

Letter

Isometric Signal Processing under Information Geometric Framework

Hao Wu *, Yongqiang Cheng * and Hongqiang Wang

College of Electronic Science, National University of Defense Technology, Changsha 410073, China; wanghongqiang@nudt.edu.cn

* Correspondence: wuhao13c@nudt.edu.cn (H.W.); nudtyqcheng@gmail.com (Y.C.)

Received: 28 February 2019 ; Accepted: 25 March 2019; Published: 27 March 2019



Abstract: Information geometry is the study of the intrinsic geometric properties of manifolds consisting of a probability distribution and provides a deeper understanding of statistical inference. Based on this discipline, this letter reports on the influence of the signal processing on the geometric structure of the statistical manifold in terms of estimation issues. This letter defines the intrinsic parameter submanifold, which reflects the essential geometric characteristics of the estimation issues. Moreover, the intrinsic parameter submanifold is proven to be a tighter one after signal processing. In addition, the necessary and sufficient condition of invariant signal processing of the geometric structure, i.e., isometric signal processing, is given. Specifically, considering the processing with the linear form, the construction method of linear isometric signal processing is proposed, and its properties are presented in this letter.

Keywords: information geometry; intrinsic parameter submanifold; isometric signal processing

1. Introduction

Information geometry was pioneered by Rao [1] in 1945, and the more concise framework was built up by Chentsov [2], Efron [3,4], and Amari [5]. In information geometry, the research object is the statistical manifold, which consists of a parameterized family of probability distributions with a topological structure, $\mathcal{M} = \{p(x; \xi)\}$. Given the Fisher information matrix as the Riemannian metric, the distance between any two points (probability distributions) can be calculated [6]. In such a manifold, the distance between two points stands for the intrinsic measure for the dissimilarity between two probability distributions [7]. As information geometry provides a new perspective on signal processing, there are many applications of it. In estimation issues, based on the Riemannian distance, the natural gradient has been employed [8–10]. The intrinsic Cramér–Rao bound is a tighter bound of both biased and unbiased estimators and derives from the Grassmann manifold [11]. In addition, the geometric structure (considering the distance between all pairs of points) can be used as an evaluation of the quality of the observation model, which has been applied in waveform optimization [12]. In optimization problems under the matrix constraint, the geometric structure was utilized [13–15]. Moreover, there are also many significant works of detection based on the distance [16–20]. Furthermore, in image processing, based on the Grassmann manifold, the target recognition in the SAR (Synthetic Aperture Radar) image is proposed [21].

As this new general theory has revealed the capability to solve statistical problems, the further development of information geometry demands the unambiguous relationship between the geometric structure and the intrinsic characteristic of common issues. This letter focuses on the influence of the signal processing on the statistical manifold in terms of estimation issues. In the estimation issues, the signal processing is the common means to mine for the information of a desired parameter. Accompanying signal processing, the geometric structure of the considered statistical manifold,

to which the distribution of the observed data belongs, would change. The purpose of this letter is studying the geometric structure change accompanying signal processing and proposing an appropriate processing based on the change of the structure.

This research will be presented in the following way. At first, according to the essence of the estimation issues, the intrinsic parameter submanifold, which reflects the geometric characteristic of the issues, has been defined. Then, we show that the statistical manifold will become a tighter one after processing and give the necessary and sufficient condition of the invariant signal processing of the geometric structure (named isometric signal processing). Considering the more specific condition that the processing is linear, the construction method of linear isometric processing is proposed. Moreover, the properties of the constructed processing are presented.

The following notations are adopted in this paper: the math italic x , lowercase bold italic \mathbf{x} , and uppercase bold \mathbf{A} denote the scalars, vectors, and matrices, respectively. Constant matrix \mathbf{I} indicates the identity matrix. Symbols $(\cdot)^H$, $(\cdot)^T$, and $(\cdot)^*$ indicate the conjugate transpose operator, transpose operator, and the complex conjugate, respectively. In addition, $[\mathbf{A}]_{ij}$ indicates the i th row j th column element of matrix \mathbf{A} , and $\text{rk}(\mathbf{A})$ is the rank of matrix \mathbf{A} . Moreover, $\mathbf{A} \geq 0$ means that the matrix \mathbf{A} is a positive semidefinite matrix. Finally, $\mathbb{E}(\cdot)$ indicates the statistical expectation of a random variable.

2. Intrinsic Parameter Submanifold

Let $\mathcal{M} = \{p(x, \xi)\}$ be a statistical manifold with coordinate system ξ , which consists of a family of probability distributions. Consider an estimation issue on the statistical manifold \mathcal{M} ; the observed data $\mathbf{x} = (x_1, x_2, \dots, x_N)$ belong to one of the probability distributions $p(x, \xi)$ in \mathcal{M} . Suppose the desired parameter θ is implied in parameter ξ and the relation between θ and ξ can be expressed as a mapping, $h : \theta \mapsto \xi$. As an instance, in the distance measurement of the pulse-Doppler radar, the desired distance r is embedded in the statistical mean μ of the observed data, i.e., $\mu = h(r) = P(t - 2r/c)$ ($P(t)$ means the pulse signal, and c is the velocity of light).

