Geometry of Positive Operators, Infinite-dimensional Probability Measures, and Stochastic Processes

Hà Quang Minh

September 9, 2024

The set $\mathrm{Sym}^{++}(n)$ of $n \times n$ symmetric positive definite (SPD) matrices possesses rich geometrical structures, including the affine-invariant Riemannian metric, Wasserstein distance, and divergence functions such as Alpha and Alpha-Beta Log-Determinant (Log-Det) divergences. By the one-to-one correspondence between $\mathrm{Sym}^{++}(n)$ and the set of zero-mean Gaussian densities $\mathrm{Gauss}(\mathbb{R}^n)$ on $\mathbb{R}^n$, the distances/divergences on $\mathrm{Sym}^{++}(n)$ correspond to those on $\mathrm{Gauss}(\mathbb{R}^n)$. In particular, the affine-invariant Riemannian metric corresponds to the Fisher-Rao metric and the Alpha Log-Det divergences correspond to the Rényi divergences, including the Kullback-Leibler (KL) divergence. From a mathematical viewpoint, it is natural to ask whether and how the geometrical structures on $\mathrm{Sym}^{++}(n)$ and $\mathrm{Gauss}(\mathbb{R}^n)$ can be generalized to the setting of positive definite operators and zero-mean Gaussian measures $\mathrm{Gauss}(\mathcal{H})$ on an infinite-dimensional Hilbert space $\mathcal{H}$, e.g. Gaussian measures corresponding to Gaussian processes. For more general stochastic processes, one would obtain the geometry of their covariance operators. From a practical viewpoint, the study of these infinite-dimensional distances/divergences is motivated by various recent applications of covariance operators and Gaussian processes in machine learning, statistics, and computer vision, e.g. [36, 27, 12, 30, 25, 24, 35].

To motivate our discussion, consider the following concrete setting for Gaussian measures on Hilbert space, namely Gaussian processes with squared integrable paths. Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let T be a compact metric space, ν a non-negative, σ-finite measure on T, $\mathscr{B}(T)$ the Borel σ-algebra of T. Let $\xi = (\xi(t))_{t\in T} = (\xi(t,\omega))_{t\in T}$ be a real Gaussian process on $(\Omega, \mathcal{F}, P)$, i.e. a stochastic process so that for every finite set $\{x_j\}_{j=1}^N \subset T$, $N \in \mathbb{N}$, the random vector $(\xi(x_j))_{j=1}^N$ is distributed according to a Gaussian measure on $\mathbb{R}^N$. We assume that ξ is measurable, i.e. the map $(t,\omega) \to \xi(t,\omega)$ is measurable with respect to $\mathscr{B}(\mathbb{R})$ and $\mathscr{B}(T) \times \mathcal{F}$. Let m be the mean function and K be the covariance function of ξ, respectively, where $m(t) = \mathbb{E}\xi(t)$, $K(s,t) = \mathbb{E}[(\xi(s) - m(s))(\xi(t) - m(t))]$, $s,t \in T$. We then write $\xi \sim \mathrm{GP}(m, K)$. The sample paths $\xi(\cdot,\omega) \in \mathcal{H} = \mathcal{L}^2(T,\nu)$ almost P-surely, i.e. $\int_T \xi^2(t,\omega)d\nu(t) < \infty$ almost P-surely, if and only if [33] $\int_T m^2(t)d\nu(t) < \infty$ and $\int_T K(t,t)d\nu(t) < \infty$. In this case, ξ induces the following Gaussian measure P_ξ on $(\mathcal{H}, \mathscr{B}(\mathcal{H}))$: $P_\xi(B) = P\{\omega \in \Omega : \xi(\cdot,\omega) \in B\}$, $B \in \mathscr{B}(\mathcal{H})$, with mean $m \in \mathcal{H}$ and covariance operator $C_K : \mathcal{H} \to \mathcal{H}$ defined by $(C_K f)(s) = \int_T K(s,t) f(t) d\nu(t)$, $f \in \mathcal{H} = \mathcal{L}^2(T,\nu)$. Conversely, let μ be a Gaussian measure on $(\mathcal{H}, \mathscr{B}(\mathcal{H}))$, then there is a Gaussian process $\xi = (\xi(t))_{t\in T}$ with sample paths in $\mathcal{H}$, with induced probability measure $P_\xi = \mu$. Thus there is a one-to-one correspondence between measurable Gaussian processes with sample paths in $\mathcal{H} = \mathcal{L}^2(T,\nu)$ and Gaussian measures on $\mathcal{L}^2(T,\nu)$.

