

---

# Log-Hilbert-Schmidt metric between positive definite operators on Hilbert spaces

## Supplementary Material

---

**Hà Quang Minh   Marco San Biagio   Vittorio Murino**  
Istituto Italiano di Tecnologia  
Via Morego 30, Genova 16163, ITALY  
{minh.haquang, marco.sanbiagio, vittorio.murino}@iit.it

### Abstract

The supplementary material contains four parts. First, we provide some sample images from the three datasets that we tested our method on. Second, we provide some background material on Hilbert-Schmidt operators. The third and main part contains the proofs for all mathematical results in the paper. The final part contains further discussions and interpretations of our framework, including its computational complexity in the kernel setting.

## A Experiments

In this section we give some example images extracted from each of the three datasets that we tested our method on. Figure 1 shows some samples extracted from the Kylberg Texture dataset. Figure 2 shows 4 samples, 1 for each split, of the KTH-TIPS2b Material dataset. Finally, Figure 3 shows examples extracted from the Fish Recognition dataset.

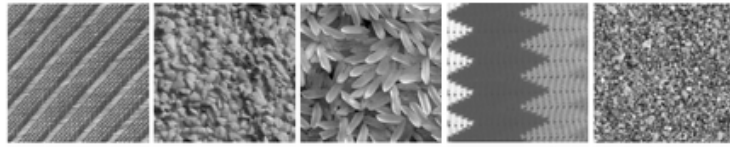


Figure 1: Some samples from the Kylberg texture dataset [6].



Figure 2: Samples extracted from the 4 splits of the KTH-TIPS2b dataset [3].

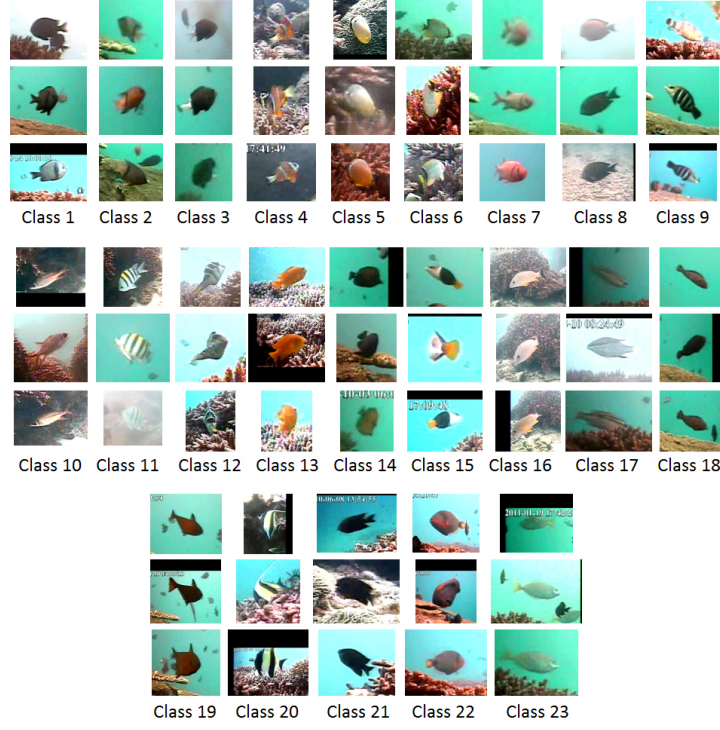


Figure 3: 3 samples from each of the 23 classes of the Fish Recognition dataset [2].

## B Background on Hilbert-Schmidt operators

We briefly give some facts about Hilbert-Schmidt operators that we need in the current work here, for more detail see e.g. [5]. Let  $\mathcal{H}$  be a separable Hilbert space. Let  $\mathcal{L}(\mathcal{H})$  be the Banach space of bounded linear operators on  $\mathcal{H}$ . The set of Hilbert-Schmidt operators, denoted by  $\text{HS}(\mathcal{H})$ , is defined by

$$\text{HS}(\mathcal{H}) = \{A \in \mathcal{L}(\mathcal{H}) : \|A\|_{\text{HS}}^2 = \text{tr}(A^*A) = \sum_{k=1}^{\infty} \|Ae_k\|^2 < \infty\},$$

where  $\{e_k\}_{k=1}^{\infty}$  is any orthonormal basis for  $\mathcal{H}$ . Under the Hilbert-Schmidt inner product,

$$\langle A, B \rangle_{\text{HS}} = \text{tr}(A^*B),$$

the set  $\text{HS}(\mathcal{H})$  becomes a Hilbert space. It is a *non-unital Banach algebra* of operators, since it is also closed under multiplication, with  $A, B \in \text{HS}(\mathcal{H}) \Rightarrow AB \in \text{HS}(\mathcal{H})$  and

$$\|AB\|_{\text{HS}} \leq \|A\|_{\text{HS}} \|B\|_{\text{HS}}.$$

This is called the Hilbert-Schmidt algebra of operators on  $\mathcal{H}$ . Furthermore, it is a two-sided ideal in  $\mathcal{L}(\mathcal{H})$ , that is

$$A \in \text{HS}(\mathcal{H}), B \in \mathcal{L}(\mathcal{H}) \Rightarrow BA \in \text{HS}(\mathcal{H}), AB \in \text{HS}(\mathcal{H}).$$

A Hilbert-Schmidt operator is compact and thus has a countable spectrum. In particular, if  $A$  is self-adjoint Hilbert-Schmidt, that is  $A \in \text{Sym}(\mathcal{H}) \cap \text{HS}(\mathcal{H})$ , then  $A$  has countably many real eigenvalues  $\{\lambda_k\}_{k=1}^{\infty}$  and we have

$$\|A\|_{\text{HS}}^2 = \sum_{k=1}^{\infty} \lambda_k^2 < \infty.$$

Clearly, the identity operator  $I \notin \text{HS}(\mathcal{H})$ , since

$$\|I\|_{\text{HS}}^2 = \sum_{k=1}^{\infty} \|e_k\|^2 = \infty.$$

## C Proofs of main results

For clarity, we restate all the results from the main paper that we prove here.

### C.1 Proofs for the Log-Hilbert-Schmidt metric - the general setting

**Theorem 1.** *Under the two operations  $\odot$  and  $\otimes$ ,  $(\Sigma(\mathcal{H}), \odot, \otimes)$  becomes a vector space, with  $\odot$  acting as vector addition and  $\otimes$  acting as scalar multiplication. The zero element in  $(\Sigma(\mathcal{H}), \odot, \otimes)$  is the identity operator  $I$  and the inverse of  $(A + \gamma I)$  is  $(A + \gamma I)^{-1}$ . Furthermore, the map*

$$\psi : (\Sigma(\mathcal{H}), \odot, \otimes) \rightarrow (\mathcal{H}_{\mathbb{R}}, +, \cdot) \text{ defined by } \psi(A + \gamma I) = \log(A + \gamma I), \quad (\text{C.1})$$

*is a vector space isomorphism, so that for all  $(A + \gamma I), (B + \mu I) \in \Sigma(\mathcal{H})$  and  $\lambda \in \mathbb{R}$ ,*

$$\begin{aligned} \psi((A + \gamma I) \odot (B + \mu I)) &= \log(A + \gamma I) + \log(B + \mu I), \\ \psi(\lambda \otimes (A + \gamma I)) &= \lambda \log(A + \gamma I), \end{aligned} \quad (\text{C.2})$$

*where  $+$  and  $\cdot$  denote the usual operator addition and multiplication operations, respectively.*

To prove Theorem 1, we need the following result.

**Lemma 1.** *Assume that  $\dim(\mathcal{H}) = \infty$ . For each operator  $(A + \gamma I) \in \Sigma(\mathcal{H})$ , there exist a unique operator  $A_1 \in \text{HS}(\mathcal{H}) \cap \text{Sym}(\mathcal{H})$  and a unique scalar  $\gamma_1 \in \mathbb{R}$  such that*

$$A + \gamma I = \exp(A_1 + \gamma_1 I). \quad (\text{C.3})$$

**Proof of Lemma 1.** It was shown in [7] that for a fixed  $P \in \Sigma(\mathcal{H})$ , the exponential map  $\text{Exp}_P : T_P(\Sigma(\mathcal{H})) \rightarrow \Sigma(\mathcal{H})$  and its inverse, the logarithm map  $\text{Log}_P : \Sigma(\mathcal{H}) \rightarrow T_P(\Sigma(\mathcal{H}))$ , are diffeomorphisms given by

$$\text{Exp}_P(U) = P^{1/2} \exp(P^{-1/2} U P^{-1/2}) P^{1/2}, \quad U \in T_P(\Sigma(\mathcal{H})), \quad (\text{C.4})$$

$$\text{Log}_P(V) = P^{1/2} \log(P^{-1/2} V P^{-1/2}) P^{1/2}, \quad V \in \Sigma(\mathcal{H}). \quad (\text{C.5})$$

The tangent space  $T_P(\Sigma(\mathcal{H}))$  at any point  $P$  can be identified with  $\mathcal{H}_{\mathbb{R}}$ , that is  $T_P(\Sigma(\mathcal{H})) \simeq \mathcal{H}_{\mathbb{R}}$ . Thus in particular, for  $P = I$ , the map

$$\text{Exp}_I(U) = \exp(U), \quad U \in \mathcal{H}_{\mathbb{R}}, \quad (\text{C.6})$$

and its inverse

$$\text{Log}_I(V) = \log(V), \quad V \in \Sigma(\mathcal{H}), \quad (\text{C.7})$$

are diffeomorphisms. Hence for each operator  $(A + \gamma I) \in \Sigma(\mathcal{H})$ , there exist a unique operator  $A_1 \in \text{HS}(\mathcal{H}) \cap \text{Sym}(\mathcal{H})$  and a unique scalar  $\gamma_1 \in \mathbb{R}$  such that

$$A + \gamma I = \exp(A_1 + \gamma_1 I).$$

This completes the proof.  $\square$

**Proof of Theorem 1.** The case  $\dim(\mathcal{H}) < \infty$  was treated by [1]. Let us assume that  $\dim(\mathcal{H}) = \infty$ .