Actually, not all $p(x, \xi)$ in \mathcal{M} are concerned with the estimation issue; the considered probability distributions $\{p(x, h(\theta))\}$ not cover the whole manifold, they are only from a submanifold, which is the essential manifold in the issue. In the above example, the considered distributions are screened by the pulse signal $P(t)$ (the statistical mean μ is able to be expressed as $P(t - 2r/c)$).

Definition 1 (Intrinsic parameter submanifold). *The manifold $\mathcal{S} = \{p(x, h(\theta))\}$ is the intrinsic parameter submanifold of $\mathcal{M} = \{p(x, \xi)\}$, with coordinate system θ .*

The Riemannian metric of submanifold \mathcal{S} is defined as $\mathbf{I}_x(\theta)$, the Fisher information matrix associated with parameter θ , as in Figure 1. Actually, the distance of two points on the submanifold is defined by using the Riemannian metric [6].

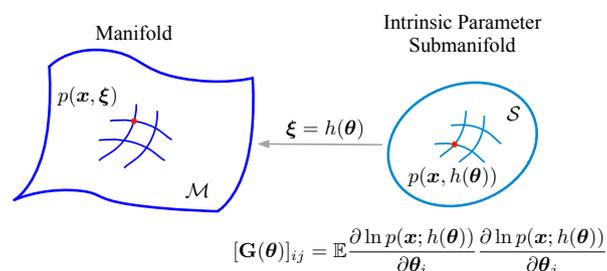


Figure 1. The intrinsic parameter submanifold.

Remark 1. *When the Fisher information matrices $\mathbf{G}_1, \mathbf{G}_2$ belonging to two observation models satisfy $\mathbf{G}_1 \geq \mathbf{G}_2$, the observation model with \mathbf{G}_1 is suggested to be better than another in terms of the estimation problem. The reason is that the distance $\mathcal{D}_1(\theta_1, \theta_2)$ (defined by \mathbf{G}_1) is larger than $\mathcal{D}_2(\theta_1, \theta_2)$ (defined by \mathbf{G}_2), because of*

the definition of the distance on the manifold. That means the two parameters θ_1, θ_2 are easier to discriminate in the manifold with \mathbf{G}_1 than \mathbf{G}_2 .

Furthermore, the above remark also can be explained in traditional statistical signal processing. In estimation theory, the Fisher information also plays an important role, as the CRLB (Cramér–Rao Lower Bound) inequality. Therefore, in the traditional estimation theory, the same conclusion can be deduced.

3. Signal Processing on the Intrinsic Parameter Submanifold

3.1. Geometric Structure Change by Signal Processing

In estimation issues, the signal is often processed to another form to obtain accurate estimates. Consider the signal processing $\mathbf{y} = g(\mathbf{x})$, where \mathbf{x} indicates the original signal and \mathbf{y} is the processed signal. The signal processing often accompanies the varying of the statistical manifold, specially the varying of the Riemannian metric.

One of the most vital factors of the submanifold in terms of estimation issues is its Riemannian metric, because the distance, representing the similarity, between two parameters is defined by it. Suppose the intrinsic parameter submanifold of x and y are \mathcal{S} and \mathcal{S}' , respectively. The Riemannian metrics of \mathcal{S} and \mathcal{S}' are $\mathbf{G}_{\mathcal{S}}$ and $\mathbf{G}_{\mathcal{S}'}$, respectively. If the PDFs (Probability Density Functions) $p_x(\mathbf{x}; \theta)$, $p_y(\mathbf{y}; \theta)$, and $p_{xy}(\mathbf{x}, \mathbf{y}; \theta)$ obey the boundary condition [22], then the Fisher information satisfies the following equation [22,23],

$$-\mathbb{E} \frac{\partial^2 \ln p_{x,y}(\mathbf{x}, \mathbf{y}; \theta)}{\partial \theta_i \partial \theta_j} = -\mathbb{E} \frac{\partial^2 \ln p_{x|y}(\mathbf{x}|\mathbf{y}; \theta)}{\partial \theta_i \partial \theta_j} - \mathbb{E} \frac{\partial^2 \ln p_y(\mathbf{y}; \theta)}{\partial \theta_i \partial \theta_j}, \tag{1}$$

$$-\mathbb{E} \frac{\partial^2 \ln p_{x,y}(\mathbf{x}, \mathbf{y}; \theta)}{\partial \theta_i \partial \theta_j} \geq -\mathbb{E} \frac{\partial^2 \ln p_y(\mathbf{y}; \theta)}{\partial \theta_i \partial \theta_j}. \tag{2}$$

