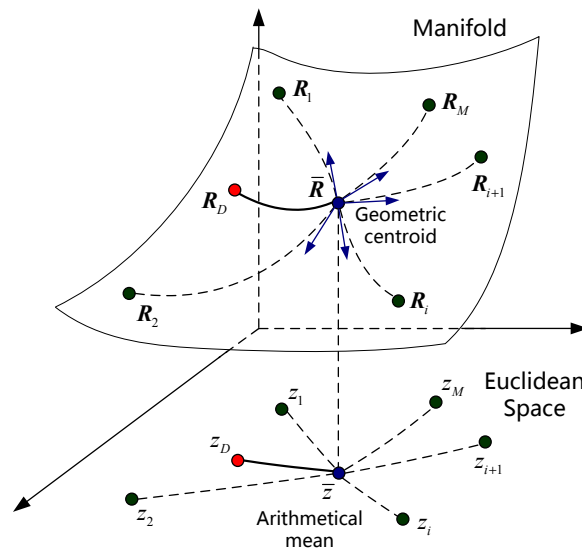
The performance of: (i) the classical CA-CFAR detector based on Doppler spectral estimation, (ii) the Riemannian distance-based matrix CFAR detector, and (iii) the Kullback–Leibler divergence-based matrix CFAR detector are compared via Monte Carlo simulations. The data samples are from $N = 7$ received pulses. A total of 50 range cells are considered, while $M = 14$ range cells are used for averaging. One moving target with Doppler frequency $f_d = 2.65$ Hz is located at the 19th range cell. As the Gaussian assumption is no longer met in many situations of practical interest, such as the high-resolution radar and sea clutter, a compound-Gaussian model is more appropriate to describe the non-Gaussian radar clutter. The compound-Gaussian clutter can be written as the product of two mutually independent random processes, where the fast fluctuating 'speckle' component is a zero-mean complex Gaussian process and the comparatively slow varying 'texture' component is a nonnegative real random process that describes the underlying mean power level of the resultant clutter process [41,42]. The most common clutter models, such as the Weibull and K-distribution, are compatible with the compound-Gaussian model. In the simulation, the clutter is assumed as a Weibull distribution with scale parameter $\alpha = 1$ and shape parameter $\beta = 3$.

The performance of the three detectors are presented in Figure 7. In addition, 100,000 Monte Carlo runs are performed in the simulation. The probabilities of false alarm P_F are assumed to be 10^{-3}, 10^{-4}, respectively. For the underlying detectors, the detection thresholds are not only related to the background clutter power but are also associated with the threshold multiplier, and a closed form expression of the threshold multiplier does not usually exist. Moreover, for Weibull distributed clutter, the threshold multiplier of the underlying detector depends on the desired probability of false alarm P_F, the number of bins averaged M, as well as the shape parameter β [43]. Consequently, it is difficult to determine the detection threshold analytically. In the simulation, the detection threshold is obtained via Monte Carlo runs. The test statistic in the absence of target is calculated via 100,000 Monte Carlo runs. As the effects of the clutter power, the number of bins averaged M and the shape parameter β are included in the test statistic; then, the detection threshold is determined according to the given probability of false alarm P_F. The result illustrates that both the two geometry based detectors outperform the classical CA-CFAR detector. Moreover, the performance of the Kullback–Leibler divergence-based matrix CFAR detector is better than the Riemannian distance-based matrix CFAR detector with 4–6 dB signal-to-clutter ratio (SCR) improvement.

