Weierstraß-Institut
für Angewandte Analysis und Stochastik

Leibniz-Institut im Forschungsverbund Berlin e. V.

Preprint ISSN 2198-5855

Some notes on the Hellinger distance and various Fisher-Rao distances

Alexander Mielke

submitted: October 15, 2025

Weierstraß-Institut
Anton-Wilhelm-Amo-Str. 39
10117 Berlin
Germany
E-Mail: alexander.mielke@wias-berlin.de

No. 3222
Berlin 2025

2020 *Mathematics Subject Classification.* 46G99, 01A60, 53C22, 58E10, 94A17.

Key words and phrases. Hellinger distance, Fisher-Rao distance, Bhattacharya distance, geodesic curves, product measures, Gaussian measures, exponential distributions, Poisson distributions, metric cone, absolutely continuous curves, growth equation.

The author is very grateful to François-Xavier Vialard for stimulating and helpful discussion about the historical background. The research was partially supported by Deutsche Forschungsgemeinschaft (DFG) through the Berlin Mathematics Research Center MATH+ (EXC-2046/1, DFG project no. 390685689) subproject “DistFell”.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Leibniz-Institut im Forschungsverbund Berlin e. V.
Anton-Wilhelm-Amo-Straße 39
10117 Berlin
Germany

Fax: +49 30 20372-303
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

Some notes on the Hellinger distance and various Fisher-Rao distances

Alexander Mielke

Abstract

These expository notes introduce the Hellinger distance on the set of all measures and the induced Fisher-Rao distances for subsets of measures, such as probability measures or Gaussian measures. The historical background is highlighted and the relations and the distinct features of the two distances are discussed. Moreover, we provide a dynamic characterization of absolutely continuous curves in the Hellinger spaces in terms of the growth equation, which replaces the continuity equation in the theory of optimal transport.

1 Introduction

The initial motivation for writing this mainly expository notes was the question what are the origin of the names "Hellinger distance" and "Fisher-Rao distance". Hence, we will introduce these two concepts, which are indeed closely related, in simple terms and explain the historical developments that go back to Hellinger [Hel07, Hel09] and Kakutani [Kak48] for the Hellinger distance and to Fisher [Fis21] and Rao [Rao45] for the Fisher-Rao distance. Unfortunately, these names are sometimes mixed up and these notes provide a guideline for distinguishing the two objects such that future mathematical discussion can made more precise by avoiding unnecessary confusion.

A second goal arises from the recent interest in gradient flows in the Hellinger space (e.g. [CH*24, MiZ25]) and in the combination of the Wasserstein distance and the Hellinger distance in the transport growth distance called Hellinger-Kantorovich distance in [LMS16, LMS18] and and Wasserstein-Fisher-Rao distance in [CP*18a, CP*18b]. While there is a large body of work in characterizing absolutely continuous curves in the Wasserstein space via the dynamic theory of Benamou-Brenier [BeB00] and Otto [Ott01] using the continuity equation, there is no counterpart available for the absolutely continuous curves in the Hellinger space. In principle, the corresponding theory can be extracted from the analysis of the Hellinger-Kantorovich theory in [LMS18], but this would lead to a huge and inscrutable overhead. In Section 2.3 we provide a short and mathematically complete characterization, which shows the relations between the metric derivative and the growth equation, which replaces the continuity equation.

The Hellinger distance $\mathsf{He}(\mu_0,\mu_1)$ between arbitrary measures $\mu_0,\mu_1\in\mathfrak{M}(\Omega)$ on a measure space Ω are is defined by

$$\mathsf{He}(\mu_0,\mu_1)^2=\sigma^2\int_\Omega\Big(\big(\frac{\mathrm{d}\mu_1}{\mathrm{d}\lambda}\big)^{1/2}-\big(\frac{\mathrm{d}\mu_0}{\mathrm{d}\lambda}\big)^{1/2}\Big)^2\mathrm{d}\lambda=\sigma^2\Big(\mu_0(\Omega)+\mu_1(\Omega)-2\sqrt{\mu_0\mu_1}\,(\Omega)\Big),\quad(1.1)$$

where $\lambda\in\mathfrak{M}(\Omega)$ is any measure such that $\mu_0\ll\lambda$ and $\mu_1\ll\lambda$, and $\frac{\mathrm{d}\mu_j}{\mathrm{d}\lambda}\in\mathrm{L}^1(\Omega,\lambda)$ denotes the Radon-Nikodym derivative. Here we have introduced a scaling factor $\sigma>0$ into the definition because in various places in the literature different factors are chosen. We keep the factor throughout to facility the comparison with the literature, but it is also helpful to understand the structure better.

DOI 10.20347/WIAS.PREPRINT.3222 Berlin 2025

The definition of He goes back to Kakutani [Kak48] and was chosen to honor the contribution of Hellinger in [Hel07, Hel09] which showed how to define the Hellinger integral

$$\sqrt{\mu_0\mu_1}(B) = \int_B \Big(\frac{\mathrm{d}\mu_0}{\mathrm{d}\lambda}\,\frac{\mathrm{d}\mu_1}{\mathrm{d}\lambda}\Big)^{1/2}\,\mathrm{d}\lambda \quad \text{for measurable } B\subset\Omega, \tag{1.2}$$

much before the introduction of Radon-Nikodym derivatives. Thus, the geometric mean $\sqrt{\mu_0\mu_1}$ is again a well-defined measure in $\mathfrak{M}(\Omega)$. Moreover, [Kak48] showed that for every $\lambda\in\mathfrak{M}(\Omega)$ the Hilbert space $\mathrm{L}^2(\Omega,\lambda)$ can be isometrically embedded into $(\mathfrak{M}(\Omega),\mathsf{He})$ via the mapping $\mathrm{L}^2(\Omega,\lambda)\ni g\mapsto \sqrt{g}\,\lambda\in\mathfrak{M}(\Omega)$. This embedding immediately shows that the Hellinger distance is even a geodesic distance in the sense that for every pair (μ_0,μ_1) there exists a unique constant-speed geodesic curve given by

$$\begin{aligned}\gamma^{\mathsf{He}}_{\mu_0\to\mu_1}(s) &= (1{-}s)^2\mu_0 + s^2\mu_1 + 2(s{-}s^2)\sqrt{\mu_0\mu_1}\\ &= (1{-}s)\mu_0 + s\mu_1 - (s{-}s^2)\frac{1}{\sigma^2}\mathsf{He}(\mu_0,\mu_1)^2 \quad \text{for } s\in[0,1].\end{aligned} \tag{1.3}$$

Moreover, one can define a pseudo-Riemannian structure on $\mathfrak{M}(\Omega)$ given by the quadratic form of Hellinger type (see (2.23) in the historical remarks in Section 2.8)

$$\mathbf{g}_\mu(\nu_1,\nu_2) = \begin{cases} \dfrac{\sigma^2}{4}\displaystyle\int_\Omega \frac{\mathrm{d}\nu_1}{\mathrm{d}\mu}\,\frac{\mathrm{d}\nu_2}{\mathrm{d}\mu}\,\mathrm{d}\mu & \text{if } \nu_1,\nu_2\ll\mu,\\ \infty & \text{else.}\end{cases} \tag{1.4}$$

In Theorem 2.2 we provide the mathematically connection of $\mathbf{g}_\mu$ with the metric speed in $(\mathfrak{M}(\Omega),\mathsf{He})$. Moreover, we refer to [AJ*15, BBM16] for a proof of the uniqueness of this Riemannian metric under diffeomorphisms.

The above Riemannian structure was indeed introduced by Fisher in [Fis21] when studying finite-dimensional parameterized families of measures. Considering the family

$$\mathcal{S} = \big\{\, f(p;\cdot)\in\mathrm{L}^1(\mathbb{R}^n)\cap\mathfrak{P}(\mathbb{R}^n)\;\big|\; p\in D\subset\mathbb{R}^m\,\big\}$$

Then, the Fisher information metric is defined via the matrix $\mathbb{F}(p)\in\mathbb{R}^{m\times m}_{\geq 0}$ given by

$$\begin{aligned} a\cdot\mathbb{F}(p)b &:= \frac{\sigma^2}{4}\int_{\mathbb{R}^n} \mathrm{D}_p\log\big(f(p,x)\big)[a]\,\mathrm{D}_p\log\big(f(p,x)\big)[b]\,f(p,x)\,\mathrm{d}x\\ &= \frac{\sigma^2}{4}\int_{\mathbb{R}^n}\frac{\mathrm{D}_pf(p,x)[a]\,\mathrm{D}_pf(p,x)[b]}{f(p,x)}\,\mathrm{d}x\\ &= -\frac{\sigma^2}{4}\int_{\mathbb{R}^m}\mathrm{D}_p^2\log\big(f(p,x)\big)[a,b]\,f(p,x)\,\mathrm{d}x\end{aligned} \tag{1.5}$$

The Fisher-Rao distance was introduced in [Rao45] and is defined as the distance on D induced by the metric tensor $\mathbb{F}$, namely

$$\mathsf{FR}_{\mathcal{S}}(p_0,p_1) := \inf\Big\{\int_0^1\!\big(p'(s)\cdot\mathbb{F}(p(s))p')s)\big)^{1/2}\mathrm{d}s\;\Big|\; p\in\mathrm{C}^1([0,1];D),\ p(0)=p_0,\ p(1)=p_1\Big\}.$$

The strength of the Fisher-Rao distance is that it does not depend on the particular choice of the pasteurization, but only on the subset $\mathcal{S}\subset\mathfrak{M}(\Omega)$. Thus, we will $\mathsf{FR}_{\mathcal{S}}\big(f(p_0,\cdot)\mathrm{d}x, f(p_1,\cdot)\mathrm{d}x)\big)$ instead of $\mathsf{FR}_{\mathcal{S}}(p_0,p_1)$.

The above construction is not restricted to finite dimensional submanifolds, but can be generalized to more general subsets $\mathcal{S}$ of $\mathfrak{M}(\Omega)$. Indeed, $\mathsf{FR}_{\mathcal{S}} : \mathcal{S}\times\mathcal{S} \to [0,\infty]$ can be understood as the intrinsic length or distance in $\mathcal{S}$ induced from $(\mathfrak{M}(\Omega), \mathsf{He})$. For this, we consider continuous paths $\gamma : [0,1] \to \mathcal{S}$, define their intrinsic length $L_{\mathsf{He}}(\gamma)$, see (3.1), and then define the Fisher-Rao distance in $\mathcal{S}$ by

$$\mathsf{FR}_{\mathcal{S}}(\mu_0,\mu_1) := \inf\Big\{ L_{\mathsf{He}}(\gamma) \;\Big|\; \gamma \in \mathrm{C}^0([0,1];\mathcal{S}),\ \gamma(0)=\mu_0,\ \gamma(1)=\mu_1) \Big\}.$$

Of course choosing $\mathcal{S} = \mathfrak{M}(\Omega)$, i.e., the space of all measures, we have $\mathsf{FR}_{\mathfrak{M}(\Omega)} = \mathsf{He}$, and this is the reason why sometimes the Hellinger distance is called Fisher-Rao distance. However, it is better to distinguished the refined concept of Fisher-Rao distances $\mathsf{FR}_{\mathcal{S}}$ which depends on the chosen submanifold or subset $\mathcal{S}$ of the set of all measures $\mathfrak{M}(\Omega)$.

In general, one has $\mathsf{FR}_{\mathcal{S}}(\mu_0,\mu_1) \geq \mathsf{He}(\mu_0,\mu_1)$ where equality holds only in exceptional cases, namely, if the Hellinger geodesic, as given in (1.3), totally lies in $\mathcal{S}$. In case that $\mathcal{S}$ is a smooth submanifold the local Fisher-Riemann metric is simply the restriction of the quadratic form (1.4) of Hellinger type, one can expect local closeness of He and $\mathsf{FR}_{\mathcal{S}}$, i.e.

$$\mathsf{He}(\mu_0,\mu_1)^2 \leq \mathsf{FR}_{\mathcal{S}}(\mu_0,\mu_1)^2 \leq \mathsf{He}(\mu_0,\mu_1)^2 + O\big(\mathsf{He}(\mu_0,\mu_1)^3\big) \quad \text{as } \mathsf{He}(\mu_0,\mu_1) \to 0.$$

Already the restriction of He to the probability measures in $\mathfrak{P}(\Omega)$ leads to a new distance, namely the Bhattacharya distance [Bha42, Rao45]

$$\mathsf{Bh}(\mu_0,\mu_1) = 2\sigma \arcsin\big(\tfrac{1}{2\sigma}\,\mathsf{He}(\mu_0,\mu_1)\big).$$

We refer to Section 3.2 for more details and emphasize that $\sigma > 0$ appears nonlinearly.

The plan of the paper is as follows. In Section 2 we present basic properties of the Hellinger distance such its the geodesic curves, the embedding property into a Hilbert spaces (showing that the geometry is flat), and the behavior under pushforwards. Moreover, we present some historical remarks about Hellinger's contribution and the development of the name "Hellinger distance". The only mathematically new part of these notes are the short and self-contained characterization of absolutely continuous curves in $(\mathfrak{M}(\Omega), \mathsf{He})$ using metric speed and the growth equation, see Theorem 2.2.

In Section 3 we discuss the abstract definition of the Fisher-Rao distance for general subsets $\mathcal{S} \subset \mathfrak{M}(\Omega)$. After treating the most important example $\mathcal{S} = \mathfrak{P}(\Omega)$ leading to the Bhattacharya distance, we show how the Fisher-Rao distance on a set $\mathcal{P} \subset \mathfrak{P}(\Omega)$ can be used to construct the Fisher-Rao distance on the cone $\mathcal{S} = [0,\infty[\mathcal{P} \subset \mathfrak{M}(\Omega)$ in Theorem 3.2. This is a general construction (see e.g. [BBI01] for general geodesic spaces, where the involvement of the scaling parameter $\sigma > 0$ is nontrivial. In Section 3.4 we show that the Fisher-Rao distance for product probability measures satisfies $\mathsf{FR}^2_{\mathcal{P}_1\otimes\mathcal{P}_2} = \mathsf{FR}^2_{\mathcal{P}_1} + \mathsf{FR}^2_{\mathcal{P}_2}$.

