Fisher information metric

Al-zahraa Mohamed Fahmy Farrag (alzahraa@aims.ac.za)
الزَهْرَاء مُحَمَدْ فَهْمِي فَرَاجّ حَسَنّ
African Institute for Mathematical Sciences (AIMS)

Supervised by: Doctor Jeff Murugan
University of Cape Town, South Africa
Supervised by: Doctor Jonathan Shock
University of Cape Town, South Africa

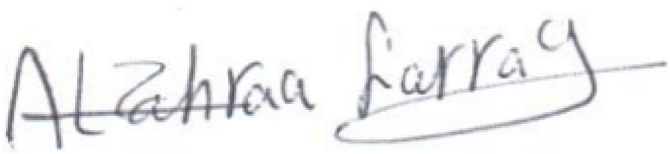
22 May 2014
Submitted in partial fulfillment of a structured masters degree at AIMS South Africa

Abstract

In this work, we undertake a preliminary exploration of the relationship between the Fisher information metric and an emergent space-time metric. In previous works, the anti–de Sitter metric usually emerges from instanton moduli spaces. U. Miyamoto and S. Yahikozawa proved in their paper *"Information metric from a linear sigma model"* that we can also obtain the flat Euclidean or Minkowskian metric as the Fisher metric from a non-trivial solution of the Klein-Gordon field. This is useful to study the relation between the space-time geometry and the information geometry (Miyamoto and Yahikozawa, 2012). I explain their paper in detail and define some of the concepts.

Declaration

I, the undersigned, hereby declare that the work contained in this research project is my original work, and that any work done by others or by myself previously has been acknowledged and referenced accordingly.



Al-zahraa Mohamed Fahmy Farrag, 22 May 2014

Contents

1. Introduction

In the Anti-de Sitter/Conformal Field Theory (AdS/CFT) correspondence, the equivalence of gravitational theory (i.e., the geometry of space-time) and Quantum Field Theory at the boundary of space-time is achieved partially. Researchers have shown that a lot of AdS geometry, where the gravitational theory is defined, can emerge as the information metric of the instanton moduli space of the boundary Yang-Mills theory.

In general, AdS/CFT correspondence has been used to study many branches of physics by translating strong coupling problems in those branches into more mathematically tractable problems in string theory and M-theory.

Although geometrical approaches have a lot of merit in statistical physics, the information metric seems like just a mathematical tool/notion that does not correspond at all to the measure of physical distance in space. One of the merits of information geometry is that at the Planckian regime ($\lesssim 10^{-33}$cm), where quantum gravitation effects have impact, the notion of space-time points and distance among them can be calculated statistically instead of using deterministic approach.

As a result, the techniques of information geometry could be used in describing space-times at the Planckian regime. This result seems like the expectation of the majority of physicists that something like non-commutative geometry will be useful in describing quantum geometry.

In such cases, where information geometry is applied to a field theory, space-time coordinates in a solution of the field theory are identified with random variables x^μ while the statistical parameters are given by θ^a. The Lagrangian density serves as the probability distribution. This approach has, thus for, yielded significant progress in understanding the emergence of space-time properties. For example, it is most surprising that a black-hole geometry, being asymptotically AdS, can also be obtained as the Fisher metric from the Yang-Mills instantons (Miyamoto and Yahikozawa, 2012).

After studying the Fisher metric of the instanton moduli spaces in the AdS/CFT correspondence, it would be important to study, from the general point of view, which kind of field theories (and their solutions) that can produce fundamental geometries (e.g. a flat space, an AdS space, and a black hole) as the information geometry. With regard to this, S. Yahikozawa showed that the AdS space can be obtained from the instanton solutions of non-linear sigma models, which was an example where the non-trivial Fisher metric is obtained from a field theory that is neither Abelian nor non-Abelian Yang-Mills theories (Miyamoto and Yahikozawa, 2012).

One of the simplest (but non-trivial) examples of this approach is obtaining an Euclidean flat space from a linear σ model. That is what we want to obtain.

2. Fisher information

2.1 What is Fisher information?

In statistical theory, **Fisher information** is a measure of the ability to estimate an unknown parameter $\vec{\theta}$ upon which the probability of an observable random variable $\vec{x}$ depends. While, in physical theory, Fisher information can be defined as a measure of the state of disorder of a system or phenomenon. According to the first definition, information (Fisher information) can be considered as a measure of the expected error e in a smart measurement (Lavis and Streater, 2002). The mean-square error e^2 in such an 'unbiased' estimator $\hat{\boldsymbol{\theta}}(\vec{x})$, obeys $\langle\hat{\boldsymbol{\theta}}(\vec{x})\rangle = \vec{\theta}$ [1], is given by

$$e^2 I \geqslant 1 \, , \tag{2.1.1}$$

where I is Fisher information given by

$$I = \int_\Omega d^D x \quad \left[\left(\frac{dp(\vec{x};\vec{\theta})}{d\vec{x}} \right)^2 \Big/ p(\vec{x};\vec{\theta}) \right] ,$$

where $\vec{x} = (x^0, x^1, \cdots, x^{D-1})$ is an arrangement of random variables distributed with the conditional probability density function (PDF) $p(\vec{x};\vec{\theta})$ (which is also the likelihood function for $\vec{\theta}$), and $\vec{\theta} = (\theta^0, \theta^1, \cdots, \theta^{N-1})$ is an arrangement of statistical parameters of $p(\vec{x};\vec{\theta})$. The domain $\Omega \in \mathbb{R}^D$ such that $\Omega = \underbrace{(-\infty,\infty) \times \ldots \times (-\infty,\infty)}_{D-times}$.

Equation (2.1.1) shows that the mean-square error (e^2) decreases as the information I increases, i.e the estimation quality is proportional to the information I.

By considering a set of unbiased estimators ($\hat{\boldsymbol{\theta}}(\vec{x})$),

$$\langle \hat{\boldsymbol{\theta}}(\vec{x}) - \vec{\theta} \rangle \equiv \int_\Omega d^D x \quad \left(\hat{\boldsymbol{\theta}}(\vec{x}) - \vec{\theta} \right) p(\vec{x};\vec{\theta}) = 0, \tag{2.1.2}$$

and differentiating with respect to θ, we get

$$\int_\Omega d^D x \quad \left(\hat{\boldsymbol{\theta}}(\vec{x}) - \vec{\theta} \right) \frac{\partial p}{\partial \vec{\theta}} - \int_\Omega d^D x \quad p = 0. \tag{2.1.3}$$

By using the normalization condition

$$\int_\Omega d^D x \quad p = 1,$$

[1] see the glossary

this becomes

$$\begin{aligned} \int_\Omega d^D x \quad \left(\hat{\boldsymbol{\theta}}(\vec{x}) - \vec{\theta}\right) \frac{\partial p}{\partial \vec{\theta}} - 1 = 0 \\ \int_\Omega d^D x \quad \left(\hat{\boldsymbol{\theta}}(\vec{x}) - \vec{\theta}\right) \frac{\partial p}{\partial \vec{\theta}} = 1 \ , \end{aligned} \tag{2.1.4}$$

then, using the identity,

$$\frac{\partial p}{\partial \vec{\theta}} = p \frac{\partial \ln p}{\partial \vec{\theta}} \ , \tag{2.1.5}$$

we find,

$$\begin{aligned} \int_\Omega d^D x \quad \left(\hat{\boldsymbol{\theta}}(\vec{x}) - \vec{\theta}\right) \ p \ \frac{\partial \ln p}{\partial \vec{\theta}} = 1 \\ \int_\Omega d^D x \quad \left[\left(\hat{\boldsymbol{\theta}}(\vec{x}) - \vec{\theta}\right) \ \sqrt{p}\right] \left[\sqrt{p} \ \frac{\partial \ln p}{\partial \vec{\theta}}\right] = 1 \\ \left[\int_\Omega d^D x \quad \left(\hat{\boldsymbol{\theta}}(\vec{x}) - \vec{\theta}\right) \ \sqrt{p}\right] \left[\int_\Omega d^D x \quad \sqrt{p} \ \frac{\partial \ln p}{\partial \vec{\theta}}\right] = 1 \\ \left[\int_\Omega d^D x \quad \left(\hat{\boldsymbol{\theta}}(\vec{x}) - \vec{\theta}\right) \ \sqrt{p}\right]^2 \left[\int_\Omega d^D x \quad \sqrt{p} \ \frac{\partial \ln p}{\partial \vec{\theta}}\right]^2 \geqslant 1 \\ \left[\int_\Omega d^D x \quad \left(\hat{\boldsymbol{\theta}}(\vec{x}) - \vec{\theta}\right)^2 \ p\right] \left[\int_\Omega d^D x \quad p \ \left(\frac{\partial \ln p}{\partial \vec{\theta}}\right)^2\right] \geqslant 1 \ . \end{aligned}$$