Figure 8 illustrates a geometrical interpretation of the classical and geometry-based detection procedures. The classical CA-CFAR detector performs in the Euclidean space, where the Euclidean distance measure is used to define the divergence between two elements. However, the geometry based detector performs in the statistical manifold, in which the intrinsic divergence of the covariance matrices can be properly exploited and utilized, thus improving the detection performance.
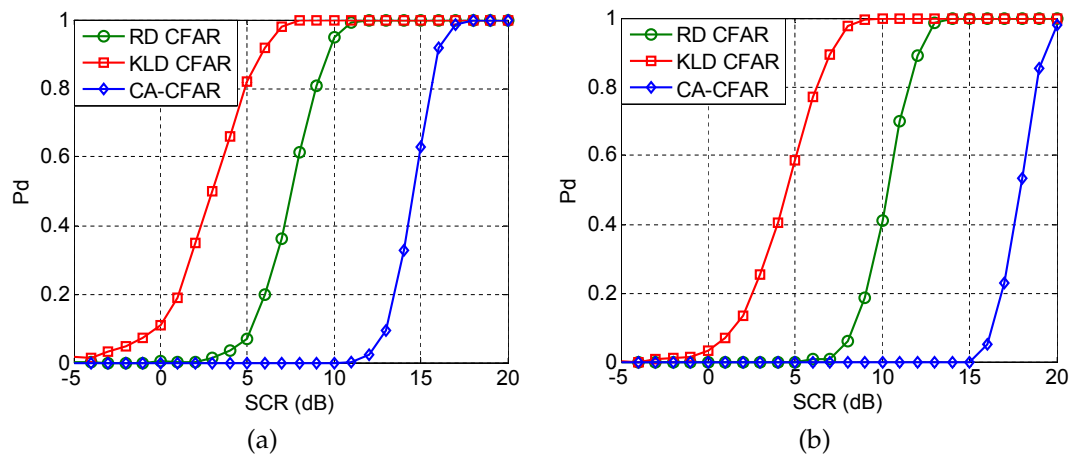
Figure 7. Performance comparison of the classical CA-CFAR detector, Riemannian distance based CFAR detector and Kullback–Leibler divergence based CFAR detector. (**a**) $P_F = 10^{-3}$; and (**b**) $P_F = 10^{-4}$.


Figure 8. Geometrical interpretation of the classical CA-CFAR detector and geometry-based detector.

Different distance measures adopted in detectors may result in different performance of detection. The Riemannian distance and Kullback–Leibler divergence are two measures used to evaluate the divergence between the data under test and clutter. The quality of the estimated Riemannian mean and Kullback–Leibler mean of SPD matrices are analyzed as follows. As the accuracy of the estimated mean matrix is a function of the number of samples (SPD matrices) used for averaging [40], when the number of samples M tends to infinity, the true mean matrix is obtained. In the simulation, 1000 samples are used to approximately estimate the true mean matrix. Then, for the case of a few number of samples available in detection, the distances between the estimated mean of few samples and the true mean matrix in terms of the Riemannian distance as well as Kullback–Leibler divergence are calculated via 300 Monte Carlo runs. The average of the 300 distances is presented as a function of the number of samples in Figure 9. It represents the distance (divergence) between the estimated mean matrix and the true mean matrix.

It can be observed that both of the Riemannian mean and Kullback–Leibler mean asymptotically approach the true mean matrix as the number of samples for averaging increases. It seems that the Kullback–Leibler mean is of smaller divergence to the true mean matrix than Riemannian mean in quantity. However, as Riemannian distance and Kullback–Leibler divergence are two different

distance measures, the values of the two measures are incomparable. As discussed in the geometric interpretation of classical likelihood ratio test presented in Section 2, the likelihood ratio test is equivalent to the generalized minimum distance detector in the sense of Kullback–Leibler divergence, rather than Riemannian distance, which provides evidence that the KLD plays a central role in the theory of statistical inference. Therefore, Kullback–Leibler divergence is probably more appropriate to the case of hypothesis testing than Riemannian distance measure, which brings a performance improvement in detection.
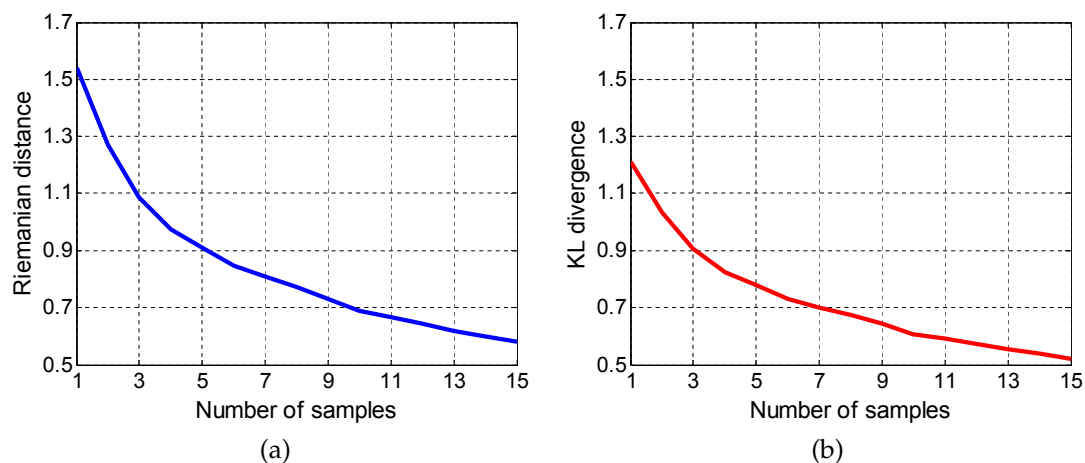
Figure 9. Estimation accuracy of mean matrix. (**a**) Riemannian mean; and (**b**) Kullback–Leibler mean.