A practical example utilizing the geometry of Gaussian processes is [27], where the authors presented a test statistic for the equality of Gaussian process covariance operators and applied this in the study of DNA minicircles, for which comparison of the covariance structure reveals the differences between two groups of DNA minicircles, in contrast to the comparison of the mean functions, which reveals no differences. Other related work for stochastic processes in this direction includes [8, 28, 4].

The infinite-dimensional setting involves the interplay of many fields, including: (i) operator theory, in particular positive definite operators; (ii) infinite-dimensional Gaussian measures and Gaussian processes; (iii) infinite-dimensional Riemannian geometry; (iv) infinite-dimensional information geometry and optimal transport; (v) reproducing kernel Hilbert spaces (RKHS); (vi) probability theory on Banach spaces.

Throughout the following, let $(\mathcal{H}, \langle,\rangle)$ be a real, separable Hilbert space, with $\dim(\mathcal{H}) = \infty$. For two separable Hilbert spaces $(\mathcal{H}_i, \langle,\rangle_i)$,$i = 1,2$, let $\mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ denote the Banach space of bounded linear operators from $\mathcal{H}_1$ to $\mathcal{H}_2$, with operator norm $\|A\| = \sup_{||x||_1 \leq 1} \|Ax\|_2$. For $\mathcal{H}_1 = \mathcal{H}_2 = \mathcal{H}$, we write $\mathcal{L}(\mathcal{H})$. Let Sym be the set of self-adjoint linear operators on $\mathcal{H}$. Let $\mathrm{Sym}^+(\mathcal{H}) \subset \mathrm{Sym}(\mathcal{H})$, $\mathrm{Sym}^{++}(\mathcal{H}) \subset \mathrm{Sym}^+(\mathcal{H})$ be the sets of self-adjoint, *positive* and *strictly positive* operators on $\mathcal{H}$, respectively, i.e. $A \in \mathrm{Sym}^{+}(\mathcal{H}) \iff A^* = A, \langle Ax, x\rangle \geq 0 \forall x \in \mathcal{H}$, $A \in \mathrm{Sym}^{++}(\mathcal{H}) \iff A^* = A, \langle Ax, x\rangle > 0\ \forall x \in \mathcal{H}, x \neq 0$. We write $A \geq 0$ for $A \in \mathrm{Sym}^{+}(\mathcal{H})$ and $A > 0$ for $A \in \mathrm{Sym}^{++}(\mathcal{H})$. If $\gamma I + A > 0$, where I is the identity operator,$\gamma \in \mathbb{R}, \gamma > 0$, then $\gamma I + A$ is also invertible, in which case it is called *positive definite*. In general, $A \in \mathrm{Sym}(\mathcal{H})$ is said to be positive definite if $\exists M_A > 0$ such that $\langle x, Ax\rangle \geq M_A ||x||^2\ \forall x \in \mathcal{H}$ - this is equivalent to A being both strictly positive and invertible, see e.g. [29]. The Banach space $\mathrm{Tr}(\mathcal{H})$ of trace class operators on $\mathcal{H}$ is defined by (see e.g. [34]) $\mathrm{Tr}(\mathcal{H}) = \{A \in \mathcal{L}(\mathcal{H}) : ||A||_{\mathrm{tr}} = \sum_{k=1}^{\infty}\langle e_k, (A^*A)^{1/2}e_k\rangle < \infty\}$, for any orthonormal basis $\{e_k\}_{k\in\mathbb{N}} \subset \mathcal{H}$. For $A \in \mathrm{Tr}(\mathcal{H})$, its trace is defined by $\mathrm{tr}(A) = \sum_{k=1}^{\infty}\langle e_k, Ae_k\rangle$, which is independent of choice of $\{e_k\}_{k\in\mathbb{N}}$. The Hilbert space $\mathrm{HS}(\mathcal{H}_1, \mathcal{H}_2)$ of Hilbert-Schmidt operators from $\mathcal{H}_1$ to $\mathcal{H}_2$ is defined by (see e.g. [13]) $\mathrm{HS}(\mathcal{H}_1, \mathcal{H}_2) = \{A \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2) : ||A||^2_{\mathrm{HS}} = \mathrm{tr}(A^*A) = \sum_{k=1}^{\infty} ||Ae_k||_2^2 < \infty\}$, for any orthonormal basis $\{e_k\}_{k\in\mathbb{N}}$ in $\mathcal{H}_1$, with inner product $\langle A, B\rangle_{\mathrm{HS}} = \mathrm{tr}(A^*B)$. For $\mathcal{H}_1 = \mathcal{H}_2 = \mathcal{H}$, we write $\mathrm{HS}(\mathcal{H})$.