Section 4 is devoted to simple examples, namely $\mathcal{S}_{\rm trans}$ containing all translations of a measure on $\mathbb{R}^n$, the family $\mathcal{S}_{\rm Poiss}$ of multivariate Poisson distributions on $\mathbb{N}_0^d$, and the family $\mathcal{S}_{\rm exp}$ of exponential distributions on $(\mathcal{R}_{\geq}0)^n$. Finally, in Section 4.4 we discuss the known results on the Fisher-Rao distance on $\mathcal{S}_{\rm Gauss}$, the set of Gaussian distributions on $\mathbb{R}^d$: only for $d=1$ an explicit formula is known, and for $d \geq 2$ only partial results are available.

2 Properties of the Hellinger distance

As in [Kak48] we start from a measure space $(\Omega, \mathfrak{A})$, i.e. $\mathfrak{A}$ is a σ-algebra over the set Ω. By $\mathfrak{M}(\Omega)$ we denote the set of all (non-negative) finite measures on $(\Omega, \mathfrak{A})$, i.e. countably additive set functions.

The subset of probability measures is denoted by $\mathfrak{P}(\Omega) = \{\, \mu \in \mathfrak{M}(\Omega) \,\big|\, \mu(\Omega) = 1 \,\}$.

2.1 Hellinger's integral

As mentioned above the Hellinger distance relies on the the so-called Hellinger integral, which is the geometric mean of two measures $\mu_0, \mu_1 \in \mathfrak{M}(\Omega)$. In modern terms the measure $\sqrt{\mu_0\mu_1} \in \mathfrak{M}(\Omega)$ is defined by the Radon-Nikodym derivative via

$$\sqrt{\mu_0\mu_1} = \Big(\frac{\mathrm{d}\mu_0}{\mathrm{d}\lambda}\Big)^{1/2}\Big(\frac{\mathrm{d}\mu_1}{\mathrm{d}\lambda}\Big)^{1/2}\lambda \quad \text{for every } \lambda \gg \mu_0, \mu_1. \tag{2.1}$$

The geometric mean can also be defined by partitions as follows

$$\sqrt{\mu_0\mu_1}(A) = \inf\Big\{ \sum_{i\in I} \mu_0(A_i)^{1/2}\mu_1(A_i)^{1/2} \,\Big|\, A = \bigcup_{i\in I} A_i,\ A_i \cap A_j = \emptyset \text{ for } i \neq j \Big\}. \tag{2.2}$$

Using that $(r,s) \mapsto (rs)^{1/2}$ is concave, it is easy to see that refining partitions of a set A leads to a smaller value (use $\sqrt{\theta_i{+}\theta_j}\sqrt{\nu_i{+}\nu_j} \geq \sqrt{\theta_i}\sqrt{\nu_i} + \sqrt{\theta_j}\sqrt{\nu_j}$). The historical Section 2.8 will explain how this construction is related to Hellinger's work in [Hel07, Hel09].

Remark 2.1 (Kolmogorov and Hellinger integrals) *More generally, for a positively one-homogeneous concave function* $\varphi : [0,\infty[^N \to [0,\infty[$ *and measures* $\mu_1, .., \mu_n \in \mathfrak{M}(\Omega)$ *the measure* $\varphi(\mu_1, \mu_1, .., \mu_n) \in \mathfrak{M}(\Omega)$ *can be defined by an infimum over partitions as in* (2.2)*. The Kolmogorov integral of* $g \in \mathrm{L}^\infty(\Omega)$ *is then defined as* $\int_\Omega g\,\mathrm{d}\varphi(\mu_1, \mu_1, .., \mu_n)$*, see [Kol30]. In particular, using* $\phi_\alpha(r,s) = r^\alpha s^{1-\alpha}$ *with* $\alpha \in [0,1]$ *and two measures* $\mu_0, \mu_1 \in \mathfrak{M}(\Omega)$*, one can define the measures* $\phi_\alpha(\mu_0,\mu_1) \in \mathfrak{M}(\Omega)$ *and the so-called* α*-Hellinger integral* $\int_\Omega g \mathrm{d}\phi_\alpha(\mu_0,\mu_1)$ *for* $g \in \mathrm{L}^\infty(\Omega)$*.*

2.2 The topology of $(\mathfrak{M}(\Omega), \mathsf{He})$

The topology induced by He on $\mathfrak{M}(\Omega)$ is the norm topology induced by the total variation

$$\|\mu_1{-}\mu_0\|_{\mathrm{TV}} = \int_\Omega \big|\frac{\mathrm{d}\mu_1}{\mathrm{d}\lambda} - \frac{\mathrm{d}\mu_0}{\mathrm{d}\lambda}\big|\,\mathrm{d}\lambda \quad \text{for } \mu_0, \mu_1\, \mathrm{d}\lambda.$$

However, we see that the total variation norm scales one-homogeneous with the mass, while the Hellinger distance scales homogeneous of degree $1/2$:

$$\|r\mu_1{-}r\mu_0\|_{\mathrm{TV}} = r\|\mu_1{-}\mu_0\|_{\mathrm{TV}} \quad \text{and} \quad \mathsf{He}(r\mu_0, r\mu_1) = r^{1/2}\,\mathsf{He}(\mu_0,\mu_1)$$

for $r \geq 0$ and $\mu_0, \mu_1 \in \mathfrak{M}(\Omega)$. This is reflected in the lower and upper estimate of the total variation norm, namely

$$\|\mu_1{-}\mu_0\|_{\mathrm{TV}} \geq \frac{1}{\sigma^2}\,\mathsf{He}(\mu_0,\mu_1)^2, \tag{2.3a}$$

$$\|\mu_1{-}\mu_0\|_{\mathrm{TV}} \leq \sqrt{2(\mu_0(\Omega){+}\mu_1(\Omega))}\,\frac{1}{\sigma}\,\mathsf{He}(\mu_0,\mu_1). \tag{2.3b}$$

To see this, choose $\lambda = \mu_0{+}\mu_1$ and write $\mu_0 = \big(\frac12 - x\big)\lambda$ and $\mu_1 = \big(\frac12 + x\big)\lambda$ with $x(\omega) \in [-\frac12, \frac12]$ λ-a.e. Using the elementary estimates

$$2x^2 \leq 1 - \sqrt{1{-}4x^2} \leq 2|x| \quad \text{for } |x| \leq 1/2, \tag{2.4}$$

the first estimate follows from

$$\|\mu_1-\mu_0\|_{\mathrm{TV}} = \int_\Omega 2|x| \,\mathrm{d}\lambda \geq \int_\Omega \big(1-\sqrt{1-4x^2}\big) \,\mathrm{d}\lambda = \frac{1}{\sigma^2}\,\mathsf{He}(\mu_0,\mu_1)^2.$$

The second estimate follows via an application of Cauchy-Schwarz' estimate:

$$\begin{aligned}\|\mu_1-\mu_0\|_{\mathrm{TV}}^2 &= \Big(\int_\Omega 2|x| \,\mathrm{d}\lambda\Big)^2 \leq \int_\Omega 2 \,\mathrm{d}\lambda \int_\Omega 2x^2 \,\mathrm{d}\lambda \\ &\leq 2\big(\mu_0(\Omega)+\mu_1(\Omega)\big) \int_\Omega \big(1-\sqrt{1-4x^2}\,\big) \,\mathrm{d}\lambda = \frac{2}{\sigma^2}\big(\mu_0(\Omega)+\mu_1(\Omega)\big)\,\mathsf{He}(\mu_0,\mu_1)^2.\end{aligned}$$

2.3 Absolutely continuous curves, metric speed and the continuity equation in $(\mathfrak{M}(\Omega),\mathsf{He})$

In this subsection we use the abstract theory developed in [AGS05, Sec. 1.1] for a general metric space $(M,\mathcal{D})$. For $p\in[1,\infty]$, a curve $\gamma:[0,1]\to M$ is called is p-absolutely continuous if there exists $g\in \mathrm{L}^p([0,1])$ such that $\mathcal{D}(\gamma(r),\gamma(t)) \leq \int_r^t g(s)\,\mathrm{d}s$ for all $0\leq r<t\leq 1$. Every rectifiable curve can be reparametrized to a Lipschitz curve (i.e. $p=\infty$), so we see that the optimal p depends on the parametrization.

In [AGS05, Thm. 1.1.2] it is shown that the metric speed

$$\lim_{h\searrow 0}\frac{1}{h}\,\mathcal{D}\big(\gamma(s),\gamma(s+h)\big) := |\dot\gamma|_{\mathcal{D}}(s)$$

exists a.e. in $[0,1]$, and that $\mathcal{D}(\gamma(r),\gamma(t)) \leq \int_r^t |\dot\gamma|_{\mathcal{D}}(s)\,\mathrm{d}s$.

The aim of this subsection is to characterize the metric speed and relate it properly to Hellinger's quadratic form. The theory is developed in analogy to [AGS05, Ch. 8], however our case of the Hellinger distance is considerably simpler than the case of the Otto-Wasserstein theory developed there. The new ingredient is the so-called generalized continuity equation, which should rather be called a *growth equation* (GE) here. Given a curve $\mu:[s_0,s_1]\to\mathfrak{M}(\Omega); s\mapsto\mu_s$ we define the measure $\mu_{[0,1]}$ on $Q:=[0,1]\times\Omega$ via

$$\int_Q h(s,\omega)\,\mathrm{d}\mu_{[0,1]}(s,\omega) := \int_{[0,1]}\int_\Omega h(s,\omega)\,\mathrm{d}\mu_s(\omega)\,\mathrm{d}s.$$

For a growth-rate function $\xi\in\mathrm{L}^1(Q;\mu_{[0,1]})$ we say that the pair (μ,ξ) is a weak solution of the growth equation $\partial_s\mu_s=\xi_s\mu_s$ if

$$\int_Q \big(\eta'(s)\mathbf{1}_A(\omega)+\xi(s,\omega)\eta(s)\mathbf{1}_A(\omega)\big)\,\mathrm{d}\mu_{[0,1]}(s,\omega) \quad \text{for all } \eta\in\mathrm{C}^1_\mathrm{c}(]s_0,s_1[) \text{ and } A\in\mathfrak{A}. \tag{2.5}$$

In some respects, the present theory is much simpler than the corresponding theory for the Otto-Wasserstein case developed in [AGS05, Ch. 8]; however that are new complications because of the change of support of the measures. From the equation $\partial_s\mu_s=\xi_s\mu_s$ one would naively guess that the solution can be written as $\mu(s)=\exp\big(\int_0^s\xi(r,\cdot)\,\mathrm{d}r\big)\mu(0)$, but that cannot be true in case where $\mu(0)$ and $\mu(1)$ have different support. For instance, choosing $\omega_0\neq\omega_1$ in Ω and considering the curve

$$\mu(s)=a_0(s)\delta_{\omega_0}+\theta a_1(s)\delta_{\omega_2} \quad \text{with } \theta\geq 0,\ a_0(s)=(1-s)^{\gamma_0} \text{ and } a_1(s)=s^{\gamma_1},$$

DOI 10.20347/WIAS.PREPRINT.3222

we have $\mu(j) = \theta^j \delta_{\omega_j}$ (Dirac measure) for $j = 1, 2$. Moreover, the growth equation is satisfied by (μ, ξ) if $\xi(s, \omega_j) = a_j'(s)/a_j(s)$. Moreover, for $\min\{\gamma_0, \gamma_1\} > 1+p$ we have

$$\int_Q |\xi|^p \,\mathrm{d}\mu_I = \int_0^1 \sum_{j=0}^{1} \theta^j |a_j'|^p / a_j^{p-1} \,\mathrm{d}s = \sum_{j=0}^{1} \theta^j \frac{\gamma_j^p}{\gamma_j - 1 - p} < \infty.$$

We also note that for the given growth-rate function ξ and the given initial condition $\mu(0) = \delta_\omega$ there are infinitely many solution pairs (μ, ξ) for the growth equation (2.5), because $\theta \geq 0$ is arbitrary.

For simplicity, we restrict to the natural case $p = 2$.

Theorem 2.2 *(A) If $\mu : [s_0, s_1] \to \mathfrak{M}(\Omega)$ is 2-absolutely continuous in $(\mathfrak{M}(\Omega), \mathsf{He})$, then there exists $\xi \in \mathrm{L}^2([s_0, s_1] \times \Omega)$ such that (μ, ξ) solve the growth equation* (2.5) *and and the metric speed satisfies*

$$|\dot{\mu}|_{\mathsf{He}}(s) = \frac{\sigma}{2} \|\xi_s\|_{\mathrm{L}^2(\Omega, \mu_s} = \frac{\sigma}{2} \bigg(\int_\Omega \xi(s, \omega)^2 \,\mathrm{d}\mu_s(\omega) \bigg)^{1/2} \quad \text{a.e. on } [0, 1]. \tag{2.6}$$

(B) Vice versa, if $\mu : [s_0, s_1] \to \mathfrak{M}(\Omega)$ is a continuous curve, $\xi \in \mathrm{L}^2([s_0, s_1] \times \Omega, \mu_{[s_0, s_1]})$, and (μ, ξ) solves (2.5)*, then μ is 2-absolutely continuous in $(\mathfrak{M}(\Omega), \mathsf{He})$ and* (2.6) *holds.*

Before going into the proof of the result we emphasize that relation (2.6) features Hellinger's quadratic form (1.4). From (the weak form of) the) growth equation $\partial_s \mu_s = \xi_s \mu_s$ we have

$$\xi_s = \frac{\mathrm{d}(\partial_s \mu_s)}{\mathrm{d}\mu_s} \quad \text{and (2.6) means} \quad \big(|\dot{\mu}|_{\mathsf{He}}(s)\big)^2 = \frac{\sigma^2}{4} \int_\Omega \Big(\frac{\mathrm{d}(\partial_s \mu_s)}{\mathrm{d}\mu_s} \Big)^2 \,\mathrm{d}\mu_s = \mathbf{g}_{\mu_s}\big(\partial_s \mu_s, \partial_s \mu_s\big).$$

Proof. To simplify notation we only consider the case $[s_0, s_1] = [0, 1] =: I$.

Proof of part (A): We proceed in analogy to [AGS05, Thm. 8.3.1].