6. Conclusions

In this paper, the problem of hypothesis testing in the Neyman–Pearson formulation is considered from a geometric viewpoint. In particular, we have described a concise geometric interpretation of deterministic and random signal detection in the philosophy of information geometry. In such a framework, both hypotheses and detectors can be regarded as geometrical objects on the statistical manifold of a parameterized family of probability distributions. Both the detector and detection performance are elucidated geometrically in terms of the Kullback–Leibler divergence. Compared to the likelihood ratio test, the geometric interpretation provides a consistent but more comprehensive means to understanding signal detection problems. Such methodology may play an important role in the development and understanding of the theoretical aspects of algorithms in radar signal processing.

Author Contributions: Yongqiang Cheng put forward the original ideas and performed the research. Xiaoqiang Hua and Hongqiang Wang conceived and designed the simulations of the matrix CFAR detector. Yuliang Qin generalized the discussion to the complex data case. Xiang Li reviewed the paper and provided useful comments. All authors have read and approved the final manuscript.

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

References

1. Kay, S.M. *Fundamentals of Statistical Signal Processing: Detection Theory*, 1st ed.; Prentice Hall PTR: Upper Saddle River, NJ, USA, 1993; Volume 2.
2. Amari, S. Information geometry on hierarchy of probability distributions. *IEEE Trans. Inf. Theory* **2001**, *47*, 1701–1711.
3. Amari, S. Information geometry of statistical inference—An overview. In Proceedings of the IEEE Information Theory Workshop, Bangalore, India, 20–25 October 2002.

4. Rao, C.R. Information and accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.* **1945**, *37*, 81–91.

5. Amari, S.; Nagaoka, H. *Methods of Information Geometry*; Kobayashi, S., Takesaki, M., Eds.; Translations of Mathematical Monographs vol. 191; American Mathematical Society: Providence, RI, USA, 2000.

6. Chentsov, N.N. *Statistical Decision Rules and Optimal Inference*; Leifman, L.J., Ed.; Translations of Mathematical Monographs vol. 53; American Mathematical Society: Providence, RI, USA, 1982.

7. Efron, B. Defining the curvature of a statistical problem (with applications to second order efficiency). *Ann. Stat.* **1975**, *3*, 1189–1242.

8. Efron, B. The geometry of exponential families. *Ann. Stat.* **1978**, *6*, 362–376.

9. Amari, S. Differential geometry of curved exponential families-curvatures and information loss. *Ann. Stat.* **1982**, *10*, 357–385.

10. Kass, R.E.; Vos, P.W. *Geometrical Foundations of Asymptotic Inference*; Wiley: New York, NY, USA, 1997.

11. Amari, S.; Kawanabe, M. Information geometry of estimating functions in semi parametric statistical models. *Bernoulli* **1997**, *3*, 29–54.

12. Amari, S.; Kurata, K.; Nagaoka, H. Information geometry of Boltzmann machines. *IEEE Trans. Neural Netw.* **1992**, *3*, 260–271.

13. Amari, S. Natural gradient works efficiently in learning. *Neural Comput.* **1998**, *10*, 251–276.

14. Amari, S. Information geometry of the EM and em algorithms for neural networks. *Neural Netw.* **1995**, *8*, 1379–1408.

15. Amari, S. Fisher information under restriction of Shannon information. *Ann. Inst. Stat. Math.* **1989**, *41*, 623–648.

16. Campbell, L.L. The relation between information theory and the differential geometry approach to statistics. *Inf. Sci.* **1985**, *35*, 199–210.

