Weyl Prior and Bayesian Statistics

Ruichao Jiang [1], Javad Tavakoli [1,]* and and Yiqiang Zhao [2]

[1] Department of Mathematics, The University of British Columbia Okanagan, Kelowna, BC V1V 1V7, Canada;
    ruichao.jiang@alumni.ubc.ca
[2] School of Mathematics and Statistics, Carlton University, Ottawa, ON K1S 5B6, Canada;
    zhao@math.carleton.ca
* Correspondence: javad.tavakoli@ubc.ca; Tel.: +1-250-807-9535

check for
updates

Abstract: When using Bayesian inference, one needs to choose a prior distribution for parameters. The well-known Jeffreys prior is based on the Riemann metric tensor on a statistical manifold. Takeuchi and Amari defined the α-parallel prior, which generalized the Jeffreys prior by exploiting a higher-order geometric object, known as a Chentsov–Amari tensor. In this paper, we propose a new prior based on the Weyl structure on a statistical manifold. It turns out that our prior is a special case of the α-parallel prior with the parameter α equaling $-n$, where n is the dimension of the underlying statistical manifold and the minus sign is a result of conventions used in the definition of α-connections. This makes the choice for the parameter α more canonical. We calculated the Weyl prior for univariate Gaussian and multivariate Gaussian distribution. The Weyl prior of the univariate Gaussian turns out to be the uniform prior.

Keywords: information geometry; Bayesian statistics; prior distributions; conformal geometry

1. Introduction

In Bayesian inference, a parameter is regarded as a random variable Θ. A density of Θ is called, by abuse of terminology, a prior distribution $p(\theta)$. After collecting some data, one obtains a conditional density $p(x|\theta)$, referred to as the likelihood function. The Bayes' theorem then computes the posterior distribution $p(\theta|x)$ using $p(\theta)$ and $p(x|\theta)$. This is interpreted as an update of the information about the unknown parameter Θ in Bayesian inference. One such choice of the prior distribution $p(\theta)$ is Jeffreys prior $^J\omega$, which is the correct choice of uniform distribution. Here, the word "uniform" means uninformative, not favorable of any particular choice of the parameter.

Information geometry, in its narrowest sense, is an attempt to use differential geometry to study statistical inference. It has found applications in statistical inference, signal processing, and machine learning [1]. References [2,3] are two elementary introductions. In information geometry, geometric structures, for example metric tensors g and affine connections ∇, can be put on the set of prior distributions $\mathcal{P}(\theta)$. These geometric structures help to single out some particular prior distributions, for example, the Jeffreys prior $^J\omega$, which, by the fundamental theorem of Riemannian geometry, is the unique volume form parallel with respect to the Levi–Civita connection $^{LC}\nabla$. Since the Jeffreys prior $^J\omega$ is provided by geometry, it is automatically invariant under reparametrization, which reflects the opinion that information can be at best not lost during a transformation of parameters and this is encoded in the notion of sufficient statistics. Similarly, if one can find a unique prior distribution satisfying some specified geometric conditions, then that prior distribution is called canonically chosen. Matsuzoe, Takeuchi, and Amari used information geometry to define the α-parallel prior $^\alpha\omega$ such that, when $\alpha = 0$, it reduces to the Jeffreys prior [4].

Historically, Weyl proposed a generalization of general relativity to unify gravity and electromagnetism. Einstein soon pointed out that Weyl's theory predicted substantial broadening of

the characteristic length of atoms, which is contradictory to the well-observed thin atomic spectra. Even though Weyl geometry failed the unification of gravity and electromagnetism, which is still an open problem, Weyl geometry has found applications in possible generalization of general relativity [5] and the differential-geometric study of defects in continuum mechanics [6]. Weyl geometry is kept in mathematics. The relation of affine differential geometry, Weyl geometry, and Riemannian geometry are shown below. Let $\pi : E \to B$ be a fibre bundle with base B and let each fibre $\pi^{-1}(x)$, $x \in B$, be a Lie group G, called the structure group of the fibre bundle. For different G, we obtain different geometries as follows:

1. $GL(n)$, the general linear group, affine differential geometry;
2. $C(n) := \{kA \mid k \in \mathbb{R}^+, \ A \in O(n)\}$, the conformal group, Weyl geometry;
3. $O(n)$, the orthogonal group, Riemannian geometry.