Geometry of positive operators. In general, most finite-dimensional geometrical structures on $\mathrm{Sym}^{++}(n)$ are *not* directly generalizable to the infinite-dimensional setting of positive definite operators on $\mathcal{H}$, since many functions, such as inverse, logarithm, trace, etc, are only well-defined on specific classes of operators. *One key and unifying framework to resolve these challenges is via regularization.* As examples, consider the affine-invariant Riemannian distance and Alpha Log-Det divergences.

- Under the affine-invariant metric, the Riemannian distance between $A, B \in \mathrm{Sym}^{++}(n)$ is given by $d_{\mathrm{ai}}(A, B) = \left\|\log(A^{-1/2}BA^{-1/2})\right\|_F$, with $\|\ \|_F$ being the Frobenius norm, $\|A\|_F^2 = \mathrm{tr}(A^TA)$. Consider the infinite-dimensional setting, where $A, B \in \mathrm{Sym}^{++}(\mathcal{H}) \cap \mathrm{HS}(\mathcal{H})$. Then A^{-1} and $\log(A)$ are both unbounded. More specifically, if $A \in \mathrm{Sym}^{++}(\mathcal{H})$ is compact, then it has a countable spectrum of eigenvalues $\{\lambda_k(A)\}_{k=1}^{\infty}$, counting multiplicities, with $\lambda_k(A) > 0\ \forall k \in \mathbb{N}$ and $\lim_{k\to\infty}\lambda_k(A) = 0$. If $\{\phi_k(A)\}_{k=1}^{\infty}$ denote the corresponding normalized eigenvectors, then A admits the spectral decomposition $A = \sum_{k=1}^{\infty}\lambda_k(A)\phi_k(A)\otimes\phi_k(A)$, where $\phi_k(A)\otimes\phi_k(A) : \mathcal{H} \to \mathcal{H}$ is defined by $(\phi_k(A)\otimes\phi_k(A))w = \langle w, \phi_k(A)\rangle\phi_k(A)$, $w \in \mathcal{H}$. The inverse and principal logarithm of A are then given by
$$A^{-1} = \sum_{k=1}^{\infty}\frac{1}{\lambda_k(A)}\phi_k(A)\otimes\phi_k(A), \quad \log(A) = \sum_{k=1}^{\infty}\log(\lambda_k(A))\phi_k(A)\otimes\phi_k(A). \tag{1}$$
Clearly, A^{-1} and $\log(A)$ are both unbounded. Thus the distance $d_{\mathrm{ai}}(A, B)$ is not directly generalizable. A proper generalization is obtained by considering the Hilbert space of *extended Hilbert-Schmidt operators* [14] $\mathrm{HS}_X(\mathcal{H}) = \{A + \gamma I : A \in \mathrm{HS}(\mathcal{H}), \gamma \in \mathbb{R}\}$ under the *extended Hilbert-Schmidt inner product* $\langle A+\gamma I, B+\mu I\rangle_{\mathrm{HS}_X} = \langle A, B\rangle_{\mathrm{HS}} + \gamma\mu$ and *norm* $\|A + \gamma I\|^2_{\mathrm{HS}_X} = \|A\|^2_{\mathrm{HS}} + \gamma^2$. The infinite-dimensional generalization of $\mathrm{Sym}^{++}(n)$ is then $\mathscr{PC}_2(\mathcal{H}) = \{A+\gamma I > 0 : A \in \mathrm{HS}(\mathcal{H}), \gamma \in \mathbb{R}\} \subset \mathrm{HS}_X(\mathcal{H})$, which is an open subset in $\mathrm{HS}_X(\mathcal{H})$ and hence is a Hilbert manifold, on which the affine-invariant Riemannian metric can be properly generalized. The corresponding Riemannian distance is
$$d_{\mathrm{ai}}[(A+\gamma I), (B+\mu I)] = \left\|\log\left[(A+\gamma I)^{-1/2}(B+\mu I)(A+\gamma I)^{-1/2}\right]\right\|_{\mathrm{HS}_X} \tag{2}$$
for $(A + \gamma I), (B + \mu I) \in \mathscr{PC}_2(\mathcal{H})$, which is always finite.