17. Amari, S. Differential geometry of a parametric family of invertible linear systems—Riemannian metric, dual affine connections and divergence. *Math. Syst. Theory* **1987**, *20*, 53–82.

18. Ohara, A.; Suda, N.; Amari, S. Dualistic differential geometry of positive definite matrices and its applications to related problems. *Linear Algebra Appl.* **1996**, *247*, 31–53.

19. Ohara, A. Information geometric analysis of an interior point method for semidefinite programming. In Proceedings of the Geometry in Present Day Science, Aarhus, Denmark, 16–18 January 1998; Barndorff-Nielsen, O.E., Vedel Jensen, E.B., Eds.; World Scientific: Singapore, Singapore, 1999; pp. 49–74.

20. Amari, S.; Shimokawa, H. *Recent Developments of Mean Field Approximation*; Opper, M., Saad, D., Eds.; MIT Press: Cambridge, MA, USA, 2000.

21. Richardson, T. The geometry of turbo-decoding dynamics. *IEEE Trans. Inf. Theory* **2000**, *46*, 9–23.

22. Li, Q.; Georghiades, C.N. On a geometric view of multiuser detection for synchronous DS/CDMA channels. *IEEE Trans. Inf. Theory* **2000**, *46*, 2723–2731.

23. Smith, S.T. Covariance, subspace, and intrinsic Cramér–Rao bounds. *IEEE Trans. Signal Process.* **2005**, *53*, 1610–1630.

24. Srivastava, A. A Bayesian approach to geometric subspace estimation. *IEEE Trans. Signal Process.* **2000**, *48*, 1390–1400.

25. Srivastava, A.; Klassen, E. Bayesian and geometric subspace tracking. *Adv. Appl. Probab.* **2004**, *36*, 43–56.

26. Westover, M.B. Asymptotic geometry of multiple hypothesis testing. *IEEE Trans. Inf. Theory* **2008**, *54*, 3327–3329.

27. Varshney, K.R. Bayes risk error is a Bregman divergence. *IEEE Trans. Signal Process.* **2011**, *59*, 4470–4472.

28. 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.

29. 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.

30. 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.

31. Etemadi, N. An elementary proof of the strong law of large numbers. *Probab. Theory Relat. Fields* **1981**, *55*, 119–122.

32. Kullback, S. *Information Theory and Statistics*; Dover Publications: New York, NY, USA, 1968.

33. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; Wiley: New York, NY, USA, 1991.

34. Pinsker, M.S. *Information and Information Stability of Random Variables and Processes*; Holden-Day: San Francisco, CA, USA, 1964.

35. Lehmann, F.L. *Testing Statistical Hypotheses*; Wiley: New York, NY, USA, 1959.

36. Sung, Y.; Tong, L.; Poor, H.V. Neyman–Pearson detection of Gauss-Markov signals in noise: Closed-form error exponent and properties. *IEEE Trans. Inf. Theory* **2006**, *52*, 1354–1365.

37. Chamberland, J.; Veeravalli, V.V. Decentralized detection in sensor networks. *IEEE Trans. Signal Process.* **2003**, *51*, 407–416.

38. Sullivant, S. Statistical Models Are Algebraic Varieties. Available online: https://www.researchgate.net/profile/Seth_Sullivant/publication/254069909_Statistical_Models_are_Algebraic_Varieties/links/552bdb0d0cf2e089a3aa889a.pdf?origin=publication_detail (accessed on 13 April 2015).

39. Moakher, M. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.* **2005**, *26*, 735–747.

40. Balaji, B.; Barbaresco, F.; Decurninge, A. Information geometry and estimation of Toeplitz covariance matrices. In Proceedings of the 2014 International Radar Conference, Lille, France, 13–17 October 2014.

41. Conte, E.; Lops, M.; Ricci, G. Asymptotically optimum radar detection in compound-Gaussian clutter. *IEEE Trans. Aerosp. Electron. Syst.* **1995**, *31*, 617–625.

42. Roy, L.P.; Kumar, R.V.R. A GLRT detector in partially correlated texture based compound-Gaussian clutter. In Proceedings of the 2010 National Conference on Communications (NCC), Chennai, India, 29–31 January 2010.

43. Vela, G.M.; Portas, J.A.B.; Corredera, J.R.C. Probability of false alarm of CA-CFAR detector in Weibull clutter. *Electron. Lett.* **1998**, *34*, 806–807.