With the reduction of structure groups $O(n) \subseteq C(n) \subseteq GL(n)$, and the fact that smaller the structure group, the more geometric properties are expected. In our case, the reduction of group gives rise to a canonical choice for the parameter α of the α-parallel prior. For more about bundle-theoretic differential geometry, see [7].

In this paper, we will use Weyl geometry to define a prior distribution for Bayesian inference, which we call the Weyl prior. We will elucidate the relation between the dimension of a statistical manifold and the parameter α in Takeuchi and Amari's α-parallel prior.

The organization of the paper is as follows: In Section 2, we review information geometry and the α-parallel prior. We discuss Weyl geometry in Section 3. We define the Weyl prior, and elucidate the relation between the Weyl prior and the α-parallel prior in Section 4. We calculate the Weyl prior for the univariate Gaussian distribution as an example in Section 5 and the multivariate Gaussian distribution in Section 6. All functions in the paper are real-valued and smooth, all connections are torsion-free, and the Einstein summation rule is used.

2. Information Geometry and α-Priors

In this section, we review some basics of information geometry. For more details, see [1].

Let us consider a statistical model $\mathcal{P}$, which is a set of parametric densities $\mathcal{P} = \{p(x|\theta)\}$. $\mathcal{P}$ can be geometrized as follows: first, we introduce the Fisher metric tensor, which is a 2nd order tensor,

Definition 1. *The Fisher metric tensor is defined by*

$$g_{ij} = E_\theta \left[\partial_i l \partial_j l \right] \tag{1}$$

where E_θ is the the transition kernel $X \times \Theta \to [0, \infty)$, l is the log likelihood function, and ∂_i is the partial derivative with respect to coordinate i.

Then, we introduce the Amari–Chentsov tensor, which is a 3rd order tensor.

Definition 2 (Amari–Chentsov Tensor). *The Amari–Chentsov tensor C is defined by*

$$C_{ijk} = E_\theta \left[\partial_i l \partial_j l \partial_k l \right] . \tag{2}$$

Remark 1. *The Amari–Chentsov tensor C defined above satisfies*

$$C = \nabla g.$$

In other words, C is the covariant derivative of the metric tensor g.

In Riemannian geometry, C vanishes everywhere, which is required if the length of a tangent vector is to be preserved under the parallel transport. In information geometry, this requirement is dropped

and thus a duality theory arises. Let ∇ be an arbitrary torsion-free affine connection on a Riemannian manifold (M, g). The dual connection ∇^* of ∇ plays an important role in information geometry.

Definition 3 (Dual Connection). *The dual connection ∇^* on a Riemannian manifold (M, g) with affine connection ∇ is defined as the unique affine connection satisfying the following equation:*

$$Xg_p(Y, Z) = g_p(\nabla_X Y, Z) + g_p(Y, \nabla^*_X Z), \tag{3}$$

where $p \in M$ and $X, Y, Z \in T_p M$.

Remark 2. *The dual connection ∇^* preserves the metric tensor g together with ∇ :*

$$g_p(X, Y) = g_q(\Pi X, \Pi^* Y),$$

where $X, Y \in T_p M$, and Π and Π^* are parallel transports induced by ∇ and ∇^*, respectively, along some curve from p to q. In general, $g_p(X, Y) \neq g_q(\Pi X, \Pi Y)$ and $g_p(X, Y) \neq g_q(\Pi^* X, \Pi^* Y)$, unless $\nabla = \nabla^* =^{LC} \nabla$. See [1].

Now, we introduce α-connections.

Definition 4. *The α-connections are defined in terms of Christoffel symbols by*

$$^\alpha\Gamma^i_{jk} =^{LC} \Gamma^i_{jk} - \frac{\alpha}{2} g^{il} C_{ljk}, \tag{4}$$

where $\alpha \in \mathbb{R}$ and LC stands for Levi–Civita.

Remark 3. *The dual connection of ∇^α is then given by*

$$\nabla^{\alpha*} = \nabla^{-\alpha}. \tag{5}$$

Remark 4. *The α-parallel prior $^\alpha\omega$ is the volume form parallel with respect to ∇^α. Unlike the Jeffreys prior, which always exists, the α-parallel prior do not necessarily exist. An α-parallel prior exists if and only if the Ricci curvature tensor is symmetric [4]. However, if $^\alpha\omega$ exists for one $\alpha \in \mathbb{R}$, then it exists for all α [8].*

The following characterization will be used in Section 5 to obtain the relation between the α-parallel prior and the Weyl prior defined therein.

