entropy

MDPI

Article

Fisher–Rao Distance for Finite-Energy Signal Manifolds: Geometric Foundations and Numerical Analysis

Franck Florin

Thales, 92190 Meudon, France; franck.florin@fr.thalesgroup.com

Abstract

This paper introduces a geometric framework for analyzing finite-energy signals observed with additive noise by representing them as points on statistical manifolds equipped with the Fisher–Rao metric. Each signal is associated with a parameter vector $\boldsymbol{\theta}$, which defines a unique probability distribution $p(\boldsymbol{x}|\boldsymbol{\theta})$ on a statistical manifold. We propose a unified approach based on the normal multivariate model to describe a raw signal mixed with additive stationary noise. In the approach considered, the background noise is typically assumed to be stationary, whereas the unknown signal is regarded as deterministic. Leveraging tools from information geometry, we compute geodesic equations for the statistical manifolds. We re-derive known results regarding the multivariate normal models and extend them to the signal processing domain. We show that in some cases, the geodesic equations can be solved to obtain a closed-form expression of the Fisher–Rao distance. This expression corresponds to a minimum bound when the sub-manifold is not geodesic, revealing a fundamental geometric constraint in signal parameter estimation. We introduce the spectral distance function, which characterizes the influence of each spectral component of the signals on the Fisher–Rao distance. Our findings provide theoretical insights for signal clustering and machine learning applications, where geometric distances can characterize classification and estimation tasks.

Keywords: Fisher–Rao distance; finite-energy signals; information geometry; statistical manifolds; Christoffel symbols; geodesic equations

1. Introduction

Information geometry bridges differential geometry and statistical inference, enabling the rigorous analysis of parametric probability distributions as statistical manifolds [1]. Time series analysis and signal processing are well-established applications of information geometry. Modeling time series as realizations of statistical populations via parametric probability distributions enables the utilization of tools from differential geometry. Information geometry has been extensively applied to multivariate normal models [2–8]. With regard to signal analysis, stationary Gaussian time series are commonly examined via the statistical properties of their Fourier expansions. For instance, Amari and Nagaoka have derived the geometry of signal processing from the spectral density function, under the assumption that the sum of the squared logarithms of the power spectrum norm is finite [9] (p. 116). Extension to the modeling of finite-energy signals in additive noise remained underexplored until recently. A recent work addressed this gap by adapting the model to finite-energy signal manifolds. The representation of signals through their Fourier expansions can be used for finite-energy signals with a possible finite duration.

Check for updates

Academic Editors: Stéphane Puechmorel, Frank Nielsen and Frédéric Barbaresco

Received: 3 April 2026
Revised: 11 May 2026
Accepted: 15 May 2026
Published: 19 May 2026

Copyright: © 2026 by the author. 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.

The Fourier expansions are seen as random variables following multivariate normal distributions. Two Riemannian manifolds for finite-energy signals have been defined: the full manifold of all finite-energy signals in a given bandwidth and the sub-manifold of finite-energy signals with a known magnitude spectrum [10–12]. This previous work derived closed-form expressions of Fisher–Rao distances on these manifolds via resolution of the geodesic equations, analyzing asymptotic behaviors and demonstrating that the sub-manifold is not fully geodesic (i.e., the distances on the sub-manifold exceed those on the full manifold).

Here, we address again the case of finite-energy signal manifolds already discussed in [10–12] with dual objectives. First, we put into perspective the results already obtained with the methods of information geometry by recalling some classical formulas and providing new evidence. The objective of this first part is to explore the geometric foundations for the finite-energy signal manifolds. The elements are treated and detailed in Sections 2–5. Although geodesic equations on the multivariate normal manifold have already been proven [4,5], we propose here a unified approach based on a common method of information geometry used by the research community [1,6–8,13]. In order to make the article self-contained, we develop new proofs for the geodesic equations and Fisher–Rao distance expressions on the multivariate normal manifold, aside from references to former works.

The second objective, addressed in the Sections 6 and 7, is to analyze the numerical behavior of the Fisher–Rao distance between finite-energy signal observations. This section provides new examples not examined in the previous publications and highlights the new concept of spectral distance. We analyze the dependence of Fisher–Rao distance values with the signal-to-noise ratio, signal bandwidth and other factors as phase variations. To demonstrate that the proposed distance gives a practical advantage for signal estimation, we use a small toy problem.

Section 2 presents the method of information geometry commonly used to compute the Fisher–Rao distance based on the Fisher metric. This section references previous work by the research community [1,6,7,13]. We recall the general approach for deriving the geodesic equations. This section also provides an opportunity to introduce relevant definitions and notation conventions.

Section 3 reviews key contributions from previous research on the Riemannian geometry of the multivariate normal model, with a particular focus on the computation of geodesic distances [2–6]. Based on the method explained in Section 2, new proofs of the already known theorems are redeveloped in order to make the development self-contained.

Section 4 presents the proposed signal model and the associated statistical distributions. This signal model was previously introduced in earlier work on finite-energy signals [10–12,14]. The model is completed here to cover several signal cases. The parametric modeling is discussed in detail, leading to the definition of the manifolds $L2(\mathcal{B}, \boldsymbol{\theta})$.

Section 5 addresses the general situation of a parametric model describing a family of finite-energy signals in any manifold $L2(\mathcal{B}, \boldsymbol{\theta})$. We outline the various steps required to compute the Fisher–Rao distance, including the formulation of the differential equations for geodesics. General formulas for the Fisher–Rao distance on the manifolds identified in Section 4 are provided. The calculations of the Christoffel symbols are detailed.

Section 6 focuses on a specific manifold $L2(\mathcal{B})$, which includes all other manifolds $L2(\mathcal{B}, \boldsymbol{\theta})$ as sub-manifolds. The geodesic equations and a closed-form expression of the Fisher–Rao distance are derived for the manifold $L2(\mathcal{B})$. This closed-form expression allows defining the spectral distance function $f_{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2}(\nu)$ and highlighting its relation to the Fisher–Rao distance.

Section 7 evaluates the Fisher–Rao distance on the previously defined manifold $L2(\mathcal{B})$. The Fisher–Rao distance on this manifold is a minimum bound for the Fisher–Rao distances

on all other manifolds $L2(\mathcal{B}, \boldsymbol{\theta})$. We analyze the distance on the manifold and examine the asymptotic behaviors. We use numerical applications to compute the distance. Figures are drawn to illustrate the numerical behavior with some examples of signals.

The conclusion summarizes the main findings, discusses the relevance and limitations of the work, and outlines directions for future research.

2. How to Compute the Fisher–Rao Distance

2.1. Notation Conventions

Before describing the methodology for computing the Fisher–Rao distance, we introduce the notation conventions that will be employed throughout this paper.

Vectors and matrices are denoted in bold, while scalar values are presented in regular typeface. For example, $[\mathbf{0}]_P$ represents the zero vector with P components. The superscript T denotes the transpose operation as in the expression $\boldsymbol{a}^T$, where $\boldsymbol{a}$ is a matrix or a vector.

We note that $\boldsymbol{e_i}$ the canonical unit vector where all the components are zero, except the ith component, which is equal to one:

$$\begin{cases} e_i^j = 0, \ i \neq j \\ e_i^j = 1, \ i = j \end{cases}$$

We note that $\boldsymbol{E_{ij}}$ is the symmetric canonical matrix where all the components are zero, except for the ijth and jith components, which are equal to one:

$$\begin{cases} \boldsymbol{E_{ij}} = \boldsymbol{e_i}\boldsymbol{e_j}^T + \boldsymbol{e_j}\boldsymbol{e_i}^T, \ i \neq j \\ \boldsymbol{E_{ii}} = \boldsymbol{e_i}\boldsymbol{e_i}^T, \ i = j \end{cases}$$

Here, δ_i^j is the Kronecker symbol ($= 1$ when $i = j$ and $= 0$ otherwise).

$Re\{c\}$ and $Im\{c\}$ design the real and imaginary parts of the complex number c, respectively, where $c = Re\{c\} + \imath \cdot Im\{c\}$ with $(\imath)^2 = -1$. The superscript * denotes the conjugate operation as in the expression $\boldsymbol{c}^* = Re\{c\} - \imath \cdot Im\{c\}$.

$Tr(\boldsymbol{A})$, $\det(\boldsymbol{A})$ and $\boldsymbol{A^{-1}}$ design the trace, the determinant and the inverse matrix (when it exists) of matrix $\boldsymbol{A}$, respectively.

We adopted Einstein's summation convention to omit summation signs in tensor expressions, which allowed a concise representation of tensor operations. However, when summing over indices, such as the frequency, where the indexation is not strictly tensorial, we retained the explicit summation symbol, as shown in Equation (36).

With regard to the parameters, it is important to distinguish between $\boldsymbol{\theta}_2$, which refers to a particular value of the parameter vector $\boldsymbol{\theta}$, and θ^2, which denotes the second component of vector $\boldsymbol{\theta}$.

To avoid any ambiguity between exponentiation and the superscript index 2, exponents are written in explicit parentheses; for example, $\left(\theta^2\right)^2$ denotes the square of the component θ^2.

Where appropriate, the vector $\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\phi} \\ \boldsymbol{\varphi} \end{pmatrix}$ can be decomposed into its sub-components $\boldsymbol{\phi}$ and $\boldsymbol{\varphi}$, each with their respective indices:

- $\boldsymbol{\theta}$ is a real vector of dimension N;
- $\boldsymbol{\phi}$ is a real vector with a length P indexed by u, v or w, where $1 \leq u, v, w \leq P$;
- $\boldsymbol{\varphi}$ is a real vector with $(N - P)$ components indexed by q, r or p, where $P + 1 \leq q, r, p \leq N$.

This convention facilitates a Schouten-like notation for differentiation: $\partial_u = \frac{\partial}{\partial \phi^u}$, $\partial_v = \frac{\partial}{\partial \phi^v}$, $\partial_w = \frac{\partial}{\partial \phi^w}$, $\partial_q = \frac{\partial}{\partial \varphi^q}$, $\partial_r = \frac{\partial}{\partial \varphi^r}$ and $\partial_p = \frac{\partial}{\partial \varphi^p}$.

The gradient with respect to the parameters is, by definition, the operator:

$$\nabla_{\boldsymbol{\theta}} = \left[\frac{\partial}{\partial \theta^i}\right]_{i \in \{1,2 \ldots N\}}$$

Under the definition of $\boldsymbol{\theta}$, $\boldsymbol{\phi}$ and $\boldsymbol{\varphi}$, we have

$$\nabla_{\boldsymbol{\phi}} = \left[\frac{\partial}{\partial \phi^u}\right]_{u \in \{1,2 \ldots P\}}$$
$$\nabla_{\boldsymbol{\varphi}} = \left[\frac{\partial}{\partial \varphi^q}\right]_{q \in \{P+1,P+2 \ldots N\}}$$
$$\nabla_{\boldsymbol{\theta}} = \begin{pmatrix} \nabla_{\boldsymbol{\phi}} \\ \nabla_{\boldsymbol{\varphi}} \end{pmatrix}$$

A specific case is the one where $\boldsymbol{\theta}$ represents the parameters of a multivariate normal distribution with a mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Assuming that $\boldsymbol{\mu}$ is of a dimension n, the covariance matrix $\boldsymbol{\Sigma}$ is definite positive and can be represented by $n(n+1)/2$ coefficients. In this case, we can assume that $N = n + n(n+1)/2$ and $P = n$. For this situation, each parameter index q corresponds to a couple of indices (i, j) with $0 \leq i \leq j \leq n$ defining a component of the matrix $\boldsymbol{\Sigma}$. With this notation, we can note that $\boldsymbol{E_q} = \boldsymbol{E_{ij}}$.

2.2. Methodology to Compute the Fisher–Rao Distance

With the Fisher metric and the Levi-Civita connection, the geodesic is the curve of minimal length connecting two points. This minimal length defines a metric distance, known as the Fisher–Rao distance [3]. Using the Fisher metric, one may evaluate this statistical distance between two points, characterized by their parameter vectors $\boldsymbol{\theta_1}$ and $\boldsymbol{\theta_2}$. This distance quantifies the dissimilarity between the two statistical populations described by the probability distributions $p(\boldsymbol{x}|\boldsymbol{\theta_1})$ and $p(\boldsymbol{x}|\boldsymbol{\theta_2})$ [1]. The Fisher–Rao distance can be used for statistical analysis, including data clustering [15] and deep learning, to name but a few [2].

For conciseness, albeit at the risk of slight imprecision, the Fisher–Rao distance between the statistical distributions corresponding to two signals parameterized by $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ will hereafter simply be referred to as the distance between the two points $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. Our objective is to compute the Fisher–Rao distance between two points, $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, on the statistical manifold.

The Fisher–Rao distance computation follows the methodology described in several foundational references [1,9,13] and is widely used in the information geometry community [6,7,10–12]. The steps are as follows.

The first step is to specify a parametric model that describes the distribution of the observations. Let $\boldsymbol{x}$ denote the observation vector and $\boldsymbol{\theta}$ denote the parameter vector. The stochastic behavior of the observations is characterized by the conditional probability distribution $p(\boldsymbol{x}|\boldsymbol{\theta})$. As detailed in [9,16], the set of all such parametric probability distributions forms a statistical model, which can be regarded as a statistical manifold.

The Fisher information matrix $[\boldsymbol{g_{ij}}]$ is derived from the log-likelihood as follows:

$$g_{ij} = E_{\boldsymbol{\theta}}\left[\frac{\partial}{\partial \theta^i} \ln p(\boldsymbol{x}|\boldsymbol{\theta}) \frac{\partial}{\partial \theta^j} \ln p(\boldsymbol{x}|\boldsymbol{\theta})\right] \quad (1)$$

Based on the Fisher information matrix, the Christoffel symbols are then computed according to

$$\forall\, m,i,j=1,\ldots,N \quad \Gamma_{ij,m} := g_{mk}\Gamma^k_{ij} = \frac{1}{2}\left(\frac{\partial g_{jm}}{\partial\theta^i} + \frac{\partial g_{mi}}{\partial\theta^j} - \frac{\partial g_{ij}}{\partial\theta^m}\right) \tag{2}$$

Once the Christoffel symbols are known, they can be employed in the following differential equations, whose solutions correspond to the geodesics $\tilde{\boldsymbol{\theta}}(\varsigma)$ joining $\boldsymbol{\theta_1} = \tilde{\boldsymbol{\theta}}(0)$ and $\boldsymbol{\theta_2} = \tilde{\boldsymbol{\theta}}(1)$ (with $\varsigma \in [0,1]$):

$$\forall\, m=1,\ldots,N \quad g_{mk}\frac{d^2\theta^k}{d\varsigma^2} + \Gamma_{ij,m}\frac{d\theta^i}{d\varsigma}\frac{d\theta^j}{d\varsigma} = 0 \tag{3}$$

Given the expressions of geodesics $\tilde{\boldsymbol{\theta}}(\varsigma)$, the Fisher–Rao distance between the distribution at $\boldsymbol{\theta_1} = \tilde{\boldsymbol{\theta}}(0)$ and that at $\boldsymbol{\theta_2} = \tilde{\boldsymbol{\theta}}(1)$ is given by the following integral:

$$d(\boldsymbol{\theta_1},\boldsymbol{\theta_2}) = \int_0^1 \sqrt{g_{ij}\frac{d\tilde{\theta}^i}{d\varsigma}\frac{d\tilde{\theta}^j}{d\varsigma}}\,d\varsigma \tag{4}$$

Constraints in the parametric model of probability distributions restrict the distribution to a sub-manifold, and the Fisher–Rao geodesic distance on the sub-manifold is generally larger than the Fisher–Rao geodesic distance measured on the global manifold.