The first term is defined as the Fisher information I

$$\boxed{I \equiv I(\vec{\theta}) \equiv \int_\Omega d^D x \quad p \ \left(\frac{\partial \ln p}{\partial \vec{\theta}}\right)^2} , \tag{2.1.6}$$

while the second term will be the mean-square error,

$$\int_\Omega d^D x \quad \left(\hat{\boldsymbol{\theta}}(\vec{x}) - \vec{\theta}\right)^2 \ p \equiv e^2 .$$

The Equation (2.1.6) shows that in general $I \equiv I(\theta)$, which means that I depends on a fixed value of parameter θ. But there is an important exception to this rule.

2.1.1 The case of shift invariance. Assume that the number of dimensions of the space-time D is equal to the number of the parameters $N \to D = N$. Furthermore, assume that the probability density function obeys $p(\vec{x};\vec{\theta}) = p(\vec{x}-\vec{\theta})$. This means that the fluctuations in x from θ are invariant to the size of θ (Lavis and Streater, 2002).

By using the identity from Equation (2.1.5) in Equation (2.1.6),

$$I(\vec{\theta}) = \int_\Omega d^D x \quad \left[\left(\frac{\partial p(\vec{x}-\vec{\theta})}{\partial \vec{\theta}} \right)^2 \Big/ p(\vec{x}-\vec{\theta}) \right],$$

$$\because \frac{\partial}{\partial \vec{\theta}} = -\frac{\partial}{\partial(\vec{x}-\vec{\theta})}$$

$$I(\vec{\theta}) = \int_\Omega d^D x \quad \left[\left(\frac{\partial p(\vec{x}-\vec{\theta})}{\partial(\vec{x}-\vec{\theta})} \right)^2 \Big/ p(\vec{x}-\vec{\theta}) \right].$$

Since the statistical parameter θ is the shift of coordinates, the information can be written as

$$\boxed{I(0) = \int_\Omega d^D x \quad \left[\left(\frac{\partial p(\vec{x})}{\partial \vec{x}} \right)^2 \Big/ p(\vec{x}) \right]} . \tag{2.1.7}$$

The Equation (2.1.7) shows that the Fisher information is independent on θ.

2.2 Other definitions for Fisher information

Under certain conditions,

* We can consider the Fisher information as the variance (the second moment) of the score function,

$$\begin{aligned} V[s](\theta) &= E\left[\left(\frac{\partial}{\partial \vec{\theta}} \ln L(\vec{\theta};\vec{x}) \right)^2 \middle| \vec{\theta} \right] - \left(E\left[\frac{\partial}{\partial \vec{\theta}} \ln L(\vec{\theta};\vec{x}) \middle| \vec{\theta} \right] \right)^2 \\ &= \int_\Omega d^D x \quad \left(\frac{\partial}{\partial \vec{\theta}} \ln p(\vec{x};\vec{\theta}) \right)^2 p(\vec{x};\vec{\theta}) - 0 \qquad \text{(from Equation (5.0.1))} \\ &= \int_\Omega d^D x \quad \left(\frac{\partial}{\partial \vec{\theta}} \ln p(\vec{x};\vec{\theta}) \right)^2 p(\vec{x};\vec{\theta}) \equiv I(\vec{\theta}). \end{aligned}$$

* Also, we can consider the Fisher information as the negative of the expectation value for the first derivative of the score function with respect to the observation x, given θ.

Firstly, let's consider the first derivative of the score function

$$
\begin{aligned}
\frac{\partial}{\partial\theta}\Big(s(\theta)\Big) &= \frac{\partial}{\partial\theta}\left(\frac{\partial}{\partial\theta}\ln L(\vec{\theta};\vec{x})\right) \\
&= \frac{\partial}{\partial\theta}\left(\frac{\partial}{\partial\theta}\ln p(\vec{x};\vec{\theta})\right) && \text{(from Equation (5.0.3))} \\
&= \frac{\partial}{\partial\theta}\left(\frac{\frac{\partial}{\partial\theta}\,p(\vec{x};\vec{\theta})}{p(\vec{x};\vec{\theta})}\right) \\
&= \frac{\left(\frac{\partial^2}{\partial\theta^2}p(\vec{x};\vec{\theta})\right)p(\vec{x};\vec{\theta}) \;-\; \left(\frac{\partial}{\partial\theta}p(\vec{x};\vec{\theta})\right)^2}{\Big(p(\vec{x};\vec{\theta})\Big)^2} \\
&= \frac{\frac{\partial^2}{\partial\theta^2}p(\vec{x};\vec{\theta})}{p(\vec{x};\vec{\theta})} \;-\; \left(\frac{\frac{\partial}{\partial\theta}(\vec{x};p\vec{\theta})}{p(\vec{x};\vec{\theta})}\right)^2 \\
&= \frac{\frac{\partial^2}{\partial\theta^2}p(\vec{x};\vec{\theta})}{p(\vec{x};\vec{\theta})} \;-\; \left(\frac{\partial}{\partial\theta}\ln p(\vec{x};\vec{\theta})\right)^2 . && \text{(from Equation (2.1.5))}
\end{aligned}
$$

Then, consider the expectation value for the first derivative of the score function

$$
\begin{aligned}
E\left[\frac{\partial}{\partial\theta}\Big(s(\theta)\Big)\right] &\equiv E\left[\frac{\partial^2}{\partial\theta^2}\ln p(\vec{x};\vec{\theta})\middle|\,\theta\right] \\
&= E\left[\frac{\frac{\partial^2}{\partial\theta^2}p(\vec{x};\vec{\theta})}{p(\vec{x};\vec{\theta})} \;-\; \left(\frac{\partial}{\partial\theta}\ln p(\vec{x};\vec{\theta})\right)^2\middle|\,\theta\right] \\
&= E\left[\frac{\frac{\partial^2}{\partial\theta^2}p(\vec{x};\vec{\theta})}{p(\vec{x};\vec{\theta})}\middle|\,\theta\right] \;-\; E\left[\left(\frac{\partial}{\partial\theta}\ln p(\vec{x};\vec{\theta})\right)^2\middle|\,\theta\right] \\
&= 0 \;-\; E\left[\left(\frac{\partial}{\partial\theta}\ln p(\vec{x};\vec{\theta})\right)^2\middle|\,\theta\right] \\
&= -\int_\Omega d^Dx \quad p\,\left(\frac{\partial\ln p}{\partial\vec{\theta}}\right)^2 \equiv -I(\vec{\theta}) \\
\therefore\; I(\vec{\theta}) &= -\,E\left[\frac{\partial^2}{\partial\theta^2}\ln p(\vec{x};\vec{\theta})\middle|\,\theta\right]
\end{aligned}
$$

$$
\boxed{I(\vec{\theta}) = -\int_\Omega d^Dx\;\frac{\partial^2\ln p(\vec{x};\vec{\theta})}{\partial\theta^2}\;p(\vec{x};\vec{\theta})}\;. \tag{2.2.1}
$$

Therefore, Fisher information is the variance of the score function, or the expected value of the first derivative of the score function (with negative sign).