For simplicity, we set $\mathscr{V} := \mathrm{L}^2(Q; \mu_I)$ where $Q = I \times \Omega$ and $\mu_I = \mu_{[0,1]}$. Moreover, we define the dense subset

$$V = \Big\{ (t, \omega) \mapsto \sum_{i=1}^{n} \eta_i(t) \mathbf{1}_{A_i}(\omega) \;\Big|\; n \in \mathbb{N},\ \eta_i \in \mathrm{C}^1(I),\ \eta_i(0) = 0 = \eta_i(1),\ A_i \in \mathfrak{A} \Big\}.$$

On V we define the linear mapping $L : V \to \mathbb{R}$ via

$$L\varphi = -\int_Q \partial_s \varphi(s, \omega) \,\mathrm{d}\mu_I(s, \omega) = -\int_I \int_\Omega \partial_s \varphi(s, \omega) \,\mathrm{d}\mu_s(\omega) \,\mathrm{d}s.$$

We now want to show that L can be extended continuously on all of $\mathscr{V}$. We define $S \subset [0, 1]$ to be the set of those s where $|\dot{\mu}|_{\mathsf{He}}(s)$ exist. Then, for all measurable and bounded $g : \Omega \to \mathbb{R}$ we have

$$\big(\langle g, \mu_{s+h} \rangle - \langle g, \mu_s \rangle\big) = \int_\Omega g(1 - 2\theta_h) \,\mathrm{d}\lambda_h \quad \text{where } \lambda_h = \mu_{s+h} + \mu_s,$$

$\gamma_s = \theta_h \lambda_h$, and $\gamma_{s+h} = (1 - \theta_h)\lambda_h$. Using Cauchy-Schwarz' estimate and (2.4) we find

$$\frac{1}{h} \big| \langle g, \mu_{s+h} \rangle - \langle g, \mu_s \rangle \big| \leq \big\| g \big\|_{\mathrm{L}^2(\Omega, \lambda_h)} \frac{\sqrt{2}}{\sigma} \frac{\mathsf{He}(\mu_s, \mu_{s+h})}{h}.$$

We now assume $s \in S$ and use that $\lambda_h \to 2\mu_s$ (by strong continuity of $t \mapsto \mu_t$. Thus we have

$$\limsup_{h \searrow 0} \frac{1}{h} \big| \langle g, \mu_{s+h} \rangle - \langle g, \mu_s \rangle \big| \leq \big\| g \|_{\mathrm{L}^2(\Omega, \mu_s)} \frac{2}{\sigma} |\dot{\mu}|_{\mathsf{He}}(s). \tag{2.7}$$

Now we consider a general $\varphi \in V$ and extend it (continuously!) by 0. Moreover, extend $s \mapsto \mu(s)$ by $\mu(1)$ for $s \geq 1$. Then, we have

$$\begin{aligned} \int_Q \partial_s \varphi \, \mathrm{d}\mu_I &= \lim_{h \searrow 0} \int_Q \frac{1}{h} \big(\varphi(s, \omega) - \varphi(s-h, , \omega) \big) \, \mathrm{d}\mu_I \\ &= \lim_{h \searrow 0} \bigg(\int_I \frac{\langle \varphi_s, \mu_s \rangle - \langle \varphi_s, \mu_{s+h} \rangle}{h} \, \mathrm{d}s - \frac{1}{h} \int_0^h \langle \varphi_{s-h}, \mu_s \rangle \, \mathrm{d}s + \frac{1}{h} \int_{1-h}^1 \langle \varphi_s, \mu_{s+h} \rangle \bigg). \end{aligned}$$

Because $\varphi_0 = 0 = \varphi_1$ the last two terms vanish with $h \searrow 0$. Hence, together with (2.7), we find

$$\big| L(\varphi) \big| = \bigg| \int_Q \partial_s \varphi \, \mathrm{d}\mu_I \bigg| \leq \int_I \big\| \varphi_s \big\|_{\mathrm{L}^2(\Omega, \mu_s)} \frac{2}{\sigma} |\dot{\mu}|_{\mathsf{He}}(s) \, \mathrm{d}s. \tag{2.8}$$

By assumption $s \mapsto |\dot{\mu}|_{\mathsf{He}}(s)$ lies in $\mathrm{L}^2(I)$, hence L can be extended continuously to $\mathscr{V}$. By Riesz' representation theorem for the Hilbert space $\mathscr{V}$, there exists $\xi \in \mathscr{V}$ such that $L(\varphi) = \int_Q \xi \varphi \, \mathrm{d}\mu_I$, but this shows that (μ, ξ) solve the growth equation (2.5).

Moreover, take $\eta = \mathbf{1}_{[s_0, s_1]}$, then using (2.8) we find

$$\begin{aligned} \int_{s_0}^{s_1} \int_\Omega \xi_s^2 \, \mathrm{d}\mu_s \, \mathrm{d}s = \int_Q \eta \xi^2 \, \mathrm{d}\mu_I = L(\eta \xi) &\overset{(2.8)}{\leq} \frac{2}{\sigma} \int_Q \eta \|\xi_s\|_{\mathrm{L}^2(\Omega, \mu_s)} |\dot{\mu}|_{\mathsf{He}}(s) \, \mathrm{d}s \\ &\leq \frac{2}{\sigma} \bigg(\int_{s_0}^{s_1} \int_\Omega \xi_s^2 \, \mathrm{d}\mu_s \, \mathrm{d}s \bigg)^{1/2} \bigg(\int_{s_0}^{s_1} \big(|\dot{\mu}|_{\mathsf{He}}(s) \big)^2 \, \mathrm{d}s \bigg)^{1/2}. \end{aligned}$$

Since s_0 and s_1 with $0 \leq s_0 < s_1 \leq 1$ are arbitrary we conclude

$$\int_\Omega \xi_s^2 \, \mathrm{d}\mu_s \leq \frac{4}{\sigma^2} \big(|\dot{\mu}|_{\mathsf{He}}(s) \big)^2 \quad \text{for a.a. } s \in [0, 1].$$

Thus, we have established (2.6) with "$\geq$" instead of "$=$". The opposite inequality will be shown via part (B).

Proof of part (B): The measure $\mu_I \in (Q)$ has two disintegration with respect to $Q = I \times \Omega$, namely into $\mathrm{d}\mu_i = \mathrm{d}\mu_s(\omega) \, \mathrm{d}s$ and $\mathrm{d}\mu_I = \mathrm{d}\nu_\omega(s) \, \mathrm{d}\overline{\mu}(\omega)$, where $\overline{\mu} \in \mathfrak{M}(\Omega)$ and $\nu_\omega \in \mathfrak{P}([0, 1])$ for $\overline{\mu}$-a.a. $\omega \in \Omega$. Here $\overline{\mu}(A) = \mu_I(I \times A)$ or $\overline{\mu} = \int_0^1 \gamma_s \, \mathrm{d}s$ (Bochner integral). From $\xi \in \mathrm{L}^2(Q; \mu_I)$ we have $\xi_\omega := \xi(\cdot, \omega) \in \mathrm{L}^2([0, 1], \nu_\omega)$ $\overline{\mu}$-a.e. in Ω. Testing weak growth equation (2.5) with $\varphi(s, \omega) = \eta(s) \mathbf{1}_A(\omega)$ we find

$$\int_A \int_I \big(\eta'(s) + \xi(s, \omega) \, \mathrm{d}\nu_\omega(s) \, \mathrm{d}\overline{\mu}(\omega) = 0 \quad \text{for all } \eta \in \mathrm{C}_0^1([0, 1]) \text{ and } A \in \mathfrak{A}.$$

As $A \in \mathfrak{A}$ is arbitrary, we conclude

$$\Big(\forall \, \eta \in \mathrm{C}_0^1([0, 1]) : \int_I \big(\eta'(s) + \xi_\omega(s) \big) \, \mathrm{d}\nu_\omega(s) = 0 \Big) \quad \overline{\mu}\text{-a.e. in } \Omega.$$

Thus, we have reduced the problem in $\mathfrak{M}(\Omega)$ to a pointwise problem scalar problem.

From $\xi_\omega \in \mathrm{L}^2([0,1],\nu_\omega)$ we see that $\zeta_\omega := \xi_\omega \nu_\omega$ is a signed measure on $[0,1]$ and we have $\partial_s \nu_\omega = \zeta_\omega$ in the distributional sense. Hence, ν_ω is absolutely continuous with respect to $\mathrm{d}s$ with $\nu_\omega = n_\omega \,\mathrm{d}s$ and $n_\omega \in \mathrm{BV}([0,1])$. Inserting this once again into the weak equation we find $n_\omega \in \mathrm{W}^{1,1}([0,1])$ with $n'_\omega(s) = \xi_\omega(s) n_\omega(s)$ a.e. in $[0,1]$.

Omitting the subscript for the moment, we set $h(s) = \sqrt{n(s)} \geq 0$ and find $2hh' = \xi h^2$, which implies either $h = 0$ or $h' = \frac{1}{2}\xi h$. Since $h(s) = 0$ and $h \geq 0$ imply $h'(s) = 0$ a.e., we obtain

$$h'(s)^2 = \frac{1}{4}\xi(s)^2 h(s) = \frac{1}{4}\xi(s)^2 n(s) \quad \text{a.e. in } [0,1].$$

Thus, we have

$$\begin{aligned} \big(\sqrt{n_\omega(1)} - \sqrt{n_\omega(0)}\big)^2 = \big(h(1){-}h(0)\big)^2 = \Big(\int_I h'(s)\,\mathrm{d}s\Big)^2 \\ \leq \int_I \big(h'(s)\big)^2\,\mathrm{d}s = \int_I \frac{1}{4}\xi_\omega(s)^2 n_\omega(s)\,\mathrm{d}s \end{aligned} \tag{2.9}$$

Noting that $\mathrm{d}\mu_s = n_\omega(s)\overline{\mu}$ we can integrate the this estimate and arrive at

$$\begin{aligned} \mathsf{He}(\mu_0,\mu_1)^2 &= \sigma^2 \int_\Omega \big(\sqrt{n_\omega(1)} - \sqrt{n_\omega(0)}\big)^2\,\mathrm{d}\overline{\mu}(\omega) \leq \frac{\sigma^2}{4}\int_\Omega\int_I \xi_\omega^2 n_\omega\,\mathrm{d}s\,\mathrm{d}\overline{\mu} \\ &= \frac{\sigma^2}{4}\int_I g(s)^2\,\mathrm{d}s \quad \text{with } g(s) := \|\xi_s\|_{\mathrm{L}^2(\Omega,\gamma_s)}. \end{aligned} \tag{2.10}$$

The same estimate can be done on each subinterval $[r,t] \subset [0,T]$ giving

$$\mathsf{He}(\mu_r,\mu_t) \leq \frac{\sigma}{2}\,(t{-}r)\Big(\frac{1}{t{-}r}\int_{[r,t]} g(s)^2\,\mathrm{d}s\Big)^{1/2}.$$

where the factors $(t{-}r)$ disappear because the interval $[0,1]$ in (2.10) needs to be rescaled to $[r,t]$. Defining the partition points $s_i = r + i(t{-}r)/N$ for $i = 0,1,\ldots,N$ and the piecewise constant function

$$G_N(s) = \sum_{i=1}^N \frac{1}{s_i{-}s_{i-1}}\int_{s_{i-1}}^{s_i} g(r)^2\,\mathrm{d}r\,\mathbf{1}_{[s_{i-1},s_i]}(s)$$

we find $G_N \to g^2$ in $\mathrm{L}^2([r,t])$ and $\sqrt{G_N} \to g$ in $\mathrm{L}^2([0,1])$. With this we have

$$\begin{aligned} \mathsf{He}(\mu_r,\mu_t) &\leq \sum_{i=1}^N \mathsf{He}(\mu_{s_{i-1}},\mu_{s_i}) \leq \frac{\sigma}{2}\sum_{i=1}^N (s_i{-}s_{i-1})\sqrt{G_N(s_{i-1/2})} \\ &= \frac{\sigma}{2}\int_r^t \sqrt{G_N(s)}\,\mathrm{d}s \to \frac{\sigma}{2}\int_r^t g(s)\,\mathrm{d}s \text{ for } N \to \infty. \end{aligned}$$

This shows that $\mu : [0,1] \to \mathfrak{M}(\Omega)$ is 2-absolutely continuous and $|\dot\mu|_{\mathsf{He}}(s) \leq \frac{\sigma}{2}\,g(s)$ a.e. in $[0,1]$, which is the "$\leq$" part of (2.6).

With this Theorem 2.2 is established. ∎

2.4 Geodesics curves

According to [Kak48] (using the choice $\sigma = 1$) the Hellinger distance is given by

$$\mathsf{He}(\mu_0, \mu_1)^2 = \sigma^2\big(\mu_0(\Omega) + \mu_1(\Omega) - 2\sqrt{\mu_0\mu_1}(\Omega)\big).$$

One importance starting point of this paper is that this distance is a geodesic distance. For each pair $(\mu_0, \mu_1) \in \mathfrak{M}(\Omega)^2$ there exists a unique constant-speed geodesic (simply called geodesic in the sequel), i.e. a curve $\gamma : [0, 1] \to \mathfrak{M}(\Omega)$ such that

$$\gamma(0) = \mu_0, \ \ \gamma(1) = \mu_1, \ \ \mathsf{He}\big(\gamma(s), \gamma(t)\big) = |s{-}t|\, \mathsf{He}(\mu_0, \mu_1) \text{ for all } s, t \in [0, 1].$$

This geodesic is given by

$$\begin{aligned} \gamma^{\mathsf{He}}_{\mu_0\to\mu_1}(s) &= \Big(\big((1{-}s)\big(\frac{\mathrm{d}\mu_0}{\mathrm{d}\lambda}\big)^{1/2} + s\,\big(\frac{\mathrm{d}\mu_1}{\mathrm{d}\lambda}\big)^{1/2}\Big)^2\lambda \\ &= (1{-}s)^2\mu_0 + s^2\mu_1 + 2(s{-}s^2)\sqrt{\mu_0\mu_1} \\ &= (1{-}s)\mu_0 + s\mu_1 - (s{-}s^2)\frac{1}{\sigma^2}\mathsf{He}(\mu_0, \mu_1)^2 \ \text{ for } s \in [0, 1], \end{aligned}$$

where $\lambda \in \mathfrak{M}(\Omega)$ is arbitrary as long as $\mu_0, \mu_1 \ll \lambda$.