3. The Case of Multivariate Normal Distributions

3.1. The General Case

Finding closed-form expressions for the Fisher–Rao distance is recognized as a non-trivial endeavor, and such expressions are only available for a limited number of families of probability distributions [7]. Here, we consider multivariate normal distributions $\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$ given by

$$p(\boldsymbol{X}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{2^{\frac{n}{2}}\pi^{\frac{n}{2}}\sqrt{\det(\boldsymbol{\Sigma})}}\exp\left(-\frac{(\boldsymbol{X}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})}{2}\right) \tag{5}$$

Here, the following definitions apply:

- n is an integer, the number of components;
- $\boldsymbol{X}$ is a real random vector with n components;
- $\boldsymbol{\Sigma}$ is an $n \times n$ positive-definite matrix;
- $\boldsymbol{\mu}$ is a real vector with n components.

When the parameters are all the coefficients of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, geodesic equations corresponding to multivariate normal distributions $\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$ take the form of second-order ordinary differential equations (with $\varsigma \in [0,1]$) [4–6] :

$$\begin{cases} \frac{d^2\boldsymbol{\mu}}{d\varsigma^2} - \frac{d\boldsymbol{\Sigma}}{d\varsigma}\boldsymbol{\Sigma}^{-1}\frac{d\boldsymbol{\mu}}{d\varsigma} = [\mathbf{0}]_{\boldsymbol{n}} \\ \frac{d^2\boldsymbol{\Sigma}}{d\varsigma^2} + \frac{d\boldsymbol{\mu}}{d\varsigma}\frac{d\boldsymbol{\mu}^T}{d\varsigma} - \frac{d\boldsymbol{\Sigma}}{d\varsigma}\boldsymbol{\Sigma}^{-1}\frac{d\boldsymbol{\Sigma}}{d\varsigma} = [\mathbf{0}]_{\boldsymbol{n}\times\boldsymbol{n}} \end{cases} \tag{6}$$

Proof. The proof relies on the method explained in Section 2.

First, we compute the Fisher matrix. From Equation (5), we can write

$$\ln p(\boldsymbol{x}|\boldsymbol{\theta}) = -\frac{n}{2}\ln 2\pi - \frac{1}{2}\ln\det(\boldsymbol{\Sigma}) - \frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})$$

Assuming that u, v and w represent the parameter indices (varying from 1 to P) of the mean $\boldsymbol{\mu}$ and q, r and p represent the parameter indices (varying from P + 1 to N) of the covariance matrix $\boldsymbol{\Sigma}$, we can write the following:

$$\partial_u \ln p(\boldsymbol{x}|\boldsymbol{\theta}) = \partial_u \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) \tag{7}$$

$$\partial_q \ln p(\boldsymbol{x}|\boldsymbol{\theta}) = -\frac{1}{2}\partial_q \ln \det(\boldsymbol{\Sigma}) - \frac{1}{2}(\boldsymbol{X} - \boldsymbol{\mu})^T \partial_q \boldsymbol{\Sigma}^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) \tag{8}$$

From Equation (7), we can see that

$$\partial_q \partial_u \ln p(\boldsymbol{x}|\boldsymbol{\theta}) = \partial_u \boldsymbol{\mu}^T \partial_q \boldsymbol{\Sigma}^{-1}(\boldsymbol{X} - \boldsymbol{\mu})$$

We know from [9] (p. 28) that the Fisher metric can also be written as $g_{qu} = E_{\boldsymbol{\theta}}\left[-\partial_q \partial_u \ln p(\boldsymbol{x}|\boldsymbol{\theta})\right]$. Thus, we can easily obtain

$$\begin{aligned} g_{qu} &= -\partial_u \boldsymbol{\mu}^T \partial_q \boldsymbol{\Sigma}^{-1} E_{\boldsymbol{\theta}}[\boldsymbol{X} - \boldsymbol{\mu}] \\ &= 0 \end{aligned}$$

As $g_{uv} = E_{\boldsymbol{\theta}}[-\partial_v \partial_u \ln p(\boldsymbol{x}|\boldsymbol{\theta})]$, we also have

$$\begin{aligned} g_{uv} &= -\partial_u \partial_v \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} E_{\boldsymbol{\theta}}[\boldsymbol{X} - \boldsymbol{\mu}] + \partial_u \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \partial_v \boldsymbol{\mu} \\ &= \partial_u \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \partial_v \boldsymbol{\mu} \\ &= \boldsymbol{e_u}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{e_v} \end{aligned}$$

Knowing from algebra rules that $d \ln \det(\boldsymbol{A}) = Tr\left(\boldsymbol{A^{-1}} d\boldsymbol{A}\right)$, $d\boldsymbol{A^{-1}} = -\boldsymbol{A^{-1}} d\boldsymbol{A} \boldsymbol{A^{-1}}$ and $Tr(\boldsymbol{AB}) = Tr(\boldsymbol{BA})$, we can deduce from Equation (8) the following equations:

$$\begin{aligned} \partial_q \ln p(\boldsymbol{x}|\boldsymbol{\theta}) &= -\frac{1}{2} Tr\left(\boldsymbol{\Sigma^{-1}} \partial_q \boldsymbol{\Sigma}\right) + \frac{1}{2}(\boldsymbol{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma^{-1}} \partial_q \boldsymbol{\Sigma} \boldsymbol{\Sigma^{-1}} (\boldsymbol{X} - \boldsymbol{\mu}) \\ &= -\frac{1}{2} Tr\left(\boldsymbol{\Sigma^{-1}} \partial_q \boldsymbol{\Sigma} - (\boldsymbol{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma^{-1}} \partial_q \boldsymbol{\Sigma} \boldsymbol{\Sigma^{-1}} (\boldsymbol{X} - \boldsymbol{\mu})\right) \\ &= -\frac{1}{2} Tr\left(\boldsymbol{\Sigma^{-1}} \partial_q \boldsymbol{\Sigma} - \boldsymbol{\Sigma^{-1}} \partial_q \boldsymbol{\Sigma} \boldsymbol{\Sigma^{-1}} (\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^T\right) \\ &= -\frac{1}{2} Tr\left(\partial_q \boldsymbol{\Sigma} \boldsymbol{\Sigma^{-1}} \left(\boldsymbol{\Sigma} - (\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^T\right) \boldsymbol{\Sigma^{-1}}\right) \end{aligned}$$

As each q corresponds to a couple of indices (i, j) with $i \leq j$ defining a component of the matrix $\boldsymbol{\Sigma}$, we can note $\boldsymbol{E_q} = \boldsymbol{E_{ij}}$, and we have the following:

$$\partial_q \boldsymbol{\Sigma} = \boldsymbol{E_q} \tag{9}$$

and

$$\partial_q \boldsymbol{\Sigma^{-1}} = -\boldsymbol{\Sigma^{-1}} \boldsymbol{E_q} \boldsymbol{\Sigma^{-1}} \tag{10}$$

This leads to the following equations:

$$\begin{aligned} \partial_q \ln p(\boldsymbol{x}|\boldsymbol{\theta}) &= -\frac{1}{2} Tr\left(\boldsymbol{E_q} \boldsymbol{\Sigma^{-1}} \left(\boldsymbol{\Sigma} - (\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^T\right) \boldsymbol{\Sigma^{-1}}\right) \\ &= -\frac{1}{2} Tr\left(\boldsymbol{E_q} \boldsymbol{\Sigma^{-1}} - \boldsymbol{E_q} \boldsymbol{\Sigma^{-1}} (\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma^{-1}}\right) \end{aligned}$$

With r corresponding to a couple (k, l) of parameter indices in the matrix $\boldsymbol{\Sigma}$ with $k \leq l$, we can write

$$\begin{aligned}\partial_r\partial_q \ln p(\boldsymbol{x}|\boldsymbol{\theta}) = &-\frac{1}{2}Tr\Big(\boldsymbol{E_q}\partial_r\boldsymbol{\Sigma^{-1}}\Big)\\ &+\frac{1}{2}Tr\Big(\boldsymbol{E_q}\partial_r\boldsymbol{\Sigma^{-1}}(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^T\boldsymbol{\Sigma^{-1}}\Big)\\ &+\frac{1}{2}Tr\Big(\boldsymbol{E_q}\boldsymbol{\Sigma^{-1}}(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^T\partial_r\boldsymbol{\Sigma^{-1}}\Big)\end{aligned}$$

We know from [9] (p. 28) that the Fisher metric can also be writen as $g_{rq} = E_{\boldsymbol{\theta}}\big[-\partial_r\partial_q \ln p(\boldsymbol{x}|\boldsymbol{\theta})\big]$, and by definition, the covariance matrix $\boldsymbol{\Sigma} = E_{\boldsymbol{\theta}}\Big[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^T\Big]$. Therefore, we can write the following:

$$\begin{aligned}g_{rq} = &+\frac{1}{2}Tr\Big(\boldsymbol{E_q}\partial_r\boldsymbol{\Sigma^{-1}}\Big)\\ &-\frac{1}{2}Tr\Big(\boldsymbol{E_q}\partial_r\boldsymbol{\Sigma^{-1}}\boldsymbol{\Sigma}\boldsymbol{\Sigma^{-1}}\Big)\\ &-\frac{1}{2}Tr\Big(\boldsymbol{E_q}\boldsymbol{\Sigma^{-1}}\boldsymbol{\Sigma}\partial_r\boldsymbol{\Sigma^{-1}}\Big)\\ =&-\frac{1}{2}Tr\Big(\boldsymbol{E_q}\partial_r\boldsymbol{\Sigma^{-1}}\Big)\\ =&-\frac{1}{2}Tr\Big(\boldsymbol{E_q}\Big(-\boldsymbol{\Sigma^{-1}}\partial_r\boldsymbol{\Sigma}\boldsymbol{\Sigma^{-1}}\Big)\Big)\\ =&\frac{1}{2}Tr\Big(\boldsymbol{E_q}\boldsymbol{\Sigma^{-1}}\boldsymbol{E_r}\boldsymbol{\Sigma^{-1}}\Big)\end{aligned}$$

As a consequence, the Fisher matrix is a block diagonal matrix:

$$[\boldsymbol{g_{ij}}] = \begin{pmatrix}[\boldsymbol{g_{uv}}] & [\boldsymbol{0}]\\ [\boldsymbol{0}] & [\boldsymbol{g_{qr}}]\end{pmatrix} \tag{11}$$

with

$$g_{uv} = \boldsymbol{e_u}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{e_v} \tag{12}$$

$$g_{qu} = 0 \tag{13}$$

$$g_{rq} = \frac{1}{2}Tr\Big(\boldsymbol{E_q}\boldsymbol{\Sigma^{-1}}\boldsymbol{E_r}\boldsymbol{\Sigma^{-1}}\Big) \tag{14}$$

The second step of the proof consists of computing the $\Gamma_{ij,m}$.
Regarding $\Gamma_{ij,m}$, from Equation (2), we have the following relationships:

$$\begin{aligned}\partial_w g_{uv} &= \partial_w\Big(\boldsymbol{e_u}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{e_v}\Big)\\ &= 0\\ \partial_w g_{rq} &= \frac{1}{2}Tr\Big(\partial_w\Big(\boldsymbol{E_q}\boldsymbol{\Sigma^{-1}}\boldsymbol{E_r}\boldsymbol{\Sigma^{-1}}\Big)\Big)\\ &= 0\\ \partial_w g_{qu} &= 0\end{aligned}$$

With p corresponding to a couple (f,g) of parameter indices in the matrix $\boldsymbol{\Sigma}$ with $f \leq g$, we can write

$$
\begin{aligned}
\partial_p g_{uv} &= \boldsymbol{e_u}^T \partial_p \boldsymbol{\Sigma}^{-1} \boldsymbol{e_v} \\
&= -\boldsymbol{e_u}^T \boldsymbol{\Sigma^{-1}} \boldsymbol{E_p} \boldsymbol{\Sigma^{-1}} \boldsymbol{e_v} \\
\partial_p g_{qu} &= 0 \\
\partial_p g_{rq} &= \frac{1}{2} Tr\left(\partial_p \left(\boldsymbol{E_q} \boldsymbol{\Sigma^{-1}} \boldsymbol{E_r} \boldsymbol{\Sigma^{-1}}\right)\right) \\
&= \frac{1}{2} Tr\left(\boldsymbol{E_q} \partial_p \boldsymbol{\Sigma^{-1}} \boldsymbol{E_r} \boldsymbol{\Sigma^{-1}} + \boldsymbol{E_q} \boldsymbol{\Sigma^{-1}} \boldsymbol{E_r} \partial_p \boldsymbol{\Sigma^{-1}}\right) \\
&= -\frac{1}{2} Tr\left(\boldsymbol{E_q} \boldsymbol{\Sigma^{-1}} \boldsymbol{E_p} \boldsymbol{\Sigma^{-1}} \boldsymbol{E_r} \boldsymbol{\Sigma^{-1}} + \boldsymbol{E_q} \boldsymbol{\Sigma^{-1}} \boldsymbol{E_r} \boldsymbol{\Sigma^{-1}} \boldsymbol{E_p} \boldsymbol{\Sigma^{-1}}\right) \\
&= -\frac{1}{2} Tr\left(\boldsymbol{E_q} \boldsymbol{\Sigma^{-1}} \boldsymbol{E_p} \boldsymbol{\Sigma^{-1}} \boldsymbol{E_r} \boldsymbol{\Sigma^{-1}} + \boldsymbol{\Sigma^{-1}} \boldsymbol{E_p} \boldsymbol{\Sigma^{-1}} \boldsymbol{E_r} \boldsymbol{\Sigma^{-1}} \boldsymbol{E_q}\right) \\
&= -\frac{1}{2} Tr\left(\boldsymbol{E_q} \boldsymbol{\Sigma^{-1}} \boldsymbol{E_p} \boldsymbol{\Sigma^{-1}} \boldsymbol{E_r} \boldsymbol{\Sigma^{-1}} + \boldsymbol{E_q} \boldsymbol{\Sigma^{-1}} \boldsymbol{E_p} \boldsymbol{\Sigma^{-1}} \boldsymbol{E_r} \boldsymbol{\Sigma^{-1}}\right) \\
&= -Tr\left(\boldsymbol{E_q} \boldsymbol{\Sigma^{-1}} \boldsymbol{E_p} \boldsymbol{\Sigma^{-1}} \boldsymbol{E_r} \boldsymbol{\Sigma^{-1}}\right) \\
&= \partial_q g_{pr} = \partial_r g_{pq}
\end{aligned}
$$