2.3 The relation between Fisher and other informations

2.3.1 Fisher information from Shannon information. :

The Shannon information is given by

$$H = -\int_{\Omega} d^D x \quad p(\vec{x};\vec{\theta})\ \ln p(\vec{x};\vec{\theta}).$$

The Hessian of the Shannon information is given by

$$\begin{aligned} h_{\,ab}(\theta) &= \frac{\partial^2}{\partial\theta^a\,\partial\theta^b} H \\ &= -\int_{\Omega} d^D x \quad \left[\frac{\partial^2 p(\vec{x};\vec{\theta})}{\partial\theta^a\,\partial\theta^b}\Big(1+\ln p(\vec{x};\vec{\theta})\Big) + \frac{1}{p(\vec{x};\vec{\theta})}\frac{\partial p(\vec{x};\vec{\theta})}{\partial\theta^a}\frac{\partial p(\vec{x};\vec{\theta})}{\partial\theta^b}\right]. \end{aligned}$$

The Hessian of the Shannon information is the same as the negative of Fisher Information matric if $\frac{\partial^2 p(\vec{x};\vec{\theta})}{\partial\theta^a\,\partial\theta^b} = 0$ for all $a, b, x,$ and θ.

For example, in two-dimension, if the probability density function given by

$$p(\vec{x};\vec{\theta}) = \theta_1 f_1(x) + \theta_2 f_2(x) + (1-\theta_1-\theta_2) f_3(x)\,,$$

where $f_1(x), f_2(x)$, and $f_3(x)$ are probability distributions.

$$\begin{aligned} \frac{\partial^2 p(\vec{x};\vec{\theta})}{\partial\theta^a\,\partial\theta^b} &= \frac{\partial^2}{\partial\theta^a\,\partial\theta^b}\left(\theta_1 f_1(x) + \theta_2 f_2(x) + (1-\theta_1-\theta_2) f_3(x)\right) \\ &= 0. \end{aligned}$$

2.3.2 Relation of Fisher information to Kullback-Leibler entropy. :

Kullback-Leibler entropy, also known as cross entropy or relative entropy because it is a ratio between PDFs (for now, two PDFs) is usually used to measure the distance between two PDFs. In one-dimension[2] it is given by

$$G = -\int dx\ \ p(x)\ \ln\left(\frac{p(x)}{r(x)}\right).$$

The relative entropy between $p(x)$ and $p(x+\Delta x)$ is given by

$$\begin{aligned} G &= -\int dx\ \ p(x)\ \ln\left(\frac{p(x+\Delta x)}{p(x)}\right) \\ &\equiv -\frac{\Delta x^2}{2} I. \end{aligned}$$

[2]The multidimensional relative entropy can be used as **the mutual information**.
In the case of two-dimension (x, y), the relative entropy can be written as

$$G = \iint dy\ dx\ \ p(x,y)\ \ln\left(\frac{p(x,y)}{p(x)p(y)}\right),$$

where $p(x, y)$ is joint probability density function of x and y, $p(x)$ is probability density functions of x and $p(y)$ is probability density functions of y.

This equation means Fisher information is proportional to the relative entropy between $p(x)$ and $p(x+\Delta x)$ (the shifted version of $p(x)$).

2.4 Fisher information metric

Fisher information metric (Information metric) can be understood to be a particular Riemannian metric which is defined on a smooth statistical manifold.

2.4.1 Information metric for Continuous PDF. Information metric is given by

$$g_{ab}(\theta) = \int_\Omega d^D x \quad \frac{\partial \ln p(\vec{x};\vec{\theta})}{\partial \theta^a} \frac{\partial \ln p(\vec{x};\vec{\theta})}{\partial \theta^b} \, p(\vec{x};\vec{\theta}) \, ,$$

where $x^\mu = (x^0, x^1, \ldots, x^{D-1})$ is an arrangement of random variables which distributed with the conditional probability density function $p(\vec{x};\vec{\theta})$ and $\theta^a = (\theta^0, \theta^1, \ldots, \theta^{N-1})$ is an arrangement of statistical parameters of a probability distribution. Also θ is used as a coordinate on the Riemannian manifold. The labels (a) and (b) index the local coordinate axes on the manifold.

The two parameters θ^a and θ^b are **orthogonal parameters** because the inner product

$$\langle \theta^a | \theta^b \rangle = \theta^a . \, \theta^b \delta_{ab},$$

that means the element of the row number a and column number b of the Fisher information metric is zero except when $a = b$.

2.4.1.1 Remark. From Equation (2.2.1) we can rewrite information metric as

$$g_{ab}(\theta) = \int_\Omega d^D x \quad \frac{\partial^2 \left(-\ln p(\vec{x};\vec{\theta}) \right)}{\partial \theta^a \, \partial \theta^b} \, p(\vec{x};\vec{\theta}) \, .$$

But that only under certain conditions.

2.4.1.2 Remark. From Equation (2.1.7) we can rewrite Fisher information metric as

$$g_{ab}(\theta) \equiv g_{ab}(0) = \int_\Omega d^D x \quad \left[\frac{\partial p(\vec{x})}{\partial x^a} \frac{\partial p(\vec{x})}{\partial x^b} \Big/ p(\vec{x}) \right],$$

thus:

$$\boxed{g_{ab}(0) = \int_\Omega d^D x \quad \left[\partial_{x^a} p(\vec{x}) \;\; \partial_{x^b} p(\vec{x}) \Big/ p(\vec{x}) \right]} \, . \tag{2.4.1}$$

These equations show that there are Cartesian coordinates for a flat space, in which the information metric is not depend on θ.

2.4.2 Information metric for discrete PDF. We can rewrite Fisher information metric as

$$g_{ab}(\theta) = \sum_{i=0}^{D-1} \frac{\partial \ln p_i(\vec{x};\vec{\theta})}{\partial \theta^a} \frac{\partial \ln p_i(\vec{x};\vec{\theta})}{\partial \theta^b} \; p_i(\vec{x};\vec{\theta})$$
$$= \sum_{i=0}^{D-1} \left[\frac{\partial p_i(\vec{x};\vec{\theta})}{\partial \theta^a} \frac{\partial p_i(\vec{x};\vec{\theta})}{\partial \theta^b} \Big/ p_i(\vec{x};\vec{\theta}) \right] . \tag{2.4.2}$$

This form of information metric is useful to prove that we can obtain an information metric as flat Euclidean metric.

2.4.3 Information metric as flat Euclidean metric. The flat Euclidean metric can be obtained as the Fisher metric after changes of variable.

The flat metric on the surface of the N-dimensional sphere is given by $h = \sum_i dx_i \, dx_i$. If we change the sphere condition $(\sum_i x_i^2 = 1)$ to the normalization condition $(\sum_i p_i(\vec{\theta}) = 1) \Rightarrow \; p_i = x_i^2$.

The metric will then be written as

$$\begin{aligned} h &= \sum_i \; dx_i \; dx_i = \sum_i d\sqrt{p_i(\vec{\theta})} \; d\sqrt{p_i(\vec{\theta})} \\ &= \sum_i \frac{dp_i(\vec{\theta})}{2\sqrt{p_i(\vec{\theta})}} \frac{dp_i(\vec{\theta})}{2\sqrt{p_i(\vec{\theta})}} \\ &= \frac{1}{4} \sum_i \frac{dp_i(\vec{\theta}) \, dp_i(\vec{\theta})}{p_i(\vec{\theta})} \\ &= \sum_i \sum_a \sum_b \left[\frac{\partial p_i(\vec{\theta})}{\partial \theta^a} \frac{\partial p_i(\vec{x};\vec{\theta})}{\partial \theta^b} \Big/ p_i(\vec{x};\vec{\theta}) \right] \quad d\theta^a \, d\theta^b \\ &= \sum_i \left[\frac{\partial p_i(\vec{x};\vec{\theta})}{\partial \theta^a} \frac{\partial p_i(\vec{x};\vec{\theta})}{\partial \theta^b} \Big/ p_i(\vec{x};\vec{\theta}) \right] . \end{aligned}$$

From the Equation (2.4.2),

$$h \equiv g_{ab}(\theta).$$

$\therefore$ The flat Euclidean metric can be obtained as information metric.