Proposition 1. *[4] Let $(M, g, ^\alpha\nabla)$ be a statistical manifold. If there exists an exact 1-form $T = d\Omega$ for some function Ω determined by ∇ and g, then the α-parallel prior is $^\alpha\omega = \exp\{-\frac{\alpha}{2}\Omega\}\sqrt{\det g}$.*

Remark 5. *$d\Omega$ is known as the Chebyshev 1-form. A differential form ϕ is called closed if the exterior derivative vanishes i.e., $d\phi = 0$, and is called exact if there exists a differential form φ such that $\phi = d\varphi$. By definition, every exact form is closed. By Poincare's lemma, every closed form is locally exact. Because statistical manifolds are simply connected, closedness implies exactness.*

3. Weyl Geometry

In this section, we review some concepts of Weyl geometry which are needed in the next section. For more details, see [9].

Two Riemannian metrics g and g' on a manifold M are said to be conformally equivalent if $g' = e^\lambda g$ for some smooth function λ on M.

A conformal structure $\mathcal{C}$ on M is an equivalent class of conformally equivalent Riemannian metrics, i.e., $\mathcal{C} := \{g' | g' = e^\lambda g\}$.

A Weyl structure is a map $F : \mathcal{C} \to \Lambda^1(M)$ from the conformal structure $\mathcal{C}$ to the set of 1-forms on M, satisfying

$$F(e^\lambda g) = F(g) - d\lambda.$$

The image of g under F is called the Weyl 1-form $F(g) := \varphi$.

A Weyl structure enables us to translate a scalar product $(\,,\,)_p$ at p to $(\,,\,)_q$ at q along a curve $c : [0,1] \to M$:

$$(\,,\,)_q = \exp\left[\int_0^1 c^* \varphi\right] g_q, \tag{6}$$

where $c^* \varphi$ is the pullback of the Weyl 1$-$form φ along curve c. A Weyl manifold is a manifold with a Weyl structure.

Remark 6. *The meaning of this equation is: If we start with a scalar product $(,)_p$ at a point p arising from the conformal class $\mathcal{C}$, then there exists a metric tensor $g \in \mathcal{C}$ extending $(,)_p$, i.e., $g_p = (,)_p$. The value of this particular choice of g at another point q is g_q. However, different choice of g gives rise to different g_q. The scalar product $(,)_q$ determined by Weyl translation is proven to be independent of g [9]. Hence, by Weyl translation, we can compare lengths of vectors at different points on a Weyl manifold, whereas, with only the conformal structure $\mathcal{C}$, we can only compare ratios of lengths.*

An affine connection ∇ is said to be a Weyl connection if the parallel transport of a scalar product under ∇ coincides with the Weyl translation.

The Weyl connection is characterized by the following propositions.

Proposition 2 ([9]). *An affine connection ∇ is a Weyl connection if and only if $\nabla g + \varphi \otimes g = 0$ for all $g \in \mathcal{C}$.*

Proposition 3 (Fundamental Theorem of Weyl Geometry [9]). *There exists a unique torsion-free Weyl connection $^W\nabla$ on a Weyl manifold M. The Christoffel symbols of $^W\nabla$ are given by*

$$^W\Gamma^i_{jk} =^{LC}\Gamma^i_{jk} + \frac{1}{2}\left(\delta^i_j \varphi_k + \delta^i_k \varphi_j - g^{im} g_{jk} \varphi_m\right), \tag{7}$$

where δ^i_j is the Kronecker delta.

4. Weyl Prior

In this section, we define the Weyl prior and show its relation to the α-parallel prior.

First, we define the Weyl prior as follows.

Definition 5 (Weyl Prior). *Let (M, g) be an n-dimensional Riemannian manifold with the conformal structure $\mathcal{C} = [g]$ and the Weyl structure F. Let $^W\nabla$ be the Weyl connection. The Weyl prior $^W\omega$ is defined as the unique volume form parallel with respect to $^W\nabla$.*

Remark 7. *The uniqueness of the Weyl prior is the result of the uniqueness of the Weyl connection.*

Now, we prove the main result of this paper.