- The Alpha Log-Det divergence with $\alpha = 1$ is given by $d^1_{\mathrm{logdet}}(A, B) = \mathrm{tr}(B^{-1}A - I) - \log\det(B^{-1}A)$, which does not generalize directly to the infinite-dimensional setting. A proper generalization is obtained in [18] by considering the Banach space of *extended trace class operators* $\mathrm{Tr}_X(\mathcal{H}) = \{A + \gamma I : A \in \mathrm{Tr}(\mathcal{H}), \gamma \in \mathbb{R}\}$ under the *extended trace norm* $\|A + \gamma\|_{\mathrm{tr}_X} = ||A||_{\mathrm{tr}} + |\gamma|$, along with the *extended trace* $\mathrm{tr}_X(A + \gamma I) = \mathrm{tr}(A) + \gamma$ and *extended Fredholm determinant* $\det_{\mathrm{X}}(A + \gamma I) = \gamma\det[(A/\gamma) + I]$, $\gamma \neq 0$, where det is the Fredholm determinant. On the open subset $\mathscr{PC}_1(\mathcal{H}) = \{A + \gamma I > 0 : A \in \mathrm{Tr}(\mathcal{H}), \gamma \in \mathbb{R}\} \subset \mathrm{Tr}_X(\mathcal{H})$, d^1_{logdet} admits the following generalization, which is a valid divergence,
$$d^1_{\mathrm{logdet}}[(A+\gamma I), (B+\mu I)] = \left(\frac{\gamma}{\mu} - 1\right)\log\frac{\gamma}{\mu} + \mathrm{tr}_X[(B+\mu I)^{-1}(A+\gamma I) - I] - \frac{\gamma}{\mu}\log\det{}_{\mathrm{X}}[(B+\mu I)^{-1}(A+\gamma I)]. \tag{3}$$

In general, for $-1 \leq \alpha \leq 1$, $d^{\alpha}_{\mathrm{logdet}}[(A+\gamma I),(B+\mu I)]$ is defined on $\mathscr{PC}_1(\mathcal{H})$ in a similar way [18]. This formulation was generalized to the Alpha-Beta Log-Det divergences on $\mathscr{PC}_1(\mathcal{H}) \subsetneq \mathscr{PC}_2(\mathcal{H})$ and to the entire Hilbert manifold $\mathscr{PC}_2(\mathcal{H})$ in [19]. These divergences encompass both the Alpha Log-Det divergences and the squared affine-invariant Riemannian distance as special cases. Furthermore, they all induce, via their Hessian operators, the affine-invariant Riemannian metric.