2.4.4 Properties of Fisher information metric . To summarize them the Fisher information metric satisfies:

1. It satisfying the requirements for Riemannian metric (i.e. symmetric, positive-definite and bilinear).
2. It expresses information inequality (i.e. the bound of the ability to estimate parameter).
3. It is an inner product of tangent vectors so, it satisfy conditions of the inner product.
4. It is invariant under reparametrizations of the sample space $\vec{x}$ (Wagenaar, 1998).

5. It is covariant under reparametrizations of the parameter space $\vec{\theta}$ (Wagenaar, 1998).
6. It is often used to measure the difference between two probability distributions by calculating the Rao distance $|g_{ab}(\theta)d\theta^a d\theta^b|^{1/2}$.

3. Emergent flat space from scalar field

3.1 Massive Klein-Gordon scalar field

The equation of motion of a D-dimensional classical scalar field ϕ is given by

$$\frac{1}{c^2}\frac{\partial^2}{\partial t^2}\phi - \nabla^2\phi + \frac{m^2c^2}{\hbar^2}\phi = 0 \ .$$

which is the massive Klein-Gordon equation (the relativistic version of the Schrödinger equation). It can be written in natural units as

$$\frac{\partial^2}{\partial t^2}\phi - \nabla^2\phi + m^2\phi = 0 \ , \tag{3.1.1}$$

or in covariant notation as

$$(\Box + m^2)\phi = 0,$$

where $\Box := \frac{\partial^2}{\partial t^2} - \nabla^2$ is the d'Alembert operator. By using the D-dimensional Minkowski metric

$$\eta^{\mu\nu} = \eta_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & -1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & -1 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \end{pmatrix},$$

the Equation (3.1.1) can be written as

$$\eta^{\mu\nu}\partial_\mu\phi\partial_\nu\phi + m^2\phi^2 = 0.$$

This equation describes a scalar field in D space-time dimension. In the linear sigma model (Miyamoto and Yahikozawa, 2012) a non-trivial solution of the massive Klein-Gordon equation is given by

$$\phi = Ae^{\left(\frac{-m}{\sqrt{D-2}}|\vec{x}|\right)}, \qquad |\vec{x}| := \sum_{\mu=0}^{D-1}|x^\mu|, \qquad D \geqslant 3 \ ^{1}, \tag{3.1.2}$$

which is solitary wave (the Figure 3.1 shows an example), where A is an arbitrary constant and m is the mass of scalar field. The Lagrangian density for the Klein-Gordon scalar field ϕ is

$$\mathscr{L}(x) = \frac{1}{2}\,\eta^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - V(\phi),$$

[1] for the massive case $(m > 0)$, if $D \leqslant 2$ the solution will not exist.

with potential $V(\phi) = m^2\phi^2$, so that we can write

$$\mathscr{L}(x) = \frac{1}{2}\left(\eta^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - m^2\phi^2\right). \tag{3.1.3}$$

By substituting ϕ into Equation (3.1.3)

$$\begin{aligned}
\mathscr{L}(x) &= \frac{1}{2}\left((\partial_{x^0}\phi)^2 - \left(\sum_{\mu=1}^{D-1}\partial_{x^\mu}\phi\right)^2 - m^2\phi^2\right)\\
&= \frac{1}{2}\left(\left(Ae^{\left(\frac{-m}{\sqrt{D-2}}|\vec{x}|\right)}\left(\frac{-m}{\sqrt{D-2}}\right)\right)^2 - (D-1)\left(Ae^{\left(\frac{2m}{\sqrt{D-2}}|\vec{x}|\right)}\left(\frac{-m}{\sqrt{D-2}}\right)\right)^2 - m^2\left(Ae^{\left(\frac{-m}{\sqrt{D-2}}|\vec{x}|\right)}\right)^2\right)\\
&= \frac{1}{2}\left(-(D-2)\left(Ae^{\left(\frac{-m}{\sqrt{D-2}}|\vec{x}|\right)}\left(\frac{-m}{\sqrt{D-2}}\right)\right)^2 - m^2\left(Ae^{\left(\frac{-m}{\sqrt{D-2}}|\vec{x}|\right)}\right)^2\right)\\
&= -\frac{1}{2}\left((D-2)\left(Ae^{\left(\frac{-m}{\sqrt{D-2}}|\vec{x}|\right)}\left(\frac{-m}{\sqrt{D-2}}\right)\right)^2 + m^2\left(Ae^{\left(\frac{-m}{\sqrt{D-2}}|\vec{x}|\right)}\right)^2\right)\\
&= -\frac{1}{2}\left((D-2)A^2e^{\left(\frac{-2m}{\sqrt{D-2}}|\vec{x}|\right)}\left(\frac{-m}{\sqrt{D-2}}\right)^2 + m^2A^2e^{\left(\frac{-2m}{\sqrt{D-2}}|\vec{x}|\right)}\right)\\
&= -\frac{1}{2}A^2e^{\left(\frac{-2m}{\sqrt{D-2}}|\vec{x}|\right)}\left((D-2)\left(\frac{m^2}{D-2}\right) + m^2\right)\\
&= -A^2m^2e^{\left(\frac{-2m}{\sqrt{D-2}}|\vec{x}|\right)}.
\end{aligned} \tag{3.1.4}$$

3.2 Emergent information metric

3.2.1 Lagrangian density as probability distribution. Miyamoto and Yahikozawa regarded the negative of the Lagrangian density for the Klein-Gordon field evaluated on a solution as the probability distribution,

$$p(\vec{x} - \vec{\theta}) := -\mathscr{L}(\vec{x} - \vec{\theta}).$$

From Equation (3.1.4)

$$p(\vec{x} - \vec{\theta}) = A^2m^2e^{\left(\frac{-2m}{\sqrt{D-2}}|\vec{x}-\vec{\theta}|\right)}. \tag{3.2.1}$$

To evaluate the constant A we will use the normalization condition for the probability density function $P(\vec{x};\vec{\theta})$,

$$
\begin{aligned}
\int_{\Omega} d^D x \quad p(\vec{x}-\vec{\theta}) &= 1 \\
\int_{\Omega} d^D x \quad m^2 A^2 e^{\left(\frac{-2m}{\sqrt{D-2}}|\vec{x}-\vec{\theta}|\right)} &= 1 \\
m^2 A^2 \left[\int_{\theta}^{\infty} \cdots \int d^D x \quad e^{\left(\frac{-2m}{\sqrt{D-2}}(\vec{x}-\vec{\theta})\right)} + \int_{-\infty}^{\theta} \cdots \int d^D x \quad e^{\left(\frac{-2m}{\sqrt{D-2}}(-\vec{x}+\vec{\theta})\right)} \right] &= 1 \\
m^2 A^2 \left[e^{\left(\frac{-2m}{\sqrt{D-2}}(x-\vec{\theta})\right)} \Bigg|_{\theta}^{\infty} \left(-\frac{\sqrt{D-2}}{2m}\right) + e^{\left(\frac{2m}{\sqrt{D-2}}(\vec{x}-\vec{\theta})\right)} \Bigg|_{-\infty}^{\theta} \left(\frac{\sqrt{D-2}}{2m}\right) \right]^D &= 1 \\
m^2 A^2 \left[(0-1)\left(-\frac{\sqrt{D-2}}{2m}\right) + (1-0)\left(\frac{\sqrt{D-2}}{2m}\right) \right]^D &= 1 \\
m^2 A^2 \left[\left(\frac{\sqrt{D-2}}{2m}\right) + \left(\frac{\sqrt{D-2}}{2m}\right) \right]^D &= 1 \\
m^2 A^2 \left[\frac{\sqrt{D-2}}{m} \right]^D &= 1 \\
A^2 \left[\frac{(\sqrt{D-2})^D}{m^{D-2}} \right] &= 1 \\
\Rightarrow A^2 = \frac{m^{D-2}}{(\sqrt{D-2})^D} & \\
A^2 = \frac{m^{D-2}}{(D-2)^{D/2}} & \\
\therefore \boxed{A = \frac{m^{(D-2)/2}}{(D-2)^{D/4}}} & .
\end{aligned} \tag{3.2.2}
$$

3.2.1.1 Remark. Considering the link between the physical/geometrical meaning of θ and the mass m by calculating the expectation value, the variance and the standard deviation.