Because \mathbf{y} is produced by \mathbf{x} via $\mathbf{y} = g(\mathbf{x})$, the following equation has been established.

$$-\mathbb{E} \frac{\partial^2 \ln p_{x,y}(\mathbf{x}, \mathbf{y}; \theta)}{\partial \theta_i \partial \theta_j} = -\mathbb{E} \frac{\partial^2 \ln p_x(\mathbf{x}; \theta)}{\partial \theta_i \partial \theta_j} \quad (\mathbf{y} = g(\mathbf{x})) \tag{3}$$

Proof. Because $p_{xy}(\mathbf{x}, \mathbf{y}; \theta) = 0$ for $\mathbf{y} \neq g(\mathbf{x})$, then the $p_{x,y}(\mathbf{x}, \mathbf{y}; \theta)$ can be expressed as $p_{x,y}(\mathbf{x}, \mathbf{y}; \theta) = p_x(\mathbf{x}; \theta) \delta(\mathbf{y} - g(\mathbf{x}))$, the Fisher information can be simplified:

$$-\mathbb{E} \frac{\partial^2 \ln p_{x,y}(\mathbf{x}, \mathbf{y}; \theta)}{\partial \theta_i \partial \theta_j} = -\mathbb{E} \frac{\partial^2 \ln p_x(\mathbf{x}; \theta)}{\partial \theta_i \partial \theta_j} - \mathbb{E} \frac{\partial^2 \ln \delta(\mathbf{y} - g(\mathbf{x}))}{\partial \theta_i \partial \theta_j} = -\mathbb{E} \frac{\partial^2 \ln p_x(\mathbf{x}; \theta)}{\partial \theta_i \partial \theta_j}. \tag{4}$$

□

Then, the following lemma holds.

Lemma 1. The Riemannian metrics $\mathbf{G}_{\mathcal{S}}$ and $\mathbf{G}_{\mathcal{S}'}$ satisfy, $\forall \theta$:

$$[\mathbf{G}_{\mathcal{S}'}(\theta)]_{ij} - \mathbb{E} \frac{\partial^2 \ln p_{x|y}(\mathbf{x}|\mathbf{y}; \theta)}{\partial \theta_i \partial \theta_j} = [\mathbf{G}_{\mathcal{S}}(\theta)]_{ij}. \tag{5}$$

Proof. By Equations (1) and (3), and the definitions of $\mathbf{G}_{\mathcal{S}}$ and $\mathbf{G}_{\mathcal{S}'}$, the lemma has been established. □

Corollary 1. For each θ , $\mathbf{G}_{\mathcal{S}}(\theta) - \mathbf{G}_{\mathcal{S}'}(\theta)$ is a positive semidefinite matrix, i.e., $\mathbf{G}_{\mathcal{S}}(\theta) \geq \mathbf{G}_{\mathcal{S}'}(\theta)$.

Proof. By Equation (2), Equation (3), and the definitions of \mathbf{G}_S and $\mathbf{G}_{S'}$, the corollary has been established. \square

Therefore, according to Lemma 1 and its corollary, the signal processing would result in Fisher information loss. As Figure 2 shows, the signal processing would turn the intrinsic parameter submanifold into a tighter one, i.e., discriminating two parameters turns out to be more difficult.

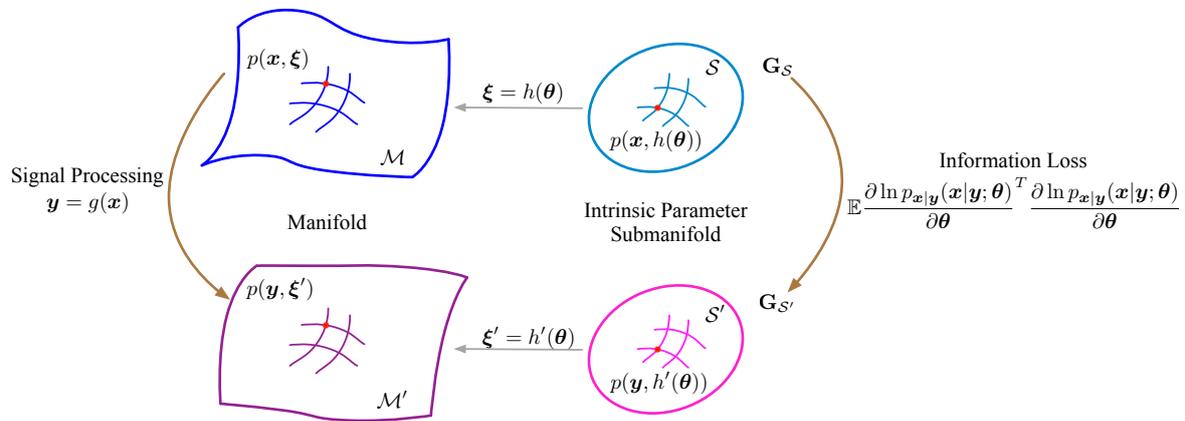


Figure 2. The signal processing on the intrinsic parameter submanifold.