As a consequence, Equation (2) can be expressed as follows:

$$
\begin{aligned}
\Gamma_{uv,w} &= 0 \\
\Gamma_{uv,q} &= \frac{1}{2}\left(\partial_u g_{vq} + \partial_v g_{uq} - \partial_q g_{uv}\right) \\
&= \frac{1}{2}\left(0 + 0 + \boldsymbol{e_u}^T \boldsymbol{\Sigma^{-1}} \boldsymbol{E_q} \boldsymbol{\Sigma^{-1}} \boldsymbol{e_v}\right) \\
&= \frac{1}{2} \boldsymbol{e_u}^T \boldsymbol{\Sigma^{-1}} \boldsymbol{E_q} \boldsymbol{\Sigma^{-1}} \boldsymbol{e_v} \\
\Gamma_{uq,w} &= \frac{1}{2}\left(\partial_u g_{wq} + \partial_q g_{uw} - \partial_w g_{uq}\right) \\
&= \frac{1}{2}\left(0 + \partial_q g_{uw} - 0\right) \\
&= \frac{1}{2}\left(0 - \boldsymbol{e_u}^T \boldsymbol{\Sigma^{-1}} \boldsymbol{E_q} \boldsymbol{\Sigma^{-1}} \boldsymbol{e_w} - 0\right) \\
&= -\frac{1}{2} \boldsymbol{e_u}^T \boldsymbol{\Sigma^{-1}} \boldsymbol{E_q} \boldsymbol{\Sigma^{-1}} \boldsymbol{e_w}
\end{aligned}
$$

$$
\begin{aligned}
\Gamma_{qu,w} &= \frac{1}{2}\left(\partial_q g_{wu} + \partial_u g_{qw} - \partial_w g_{qu}\right) \\
&= \frac{1}{2}\left(\partial_q g_{wu} + 0 - 0\right) \\
&= \frac{1}{2}\left(-\boldsymbol{e_w}^T \boldsymbol{\Sigma^{-1}} \boldsymbol{E_q} \boldsymbol{\Sigma^{-1}} \boldsymbol{e_u} + 0 - 0\right) \\
&= -\frac{1}{2} \boldsymbol{e_w}^T \boldsymbol{\Sigma^{-1}} \boldsymbol{E_q} \boldsymbol{\Sigma^{-1}} \boldsymbol{e_u} \\
&= \Gamma_{uq,w}
\end{aligned}
$$

$$\begin{aligned}
\Gamma_{uq,r} &= \frac{1}{2}\left(\partial_u g_{qr} + \partial_q g_{ur} - \partial_r g_{uq}\right) \\
&= \frac{1}{2}(0+0-0) \\
&= 0 \\
\Gamma_{qr,w} &= 0 \\
\Gamma_{qr,p} &= \frac{1}{2}\left(\partial_q g_{pr} + \partial_r g_{qp} - \partial_p g_{qr}\right) \\
&= \frac{1}{2}\left(\partial_q g_{pr} + \partial_r g_{qp} - \partial_p g_{qr}\right) \\
&= \frac{1}{2}\left(\partial_q g_{pr}\right) \\
&= -\frac{1}{2} Tr\left(\boldsymbol{E_q}\boldsymbol{\Sigma^{-1}}\boldsymbol{E_p}\boldsymbol{\Sigma^{-1}}\boldsymbol{E_r}\boldsymbol{\Sigma^{-1}}\right)
\end{aligned}$$

The third step of the proof consists of replacing the values of $\Gamma_{ij,m}$ in the geodesic equation (Equation (3)).

As a consequence, Equation (3) becomes

$$\begin{cases} \forall u & g_{uv}\frac{d^2\theta^v}{d\varsigma^2} + \Gamma_{wq,u}\frac{d\theta^w}{d\varsigma}\frac{d\theta^q}{d\varsigma} + \Gamma_{qw,u}\frac{d\theta^q}{d\varsigma}\frac{d\theta^w}{d\varsigma} = 0 \\ \forall q & g_{qr}\frac{d^2\theta^r}{d\varsigma^2} + \Gamma_{uv,q}\frac{d\theta^u}{d\varsigma}\frac{d\theta^v}{d\varsigma} + \Gamma_{pr,q}\frac{d\theta^p}{d\varsigma}\frac{d\theta^r}{d\varsigma} = 0 \end{cases}$$

These equations can be rewritten as follows:

$$\begin{cases} \forall u & \boldsymbol{e_u}^T\boldsymbol{\Sigma^{-1}}\boldsymbol{e_v}\frac{d^2\theta^v}{d\varsigma^2} - \frac{1}{2}\boldsymbol{e_u}^T\boldsymbol{\Sigma^{-1}}\boldsymbol{E_q}\boldsymbol{\Sigma^{-1}}\boldsymbol{e_w}\frac{d\theta^w}{d\varsigma}\frac{d\theta^q}{d\varsigma} - \frac{1}{2}\boldsymbol{e_u}^T\boldsymbol{\Sigma^{-1}}\boldsymbol{E_q}\boldsymbol{\Sigma^{-1}}\boldsymbol{e_w}\frac{d\theta^w}{d\varsigma}\frac{d\theta^q}{d\varsigma} = 0 \\ \forall q & Tr\left(\boldsymbol{E_q}\boldsymbol{\Sigma^{-1}}\boldsymbol{E_r}\boldsymbol{\Sigma^{-1}}\right)\frac{d^2\theta^r}{d\varsigma^2} + \boldsymbol{e_u}^T\boldsymbol{\Sigma^{-1}}\boldsymbol{E_q}\boldsymbol{\Sigma^{-1}}\boldsymbol{e_v}\frac{d\theta^u}{d\varsigma}\frac{d\theta^v}{d\varsigma} - Tr\left(\boldsymbol{E_q}\boldsymbol{\Sigma^{-1}}\boldsymbol{E_p}\boldsymbol{\Sigma^{-1}}\boldsymbol{E_r}\boldsymbol{\Sigma^{-1}}\right)\frac{d\theta^p}{d\varsigma}\frac{d\theta^r}{d\varsigma} = 0 \end{cases}$$

In other words, we have

$$\begin{cases} \forall u & \boldsymbol{e_u}^T\boldsymbol{\Sigma^{-1}}\left(\boldsymbol{e_v}\frac{d^2\theta^v}{d\varsigma^2} - 2\cdot\frac{1}{2}\frac{d\theta^q}{d\varsigma}\boldsymbol{E_q}\boldsymbol{\Sigma^{-1}}\boldsymbol{e_w}\frac{d\theta^w}{d\varsigma}\right) = 0 \\ \forall q & Tr\left(\boldsymbol{\Sigma^{-1}}\boldsymbol{E_q}\boldsymbol{\Sigma^{-1}}\boldsymbol{E_r}\frac{d^2\theta^r}{d\varsigma^2} + \boldsymbol{\Sigma^{-1}}\boldsymbol{E_q}\boldsymbol{\Sigma^{-1}}\boldsymbol{e_v}\boldsymbol{e_u}^T\frac{d\theta^u}{d\varsigma}\frac{d\theta^v}{d\varsigma} - \boldsymbol{\Sigma^{-1}}\boldsymbol{E_q}\boldsymbol{\Sigma^{-1}}\boldsymbol{E_p}\boldsymbol{\Sigma^{-1}}\boldsymbol{E_r}\frac{d\theta^p}{d\varsigma}\frac{d\theta^r}{d\varsigma}\right) = 0 \end{cases}$$

and finally

$$\begin{cases} \forall u & \boldsymbol{e_u}^T\boldsymbol{\Sigma^{-1}}\left(\boldsymbol{e_v}\frac{d^2\theta^v}{d\varsigma^2} - \frac{d\theta^q}{d\varsigma}\boldsymbol{E_q}\boldsymbol{\Sigma^{-1}}\boldsymbol{e_w}\frac{d\theta^w}{d\varsigma}\right) = 0 \\ \forall q & Tr\left[\boldsymbol{\Sigma^{-1}}\boldsymbol{E_q}\boldsymbol{\Sigma^{-1}}\left(\boldsymbol{E_r}\frac{d^2\theta^r}{d\varsigma^2} + \frac{d\theta^v}{d\varsigma}\boldsymbol{e_v}\boldsymbol{e_u}^T\frac{d\theta^u}{d\varsigma} - \boldsymbol{E_p}\frac{d\theta^p}{d\varsigma}\boldsymbol{\Sigma^{-1}}\boldsymbol{E_r}\frac{d\theta^r}{d\varsigma}\right)\right] = 0 \end{cases}$$

We know that $\boldsymbol{a}^T\boldsymbol{b}$ defines a scalar product between the two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, while $Tr[\boldsymbol{A}^T\boldsymbol{B}]$ defines a scalar product between the two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$.

As $\left\{\forall u \;\; \boldsymbol{\Sigma^{-1}}\boldsymbol{e_u}\right\}$ is the basis of $\mathbb{R}^n$, and $\left\{\forall q \;\; \boldsymbol{\Sigma^{-1}}\boldsymbol{E_q}\boldsymbol{\Sigma^{-1}}\right\}$ is a basis of the vector space of all symmetric $n \times n$ matrices, the two equations become

$$\begin{cases} \boldsymbol{e_v}\frac{d^2\theta^v}{d\varsigma^2} - \frac{d\theta^q}{d\varsigma}\boldsymbol{E_q}\boldsymbol{\Sigma^{-1}}\boldsymbol{e_w}\frac{d\theta^w}{d\varsigma} = [\boldsymbol{0}]_n \\ \boldsymbol{E_r}\frac{d^2\theta^r}{d\varsigma^2} + \frac{d\theta^v}{d\varsigma}\boldsymbol{e_v}\boldsymbol{e_u}^T\frac{d\theta^u}{d\varsigma} - \boldsymbol{E_p}\frac{d\theta^p}{d\varsigma}\boldsymbol{\Sigma^{-1}}\boldsymbol{E_r}\frac{d\theta^r}{d\varsigma} = [\boldsymbol{0}]_{n\times n} \end{cases}$$

Due to the definitions of $\boldsymbol{e_v}$ and $\boldsymbol{E_r}$, we can write

$$\begin{aligned} \boldsymbol{\mu} &= \theta^u\boldsymbol{e_u} \\ \boldsymbol{\Sigma} &= \theta^q\boldsymbol{E_q} \end{aligned}$$

https://doi.org/10.3390/e28050569

$$\frac{d\boldsymbol{\mu}}{d\varsigma} = \frac{d\theta^u}{d\varsigma}\boldsymbol{e_u}$$
$$\frac{d\boldsymbol{\Sigma}}{d\varsigma} = \frac{d\theta^q}{d\varsigma}\boldsymbol{E_q}$$

$$\frac{d^2\boldsymbol{\mu}}{d\varsigma^2} = \frac{d^2\theta^u}{d\varsigma^2}\boldsymbol{e_u}$$
$$\frac{d^2\boldsymbol{\Sigma}}{d\varsigma^2} = \frac{d^2\theta^q}{d\varsigma^2}\boldsymbol{E_q}$$

and the two geodesic equations can be written as follows:

$$\begin{cases} \frac{d^2\boldsymbol{\mu}}{d\varsigma^2} - \frac{d\boldsymbol{\Sigma}}{d\varsigma}\boldsymbol{\Sigma^{-1}}\frac{d\boldsymbol{\mu}}{d\varsigma} = [\mathbf{0}]_{\boldsymbol{n}} \\ \frac{d^2\boldsymbol{\Sigma}}{d\varsigma^2} + \frac{d\boldsymbol{\mu}}{d\varsigma}\frac{d\boldsymbol{\mu}^T}{d\varsigma} - \frac{d\boldsymbol{\Sigma}}{d\varsigma}\boldsymbol{\Sigma^{-1}}\frac{d\boldsymbol{\Sigma}}{d\varsigma} = [\mathbf{0}]_{\boldsymbol{n}\times\boldsymbol{n}} \end{cases}$$

□

These geodesic equations have been solved by fixing the initial value conditions in $\boldsymbol{\mu}$ and $\frac{d\boldsymbol{\mu}}{dt}$ [3,8]. However, they were not solved in the general case for boundary conditions, i.e., fixing $\boldsymbol{\mu}$ at $\varsigma = 0$, $\varsigma = 1$ ($\boldsymbol{\mu}(0) = \boldsymbol{\mu}_1$ and $\boldsymbol{\mu}(1) = \boldsymbol{\mu}_2$).