Geometry of Gaussian measures on Hilbert space. The geometry of positive definite Hilbert-Schmidt operators was obtained from a purely geometrical viewpoint. By the one-to-one correspondence between the subset $\mathrm{Sym}^+(\mathcal{H}) \cap \mathrm{Tr}(\mathcal{H})$ with the set $\mathrm{Gauss}(\mathcal{H})$ via their covariance operators, it is clear that there is a correspondence with the geometry of $\mathrm{Gauss}(\mathcal{H})$. As with $\mathrm{Sym}^{++}(n)$, most finite-dimensional geometrical structures on $\mathrm{Gauss}(\mathbb{R}^n)$ *defined via probability density functions with respect to the Lebesgue measure* do not directly generalize to $\mathrm{Gauss}(\mathcal{H})$. This is because, in contrast to $\mathbb{R}^n$, there is no Lebesgue measure on $\mathcal{H}$. Furthermore, in contrast to $\mathrm{Gauss}(\mathbb{R}^n)$, by the Feldman-Hajek Theorem [7, 11], any two measures $\mu,\nu \in \mathrm{Gauss}(\mathcal{H})$ are either equivalent ($\mu \sim \nu$) or mutually singular ($\mu \perp \nu$). Consider the setting of *equivalent Gaussian measures*, where the Fisher-Rao metric and KL divergence can be computed explicitly. Let $\mu_0 = \mathcal{N}(0, C_0)$ be fixed. The set of all Gaussian measures equivalent to μ_0 is $\mathrm{Gauss}(\mathcal{H}, \mu_0) = \{\mu = \mathcal{N}(0, C) : C = C_0^{1/2}(I-S)C_0^{1/2},\ S \in \mathrm{SymHS}(\mathcal{H})_{<I}\}$, parameterized by the Hilbert manifold $\mathrm{SymHS}(\mathcal{H})_{<I} = \{S \in \mathrm{Sym}(\mathcal{H}) \cap \mathrm{HS}(\mathcal{H}), I - S > 0\}$. On this manifold, the Fisher-Rao metric and the corresponding Riemannian manifold structure are well-defined and can be computed explicitly, with the Riemannian distance between $\mu = \mathcal{N}(0,A) \sim \nu = \mathcal{N}(0,B)$, $B = A^{1/2}(I-S)A^{1/2}$, given by [17] $d_{\mathrm{FR}}(\mu,\nu) = \frac{1}{\sqrt{2}}\left\|\log(A^{-1/2}BA^{-1/2})\right\|_{\mathrm{HS}} = \frac{1}{\sqrt{2}}\left\|\log(I-S)\right\|_{\mathrm{HS}}$. It is closely connected with the affine-invariant Riemannian distance on $\mathscr{PC}_2(\mathcal{H})$ by

$$\lim_{\gamma \to 0^+} \left\|\log\left[(A+\gamma I)^{-1/2}(B+\gamma I)(A+\gamma I)^{-1/2}\right]\right\|_{\mathrm{HS}_X} = \sqrt{2} d_{\mathrm{FR}}(\mu,\nu), \quad \text{if } \mu \sim \nu. \tag{4}$$

The left hand side of (4) thus can be considered as a *regularized Fisher-Rao distance*. Computationally, it has two advantages over the exact Fisher-Rao distance: (i) it is valid for *any pair* of Gaussian measures μ,ν; (ii) it is much more straightforward to obtain its finite-dimensional approximations. Similarly, for the KL divergence, if $\mu \perp \nu$ then $\mathrm{KL}(\nu||\mu) = \infty$, otherwise if $\mu \sim \nu$ then [20]

$$\lim_{\gamma \to 0^+} d^1_{\mathrm{logdet}}[(B+\gamma I),(A+\gamma I)] = 2\mathrm{KL}(\nu||\mu) = -\log\det\nolimits_2(I-S), \tag{5}$$

where $\det_2$ denotes the Hilbert-Carleman determinant. The left hand side of (5) can thus be considered as a *regularized KL divergence*, with the same computational advantages as above. For work on infinite-dimensional geometry in more general and abstract settings, we refer to [32, 31, 10, 5, 26, 1, 2, 3].