The growth equation along a geodesic can also be given explicitly, namely

$$\partial_s\gamma^{\mathsf{He}}_{\mu_0\to\mu_1}(s) = \xi(s,\cdot)\gamma^{\mathsf{He}}_{\mu_0\to\mu_1}(s) \quad \text{with } \xi(s, \omega) = \frac{2\,(f_1(\omega) - f_0(\omega))}{(1{-}s)f_0(\omega) + sf_1(\omega)}, \tag{2.11a}$$

where $f_j = \big(\,\mathrm{d}\mu_j/\,\mathrm{d}\lambda\big)^{1/2}$. Note that ξ satisfies the equation

$$\partial_s\xi(s, \omega) + \frac{1}{2}\xi(s, \omega)^2 = 0. \tag{2.11b}$$

which is completely independent of μ_0 and μ_1. The system (2.11) form the geodesic equations, which are a special case of the equations for the Hellinger-Kantorovich geodesics derived in [LMS16, Eqn. (5.1)] and [LMS18, Eqn. (8.72)].

An important feature is that along the geodesics the total mass of the measure is exactly quadratic, namely

$$\begin{aligned} \gamma^{\mathsf{He}}_{\mu_0\to\mu_1}(s)(\Omega) &= (1{-}s)^2\mu_0(s) + s^2\mu_1(\Omega) + 2(s{-}s^2)\sqrt{\mu_0\mu_1}(\Omega) \\ &= (1{-}s)\mu_0(\Omega) + s\mu_1(\Omega) - (s{-}s^2)\frac{1}{\sigma^2}\mathsf{He}(\mu_0, \mu_1)^2. \end{aligned} \tag{2.12}$$

In particular, one can define a Hellinger average by taking the midpoint of the geodesics:

$$A^{\mathsf{He}}(\mu_0, \mu_1) := \gamma^{\mathsf{He}}_{\mu_0\to\mu_1}(1/2) = \frac{1}{4}\mu_0 + \frac{1}{4}\mu_1 + \frac{1}{2}\sqrt{\mu_0\mu_1}. \tag{2.13}$$

2.5 Properties similar to Hilbert-space geometry

The last form of the geodesic already indicates that $(\mathfrak{M}(\Omega), \mathsf{He})$ is somehow related to the positive cone in the Hilbert space $\mathrm{L}^2(\Omega, \lambda)$. However, here the measure λ depends on the measures μ_j that are relevant for the current construction. This embedding is already included in [Kak48, Sec. 4], see Remark 2.3.

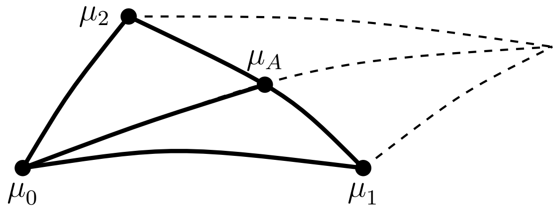

Figure 2.1: A visualization of the parallelogram identity in $(\mathfrak{M}(\Omega), \mathsf{He})$ where all edges (full lines) are geodesics, while the broken lines contain curves that may lead to signed measures lying in $\mathrm{L}^2(\Omega, \lambda)$. The center of the parallelogram is $\mu_A = A^{\mathsf{He}}(\mu_1, \mu_2)$.

(H1) We have a counterpart to the *parallelogram identity* for all $\mu_0, \mu_1, \mu_2 \in \mathfrak{M}(\Omega)$:

$$2\mathsf{He}(\mu_0, \mu_1)^2 + 2\mathsf{He}(\mu_0, \mu_2)^2 = \mathsf{He}(\mu_1, \mu_2)^2 + 4\,\mathsf{He}\big(\mu_0, A^{\mathsf{He}}(\mu_1, \mu_2)\big)^2, \tag{2.14}$$

where A denotes the average defined in (2.13). See Figure 2.1 for a visualization.

(H2) A second instance occurs when looking at the *squared distance* along geodesic curves. For every three points $\mu_0, \mu_1, \eta \in \mathfrak{M}(\Omega)$ we have *geodesic* 2*-convexity* as well as *geodesic* 2*-concavity*, i.e. for all $s \in [0, 1]$ we have

$$\mathsf{He}\big(\gamma^{\mathsf{He}}_{\mu_0\to\mu_1}(s), \eta\big)^2 = (1{-}s)\mathsf{He}(\mu_0, \eta)^2 + s\mathsf{He}(\mu_1, \eta)^2 - 2\,\frac{s{-}s^2}{2}\mathsf{He}(\mu_0, \mu_1)^2. \tag{2.15}$$

(H3) Finally, we may define *angles* between geodesics emanating from a point μ_0. Setting $\gamma_j(s) = \gamma^{\mathsf{He}}_{\mu_0\to\mu_j}(s)$ for $j = 1, 2$ the angle between geodesics is defined in the sense of geodesic spaces (see [BBI01, LaM19]) via

$$\sphericalangle(\gamma_1, \gamma_2) := \arccos\Big(\lim_{s,t\searrow 0} \frac{\mathsf{He}(\mu_0, \gamma_1(s))^2{+}\mathsf{He}(\mu_0, \gamma_2(t))^2{-}\mathsf{He}(\gamma_1(s), \gamma_2(t))^2}{2\,\mathsf{He}(\mu_0, \gamma_1(s))\,\mathsf{He}(\mu_0, \gamma_2(t))}\Big) \in [0, \pi]$$

whenever it exists. Exploiting the quadratic formula (2.15) a straightforward calculation shows that the fraction in the above definition is indeed constant (as in Hilbert spaces) and we find

$$\sphericalangle(\gamma_1, \gamma_2) = \arccos\Big(\frac{\mathsf{He}(\mu_0, \mu_1)^2{+}\mathsf{He}(\mu_0, \mu_2)^2{-}\mathsf{He}(\mu_1, \mu_2)^2}{2\,\mathsf{He}(\mu_0, \mu_1)\,\mathsf{He}(\mu_0, \mu_2)}\Big). \tag{2.16}$$

Clearly this is the same formula as in planar geometry, which is valid in all Hilbert spaces.

Remark 2.3 (Embedding into Euclidean space) *The observation that certain subsets of $\mathfrak{M}(\Omega)$ can be embedded into a Euclidean (Hilbert) space was crucial for the work in [Kak48]. There Section 4 is entitled "*Embedding into Euclidean space*". Moreover, the introduction contains the following text:*

> *The results of this paper were much amplified and the arguments used below much simplified, thanks to certain suggestions kindly made by Professor John von Neumann. In particular, the introduction of inner product and isometric embedding of $\mathfrak{M}(\Omega)$ into a general Euclidean space, as well as the indication of relationship of this paper with earlier works of E. Hellinger, as are discussed in §4, are due to Professor J. von Neumann. For all these I wish to express my heartiest thanks.*

Indeed, given any $\lambda \in \mathfrak{M}(\Omega)$ the subspace $\mathrm{L}^1(\Omega, \lambda) \subset \mathfrak{M}(\Omega)$ equipped with the Hellinger distance can be embedded isometrically into the Hilbert space $\mathrm{L}^2(\Omega, \lambda)$ via the mapping $g\lambda \mapsto \sqrt{g}\,\lambda$ (using the normalization $\sigma = 1$).

2.6 Hellinger distance for product measures

Already the paper [Kak48] shows that the Hellinger integral is very useful when studying product measures. Assuming $\Omega = \Omega_1 \times \Omega_2$ with associated σ-algebras we can consider measures $\nu_1 \in \mathfrak{M}(\Omega_1)$ and $\nu_2 \in \mathfrak{M}(\Omega_2)$ and define the product measure $\mu = \nu_1 \otimes \nu_2$ as the unique measure $\mu \in \mathfrak{M}(\Omega)$ satisfying

$$\nu_1 \otimes \nu_2 \big(A_1 \times A_2\big) = \nu_1(A_1)\, \nu_2(A_2).$$

It now follows easily from the definition of the Hellinger integral that it is compatible with the product structure in the sense that

$$\sqrt{(\nu_1 \otimes \nu_2)\, (\eta_1 \otimes \eta_2)} = \sqrt{\mu_1 \eta_1} \otimes \sqrt{\nu_2 \eta_2}. \tag{2.17}$$

Indeed, this relation and its generalization to infinite product measures is the basis of the analysis in [Kak48].

From this we can derive a corresponding formula for the Hellinger distance between two product measures, namely

$$\mathsf{He}\big(\nu_1 \otimes \nu_2, \eta_1 \otimes \eta_2\big)^2 = \sigma^2 \big(\overline{\nu}_1 \overline{\nu}_2 + \overline{\eta}_1 \overline{\eta}_2\big) - \frac{\sigma^2}{2} \big(\overline{\nu}_1 + \overline{\eta}_1 - \frac{1}{\sigma^2} \mathsf{He}(\nu_1, \eta_1)^2\big) \big(\overline{\nu}_2 + \overline{\eta}_2 - \frac{1}{\sigma^2} \mathsf{He}(\nu_2, \eta_2)^2\big), \tag{2.18}$$

where we abbreviated $\overline{\nu}_j = \nu_j(\Omega_j)$ and $\overline{\eta}_j = \eta_j(\Omega_j)$. We see that the Hellinger distance of the product measures can be expressed solely in therms of the total masses of the individual measures and the Hellinger distances between the corresponding factors.

The above result takes a much simpler form if we restrict to probability measures $\nu_1, \eta_1 \in \mathfrak{P}(\Omega_1)$ and $\nu_2, \eta_2 \in \mathfrak{P}(\Omega_2)$:

$$\mathsf{He}\big(\nu_1 \otimes \nu_2, \eta_1 \otimes \eta_2\big)^2 = \mathsf{He}(\nu_1, \eta_1)^2 + \mathsf{He}(\nu_2, \eta_2)^2 - \frac{1}{2\sigma^2} \mathsf{He}(\nu_1, \eta_1)^2 \mathsf{He}(\nu_2, \eta_2)^2. \tag{2.19}$$

We already emphasize at this point that in general a geodesic curve between two product measures does not stay a product measure any more. It is rather a convex combination of three product measures, namely

$$\gamma^{\mathsf{He}}_{\nu_0 \otimes \eta_0 \to \nu_1 \otimes \eta_1}(s) = (1-s)^2 \nu_0 \otimes \eta_0 + s^2 \nu_1 \otimes \eta_1 + 2(s-s^2) \sqrt{\nu_0 \eta_0} \otimes \sqrt{\nu_1 \eta_1}.$$

To find the shortest connecting path consisting of product measures will be treated as a special case of a Fisher-Rao distance in Section (3.4).

The case of product measures with more than two factors (as in [Kak48]) works analogously: For $\nu_j, \eta_j \in \mathfrak{P}(\Omega_j)$ we have

$$\begin{aligned} 1 - \frac{1}{2\sigma^2} \mathsf{He}\Big(\bigotimes_{i=1}^n \nu_i, \bigotimes_{j=1}^n \eta_j \Big)^2 &= \sqrt{(\otimes \nu_i)(\otimes \eta_j)} \Big(\mathop{\times}_{k=1}^n \Omega_k \Big) \\ &= \prod_{k=1}^n \sqrt{\nu_k \eta_k}(\Omega_k) = \prod_{k=1}^n \big(1 - \frac{1}{2\sigma^2} \mathsf{He}(\nu_k, \eta_k)^2 \big). \end{aligned} \tag{2.20}$$

2.7 Invariance under pushforwards of the Hellinger distance

Considering two measure spaces $(\Omega, \mathfrak{A})$ and $(\Sigma, \mathfrak{B})$ and a measurable mapping $\Phi : \Omega \to \Sigma$, the pushforward $\Phi_\# \mu \in \mathfrak{M}(\Sigma)$ of μ is defined via

$$\Phi_\# \mu(B) := \mu\big(\Phi^{-1}(B)\big) \text{ for all } B \in \mathfrak{B} \text{ and } \mu \in \mathfrak{M}(\Omega),$$

where $\Phi^{-1}(B) := \{ x \in \Omega \,|\, \Phi(x) \in B \} \subset \mathfrak{A}$ such that Φ does not need to be injective. See [AGS05, Sec. 5.2] for this and further properties of pushforward measures.

Using the infimum characterization (2.2) of $\sqrt{\mu_0\mu_1}$ and comparing the admissible partitions for the two sides we easily find

$$\Phi_\#\big(\sqrt{\mu_0\mu_1}\big)(B) \leq \sqrt{(\Phi_\#\mu_0)(\Phi_\#\mu_1)}(B) \text{ for all } B \in \mathfrak{B}.$$

Clearly, we have $\Phi_\#\mu_j(\Sigma) = \mu_j(\Omega)$, hence we immediately find the monotonicity of the Hellinger distance under pushforward, with equality if the operation can be reversed.

Lemma 2.4 (Hellinger distance and pushforward) *For measure spaces $(\Omega, \mathfrak{A})$ and $(\Sigma, \mathfrak{B})$ and a measurable mapping $\Phi : \Omega \to \Sigma$ we have*

$$\forall\, \mu_0, \mu_1 \in \mathfrak{M}(\Omega) : \ \mathsf{He}_\Sigma\big(\Phi_\#\mu_0, \Phi_\#\mu_1\big) \leq \mathsf{He}_\Omega(\mu_0, \mu_1). \tag{2.21}$$

If additionally Ψ is one-to-one with measurable inverse Ψ^{-1}, then

$$\forall\, \mu_0, \mu_1 \in \mathfrak{M}(\Omega) : \ \mathsf{He}_\Sigma\big(\Psi_\#\mu_0, \Psi_\#\mu_1\big) = \mathsf{He}_\Omega(\mu_0, \mu_1). \tag{2.22}$$

In [BBM16] the case $\Sigma = \Omega$ being a smooth finite-dimensional manifold without boundary and dimension ≥ 2 is studied, and it is shown that the Hellinger distance (called Fisher-Rao metric there) restricted to probability measures is the only "Riemannian distance" that has the invariance property (2.22). There the theory is restricted to smooth densities (instead of measures) and diffeomorphisms.

In [BBM16, last parag.] it is shown that asking (2.22) only for smooth diffeomorphisms on $\Omega = \mathbb{S}^1$ allows for more general distances $\mathcal{D}$ in $\mathfrak{P}(\Omega)$ than multiplies of He. Hence, it would be interesting to know whether enforcing (2.22) also for general measurable homeomorphism rules out this pathology. More generally, one might conjecture every distance $\mathcal{D} : \mathfrak{P}(\Omega)\times\mathfrak{P}(\Omega) \to [0, \infty[$ satisfying the invariance (2.22) and the properties (H1), (H2), and (H3) in Section 2.5 is a multiple of He.