Once we have the expressions of the solutions of the geodesic equations $\tilde{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{\Sigma}}$, we can compute the Fisher–Rao distance using the following formula:

$$d(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) = \int_0^1 \sqrt{\frac{d\tilde{\boldsymbol{\mu}}}{d\varsigma}^T \tilde{\boldsymbol{\Sigma}}^{-1}\frac{d\tilde{\boldsymbol{\mu}}}{d\varsigma} - \frac{1}{2}Tr\left(\frac{d\tilde{\boldsymbol{\Sigma}}}{d\varsigma}\frac{d\tilde{\boldsymbol{\Sigma}}^{-1}}{d\varsigma}\right)}\, d\varsigma \tag{15}$$

Proof. From Equations (4) and (11), we can write

$$\begin{aligned} d(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) &= \int_0^1 \sqrt{g_{uv}\frac{d\tilde{\theta}^u}{d\varsigma}\frac{d\tilde{\theta}^v}{d\varsigma} + g_{qr}\frac{d\tilde{\theta}^q}{d\varsigma}\frac{d\tilde{\theta}^r}{d\varsigma}}\, d\varsigma \\ &= \int_0^1 \sqrt{\frac{d\tilde{\theta}^u}{d\varsigma}\boldsymbol{e_u}^T\tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{e_v}\frac{d\tilde{\theta}^v}{d\varsigma} + \frac{1}{2}Tr\left(\frac{d\tilde{\theta}^q}{d\varsigma}\boldsymbol{E_q}\tilde{\boldsymbol{\Sigma}}^{-1}\frac{d\tilde{\theta}^r}{d\varsigma}\boldsymbol{E_r}\tilde{\boldsymbol{\Sigma}}^{-1}\right)}\, d\varsigma \\ &= \int_0^1 \sqrt{\frac{d\tilde{\boldsymbol{\mu}}}{d\varsigma}^T\tilde{\boldsymbol{\Sigma}}^{-1}\frac{d\tilde{\boldsymbol{\mu}}}{d\varsigma} + \frac{1}{2}Tr\left(\frac{d\tilde{\boldsymbol{\Sigma}}}{d\varsigma}\tilde{\boldsymbol{\Sigma}}^{-1}\frac{d\tilde{\boldsymbol{\Sigma}}}{d\varsigma}\tilde{\boldsymbol{\Sigma}}^{-1}\right)}\, d\varsigma \\ &= \int_0^1 \sqrt{\frac{d\tilde{\boldsymbol{\mu}}}{d\varsigma}^T\tilde{\boldsymbol{\Sigma}}^{-1}\frac{d\tilde{\boldsymbol{\mu}}}{d\varsigma} - \frac{1}{2}Tr\left(\frac{d\tilde{\boldsymbol{\Sigma}}}{d\varsigma}\frac{d\tilde{\boldsymbol{\Sigma}}^{-1}}{d\varsigma}\right)}\, d\varsigma \end{aligned}$$

□

3.2. The Case Where $\boldsymbol{\Sigma}$ *Is Constant*

As stated above, in the general case, there is no explicit closed-form expression for the Fisher–Rao distance between multivariate normal probability distributions. Nevertheless, a closed-form expression can be derived in particular instances, such as multivariate normal distributions that share a common covariance matrix. The solutions for the equations with the boundary conditions are known for the distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ in the case where $\boldsymbol{\Sigma}$ is the constant [6]. Below, we examine this case, which is of interest for the signal model.

When we consider the distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with the $\boldsymbol{\Sigma}$ constant, and when we have $\boldsymbol{\theta} = \boldsymbol{\mu}$, the second line of Equation (6) does not exist, and the geodesic equation becomes

$$\frac{d^2\boldsymbol{\mu}}{d\varsigma^2} = [\mathbf{0}]_{\boldsymbol{n}} \tag{16}$$

Proof. From Equation (5), we can write

$$\ln p(\boldsymbol{x}|\boldsymbol{\theta}) = -\frac{n}{2}\ln 2\pi - \frac{1}{2}\ln\det(\boldsymbol{\Sigma}) - \frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})$$

Then, using the Leibniz rule, we have

$$\frac{\partial}{\partial\theta^i}\ln p(\boldsymbol{x}|\boldsymbol{\theta}) = -\frac{1}{2}\left(-\frac{\partial\boldsymbol{\mu}}{\partial\theta^i}\right)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu}) - \frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}\left(-\frac{\partial\boldsymbol{\mu}}{\partial\theta^i}\right)$$

And because $\boldsymbol{\Sigma}$ is symmetric, we also have

$$\frac{\partial}{\partial\theta^i}\ln p(\boldsymbol{x}|\boldsymbol{\theta}) = \left(\frac{\partial\boldsymbol{\mu}}{\partial\theta^i}\right)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})$$

As we have $\boldsymbol{\Sigma} = E_{\boldsymbol{\theta}}\left[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^T\right]$, we can write in Equation (1)

$$\begin{aligned} g_{ij} &= E_{\boldsymbol{\theta}}\left[\left(\frac{\partial\boldsymbol{\mu}}{\partial\theta^i}\right)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}\left(\frac{\partial\boldsymbol{\mu}}{\partial\theta^j}\right)\right] \\ &= \left(\frac{\partial\boldsymbol{\mu}}{\partial\theta^i}\right)^T\boldsymbol{\Sigma}^{-1}E_{\boldsymbol{\theta}}\left[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^T\right]\boldsymbol{\Sigma}^{-1}\left(\frac{\partial\boldsymbol{\mu}}{\partial\theta^j}\right) \\ &= \left(\frac{\partial\boldsymbol{\mu}}{\partial\theta^i}\right)^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\left(\frac{\partial\boldsymbol{\mu}}{\partial\theta^j}\right) \\ &= \frac{\partial\boldsymbol{\mu}^{\boldsymbol{T}}}{\partial\theta^i}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\mu}}{\partial\theta^j} \end{aligned}$$

As $\boldsymbol{\theta} = \boldsymbol{\mu}$, we have $\frac{\partial\boldsymbol{\mu}}{\partial\theta^i} = \boldsymbol{e_i}$, where all the components of $\boldsymbol{e_i}$ are zero except the ith component, which is equal to one:

$$\begin{cases} e_i^j = 0, \ i \neq j \\ e_i^j = 1, \ i = j \end{cases}$$

and thus

$$g_{ij} = \boldsymbol{e_i}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{e_j}$$

This means that $\boldsymbol{G} = \boldsymbol{\Sigma^{-1}}$, and the partial derivatives are zero: $\frac{\partial g_{ij}}{\partial\theta^k} = 0$.
And due to Equation (2), we have

$$\forall\, m,i,j = 1,\ldots,n \quad \Gamma_{ij,m} := 0$$

This leads to the following transformation of Equation (3):

$$\forall\, m = 1,\ldots,n \quad g_{mk}\frac{d^2\mu^k}{d\varsigma^2} = 0$$

This last equation is equivalent to (16).

□

In this scenario, assuming Equation (16), the Fisher–Rao distance reduces to the Mahalanobis distance between the means of the multivariate normal distributions [6]; that

is, when we consider the distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with the $\boldsymbol{\Sigma}$ constant, and when we have $\boldsymbol{\theta} = \boldsymbol{\mu}$, we can write

$$d(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) = d(\boldsymbol{\mu_1}, \boldsymbol{\mu_2}) = \sqrt{(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)} \tag{17}$$

Proof. From Equation (16), we deduce that $\frac{d\boldsymbol{\mu}}{d\varsigma}$ is a constant vector:

$$\frac{d\boldsymbol{\mu}}{d\varsigma} = \boldsymbol{\mu}_2 - \boldsymbol{\mu}_1$$

and by replacing it in Equation (4) we get

$$\begin{aligned} d(\boldsymbol{\mu_1}, \boldsymbol{\mu_2}) &= \int_0^1 \sqrt{\frac{d\boldsymbol{\mu}^T}{d\varsigma} \boldsymbol{G} \frac{d\boldsymbol{\mu}}{d\varsigma}} d\varsigma \\ &= \int_0^1 \sqrt{(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma^{-1}} (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)} d\varsigma \\ &= \sqrt{(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma^{-1}} (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)} \end{aligned}$$

□

For more general cases, the distance generally does not admit an explicit closed form and must be approximately evaluated with numerical techniques such as geodesic shooting or bound approximations [3].

4. Signal Parametric Modeling

4.1. Observation of a Finite-Energy Signal

The model used to derive the expression of the statistical distribution has already been applied in previous works [10–12,14]. We detail here the hypotheses and give the expression for the law of probability of the observations and its dependence on the parameters.

We are interested in applications recording finite-energy signals. As examples, we can consider chirp frequency-modulated (FM) signals.

At the sensor level, the received signal $s_r(t)$ is a real signal with finite energy, depending on the time t, for which we have $\int_{-\infty}^{+\infty} (s_r(t))^2 dt < \infty$. This signal admits a Fourier expansion $\hat{s}_r(\nu)$, depending on the frequency ν. In the real world, the signal is only observed during a limited time interval—$t \in [0, T]$—and most of the time, it has a finite duration.

We examine the observations after sampling and applying the Fourier transform. Because, for real signals, the Fourier transform values at negative frequencies are the complex conjugates of the Fourier transform values at positive frequencies, we limited the observations to the positive frequencies. The sensor has a limited bandwidth $\check{B}$, which is supposed to be a continuous interval in $\mathbb{R}^+$. Due to the limited frequency bandwidth $\check{B}$ of the sensor, time sampling effect and discrete Fourier transform (DFT), the observation bandwidth is limited to a subset $\mathcal{B}$ of positive frequencies: $\mathcal{B} \subset \check{B} \subset \mathbb{R}^+$. In practice, this subset $\mathcal{B}$ is made up of $N_\mathcal{B}$ regularly spaced discrete values $\mathcal{B} = \{\nu_1, \nu_2, \ldots \nu_{N_\mathcal{B}}\}$. After time sampling and the DFT, and from the receiver point of view, the signal is parameterized by a real vector $\boldsymbol{\theta}$, and for frequency dependence, we note that $s_{\boldsymbol{\theta}}(\nu)$.

As a consequence of Parseval's theorem, we can write $\sum_{\nu \in \mathcal{B}} |s_{\boldsymbol{\theta}}(\nu)|^2 \delta\nu < \infty$ (with $\delta\nu = B/N_\mathcal{B}$, where B is the length of the interval $\check{B}$).

In addition to the above hypotheses, we assume that the observation is composed of the signal mixed with additive stationary random noise. Thus, the observation takes the form.

$$\forall \nu \in \mathcal{B} \quad x(\nu) = s_{\boldsymbol{\theta}}(\nu) + n(\nu) \tag{18}$$

In this equation, $x(\nu)$, $s_{\boldsymbol{\theta}}(\nu)$ and $n(\nu)$ are complex numbers. The total vector of observation $\boldsymbol{x}$ is made up of all complex variables $x(\nu)$ in the observation bandwidth: $\boldsymbol{x} = (x(\nu))_{\nu \in \mathcal{B}}$. This vector has a total of $N_{\mathcal{B}}$ complex coordinates, each corresponding to one of the frequency values in the bandwidth $\mathcal{B} = \{\nu_1, \nu_2, \ldots \nu_{N_{\mathcal{B}}}\}$.

The signal is assumed to be deterministic, which means that each value of the parameter vector $\boldsymbol{\theta}$ corresponds to a single frequency function $\nu \mapsto s_{\boldsymbol{\theta}}(\nu)$.

Looking at the noise contributions in the frequency domain $n(\nu)$, we assume that the noise components after the Fourier transform $n(\nu)$ are centered, circularly symmetric complex Gaussian random variables which are independent across frequencies. The Gaussianity is true if the noise is naturally Gaussian, but it is also asymptotically true for other types of noise distributions. The asymptotic independence for any pair of frequencies is reported in [9].

We assume that the noise spectral power density in the bandwidth is known and defined by $\forall \nu \in \mathcal{B} \quad \gamma_0(\nu) = E(n(\nu)n^*(\nu))$. The hypothesis that the Gaussian complex noise is circularly symmetric means that $E(n(\nu)n(\nu)) = 0$.

When assuming all previous hypotheses, the law of the total vector of the observations $\boldsymbol{x} = (x(\nu))_{\nu \in \mathcal{B}}$ can be expressed as follows:

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{\nu \in \mathcal{B}} \frac{1}{\pi \gamma_0(\nu)} \exp\left(- \frac{|x(\nu) - s_{\boldsymbol{\theta}}(\nu)|^2}{\gamma_0(\nu)} \right) \tag{19}$$

The total vector of observation $\boldsymbol{x}$ has a total number $N_{\mathcal{B}}$ of complex components and can be rewritten as a real vector $\boldsymbol{X}$ with $2N_{\mathcal{B}}$ real components. With this notation, Equation (19) is similar to the normal multivariate distribution expressed in [6] and in Equation (5):

$$p(\boldsymbol{X}|\boldsymbol{\theta}) = \frac{1}{2^{N_{\mathcal{B}}} \pi^{N_{\mathcal{B}}} \sqrt{\det(\boldsymbol{\Sigma})}} \exp\left(- \frac{(\boldsymbol{X} - \boldsymbol{\mu}_{\boldsymbol{\theta}})^T {\boldsymbol{\Sigma}_{\mathbf{0}}}^{-1} (\boldsymbol{X} - \boldsymbol{\mu}_{\boldsymbol{\theta}})}{2} \right) \tag{20}$$

where the following definitions apply:

- $N_{\mathcal{B}}$ is the number of frequencies ν in $\mathcal{B}$;
- $\boldsymbol{\Sigma_0} = diag\left(\sigma_0^2(n)\right)$ is a diagonal $2N_{\mathcal{B}} \times 2N_{\mathcal{B}}$ matrix such that

$$\forall k = 1, \ldots N_{\mathcal{B}} \quad \sigma_0^2(2k-1) = \sigma_0^2(2k) = \tfrac{\gamma_0(\nu_k)}{2},$$

- $\boldsymbol{X}$ is a $2N_{\mathcal{B}}$ vector, with $\forall k = 1, \ldots N_{\mathcal{B}} \;\; X^{2k-1} = Re\{x(k)\}$ and $X^{2k} = Im\{x(k)\}$;
- $\boldsymbol{\mu_\theta}$ is a $2N_{\mathcal{B}}$ vector, with $\forall k = 1, \ldots N_{\mathcal{B}} \;\; \mu_{\boldsymbol{\theta}}^{2k-1} = Re\{s_{\boldsymbol{\theta}}(k)\}$ and $\mu_{\boldsymbol{\theta}}^{2k} = Im\{s_{\boldsymbol{\theta}}(k)\}$.

4.2. Signal Dependence on Parameters and the Manifolds L2$(\mathcal{B}, \boldsymbol{\theta})$

Equation (19) (or equivalently Equation (20)) defines a family of probability distributions S_Θ when the parameter $\boldsymbol{\theta}$ varies in a predefined set Θ:

$$S_\Theta = \{p_{\boldsymbol{\theta}} = p(\boldsymbol{x}|\boldsymbol{\theta}) \mid \boldsymbol{\theta} \in \Theta\} \tag{21}$$

With additional regularity conditions, S_Θ can be considered a statistical manifold (see [9,12], Section 2.1 for more details).

As a consequence, the geometry of the manifold lies in Equation (19) (or in Equation (20)) and in the definition of the parameter set Θ. We call such a manifold manifold $L2(\mathcal{B}, \boldsymbol{\theta})$. As $\boldsymbol{\theta}$ has N components, with $N \leq 2N_{\mathcal{B}}$, N is the dimension of the manifold $L2(\mathcal{B}, \boldsymbol{\theta})$.

For all ν in $\mathcal{B}$, the signal $s_{\boldsymbol{\theta}}(\nu)$ is a complex value, which can be represented by a modulus (or magnitude) $\rho_{\boldsymbol{\theta}}(\nu)$ and a phase $\psi_{\boldsymbol{\theta}}(\nu)$:

$$s_{\boldsymbol{\theta}}(\nu) = \rho_{\boldsymbol{\theta}}(\nu) \cdot \exp(\imath\psi_{\boldsymbol{\theta}}(\nu)) \tag{22}$$

with $\psi_{\boldsymbol{\theta}}(\nu) \in]-\pi, \pi]$ and $\rho_{\boldsymbol{\theta}}(\nu) \in]0, +\infty[$.