In contrast to the Fisher-Rao metric and KL divergence, *Optimal Transport distances* are defined via joint probability measures, not density functions. As an example, consider the 2-Wasserstein distance [6, 9] and its entropic regularization, the Sinkhorn divergence on $\mathrm{Gauss}(\mathcal{H})$ [23]. For $\mu_0 = \mathcal{N}(m_0, C_0)$, $\mu_1 = \mathcal{N}(m_1, C_1)$,

$$W_2^2(\mu_0,\mu_1) = \|m_0 - m_1\|^2 + \mathrm{tr}(C_0) + \mathrm{tr}(C_1) - 2\mathrm{tr}\left(C_1^{1/2}C_0C_1^{1/2}\right)^{1/2}. \tag{6}$$

$$\mathrm{S}_2^{\epsilon}(\mu_0,\mu_1) = ||m_0 - m_1||^2 + \frac{\epsilon}{4}\mathrm{tr}\left[M_{00}^{\epsilon} - 2M_{01}^{\epsilon} + M_{11}^{\epsilon}\right] + \frac{\epsilon}{4}\log\left[\frac{\det\left(I + \frac{1}{2}M_{01}^{\epsilon}\right)^2}{\det\left(I+\frac{1}{2}M_{00}^{\epsilon}\right)\det\left(I+\frac{1}{2}M_{11}^{\epsilon}\right)}\right], \epsilon > 0. \tag{7}$$

where $M_{ij}^{\epsilon} = -I + \left(I + \frac{16}{\epsilon^2}C_i^{1/2}C_jC_i^{1/2}\right)^{1/2} \in \mathrm{Tr}(\mathcal{H})$. Both (6) and (7) have the same forms as on $\mathrm{Gauss}(\mathbb{R}^n)$. In particular, they are valid when C_0, C_1 are both singular, even when $C_0 = C_1 = 0$. **Note**: The Sinkhorn divergence S_2^{ϵ} in (7) is valid in the more general setting $C_0, C_1 \in \mathrm{Sym}^+(\mathcal{H}) \cap \mathrm{HS}(\mathcal{H})$. *Entropic regularization*, while not necessary for the formulation of infinite-dimensional optimal transport, can lead to many favorable theoretical properties [23, 22, 21], as in the Gaussian process setting below.

Geometry of stochastic processes. Consider now two measurable Gaussian processes $\xi^i = (\xi^i(t,\omega))_{t \in T}$, $i = 1,2$, $\xi^i \sim \mathrm{GP}(m^i, K^i)$, satisfying $\int_T (m^i(t))^2 d\nu(t) < \infty$, $\int_T K^i(t,t)d\nu(t) < \infty$. Then the sample paths $\xi^i(\cdot,\omega) \in \mathcal{H} = \mathcal{L}^2(T,\nu)$ P-almost surely and ξ^i corresponds to the Gaussian measure $\mu^i = \mathcal{N}(m^i, C_{K^i})$, where $C_{K^i} : \mathcal{L}^2(T,\nu) \to \mathcal{L}^2(T,\nu)$ is defined by $C_{K^i}f(x) = \int_T K^i(x,t)f(t)d\nu(t)$. As an example, consider the optimal transport distances with $m^i = 0$ for simplicity. Let $\mathbf{X} = (x_j)_{j=1}^N$ be sampled independently from (T,ν) and define the Gram matrix $K^i[\mathbf{X}]$ by $(K^i[\mathbf{X}])_{j,k} = K^i(x_j, x_k)$, $j,k = 1,\ldots,N$. Then $W_2[\mathcal{N}(0, C_{K^1}), \mathcal{N}(0, C_{K^2})]$ and $S_2^{\epsilon}[\mathcal{N}(0, C_{K^1}), \mathcal{N}(0, C_{K^2})]$ both have closed form expressions and can be consistently estimated from $W_2[\mathcal{N}(0, \frac{1}{N}K^1[\mathbf{X}]), \mathcal{N}(0, \frac{1}{N}K^2[\mathbf{X}])]$ and $S_2^{\epsilon}[\mathcal{N}(0, \frac{1}{N}K^1[\mathbf{X}]), \mathcal{N}(0, \frac{1}{N}K^2[\mathbf{X}])]$, respectively [21]. The convergence is obtained via their representations using operators between the RKHS $\mathcal{H}_{K^1}, \mathcal{H}_{K^2}$ induced by K^1, K^2, respectively. In particular, a sequence of Gaussian measures converges in S_2^{ϵ} if the corresponding covariance operators converges in the $|| \; ||_{\mathrm{HS}}$ norm [22] and one can apply the law of large numbers for Hilbert space-valued random variables to obtain dimension-independent sample complexity for the estimation of S_2^{ϵ} (this is *not* the case with W_2, where the convergence happens in the Banach $|| \; ||_{\mathrm{tr}}$ norm). For related work on optimal transport distances for covariance operators and Gaussian processes, see also [16, 15].