2.8 Historical remarks

In his dissertation [Hel07] and habilitation thesis [Hel09], Hellinger introduced integrals of the type $\int_a^b u(t) \frac{\mathrm{d}f_1 \mathrm{d}f_2}{\mathrm{d}g}$ for functions $u, f_1, f_2, g, h \in \mathrm{C}^0([a, b])$ where additionally g and h are increasing and satisfy $\big(f_j(t_2){-}f_j(t_1)\big)^2 \leq \big(g(t_2){-}g(t_1)\big)\big(h(t_2){-}h(t_1)\big)$ for all t_1, t_2 with $a \leq t_1 < t_2 \leq b$. In modern terns using the Radon-Nikodým derivative, we would introduce a dominating measure $\lambda \in ([a, b])$ and assume $\mathrm{d}f_j = \phi_j\, \mathrm{d}\lambda$, $\mathrm{d}g = \gamma\, \mathrm{d}\lambda$, and $\mathrm{d}h = \eta\, \mathrm{d}\lambda$ with the restriction $\phi_j^2 \leq \gamma\eta$. Then, $\phi_1\phi_2/\gamma \leq \eta$ a.e. with respect to λ, and Hellinger's integral can be interpreted in the form

$$\int_a^b u(t) \frac{\mathrm{d}f_1 \mathrm{d}f_2}{\mathrm{d}g} := \int_a^b u(t) \frac{\phi_1(t)\phi_2(t)}{\gamma(t)}\, \mathrm{d}\lambda(t). \tag{2.23}$$

However, as Hellinger's construction was much before the introduction of the Radon-Nikodým derivative, he used a convexity argument that is reminiscent to the concavity in the definition of $\sqrt{\mu_1\mu_2}$ in (2.2): Restricting the integral in (2.23) to the case $u \equiv 1$ and $f = f_1 = f_2$ one shows that

$$\int_a^b \frac{\mathrm{d}f^2}{\mathrm{d}g} = \lim_{\Delta(\Pi)\to 0} \sum_{t_i \in \Pi} \frac{\big(f(t_i){-}f(t_{i-1})\big)^2}{g(t_i){-}g(t_{i-1})} = \sup_{\Pi \in \mathrm{Part}([a,b])} \sum_{t_i \in \Pi} \frac{\big(f(t_i){-}f(t_{i-1})\big)^2}{g(t_i){-}g(t_{i-1})},$$

see [Hel09, §4, p. 234], because the discrete sum is increasing under refinements of the partitions Π of the interval $[a,b]$, by using the estimate

$$\frac{\big(f(t_i)-f(t_{i-2})\big)^2}{g(t_i)-g(t_{i-2})} \leq \frac{\big(f(t_i)-f(t_{i-1})\big)^2}{g(t_i)-g(t_{i-1})} + \frac{\big(f(t_{i-1})-f(t_{i-2})\big)^2}{g(t_{i-1})-g(t_{i-2})}.$$

In [Kol30] the argument was generalized to arbitrary measure spaces, defining so-called Kolmogorov integrals by using the infimum construction, but an explicit reference of Hellinger's work is given [Kol30, p. 679], referring explicitly to [Hel09, p.234].

Using the modern tool of the Radon-Nikodým derivative, [Kak48, Eqn. (11)] introduces the so-called Hellinger integral $\rho(\mu,\nu) = \int_\Omega \sqrt{\mu(\mathrm{d}\omega)\nu(\mathrm{d}\omega)}$ and defines what is nowadays called the Hellinger distance on probability measures via $\mathsf{He}(\mu,\nu) = \big(2-2\rho(\mu,\nu)\big)^{1/2}$.

Since the early 1960s, the name Hellinger distance is consistently used in probability theory and statistics (see. e.g. [LeC70]), which was checked by a search of "Hellinger distance" in MathSciNet in 2023, which led to more than 600 hits in abstracts or titles. In particular, Rao's paper [RaV63, §3, p. 304] introduces the Hellinger integral and the Hellinger distance explicitly by name.

The Fisher-Rao distance was popularized by [Rao45] as geodesic distance for the Fisher information metric. It is interesting to see that the abstract version of the Fisher metric given in (1.4) is exactly of the form of the Hellinger integrals (2.23) introduced already in [Hel09], however in a rather restrictive setting.

3 Various Fisher-Rao distances

We first discuss the general construction of the Fisher-Rao distance $\mathsf{FR}_\mathcal{S}$ for general subset $\mathcal{S} \subset \mathfrak{M}(\Omega)$ without direct reference to the local Fisher information metric $\mathbf{g}_\mu(\nu_1,\nu_2)$ defined in (1.4).

3.1 The general construction for subsets $\mathcal{S} \subset \mathfrak{M}(\Omega)$

Throughout, our subsets $\mathcal{S}$ will be path-connected, i.e. between any to points $\mu_0, \mu_1 \in \mathcal{S}$ there exists a continuous path $\gamma \in \mathrm{C}^0([0,1];\mathfrak{M}(\Omega))$ with $\gamma(s) \in \mathcal{S}$ for all $s \in [0,1]$. The intrinsic length of γ is defined by

$$L_{\mathsf{He}}(\gamma) = \sup\Big\{ \sum_{i=1}^N \mathsf{He}\big(\gamma(s_i),\gamma(s_{i-1})\big) \;\Big|\; N\in\mathbb{N},\ 0=s_0<s_1<\cdots<s_{N-1}<s_N=1 \Big\}, \tag{3.1}$$

and $L_{\mathsf{He}}(\gamma) < \infty$ means that γ is rectifiable in $(\mathfrak{M}(\Omega),\mathsf{He})$. In that case, we can change the parametrization with a monotone function $t : [0,1] \to [0,1]$ such that $\widetilde\gamma = \gamma\circ t$ has constant speed, namely

$$\mathsf{He}\big(\widetilde\gamma(r),\widetilde\gamma(s)\big) = \mathsf{He}\big(\gamma(t(r)),\gamma(t(s))\big) = |r-s|\, L_{\mathsf{He}}(\gamma) \quad \text{for all } r,s\in[0,1]. \tag{3.2}$$

To see this, consider the function $\ell : [0,1] \to [0,1]$ with $\ell(t) = L_{\mathsf{He}}\big(\gamma|_{[0,t]}\big)/L_{\mathsf{He}}(\gamma)$, which is continuous, non-decreasing and surjective. Now, we can choose any $t : [0,1] \to [0,1]$ such that $t\circ\ell = \mathrm{id}_{[0,1]}$, i.e. $t(s) \in \ell^{-1}(\{s\})$. Using the metric derivative or speed) $|\widetilde\gamma|_{\mathsf{He}}(s)$ as defined in Section 2.3, one then has the relation

$$L_{\mathsf{He}}(\gamma)^2 = L_{\mathsf{He}}(\widetilde\gamma)^2 = \lim_{N\to\infty} \sum_{i=1}^N N\,\mathsf{He}\big(\widetilde\gamma(i/N),\widetilde\gamma((i-1)/N)\big)^2. \tag{3.3}$$

Using the length L_{He}, the Fisher-Rao distance $\mathsf{FR}_{\mathcal{S}}$ for the subset $\mathcal{S}$ is defined by

$$\mathsf{FR}_{\mathcal{S}}(\mu_0,\mu_1) := \inf\Big\{ \, L_{\mathsf{He}}(\gamma) \, \Big| \, \gamma\in \mathrm{C}^0([0,1];\mathcal{S}),\ \gamma(0)=\mu_0,\ \gamma(1)=\mu_1) \,\Big\} \tag{3.4a}$$

$$= \inf\Big\{ \int_0^1 \big(|\dot\gamma|_{\mathsf{He}}(s)\big)^2 \,\mathrm{d}s \, \Big| \, \gamma \text{ 2-absol. contin.}, \gamma(0)=\mu_0,\ \gamma(1)=\mu_1) \,\Big\}. \tag{3.4b}$$

From the definition we immediately obtain the lower estimate

$$\mathsf{FR}_{\mathcal{S}}(\mu_0,\mu_1) \geq \mathsf{He}(\mu_0,\mu_1) \quad \text{for all } \mu_0,\mu_1\in\mathcal{S}, \tag{3.5}$$

and equality can only hold if the geodesic curve $\gamma^{\mathsf{He}}_{\mu_0\to\mu_1}$ (cf. (1.3)) is contained in $\mathcal{S}$.

If $\mathcal{S}$ is a smooth manifold that is given by a parameter $p\in D\subset X$, where X is a Banach space (e.g. $\mathbb{R}^m$) in the form $\mathcal{S} = \big\{ \widehat\mu(p) \,\big|\, p\in D\subset X \,\big\}$, then the induced metric tensor, also called Fisher's information matrix. can be reconstructed via

$$\big\langle \widehat{\mathbb{G}}(p)v,v\big\rangle = \lim_{\varepsilon\to 0^+} \frac{1}{\varepsilon^2} \mathsf{He}\big(\widehat\mu(p),\widehat\mu(p{+}\varepsilon v)\big)^2 \tag{3.6}$$

3.2 Bhattacharya distance alias spherical Hellinger distance

The simplest and still very important submanifold in $\mathfrak{M}(\Omega)$ is the set of probability measures $\mathfrak{P}(\Omega)\subset \mathfrak{M}(\Omega)$.

Since the Hellinger distance satisfies the general scaling property

$$\mathsf{He}(r_0^2\mu_0,r_1^2\mu_1)^2 = r_0r_1\mathsf{He}(\mu_0,\mu_1)^2 + \sigma^2(r_0^2{-}r_0r_1)\mu_0(\Omega) + \sigma^2(r_1^2{-}r_0r_1)\mu_1(\Omega), \tag{3.7}$$

we can interpret the set $\mathfrak{M}(\Omega)$ of all (non-negative) measures as a metric cone over the base space $\mathfrak{P}(\Omega)$, in the sense of [BBI01, §3.6]. This general geometric construction implies that the induced Fisher-Rao distance $\mathsf{FR}_{\mathcal{P}}$ on $\mathfrak{P}(\Omega)$, which is also called *Bhattacharya distance* Bh (cf. [Rao45]), as well as the exact form of the geodesics can be given explicitly, see [LaM19, Sec. 2] for the details. In the latter work this distance is called the *spherical Hellinger distance* because the base space $\mathcal{P}(\Omega)$ is called the spherical space of the cone. We obtain

$$\mathsf{Bh}(\nu_0,\nu_1) = 2\sigma \arcsin\big(\frac{1}{2\sigma}\mathsf{He}(\nu_0,\nu_1)\big) = \sigma\arccos\big(1{-}\frac{1}{2\sigma^2}\mathsf{He}(\nu_0,\nu_1)^2\big),$$

Note that He takes values in $[0,\sigma\sqrt{2}]$ (because $\nu_j(\Omega)=1$), whereas Bh takes values in $[0,\sigma\,\pi/2]$. The maximum values are achieved if ν_0 and ν_1 are mutual singular such that $\sqrt{\nu_0\nu_1}=0$, e.g. for Dirac measures $\nu_j=\delta_{\omega_j}$ with $\omega_0\neq\omega_1$.

According to [LaM19, Thm. 2.7], the geodesics take the form

$$\gamma^{\mathsf{Bh}}_{\nu_0\to\nu_1}(s) = n(s)\gamma^{\mathsf{He}}_{\nu_0\to\nu_1}(t(s)) \quad \text{with } t(s) = \frac{\sin(s\delta)}{\sin\big((1{-}s)\delta\big)+\sin(s\delta)} \in [0,1],$$

$$\delta = \frac{1}{\sigma}\,\mathsf{Bh}(\nu_0,\nu_1)\in[0,\tfrac{\pi}{2}], \text{ and } n(s) = \Big(\frac{\sin\big((1{-}s)\delta\big)+\sin(s\delta)}{\sin(\delta)}\Big)^2 \in [1,2].$$

Recalling $\gamma^{\mathsf{He}}_{\nu_0\to\nu_1}(t)(\Omega) = 1-(t{-}t^2)\frac{1}{\sigma^2}\mathsf{He}(\nu_0,\nu_1)^2 = 1-2(t{-}t^2)(1{-}\cos\delta)$ we indeed find $n(s)\gamma^{\mathsf{He}}_{\nu_0\to\nu_1}(t(s))(\Omega)\equiv 1$, i.e. $\gamma^{\mathsf{Bh}}_{\nu_0\to\nu_1}(s)\in\mathfrak{P}(\Omega)$.