(I) First, we need to show that  $\Sigma(\mathcal{H})$  is closed under the operations  $\odot$  and  $\otimes$ , that is

$$(A + \gamma I) \odot (B + \mu I) \in \Sigma(\mathcal{H})$$

and

$$\lambda \otimes (A + \gamma I) \in \Sigma(\mathcal{H})$$

for all operators  $(A + \gamma I), (B + \mu I) \in \Sigma(\mathcal{H})$  and all  $\lambda \in \mathbb{R}$ .

By Lemma 1, there exist unique  $A_1, B_1 \in \text{HS}(\mathcal{H}) \cap \text{Sym}(\mathcal{H})$  and  $\gamma_1, \mu_1 \in \mathbb{R}$  such that

$$A + \gamma I = \exp(A_1 + \gamma_1 I), \quad B + \mu I = \exp(B_1 + \mu_1 I).$$

It follows that

$$(A + \gamma I) \odot (B + \mu I) = \exp((A_1 + B_1) + (\gamma_1 + \mu_1)I) \in \Sigma(\mathcal{H}).$$

Similarly,

$$\lambda \otimes (A + \gamma I) = \exp(\lambda(A_1 + \gamma_1 I)) = \exp(\lambda A_1 + \lambda \gamma_1 I) \in \Sigma(\mathcal{H}).$$

Thus  $\Sigma(\mathcal{H})$  are closed under the two operations  $\odot$  and  $\otimes$ .

(II) Let us show that  $(\Sigma(\mathcal{H}), \odot)$  is an abelian group by verifying all the axioms.

i) Associativity is satisfied, since

$$\begin{aligned} [(A + \gamma I) \odot (B + \mu I)] \odot (C + \eta I) &= (A + \gamma I) \odot [(B + \mu I) \odot (C + \eta I)] \\ &= \exp(\log(A + \gamma I) + \log(B + \mu I) + \log(C + \eta I)). \end{aligned}$$

ii) The neutral element is the identity operator  $I$ , since it is clear that

$$(A + \gamma I) \odot I = I \odot (A + \gamma I) = (A + \gamma I).$$

iii) The inverse of  $(A + \gamma I)$  is  $(A + \gamma I)^{-1}$ , since

$$(A + \gamma I) \odot (A + \gamma I)^{-1} = \exp(0) = I = (A + \gamma I)^{-1} \odot (A + \gamma I).$$

iv) Commutativity is satisfied, since

$$(A + \gamma I) \odot (B + \mu I) = (B + \mu I) \odot (A + \gamma I) = \exp(\log(A + \gamma I) + \log(B + \mu I)).$$

Thus  $\Sigma(\mathcal{H})$  is an abelian group.

(III) Let us now verify the axioms showing that  $(\Sigma(\mathcal{H}), \odot, \otimes)$  is a vector space, with the operation  $\odot$  acting as vector addition and  $\otimes$  acting as scalar multiplication.

(i) First distributive property:

$$\begin{aligned} \lambda \otimes [(A + \gamma I) \odot (B + \mu I)] &= \exp(\lambda \log[(A + \gamma I) \odot (B + \mu I)]) \\ &= \exp(\lambda [\log(A + \gamma I) + \log(B + \mu I)]) = \exp(\log[(A + \gamma I)^\lambda + \log[(B + \mu I)^\lambda]]) \\ &= \exp(\log[\lambda \otimes (A + \gamma I)] + \log[\lambda \otimes (B + \mu I)]) = [\lambda \otimes (A + \gamma I)] \odot [\lambda \otimes (B + \mu I)]. \end{aligned}$$

(ii) Second distributive property:

$$\begin{aligned} (\lambda + \mu) \otimes (A + \gamma I) &= \exp((\lambda + \mu) \log(A + \gamma I)) = \exp(\lambda \log(A + \gamma I) + \mu \log(A + \gamma I)) \\ &= \exp(\log[(A + \gamma I)^\lambda + \log[(A + \gamma I)^\mu]]) = \exp(\log[\lambda \otimes (A + \gamma I)] + \log[\mu \otimes (A + \gamma I)]) \\ &= [\lambda \otimes (A + \gamma I)] \odot [\mu \otimes (A + \gamma I)]. \end{aligned}$$

(iii) Associativity of scalar multiplication:

$$\begin{aligned} \lambda \otimes [\mu \otimes (A + \gamma I)] &= \lambda \otimes \exp(\mu \log(A + \gamma I)) = \lambda \otimes \exp(\log[(A + \gamma I)^\mu]) \\ &= \lambda \otimes [(A + \gamma I)^\mu] = \exp(\lambda \log[(A + \gamma I)^\mu]) = \exp(\lambda \mu \log(A + \gamma I)) = (\lambda \mu) \otimes (A + \gamma I). \end{aligned}$$

(iv) Multiplication by the unit scalar:

$$1 \otimes (A + \gamma I) = \exp(\log(A + \gamma I)) = (A + \gamma I).$$

These axioms, together with the axioms showing that  $(\Sigma(\mathcal{H}), \odot)$  is abelian, show that  $(\Sigma(\mathcal{H}), \odot, \otimes)$  is a vector space.

(IV) Consider now the map

$$\psi : (\Sigma(\mathcal{H}), \odot, \otimes) \rightarrow (\mathcal{H}_{\mathbb{R}}, +, \cdot),$$

defined by

$$\psi(A + \gamma I) = \log(A + \gamma I).$$

We have

$$\psi([(A + \gamma I) \odot (B + \mu I)]) = \log(A + \gamma I) + \log(B + \mu I),$$

$$\psi(\lambda \otimes (A + \gamma I)) = \lambda \log(A + \gamma I).$$

This shows that  $\psi$  is a homomorphism. We already know that the map  $\log : \Sigma(\mathcal{H}) \rightarrow \mathcal{H}_{\mathbb{R}}$  is a bijection. Thus  $\psi$  is a vector space isomorphism. This completes the proof.  $\square$

Consider the *Log-Hilbert-Schmidt distance* between two operators  $(A + \gamma I) \in \Sigma(\mathcal{H})$ ,  $(B + \mu I) \in \Sigma(\mathcal{H})$ , defined by

$$d_{\log\text{HS}}[(A + \gamma I), (B + \mu I)] = \|\log[(A + \gamma I) \odot (B + \mu I)^{-1}]\|_{\text{eHS}}. \quad (\text{C.8})$$

**Theorem 2.** *The Log-Hilbert-Schmidt distance as defined in (C.8) is a metric, making  $(\Sigma(\mathcal{H}), d_{\log\text{HS}})$  a metric space. Let  $(A + \gamma I) \in \Sigma(\mathcal{H})$ ,  $(B + \mu I) \in \Sigma(\mathcal{H})$ . If  $\dim(\mathcal{H}) = \infty$ , then there exist unique operators  $A_1, B_1 \in \text{HS}(\mathcal{H}) \cap \text{Sym}(\mathcal{H})$  and scalars  $\gamma_1, \mu_1 \in \mathbb{R}$  such that*

$$A + \gamma I = \exp(A_1 + \gamma_1 I), \quad B + \mu I = \exp(B_1 + \mu_1 I), \quad (\text{C.9})$$

and

$$d_{\log\text{HS}}^2[(A + \gamma I), (B + \mu I)] = \|A_1 - B_1\|_{\text{HS}}^2 + (\gamma_1 - \mu_1)^2. \quad (\text{C.10})$$

If  $\dim(\mathcal{H}) < \infty$ , then (C.9) and (C.10) hold with  $A_1 = \log(A + \gamma I)$ ,  $B_1 = \log(B + \mu I)$ ,  $\gamma_1 = \mu_1 = 0$ .

**Proof of Theorem 2.** If  $\dim(\mathcal{H}) < \infty$ , then by definition

$$\begin{aligned} d_{\log\text{HS}}[(A + \gamma I), (B + \mu I)] &= \|\log[(A + \gamma I) \odot (B + \mu I)^{-1}]\|_{\text{HS}} \\ &= \|\log(A + \gamma I) - \log(B + \mu I)\|_{\text{HS}}, \end{aligned}$$

which is simply the Log-Euclidean metric between  $(A + \gamma I)$  and  $(B + \mu I)$ .