References

[1] N. Ay, J. Jost, H.V. Lê, and L. Schwachhöfer. Information geometry and sufficient statistics. *Probability Theory and Related Fields*, 162:327–364, 2015.

[2] N. Ay, J. Jost, H.V. Lê, and L. Schwachhöfer. *Information geometry*, volume 64. Springer, 2017.

[3] N. Ay, J. Jost, H.V. Lê, and L. Schwachhöfer. Parametrized measure models. *Bernoulli*, 24(3):1692–1725, 2018.

[4] G. Boente, D. Rodriguez, and M. Sued. Testing equality between several populations covariance operators. *Annals of the Institute of Statistical Mathematics*, 70(4):919–950, 2018.

[5] A. Cena and G. Pistone. Exponential statistical manifold. *Annals of the Institute of Statistical Mathematics*, 59:27–56, 2007.

[6] J. Cuesta-Albertos, C. Matrán-Bea, and A. Tuero-Diaz. On lower bounds for the L2-Wasserstein metric in a Hilbert space. *Journal of Theoretical Probability*, 9(2):263–283, 1996.

[7] J. Feldman. Equivalence and perpendicularity of Gaussian processes. *Pacific Journal of Mathematics*, 8(4):699–708, 1958.

[8] S. Fremdt, J. Steinebach, L. Horváth, and P. Kokoszka. Testing the equality of covariance operators in functional samples. *Scandinavian Journal of Statistics*, 40(1):138–152, 2013.

[9] M. Gelbrich. On a formula for the L2 Wasserstein metric between measures on Euclidean and Hilbert spaces. *Mathematische Nachrichten*, 147(1):185–203, 1990.

[10] P. Gibilisco and G. Pistone. Connections on non-parametric statistical manifolds by Orlicz space geometry. *Infinite Dimensional Analysis, Quantum Probability and Related Topics*, 1(02):325–347, 1998.

[11] J. Hájek. On a property of normal distributions of any stochastic process. *Czechoslovak Mathematical Journal*, 08(4):610–618, 1958.

[12] M. Harandi, M. Salzmann, and F. Porikli. Bregman divergences for infinite dimensional covariance matrices. In *IEEE CVPR*, 2014.

[13] R.V. Kadison and J.R. Ringrose. *Fundamentals of the theory of operator algebras. Volume I: Elementary Theory.* Academic Press, 1983.

[14] G. Larotonda. Nonpositive curvature: A geometrical approach to Hilbert-Schmidt operators. *Differential Geometry and its Applications*, 25:679–700, 2007.

[15] V. Masarotto and G. Masarotto. Covariance-based soft clustering of functional data based on the Wasserstein–Procrustes metric. *Scandinavian Journal of Statistics*, 51(2):485–512, 2024.