Of course, one can also go opposite and consider $\mathfrak{M}(\Omega)$ as a cone over $\mathfrak{P}(\Omega)$, i.e. $\mathfrak{M}(\Omega) = [0,\infty[\mathfrak{P}(\Omega)$. Then, we have the relation

$$\mathsf{He}(r_0^2\nu_0, r_1^2\nu_1)^2 = r_0 r_1 \mathsf{He}(\nu_0,\nu_1)^2 + \sigma^2(r_1{-}r_0)^2 \tag{3.8a}$$
$$= \sigma^2\big((r_1{-}r_0)^2 + 2r_0r_1\big(1{-}\cos\big(\tfrac{1}{\sigma}\mathsf{Bh}(\nu_0,\nu_1)\big)\big)\big). \tag{3.8b}$$

Remark 3.1 (Unique characterization) *In [KL*13] the question is studied whether* Bh *is the only Riemannian distance (up to a positive scalar factor) on a finite-dimensional, smooth manifold that is invariant under all pushforwards (cf. Section 2.7) with respect to smooth diffeomorphisms. It is shown that this is true for dimension* $n \geq 2$ *but it may fail for* $n = 1$. *It is unclear whether the pathology for* $n = 1$ *disappears if pushforwards for all measurable homeomorphisms are considered.*

3.3 General cones

A special property of the Hellinger distance is the scaling property (3.7) that suggests that the transition between $\mathfrak{M}(\Omega)$ and $\mathfrak{P}(\Omega)$ can be seen as the transition between the cone $\mathbb{C}_\mathcal{P} \subset \mathfrak{M}(\Omega)$ and its base space $\mathcal{P} \subset \mathfrak{P}(\Omega)$:

$$\mathcal{P} \subset \mathfrak{P}(\Omega) \quad \text{and} \quad \mathbb{C}_\mathcal{P} := [0,\infty[\mathcal{P} := \big\{\, r^2\nu \,\big|\, r \in [0,\infty[,\ \nu \in \mathcal{P}\,\big\} = \mathcal{S},$$

Here $\mathcal{P}$ is chosen arbitrarily such that $(\mathcal{P}, \mathsf{FR}_\mathcal{P})$ is a length space. The following result gives an explicit formula for $\mathsf{FR}_\mathcal{S}$ in terms of $\mathsf{FR}_\mathcal{P}$. Whenever $\mathsf{FR}_\mathcal{P}(\nu_0,\nu_1) \geq \sigma\pi$ we will find $\mathsf{FR}_\mathcal{S}(r_0^2\nu_0, r_1^2\nu_1) = \sigma(r_0+r_1)$ and the corresponding geodesic curve is given by

$$\gamma^\mathcal{S}_{r_0^2\nu_0\to r_1^2\nu_1}(s) = \begin{cases} \big(r_0 - (r_0{+}r_1)s\big)^2\nu_0 & \text{for } s \in [0, r_0/(r_0{+}r_1)], \\ \big((r_0{+}r_1)s - r_0\big)^2\nu_1 & \text{for } s \in [r_0/(r_0{+}r_1), 1]. \end{cases}$$

For $\delta := \dfrac{1}{\sigma}\,\mathsf{FR}_\mathcal{P}(\nu_0,\nu_1) < \pi$ geodesics in $\mathcal{S}$ can be expressed by geodesics in $\mathcal{P}$ via

$$\gamma^\mathcal{S}_{r_0^2\nu_0\to r_1^2\nu_1}(s) = \widehat{r}(s)^2\,\gamma^\mathcal{P}_{\nu_0\to\nu_1}\big(\zeta(s)\big) \text{ with } \zeta(s) = \frac{1}{\delta}\arcsin\Big(s\,\frac{r_1\sin\delta}{\widehat{r}(s)}\Big)$$
$$\text{and } \widehat{r}(s)^2 = (1{-}s)^2r_0^2 + s^2r_1^2 + 2(s{-}s^2)r_0r_1\cos_\pi\delta.$$

We refer to [LaM19, Sec. 2.3] for these formulas of the geodesics, while the formula for FR_S given below is from [BBI01, §3.6]. Here we give a sketch of an alternative proof using metric speeds.

Theorem 3.2 (Fisher-Rao distance on cones) *If* $(\mathcal{P}, \mathsf{FR}_\mathcal{P})$ *is a length space and* $S = \mathbb{C}_\mathcal{P} \subset \mathfrak{M}(\Omega)$*, then* $(\mathcal{S}, \mathsf{FR}_S)$ *is a length space with*

$$\mathsf{FR}_\mathcal{S}(r_0^2\nu_0, r_1^2\nu_1)^2 = \sigma^2\Big(r_0^2 + r_1^2 - 2r_0r_1\cos_\pi\big(\tfrac{1}{\sigma}\mathsf{FR}_\mathcal{P}(\nu_0,\nu_1)\big)\Big), \tag{3.9}$$

where $\cos_\pi(r) = \cos\big(\min\{|r|,\pi\}\big)$.

Sketch of proof. A curve $s \mapsto \mu(s) = r(s)^2\nu(s) \in \mathcal{S} \subset \mathfrak{M}(\Omega)$ can having finite length has a metric speed a.e. in $[0,1]$. According to Theorem 2.2 we can calculate the speed via Hellinger's quadratic form

$$\big(|\dot\mu|_{\mathsf{He}}(s)\big)^2 = \int_\Omega \xi_s^2 \,\mathrm{d}\mu_s.$$

Similarly, we can calculate the metric speed of $s \mapsto \nu(s) \in \mathcal{P} \subset \mathfrak{P}(\Omega)$. From (3.8a) we obtain

$$
\begin{aligned}
|\dot\mu|_{\mathsf{He}}(s)^2 &= \lim_{h\searrow} \frac{1}{h^2}\, \mathsf{He}(\mu(s),\mu(s{+}h))^2 \\
&= \lim_{h\searrow} \frac{1}{h^2}\Big(r(s{+}h)r(s)\mathsf{He}(\nu(s),\nu(s{+}h))^2 + \sigma^2\big(r(s{+}h)-r(s)\big)^2\Big) \\
&= r(s)^2|\dot\nu|_{\mathsf{He}}(s)^2 + \big(r'(s)\big)^2.
\end{aligned}
$$

Since $\mathsf{FR}_{\mathcal{S}}(r_0^2\nu_0, r_1^2\nu_1)^2$ is given of the infimum over

$$
\int_0^1 |\dot\mu|_{\mathsf{He}}(s)^2 \,\mathrm{d}s = \int_0^1 \big(r(s)^2|\dot\nu|_{\mathsf{He}}(s)^2 + \sigma^2\big(r'(s)\big)^2\big)\,\mathrm{d}s
$$

subject to the boundary conditions $\mu(j) = \mu_J = r_j^2\nu_j$ for $j = 0, 1$, we see obtain

$$
\mathsf{FR}_{\mathcal{S}}(r_0^2\nu_0, r_1^2\nu_1)^2 = \inf\Big\{ \int_0^1 \big(r^2y^2{+}(\sigma r')^2\big)\mathrm{d}s \;\Big|\; r(0)=r_0,\ r(1)=r_1,\ \int_0^1 y\mathrm{d}s = \mathsf{FR}_{\mathcal{P}}(\nu_0,\nu_1)\Big\}.
$$

This minimization problem has been analyzed explicitly by [LMS16, Thm. 2] for the case $\sigma = 1$ (if one sets $\alpha = 1$ and $\beta = 4$ there). A crucial point is to realize that $ry = \mathrm{const}$ along minimizers. The case of general σ follows by scaling replacing y by σy, thus rescaling $\mathsf{FR}_{\mathcal{P}}$ by a factor σ, and pulling out the factor σ^2.

This yields the desired formula (3.9). ■

3.4 Product measures

In applications one is often interested in situations where the basic measure space is a product space, viz. $\Omega = \Omega_1{\times}\Omega_2$. Given subsets $\mathcal{S}_1 \subset \mathfrak{M}(\Omega_1)$ and $\mathcal{S}_2 \subset \mathfrak{M}(\Omega_2)$ one is then interested in the Fisher-Rao distance for the subset

$$
\mathcal{S}_1 \otimes \mathcal{S}_2 := \big\{ \mu_1{\otimes}\mu_2 \;\big|\; \mu_1 \in \mathcal{S}_1, \mu_2 \in \mathcal{S}_2 \big\}.
$$

The natural question is whether $\mathsf{FR}_{\mathcal{S}_1\otimes\mathcal{S}_2}$ can be expressed of estimated by $\mathsf{FR}_{\mathcal{S}_1}$ and $\mathsf{FR}_{\mathcal{S}_2}$.

A positive and simple answer can be given in the case that $\mathcal{S}_j$ are contained in the probability measures $\mathfrak{P}(\Omega_j)$.

Proposition 3.3 (Product probability measures) *Assume that $\mathcal{P}_j \subset \mathfrak{P}(\Omega_j)$ and that $\mathsf{FR}_{\mathcal{P}_j}$ are finite for $j = 1, 2$, then we have*

$$
\mathsf{FR}_{\mathcal{P}_1\otimes\mathcal{P}_2}(\nu_1{\otimes}\nu_2, \eta_1{\otimes}\eta_2)^2 = \mathsf{FR}_{\mathcal{P}_1}(\nu_1,\eta_1)^2 + \mathsf{FR}_{\mathcal{P}_2}(\nu_2,\eta_2)^2 \quad \textit{for all } \nu_j,\eta_j \in \mathfrak{P}(\Omega_j). \tag{3.10}
$$

Proof. Since we are working with probability measures we can use the simple representation (2.19) for the Hellinger distance of product measures. We first observe that $\mathsf{He}(\nu_j,\eta_j)^2 \leq 2\sigma^2$ implies

$$
\begin{aligned}
\frac12\big(\mathsf{He}(\nu_1,\eta_1)^2 + \mathsf{He}(\nu_2,\eta_2)^2\big) &\leq \mathsf{He}(\nu_1,\eta_1)^2 + \mathsf{He}(\nu_2,\eta_2)^2 - \frac{1}{2\sigma^2}\mathsf{He}(\nu_1,\eta_1)^2\mathsf{He}(\nu_2,\eta_2)^2 \\
&= \mathsf{He}(\nu_1{\otimes}\eta_1, \nu_2,\eta_2) \leq \mathsf{He}(\nu_1,\eta_1)^2 + \mathsf{He}(\nu_2,\eta_2)^2.
\end{aligned}
$$

Hence the curve $s \mapsto \mu(s)$ is rectifiable if and only if the two curves $s \mapsto \nu_j(s)$ are rectifiable.

This allows us to calculate the metric speed for curves $\mu(s) = \nu_1(s) \otimes \nu_2(s)$ as follows. For a.a. $s \in [0,1]$ the three metric derivatives $|\dot\mu|_{\mathsf{He}}$, $|\dot\nu_1|_{\mathsf{He}}$, and $|\dot\nu_2|_{\mathsf{He}}$ exists and for those points we have .

$$\begin{aligned} |\dot\mu|_{\mathsf{He}}(s)^2 &= \lim_{h\searrow 0} \frac{1}{h^2} \mathsf{He}\big(\mu(s), \mu(s{+}h)\big)^2 \\ &\overset{(2.19)}{=} \lim_{h\searrow 0} \frac{1}{h^2} \Big(\mathsf{He}\big(\nu_1(s), \nu_1(s{+}h)\big)^2 + \mathsf{He}\big(\nu_2(s), \nu_2(s{+}h)\big)^2 \\ &\qquad\qquad - \frac{1}{2\sigma^2} \mathsf{He}\big(\nu_1(s), \nu_1(s{+}h)\big)^2 \mathsf{He}\big(\nu_2(s), \nu_2(s{+}h)\big)^2 \Big) \\ &= \lim_{h\searrow 0} \frac{1}{h^2} \mathsf{He}\big(\nu_1(s), \nu_1(s{+}h)\big)^2 + \lim_{h\searrow 0} \frac{1}{h^2} \mathsf{He}\big(\nu_2(s), \nu_2(s{+}h)\big)^2 - 0 \\ &= |\dot\nu_1|_{\mathsf{He}}(s)^2 + |\dot\nu_2|_{\mathsf{He}}(s)^2. \end{aligned}$$

Thus, formula (3.10) follows immediately as calculating the Fisher-Rao distances via the characterization in (3.4b). ■

This simple additive structure for the Fisher-Rao distance of product measures disappears if one leaves the realm of probability measures. Relation (2.18) can be rewritten in the form

$$\begin{aligned} \mathsf{He}\big(\nu_1 \otimes \nu_2, \eta_1 \otimes \eta_2\big)^2 &= \tfrac{1}{2}\big(\overline\nu_2 + \overline\eta_2\big) \mathsf{He}(\nu_1, \eta_1)^2 + \tfrac{1}{2}\big(\overline\nu_1 + \overline\eta_1\big) \mathsf{He}(\nu_2, \eta_2)^2 \\ &\quad + \sigma^2 \big(\overline\nu_1 - \overline\eta_1\big)\big(\overline\nu_2 - \overline\eta_2\big) - \frac{1}{2\sigma^2} \mathsf{He}(\nu_1, \eta_1)^2 \mathsf{He}(\nu_2, \eta_2)^2, \end{aligned} \tag{3.11}$$

where $\overline\nu_j = \nu_j(\Omega_j)$ and $\overline\eta_j = \eta_j(\Omega_j)$. Hence, for curves $\mu(s) = \nu(s) \otimes \eta(s)$ we obtain the metric speed

$$|\dot\mu|^2_{\mathsf{He}} = m_\eta(s) |\dot\nu|^2_{\mathsf{He}} + m_\nu(s) |\dot\eta|^2_{\mathsf{He}} + \sigma^2 m_\nu'(s) m_\eta'(s)$$

with $m_\mu(s) = \nu_s(\Omega)$ and $m_\eta(s) = \eta_s(\Omega)$. Hence, there is a much stronger interaction between the measures in $\mathcal{S}_1 \subset \mathfrak{M}(\Omega_1)$ and those in $\mathcal{S}_2 \subset \mathfrak{M}(\Omega_2)$, and in the general case an explicit form for $\mathsf{FR}_{\mathcal{S}_1 \otimes \mathcal{S}_2}$ seems out of reach.

There is one case that can be treated, namely if $\mathcal{S}_1$ and $\mathcal{S}_2$ are cones over $\mathcal{P}_1$ and $\mathcal{P}_2$, respectively. In this case, we have

$$S_1 \otimes S_2 = \mathbb{C}_{\mathcal{P}_1} \otimes \mathbb{C}_{\mathcal{P}_2} = \mathbb{C}_{\mathcal{P}_1 \otimes \mathcal{P}_2}.$$

Thus, we can first apply Proposition 3.3 to obtain $\mathsf{FR}_{\mathcal{P}_1 \otimes \mathcal{P}_2}$ and afterwards invoke Theorem 3.2.

4 Classical families of probability distributions

As applications of the above theory we treat a few classical examples. For further applications we refer to [May16].

4.1 Translations of a measure

As a first example we treat the case $\Omega = \mathbb{R}^n$, fix a $\mu \in (\mathbb{R}^n)$, and using the diffeomorphisms $\Phi^y : x \mapsto x{+}y$ we define

$$\mathcal{S}_{\mathrm{trans}}(\mu) := \big\{ \Phi^y_\# \mu \;\big|\; y \in \mathbb{R}^n \big\}.$$

Observe that $\mathcal{S}_{\rm trans}(\mu)$ is not path connected if μ has a discrete part, because for Dirac measures we have $\Phi^y\delta_z = \delta_{z-y}$ and $\mathsf{He}(\delta_{x_1},\delta_{x_2}) = \sqrt{2}$ for $x_1 \neq x_2$. If μ is absolutely continuous with respect to the Lebesgue measure, $\mathcal{S}_{\rm trans}(\mu)$ is path connected but the intrinsic length may still be infinite for nontrivial curves. For this consider $n = 1$ and $\widetilde\mu = \mathbf{1}_{[0,1]}\,\mathrm{d}x$ giving the distances

$$\mathsf{He}(\Phi^y_\#\widetilde\mu,\Phi^z_\#\widetilde\mu)^2 = \sigma^2\int_{\mathbb{R}}\big(\mathbf{1}_{[0,1]}(x{-}y) - \mathbf{1}_{[0,1]}(x{-}z)\big)\,\mathrm{d}x = 2\sigma^2\min\big\{|y{-}x|,1\big\}.$$

Hence, for every non-constant curve γ in $\mathcal{S}_{\rm trans}(\widetilde\mu)$ we have $L_{\mathsf{He}}(\gamma) = \infty$, which implies $\mathsf{FR}_{\mathcal{S}_{\rm trans}(\widetilde\mu)}(\Phi^y_\#\widetilde\mu,\Phi^z_\#\widetilde\mu) = \infty$ for $y \neq z$.