Theorem 1. *Let (M, g) be a Riemannian manifold. Let $^W\nabla$ and $^{-n}\nabla$ be the Weyl connection and the $-n$-connection, i.e., the α-connection with $\alpha = -n$, where n is the dimension of M. Suppose that the $-n$-prior $^{-n}\omega$ exists, then*

$$^W\omega =^{-n}\omega.$$

Proof. Consider an arbitrary volume form $f\sqrt{\det g}$, where f is a positive function on M. For $f\sqrt{\det g}$ to be parallel with respect to $^W\omega$, it is necessary and sufficient that

$$^W\nabla\left(f\sqrt{\det g}\right) = f^W\nabla\sqrt{\det g} + ^W\nabla f\sqrt{\det g} = 0. \tag{8}$$

Componentwise, Equation (8) becomes

$$f^W\nabla_j\sqrt{\det g} + ^W\nabla_j f\sqrt{\det g} = f^W\nabla_j\sqrt{\det g} + \partial_j f\sqrt{\det g}, \tag{9}$$

since covariant derivative coincides with partial derivative for functions.

Since $\sqrt{\det g}$ is a scalar density of weight 1, its covariant derivative is given by

$$^W\nabla_j\sqrt{\det g} = \partial_j\sqrt{\det g} - ^W\Gamma_j\sqrt{\det g}, \tag{10}$$

where Γ_j is obtained by the contraction of Equation (7) over i and k :

$$\begin{aligned}
^W\Gamma_j = ^W\Gamma_{ji}^i = ^{LC}\Gamma_{ji}^i &+ \frac{1}{2}\left(\delta_j^i\varphi_i + \delta_i^i\varphi_j - g^{im}g_{ji}\varphi_m\right) \\
&= \partial_j\ln\sqrt{\det g} + \frac{1}{2}\left(\varphi_j + n\varphi_j - \delta_j^m\varphi_m\right) \\
&= \partial_j\ln\sqrt{\det g} + \frac{n}{2}\varphi_j.
\end{aligned}$$

Substituting Equation (10) into Equation (9), we obtain

$$\partial_j f = \frac{n}{2}\varphi_j f. \tag{11}$$

Since the covariant derivative coincides with exterior derivative for functions, collect indices in Equation (11)

$$\varphi = \frac{2}{n}d\ln f. \tag{12}$$

Assume for now that the Weyl 1-form φ is exact, that is, $\varphi = d\Omega$ for some function Ω on M. Then, from Equation (29), the Weyl prior is given by

$$^W\omega = \exp\{\frac{n}{2}\Omega\}\sqrt{\det g}. \tag{13}$$

By comparison of Equation (13) with Proposition 1, the theorem is proved under the assumption of the exactness of the Weyl 1-form.

Since we proved that the Weyl prior $^W\omega$ is the $\alpha-$prior with $\alpha = -n$, and we required the existence of $^n\omega$, our assumption of the exactness of the Weyl 1$-$form φ is indeed true by Remark 4. $\quad\square$

Remark 8. *The minus sign in $\alpha = -n$ is a result of the definition of α-connection. By Remark 3, the dual connection of $^\alpha\nabla$ is $^{-\alpha}\nabla$, we would have $\alpha = n$ here, had we defined the α-connection to be its dual connection in Definition 4. It would seem more natural to consider the dual prior of the Weyl prior.*

5. Weyl Prior for Gaussian Family

In this section, we calculate the Weyl prior of the Gaussian family as an example.

Example 1 (Gaussian Family)**.** *Consider the Gaussian family*

$$\mathcal{P} = \left\{p(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right] \mid (\mu,\sigma) \in M\right\}.$$

Choosing (μ, σ^2) as a coordinate system, we have

$$\partial_\mu l = \frac{x - \mu}{\sigma^2},$$

$$\partial_{\sigma^2} l = \frac{(x - \mu)^2}{2\sigma^4} - \frac{1}{2\sigma^2},$$

where l is the log likelihood function.