[16] V. Masarotto, V. Panaretos, and Y. Zemel. Procrustes metrics on covariance operators and optimal transportation of Gaussian processes. *Sankhya A*, 81(1):172–213, 2019.

[17] H. Q. Minh. Fisher-Rao geometry of equivalent Gaussian measures on infinite-dimensional Hilbert spaces. *Information Geometry*, 2024.

[18] H.Q. Minh. Infinite-dimensional Log-Determinant divergences between positive definite trace class operators. *Linear Algebra and Its Applications*, 528:331–383, 2017.

[19] H.Q. Minh. Infinite-dimensional Log-Determinant divergences between positive definite Hilbert-Schmidt operators. *Positivity*, 24:631–662, 2020.

[20] H.Q. Minh. Regularized divergences between covariance operators and Gaussian measures on Hilbert spaces. *Journal of Theoretical Probability*, 34:580–643, 2021.

[21] H.Q. Minh. Finite sample approximations of exact and entropic Wasserstein distances between covariance operators and Gaussian processes. *SIAM/ASA Journal on Uncertainty Quantification*, 10:96–124, 2022.

[22] H.Q. Minh. Convergence and finite sample approximations of entropic regularized Wasserstein distances in Gaussian and RKHS settings. *Analysis and Applications*, 21(3):719–775, 2023.

[23] H.Q. Minh. Entropic regularization of Wasserstein distance between infinite-dimensional Gaussian measures and Gaussian processes. *Journal of Theoretical Probability*, 36:201–296, 2023.

[24] H.Q. Minh and V. Murino. Covariances in computer vision and machine learning. *Synthesis Lectures on Computer Vision*, 7(4):1–170, 2017.

[25] H.Q. Minh, M. San Biagio, and V. Murino. Log-Hilbert-Schmidt metric between positive definite operators on Hilbert spaces. In *Advances in Neural Information Processing Systems 27 (NIPS 2014)*, pages 388–396, 2014.

[26] N.J. Newton. An infinite-dimensional statistical manifold modelled on Hilbert space. *Journal of Functional Analysis*, 263(6):1661–1681, 2012.

[27] V. Panaretos, D. Kraus, and J. Maddocks. Second-order comparison of Gaussian random functions and the geometry of DNA minicircles. *Journal of the American Statistical Association*, 105(490):670–682, 2010.

[28] E. Paparoditis and T. Sapatinas. Bootstrap-based testing of equality of mean functions or equality of covariance operators for functional data. *Biometrika*, 103(3):727–733, 2016.

[29] W.V. Petryshyn. Direct and iterative methods for the solution of linear operator equations in Hilbert spaces. *Transactions of the American Mathematical Society*, 105:136–175, 1962.

[30] D. Pigoli, J. Aston, I.L. Dryden, and P. Secchi. Distances and inference for covariance operators. *Biometrika*, 101(2):409–422, 2014.

[31] G. Pistone and M.P. Rogantin. The exponential statistical manifold: mean parameters, orthogonality and space transformations. *Bernoulli*, 5(4):721–760, 1999.

[32] G. Pistone and C. Sempi. An infinite-dimensional geometric structure on the space of all the probability measures equivalent to a given one. *The Annals of Statistics*, 23(5):1543–1561, 1995.

[33] B. Rajput and S. Cambanis. Gaussian processes and Gaussian measures. *The Annals of Mathematical Statistics*, 43(6):1944–1952, 1972.

[34] M. Reed and B. Simon. *Methods of Modern Mathematical Physics: Functional analysis*. Academic Press, 1975.

[35] S. Sun, G. Zhang, J. Shi, and R. Grosse. Functional variational Bayesian neural networks. *International Conference on Learning Representation*, 2019.

[36] S. K. Zhou and R. Chellappa. From sample similarity to ensemble similarity: Probabilistic distance measures in reproducing kernel Hilbert space. *TPAMI*, 28(6):917–929, 2006.