However, considering $\widehat\mu = f\,\mathrm{d}x$ with $\sqrt{f}\in \mathrm{H}^1(\mathbb{R}^d)$, it is straightforward to show that

$$\big(\mathsf{FR}_{\mathcal{S}_{\rm trans}(\widehat\mu)}(\Phi^y_\#\widehat\mu,\Phi^z_\#\widehat\mu)\big)^2 = (z{-}y)\cdot\mathbb{A}(z{-}y),$$

where the induced translation invariant Riemannian metric on the parameter space $D = \mathbb{R}^n$ is given by

$$\mathbb{A} := \int_{\mathbb{R}^d}\frac{\sigma^2}{f}\,\nabla f\otimes\nabla f\,\mathrm{d}x = \int_{\mathbb{R}^d}4\sigma^2\,\nabla\sqrt{f}\otimes\nabla\sqrt{f}\,\mathrm{d}x.$$

Clearly, the induced distance on $\mathbb{R}^n$ is translation invariant. If there are further symmetries (reflections or rotations) of the measure $\mu = f\,\mathrm{d}x$ they are reflected in the induced matrix $\mathbb{A}$ by applying Lemma 2.4.

4.2 Submanifold of Poisson measures on $\mathbb{N}_0^d$

For $\Omega = \mathbb{N}_0^d$ the multivariate Poisson distribution π_α for $\alpha\in[0,\infty[^d$ is given by (where $n = (n_i)_i\in\mathbb{N}_0^d$)

$$\pi_\alpha\big(\{n\}\big) = \frac{\mathrm{e}^{-\overline\alpha}\,\alpha^n}{n!}\quad\text{with }\ \overline\alpha = \sum_{i=1}^d\alpha_i,\ \ \alpha^n = \prod_{i=1}^d\alpha_i^{n_i},\ \text{ and } n! = \prod_{i=1}^d n_i!\,.$$

The Hellinger distance between π_α and π_β can easily calculated by observing that $\sqrt{\pi_\alpha\pi_\beta}$ is a multiple of $\pi_{\frac12(\alpha+\beta)}$, giving

$$\mathsf{He}(\pi_\alpha,\pi_\beta)^2 = 2\sigma^2\big(1-\mathrm{e}^{b(\alpha,\beta)}\big)\ \text{ with } b(\alpha,\beta) = \sum_{i=1}^d\sqrt{\alpha_i\beta_i} - \tfrac{\alpha_i+\beta_i}{2} = -\frac12\sum_{i=1}^d\big(\sqrt{\alpha_i}-\sqrt{\beta_i}\big)^2.$$

The Fisher metric tensor on the d-dimensional manifold

$$\mathcal{S}_{\rm Poiss}(\mathbb{N}_0^d) = \big\{\,\pi_\alpha\;\big|\;\alpha\in[0,\infty[^d\,\big\}$$

is now most easily constructed by applying (3.6), giving

$$\big\langle\mathbb{G}_{\rm Poiss}(\alpha)v,v\big\rangle = \frac{\sigma^2}{4}\sum_{i=1}^d\frac{v_i^2}{\alpha_i}.$$

With this, the associated Fisher-Rao distance for $\mathcal{S}_{\rm Poiss}(\mathbb{N}_0^d)$ can be calculated explicitly and gives the Hellinger distance on the positive orthant $[0,\infty[^d = \mathcal{M}(\{1,..,d\})$ of $\mathbb{R}^d$, viz.

$$\mathsf{FR}_{\rm Poiss}(\pi_\alpha,\pi_\beta)^2 = \sigma^2\sum_{i=1}^d\big(\sqrt{\alpha_i}-\sqrt{\beta_i}\big)^2.$$

The sum structure of this formula is to be expected because the Poisson distributions are tensor products of scalar Poisson distributions, such that Proposition 3.3 applies.

Surprisingly, as in the case of the Bhattacharya (a.k.a. spherical Hellinger distance), the Fisher-Rao distance can be expressed in terms of the Hellinger distance itself, namely

$$\mathsf{FR}_{\mathrm{Poiss}}(\pi_\alpha,\pi_\beta)^2 = -2\sigma^2 \log\big(1-\frac{1}{2\sigma^2}\mathsf{He}(\pi_\alpha,\pi_\beta)^2\big) \geq \mathsf{He}(\pi_\alpha,\pi_\beta)^2.$$

Note that $\mathsf{FR}_{\mathrm{Poiss}}(\pi_\alpha,\pi_\beta)$ can become arbitrarily large, while $\mathsf{He}(\pi_\alpha,\pi_\beta)\in[0,\sqrt{2}\,\sigma]$.

4.3 Submanifold of exponential distributions

We now consider the case $\Omega=[0,\infty[^n$, the parameter space $A=]0,\infty[^n\subset\mathbb{R}^n$, and the probability densities

$$\varepsilon_\alpha(x)=p(\alpha)\mathrm{e}^{-\alpha\cdot x} \quad \text{with } p(\alpha)=\prod_{i=1}^n \alpha_i.$$

The exponential submanifold is then given by

$$\mathcal{S}_{\exp}=\big\{\,\varepsilon_\alpha\,\big|\,\alpha\in A=]0,\infty[^n\subset\mathbb{R}^n\,\big\}\subset \mathrm{L}^1(\Omega)\cap\mathfrak{P}(\Omega).$$

Again the Hellinger distances are easily calculated in terms of the arithmetic mean $a(\alpha,\beta)=\frac{1}{2}(\alpha{+}\beta)\in A$ and the geometric mean $g(\alpha,\beta)=\big(\sqrt{\alpha_i\beta_I}\big)_i\in A$, namely

$$\mathsf{He}(\varepsilon_\alpha,\varepsilon_\beta)^2=2\sigma^2\Big(1-\frac{p\big(g(\alpha,\beta)\big)}{p\big(a(\alpha,\beta)\big)}\Big)=2\sigma^2\Big(1-\prod_{i=1}^n\frac{\sqrt{\alpha_i\beta_i}}{\frac{1}{2}(\alpha_i{+}\beta_i)}\Big).$$

The Fisher information matrix is easily obtained by using Fisher's logarithmic derivative, namely

$$\begin{aligned}\big\langle\mathbb{G}_{\exp}(\alpha)v,v\big\rangle &= \frac{\sigma^2}{4}\int_\Omega \big|v\cdot\nabla_\alpha\log\big(\varepsilon_\alpha(x)\big)\big|^2\varepsilon_\alpha(x)\,\mathrm{d}x\\ &=\frac{\sigma^2}{4}\int_\Omega\big|\sum_{i=1}^n v_i(\frac{1}{\alpha_i}{-}x_i)\big|^2\varepsilon_\alpha(x)\,\mathrm{d}x=\frac{\sigma^2}{4}\sum_{k=1}^n\big(\frac{v_k}{\alpha_k}\big)^2.\end{aligned}$$

Hence, the Fisher-Rao distance takes the form

$$\mathsf{FR}_{\exp}(\varepsilon_\alpha,\varepsilon_\beta)^2=\frac{\sigma^2}{4}\sum_{k=1}^n\big(\log\alpha_k-\log\beta_k\big)^2=\frac{\sigma^2}{4}\sum_{k=1}^n\big(\log(\alpha_k/\beta_k)\big)^2.$$

Again the sum structure follows because the multivariate exponential distribution is the tensor product of one-dimensional exponential distributions.

Moreover, we can use the scaling invariance of the exponential distributions under mappings $\Phi^D(x)=Dx$ with $D=\mathrm{diag}\big(\delta_i\big)_{i=1,..,n}$. Using that $\Phi^D_\#\varepsilon_\alpha=\varepsilon_{D\alpha}$ we find the invariance $\mathsf{FR}_{\exp}(\varepsilon_\alpha,\varepsilon_\beta)=\mathsf{FR}_{\exp}(\varepsilon_{D\alpha},\varepsilon_{D\beta})$. This implies that $\mathsf{FR}_{\exp}(\varepsilon_\alpha,\varepsilon_\beta)$ can only depend on α_k/β_k.

4.4 Gaussian distributions or multivariate normal distributions

The Fisher-Rao distance between Gaussian distributions was one of the motivation to introduce the the concept of Fisher matrix and Fisher-Rao distance and the one-dimensional case (univariate normal distributions) was already discussed in [Rao45]. For the general multivariate case, no explicit formula is known so far, and estimating the Fisher-Rao distance between Gaussians from above or below is an active field of research in the area of information geometry. We refer to [SPC15, PCS19, NiB19, PSC20, Nie23] for some recent works in this direction.

Remark 4.1 *The situation is different if the Hellinger distance is replaced by the Wasserstein distance, because in the Otto-Wasserstein geometry the geodesic curves between Gaussians remains in the class of Gaussians. This leads to the so-called Bures-Wasserstein distance, see e.g. [BJL19, KSS21, LC*22].*

We consider the case $\Omega = \mathbb{R}^d$ and use the standard representation of Gaussian measures $G_{m,\Sigma} \in \mathfrak{P}(\mathbb{R}^d)$ with density

$$p_{m,\Sigma}(x) = \frac{1}{\sqrt{\det(2\pi\Sigma)}} \exp\Big(-\frac{1}{2}(x-m)\cdot\Sigma^{-1}(x-m)\Big),$$

where $m \in \mathbb{R}^d$ is the mean and $\Sigma \in \mathbb{R}^{d\times d}_{\mathrm{spd}}$ is the symmetric and positive definite covariance matrix.

The Hellinger distance can be calculated easily as $\sqrt{G_{m_0,\Sigma_0}G_{m_1,\Sigma_1}}$ is a multiple of a Gaussian with covariance $2(\Sigma_0^{-1}+\Sigma_1^{-1})^{-1}$:

$$\mathsf{He}(G_{m_0,\Sigma_0}, G_{m_1,\Sigma_1})^2 = 2\sigma^2\Big(1 - \frac{\exp\big(-\frac{1}{4}(m_1-m_0)\cdot(\Sigma_0^{-1}+\Sigma_1^{-1})(m_1-m_0)\big)}{\big(\det\Sigma_0\,\det\Sigma_1\big)^{1/4}\big(\det(\frac{1}{2}\Sigma_0^{-1}+\frac{1}{2}\Sigma_1^{-1})\big)^{1/2}}\Big).$$

One can check that this formula is invariant under pushforwards (in the sense of Section 2.7) under all affine transformations $\Phi(x) = Ax + x_*$ by noting that $\Phi_\# G_{m,\Sigma} = G_{\overline{m},\overline{\Sigma}'}$ with $\overline{m} = Am + x_*$ and $\overline{\Sigma} = A\Sigma A^*$. For instance, the distance only depends on m_1-m_0. i.e. it is translation invariant.

Setting $\mathcal{P}_{\mathrm{Gauss}} := \big\{\, G_{m,\Sigma} \,\big|\, m \in \mathbb{R}^d,\ \Sigma \in \mathbb{R}^{d\times d}_{\mathrm{spd}} \,\big\}$ and $\mathsf{FR}_{\mathsf{Gauss}} = \mathsf{FR}_{\mathcal{P}_{\mathrm{Gauss}}}$, the invariance under pushforwards with respect to affine transformations is sill true for $\mathsf{FR}_{\mathsf{Gauss}}$, see also [PSC20] or "Property 1" in [Nie23, P. 4]. This is best seen by looking at the equation for the geodesic curves inside $\mathcal{P}_{\mathrm{Gauss}}$. For this one calculates the Fisher information matrix; using the parameters (m,Σ) it takes the form

$$\begin{pmatrix} v \\ V \end{pmatrix} \cdot \mathbb{G}_{\mathrm{Gauss}}(m,\Sigma) \begin{pmatrix} v \\ V \end{pmatrix} = \sigma^2\big(v\cdot\Sigma^{-1}v + \frac{1}{2}\,\mathrm{tr}(\Sigma^{-1}V\Sigma^{-1}V)\big). \tag{4.1}$$

From this the geodesic equations for $s \mapsto (m(s),\Sigma(s))$ can be derived by treating the inverse quadratic form

$$H(m,\Sigma,\xi,\Xi) := \begin{pmatrix} \xi \\ \Xi \end{pmatrix} \cdot \mathbb{K}_{\mathrm{Gauss}}(m,\Sigma) \begin{pmatrix} \xi \\ \Xi \end{pmatrix} = \frac{1}{4\sigma^2}\Big(\xi\cdot\Sigma\xi + 2\,\mathrm{tr}(\Sigma\Xi\Sigma\xi)\Big)$$

as a Hamiltonian H, see [LM*25, Sec. 4.1] for the details. Here $\xi(s) \in \mathbb{R}^d$ and $\Xi(s) \in \mathbb{R}^{d\times d}_{\mathrm{sym}}$ are the dual variables corresponding to $\xi_s = \xi(s,\cdot) \in \mathrm{L}^2(\Omega,\gamma_s)$ in Theorem 2.2. After scaling the parameter along the geodesic curves by the prefactor $\sigma^2/2$, one arrives at

$$m' = \tfrac{\sigma^2}{2}\mathrm{D}_\xi H(m,\Sigma,\xi,\Xi) = \Sigma\xi, \quad \Sigma' = \tfrac{\sigma^2}{2}\mathrm{D}_\Xi H(m,\Sigma,\xi,\Xi) = 2\Sigma\Xi\Sigma, \tag{4.2a}$$

$$\xi' = -\tfrac{\sigma^2}{2}\mathrm{D}_m H(m,\Sigma,\xi,\Xi) = 0, \quad \Xi' = -\tfrac{\sigma^2}{2}\mathrm{D}_\Sigma H(m,\Sigma,\xi,\Xi) = -2\Xi\Sigma\Xi - \frac{1}{2}\xi\otimes\xi. \tag{4.2b}$$

The translation invariance is seen in the fact that H does not depend on m, which implies that ξ is a constant along solutions by Noether's theorem. Similarly, for all $A \in \mathrm{GL}(\mathbb{R}^d)$ the mapping

$$(m, \Sigma, \xi, \Xi) \mapsto \big(Am, A\Sigma A^*, A^{-*}\xi, A^{-*}\Xi A^{-1}\big)$$

leaves the Hamiltonian invariant, and Noether's theory leads to the conserved quantities

$$J = \Sigma\Xi + \frac{1}{2}\, m\otimes\xi \in \mathbb{R}^{d\times d} \ \text{ (no symmetry).}$$

The $(1{+}d{+}d^2)$ conserved scalar quantities defined via $H \in \mathbb{R}^1$, $\mu \in \mathbb{R}^d$ and $J \in \mathbb{R}^{d\times d}$ are enough to show that the geodesic curves can be found, see [CaO91] or [PSC20, Eqn. (15)]. However, knowing the geodesics means solving an initial-value problem, while calculating the Fisher-Rao distance $\mathsf{FR}_{\mathrm{Gauss}}\big((m_0, \Sigma_0), (m_1, \Sigma_1)\big)$ means to solve a boundary-value problem.