As explained in [12], the parameters $\boldsymbol{\theta}$ of the signal can be split into two parts: $\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\phi} \\ \boldsymbol{\varphi} \end{pmatrix}$. One part, $\boldsymbol{\phi}$, with P components, is related to the magnitude $\rho_{\boldsymbol{\phi}}(\nu)$ of the signal, and the other part, $\boldsymbol{\varphi}$, with $N - P$ components, is related to the phase $\psi_{\boldsymbol{\varphi}}(\nu)$ of the signal:

$$\forall \nu \in \mathcal{B} \quad s_{\boldsymbol{\theta}}(\nu) = \rho_{\boldsymbol{\phi}}(\nu) \cdot \exp\left(\imath\psi_{\boldsymbol{\varphi}}(\nu)\right) \tag{23}$$

In the following, any manifold $L2(\mathcal{B}, \boldsymbol{\theta})$ is described by Equation (19) completed with Equation (23), which refers to the case where $\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\phi} \\ \boldsymbol{\varphi} \end{pmatrix}$.

As the magnitudes and phases can be determined separately from the $s_{\boldsymbol{\theta}}(\nu)$, we assume that the numbers of coordinates of the parameter vectors $\boldsymbol{\theta}$ verify the following properties:

$$\begin{cases} P \leq N_{\mathcal{B}} \\ N - P \leq N_{\mathcal{B}} \end{cases} \tag{24}$$

In a manifold $L2(\mathcal{B}, \boldsymbol{\theta})$, the expression of the distribution is given by

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{\nu \in \mathcal{B}} \frac{1}{\pi\gamma_0(\nu)} \exp\left(- \frac{\left| x(\nu) - \rho_{\boldsymbol{\phi}}(\nu) \cdot \exp\left(\imath\psi_{\boldsymbol{\varphi}}(\nu)\right) \right|^2}{\gamma_0(\nu)} \right) \tag{25}$$

More details regarding the regularity conditions and dependence of the phase $\psi_{\boldsymbol{\varphi}}(\nu)$ on the frequency ν are available in [12].

5. General Expressions of the Fisher–Rao Distances on the Manifolds $L2(\mathcal{B}, \boldsymbol{\theta})$

5.1. Geodesic Equations on the Manifolds $L2(\mathcal{B}, \theta)$

To compute the geodesic equations on a manifold $L2(\mathcal{B}, \theta)$, we use the method described in Section 2 and Equations (1)–(3) and (25). We obtained the following theorem already expressed in [10–12].

Theorem 1 (Linearly dependant gradient (LDG) theorem)**.** *The (Fisher-)geodesic equations in the submanifold $L2(\mathcal{B}, \theta)$ reduce to the following two vectorial equations:*

$$\begin{cases} \sum_{\nu \in \mathcal{B}} \dfrac{2(\rho_{\boldsymbol{\theta}}(\nu))^2}{\gamma_0(\nu)} \left(\frac{d^2 \ln \rho_{\boldsymbol{\theta}}}{d\varsigma^2}(\nu) + \left(\frac{d \ln \rho_{\boldsymbol{\theta}}}{d\varsigma}(\nu) \right)^2 - \left(\frac{d\psi_{\boldsymbol{\theta}}}{d\varsigma}(\nu) \right)^2 \right) \nabla_{\boldsymbol{\theta}} \ln \rho_{\boldsymbol{\theta}}(\nu) = [\mathbf{0}]_N \\ \sum_{\nu \in \mathcal{B}} \dfrac{2(\rho_{\boldsymbol{\theta}}(\nu))^2}{\gamma_0(\nu)} \left(\frac{d^2 \psi_{\boldsymbol{\theta}}}{d\varsigma^2}(\nu) + 2 \frac{d \ln \rho_{\boldsymbol{\theta}}}{d\varsigma}(\nu) \frac{d\psi_{\boldsymbol{\theta}}}{d\varsigma}(\nu) \right) \nabla_{\boldsymbol{\theta}} \psi_{\boldsymbol{\theta}}(\nu) = [\mathbf{0}]_N \end{cases} \tag{26}$$

Proof. The method explained in Section 2.2 consists of three steps:

- Calculate the Fisher information matrix $\boldsymbol{G}$.
- Calculate the partial derivatives $\partial_i \boldsymbol{G}$ and the Christoffel symbols $\Gamma_{ij,k}$.
- Replace the Christoffel symbols to express the geodesic equation.

The calculations are detailed in [12].
The first step consists of calculating the Fisher information matrix.
It can be shown that the Fisher matrix is a block diagonal matrix:

$$\boldsymbol{G} = [\boldsymbol{g_{ij}}] = \begin{pmatrix} [g_{uv}] & [0] \\ [0] & [g_{qr}] \end{pmatrix}$$

The second step consists of calculating the Christoffel symbols, based on the partial derivative of the Fisher matrix coefficient with respect to the parameters. In the third step of the proof, we replace the values of $\Gamma_{ij,m}$ in the geodesic equation (Equation (3)) in order to obtain geodesic equations expressed on the manifolds $L2(\mathcal{B}, \theta)$.

Finally, we can write the geodesic equations as follows:

$$\begin{cases} \sum_{\nu \in \mathcal{B}} \frac{2}{\gamma_0(\nu)} \left(\frac{d^2 \rho_{\boldsymbol{\phi}}}{d\varsigma^2}(\nu) - \rho_{\boldsymbol{\phi}}(\nu) \left(\frac{d\psi_{\boldsymbol{\varphi}}}{d\varsigma}(\nu) \right)^2 \right) \nabla_{\boldsymbol{\phi}} \rho_{\boldsymbol{\phi}}(\nu) = [\mathbf{0}]_{\boldsymbol{P}} \\ \sum_{\nu \in \mathcal{B}} \frac{2}{\gamma_0(\nu)} \left(\frac{d^2 \psi_{\boldsymbol{\varphi}}}{d\varsigma^2}(\nu) + 2\rho_{\boldsymbol{\phi}}(\nu) \frac{d\rho_{\boldsymbol{\phi}}}{d\varsigma}(\nu) \frac{d\psi_{\boldsymbol{\varphi}}}{d\varsigma}(\nu) \right) \nabla_{\boldsymbol{\varphi}} \psi_{\boldsymbol{\varphi}}(\nu) = [\mathbf{0}]_{\boldsymbol{N-P}} \end{cases} \tag{27}$$

Equation (27) can easily be transformed into Equation (26).

□

5.2. Fisher–Rao Distances on the Manifolds $L2(\mathcal{B}, \theta)$

Theorem 2. *With $\tilde{\rho}_{\boldsymbol{\theta}}$ and $\tilde{\psi}_{\boldsymbol{\theta}}$ as solutions of the geodesic equations, the Fisher–Rao distance on the submanifold $L2(\mathcal{B}, \theta)$ can be calculated as the following integral:*

$$d_{L2,\mathcal{B},\boldsymbol{\theta}}(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) = \int_0^1 \sqrt{\sum_{\nu \in \mathcal{B}} \frac{2(\tilde{\rho}_{\boldsymbol{\theta}}(\nu))^2}{\gamma_0(\nu)} \left(\left(\frac{d \ln \tilde{\rho}_{\boldsymbol{\theta}}}{d\varsigma}(\nu) \right)^2 + \left(\frac{d\tilde{\psi}_{\boldsymbol{\theta}}}{d\varsigma}(\nu) \right)^2 \right)} d\varsigma \tag{28}$$

Proof. Detailed calculations are available in [12].

We note that $\tilde{\phi}^u$ (or $\tilde{\phi}^v$) and $\tilde{\varphi}^q$ (or $\tilde{\varphi}^r$) are the solutions of the geodesic equation (Equation (27)). They are expressed as functions of ς. From Equation (4), taking into account the block diagonal structure of the Fisher information matrix as demonstrated in Section 5.1, we derive the Fisher–Rao distance from the following integral:

$$d_{L2,\mathcal{B},\boldsymbol{\theta}}(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) = \int_0^1 \sqrt{g_{uv} \frac{d\tilde{\phi}^u}{d\varsigma} \frac{d\tilde{\phi}^v}{d\varsigma} + g_{qr} \frac{d\tilde{\varphi}^q}{d\varsigma} \frac{d\tilde{\varphi}^r}{d\varsigma}} d\varsigma \tag{29}$$

We replace the g_{uv} and g_{qr} with their expressions and we use the following equalities:

$$\partial_u \rho_{\boldsymbol{\phi}}(\nu) \cdot \frac{d\phi^u}{d\varsigma} = \frac{\partial \rho_{\boldsymbol{\phi}}}{\partial \phi^u}(\nu) \frac{d\phi^u}{d\varsigma} = \frac{d\rho_{\boldsymbol{\phi}}}{d\varsigma}(\nu)$$

$$\partial_r \psi_{\boldsymbol{\varphi}}(\nu) \cdot \frac{d\varphi^r}{d\varsigma} = \frac{\partial \psi_{\boldsymbol{\varphi}}}{\partial \varphi^r}(\nu) \frac{d\varphi^r}{d\varsigma} = \frac{d\psi_{\boldsymbol{\varphi}}}{d\varsigma}(\nu)$$

We then obtain

$$d_{L2,\mathcal{B},\boldsymbol{\theta}}(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) = \int_0^1 \sqrt{\sum_{\nu \in \mathcal{B}} \frac{2}{\gamma_0(\nu)} \left(\left(\frac{d\tilde{\rho}_{\boldsymbol{\phi}}}{d\varsigma}(\nu) \right)^2 + \left(\tilde{\rho}_{\boldsymbol{\phi}}(\nu) \frac{d\tilde{\psi}_{\boldsymbol{\varphi}}}{d\varsigma}(\nu) \right)^2 \right)} d\varsigma \tag{30}$$

By definition, $\tilde{\rho}_{\boldsymbol{\theta}} = \tilde{\rho}_{\boldsymbol{\phi}}$, and $\tilde{\psi}_{\boldsymbol{\theta}} = \tilde{\psi}_{\boldsymbol{\varphi}}$.

Due to the derivative properties, we have $\frac{1}{\tilde{\rho}_{\boldsymbol{\theta}}(\nu)} \frac{d\tilde{\rho}_{\boldsymbol{\theta}}}{d\varsigma}(\nu) = \frac{d \ln \tilde{\rho}_{\boldsymbol{\theta}}}{d\varsigma}(\nu)$.

Using these equalities, Equation (30) is easily transformed into Equation (28).

□

To compute the expression in Equation (28) and give a closed-form expression of the Fisher–Rao distance, it is necessary to solve Equation (26). This can only be accomplished

when the parametric expressions of a magnitude $\rho_{\boldsymbol{\theta}}(\nu)$ and phase $\psi_{\boldsymbol{\theta}}(\nu)$ are expressed in closed forms, i.e., when the model is particularized. This is performed with the specific manifold in the next section.

6. General Expressions of the Fisher–Rao Distances on the Manifold $L2(\mathcal{B})$

6.1. The Finite-Energy Signals and the Manifold $L2(\mathcal{B})$

Here, we look at a specific manifold corresponding to the situation for which, in Equation (20), the parameters are the components of the vector $\boldsymbol{\mu}_{\boldsymbol{\theta}}$. This means that $\boldsymbol{\mu}_{\boldsymbol{\theta}} = \boldsymbol{\theta}$.

In this case, $\boldsymbol{\theta}$ is a vector with $2N_{\mathcal{B}}$ components, and $\forall k = 1, \dots 2N_{\mathcal{B}} \quad \theta^k = \mu^k$. In other words, $\Theta = \mathbb{R}^{2N_{\mathcal{B}}}$. This means that all the signal components $s_{\boldsymbol{\theta}}(\nu)$ vary freely in $\mathbb{C}$.

Variations in all signal components determine a manifold, which we call the manifold $L2(\mathcal{B})$. This corresponds to all signals with finite energy observed in the frequency bandwidth $\mathcal{B}$. The dimension of the manifold $L2(\mathcal{B})$ is $2N_{\mathcal{B}}$. The manifold $L2(\mathcal{B})$ corresponds exactly to the manifold described by $2N_{\mathcal{B}}$ multivariate normal distributions with the same diagonal covariance matrix $\boldsymbol{\Sigma_0}$ and different mean values.

Any manifold $L2(\mathcal{B}, \boldsymbol{\theta})$ is a sub-manifold of the manifold $L2(\mathcal{B})$.