3.2. Isometric Signal Processing

As the above discussion, the appropriate signal processing should satisfy that the intrinsic parameter submanifold of processed signal is isometric to the original submanifold, i.e., the difference between any two parameters is unreduced.

Definition 2 (Isometry). When $\mathbf{G}_S(\theta) = \mathbf{G}_{S'}(\theta)$, the two intrinsic parameter submanifolds S and S' are isometric.

Actually, the sufficient and necessary condition of the isometry of S and S' is as follows.

Theorem 1. If and only if \mathbf{y} is the sufficient statistic of \mathbf{x} , $\mathbf{G}_S(\theta) = \mathbf{G}_{S'}(\theta)$.

Proof. For Lemma 1, the following relations are equivalent,

$$\mathbf{G}_S(\theta) = \mathbf{G}_{S'}(\theta) \iff \mathbb{E} \left(\frac{\partial \ln p_{x|y}(x|y; \theta)}{\partial \theta_i} \right)^2 = 0 (\forall i) \iff \frac{\partial p_{x|y}(x|y; \theta)}{\partial \theta_i} \stackrel{a.e.}{=} 0 (\forall i). \tag{6}$$

That means $p_{x|y}(x|y; \theta)$ is irrelevant to parameter θ , i.e., \mathbf{y} is the sufficient statistic of \mathbf{x} . \square

The theorem suggests to use the test statistic to estimate the desired parameter, in the information geometry view. Actually, this conclusion also can be ensured in traditional estimation theory. For the Rao–Blackwell theorem [24], for any estimator $\hat{\theta}(x)$, the estimator $\check{\theta}(y) = \mathbb{E}(\hat{\theta}(x)|y)$ is the better estimator, i.e., $\mathbb{E}(\check{\theta}(y) - \theta)^2 \leq \mathbb{E}(\hat{\theta}(x) - \theta)^2$, when $\mathbf{y} = g(x)$ is the sufficient statistic. This theorem indicates that designing the estimator using the sufficient statistic \mathbf{y} is more appropriate, because for each estimator $\hat{\theta}(x)$ using the original signal \mathbf{x} as input, there exists the estimator $\check{\theta}(y) = \mathbb{E}(\hat{\theta}(x)|y)$ using the sufficient statistic \mathbf{y} as the input that is better than $\hat{\theta}(x)$. Furthermore, for the Lehmann–Scheffé theorem [25,26], when the sufficient statistic \mathbf{y} is complete, if the estimator $\check{\theta}(y)$ is unbiased, i.e., $\mathbb{E}(\check{\theta}(y)) = \theta$, the estimator $\check{\theta}(y)$ is the minimum-variance unbiased estimator.

Corollary 2. If $g(x)$ is a reversible function, $\mathbf{G}_S(\theta) = \mathbf{G}_{S'}(\theta)$.

Proof. If $g(x)$ is a reversible function, the PDF of x and y satisfy:

$$p_y(\mathbf{y}; \boldsymbol{\theta}) \frac{dg(x)}{dx} = p_x(x; \boldsymbol{\theta}). \quad (7)$$

According to the Fisher–Neyman factorization theorem [27], y is the sufficient statistic of x , so $\mathbf{G}_S(\boldsymbol{\theta}) = \mathbf{G}_{S'}(\boldsymbol{\theta})$. \square

When the processed signal $y = g(x)$ is the sufficient statistic of x , the signal processing $g(x)$ is the isometric signal processing. Specifically, the reversible processing is definitely isometric processing, such as DFT (Discrete Fourier Transformation, because the inverse discrete Fourier transformation can recover the original signal, i.e., DFT is a reversible process). Moreover, this conclusion is also encountered in traditional estimation theory as the Rao–Blackwell theorem and Lehmann–Scheffé theorem.

4. Linear Form of Signal Processing

In real works, the noise is often Gaussian or asymptotically Gaussian, and the common signal processing is linear, such as DFT, matched filter, coherent integration, etc. This section will discuss the linear form of signal processing on the Gaussian statistical manifold.

4.1. Model Formulation

The information, as the desired parameter, is usually embedded in the signal, and the signal is often contaminated by noise, which can be described as $x = s(\boldsymbol{\theta}) + w$, where $s(\boldsymbol{\theta})$ is the uncontaminated signal waveform, w is the Gaussian noise, and x is the signal. The linear signal processing can be expressed as a matrix form, $y = \mathbf{H}x$.