The case $d = 1$, which was already treated in [Rao45], is by now classical and can be related to hyperbolic geometry by introduction $N = \sqrt{2\Sigma}$ and using the conservation laws

$$4\Sigma^2\Xi^2 + 2\Sigma\xi^2 = h_* = \mathrm{const}, \quad \xi = \xi_*, \ \text{ and } 2\Sigma\xi + m\xi = j_*$$

leads to the condition that (m, N) lies on the semi-circle

$$\big(m - j_*/\xi_*\big)^2 + N^2 = h_*/\xi_*^2.$$

Another easy case occurs for $\xi = \xi_* = 0 \in \mathbb{R}^d$, where now $d \in \mathbb{N}_*$ is general. This implies $m(s) = m_0 = m_1$ and corresponds to Gaussians with the same center. From $\Sigma' = 2J\Sigma$ and $\Xi' = -2\Xi J$ and the boundary conditions $\Sigma(i) = \Sigma_i$ for $i = 0, 1$, we obtain $J = -\frac{1}{2}\log\big(\Sigma_1\Sigma_0^{-1}\big) = -\frac{1}{2}\Sigma_0^{1/2}\log\big(\Sigma_0^{-1/2}\Sigma_1\Sigma_0^{-1/2}\big)\Sigma_0^{-1/2}$ and find

$$\Sigma(s) = \Sigma_0^{1/2}\big(\Sigma_0^{-1/2}\Sigma_1\Sigma_0^{-1/2}\big)^s\Sigma_0^{1/2}.$$

A third case can be handled by using the cross-product theory of Section 3.4 and joining the two cases from above, namely the case where $m_1{-}m_0$ is an eigenvector of Σ_1 and Σ_0. Using the invariance under affine transformations (rotations suffice) we can assume $m_1{-}m_0 = \delta e_1$. In that case, we can find a solution of the Hamiltonian system in the form

$$\begin{aligned} m(s) &= m_0 + \alpha(s)e_1, \qquad & \Sigma(s) &= \begin{pmatrix} \gamma(s) & 0 \\ 0 & \Gamma(s) \end{pmatrix} \in \mathbb{R}^{1\times 1}\otimes\mathbb{R}^{(d-1)\times(d-1)}. \\ \mu(s) &= \beta e_1, & \Xi(s) &= \begin{pmatrix} \zeta(s) & 0 \\ 0 & \Pi(s) \end{pmatrix} \in \mathbb{R}^{1\times 1}\otimes\mathbb{R}^{(d-1)\times(d-1)}. \end{aligned} \tag{4.3}$$

For (α, γ) the one-dimensional theory applies, while for (Γ, Π) the theory with the same center 0 can be used.

In these three cases the following results for $\mathsf{FR}_{\mathrm{Gauss}}(G_{m_0,\Sigma_0}, G_{m_1,\Sigma_1})$ are obtained, see [PSC20]. To present the results in a unified way we use the function $M : \mathbb{R}_\geq{\times}\mathbb{R}_> \to \mathbb{R}_\geq$ with

$$M(\Delta, \Lambda) := \sqrt{2}\,\log\left(\frac{1}{8}\Big(\sqrt{\Delta + 2(\Lambda^{1/4}{+}\Lambda^{-1/4})^2} + \sqrt{\Delta + 2(\Lambda^{1/4}{-}\Lambda^{-1/4})^2}\Big)^2\right)$$

which essentially arises from hyperbolic theory and satisfies $M(0, \Lambda) = 2^{-1/2}\,|\log\Lambda|$, $M(\Delta, 1) = \sqrt{\Delta} + O(\Delta)_{\Delta\to 0}$, and $M(\Delta, 1) = \sqrt{2}\,\log\Delta + O(1)_{\Delta\to\infty}$.

Theorem 4.2 (Fisher-Rao distance within Gaussians)

(A) In the one-dimensional case $d = 1$ we have the formula

$$d = 1: \quad \mathsf{FR}_{\mathrm{Gauss}}(G_{m_0,\Sigma_0}, G_{m_1,\Sigma_1}) = \frac{\sigma}{2} M\Big(\frac{(m_1{-}m_0)^2}{\sqrt{\Sigma_0\Sigma_1}}, \frac{\Sigma_1}{\Sigma_0}\Big). \tag{4.4}$$

(B) In the case with the same center $m_1 = m_0$ and $d \geq 1$ we have

$$d \geq 1: \quad \mathsf{FR}_{\mathrm{Gauss}}(G_{m_0,\Sigma_0}, G_{m_0,\Sigma_1})^2 = \frac{\sigma^2}{4} \sum_{n=1}^{d} M(0,\Lambda_n)^2 = \frac{\sigma^2}{2} \sum_{n=1}^{d} \big(\log \Lambda_n\big)^2, \tag{4.5}$$

where $\Lambda_n = \lambda_n(\Sigma_0^{-1/2}\Sigma_1\Sigma_0^{-1/2})$ is the n-th eigenvalue of the symmetric matrix $\Sigma_0^{-1/2}\Sigma_1\Sigma_0^{-1/2}$.

(C) Let $\Lambda_n > 0$ be as in (B) and assume $\Sigma_0^{-1}(m_1{-}m_0) = \Lambda_0\Sigma_1^{-1}(m_1{-}m_0)$, then,

$$\begin{aligned} \mathsf{FR}_{\mathrm{Gauss}}(G_{m_0,\Sigma_0}, G_{m_1,\Sigma_1})^2 = \frac{\sigma^2}{4} M\Big(|\Sigma_0^{-1/2}(m_1{-}m_0)|\,|\Sigma_1^{-1/2}(m_1{-}m_0)|, \Lambda_1\Big)^2 \\ + \frac{\sigma^2}{4} \sum_{n=2}^{d} M(0,\Lambda_n)^2. \end{aligned} \tag{4.6}$$

We refer to [PSC20, Sec. 2.1] and [Nie23, Sec. 1.2] and the references therein for the details of proofs and the corresponding the calculations. The sum structure for FR^2 on formulas (4.5) and (4.6) are again consequences of the cross-product theory in Proposition 3.3 because the Gaussians can be simultaneously transformed affinely to have the same eigenbasis.

However, we warn the reader that the general case cannot be handled by cross products in all the cases that the initial and final Gaussian have the same product structure. The reason is that the geodesic curves may lead the space of such product measures, see the discussion in [SPC15, PCS19, PSC20]. In the present context this can be seen by looking at the geodesic equations (4.2) involving the rank-one matrix $\xi{\otimes}\xi \in \mathbb{R}^{d\times d}_{\mathrm{sym}}$. A cross-product structure would mean that $\Sigma(s)$ and $\Xi(s)$ would have the same block structure for all $s \in [0,1]$. Then, this would also hold for $\xi{\otimes}\xi$, but together with the rank-one condition this implies that only one block can be non-trivial. This essentially explains the condition in case (C) of the above theorem.

References

[AGS05] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures.* Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2005.

[AJ*15] N. Ay, J. Jost, H. V. Lê, and L. Schwachhöfer. Information geometry and sufficient statistics. *Probab. Theor. Relat. Fields*, 162, 327–364, 2015.

[BBI01] D. Burago, Y. Burago, and S. Ivanov. *A course in metric geometry*, volume 33 of *Graduate Studies in Mathematics.* American Mathematical Society, Providence, RI, 2001.

[BBM16] M. Bauer, M. Bruveris, and P. W. Michor. Uniqueness of the Fisher-Rao metric on the space of smooth densities. *Bull. Lond. Math. Soc.*, 48(3), 499–506, 2016.

[BeB00] J.-D. Benamou and Y. Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numer. Math.*, 84(3), 375–393, 2000.

[Bha42] A. BHATTACHARYA. On discrimination and divergence. *Proc. Indian Sci. Congress*, Part III, 13, 1942.

[BJL19] R. BHATIA, T. JAIN, and Y. LIM. On the Bures-Wasserstein distance between positive definite matrices. *Expositiones Mathematicae*, 37(2), 165–191, 2019.

[CaO91] M. CALVO and J. M. OLLER. An explicit solution of information geodesic equations for the multivariate normal model. *Stat. Risk Model.*, 9(1-2), 119–138, 1991.

[CH*24] J. A. CARRILLO, D. Z. HUANG, J. HUANG, and D. WEI. Fisher-Rao gradient flow: geodesic convexity and functional inequalities. *Preprint*, arXiv:2407.15693, 2024.

[CP*18a] L. CHIZAT, G. PEYRÉ, B. SCHMITZER, and F.-X. VIALARD. An interpolating distance between optimal transport and Fisher–Rao metrics. *Found. Comput. Math.*, 18(1), 1–44, 2018. (arXiv 2015).

[CP*18b] L. CHIZAT, G. PEYRÉ, B. SCHMITZER, and F.-X. VIALARD. Unbalanced optimal transport: geometry and Kantorovich formulation. *J. Funct. Analysis*, 274(11), 3090–3123, 2018.

[Fis21] R. A. FISHER. On the mathematical foundations of theoretical statistics. *Phil. Trans. Royal Soc. A*, 222, 309–368, 1921.

[Hel07] E. HELLINGER. *Die Orthogonalinvarianten quadratischer Formen von unendlich vielen Variablen*. PhD thesis, Universität Göttingen, 1907.

[Hel09] E. HELLINGER. Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen (in German). *J. Reine Angew. Math.*, 136, 210–271, 1909.

[Kak48] S. KAKUTANI. On equivalence of infinite product measures. *Annals Math.*, 49(1), 214–224, 1948.

[KL*13] B. KHESIN, J. LENELLS, G. MISIOŁEK, and S. C. PRESTON. Geometry of diffeomorphism groups, complete integrability and geometric statistics. *Geom. Funct. Anal.*, 23, 334–366, 2013.

[Kol30] A. KOLMOGOROFF. Untersuchungen über den Integralbegriff. *Math. Annalen*, 103, 654–696, 1930.

[KSS21] A. KROSHNIN, V. SPOKOINY, and A. SUVORIKOVA. Statistical inference for Bures-Wasserstein barycenters. *Ann. Appl. Probab.*, 31(3), 1264–1298, 2021.

[LaM19] V. LASCHOS and A. MIELKE. Geometric properties of cones with applications on the Hellinger–Kantorovich space, and a new distance on the space of probability measures. *J. Funct. Analysis*, 276(11), 3529–3576, 2019.

[LC*22] M. LAMBERT, S. CHEWI, F. BACH, S. BONNABEL, and P. RIGOLLET. Variational inference via Wasserstein gradient flows. *arXiv*, 2205.15902, 2022.

[LeC70] L. LECAM. On the assumptions used to prove asymptotic normality of maximum likelihood estimates. *Ann. Math. Stat.*, 41(3), 802–828, 1970.

[LM*25] M. LIERO, A. MIELKE, O. TSE, and J.-J. ZHU. Evolution of Gaussians in the Hellinger-Kantorovich-Boltzmann gradient flow. *Comm. Pure Appl. Anal.*, 2025. To appear, arXiv:250420400, WIAS preprint 3198.

[LMS16] M. LIERO, A. MIELKE, and G. SAVARÉ. Optimal transport in competition with reaction – the Hellinger–Kantorovich distance and geodesic curves. *SIAM J. Math. Analysis*, 48(4), 2869–2911, 2016.

[LMS18] M. LIERO, A. MIELKE, and G. SAVARÉ. Optimal entropy-transport problems and a new Hellinger–Kantorovich distance between positive measures. *Invent. math.*, 211, 969–1117, 2018.

[May16] S. J. MAYBANK. A Fisher-Rao metric for curves using the information in edges. *J. Math. Imaging Vis.*, 54, 287–300, 2016.

[MiZ25] A. MIELKE and J.-J. ZHU. Hellinger-Kantorovich gradient flows: Global exponential decay of entropy functionals. *Preprint arXiv:2501.17049*, 2025. WIAS preprint 3167.

[NiB19] F. Nielsen and F. Barbaresco, editors. *Geometric Science of Information*, Lecture Notes in Computer Science. Springer Nature, 2019.

[Nie23] F. NIELSEN. A simple approximation method for the Fisher-Rao distance between multivariate normal distributions. *Entropy*, 25(654), 1–42, 2023.

[Ott01] F. OTTO. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Diff. Eqns.*, 26, 101–174, 2001.

[PCS19] J. PINELE, S. I. R. COSTA, and J. E. STRAPASSON. On the Fisher-Rao information metric in the space of normal distributions. In Nielsen and Barbaresco [NiB19], pages 676–684.

[PSC20] J. PINELE, J. E. STRAPASSON, and S. I. R. COSTA. The Fisher-Rao distance between multivariate normal distributions: special cases, bounds and applications. *Entropy*, 22(404), 1–24, 2020.

[Rao45] C. R. RAO. Information and the accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.*, 37, 81–91, 1945.

[RaV63] C. R. RAO and V. S. VARADARAJAN. Discrimination of Gaussian processes. *Sankhya: Indian J. Stat. A*, 25(3), 303–330, 1963.

[SPC15] J. E. STRAPASSON, J. P. S. PORTO, and S. I. R. COSTA. On bounds for the Fisher-Rao distance between multivariate normal distributions. In *AIP Conf. Proc. 1641*, pages 313–320. AIP Publishung LLC, 2015. Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MaxEnt 2014).