6.2. Geodesic Equations on the Manifold $L2(\mathcal{B})$

The geodesic equation on the manifold $L2(\mathcal{B})$ is derived directly from Equations (6) and (20) when the coefficients of the matrix $\boldsymbol{\Sigma_0}$ are not part of the coordinates. In this case, the second line of Equation (6) does not exist, and the geodesic equation becomes

$$\left\{ \frac{d^2 \mu_{\boldsymbol{\theta}}}{d\varsigma^2} = [\mathbf{0}]_N \right. \tag{31}$$

The solution between $\boldsymbol{\mu}_{\boldsymbol{\theta_1}}$ and $\boldsymbol{\mu}_{\boldsymbol{\theta_2}}$ is an affine function of ς [6]:

$$\forall \varsigma \in [0,1] \quad \tilde{\boldsymbol{\mu}}_{\boldsymbol{\theta}}(\varsigma) = \boldsymbol{\mu}_{\boldsymbol{\theta_1}} + \varsigma(\boldsymbol{\mu}_{\boldsymbol{\theta_2}} - \boldsymbol{\mu}_{\boldsymbol{\theta_1}}) \tag{32}$$

6.3. Fisher–Rao Distance on the Manifold $L2(\mathcal{B})$

From Equation (20) and the definition of the Fisher matrix in Equation (1), we can easily derive the expression of the Fisher information matrix $\boldsymbol{G} = \left[g_{ij}\right]_{N \times N}$ in the manifold $L2(\mathcal{B})$:

$$g_{ij} = \frac{\partial \boldsymbol{\mu}_{\boldsymbol{\theta}}^{\boldsymbol{T}}}{\partial \theta^i} {\boldsymbol{\Sigma_0}}^{-1} \frac{\partial \boldsymbol{\mu}_{\boldsymbol{\theta}}}{\partial \theta^j} \tag{33}$$

If we report this expression of $\boldsymbol{G}$ in Equation (4), then we get the expression of the Fisher–Rao distance on the manifold $L2(\mathcal{B})$:

$$d_{L2,\mathcal{B}}(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) = \int_0^1 \sqrt{\frac{d\tilde{\boldsymbol{\mu}}_{\boldsymbol{\theta}}^{\boldsymbol{T}}}{d\varsigma} {\boldsymbol{\Sigma_0}}^{-1} \frac{d\tilde{\boldsymbol{\mu}}_{\boldsymbol{\theta}}}{d\varsigma}} d\varsigma \tag{34}$$

It is known from [6] that in the $2N_{\mathcal{B}}$-dimensional manifold composed of multivariate normal distributions with a common covariance matrix $\boldsymbol{\Sigma_0}$, the Fisher–Rao distance between two distributions parametrised by $\boldsymbol{\mu}_{\boldsymbol{\theta_1}}$ and $\boldsymbol{\mu}_{\boldsymbol{\theta_2}}$ is equal to the Mahalanobis distance. This can be easily verified here. When changing $\tilde{\boldsymbol{\mu}}$ in Equation (34) by its value from Equation (32), we have

$$d_{L2,\mathcal{B}}(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) = \sqrt{\left(\boldsymbol{\mu}_{\boldsymbol{\theta_2}} - \boldsymbol{\mu}_{\boldsymbol{\theta_1}}\right)^T {\boldsymbol{\Sigma_0}}^{-1} \left(\boldsymbol{\mu}_{\boldsymbol{\theta_2}} - \boldsymbol{\mu}_{\boldsymbol{\theta_1}}\right)} \tag{35}$$

With the notation of Equations (20) and (22), this Fisher–Rao distance on the manifold $L2(\mathcal{B})$ can be rewritten as follows:

$$d_{L2,\mathcal{B}}(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) = \sqrt{\sum_{\nu \in \mathcal{B}} \frac{2}{\gamma_0(\nu)} \left((\rho_{\boldsymbol{\theta_2}}(\nu))^2 + (\rho_{\boldsymbol{\theta_1}}(\nu))^2 - 2\rho_{\boldsymbol{\theta_2}}(\nu) \cdot \rho_{\boldsymbol{\theta_1}}(\nu) \cdot \cos(\Delta\psi(\nu)) \right)} \tag{36}$$

with

$$\Delta\psi(\nu) = \psi_{\boldsymbol{\theta_2}}(\nu) - \psi_{\boldsymbol{\theta_1}}(\nu) \tag{37}$$

Proof. The $N_{\mathcal{B}}$ frequencies in $\mathcal{B}$ are $\nu_1, \nu_2, \ldots, \nu_{N_{\mathcal{B}}}$.

$\boldsymbol{\Sigma_0}$ is a diagonal $2N_{\mathcal{B}} \times 2N_{\mathcal{B}}$ real matrix. Each value ν_i in the frequency band is represented by two diagonal values, $(\sigma_0(2i-1))^2$ and $(\sigma_0(2i))^2$, and we have by definition

$$(\sigma_0(2i-1))^2 = (\sigma_0(2i))^2 = \frac{\gamma(\nu_i)}{2}$$

The vector $\boldsymbol{\mu_\theta}$ has $2N_{\mathcal{B}}$ components arranged in pairs:

$$\begin{cases} \mu_\theta^{2i-1} = Re\{s_{\boldsymbol{\theta}}(\nu_i)\} = \rho_{\boldsymbol{\theta}}(\nu_i) \cdot \cos(\psi_{\boldsymbol{\theta}}(\nu_i)) \\ \mu_\theta^{2i} = Im\{s_{\boldsymbol{\theta}}(\nu_i)\} = \rho_{\boldsymbol{\theta}}(\nu_i) \cdot \sin(\psi_{\boldsymbol{\theta}}(\nu_i)) \end{cases}$$

At frequency ν_i, the contribution of the components indexed by $2i-1$ and $2i$ is as follows:

$$\begin{aligned}
&\frac{\left(\mu_2^{2i-1} - \mu_1^{2i-1}\right)^2}{(\sigma_0(2i-1))^2} + \frac{\left(\mu_2^{2i} - \mu_1^{2i}\right)^2}{(\sigma_0(2i))^2} \\
&= \frac{2}{\gamma(\nu_i)} \cdot \left(\left(\rho_{\boldsymbol{\theta_2}}(\nu_i)\cos\left(\psi_{\boldsymbol{\theta_2}}(\nu_i)\right) - \rho_{\boldsymbol{\theta_1}}(\nu_i)\cos\left(\psi_{\boldsymbol{\theta_1}}(\nu_i)\right)\right)^2 \right) \\
&+ \frac{2}{\gamma(\nu_i)} \cdot \left(\left(\rho_{\boldsymbol{\theta_2}}(\nu_i)\sin\left(\psi_{\boldsymbol{\theta_2}}(\nu_i)\right) - \rho_{\boldsymbol{\theta_1}}(\nu_i)\sin\left(\psi_{\boldsymbol{\theta_1}}(\nu_i)\right)\right)^2 \right) \\
&= \frac{2}{\gamma(\nu_i)} \cdot \left(\left(\rho_{\boldsymbol{\theta_2}}(\nu_i)\right)^2 + \left(\rho_{\boldsymbol{\theta_1}}(\nu_i)\right)^2 \right) \\
&- \frac{2}{\gamma(\nu_i)} \cdot 2\rho_2(\nu_i)\rho_1(\nu_i) \cdot \left(\cos(\psi_2(\nu_i))\cos(\psi_1(\nu_i)) + \sin\left(\psi_{\boldsymbol{\theta_2}}(\nu_i)\right)\sin\left(\psi_{\boldsymbol{\theta_1}}(\nu_i)\right)\right) \\
&= \frac{2}{\gamma(\nu_i)} \cdot \left(\left(\rho_{\boldsymbol{\theta_2}}(\nu_i)\right)^2 + \left(\rho_{\boldsymbol{\theta_1}}(\nu_i)\right)^2 \right) - \frac{2}{\gamma(\nu_i)} \cdot 2\rho_{\boldsymbol{\theta_2}}(\nu_i)\rho_{\boldsymbol{\theta_1}}(\nu_i)\cos\left(\psi_{\boldsymbol{\theta_2}}(\nu_i) - \psi_{\boldsymbol{\theta_1}}(\nu_i)\right) \\
&= \frac{2}{\gamma(\nu_i)} \cdot \left(\left(\rho_{\boldsymbol{\theta_2}}(\nu_i)\right)^2 + \left(\rho_{\boldsymbol{\theta_1}}(\nu_i)\right)^2 - 2\rho_{\boldsymbol{\theta_2}}(\nu_i)\rho_{\boldsymbol{\theta_1}}(\nu_i)\cos\left(\psi_{\boldsymbol{\theta_2}}(\nu_i) - \psi_{\boldsymbol{\theta_1}}(\nu_i)\right) \right)
\end{aligned}$$

□

6.4. Expression with the Spectral Distance Function

We define the spectral distance function as follows:

$$\forall \nu \in \mathcal{B} \quad f_{\boldsymbol{\theta_1},\boldsymbol{\theta_2}}(\nu) = \sqrt{\frac{2}{\gamma_0(\nu)} \left((\rho_{\boldsymbol{\theta_2}}(\nu))^2 + (\rho_{\boldsymbol{\theta_1}}(\nu))^2 - 2\rho_{\boldsymbol{\theta_2}}(\nu) \cdot \rho_{\boldsymbol{\theta_1}}(\nu) \cdot \cos(\Delta\psi(\nu)) \right)} \tag{38}$$

This function describes the contribution of each spectral component of the signal observations $\boldsymbol{\theta_1}$ and $\boldsymbol{\theta_2}$ to their Fisher–Rao distance.

According to Equation (36), the Fisher–Rao distance can be expressed from the spectral distance function by

$$d_{L2,\mathcal{B}}(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) = \sqrt{\sum_{\nu \in \mathcal{B}} \left(f_{\boldsymbol{\theta_1},\boldsymbol{\theta_2}}(\nu)\right)^2} \tag{39}$$

When the signals have a continuous spectrum, this expression takes the form

$$d_{L2,\mathcal{B}}(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) = \sqrt{\frac{1}{B}\int_{\nu\in\breve{B}} (f_{\boldsymbol{\theta_1},\boldsymbol{\theta_2}}(\nu))^2 d\nu} \tag{40}$$

7. Evaluation of the Fisher–Rao Distances

7.1. Comparison of the Fisher–Rao Distances on the Manifold $L2(\mathcal{B})$ and on Any Sub-Manifold $L2(\mathcal{B}, \boldsymbol{\theta})$

The N-dimensional sub-manifolds $L2(\mathcal{B}, \boldsymbol{\theta})$ are included in the $2N_{\mathcal{B}}$-dimensional manifold $L2(\mathcal{B})$ composed by multivariate normal distributions with a common diagonal covariance matrix $\boldsymbol{\Sigma_0}$. The metric on any sub-manifold $L2(B, \boldsymbol{\theta})$ is the induced metric from the ambient manifold $L2(B)$. When two endpoints are connected with a geodesic path on the sub-manifold, they are also connected with the same path on the ambient manifold. However, this path on the ambient manifold is not always the minimum (geodesic) path for the ambient manifold. As a consequence, we expect that the Fisher–Rao distance $d_{L2,\mathcal{B},\boldsymbol{\theta}}(\boldsymbol{\theta_1}, \boldsymbol{\theta_2})$ measured on any sub-manifold $L2(\mathcal{B}, \boldsymbol{\theta})$ is greater than or equal to the Fisher–Rao distance measured on the manifold $L2(\mathcal{B})$:

$$d_{L2,\mathcal{B},\boldsymbol{\theta}}(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) \geq d_{L2,\mathcal{B}}(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) \tag{41}$$

When both quantities are always equal, the sub-manifold $L2(\mathcal{B}, \boldsymbol{\theta})$ is said to be fully geodesic on the manifold $L2(\mathcal{B})$. As a consequence of Equation (41), the value of the Fisher–Rao distance between two finite-energy signals on the manifold $L2(\mathcal{B})$ is a minimum bound for any Fisher–Rao distance between both finite-energy signals calculated on any sub-manifold $L2(\mathcal{B}, \boldsymbol{\theta})$.

7.2. Asymptotic Behavior of the Fisher–Rao Distances

Theorem 3. *The Fisher-Rao distances belong to the following interval:*

$$\sqrt{\sum_{\nu\in\mathcal{B}} \frac{2(\rho_{\boldsymbol{\theta_2}}(\nu) - \rho_{\boldsymbol{\theta_1}}(\nu))^2}{\gamma_0(\nu)}} \leq d_{L2,\mathcal{B}}(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) \leq \sqrt{\sum_{\nu\in\mathcal{B}} \frac{2(\rho_{\boldsymbol{\theta_2}}(\nu) + \rho_{\boldsymbol{\theta_1}}(\nu))^2}{\gamma_0(\nu)}} \tag{42}$$

Proof. The value of $\cos(\Delta\psi(\nu))$ in Equation (36) varies between -1 and $+1$. □

Theorem 4. *When the phase differences $\Delta\psi(\nu)$ are small, distances tend toward the minimum values:*

$$d_{L2,\mathcal{B}}(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) \underset{\Delta\psi(\nu)\to 0}{\longrightarrow} \sqrt{\sum_{\nu\in\mathcal{B}} \frac{2(\rho_{\boldsymbol{\theta_2}}(\nu) - \rho_{\boldsymbol{\theta_1}}(\nu))^2}{\gamma_0(\nu)}} \tag{43}$$

Proof. The value of $\cos(\Delta\psi(\nu))$ tends toward one in Equation (36). □

Theorem 5. *When the signal at $\boldsymbol{\theta_1}$ and the signal $\boldsymbol{\theta_2}$ have separate bandwidths, or when the phase differences vary significantly across the bandwidth, the Fisher–Rao distance becomes*

$$d_{L2,\mathcal{B}}(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) \approx \sqrt{SNR_1 + SNR_2} \tag{44}$$

With:

$$SNR_1 = \sum_{\nu \in \mathcal{B}} \frac{2\left(\rho_{\boldsymbol{\theta}_1}(\nu)\right)^2}{\gamma_0(\nu)} \tag{45}$$

$$SNR_2 = \sum_{\nu \in \mathcal{B}} \frac{2\left(\rho_{\boldsymbol{\theta}_2}(\nu)\right)^2}{\gamma_0(\nu)} \tag{46}$$

Proof. The value of $\sum_{\nu \in \mathcal{B}} \frac{2\rho_{\boldsymbol{\theta}_1}(\nu)\rho_{\boldsymbol{\theta}_2}(\nu)}{\gamma_0(\nu)} \cos(\Delta\psi(\nu))$ tends toward zero in Equation (36). □

In this expressions, SNR_1 and SNR_2 represent the signal-to-noise ratios at $\boldsymbol{\theta_1}$ and $\boldsymbol{\theta_2}$, respectively.

7.3. Numerical Application

We see with Equation (36) that the Fisher–Rao distance is proportional to the square root of the signal-to-noise ratio (SNR), meaning that when the signal-to-noise ratios SNR_1 and SNR_2 are multiplied by a, the Fisher–Rao distance is multiplied by $\sqrt{a}$ (if the phase differences remain the same).

Furthermore, we observe that the Fisher–Rao distance is symmetric under the permutations of $\boldsymbol{\theta_1}$ and $\boldsymbol{\theta_2}$.

As a consequence, we can explore the distance at a reference signal-to-noise ratio $SNR_1 = 1$ with the hypothesis that $SNR_2 \leq 1$.

We assume that the noise is Gaussian white with a constant spectral density $\gamma_0(\nu)$.

We examine several signal configurations and their respective distances.

We compare two linear frequency-modulated (LFM) pulses with the same time duration but different bandwidths. The pulses were weighted by a Hann window. The parametric expressions of the signals are as follows:

$$\forall t \in [0, T] \quad s_r(t) = \left(0.5 - 0.5 * \cos(2\pi \frac{t}{T})\right) \cdot \sin\left(\eta + 2\pi\left((f_0 - \epsilon\frac{B_{FM}}{2})t + \epsilon\frac{B_{FM}}{2T}t^2\right)\right) \tag{47}$$

In Equation (47), the meanings of the parameters are as follows:

- f_0 is the middle frequency of the modulation;
- B_{FM} is the frequency bandwidth of the modulation;
- η is a phase parameter between 0 and 2π which introduces a phase shift;
- ϵ takes the values $+1$ and -1, indicating whether the slope of the LFM frequency variation is upward or downward, respectively;
- T is the time length of the impulsion.