4.2. Fisher Information Loss of Linear Signal Processing

Suppose the linear form of signal processing is formed as $y = \mathbf{H}x$; x is the m dimension, and y is the n dimension, then the matrix \mathbf{H} is the $n \times m$ dimension. If $\text{rk}(\mathbf{H}) < n$, there are $n - \text{rk}(\mathbf{H})$ rows, which are the linear combination of the rest of the $\text{rk}(\mathbf{H})$ rows. Therefore, the PDF of y only depends on the $\text{rk}(\mathbf{H})$ corresponding elements, and the Fisher information loss is equivalent to the loss of the submatrix consisting of such $\text{rk}(\mathbf{H})$ rows. Therefore, for a convenient statement, $\text{rk}(\mathbf{H})$ is assumed to be n , i.e., matrix \mathbf{H} is row full rank.

The Fisher information loss will be discussed under WGN (White Gaussian Noise), at first. Then, the Fisher information under CGN (Colored Gaussian Noise) will be presented based on the results under WGN.

4.2.1. White Gaussian Noise

Suppose the noise is WGN and with power σ^2 , then the signal also obeys normal distribution $x \sim \mathcal{N}(s(\boldsymbol{\theta}), \sigma^2 \mathbf{I})$. As the property of the normal distribution, the distribution of y is also the normal distribution, but with different parameter $\mathcal{N}(\mathbf{H}s(\boldsymbol{\theta}), \sigma^2 \mathbf{H}\mathbf{H}^H)$. Calculate the Fisher information of x and $\mathbf{H}x$; the loss of information is:

$$\mathbf{G}_{\Delta}(\boldsymbol{\theta}) = \mathbf{G}_S(\boldsymbol{\theta}) - \mathbf{G}_{S'}(\boldsymbol{\theta}) = \frac{1}{\sigma^2} \frac{\partial s(\boldsymbol{\theta})^H}{\partial \boldsymbol{\theta}} (\mathbf{I} - \mathbf{H}^H (\mathbf{H}\mathbf{H}^H)^{-1} \mathbf{H}) \frac{\partial s(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}. \quad (8)$$

4.2.2. Colored Gaussian Noise

Suppose the noise is CGN and with covariance matrix \mathbf{C} . According to the property of the Hermite positive definite matrix, the covariance matrix can be expressed as $\mathbf{C} = \mathbf{D}\mathbf{D}^H$, where \mathbf{D} is a reversible matrix.

According to Theorem 1, perform the reversible transformation $\mathbf{x}^* = \mathbf{D}^{-1}\mathbf{x}$; the Fisher information is invariant, i.e., $\mathbf{G}_S(\boldsymbol{\theta}) = \mathbf{G}_{S'}(\boldsymbol{\theta})$, and the noise in \mathbf{x}^* is WGN. Performing the linear processing \mathbf{HD} to \mathbf{x}^* , the result is:

$$\mathbf{HD}\mathbf{x}^* = \mathbf{HDD}^{-1}\mathbf{x} = \mathbf{y}, \tag{9}$$

and the information loss can be calculated by Equation (8). Therefore, the loss of information is:

$$\mathbf{G}_\Delta(\boldsymbol{\theta}) = \frac{\partial \mathbf{s}(\boldsymbol{\theta})^H}{\partial \boldsymbol{\theta}} (\mathbf{C}^{-1} - \mathbf{H}^H(\mathbf{HCH}^H)^{-1}\mathbf{H}) \frac{\partial \mathbf{s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}. \tag{10}$$

4.3. The Construction of the Isometric Linear Form of Signal Processing

In the previous section, the sufficient and necessary condition of isometric signal processing was that $\mathbf{y} = g(\mathbf{x})$ is the sufficient statistic of \mathbf{x} . However, the sufficient statistic of \mathbf{x} is often difficult to obtain, and the isometric processing should be constructed in another way. This part will introduce the construction method of linear isometric signal processing.

As regards the previous discussion, the signal under CGN can be transformed to the signal under WGN without information loss. Therefore, the signal under WGN is discussed in this part. As for the condition of CGN, the signal can be white at first, then the next steps are the same as the WGN condition.

The linear isometric processing can be obtained in the following way. Firstly, solve the equation:

$$\forall \boldsymbol{\theta}, \frac{\partial \mathbf{s}(\boldsymbol{\theta})^H}{\partial \boldsymbol{\theta}} \mathbf{v} = \mathbf{0} \quad (\mathbf{v} \in \mathbb{R}^m). \tag{11}$$

Suppose the solution space is $\mathbf{V} = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_l\}$ with dimension l and the orthogonal complement of \mathbf{V} is \mathbf{V}^\perp with dimension $n = m - l$. Then, the desired signal processing is formed as:

$$\mathbf{H} = [\mathbf{v}'_1, \mathbf{v}'_2, \dots, \mathbf{v}'_n]^H, \tag{12}$$

where $\mathbf{v}'_1, \mathbf{v}'_2, \dots, \mathbf{v}'_n$ is the bias of \mathbf{V}^\perp .