Consider now the case  $\dim(\mathcal{H}) = \infty$ . By Lemma 1, for operators  $(A + \gamma I) \in \Sigma(\mathcal{H})$ ,  $(B + \mu I) \in \Sigma(\mathcal{H})$ , there exist unique operators  $A_1 \in \text{HS}(\mathcal{H}) \cap \text{Sym}(\mathcal{H})$  and unique scalars  $\gamma_1, \mu_1 \in \mathbb{R}$  such that

$$A + \gamma I = \exp(A_1 + \gamma_1 I), \quad B + \mu I = \exp(B_1 + \mu_1 I). \quad (\text{C.11})$$

It follows that

$$\begin{aligned} (A + \gamma I) \odot (B + \mu I)^{-1} &= \exp(\log(A + \gamma I) - \log(B + \mu I)) = \exp((A_1 + \gamma_1 I) - (B_1 + \mu_1 I)) \\ &= \exp((A_1 - B_1) + (\gamma_1 - \mu_1)I). \end{aligned}$$

Consequently,

$$\log[(A + \gamma I) \odot (B + \mu I)^{-1}] = (A_1 - B_1) + (\gamma_1 - \mu_1)I \in \mathcal{H}_{\mathbb{R}}.$$

By definition of the extended Hilbert-Schmidt norm, we have

$$d_{\log\text{HS}}^2[(A + \gamma I), (B + \mu I)] = \|\log[(A + \gamma I) \odot (B + \mu I)^{-1}]\|_{\text{eHS}}^2 = \|A_1 - B_1\|_{\text{HS}}^2 + (\gamma_1 - \mu_1)^2.$$

Let us show that  $d_{\log\text{HS}}$  is indeed a metric by verifying all the axioms of metric space.

i) Positivity: clearly we have

$$d_{\log\text{HS}}[(A + \gamma I), (B + \mu I)] = \sqrt{\|A_1 - B_1\|_{\text{HS}}^2 + (\gamma_1 - \mu_1)^2} \geq 0$$

for all  $(A + \gamma I), (B + \mu I) \in \Sigma(\mathcal{H})$ . Equality happens if and only if  $A_1 = B_1$  and  $\gamma_1 = \mu_1$ , that is if and only if  $A = B$  and  $\gamma = \mu$ .

ii) Symmetry: this is also clear from the above expression.

iii) Triangle inequality: by definition,

$$\begin{aligned} d_{\log\text{HS}}[(A + \gamma I), (B + \mu I)] &= \|\log[(A + \gamma I) \odot (B + \mu I)^{-1}]\|_{\text{eHS}} \\ &= \|\log(A + \gamma I) - \log(B + \mu I)\|_{\text{eHS}}. \end{aligned}$$

The triangle inequality for  $d_{\log\text{HS}}$  then follows from the triangle inequality for the extended Hilbert-Schmidt norm  $\|\cdot\|_{\text{eHS}}$ . Thus  $(\Sigma(\mathcal{H}), d_{\log\text{HS}})$  is a metric space. This completes the proof of the theorem.  $\square$

Consider the *Log-Hilbert-Schmidt inner product* between  $(A + \gamma I)$  and  $(B + \mu I)$ , defined by

$$\langle A + \gamma I, B + \mu I \rangle_{\log\text{HS}} = \langle \log(A + \gamma I), \log(B + \mu I) \rangle_{\text{eHS}} = \langle A_1, B_1 \rangle_{\text{HS}} + \gamma_1 \mu_1, \quad (\text{C.12})$$

where  $A_1, B_1 \in \text{Sym}(\mathcal{H}) \cap \text{HS}(\mathcal{H})$  and  $\gamma_1, \mu_1 \in \mathbb{R}$  are such that

$$(A + \gamma I) = \exp(A_1 + \gamma_1 I), \quad (B + \mu I) = \exp(B_1 + \mu_1 I),$$

as in Theorem 2.

**Theorem 3.** The inner product  $\langle \cdot, \cdot \rangle_{\log\text{HS}}$  as given in (C.12) is well-defined on  $(\Sigma(\mathcal{H}), \odot, \otimes)$ . Endowed with this inner product,  $(\Sigma(\mathcal{H}), \odot, \otimes, \langle \cdot, \cdot \rangle_{\log\text{HS}})$  becomes a Hilbert space. The corresponding Log-Hilbert-Schmidt norm is given by

$$\|A + \gamma I\|_{\log\text{HS}}^2 = \|\log(A + \gamma I)\|_{\text{eHS}}^2 = \|A_1\|_{\text{HS}}^2 + \gamma_1^2. \quad (\text{C.13})$$

In terms of this norm, the Log-Hilbert-Schmidt distance is given by

$$d_{\log\text{HS}}[(A + \gamma I), (B + \mu I)] = \|(A + \gamma I) \odot (B + \mu I)^{-1}\|_{\log\text{HS}}. \quad (\text{C.14})$$

**Proof of Theorem 3.** We first need to show that the inner product

$$\langle A + \gamma I, B + \mu I \rangle_{\log\text{HS}} = \langle A_1, B_1 \rangle_{\text{HS}} + \gamma_1 \mu_1.$$

is well-defined on  $(\Sigma(\mathcal{H}), \odot, \otimes)$ , by verifying all the necessary axioms.

i) Symmetry is obvious.

ii) First linear property:

$$\begin{aligned} \langle (A + \gamma I) \odot (B + \mu I), (C + \eta I) \rangle_{\log\text{HS}} &= \langle \exp(\log(A + \gamma I) + \log(B + \mu I)), C + \eta I \rangle_{\log\text{HS}} \\ &= \langle \exp((A_1 + B_1) + (\gamma_1 + \mu_1)I), \exp(C_1 + \eta_1 I) \rangle_{\log\text{HS}} \\ &= \langle (A_1 + B_1), C_1 \rangle_{\text{HS}} + (\gamma_1 + \mu_1)\eta_1 = (\langle A_1, C_1 \rangle_{\text{HS}} + \gamma_1 \eta_1) + (\langle B_1, C_1 \rangle_{\text{HS}} + \mu_1 \eta_1) \\ &= \langle (A + \gamma I), (C + \eta I) \rangle_{\log\text{HS}} + \langle (B + \mu I), (C + \eta I) \rangle_{\log\text{HS}}. \end{aligned}$$

iii) Second linear property:

$$\begin{aligned} \langle [\lambda \otimes (A + \gamma I)], (B + \mu I) \rangle_{\log\text{HS}} &= \langle \exp(\lambda \log(A + \gamma I)), (B + \mu I) \rangle_{\log\text{HS}} \\ &= \langle \exp(\lambda(A_1 + \gamma_1 I)), \exp(B_1 + \mu_1 I) \rangle_{\log\text{HS}} = \langle \lambda A_1, B_1 \rangle_{\text{HS}} + \lambda \gamma_1 \mu_1 \\ &= \lambda[\langle A_1, B_1 \rangle_{\text{HS}} + \gamma_1 \mu_1] = \lambda \langle (A + \gamma I), (B + \mu I) \rangle_{\log\text{HS}}. \end{aligned}$$

iv) Positivity:

$$\langle A + \gamma I, A + \gamma I \rangle_{\log\text{HS}} = \|A_1\|_{\text{HS}}^2 + \gamma_1^2 \geq 0.$$

Equality happens if and only if  $A_1 = 0$  and  $\gamma_1 = 0$ , that is if and only if  $A + \gamma I = \exp(A_1 + \gamma_1 I) = I$ . Thus  $\langle \cdot, \cdot \rangle_{\log\text{HS}}$  is a well-defined inner product, giving rise to the norm

$$\|A + \gamma I\|_{\log\text{HS}}^2 = \|\log(A + \gamma I)\|_{\text{eHS}}^2 = \|A_1\|_{\text{HS}}^2 + \gamma_1^2.$$

By definition, we have

$$d_{\log\text{HS}}[(A + \gamma I), (B + \mu I)] = \|\log[(A + \gamma I) \odot (B + \mu I)^{-1}]\|_{\text{eHS}} = \|(A + \gamma I) \odot (B + \mu I)^{-1}\|_{\log\text{HS}},$$

as we claimed.

To show that  $(\Sigma(\mathcal{H}), \odot, \otimes, \langle \cdot, \cdot \rangle_{\log\text{HS}})$  is a Hilbert space, we need to show that it is complete under the norm  $\|\cdot\|_{\log\text{HS}}$ . This is obvious if  $\dim(\mathcal{H}) < \infty$ . If  $\dim(\mathcal{H}) = \infty$ , let  $\{(A^n + \gamma^n I)\}_{n \in \mathbb{N}}$  be a Cauchy sequence in  $\|\cdot\|_{\log\text{HS}}$ , that is

$$\lim_{n, m \rightarrow \infty} \|(A^n + \gamma^n I) \odot (A^m + \gamma^m I)^{-1}\|_{\log\text{HS}}^2 = \lim_{n, m \rightarrow \infty} \|A_1^n - A_1^m\|_{\text{HS}}^2 + (\gamma_1^n - \gamma_1^m)^2 = 0,$$

where  $A_1^n \in \text{Sym}(\mathcal{H}) \cap \text{HS}(\mathcal{H})$  and  $\gamma_1^n \in \mathbb{R}$  are such that  $(A^n + \gamma^n I) = \exp(A_1^n + \gamma_1^n I)$ . The convergence on the right hand side above implies that there is a unique  $A_1 \in \text{Sym}(\mathcal{H}) \cap \text{HS}(\mathcal{H})$  and a unique  $\gamma_1 \in \mathbb{R}$  such that

$$\lim_{n \rightarrow \infty} \|A_1^n - A_1\|_{\text{HS}}^2 = 0, \quad \lim_{n \rightarrow \infty} \gamma_1^n = \gamma_1.$$

Hence if  $(A + \gamma I) = \exp(A_1 + \gamma_1 I)$ , then

$$\lim_{n \rightarrow \infty} \|(A^n + \gamma^n I) \odot (A + \gamma I)^{-1}\|_{\log\text{HS}}^2 = \lim_{n \rightarrow \infty} \|A_1^n - A_1\|_{\text{HS}}^2 + (\gamma_1^n - \gamma_1)^2 = 0.$$

Thus the Cauchy sequence  $\{(A^n + \gamma^n I)\}_{n \in \mathbb{N}}$  converges under the  $\|\cdot\|_{\log\text{HS}}$  norm, showing that  $(\Sigma(\mathcal{H}), \odot, \otimes, \langle \cdot, \cdot \rangle_{\log\text{HS}})$  is a complete inner product space, that is it is a Hilbert space. This completes the proof.  $\square$

**Corollary 1.** *The following kernels defined on  $\Sigma(\mathcal{H}) \times \Sigma(\mathcal{H})$  are positive definite:*

$$K[(A + \gamma I), (B + \mu I)] = (c + \langle A + \gamma I, B + \mu I \rangle_{\log \text{HS}})^d, \quad c > 0, \quad d \in \mathbb{N}, \quad (\text{C.15})$$

$$K[(A + \gamma I), (B + \mu I)] = \exp(-d_{\log \text{HS}}^p[(A + \gamma I), (B + \mu I)]/\sigma^2), \quad 0 < p \leq 2. \quad (\text{C.16})$$

**Proof of Corollary 1.** For the first kernel, we have the property that the sum and product of positive definite kernels are also positive definite. Thus from the positivity of the inner product  $\langle A + \gamma I, B + \mu I \rangle_{\log \text{HS}}$ , it follows that  $K[(A + \gamma I), (B + \mu I)] = (c + \langle A + \gamma I, B + \mu I \rangle_{\log \text{HS}})^d$  is positive definite, as in the Euclidean setting.