The first element of the Fisher metric tensor g in the (μ, σ^2)-coordinate is given by

$$\begin{aligned}
g_{\mu\mu} &= E_\theta \left[\partial_\mu l \partial_\mu l \right] \\
&= \int_{-\infty}^{\infty} \frac{(x - \mu)^2}{\sigma^4} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}(x - \mu)^2 \right] dx \\
&= \frac{1}{\sigma^2},
\end{aligned}$$

where E_θ is the conditional expectation of X given Θ. The other elements of the Fisher metric tensor are

$$g_{\mu\sigma^2} = g_{\sigma^2\mu} = 0,$$

and

$$g_{\sigma^2\sigma^2} = \frac{1}{2\sigma^4}.$$

Hence,

$$\sqrt{\det g} = \frac{1}{\sqrt{2}\sigma^3}. \tag{14}$$

To calculate the Weyl 1-form, we first calculate the Amari–Chentsov tensor C,

$$\begin{aligned}
C_{\mu\mu\mu} &= E_\theta \left[\partial_\mu l \partial_\mu l \partial_\mu l \right] \\
&= \int_{-\infty}^{\infty} \frac{(x - \mu)^3}{\sigma^6} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}(x - \mu)^2 \right] dx \\
&= 0.
\end{aligned}$$

Similarly,

$$C_{\sigma^2\mu\mu} = C_{\mu\sigma^2\mu} = C_{\mu\mu\sigma^2} = \frac{1}{\sigma^4},$$

$$C_{\sigma^2\sigma^2\mu} = C_{\sigma^2\mu\sigma^2} = C_{\mu\sigma^2\sigma^2} = 0,$$

and

$$C_{\sigma^2\sigma^2\sigma^2} = \frac{1}{\sigma^6}.$$

Hence, the Weyl 1-form is given by

$$\begin{aligned}
\varphi &= \frac{1}{2} C_{ijk} g^{jk} d\theta^i \\
&= \frac{3}{2\sigma^2} d\sigma^2.
\end{aligned} \tag{15}$$

Now, it is easy to check that $\varphi = \mathrm{d}(\frac{3}{2}\ln\sigma^2)$ is an exact form. Hence, for Gaussian family $\mathcal{P}$, a Weyl prior exists and is given by

$$
\begin{aligned}
^W\omega &= \exp\left\{\frac{2}{2}\frac{3}{2}\ln\sigma^2\right\}\frac{1}{\sqrt{2}\sigma^3} \\
&= \frac{1}{\sqrt{2}}.
\end{aligned}
\tag{16}
$$

Remark 9. *Based on our calculation, we find that the Weyl prior for the univariate Gaussian distribution with unknown mean and unknown variance is just the uniform prior. This shows that the uniform prior is in fact a uninformative prior. This counter-intuitive result is related to the fact that every two-dimensional manifold is conformally-flat, which can be proved using the existence of isothermal coordinates in two dimensions [10].*

6. Multivariate Gaussian

The above example can be extended to the multivariate case. Consider the multivariate Gaussian distribution

$$
f(x|\mu,\Sigma) = \frac{1}{(2\pi)^{n/2}\sqrt{\det\Sigma}}\exp\left\{-\frac{1}{2}(x-\mu)^{\mathsf{T}}\Sigma^{-1}(x-\mu)\right\},
\tag{17}
$$

where μ is the mean vector and Σ is the covariance matrix.

Using matrix calculus, we have

$$
\partial_\mu l = \Sigma^{-1}(x-\mu)
\tag{18}
$$

and

$$
\begin{aligned}
\partial_\Sigma l &= -\frac{1}{2}\Sigma^{-1} + \frac{1}{2}\Sigma^{-1}(x-\mu)(x-\mu)^{\mathsf{T}}\Sigma^{-1} \\
&= -\frac{1}{2}\Sigma^{-1} + \frac{1}{2}[\Sigma^{-1}(x-\mu)]\otimes[\Sigma^{-1}(x-\mu)] \\
&= -\frac{1}{2}\Sigma^{-1} + \frac{1}{2}(\Sigma^{-1}\otimes\Sigma^{-1})[(x-\mu)\otimes(x-\mu)]
\end{aligned}
\tag{19}
$$

We can now compute the Fisher information matrix.

$$
\begin{aligned}
g_{\mu\mu} &= \mathrm{E}_\theta[\partial_\mu\otimes\partial_\mu] \\
&= (\Sigma^{-1}\otimes\Sigma^{-1})\mathrm{E}_\theta[(x-\mu)\otimes(x-\mu)] \\
&= (\Sigma^{-1}\otimes\Sigma^{-1})\Sigma \\
&= \Sigma^{-1}\Sigma\Sigma^{-\mathsf{T}} \\
&= \Sigma^{-1}
\end{aligned}
\tag{20}
$$

where the second last line is by the action of matrix tensor product and the last line is by the definition of covariance matrix.