Proposition 1. $\mathbf{H} = [\mathbf{v}'_1, \mathbf{v}'_2, \dots, \mathbf{v}'_n]^H$ is the isometric processing.

Proof. Let $\mathbf{Q} = \mathbf{I} - \mathbf{H}^H(\mathbf{HH}^H)^{-1}\mathbf{H}$. Because the non-zero eigenvalue of $\mathbf{H}^H(\mathbf{HH}^H)^{-1}\mathbf{H}$ is equivalent to that of $\mathbf{HH}^H(\mathbf{HH}^H)^{-1} = \mathbf{I}$, the eigenvalue of $\mathbf{H}^H(\mathbf{HH}^H)^{-1}\mathbf{H}$ is one (n multiplicity) and zero ($m - n$ multiplicity). Therefore, the eigenvalue of \mathbf{Q} is one ($m - n$ multiplicity) and zero (n multiplicity). Then, as the matrix \mathbf{Q} is the Hermitian symmetric matrix, it can be expressed as:

$$\mathbf{Q} = \mathbf{L} \text{diag}(1, \dots, 1, 0, \dots, 0) \mathbf{L}^H. \tag{13}$$

Consider the fact $\mathbf{QH}^H = \mathbf{0}$; the first n columns of \mathbf{HL} must equal zero. That means the first n columns of \mathbf{L} are the bias of \mathbf{V} , and the rest of the columns are the bias of \mathbf{V}^\perp , i.e.,

$$\mathbf{L} = [\mathbf{v}_1, \dots, \mathbf{v}_{m-n}, \mathbf{v}''_1, \dots, \mathbf{v}''_n]. \tag{14}$$

Because $\mathbf{v}_1, \dots, \mathbf{v}_{m-n}$ is the solution of Equation (12),

$$\mathbf{L}^H \frac{\partial \mathbf{s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = [\mathbf{v}_1, \dots, \mathbf{v}_{m-n}, \mathbf{v}''_1, \dots, \mathbf{v}''_n]^H \frac{\partial \mathbf{s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = [\mathbf{0}, \dots, \mathbf{0}, \mathbf{v}''_1{}^H \frac{\partial \mathbf{s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \dots, \mathbf{v}''_n{}^H \frac{\partial \mathbf{s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}]^T, \tag{15}$$

then:

$$\frac{\partial \mathbf{s}(\boldsymbol{\theta})^H}{\partial \boldsymbol{\theta}} \mathbf{L} \text{diag}(1, \dots, 1, 0, \dots, 0) \mathbf{L}^H \frac{\partial \mathbf{s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}, \tag{16}$$

i.e., the Fisher information loss is zero. \square

According to the proposed construction method, the following proposition can be obtained.

Proposition 2. *The matrix \mathbf{H} is the isometric matrix with the minimal rows, i.e., the processed signal has the minimal length.*

Proof. Let \mathbf{H}' be the isometric matrix with dimension n' and $\mathbf{Q}' = \mathbf{I} - \mathbf{H}'^H(\mathbf{H}'\mathbf{H}'^H)^{-1}\mathbf{H}'$. Similarly, the matrix also can be expressed as:

$$\mathbf{Q}' = \mathbf{L}' \text{diag}(1, \dots, 1, 0, \dots, 0)\mathbf{L}'^H, \tag{17}$$

where the multiplicity of eigenvalue one is $m - n'$.

As:

$$\frac{\partial \mathbf{s}(\boldsymbol{\theta})^H}{\partial \boldsymbol{\theta}} \mathbf{L}' \text{diag}(1, \dots, 1, 0, \dots, 0)\mathbf{L}'^H \frac{\partial \mathbf{s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}, \tag{18}$$

the first $m - n'$ rows of $\mathbf{L}'^H \frac{\partial \mathbf{s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ must be zero, which means the first $m - n'$ columns of \mathbf{L} is the linear independent solution of Equation (12). However, the solution space $\mathbf{V} = \text{span}\{v_1, v_2, \dots, v_l\}$ has dimension $m - n$, so we can get $m - n' \leq m - n$, i.e., $n' \geq n$.