For the second kernel, we have by definition,

$$d_{\log \text{HS}}[(A + \gamma I), (B + \mu I)] = \|\log(A + \gamma I) - \log(B + \mu I)\|_{\text{eHS}}.$$

It follows that

$$\exp(-d_{\log \text{HS}}^p[(A + \gamma I), (B + \mu I)]/\sigma^2) = \exp(-\|\log(A + \gamma I) - \log(B + \mu I)\|_{\text{eHS}}^p/\sigma^2),$$

which is positive definite for  $0 < p \leq 2$  by a classical result due to Schoenberg [9].  $\square$

## C.2 Log-Hilbert-Schmidt metric between regularized positive operators

**Theorem 4.** *Assume that  $\dim(\mathcal{H}) = \infty$ . Let  $A, B \in \text{HS}(\mathcal{H}) \cap \text{Sym}^+(\mathcal{H})$ . Let  $\gamma, \mu > 0$ . Then  $(A + \gamma I), (B + \mu I) \in \Sigma(\mathcal{H})$  and*

$$d_{\log \text{HS}}^2[(A + \gamma I), (B + \mu I)] = \left\| \log\left(\frac{1}{\gamma}A + I\right) - \log\left(\frac{1}{\mu}B + I\right) \right\|_{\text{HS}}^2 + (\log \gamma - \log \mu)^2. \quad (\text{C.17})$$

In particular, for  $\mu = \gamma > 0$ ,

$$d_{\log \text{HS}}[(A + \gamma I), (B + \gamma I)] = \left\| \log\left(\frac{1}{\gamma}A + I\right) - \log\left(\frac{1}{\gamma}B + I\right) \right\|_{\text{HS}}. \quad (\text{C.18})$$

Their Log-Hilbert-Schmidt inner product is given by

$$\langle (A + \gamma I), (B + \mu I) \rangle_{\log \text{HS}} = \left\langle \log\left(\frac{1}{\gamma}A + I\right), \log\left(\frac{1}{\mu}B + I\right) \right\rangle_{\text{HS}} + (\log \gamma)(\log \mu). \quad (\text{C.19})$$

The corresponding norm is given by

$$\|(A + \gamma I)\|_{\log \text{HS}}^2 = \left\| \log\left(\frac{1}{\gamma}A + I\right) \right\|_{\text{HS}}^2 + (\log \gamma)^2. \quad (\text{C.20})$$

To prove Theorem 4, we first note that for each compact operator  $A \in \text{Sym}^+(\mathcal{H})$ , the following operator is well-defined:

$$\log(I + \gamma A) = \sum_{k=1}^{\dim(\mathcal{H})} \log(1 + \gamma \lambda_k(A)) \phi_k(A) \otimes \phi_k(A), \quad \gamma \geq 0. \quad (\text{C.21})$$

Furthermore, we have the following result.

**Lemma 2.** *If  $A \in \text{HS}(\mathcal{H}) \cap \text{Sym}^+(\mathcal{H})$ , then  $\log(I + \gamma A) \in \text{HS}(\mathcal{H}) \cap \text{Sym}(\mathcal{H})$  for all  $\gamma \geq 0$ , with the Hilbert-Schmidt norm of  $\log(I + \gamma A)$  given by*

$$\|\log(I + \gamma A)\|_{\text{HS}}^2 = \sum_{k=1}^{\dim(\mathcal{H})} [\log(1 + \gamma \lambda_k(A))]^2 \leq \gamma^2 \|A\|_{\text{HS}}^2. \quad (\text{C.22})$$

**Proof of Lemma 2.** Since  $A$  is compact, self-adjoint, and positive, the operator  $\log(I + \gamma A)$  is well-defined for all  $\gamma \geq 0$ , as noted above. It is also clear that  $\log(I + \gamma A) \in \text{Sym}(\mathcal{H})$ . To show that  $\log(I + \gamma A) \in \text{HS}(\mathcal{H})$ , we make use of the identity

$$\log(1 + x) \leq x \quad \text{for all } x \geq 0. \quad (\text{C.23})$$

It follows that

$$\|\log(I + \gamma A)\|_{\text{HS}}^2 = \sum_{k=1}^{\dim(\mathcal{H})} [\log(1 + \gamma \lambda_k(A))]^2 \leq \sum_{k=1}^{\dim(\mathcal{H})} \gamma^2 \lambda_k^2(A) = \gamma^2 \|A\|_{\text{HS}}^2 < \infty,$$

by the assumption that  $A \in \text{HS}(\mathcal{H})$ . This completes the proof.  $\square$

**Proof of Theorem 4.** Since the identity operator  $I$  commutes with all other operators, from the definition of the log function, we have

$$\log(A + \gamma I) = \log \left[ \gamma I \left( \frac{1}{\gamma} A + I \right) \right] = \log(\gamma I) + \log \left( \frac{1}{\gamma} A + I \right), \quad (\text{C.24})$$

where  $\log \left( \frac{1}{\gamma} A + I \right) \in \text{HS}(\mathcal{H}) \cap \text{Sym}(\mathcal{H})$  by Lemma 2. It follows that in Theorem 2,

$$(A + \gamma I) = \exp(A_1 + \gamma_1 I),$$

where

$$A_1 = \log \left( \frac{1}{\gamma} A + I \right), \quad \gamma_1 = \log \gamma.$$

Similarly,

$$(B + \mu I) = \exp(B_1 + \mu_1 I),$$

where

$$B_1 = \log \left( \frac{1}{\mu} B + I \right), \quad \mu_1 = \log \mu.$$

Thus from Theorem 2, we have

$$\begin{aligned} d_{\log \text{HS}}^2[(A + \gamma I), (B + \mu I)] &= \|A_1 - B_1\|_{\text{HS}}^2 + (\gamma_1 - \mu_1)^2 \\ &= \left\| \log \left( \frac{1}{\gamma} A + I \right) - \log \left( \frac{1}{\mu} B + I \right) \right\|_{\text{HS}}^2 + (\log(\gamma) - \log(\mu))^2, \end{aligned}$$

which is the desired formula for the distance  $d_{\log \text{HS}}$ . By definition of the Log-Hilbert-Schmidt inner product, we have

$$\begin{aligned} \langle (A + \gamma I), (B + \mu I) \rangle_{\log \text{HS}} &= \langle A_1, B_1 \rangle_{\text{HS}} + \gamma_1 \mu_1 \\ &= \left\langle \log \left( \frac{1}{\gamma} A + I \right), \log \left( \frac{1}{\mu} B + I \right) \right\rangle_{\text{HS}} + (\log \gamma)(\log \mu), \end{aligned}$$

as we claimed. The formula for the Log-Hilbert-Schmidt norm then follows immediately. This completes the proof.  $\square$

**Theorem 5.** Assume that  $\dim(\mathcal{H}) < \infty$ . Let  $A, B \in \text{Sym}^+(\mathcal{H})$ . Let  $\gamma, \mu > 0$ . Then  $(A + \gamma I), (B + \mu I) \in \Sigma(\mathcal{H})$  and

$$\begin{aligned} d_{\log \text{HS}}^2[(A + \gamma I), (B + \mu I)] &= \left\| \log \left( \frac{A}{\gamma} + I \right) - \log \left( \frac{B}{\mu} + I \right) \right\|_{\text{HS}}^2 \\ &+ 2(\log \gamma - \log \mu) \text{tr} \left( \log \left( \frac{A}{\gamma} + I \right) - \log \left( \frac{B}{\mu} + I \right) \right) + (\log \gamma - \log \mu)^2 \dim(\mathcal{H}). \quad (\text{C.25}) \end{aligned}$$

The Log-Hilbert-Schmidt inner product between  $(A + \gamma I)$  and  $(B + \mu I)$  is given by

$$\begin{aligned} \langle (A + \gamma I), (B + \mu I) \rangle_{\log \text{HS}} &= \left\langle \log \left( \frac{A}{\gamma} + I \right), \log \left( \frac{B}{\mu} + I \right) \right\rangle_{\text{HS}} \\ &+ (\log \gamma) \text{tr} \left( \log \left( \frac{B}{\mu} + I \right) \right) + (\log \mu) \text{tr} \left( \log \left( \frac{A}{\gamma} + I \right) \right) + (\log \gamma \log \mu) \dim(\mathcal{H}). \quad (\text{C.26}) \end{aligned}$$