* The expectation value of variable x^μ is given by

$$
\begin{aligned}
E\left[x^\mu\right](\theta) &= \int_{\Omega} d^D x \, x^\mu \, p(\vec{x}-\vec{\theta}) \\
&= -\int_{\Omega} d^D x \, x^\mu \, \mathscr{L}(\vec{x}-\vec{\theta}) \\
&= E\left[x^\mu\right](0) - \theta^\mu \int_{\Omega} d^D x \;\; \mathscr{L}(\vec{x}) \\
&= E\left[x^\mu\right](0) + \theta^\mu \int_{\Omega} d^D x \;\; p(\vec{x}).
\end{aligned}
$$

From the normalization condition,

$$E\left[x^{\mu}\right](\theta) = E\left[x^{\mu}\right](0) + \theta^{\mu},$$

or since $x^{\mu}\, p(\vec{x})$ is an odd function, $E\left[x^{\mu}\right](0) = 0.$

Consequently,

$$\boxed{E\left[x^{\mu}\right](\theta) = \theta^{\mu}}\ . \tag{3.2.3}$$

* The variance of variable x^{μ} is given by

$$\begin{aligned}
V\left[x^{\mu}\right](\theta) &= E\left[(x^{\mu})^2\right](\theta) \ -\ \left(E\left[x^{\mu}\right](\theta)\right)^2 \\
&= \int_{\Omega} d^D x\ (x^{\mu})^2\ p(\vec{x}-\vec{\theta}) \ -\ (\theta^{\mu})^2 \\
&= E\left[(x^{\mu})^2\right](0) \ +\ \left(\theta^{\mu}\int_{\Omega} d^D x\ \ p(\vec{x})\right)^2 - (\theta^{\mu})^2 \\
&= E\left[(x^{\mu})^2\right](0) \ + (\theta^{\mu})^2 - (\theta^{\mu})^2 \\
&= E\left[(x^{\mu})^2\right](0) \\
&= \frac{1}{2(D-2)^{(D-6)/2} m^2}\,.
\end{aligned} \tag{3.2.4}$$

* The standard deviation $\sigma_{x^{\mu}}$ is the square root of variance of the variable x^{μ}.

From Equation (3.2.4) it becomes,

$$\boxed{\sigma_{x^{\mu}} = \frac{1}{\sqrt{2}(D-2)^{(D-6)/4} m}}\ . \tag{3.2.5}$$

The Equations (3.2.3) and (3.2.5) show that the distribution is centered at $x^{\mu} = \theta^{\mu}$ and localized with the scale m^{-1}.

For example, considering the probability distribution in 3-dimensions.

From Equation (3.2.1)

$$p(\vec{x}-\vec{\theta})\big|_{D=3} = m^3 e^{\left(-2m|\vec{x}-\vec{\theta}|\right)}.$$

The standard deviation from Equation (3.2.5) becomes

$$\sigma\big|_{D=3} = \frac{1}{\sqrt{2}m} \ \approx m^{-1}.$$

Assume that the mass $m = 1$ and the statistical parameters $\theta^0 = 0,\ \theta^1 = 1.$

From Equation (3.2.3)

$$\begin{aligned}
(x^0, x^1) &= (\theta^0, \theta^1) \\
&= (0, 1).
\end{aligned}$$

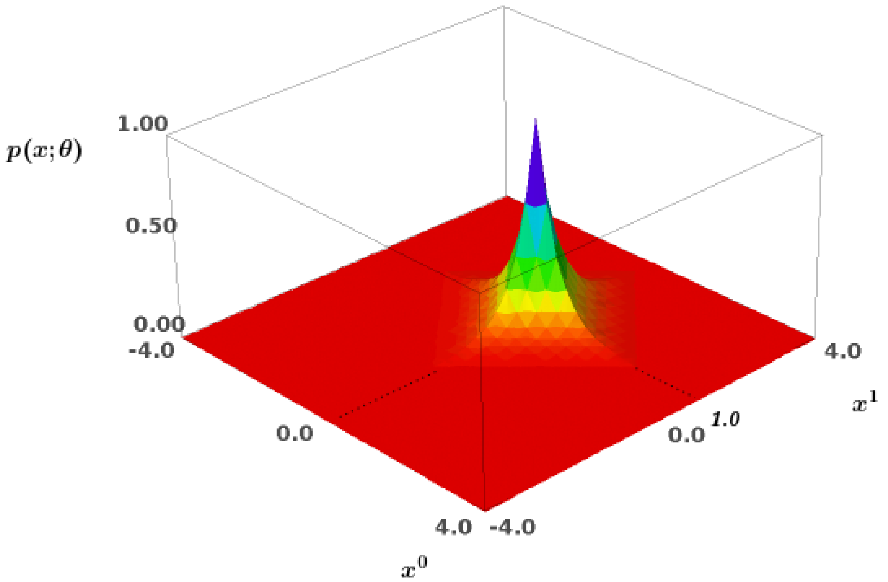


Figure 3.1: Probability distribution for the D = 3

The figure 3.1 shows that the distribution is centred at $(x^0, x^1) = (\theta^0, \theta^1) = (0, 1)$ and, it is localized with the scale $m^{-1} = 1$.

Calculating information metric.

From Equation (3.2.1)

$$p(\vec{x}) = A^2 m^2 e^{\left(\frac{-2m}{\sqrt{D-2}}|\vec{x}|\right)}.$$

$$\begin{aligned}
\partial_{x^a}\, p(\vec{x}) &= A^2 m^2 \partial_{x^a}\, e^{\left(\frac{-2m}{\sqrt{D-2}}|\vec{x}|\right)} \\
&= A^2 m^2 \left(\frac{-2m}{\sqrt{D-2}}\right)\left(\frac{x^a}{|x^a|}\right) e^{\left(\frac{-2m}{\sqrt{D-2}}|\vec{x}|\right)}. \\
\partial_{x^b}\, p(\vec{x}) &= A^2 m^2 \partial_{x^b}\, e^{\left(\frac{-2m}{\sqrt{D-2}}|\vec{x}|\right)} \\
&= A^2 m^2 \left(\frac{-2m}{\sqrt{D-2}}\right)\left(\frac{x^b}{|x^b|}\right) e^{\left(\frac{-2m}{\sqrt{D-2}}|\vec{x}|\right)}\ .
\end{aligned}$$

Because of the orthogonality of parameters,

$$\begin{aligned}
\partial_{x^a}\, p(\vec{x})\, \partial_{x^b}\, p(\vec{x}) &= A^2 m^2 \left(\frac{-2m}{\sqrt{D-2}}\right)\left(\frac{x^a}{|x^a|}\right) e^{\left(\frac{-2m}{\sqrt{D-2}}|\vec{x}|\right)} A^2 m^2 \left(\frac{-2m}{\sqrt{D-2}}\right)\left(\frac{x^b}{|x^b|}\right) e^{\left(\frac{-2m}{\sqrt{D-2}}|\vec{x}|\right)} \quad \delta_{ab} \\
&= A^4 m^4 \left(\frac{4m^2}{D-2}\right)\left(\frac{x^a}{|x^a|}\right)\left(\frac{x^b}{|x^b|}\right) e^{\left(\frac{-4m}{\sqrt{D-2}}|\vec{x}|\right)} \quad \delta_{ab}\ .
\end{aligned}$$