Therefore, the matrix \mathbf{H} is the isometric matrix with the minimal rows. \square

Remark 2. *Because the first $m - n$ columns of \mathbf{L} are the linear independent solution of Equation (12), that means any element v' from \mathbf{V}^\perp satisfies that the first $m - n$ elements of $\mathbf{L}^H v'$ equal zero. Therefore, the solution space of $\mathbf{Q}'x = \mathbf{0}$ is \mathbf{V}^\perp . Moreover, \mathbf{H}' satisfies $\mathbf{Q}'\mathbf{H}'^H = \mathbf{0}$, so \mathbf{H}' consists of the bias of \mathbf{V}^\perp .*

In other words, the isometric matrix with dimension n is the equivalent matrix of \mathbf{H} , which indicates that the proposed construction method can generate any isometric matrix with minimal rows.

Sample of the Construction

Consider the radar target detection scene: the radar emits the single frequency signal and receives the echo to obtain the distance and RCS (Radar-Cross-Section) information of the target. The observation model can be formulated as:

$$x_k = A \exp(j2\pi f(kt_\Delta - \frac{2r}{c})) + w_k \quad k = 1, \dots, N, \tag{19}$$

where j indicates the unit of the imaginary part, t_Δ is the sampling interval, f is the frequency of the emitted signal, c is the velocity of light, w_k denotes WGN, r indicates the distance of the target, and A is the unknown amplitude, which contains the information of RCS. The desired parameter is $\boldsymbol{\theta} = (A, r)$.

Firstly, the derivative is:

$$\frac{\partial \mathbf{s}(\boldsymbol{\theta})^H}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \exp(j2\pi f(t_\Delta - \frac{2r}{c})) & \dots & \exp(j2\pi f(Nt_\Delta - \frac{2r}{c})) \\ \frac{4jA\pi f}{c} \exp(j2\pi f(t_\Delta - \frac{2r}{c})) & \dots & \frac{4jA\pi f}{c} \exp(j2\pi f(Nt_\Delta - \frac{2r}{c})) \end{bmatrix}. \tag{20}$$

Solve Equation (12); the orthogonal complement of the solution space is:

$$\text{span}\{(\exp(j2\pi f t_\Delta), \dots, \exp(j2\pi f N t_\Delta))\}. \tag{21}$$

Therefore, the isometric processing is:

$$y = \sum_{k=1}^N x_k \exp(j2\pi f k t_\Delta). \tag{22}$$

5. Conclusions

This letter focuses on the influence of signal processing on the geometric structure of the statistical manifold in estimation issues. Based on the intrinsic characteristics of the estimation issues, the intrinsic parameter submanifold is defined in this letter. Then, the intrinsic parameter submanifold is proven, which turns into a tighter one after signal processing. Moreover, we show that if and only if the processed signal is the sufficient statistic, the geometric structure of the intrinsic parameter submanifold is invariant. In addition, the construction method of the linear isometric signal processing is proposed. Moreover, the linear processing produced by the proposed method is shown with minimal rows (when it is represented as a matrix), i.e., the processed signal has the minimal length, and the proposed method can generate all linear isometry with minimal rows.

Author Contributions: H.W. put forward the original ideas and performed the research. Y.C. raised the research question, reviewed this paper, and provided improvement suggestions. H.W. reviewed the paper and provided useful comments. All authors have read and approved the final manuscript.

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

Acknowledgments: This work was supported by the National Natural Science Foundation of China under Grant No. 61871472. The authors are grateful for the valuable comments made by the reviewers, which have assisted us in having a better understanding of the underlying issues and therefore resulting in a significant improvement in the quality of the paper.

Conflicts of Interest: The authors declare that there is no conflict of interest regarding the publication of this paper.