The Log-Hilbert-Schmidt norm of  $(A + \gamma I)$  is given by

$$\|(A + \gamma I)\|_{\log \text{HS}}^2 = \left\| \log \left( \frac{1}{\gamma} A + I \right) \right\|_{\text{HS}}^2 + 2(\log \gamma) \text{tr} \log \left( \frac{1}{\gamma} A + I \right) + (\log \gamma)^2 \dim(\mathcal{H}). \quad (\text{C.27})$$



**Proof of Theorem 5.** If  $\dim(\mathcal{H}) < \infty$ , then in Theorem 2, we have

$$(A + \gamma I) = \exp(A_1 + \gamma_1 I), \quad (B + \mu I) = \exp(B_1 + \mu_1 I)$$

where

$$A_1 = \log(A + \gamma I), \quad B_1 = \log(B + \mu I), \quad \gamma_1 = \mu_1 = 0.$$

The Log-Hilbert-Schmidt distance is now the Log-Euclidean distance and is given by

$$\begin{aligned} d_{\log\text{HS}}^2[(A + \gamma I), (B + \mu I)] &= \|A_1 - B_1\|_{\text{HS}}^2 = \|\log(A + \gamma I) - \log(B + \mu I)\|_{\text{HS}}^2 \\ &= \left\| \log\left(\frac{A}{\gamma} + I\right) - \log\left(\frac{B}{\mu} + I\right) + (\log \gamma - \log \mu)I \right\|_{\text{HS}}^2 \\ &= \left\| \log\left(\frac{A}{\gamma} + I\right) - \log\left(\frac{B}{\mu} + I\right) \right\|_{\text{HS}}^2 + 2(\log \gamma - \log \mu) \text{tr} \left( \log\left(\frac{A}{\gamma} + I\right) - \log\left(\frac{B}{\mu} + I\right) \right) \\ &\quad + (\log \gamma - \log \mu)^2 \dim(\mathcal{H}). \end{aligned}$$

Similarly, the Log-Hilbert-Schmidt inner product is

$$\begin{aligned} \langle (A + \gamma I), (B + \mu I) \rangle_{\log\text{HS}} &= \langle A_1, B_1 \rangle_{\text{HS}} = \langle \log(A + \gamma I), \log(B + \mu I) \rangle_{\text{HS}} \\ &= \left\langle \log\left(\frac{A}{\gamma} + I\right) + (\log \gamma)I, \log\left(\frac{B}{\mu} + I\right) + (\log \mu)I \right\rangle_{\text{HS}} \\ &= \left\langle \log\left(\frac{A}{\gamma} + I\right), \log\left(\frac{B}{\mu} + I\right) \right\rangle_{\text{HS}} + (\log \gamma) \text{tr} \left( \log\left(\frac{B}{\mu} + I\right) \right) + (\log \mu) \text{tr} \left( \log\left(\frac{A}{\gamma} + I\right) \right) \\ &\quad + (\log \gamma \log \mu) \dim(\mathcal{H}). \end{aligned}$$

Finally, for the Log-Hilbert-Schmidt norm,

$$\begin{aligned} \|(A + \gamma I)\|_{\log\text{HS}}^2 &= \|\log(A + \gamma I)\|_{\text{HS}}^2 = \left\| \log\left(\frac{1}{\gamma}A + I\right) + (\log \gamma)I \right\|_{\text{HS}}^2 \\ &= \left\| \log\left(\frac{1}{\gamma}A + I\right) \right\|_{\text{HS}}^2 + 2(\log \gamma) \text{tr} \log\left(\frac{1}{\gamma}A + I\right) + (\log \gamma)^2 \dim(\mathcal{H}). \end{aligned}$$

This completes the proof.  $\square$

### C.3 Log-Hilbert-Schmidt metric between regularized covariance operators

Let  $\mathcal{X}$  be an arbitrary non-empty set. Let  $K$  be a positive definite kernel on  $\mathcal{X} \times \mathcal{X}$  and  $\mathcal{H}_K$  be its corresponding RKHS. Let  $\mathbf{x} = [x_i]_{i=1}^m$ ,  $\mathbf{y} = [y_i]_{i=1}^m$ ,  $m \in \mathbb{N}$ , be two data matrices sampled from  $\mathcal{X}$  and  $C_{\Phi(\mathbf{x})}$ ,  $C_{\Phi(\mathbf{y})}$  be the corresponding covariance operators induced by the kernel  $K$ . Let  $K[\mathbf{x}]$ ,  $K[\mathbf{y}]$ , and  $K[\mathbf{x}, \mathbf{y}]$  be the  $m \times m$  Gram matrices defined by  $(K[\mathbf{x}])_{ij} = K(x_i, x_j)$ ,  $(K[\mathbf{y}])_{ij} = K(y_i, y_j)$ ,  $(K[\mathbf{x}, \mathbf{y}])_{ij} = K(x_i, y_j)$ ,  $1 \leq i, j \leq m$ . Let  $A = \frac{1}{\sqrt{\gamma m}} \Phi(\mathbf{x}) J_m : \mathbb{R}^m \rightarrow \mathcal{H}_K$ ,  $B = \frac{1}{\sqrt{\mu m}} \Phi(\mathbf{y}) J_m : \mathbb{R}^m \rightarrow \mathcal{H}_K$ , so that

$$A^T A = \frac{1}{\gamma m} J_m K[\mathbf{x}] J_m, \quad B^T B = \frac{1}{\mu m} J_m K[\mathbf{y}] J_m, \quad A^T B = \frac{1}{\sqrt{\gamma \mu m}} J_m K[\mathbf{x}, \mathbf{y}] J_m. \quad (\text{C.28})$$

Let  $N_A$  and  $N_B$  be the numbers of nonzero eigenvalues of  $A^T A$  and  $B^T B$ , respectively. Let  $\Sigma_A$  and  $\Sigma_B$  be the diagonal matrices of size  $N_A \times N_A$  and  $N_B \times N_B$ , respectively, and  $U_A$  and  $U_B$  be the matrices of size  $m \times N_A$  and  $m \times N_B$ , respectively, which are obtained from the spectral decompositions

$$\frac{1}{\gamma m} J_m K[\mathbf{x}] J_m = U_A \Sigma_A U_A^T, \quad \frac{1}{\mu m} J_m K[\mathbf{y}] J_m = U_B \Sigma_B U_B^T. \quad (\text{C.29})$$

In the following, let  $\circ$  denote the Hadamard (element-wise) matrix product. Define

$$C_{AB} = \mathbf{1}_{N_A}^T \log(I_{N_A} + \Sigma_A) \Sigma_A^{-1} (U_A^T A^T B U_B \circ U_A^T A^T B U_B) \Sigma_B^{-1} \log(I_{N_B} + \Sigma_B) \mathbf{1}_{N_B}. \quad (\text{C.30})$$

**Theorem 6.** Assume that  $\dim(\mathcal{H}_K) = \infty$ . Let  $\gamma > 0$ ,  $\mu > 0$ . Then

$$d_{\log \text{HS}}^2[(C_{\Phi(\mathbf{x})} + \gamma I), (C_{\Phi(\mathbf{y})} + \mu I)] = \text{tr}[\log(I_{N_A} + \Sigma_A)]^2 + \text{tr}[\log(I_{N_B} + \Sigma_B)]^2 - 2C_{AB} + (\log \gamma - \log \mu)^2. \quad (\text{C.31})$$

The Log-Hilbert-Schmidt inner product between  $(C_{\Phi(\mathbf{x})} + \gamma I)$  and  $(C_{\Phi(\mathbf{y})} + \mu I)$  is

$$\langle (C_{\Phi(\mathbf{x})} + \gamma I), (C_{\Phi(\mathbf{y})} + \mu I) \rangle_{\log \text{HS}} = C_{AB} + (\log \gamma)(\log \mu). \quad (\text{C.32})$$

The Log-Hilbert-Schmidt norm of the operator  $(C_{\Phi(\mathbf{x})} + \gamma I)$  is given by

$$\|(C_{\Phi(\mathbf{x})} + \gamma I)\|_{\log \text{HS}}^2 = \text{tr}[\log(I_{N_A} + \Sigma_A)]^2 + (\log \gamma)^2. \quad (\text{C.33})$$

**Theorem 7.** Assume that  $\dim(\mathcal{H}_K) < \infty$ . Let  $\gamma > 0$ ,  $\mu > 0$ . Then

$$\begin{aligned} d_{\log \text{HS}}^2[(C_{\Phi(\mathbf{x})} + \gamma I), (C_{\Phi(\mathbf{y})} + \mu I)] &= \text{tr}[\log(I_{N_A} + \Sigma_A)]^2 + \text{tr}[\log(I_{N_B} + \Sigma_B)]^2 \\ &\quad - 2C_{AB} + 2(\log \gamma - \log \mu)(\text{tr}[\log(I_{N_A} + \Sigma_A)] - \text{tr}[\log(I_{N_B} + \Sigma_B)]) \\ &\quad + (\log \gamma - \log \mu)^2 \dim(\mathcal{H}_K). \end{aligned} \quad (\text{C.34})$$

The Log-Hilbert-Schmidt inner product between  $(C_{\Phi(\mathbf{x})} + \gamma I)$  and  $(C_{\Phi(\mathbf{y})} + \mu I)$  is

$$\begin{aligned} \langle (C_{\Phi(\mathbf{x})} + \gamma I), (C_{\Phi(\mathbf{y})} + \mu I) \rangle_{\log \text{HS}} &= C_{AB} + (\log \mu) \text{tr}[\log(I_{N_A} + \Sigma_A)] \\ &\quad + (\log \gamma) \text{tr}[\log(I_{N_B} + \Sigma_B)] + (\log \gamma \log \mu) \dim(\mathcal{H}_K). \end{aligned} \quad (\text{C.35})$$

The Log-Hilbert-Schmidt norm of  $(C_{\Phi(\mathbf{x})} + \gamma I_{\mathcal{H}_K})$  is given by

$$\begin{aligned} \|(C_{\Phi(\mathbf{x})} + \gamma I)\|_{\log \text{HS}}^2 &= \text{tr}[\log(I_{N_A} + \Sigma_A)]^2 + 2(\log \gamma) \text{tr}[\log(I_{N_A} + \Sigma_A)] \\ &\quad + (\log \gamma)^2 \dim(\mathcal{H}_K). \end{aligned} \quad (\text{C.36})$$

To prove Theorems 6 and 7, we need the following results.