From Equation (2.4.1)

$$
\begin{aligned}
g_{ab}(0) &= \int_\Omega d^Dx \; \frac{A^4 m^4 \left(\frac{4m^2}{D-2}\right)\left(\frac{x^a}{|x^a|}\right)\left(\frac{x^b}{|x^b|}\right) e^{\left(\frac{-4m}{\sqrt{D-2}}|\vec{x}|\right)}}{A^2 m^2 e^{\left(\frac{-2m}{\sqrt{D-2}}|\vec{x}|\right)}} \;\; \delta_{ab} \\
&= \int_\Omega d^Dx \; A^2 m^2 \left(\frac{4m^2}{D-2}\right)\left(\frac{x^a}{|x^a|}\right)\left(\frac{x^b}{|x^b|}\right) e^{\left(\frac{-2m}{\sqrt{D-2}}|\vec{x}|\right)} \;\; \delta_{ab} \\
&= A^2 m^2 \left(\frac{4m^2}{D-2}\right) \int_\Omega d^Dx \; \left(\frac{x^a}{|x^a|}\right)\left(\frac{x^b}{|x^b|}\right) e^{\left(\frac{-2m}{\sqrt{D-2}}|\vec{x}|\right)} \;\; \delta_{ab} \; .
\end{aligned} \tag{3.2.6}
$$

By substitute A from Equation (3.2.2) into Equation, (3.2.6)

$$
\begin{aligned}
g_{ab}(0) &= \frac{4m^{D+2}}{(D-2)^{(D+2)/2}} \int_\Omega d^Dx \; \left(\frac{x^a}{|x^a|}\right)\left(\frac{x^b}{|x^b|}\right) e^{\left(\frac{-2m}{\sqrt{D-2}}|\vec{x}|\right)} \;\; \delta_{ab} \\
&= \frac{4m^{D+2}}{(D-2)^{(D+2)/2}} \left[\int_0^\infty dx \; e^{\left(\frac{-2m}{\sqrt{D-2}}x\right)} + \int_{-\infty}^0 dx \; e^{\left(\frac{2m}{\sqrt{D-2}}x\right)}\right]^D \;\; \delta_{ab} \\
&= \frac{4m^{D+2}}{(D-2)^{(D+2)/2}} \left[\left(\frac{\sqrt{D-2}}{2m}\right) + \left(\frac{\sqrt{D-2}}{2m}\right)\right]^D \;\; \delta_{ab} \\
&= \frac{4m^{D+2}}{(D-2)^{(D+2)/2}} \left(\frac{\sqrt{D-2}}{m}\right)^D \;\; \delta_{ab} \\
&= \frac{4m^2}{D-2} \; \delta_{ab} \; .
\end{aligned} \tag{3.2.7}
$$

Hence, the flat space-time metric has been obtained from the linear sigma model as information metric. By multiplying the coordinates θ^a by the scaling factor $\frac{2m}{\sqrt{D-2}}$ we can obtain normalized metric as following

$$
\theta^a \rightarrow \bar{\theta^a} = \left(\frac{2m}{\sqrt{D-2}}\right) \; \theta^a
$$

From equation (3.2.3)

$$
\begin{aligned}
x^a &\rightarrow \bar{x^a} = \left(\frac{2m}{\sqrt{D-2}}\right) \; x^a \\
\Rightarrow \frac{\partial}{\partial x^a} &\rightarrow \left(\frac{2m}{\sqrt{D-2}}\right) \; \frac{\partial}{\partial x^a} \\
\therefore \; g_{ab} &\rightarrow \bar{g_{ab}} = \delta_{ab}
\end{aligned}
$$

Therefore the information metric is proportional to the Kronecker delta function δ_{ab}.

3.2.2 Gaussian distribution as probability distribution. It is very hard to find Lagrangian in form Gaussian function to describe physically reasonable relativistic field, but it will be highly non-trivial.

Gaussian distribution given by

$$G(x) = \pi^{-D/2}\; e^{-\left(\vec{x}-\vec{\theta}\right)^2}.$$

Let's regard Gaussian distribution as probability distribution

$$\begin{aligned} p(\vec{x}-\vec{\theta}) &= \pi^{-D/2}\; e^{-\left(\vec{x}-\vec{\theta}\right)^2} \\ p(\vec{x}) &= \pi^{-D/2}\; e^{-(\vec{x})^2}. \end{aligned}$$

thus,

$$\begin{aligned} \partial_{x^a}\; p(\vec{x}) &= \pi^{-D/2}\; \partial_{x^a}\; e^{-(\vec{x})^2} \\ &= \left(\pi^{-D/2}\right)(-2x^a)\, e^{-(\vec{x})^2}. \\ \partial_{x^b}\; p(\vec{x}) &= \pi^{-D/2}\; \partial_{x^b}\; e^{-(\vec{x})^2} \\ &= \left(\pi^{-D/2}\right)\left(-2x^b\right) e^{-(\vec{x})^2}. \end{aligned}$$

Because of the orthogonality of parameters,

$$\begin{aligned} \partial_{x^a}\, p(\vec{x})\, \partial_{x^b}\, p(\vec{x}) &= \left(\pi^{-D/2}\right)(-2x^a)\, e^{-(\vec{x})^2} \left(\pi^{-D/2}\right)\left(-2x^b\right) e^{-(\vec{x})^2} \quad \delta_{ab} \\ &= 4\left(\pi^{-D}\right) x^a x^b e^{-2(\vec{x})^2} \quad \delta_{ab}\;\; . \end{aligned}$$

From Equation (2.4.1)

$$\begin{aligned} \Rightarrow g_{ab}(0) &= \int_\Omega d^D x\; \frac{4\left(\pi^{-D}\right) x^a x^b e^{-2(\vec{x})^2}}{\pi^{-D/2}\; e^{-(\vec{x})^2}} \quad \delta_{ab} \\ &= 4\left(\pi^{-D/2}\right) \int_\Omega d^D x \quad x^a x^b e^{-(\vec{x})^2} \quad \delta_{ab} \\ &= 4\left(\pi^{-D/2}\right)\left[\int_{-\infty}^{\infty} dx^\mu \quad (x^\mu)^2 e^{-(x^\mu)^2}\right] \left[\prod_{i=0}^{D-2} \int_{-\infty}^{\infty} dx^i \quad e^{-\left(x^i\right)^2}\right] \delta_{ab}. \end{aligned}$$

By changing the variables $x^2 \to y$

$$
\begin{aligned}
&= 4\left(\pi^{-D/2}\right)\left[\int_{-\infty}^{\infty} dy \quad (y)^{1/2}\, e^{-y}\right]\left[\prod_{i=0}^{D-2}\int_{-\infty}^{\infty} dy_i\ (y_i)^{-1/2}\ \ e^{-y_i}\right]\delta_{ab} \\
&= 4\left(\pi^{-D/2}\right)\left[\Gamma\left(\frac{3}{2}\right)\ {}^{2}\right]\left[\prod_{i=0}^{D-2}\Gamma\left(\frac{1}{2}\right)\right]\delta_{ab} \\
&= 4\left(\pi^{-D/2}\right)\left[\frac{\sqrt{\pi}}{2}\right]\ \left[\left(\sqrt{\pi}\right)^{(D-1)}\right]\delta_{ab} \\
&= 2\ \delta_{ab}\ .
\end{aligned}
$$

Therefore the information metric is proportional to the Kronecker delta function δ_{ab} .

[2] $\Gamma(t) = \int_0^\infty y^{t-1} e^{-y}\, dy$

4. Concluding remarks

We have obtained a flat space-time metric as Fisher metric which is definite and non-degenerate. This realization is useful to study the correspondence between space-time geometry and the information geometry. We showed that this metric is proportional to the Kronecker delta function δ_{ab}.