Similarly,

$$
g_{\mu\Sigma} = g_{\Sigma\mu} = 0,
\tag{21}
$$

and

$$
g_{\Sigma\Sigma} = \frac{1}{2}\Sigma^{-1}\otimes\Sigma^{-1}.
\tag{22}
$$

The Amari–Chentsov tensor can be computed in the same way:

$$
C_{\mu\mu\mu} = \mathrm{E}_\theta[\partial_\mu l\otimes\partial_\mu l\otimes\partial_\mu l] = 0.
\tag{23}
$$

$$
C_{\Sigma\mu\mu} = C_{\mu\Sigma\mu} = C_{\mu\mu\Sigma} = \Sigma^{-1}\otimes\Sigma^{-1}.
\tag{24}
$$

$$C_{\mu\Sigma\Sigma} = C_{\Sigma\mu\Sigma} = C_{\Sigma\Sigma\mu} = \mathbf{0}. \tag{25}$$

$$C_{\Sigma\Sigma\Sigma} = \Sigma^{-1} \otimes \Sigma^{-1} \otimes \Sigma^{-1}. \tag{26}$$

The detail computation of Equation (26) is as follows:

$$
\begin{aligned}
C_{\Sigma\Sigma\Sigma} &= E_\theta[\partial_\Sigma l \otimes \partial_\Sigma l \otimes \partial_\Sigma l] \\
&= E_\theta\{-\frac{1}{8}\Sigma^{-1} \otimes \Sigma^{-1} \otimes \Sigma^{-1} + \frac{1}{8}\Sigma^{-1} \otimes (\Sigma^{-1} \otimes \Sigma^{-1})[(x-\mu) \otimes (x-\mu)] \otimes \Sigma^{-1} \\
&\quad + \frac{1}{8}\left(\Sigma^{-1} \otimes \Sigma^{-1}\right)[(x-\mu) \otimes (x-\mu)] \otimes \left(\Sigma^{-1} \otimes \Sigma^{-1}\right) - \frac{1}{8}\left(\Sigma^{-1} \otimes \Sigma^{-1} \otimes \Sigma^{-1} \otimes \Sigma^{-1}\right) \\
&\quad (x-\mu) \otimes (x-\mu) \otimes (x-\mu) \otimes (x-\mu) \otimes \Sigma^{-1} + \frac{1}{8}\Sigma^{-1} \otimes \Sigma^{-1} \otimes \left(\Sigma^{-1} \otimes \Sigma^{-1}\right) \\
&\quad [(x-\mu) \otimes (x-\mu)] - \frac{1}{8}\Sigma^{-1} \otimes \left(\Sigma^{-1} \otimes \Sigma^{-1}\right)[(x-\mu) \otimes (x-\mu)] \otimes \left(\Sigma^{-1} \otimes \Sigma^{-1}\right) \\
&\quad [(x-\mu) \otimes (x-\mu)] - \frac{1}{8}\left(\Sigma^{-1} \otimes \Sigma^{-1}\right)[(x-\mu) \otimes (x-\mu)] \otimes \Sigma^{-1} \otimes \left(\Sigma^{-1} \otimes \Sigma^{-1}\right) \\
&\quad [(x-\mu) \otimes (x-\mu)] + \frac{1}{8}\left(\Sigma^{-1} \otimes \Sigma^{-1} \otimes \Sigma^{-1} \otimes \Sigma^{-1} \otimes \Sigma^{-1} \otimes \Sigma^{-1}\right)[(x-\mu) \otimes (x-\mu) \\
&\quad \otimes (x-\mu) \otimes (x-\mu) \otimes (x-\mu) \otimes (x-\mu)]\} = \Sigma^{-1} \otimes \Sigma^{-1} \otimes \Sigma^{-1}
\end{aligned} \tag{27}
$$