**Lemma 3.** Let  $\mathcal{H}_1$  and  $\mathcal{H}_2$  be two separable Hilbert spaces. Let  $A : \mathcal{H}_1 \rightarrow \mathcal{H}_2$  and  $B : \mathcal{H}_2 \rightarrow \mathcal{H}_1$  be two bounded linear operators. Then the nonzero eigenvalues of  $BA : \mathcal{H}_1 \rightarrow \mathcal{H}_1$  and  $AB : \mathcal{H}_2 \rightarrow \mathcal{H}_2$ , if they exist, are the same.

**Proof of Lemma 3.** Let  $\lambda \neq 0$  be an eigenvalue for  $AB$  with corresponding eigenvector  $v \neq 0$ , then

$$ABv = \lambda v.$$

Multiplying both sides by  $B$  gives

$$(BA)(Bv) = \lambda(Bv).$$

This shows that  $\lambda$  is an eigenvalue of  $BA$ , with eigenvector  $Bv$ . Note that we also have  $Bv \neq 0$ , since

$$Bv = 0 \implies ABv = 0 = \lambda v \implies v = 0,$$

in contradiction to our assumption that  $v \neq 0$ .

Conversely, if  $\mu \neq 0$  is an eigenvalue of  $BA$  with eigenvector  $w \neq 0$ , then it is also an eigenvalue of  $AB$ , with eigenvector  $Aw \neq 0$ . Thus the nonzero eigenvalues of  $AB$  are the same as those of  $BA$ .  $\square$

**Lemma 4.** Let  $\mathcal{H}_1$  and  $\mathcal{H}_2$  be two separable Hilbert spaces. Let  $A : \mathcal{H}_1 \rightarrow \mathcal{H}_2$  and  $B : \mathcal{H}_2 \rightarrow \mathcal{H}_1$  be two bounded linear operators such that  $AB : \mathcal{H}_2 \rightarrow \mathcal{H}_2$  and  $BA : \mathcal{H}_1 \rightarrow \mathcal{H}_1$  are positive, self-adjoint compact operators. If  $AB$  is trace class, then  $BA$  is also trace class. Furthermore, both  $\log(AB + I_{\mathcal{H}_2})$  and  $\log(BA + I_{\mathcal{H}_1})$  are trace class operators and

$$\text{tr} \log(BA + I_{\mathcal{H}_1}) = \text{tr} \log(AB + I_{\mathcal{H}_2}) < \infty. \quad (\text{C.37})$$

**Proof of Lemma 4.** Let  $\{\lambda_k(AB)\}_{k=1}^{N_{AB}}$  be the strictly positive eigenvalues of  $AB$ , with  $1 \leq N_{AB} \leq \infty$ . Then

$$\text{tr} \log(AB + I_{\mathcal{H}_2}) = \sum_{k=1}^{N_{AB}} \log(\lambda_k(AB) + 1) \leq \sum_{k=1}^{N_{AB}} \lambda_k(AB) < \infty,$$

where we have used the inequality  $\log(1+x) \leq x$  for  $x \geq 0$  and the assumption that  $AB$  is trace class. Thus  $\log(AB + I_{\mathcal{H}_2})$  is a trace class operator. By Lemma 3, the nonzero eigenvalues of  $BA$  are the same as those of  $AB$ . Thus  $\log(BA + I_{\mathcal{H}_1})$  is also a trace class operator and

$$\text{tr} \log(BA + I_{\mathcal{H}_1}) = \text{tr} \log(AB + I_{\mathcal{H}_2}) < \infty.$$

This completes the proof.  $\square$

**Lemma 5.** *Let  $\mathcal{H}_1$  and  $\mathcal{H}_2$  be two separable Hilbert spaces. Let  $A : \mathcal{H}_1 \rightarrow \mathcal{H}_2$  and  $B : \mathcal{H}_2 \rightarrow \mathcal{H}_1$  be bounded linear operators such that  $AB : \mathcal{H}_2 \rightarrow \mathcal{H}_2$  and  $BA : \mathcal{H}_1 \rightarrow \mathcal{H}_1$  are positive, self-adjoint compact operators and that  $AB$  is Hilbert-Schmidt. Then  $BA$  is also Hilbert-Schmidt and*

$$\|\log(I_{\mathcal{H}_1} + BA)\|_{\text{HS}}^2 = \|\log(I_{\mathcal{H}_2} + AB)\|_{\text{HS}}^2. \quad (\text{C.38})$$

**Proof of Lemma 5.** By definition, for  $A : \mathcal{H}_1 \rightarrow \mathcal{H}_2$  and  $B : \mathcal{H}_2 \rightarrow \mathcal{H}_1$ ,

$$\|\log(I_{\mathcal{H}_2} + AB)\|_{\text{HS}}^2 = \sum_{k=1}^{N_{AB}} [\log(1 + \lambda_k(AB))]^2.$$

Similarly,

$$\|\log(I_{\mathcal{H}_1} + BA)\|_{\text{HS}}^2 = \sum_{k=1}^{N_{BA}} [\log(1 + \lambda_k(BA))]^2.$$

Here  $N_{AB}$  and  $N_{BA}$  denote the numbers of positive eigenvalues of  $AB$  and  $BA$ , respectively. We have shown that the nonzero eigenvalues of  $BA$  and  $AB$  are the same. Thus the above two expressions are equal to each other.  $\square$

**Lemma 6.** *Let  $\mathcal{H}$  be a separable Hilbert space and  $A : \mathcal{H} \rightarrow \mathcal{H}$ ,  $B : \mathcal{H} \rightarrow \mathcal{H}$  be two self-adjoint, positive Hilbert-Schmidt operators. Let  $\{\lambda_k(A)\}_{k=1}^{N_A}$  and  $\{\lambda_k(B)\}_{k=1}^{N_B}$  be their positive spectra, respectively, with corresponding normalized eigenvectors  $\{\phi_k(A)\}_{k=1}^{N_A}$  and  $\{\phi_k(B)\}_{k=1}^{N_B}$ . Then*

$$\text{tr}[\log(I + A) \log(I + B)] = \sum_{k=1}^{N_A} \sum_{j=1}^{N_B} \log(1 + \lambda_k(A)) \log(1 + \lambda_j(B)) |\langle \phi_k(A), \phi_j(B) \rangle|^2. \quad (\text{C.39})$$

**Proof of Lemma 6.** From the spectral expansions

$$\log(I + A) = \sum_{k=1}^{N_A} \log(1 + \lambda_k(A)) \phi_k(A) \otimes \phi_k(A),$$

$$\log(I + B) = \sum_{k=1}^{N_B} \log(1 + \lambda_k(B)) \phi_k(B) \otimes \phi_k(B),$$

we have

$$\begin{aligned} & \log(I + A) \log(I + B) \\ &= \left[ \sum_{k=1}^{N_A} \log(1 + \lambda_k(A)) \phi_k(A) \otimes \phi_k(A) \right] \left[ \sum_{k=1}^{N_B} \log(1 + \lambda_k(B)) \phi_k(B) \otimes \phi_k(B) \right] \\ &= \sum_{k=1}^{N_A} \sum_{j=1}^{N_B} \log(1 + \lambda_k(A)) \log(1 + \lambda_j(B)) \langle \phi_k(A), \phi_j(B) \rangle \phi_k(A) \otimes \phi_j(B). \end{aligned}$$

It follows that

$$\begin{aligned} & \text{tr}[\log(I + A) \log(I + B)] \\ &= \sum_{k=1}^{N_A} \sum_{j=1}^{N_B} \log(1 + \lambda_k(A)) \log(1 + \lambda_j(B)) |\langle \phi_k(A), \phi_j(B) \rangle|^2. \end{aligned}$$

This completes the proof of the lemma.  $\square$

**Lemma 7.** Let  $\mathcal{H}_1, \mathcal{H}_2$  be separable Hilbert spaces and  $A : \mathcal{H}_1 \rightarrow \mathcal{H}_2$  be a compact operator such that  $A^T A : \mathcal{H}_1 \rightarrow \mathcal{H}_1$  is Hilbert-Schmidt. Let

$$\log(I_{\mathcal{H}_1} + A^T A) = \sum_{k=1}^{N_A} \log(1 + \lambda_k(A^T A)) \phi_k(A^T A) \otimes \phi_k(A^T A)$$

be an orthogonal spectral decomposition for  $\log(I_{\mathcal{H}_1} + A^T A)$ , with  $\{\lambda_k(A^T A)\}_{k=1}^{N_A}$  being the positive eigenvalues of  $A^T A$  and  $\{\phi_k(A^T A)\}_{k=1}^{N_A}$  their corresponding normalized eigenvectors. Then

$$\log(I_{\mathcal{H}_2} + AA^T) = \sum_{k=1}^{N_A} \log(1 + \lambda_k(A^T A)) \frac{(A\phi_k(A^T A))}{\|A\phi_k(A^T A)\|_{\mathcal{H}_2}} \otimes \frac{(A\phi_k(A^T A))}{\|A\phi_k(A^T A)\|_{\mathcal{H}_2}}$$

is an orthogonal spectral decomposition for  $\log(I_{\mathcal{H}_2} + AA^T)$ , which is equivalent to

$$\log(I_{\mathcal{H}_2} + AA^T) = \sum_{k=1}^{N_A} \frac{\log(1 + \lambda_k(A^T A))}{\lambda_k(A^T A)} (A\phi_k(A^T A)) \otimes (A\phi_k(A^T A)).$$