References

1. Rao, C.R. Information and the Accuracy Attainable in the Estimation of Statistical Parameters. In *Breakthroughs in Statistics: Foundations and Basic Theory*; Kotz, S., Johnson, N.L., Eds.; Springer: New York, NY, USA, 1992; pp. 235–247.
2. Chentsov, N.N. *Statistical Decision Rules and Optimal Inference*; Number v. 53 in Translations of Mathematical Monographs; American Mathematical Society: Providence, RI, USA, 1982.
3. Efron, B. Defining the Curvature of a Statistical Problem (with Applications to Second Order Efficiency). *Ann. Stat.* **1975**, *3*, 1189–1242. [[CrossRef](#)]
4. Efron, B. The Geometry of Exponential Families. *Ann. Stat.* **1978**, *6*, 362–376. [[CrossRef](#)]
5. Amari, S.I. *Information Geometry and Its Applications*, 1st ed.; Springer Publishing Company, Incorporated: Berlin, Germany, 2016.
6. Chern, S.S.; Chen, W.H.; Lam, K.S. Lectures on Differential Geometry. *Ann. Inst. Henri Poincaré-Phys. Theor.* **2014**, *40*, 329–342.
7. Rong, Y.; Tang, M.; Zhou, J. Intrinsic Losses Based on Information Geometry and Their Applications. *Entropy* **2017**, *19*, 405. [[CrossRef](#)]
8. Cheng, Y.; Wang, X.; Caelli, T.; Moran, B. Tracking and Localizing Moving Targets in the Presence of Phase Measurement Ambiguities. *IEEE Trans. Signal Process.* **2011**, *59*, 3514–3525. [[CrossRef](#)]
9. Cheng, Y.; Wang, X.; Caelli, T.; Li, X.; Moran, B. Optimal Nonlinear Estimation for Localization of Wireless Sensor Networks. *IEEE Trans. Signal Process.* **2011**, *59*, 5674–5685. [[CrossRef](#)]
10. Cheng, Y.; Wang, X.; Moran, B. Optimal Nonlinear Estimation in Statistical Manifolds with Application to Sensor Network Localization. *Entropy* **2017**, *19*, 308. [[CrossRef](#)]
11. Smith, S.T. Covariance, subspace, and intrinsic Cramér-Rao bounds. *IEEE Trans. Signal Process.* **2005**, *53*, 1610–1630. [[CrossRef](#)]
12. Wang, L.; Wong, K.K.; Wang, H.; Qin, Y. MIMO radar adaptive waveform design for extended target recognition. *Int. J. Distrib. Sens. Netw.* **2016**, *2015*, 84. [[CrossRef](#)]
13. Manton, J.H. Optimization algorithms exploiting unitary constraints. *IEEE Trans. Signal Process.* **2002**, *50*, 635–650. [[CrossRef](#)]
14. Abrudan, T.E.; Eriksson, J.; Koivunen, V. Steepest Descent Algorithms for Optimization Under Unitary Matrix Constraint. *IEEE Trans. Signal Process.* **2008**, *56*, 1134–1147. [[CrossRef](#)]
15. Abrudan, T.; Eriksson, J.; Koivunen, V. Conjugate gradient algorithm for optimization under unitary matrix constraint. *Signal Process.* **2009**, *89*, 1704–1714. [[CrossRef](#)]

16. Barbaresco, F. Innovative Tools for Radar Signal Processing Based on Cartan's Geometry of SPD Matrices and Information Geometry. In Proceedings of the Radar Conference, Rome, Italy, 26–30 May 2008; pp. 1–6.
17. Barbaresco, F. Robust statistical Radar Processing in Fréchet metric space: OS-HDR-CFAR and OS-STAP Processing in Siegel homogeneous bounded domains. In Proceedings of the International Radar Symposium, Leipzig, Germany, 7–9 September 2011; pp. 639–644.
18. Wu, H.; Cheng, Y.; Hua, X.; Wang, H. Vector Bundle Model of Complex Electromagnetic Space and Change Detection. *Entropy* **2018**, *21*, 10. [[CrossRef](#)]
19. Hua, X.; Cheng, Y.; Wang, H.; Qin, Y.; Li, Y.; Zhang, W. Matrix CFAR detectors based on symmetrized Kullback-Leibler and total Kullback-Leibler divergences. *Digit. Signal Process.* **2017**, *69*, 106–116. [[CrossRef](#)]
20. Hua, X.; Fan, H.; Cheng, Y.; Wang, H.; Qin, Y. Information Geometry for Radar Target Detection with Total Jensen-Bregman Divergence. *Entropy* **2018**, *20*, 256. [[CrossRef](#)]
21. Dong, G.; Kuang, G.; Wang, N.; Wang, W. Classification via Sparse Representation of Steerable Wavelet Frames on Grassmann Manifold: Application to Target Recognition in SAR Image. *IEEE Trans. Image Process.* **2017**, *26*, 2892–2904. [[CrossRef](#)] [[PubMed](#)]
22. Zegers, P. Fisher Information Properties. *Entropy* **2015**, *17*, 4918–4939. [[CrossRef](#)]
23. Zamir, R. A proof of the Fisher information inequality via a data processing argument. *IEEE Trans. Inf. Theory* **1998**, *44*, 1246–1250. [[CrossRef](#)]
24. Blackwell, D. Conditional Expectation and Unbiased Sequential Estimation. *Ann. Math. Stat.* **1947**, *18*, 105–110. [[CrossRef](#)]
25. Lehmann, E.L.; Scheffé, H. Completeness, Similar Regions, and Unbiased Estimation—Part I. In *Selected Works of E. L. Lehmann*; Rojo, J., Ed.; Springer US: Boston, MA, USA, 2012; pp. 233–268.
26. Lehmann, E.L.; Scheffé, H. Completeness, Similar Regions, and Unbiased Estimation—Part II. In *Selected Works of E. L. Lehmann*; Rojo, J., Ed.; Springer US: Boston, MA, USA, 2012; pp. 269–286.
27. Kay, S.M. Fundamentals of statistical signal processing: Estimation theory. *Control Eng. Pract.* **1994**, *37*, 465–466.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).