The above expression can be evaluated by 4th and 6th moments of multivariate Gaussian. The Weyl prior is then given by:

$$
\begin{aligned}
\varphi &= \frac{1}{2}C_{ijk}g^{jk}d\theta^i \\
&= \frac{1}{2}C_{\Sigma\mu\mu}g^{\mu\mu}d\Sigma + \frac{1}{2}C_{\Sigma\Sigma\Sigma}g^{\Sigma\Sigma}d\Sigma \\
&= \frac{1}{2}\left(\Sigma^{-1} \otimes \Sigma^{-1}\right)\Sigma d\Sigma + \frac{1}{2}\left(\Sigma^{-1} \otimes \Sigma^{-1} \otimes \Sigma^{-1}\right)2\left(\Sigma \otimes \Sigma\right)d\Sigma \\
&= \frac{3}{2}\Sigma^{-1}d\Sigma \\
&= d\left(\frac{3}{2}\ln\det\Sigma\right).
\end{aligned} \tag{28}
$$

The Weyl prior is thus given by:

$$
\begin{aligned}
{}^W\omega &= \exp\left\{\frac{n + (n+1)n/2}{2}\frac{3}{2}\ln\det\Sigma\right\}\sqrt{\det\Sigma^{-1}\det\left[\frac{1}{2}\Sigma^{-1} \otimes \Sigma^{-1}\right]} \\
&= (\det\Sigma)^{(3n^2+9n)/8}\frac{(\det\Sigma)^{(-2n-1)/2}}{2^{n/2}} \\
&= \frac{(\det\Sigma)^{(n-1)(3n+4)/8}}{2^{n/2}},
\end{aligned} \tag{29}
$$

where, in the first line, $n + (n+1)n/2$ is the dimension of the statistical manifold for the multivariate Gaussian distribution.

Remark 10. *Our calculation of the Weyl prior of the multivariate Gaussian distribution is generally not a uniform prior. However, Equation (29) shows that, when $n = 1$, that is, the univariate case, the Weyl prior is indeed the uniform prior. This is in accordance with our direct calculation for the univariate case.*

7. Discussion and Conclusions

We discussed Weyl geometry and Weyl prior in this paper. We also calculated Weyl prior for the Gaussian family as an example.

The underlying principle of Jeffreys prior, α-parallel prior, and Weyl prior is the concept of invariance in statistics. Jeffreys prior is invariant under a change of the coordinate of parameters. Weyl prior and α-parallel prior, as generalizations of Jeffreys prior, automatically satisfy this invariance. Moreover, Weyl prior, as a volume form defined on a Weyl manifold, is also invariant under a gauge transformation [11]. In addition, invariant under the gauge transformation is the generalized conjugate connection [11].

One possible use of the Weyl prior is using the uniform prior for distributions with two parameters. This is because any two-dimensional manifold is conformally-flat.

References

1. Amari, S. *Information Geometry and Its Applications*; Springer: Berlin/Heidelberg, Germany, 2016; Volume 194.
2. Calin, O.; Udriste, C. *Geometric Modeling in Probability and Statistics*; Springer: Basel, Switzerland, 2014.
3. Nielsen, F. An elementary introduction to information geometry. *arXiv* **2018**, arXiv:1808.08271.
4. Matsuzoe, H.; Takeuchi, J.; Amari, S. Equiaffine structures on statistical manifolds and Bayesian statistics. *Differ. Geom. Its Appl.* **2006**, *24*, 567–578. [CrossRef]
5. Ciambelli, L.; Leigh, R.G. Weyl Connections and their Role in Holography. *arXiv* **2019**, arXiv:1905.04339.
6. Yavari, A.; Goriely, A. Weyl geometry and the nonlinear mechanics of distributed point defects. *Proc. R. Soc. A* **2012**, *468*, 3902–3922. [CrossRef]
7. Kobayashi, S.; Nomizu, K. *Foundations of Differential Geometry*; Wiley: New York, NY, USA, 1963; Volume 1.
8. Takeuchi, J.; Amari, S. α-parallel prior and its properties. *IEEE Trans. Inf. Theory* **2005**, *51*, 1011–1023. [CrossRef]
9. Folland, G.B. Weyl manifolds. *J. Differ. Geom.* **1970**, *4*, 145–153. [CrossRef]
10. Kulkarni, R. Conformally flat manifolds. *Proc. Natl. Acad. Sci. USA* **1972**, *69*, 2675. [CrossRef] [PubMed]
11. Calin, O.; Matsuzoe, H.; Zhang, J. Generalizations of conjugate connections. In *Trends in Differential Geometry, Complex Analysis and Mathematical Physics*; World Scientific: Singapore, 2009; pp. 26–34.