**Proof of Lemma 7.** By Lemma 3, the positive eigenvalues of  $A^T A$  and  $AA^T$  are the same. Furthermore, if

$$A^T A \phi_k(A^T A) = \lambda_k(A^T A) \phi_k(A^T A)$$

with  $\lambda_k(A^T A) \neq 0$  and  $\phi_k(A^T A) \in \mathcal{H}_1$ , then multiplying both sides by  $A$  gives

$$AA^T [A\phi_k(A^T A)] = \lambda_k(A^T A) [A\phi_k(A^T A)].$$

Thus  $A\phi_k(A^T A) \in \mathcal{H}_2$  is an eigenvector of  $AA^T$  with the same eigenvalue  $\lambda_k(A^T A)$ . Its norm is

$$\begin{aligned} \|A\phi_k(A^T A)\|_{\mathcal{H}_2}^2 &= \langle A\phi_k(A^T A), A\phi_k(A^T A) \rangle_{\mathcal{H}_2} = \langle \phi_k(A^T A), A^T A \phi_k(A^T A) \rangle_{\mathcal{H}_1} \\ &= \lambda_k(A^T A) \|\phi_k(A^T A)\|_{\mathcal{H}_1}^2 = \lambda_k(A^T A). \end{aligned}$$

Also, for  $k \neq j$ ,

$$\begin{aligned} \langle A\phi_k(A^T A), A\phi_j(A^T A) \rangle_{\mathcal{H}_2} &= \langle \phi_k(A^T A), A^T A \phi_j(A^T A) \rangle_{\mathcal{H}_1} \\ &= \lambda_j \langle \phi_k(A^T A), \phi_j(A^T A) \rangle_{\mathcal{H}_1} = 0. \end{aligned}$$

Thus  $\{\frac{1}{\sqrt{\lambda_k(A^T A)}} A\phi_k(A^T A)\}_{k=1}^{N_A}$  are the  $\mathcal{H}_2$ -normalized orthonormal eigenvectors of  $AA^T$  corresponding to the positive eigenvalues  $\{\lambda_k(A^T A)\}_{k=1}^{N_A}$ . From this fact, we obtain the spectral decomposition of  $AA^T$  and hence of  $\log(I_{\mathcal{H}_2} + AA^T)$ . This completes the proof of the lemma.  $\square$

**Lemma 8.** Let  $\mathcal{H}_1, \mathcal{H}_2$  be two separable Hilbert spaces and  $A : \mathcal{H}_1 \rightarrow \mathcal{H}_2, B : \mathcal{H}_1 \rightarrow \mathcal{H}_2$  be compact operators such that  $A^T A : \mathcal{H}_1 \rightarrow \mathcal{H}_1$  and  $B^T B : \mathcal{H}_1 \rightarrow \mathcal{H}_1$  are Hilbert-Schmidt. Then

$$\begin{aligned} &\text{tr}[\log(AA^T + I_{\mathcal{H}_2}) \log(BB^T + I_{\mathcal{H}_2})] \\ &= \mathbf{1}_{N_A}^T \log(I_{N_A} + \Sigma_A) \Sigma_A^{-1} (U_A^T A^T B U_B \circ U_A^T A^T B U_B) \Sigma_B^{-1} \log(I_{N_B} + \Sigma_B) \mathbf{1}_{N_B}. \end{aligned} \quad (\text{C.40})$$

In the above expression,  $\circ$  denotes the Hadamard (element-wise) matrix product,  $N_A, N_B$  are the numbers of nonzero eigenvalues of  $A^T A$  and  $B^T B$ , respectively. The diagonal matrices  $\Sigma_A$ , of size  $N_A \times N_A$ , and  $\Sigma_B$ , of size  $N_B \times N_B$ , and the matrices  $U_A$ , of size  $\dim(\mathcal{H}_1) \times N_A$ , and  $U_B$ , of size  $\dim(\mathcal{H}_1) \times N_B$ , are obtained from the spectral decompositions

$$A^T A = U_A \Sigma_A U_A^T, \quad B^T B = U_B \Sigma_B U_B^T, \quad (\text{C.41})$$

respectively.

**Proof of Lemma 8.** Let  $N_A$  be the number of nonzero eigenvalues of  $AA^T$ , which are the same as those of  $A^T A$ , and  $N_B$  be the number of nonzero eigenvalues of  $BB^T$ , which are the same as those

of  $B^T B$ . Then by applying Lemma 6 and Lemma 7, we get

$$\begin{aligned}
& \text{tr}[\log(AA^T + I_{\mathcal{H}_2}) \log(BB^T + I_{\mathcal{H}_2})] \\
&= \sum_{k=1}^{N_A} \sum_{j=1}^{N_B} \log(1 + \lambda_k(AA^T)) \log(1 + \lambda_j(BB^T)) |\langle \phi_k(AA^T), \phi_j(BB^T) \rangle_{\mathcal{H}_2}|^2 \\
&= \sum_{k=1}^{N_A} \sum_{j=1}^{N_B} \log(1 + \lambda_k(A^T A)) \log(1 + \lambda_j(B^T B)) |\langle \phi_k(AA^T), \phi_j(BB^T) \rangle_{\mathcal{H}_2}|^2 \\
&= \sum_{k=1}^{N_A} \sum_{j=1}^{N_B} \frac{\log(1 + \lambda_k(A^T A)) \log(1 + \lambda_j(B^T B))}{\lambda_k(A^T A) \lambda_j(B^T B)} |\langle A\phi_k(A^T A), B\phi_j(B^T B) \rangle_{\mathcal{H}_2}|^2 \\
&= \sum_{k=1}^{N_A} \sum_{j=1}^{N_B} \frac{\log(1 + \lambda_k(A^T A)) \log(1 + \lambda_j(B^T B))}{\lambda_k(A^T A) \lambda_j(B^T B)} |\langle \phi_k(A^T A), A^T B \phi_j(B^T B) \rangle_{\mathcal{H}_1}|^2
\end{aligned}$$

Consider the following spectral decompositions for  $A^T A$  and  $B^T B$ :

$$A^T A = U_A \Sigma_A U_A^T, \quad B^T B = U_B \Sigma_B U_B^T,$$

where the diagonal matrix  $\Sigma_A$  is of size  $N_A \times N_A$ , with diagonal consisting of the nonzero eigenvalues of  $A^T A$ , and the diagonal matrix  $\Sigma_B$  is of size  $N_B \times N_B$ , with diagonal consisting of the nonzero eigenvalues of  $B^T B$ . Then

$$\begin{aligned}
& \text{tr}[\log(AA^T + I_{\mathcal{H}_2}) \log(BB^T + I_{\mathcal{H}_2})] \\
&= \mathbf{1}_{N_A}^T \log(I_{N_A} + \Sigma_A) \Sigma_A^{-1} (U_A^T A^T B U_B \circ U_A^T A^T B U_B) \Sigma_B^{-1} \log(I_{N_B} + \Sigma_B) \mathbf{1}_{N_B},
\end{aligned}$$

where  $\circ$  denotes the Hadamard (element-wise) matrix product. This completes the proof of the lemma.  $\square$

**Proof of Theorem 6.** By Theorem 4, when  $\dim(\mathcal{H}_K) = \infty$ , we have

$$\begin{aligned}
d_{\log \text{HS}}^2[(C_{\Phi(\mathbf{x})} + \gamma I_{\mathcal{H}_K}), (C_{\Phi(\mathbf{y})} + \mu I_{\mathcal{H}_K})] &= \left\| \log\left(\frac{1}{\gamma} C_{\Phi(\mathbf{x})} + I_{\mathcal{H}_K}\right) - \log\left(\frac{1}{\mu} C_{\Phi(\mathbf{y})} + I_{\mathcal{H}_K}\right) \right\|_{\text{HS}}^2 \\
&\quad + (\log \gamma - \log \mu)^2.
\end{aligned}$$

Expanding the first term, we have

$$\begin{aligned}
& \left\| \log\left(\frac{1}{\gamma} C_{\Phi(\mathbf{x})} + I_{\mathcal{H}_K}\right) - \log\left(\frac{1}{\mu} C_{\Phi(\mathbf{y})} + I_{\mathcal{H}_K}\right) \right\|_{\text{HS}}^2 \\
&= \left\| \log\left(\frac{1}{\gamma} C_{\Phi(\mathbf{x})} + I_{\mathcal{H}_K}\right) \right\|_{\text{HS}}^2 + \left\| \log\left(\frac{1}{\mu} C_{\Phi(\mathbf{y})} + I_{\mathcal{H}_K}\right) \right\|_{\text{HS}}^2 \\
&\quad - 2 \text{tr} \left[ \log\left(\frac{1}{\gamma} C_{\Phi(\mathbf{x})} + I_{\mathcal{H}_K}\right) \log\left(\frac{1}{\mu} C_{\Phi(\mathbf{y})} + I_{\mathcal{H}_K}\right) \right].
\end{aligned}$$