After sampling at frequency F_s, as we choose T to also be the time duration of the Fourier transform, we obtain $N_T = F_s T$ samples, and following the DFT, the observations become

$$\forall \nu \in \mathcal{B} \quad s_{\boldsymbol{\theta}}(\nu) = \sum_{n=1}^{N_T} s_r\left(\frac{n-1}{F_s}\right) \exp\left(-2\imath\pi\frac{\nu}{F_s}(n-1)\right) \tag{48}$$

with the following definitions:

- $\mathcal{B} = \{\nu_1, \nu_2, \ldots \nu_{N_\mathcal{B}}\}$;
- $\forall k = 1 \ldots N_\mathcal{B} \quad \nu_k = (k-1)\frac{F_s}{N_T}$;
- If N_T is even, then $N_\mathcal{B} = \frac{N_T}{2} + 1$, or if N_T is odd, then $N_\mathcal{B} = \frac{N_T - 1}{2} + 1$.

For the numerical applications, we choose the following:

- The time duration is $T = 1.000$ s;
- The sample frequency Fs is fixed to 1.024 Hz;

- This determines $N_T = 1024$ and $N_{\mathcal{B}} = 513$.

We analyzed various scenarios; each was characterized by a set of parameters and illustrated with figures presenting the calculation results:

- Scenario 1: Similar signals with different bandwidths.
  The parameters are described in Table 1. Both signals were similar (same phase and middle frequency and same slope of modulation) except for their respective bandwidths. The figure shows variation in the distance with respect to the bandwidth of signal $\boldsymbol{\theta_2}$.

Table 1. Values of the parameters for scenario 1.

Parameters	η	f_0	ϵ	B_{FM}
$\boldsymbol{\theta_1}$	0	0.25	+1	0.25
$\boldsymbol{\theta_2}$	0	0.25	+1	$0. \cdots 0.5$ [1]

[1] Variation between the values.

The results are shown in Figure 1. We observe that th

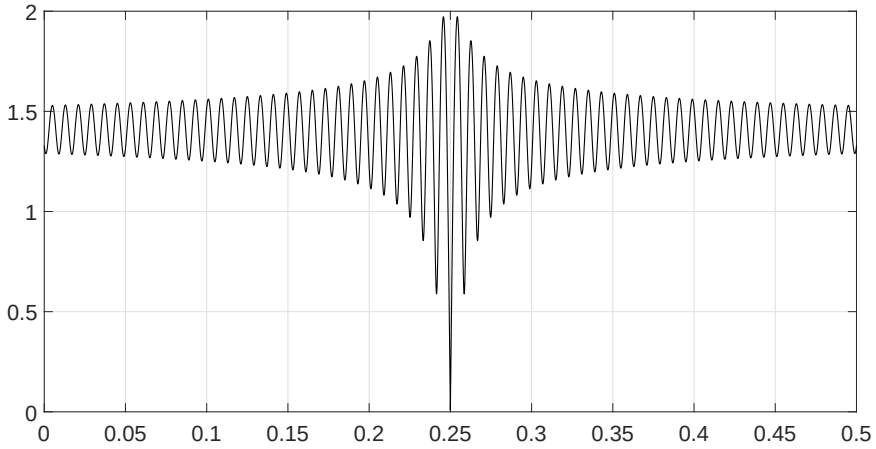


Figure 1. Value of the Fisher–Rao distance as a function of the bandwidth B_{FM} of the signal $\boldsymbol{\theta_2}$ for scenario 1.

- Scenario 2: Signals with different bandwidths and slopes of modulation.
  The parameters are described in Table 2. Both signals were similar (same phase and middle frequency) except for their respective bandwidths and their slopes of modulation. The figure shows variation in the distance with respect to the bandwidth of signal $\boldsymbol{\theta_2}$.
  The results are shown in Figure 2. We observe that the distance varied slightly around 1.4142 ($\sqrt{2}$). This is a situation close to Equation (44).

Table 2. Values of the parameters for scenario 2.

Parameters	η	f_0	ϵ	B_{FM}
$\boldsymbol{\theta_1}$	0	0.25	+1	0.25
$\boldsymbol{\theta_2}$	0	0.25	-1	$0. \cdots 0.5$ [1]

[1] Variation between the values.

- Scenario 3: Signals with opposite phases and different bandwidths.
  The parameters are described in Table 3. Both signals were similar (same phase and middle frequency and same slope of modulation) except for their respective bandwidths, and they had a phase difference of π. The figure shows variation in the distance with respect to the bandwidth of signal $\boldsymbol{\theta_2}$.

The results are shown in Figure 3. We observe that the distance varied between 0 and 2. This corresponds to the bounds of Equation (42). The maximum was reached when the two signals were identical with a phase shift of π (same bandwidth B_{FM}).

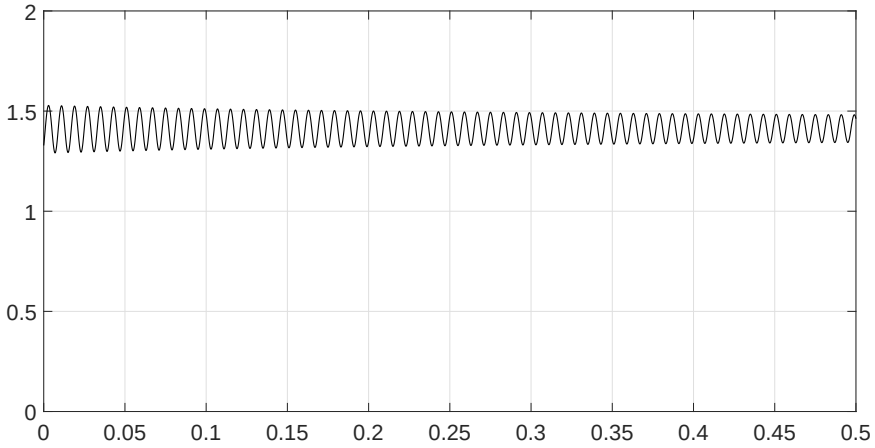


Figure 2. Value of the Fisher–Rao distance as a function of the bandwidth B_{FM} of the signal $\boldsymbol{\theta_2}$ for scenario 2.

Table 3. Values of the parameters for scenario 3.

Parameters	η	f_0	ϵ	B_{FM}
$\boldsymbol{\theta_1}$	0	0.25	+1	0.25
$\boldsymbol{\theta_2}$	π	0.25	+1	$0. \cdots 0.5$ [1]

[1] Variation between the values.

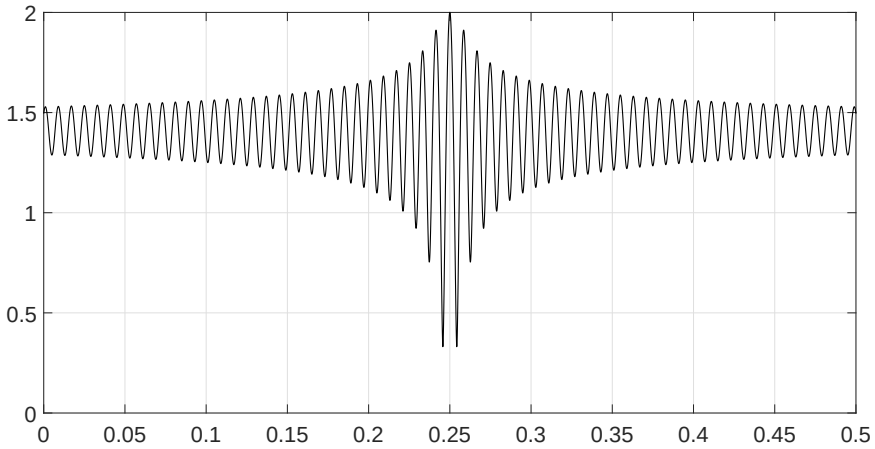


Figure 3. Value of the Fisher–Rao distance as a function of the bandwidth B_{FM} of the signal $\boldsymbol{\theta_2}$ for scenario 3.

- Scenario 4: Signals with different bandwidths and mid frequencies.
  The parameters are described in Table 4. The signals had different middle frequencies, but they had the same phase and the same slope of modulation. The bandwidth of the second signal varied. The figure shows variation in the distance with respect to the bandwidth of signal $\boldsymbol{\theta_2}$.

Table 4. Values of the parameters for scenario 4.

Parameters	η	f_0	ϵ	B_{FM}
$\boldsymbol{\theta_1}$	0	0.15	+1	0.25
$\boldsymbol{\theta_2}$	0	0.25	+1	$0. \cdots 0.5$ [1]

[1] Variation between the values.

The results are shown in Figure 4. We observe that the distance varied rather slightly around 1.4142 ($\sqrt{2}$) (less than 0.1% variation). This is a situation corresponding to Equation (44).

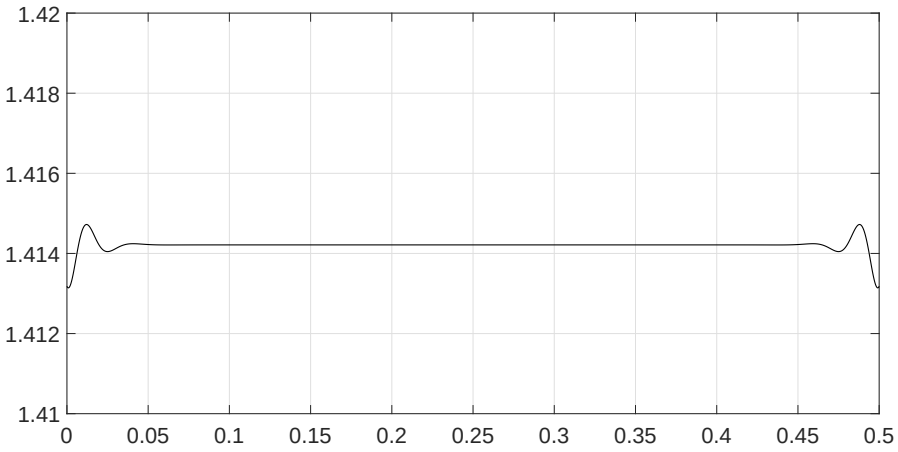


Figure 4. Value of the Fisher–Rao distance as a function of the bandwidth B_{FM} of the signal $\boldsymbol{\theta_2}$for scenario 4.

- Scenario 5: Analysis of the spectral distance function $f_{\boldsymbol{\theta_1},\boldsymbol{\theta_2}}(\nu)$ for two signals with different bandwidths.
  The parameters are described in Table 5. The signals had different bandwidths, but they had same phase, same middle frequency and same slope of modulation. We drew two figures: one representing the two signal spectra and the other representing the spectral distance between the two signals.
  The value of the Fisher–Rao distance between the two signals was $d_{L2,\mathcal{B}}(\boldsymbol{\theta_1},\boldsymbol{\theta_2}) \simeq 1.5455$. Figures 5 and 6 show the impact of the signals' frequencies on the distance. We observe that the spectral distance varied significantly within the common bandwidth.

Table 5. Values of the parameters for scenario 5.

Parameters	η	f_0	ϵ	B_{FM}
$\boldsymbol{\theta_1}$	0	0.25	+1	0.25
$\boldsymbol{\theta_2}$	0	0.25	+1	0.15

- Scenario 6: Analysis of the spectral distance $f_{\boldsymbol{\theta_1},\boldsymbol{\theta_2}}(\nu)$ for two signals with different middle frequencies.
  The parameters are described in Table 6. The signals had different middle frequencies, but they had the same phase, same bandwidth and same slope of modulation. We drew two figures: one representing the signal spectra and the other representing the spectral distance between the two signals.
  The value of the Fisher–Rao distance between the two signals was $d_{L2,\mathcal{B}}(\boldsymbol{\theta_1},\boldsymbol{\theta_2}) \simeq 1.4142$. Figures 7 and 8 show the impact of the signals' frequencies on the distance. We observe that the spectral distance varied significantly within the common bandwidth. This is a situation corresponding to Equation (44).

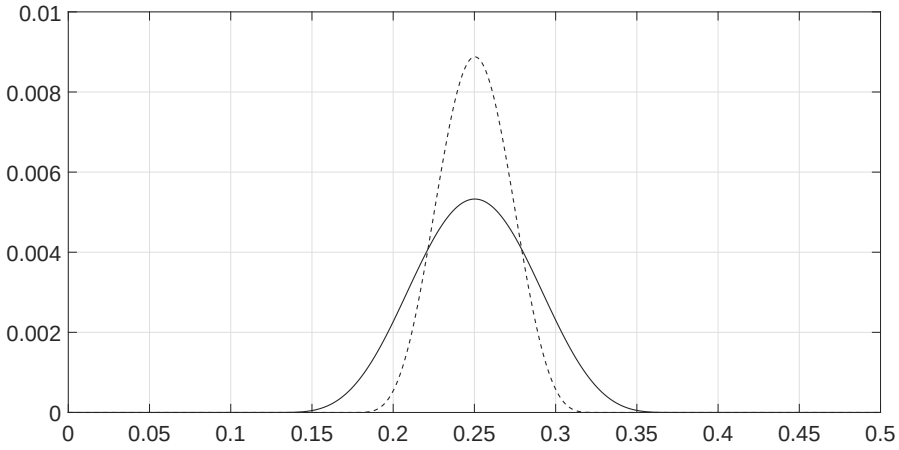


Figure 5. Values of the spectra $\rho_{\boldsymbol{\theta}_1}(\nu)$ (solid line) and $\rho_{\boldsymbol{\theta}_2}(\nu)$ (dashed line) for the two signals as a function of the frequency ν for scenario 5.

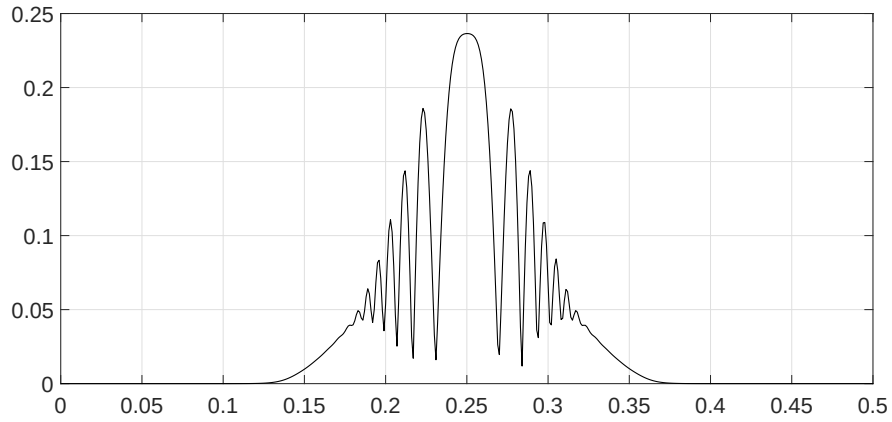


Figure 6. Values of the spectral distance $f_{\boldsymbol{\theta}_1,\boldsymbol{\theta}_2}(\nu)$ as a function of the frequency ν for scenario 5.