Future work:

It will be very useful to study the instanton moduli spaces of more basic field theories including those of grassmannian models in both the non super symmetric and supersymmetric cases. Once we have found these moduli spaces we will then be able to calculate the fisher information metric which gives the emergent geometry hidden within the field theory.

5. Concepts/definitions

* **AdS/CFT correspondence**

The Anti-de Sitter/conformal field theory conjecture (AdS/CFT) is a correspondence between the Anti-de Sitter(AdS)(e.g. string theory and M-theory) and the Conformal Field Theories (CFT) (e.g. Yang–Mills theories), which is often called a duality because it is an equivalence between two different "dual" descriptions of the same physics (Klebanov and Maldacena, 2009).

* **Anti-de Sitter space**

It is a space-time geometry, related to a particular solution of Einstein's equation (Klebanov and Maldacena, 2009).

For example, the Figure 5.1 shows the difference between the Euclidean space and AdS space in 3-dimensions.

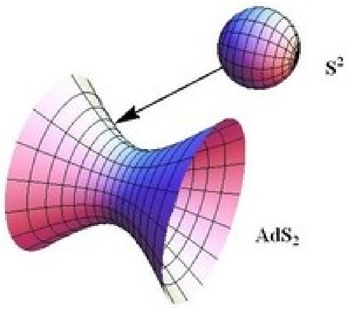


Figure 5.1: Euclidean space and AdS space
credited to: Wiegandt, Dipl.-Physiker Konstantin

* **Conformal field theory**

It refers to a particular type of quantum field theory, which is symmetric and mathematically well-behaved.

* **Estimator**

It is an optimal function of the outcome data, which estimates the parameter.

For example, consider an estimator (sample mean) for the expectation value $E[x]$;

$$\mu = E[x] \qquad \textit{is the parameter.}$$

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i \equiv \mu \qquad \textit{is the estimator.}$$

* **Expectation value**

In probability theory, for a function f, of a random variable x, distributed with a probability density function $p(x;\theta)$, the expectation value is given by

$$E[f(x)](\theta) = \int_{-\infty}^{\infty} f(x)\; p(x;\theta)\; dx\;.$$

For example, the expectation value for the score function with respect to the observation x, given θ, under certain conditions, it is given by

$$E[s](\theta) = E\left[\frac{\partial}{\partial\theta}\ln L(\theta;X)\Big|\theta\right] = \int_{-\infty}^{\infty} p(x;\theta)\;\frac{\partial}{\partial\theta}\ln L(\theta;X)\;dx,$$
$$\because L(\theta;X) = p(x;\theta)\;,$$

so,

$$\begin{aligned} E[s](\theta) &= \int_{-\infty}^{\infty}\frac{1}{p(x;\theta)}\;\frac{\partial p(x;\theta)}{\partial\theta}\;p(x;\theta)\;dx \\ &= \int_{-\infty}^{\infty}\frac{\partial p(x;\theta)}{\partial\theta}\;dx \\ &= \frac{\partial}{\partial\theta}\int_{-\infty}^{\infty}\;p(x;\theta)\;dx \\ &= \frac{\partial}{\partial\theta}1 = 0\;. \end{aligned} \tag{5.0.1}$$

And in the same way,

$$\begin{aligned} E\left[\frac{\partial^2}{\partial\theta^2}\ln L(\theta;X)\Big|\theta\right] &= \int_{-\infty}^{\infty} p(x;\theta)\;\frac{\partial^2}{\partial\theta^2}\ln L(\theta;X)\;dx \\ &= \int_{-\infty}^{\infty}\frac{1}{p(x;\theta)}\;\frac{\partial^2 p(x;\theta)}{\partial\theta^2}\;p(x;\theta)\;dx \\ &= \int_{-\infty}^{\infty}\frac{\partial^2 p(x;\theta)}{\partial\theta^2}\;dx \\ &= \frac{\partial^2}{\partial\theta^2}\int_{-\infty}^{\infty}\;p(x;\theta)\;dx \\ &= \frac{\partial^2}{\partial\theta^2}1 = 0\;. \end{aligned} \tag{5.0.2}$$

* **Flat space-time**

In the General Theory of Relativity, the geometry of space-time is curved geometrically but in the field of weak gravity (considered in the special theory of relativity), the geometry of space-time becomes flat.

* **Information geometry**

It is a mathematical branch in which we can apply the technique of differential geometry to the field of probability theory by studying the points of a Riemannian manifold as probability distributions for a statistical model. This technique is used to describe the space-time of probability distributions for a random variable $\vec{x}$, at the Planckian regime.The probability $p(\vec{x};\vec{\theta})$, must be differentiable with respect to θ, and an invertible function of θ.

* **Likelihood function**

Let's assume that the observable outcome of an experiment is $X \equiv (x_1, x_2, \ldots, x_n)$, which is considered as a sample from a joint probability density function with parameters (θ) $p(x_1, x_2, \ldots, x_n; \theta)$. So, the likelihood function is given by

$$L(\theta; X) = p(x_1, x_2, \ldots, x_n; \theta)\,. \tag{5.0.3}$$

* **Riemannian manifold**

It is a manifold which has Riemannian metric.

For example, in the smooth manifold S^2, we can obtain Riemannian metric f_p, from embedding space of scalar product, which is Euclidean scalar product $\in \mathbb{R}^3$:

$$\begin{aligned} f_p : T_pS^2 \times T_pS^2 &\to \mathbb{R} \\ (X_p, Y_p) &\mapsto f_p(X_p, Y_p) = \langle X_p, Y_p \rangle. \end{aligned}$$

$\therefore$ S^2 is Riemannian manifold.

* **Riemannian metric**

A Riemannian metric f, on a smooth manifold M, is a smooth metric which is a section of symmetric positive-definite inner product of the tangent spaces T_p. In other words, a Riemannian metric f on M, is a differentiable map $p \mapsto f_p : T_pM \times T_pM \to \mathbb{R}$, such that f_p should be bilinear, symmetric and positive-definite.

I.e,

$$\begin{aligned} f_p(aX_1 + X_2, Y) &= af_b(X_1, Y) + bf_b(X_2, Y)\,, \\ f_p(X, Y) &= f_p(Y, X)\,, \\ f_p(X, X) &> 0 \quad \textit{for } X \neq 0\,, \end{aligned}$$

where X , X_1, X_2 and Y are vector fields on M (Pfeiffer and Atariah).

* **Score function**

Also known as the efficient score, it is the first derivative of the logarithm of the likelihood function $\ln\ L(\theta; X)$ with respect to θ (Little and Rubin, 2002). So, it is a measurement of sensitivity of the dependence of a likelihood function on θ. It is given by

$$s(\theta) = \frac{\partial \ln L(\theta; X)}{\partial \theta}\,.$$

* **Smart measurement**

Suppose we have observed data $\vec{x} = (x^0, x^1, \cdots, x^{D-1})$. We want to find a function $\hat{\theta}$ of the data to estimate the parameter θ. The overall measurement procedure is called smart measurement. It is a better estimate of θ than is any one of the data observables Lavis and Streater (2002).

* **Solitary wave**

It is a wave, which moves without changing in shape or in size, and therefore without dissipation (Ryder, 1996).

* **Solitons**

It is a solitary wave, which exhibits orbital stability (Bellazzini et al., 2007). In other words, it is solution of equations in field theory, which represent stable configurations with a well-defined energy which is not singular anywhere.

5.0.2.1 Remark. There are two conditions that can tell us that the equations might produce solitons and solitary wave:

1. Complete integrability (e.g. Korteweg-de Vries equation)

$$\frac{\partial \phi}{\partial t} + \frac{\partial^3 \phi}{\partial x^3} + 6\phi \frac{\partial \phi}{\partial x} = 0 \,,$$

   which its soliton solution is given by

$$\phi(x,t) = \frac{1}{2}\, v \,\operatorname{sech}^2\left(\frac{\sqrt{v}}{2}(x - v\,t - A)\right) \,,$$

   where v is the speed of the wave ϕ, and A is an arbitrary constant.