Table 6. Values of the parameters for scenario 6.

Parameters	η	f_0	ϵ	B_{FM}
$\boldsymbol{\theta_1}$	0	0.25	+1	0.25
$\boldsymbol{\theta_2}$	0	0.15	+1	0.25

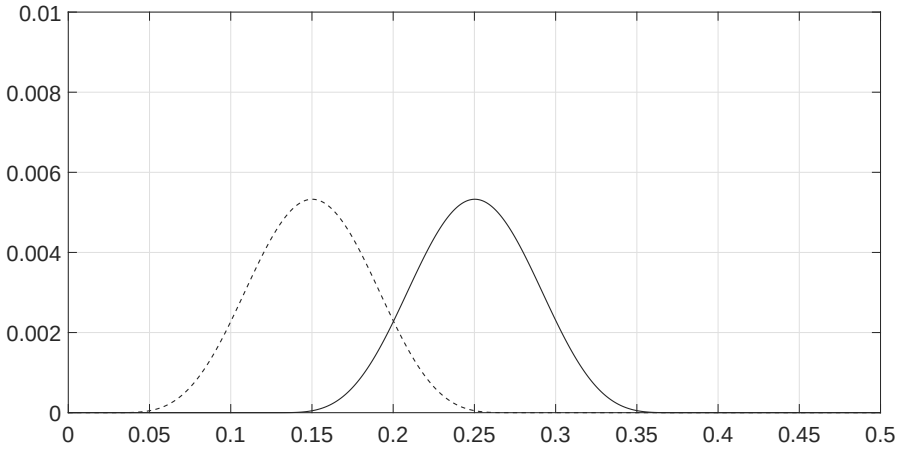


Figure 7. Values of the spectra $\rho_{\boldsymbol{\theta}_1}(\nu)$ (solid line) and $\rho_{\boldsymbol{\theta}_2}(\nu)$ (dashed line) for the two signals as a function of the frequency ν for scenario 6.

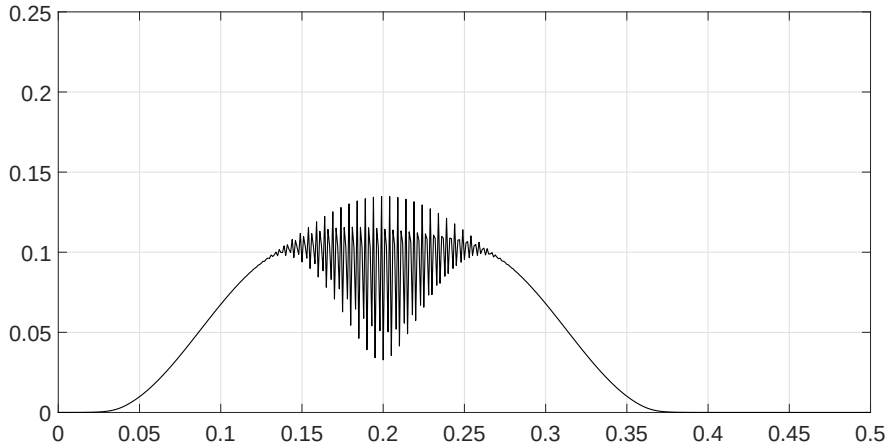


Figure 8. Values of the spectral distance $f_{\theta_1,\theta_2}(\nu)$ as a function of the frequency ν for scenario 6.

- Scenario 7: A toy problem to demonstrate the usefulness of using the Fisher–Rao distance for parameter estimation.
  We had three signals. They had the same middle frequencies, same bandwidth and same slope of modulation, but they had different phases. The problem consisted of estimating the optimal phase shift values for the three signals to facilitate their discrimination during observation.
  In order to optimally separate the signals, we wanted to ensure the maximum possible value of the Fisher–Rao distance between any two signals. To achieve this, we calculated all the Fisher–Rao distances between the signals, namely $d_{L2,\mathcal{B}}(\boldsymbol{\theta_1}, \boldsymbol{\theta_2})$, $d_{L2,\mathcal{B}}(\boldsymbol{\theta_1}, \boldsymbol{\theta_3})$ and $d_{L2,\mathcal{B}}(\boldsymbol{\theta_2}, \boldsymbol{\theta_3})$, and we determine the minimum distance:

$$d_{min} = \min(d_{L2,\mathcal{B}}(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}), d_{L2,\mathcal{B}}(\boldsymbol{\theta_1}, \boldsymbol{\theta_3}), d_{L2,\mathcal{B}}(\boldsymbol{\theta_2}, \boldsymbol{\theta_3})) \tag{49}$$

  For the signal $\boldsymbol{\theta_1}$, the phase shift $\eta(\boldsymbol{\theta_1})$ was arbitrarily set to zero. We looked for the respective values of $\eta(\boldsymbol{\theta_2})$ and $\eta(\boldsymbol{\theta_3})$, which determined the maximum value of d_{min}. The parameters of the scenario are described in Table 7.

Table 7. Values of the parameters for scenario 7.

Parameters	η	f_0	ϵ	B_{FM}
$\boldsymbol{\theta_1}$	0	0.25	+1	0.05
$\boldsymbol{\theta_2}$	$0 \cdots 2\pi$ [1]	0.25	+1	0.05
$\boldsymbol{\theta_3}$	$0 \cdots 2\pi$ [1]	0.25	+1	0.05

[1] Variation between the values.

  Figure 9 shows the variations in d_{min} as a function of both parameters $\eta(\boldsymbol{\theta_2})$ and $\eta(\boldsymbol{\theta_3})$. The maximum value of the minimum Fisher–Rao distance d_{min} was equal to $\sqrt{3} \simeq 1.7321$, and it was reached for $(\eta(\boldsymbol{\theta_2}), \eta(\boldsymbol{\theta_3})) = (2\pi/3, 4\pi/3)$, or symmetrically, $(\eta(\boldsymbol{\theta_2}), \eta(\boldsymbol{\theta_3})) = (4\pi/3, 2\pi/3)$.
  When we attempted to fix the maximum value for the distance between two signals, such as $d_{L2,\mathcal{B}}(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) = 2$, then another distance became less than $\sqrt{3}$.
  The optimal separation of the three signals was achieved when the phase difference between them was $2\pi/3$, which is a well-known result in communication theory.

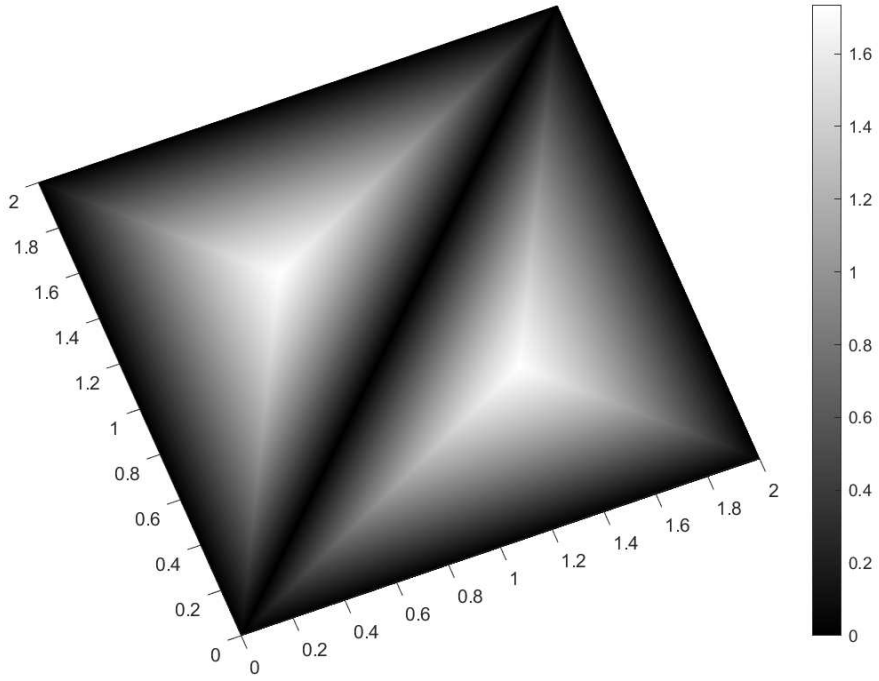


Figure 9. 3D plot of d_{min} for the toy problem of scenario 7. The X axis represents $\eta(\boldsymbol{\theta_2})$, and the Y axis represents $\eta(\boldsymbol{\theta_3})$. The color bar indicates the values of d_{min} (Z axis). The X and Y axes are graduated in π values (e.g., 2 means 2π, and 0.6 means 0.6π). We see that maximum values of d_{min} were reached for the parameter values $(\eta(\boldsymbol{\theta_2}), \eta(\boldsymbol{\theta_3})) = (2\pi/3, 4\pi/3)$ or $(\eta(\boldsymbol{\theta_2}), \eta(\boldsymbol{\theta_3})) = (4\pi/3, 2\pi/3)$.

8. Conclusions

In this article, we addressed signal processing through information geometry by computing the Fisher–Rao distance between two finite-energy signals observed in Gaussian noise. To obtain a closed-form expression for this distance, one must express and then solve the geodesic equation, given the boundary conditions imposed by the signals under consideration. To formulate the geodesic equation, the Fisher information matrix was used and differentiated with respect to the parameters to obtain the Christoffel symbols.

The calculation starts from the multivariate normal distribution, and we re-derive the geodesic equations for both distributions with unknown means and covariance matrices, as well as for the case of a known covariance matrix. The geodesic equation can be solved in the latter case, allowing application to finite-energy signals. Through the frequency-domain model of these signals, we identified the statistical manifold $L2(\mathcal{B})$ and obtained the closed-form expression for the Fisher–Rao distance on this manifold. This value appeared as a minimum bound for the Fisher–Rao distance between two finite-energy signals, regardless of the parameterization of these signals that defined a sub-manifold. The obtained expression makes it possible to highlight the spectral distance function $f_{\boldsymbol{\theta_1},\boldsymbol{\theta_2}}(\nu)$, which characterizes the impact of each frequency component on the Fisher–Rao distance value. We show that the Fisher–Rao distance varied as a function of the square roots of the signal-to-noise ratios, and for normalized signal-to-noise ratios equal to one, the distance value lied between 0 and 2. It is also shown that in the case of signals with disjoint spectral bands (or whose phase differences varied significantly within the band), the value of the distance was close to $\sqrt{2}$.

Several numerical examples related to LFM impulse signals are wenalyzed to illustrate the results obtained. It was shown that it is possible to reach the limiting values of the distance; the value zero was achieved for identical signals (confirming the concept of

distance), and the value two was achieved for identical signals in opposition of the phase. Beyond these extreme cases, the distance varied more or less significantly around the value $\sqrt{2}$, depending on the phase characteristics of the signals and the separation of their frequency bands. Using a toy problem, we provide theoretical insights into signal estimation, with an open perspective toward clustering and machine learning applications, where geometric distances can characterize classification and estimation tasks.

In the numerical analysis, signals with a strictly equal duration T and without delay were considered. The influence of delays and the signal duration could be the subject of future research. Nevertheless, we anticipate that the Fisher–Rao distance between finite energy signals will help the characterization of signal databases.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: The author thanks the anonymous reviewers for their feedback.

Conflicts of Interest: The author was employed by the company Thales. The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

DFT	Discrete Fourier transform
FM	Frequency modulation
LFM	Linear frequency modulation

References

1. Nielsen, F. An Elementary Introduction to Information Geometry. *Entropy* **2020**, *10*, 1100. [CrossRef] [PubMed]
2. Nielsen, F. A simple approximation method for the Fisher-Rao distance between multivariate normal distributions. *Entropy* **2023**, *25*, 654. [CrossRef] [PubMed]
3. Nielsen, F. Approximation and bounding techniques for the Fisher-Rao distances between parametric statistical models. In *Handbook of Statistics: Probability Models*; Elsevier: Amsterdam, The Netherlands, 2024; Volume 51, pp. 67–116. [CrossRef]
4. Skovgaard, L. *A Riemannian Geometry of the Multivariate Normal Model*; Technical Report No. 167; Stanford University: Stanford, CA, USA, 1981.
5. Skovgaard, L. A Riemannian Geometry of the Multivariate Normal Model. *Scand. J. Stat.* **1984**, *11*, 211–223.
6. Pinele, J.; Strapasson, J.E.; Costa, S. The Fisher-Rao distance between multivariate normal distributions: Special cases, bounds and applications. *Entropy* **2020**, *22*, 404. [CrossRef] [PubMed]
7. Miyamoto, H.; Meneghetti, F.; Pinele, J.; Costa, S. On Closed-Form Expressions for the Fisher-Rao Distance. *Inf. Geom.* **2024**, *7*, 311–354. [CrossRef]
8. Calvo, M.; Oller, J.M. An explicit solution of information geodesic equations for the multivariate normal model. *Stat. Risk Model.* **1991**, *9*, 119–138. [CrossRef]
9. Amari, S.I.; Nagaoka, H. *Methods of Information Geometry*; AMS: Providence, RI, USA, 2000.
10. Florin, F.; Bertrand, A. Géométrie de l'information et signaux d'énergie finie. In *Proceedings of the 30e Colloque sur le Traitement du Signal et des Images*; GRETSI-Groupe de Recherche en Traitement du Signal et des Images: Strasbourg, France, 2025; Number 2025-1419; pp. 5–8.
11. Florin, F. The Fisher-Rao distance between finite energy signals. In *Proceedings of the Geometric Science of Information, GSI 2025*; Nielsen, F., Barbaresco, F., Eds.; Springer Nature: Cham, Switzerland, 2025. [CrossRef]
12. Florin, F. The Fisher-Rao distance between finite energy signals in Gaussian Noise. *Inf. Geom.* **2026**. [CrossRef]
13. Amari, S.I. *Information Geometry and Its Applications*; Springer: Tokyo, Japan, 2016. [CrossRef]
14. Florin, F. On Fisher Information Matrix, Array Manifold Geometry and Time Delay Estimation. In *Proceedings of the Geometric Science of Information, GSI 2023*; Nielsen, F., Barbaresco, F., Eds.; Springer: Cham, Switzerland, 2023; pp. 307–317. [CrossRef]

15. Strapasson, J.E.; Pinele, J.; Costa, S.I.R. Clustering using the Fisher-Rao distance. In *Proceedings of the 2016 IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM), Rio de Janeiro, Brazil, 10–13 July 2016*; IEEE: New York, NY, USA, 2016; pp. 1–5. [CrossRef]
16. Amari, S.i.; Cichocki, A. Information geometry of divergence functions. *Bull. Pol. Acad. Sci. Tech. Sci.* **2010**, *58*, 183–195. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.