   This soliton developed by means of the inverse scattering transform.

2. Topological constraints (e.g. Sine-Gordon equation)

$$\frac{\partial^2 \phi}{\partial t^2} - \frac{\partial^2 \phi}{\partial x^2} + \frac{1}{b^2}\sin(b\,\phi) = 0 \,,$$

   which its soliton solution is given by

$$\phi(x,t) = \frac{4}{b}\ \arctan e^{\pm(\gamma/b)(x-vt)} \,,$$

   where $\gamma = (1 - v^2)^{-1/2}$, v is the speed of the wave ϕ, and b is an arbitrary constant.

   This soliton is called a kink soliton and the existence of it depends on the Topological constraints on the space.

5.0.2.2 Remark. The properties of the solitons:

1. It doesn't obey superposition principle.
2. It moves without changing in shape or in size.
3. Its speed v, depends on the size (amplitude) of the wave .

* **Statistical manifold**

  It is a Riemannian manifold, which we consider its points as probability distributions for a statistical model, i.e., in open interval $O \subseteq \mathbb{R}^N$, for a set of probability distributions $S = \{p_\theta | \theta \in O\}$, a Riemannian manifold M becomes statistical manifold if every point $p \in M$ is one to one relation with the distribution $p_\theta \in S$ (Wagenaar, 1998).

* **Tangent spaces**

  In a smooth manifold M, the tangent space T_p of M at p is the set of all tangent vector X_P in a point p to a differentiable curve $C(t)$.

  For example, by considering a sphere S^2 as a smooth manifold, and assuming that $p \in S^2$ is any arbitrary point in the sphere. If $C_1(t)$ and $C_2(t)$ are the only two differentiable curves passing through p such that $C_1(0) = C_2(0) = p$, therefore we have two tangent vectors $C_1'(0)$ and $C_2'(0)$[1] of the manifold S^2 at the point p, and the tangent space $T_p = \{C_1'(0), C_2'(0)\}$.

  .

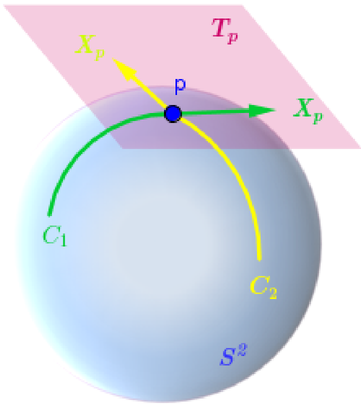


Figure 5.2: Tangent space and tangent vectors of S^2

* **Tangent vector**

  In a smooth manifold M, the tangent vector X_P at a point P to a differentiable curve $C(t)$ is defined as

$$X_p = \frac{dC}{dt}\bigg|_{t=0},$$

  where $C(0) = p$.

[1]Where $C_i'(0) = \frac{dC_i}{dt}\Big|_{t=0}$

* **Unbiased estimator**

If the systematic error ($b = \langle\hat{\boldsymbol{\theta}}(\vec{x})\rangle \;-\; \vec{\theta}$) is equal to zero that means the average of repeated measurements tends to true value. And the estimator is called an unbiased estimator.

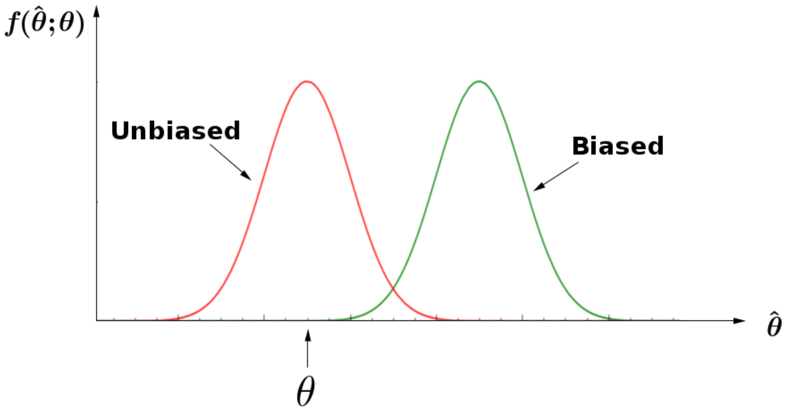


Figure 5.3: Biased and Unbiased estimator

For example, we check the estimator $\hat{\mu}$ in the previous example if it is biased or unbiased estimator.

The systematic error,

$$\begin{aligned} b &= \langle\hat{\mu}\rangle \;-\; \mu \\ &= \mu \;-\; \mu \\ &= 0. \end{aligned}$$

That means the estimator $\hat{\mu}$ is unbiased estimator.

* **Variance**

In Probability theory, the variance of a function f, of a random variable x, is a measurement of how far this function f spread out of its expectation value $E[f]$. It is equivalent to the expectation value of the square of the bias between the function and its expectation value, and it is given by

$$V[f] = E\left[(f - E[f])^2\right].$$

We can simplify this equation by

$$\begin{aligned} V[f] &= E\left[f^2 - 2fE[f] + (E[f])^2\right] \\ &= E\left[f^2\right] - 2E\left[f\right]E[f] + (E[f])^2 \\ &= E\left[f^2\right] - 2(E[f])^2 + (E[f])^2 \\ &= E\left[f^2\right] - (E[f])^2. \end{aligned}$$

5.0.2.3 Remark. The variance of a continuous/discrete function.

– The variance of a continuous function f, of a random variable x, distributed with probability function $p(x;\theta)$, is given by

$$V[f](\theta) \;=\; \int_{-\infty}^{\infty}\left(f^2\; p(x;\theta)\; dx\right) \;-\; \int_{-\infty}^{\infty}\left(f\; p(x;\theta)\; dx\right).$$

– The variance of a discrete function f, of a random variable x, distributed with probability function $p(x;\theta)$, is given by

$$V[f](\theta) \ = \ \sum_{n=1}^{N} \left(f_n^2 \, p_n(x;\theta)\right) \ - \ \sum_{n=1}^{N} \left(f_n \, p_n(x;\theta)\right).$$

For example, consider the variance of the estimator $\hat{\mu}$ for the expectation value, which is given by

$$V[\hat{\mu}] = \frac{\sigma^2}{n} \, ,$$

where $\sigma = E[x^2] - (E[x])^2$.

Acknowledgements

بِسْمِ الله الرَّحْمَـن الرَّحِيم
و الحَمْدُ لِلّه الَذي بِنِعْمَتِهِ تَتِمُّ الصَالِحَات

I would like to take this chance to express my special thank to my supervisor Dr Jeff Murugan and Dr Jonathan Shock for their infinite support and enormous work they done to complete this essay. I also thank AIMS staff for the great advices and comments they gave.

Finally, I want to thank my beloved family for their unconditional support during this journey.

References

J. Bellazzini, V. Benci, C. Bonanno, and A. M. Micheletti. Solitons for the nonlinear klein-gordon equation. *arXiv preprint arXiv:0712.1103*, 2007.

I. R. Klebanov and J. M. Maldacena. Solving quantum field theories via curved spacetimes. *Phys. Today*, 62:28, 2009.

D. Lavis and R. Streater. Physics from fisher information. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 33(2):327–343, 2002.

R. J. Little and D. B. Rubin. Statistical analysis with. 2002.

U. Miyamoto and S. Yahikozawa. Information metric from a linear sigma model. *Physical Review E*, 85 (5):051133, 2012.

T. G. Pfeiffer and D. Atariah. What is... a geodesic on a riemannian manifold?

L. H. Ryder. *Quantum field theory*. Cambridge university press, 1996.

D. Wagenaar. Information geometry for neural networks. *Term paper for reading course with ACC Coolen, King's College London*, 1998.