By Lemma 5, the first term is

$$\begin{aligned}
& \left\| \log\left(\frac{1}{\gamma} C_{\Phi(\mathbf{x})} + I_{\mathcal{H}_K}\right) \right\|_{\text{HS}}^2 = \left\| \log\left(\frac{1}{\gamma m} \Phi(\mathbf{x}) J_m^2 \Phi(\mathbf{x})^T + I_{\mathcal{H}_K}\right) \right\|_{\text{HS}}^2 \\
&= \left\| \log\left(\frac{1}{\gamma m} J_m \Phi(\mathbf{x})^T \Phi(\mathbf{x}) J_m + I_m\right) \right\|_{\text{HS}}^2 = \left\| \log\left(\frac{1}{\gamma m} J_m K[\mathbf{x}] J_m + I_m\right) \right\|_{\text{HS}}^2 \\
&= \text{tr} \left[ \log\left(\frac{1}{\gamma m} J_m K[\mathbf{x}] J_m + I_m\right) \right]^2.
\end{aligned}$$

Similarly, the second term is

$$\left\| \log\left(\frac{1}{\mu} C_{\Phi(\mathbf{y})} + I_{\mathcal{H}_K}\right) \right\|_{\text{HS}}^2 = \text{tr} \left[ \log\left(\frac{1}{\mu m} J_m K[\mathbf{y}] J_m + I_m\right) \right]^2.$$

For the third term, we have

$$\begin{aligned} & \text{tr} \left[ \log \left( \frac{1}{\gamma} C_{\Phi(\mathbf{x})} + I_{\mathcal{H}_K} \right) \log \left( \frac{1}{\mu} C_{\Phi(\mathbf{y})} + I_{\mathcal{H}_K} \right) \right] \\ &= \text{tr} \left[ \log \left( \frac{1}{\gamma m} \Phi(\mathbf{x}) J_m^2 \Phi(\mathbf{x})^T + I_{\mathcal{H}_K} \right) \log \left( \frac{1}{\mu m} \Phi(\mathbf{y}) J_m^2 \Phi(\mathbf{y})^T + I_{\mathcal{H}_K} \right) \right]. \end{aligned}$$

Let  $A = \frac{1}{\sqrt{\gamma m}} \Phi(\mathbf{x}) J_m : \mathbb{R}^m \rightarrow \mathcal{H}_K$ ,  $B = \frac{1}{\sqrt{\mu m}} \Phi(\mathbf{y}) J_m : \mathbb{R}^m \rightarrow \mathcal{H}_K$ , then

$$\begin{aligned} AA^T &= \frac{1}{\gamma} C_{\Phi(\mathbf{x})}, \quad BB^T = \frac{1}{\mu} C_{\Phi(\mathbf{y})}, \\ A^T A &= \frac{1}{\gamma m} J_m K[\mathbf{x}] J_m, \quad B^T B = \frac{1}{\mu m} J_m K[\mathbf{y}] J_m, \quad A^T B = \frac{1}{\sqrt{\gamma \mu m}} J_m K[\mathbf{x}, \mathbf{y}] J_m. \end{aligned}$$

Applying Lemma 8, we obtain the formula for the cross term  $C_{AB}$ . For the first two terms, we have

$$\text{tr} \left[ \log \left( \frac{1}{\gamma m} J_m K[\mathbf{x}] J_m + I_m \right) \right]^2 = \text{tr} [\log(A^T A + I_m)]^2 = \text{tr} [\log(I_{N_A} + \Sigma_A)]^2,$$

$$\text{tr} \left[ \log \left( \frac{1}{\mu m} J_m K[\mathbf{y}] J_m + I_m \right) \right]^2 = \text{tr} [\log(B^T B + I_m)]^2 = \text{tr} [\log(I_{N_B} + \Sigma_B)]^2.$$

Combining all the expressions, we obtain the formula for  $d_{\log \text{HS}}$ . The formulas for  $\langle \cdot, \cdot \rangle_{\log \text{HS}}$  and  $\| \cdot \|_{\log \text{HS}}$  are obtained similarly.  $\square$

**Proof of Theorem 7.** By Theorem 5, when  $\dim(\mathcal{H}_K) < \infty$ , we have

$$\begin{aligned} & d_{\log \text{HS}}^2[(C_{\Phi(\mathbf{x})} + \gamma I_{\mathcal{H}_K}), (C_{\Phi(\mathbf{y})} + \mu I_{\mathcal{H}_K})] \\ &= \left\| \log(C_{\Phi(\mathbf{x})} + \gamma I_{\mathcal{H}_K}) - \log(C_{\Phi(\mathbf{y})} + \mu I_{\mathcal{H}_K}) \right\|_{\text{HS}}^2 \\ &= \left\| \log \left( \frac{1}{\gamma} C_{\Phi(\mathbf{x})} + I_{\mathcal{H}_K} \right) - \log \left( \frac{1}{\mu} C_{\Phi(\mathbf{y})} + I_{\mathcal{H}_K} \right) \right\|_{\text{HS}}^2 \\ &+ 2(\log \gamma - \log \mu) \text{tr} \left( \log \left( \frac{1}{\gamma} C_{\Phi(\mathbf{x})} + I_{\mathcal{H}_K} \right) - \log \left( \frac{1}{\mu} C_{\Phi(\mathbf{y})} + I_{\mathcal{H}_K} \right) \right) \\ &+ (\log \gamma - \log \mu)^2 \dim(\mathcal{H}_K). \end{aligned}$$

Let  $A = \frac{1}{\sqrt{\gamma m}} \Phi(\mathbf{x}) J_m : \mathbb{R}^m \rightarrow \mathcal{H}_K$ ,  $B = \frac{1}{\sqrt{\mu m}} \Phi(\mathbf{y}) J_m : \mathbb{R}^m \rightarrow \mathcal{H}_K$ , then

$$\begin{aligned} AA^T &= \frac{1}{\gamma} C_{\Phi(\mathbf{x})}, \quad BB^T = \frac{1}{\mu} C_{\Phi(\mathbf{y})}, \\ A^T A &= \frac{1}{\gamma m} J_m K[\mathbf{x}] J_m, \quad B^T B = \frac{1}{\mu m} J_m K[\mathbf{y}] J_m, \quad A^T B = \frac{1}{\sqrt{\gamma \mu m}} J_m K[\mathbf{x}, \mathbf{y}] J_m. \end{aligned}$$

The first term in the expression for  $d_{\log \text{HS}}^2$  follows from the proof of Theorem 6. For the second term, we have

$$\begin{aligned} \text{tr} \left( \log \left( \frac{1}{\gamma} C_{\Phi(\mathbf{x})} + I_{\mathcal{H}_K} \right) \right) &= \text{tr} \log(AA^T + I_{\mathcal{H}_K}) \\ &= \text{tr} \log(A^T A + I_m) \quad \text{by Lemma 4} \\ &= \text{tr} \log \left( \frac{1}{\gamma m} J_m K[\mathbf{x}] J_m + I_m \right) = \text{tr} [\log(I_{N_A} + \Sigma_A)]. \end{aligned}$$

Similarly

$$\text{tr} \left( \log \left( \frac{1}{\mu} C_{\Phi(\mathbf{y})} + I_{\mathcal{H}_K} \right) \right) = \text{tr} \log \left( \frac{1}{\mu m} J_m K[\mathbf{y}] J_m + I_m \right) = \text{tr} [\log(I_{N_B} + \Sigma_B)].$$

Combining these expressions, we obtain the formula for  $d_{\log \text{HS}}^2$ . The formulas for  $\langle \cdot, \cdot \rangle_{\log \text{HS}}$  and  $\| \cdot \|_{\log \text{HS}}$  are obtained similarly. This completes the proof.  $\square$

## D Further discussions

### D.1 Further discussions on the theoretical framework

In this section, we provide a more detailed discussion of Eqs. (12) and (13) in the main paper. Recall that Eq. (12) defines the manifold of positive definite unitized Hilbert-Schmidt operators on a Hilbert space  $\mathcal{H}$ :

$$\Sigma(\mathcal{H}) = \mathbb{P}(\mathcal{H}) \cap \mathcal{H}_{\mathbb{R}} = \{A + \gamma I > 0 : A^* = A, A \in \text{HS}(\mathcal{H}), \gamma \in \mathbb{R}\}.$$

If  $\dim(\mathcal{H}) = \infty$ , then from the assumption that  $A$  is Hilbert-Schmidt, we have  $\lim_{k \rightarrow \infty} \lambda_k(A) = 0$  and the requirement that  $A + \gamma I > 0$  automatically implies that  $\gamma > 0$ . However, if  $\dim(\mathcal{H}) < \infty$ , then  $\gamma > 0$  can take on negative values, since all we need in this case is that the spectrum of  $A + \gamma I$  be strictly positive.

Furthermore, if  $\gamma > 0$  is sufficiently large, then we can have  $A + \gamma I > 0$  even if  $A$  has negative eigenvalues.

From Section 4.2 onwards, we deal with the subset (Eq. (26))

$$\Sigma^+(\mathcal{H}) = \{A + \gamma I : A \in \text{HS}(\mathcal{H}) \cap \text{Sym}^+(\mathcal{H}), \gamma > 0\},$$

where  $A$  is positive (examples of which are the covariance operators) and  $\gamma > 0$ . This set is a strict subset of the manifold  $\Sigma(\mathcal{H})$ .

We now show that when  $\dim(\mathcal{H}) = \infty$ , the affine-invariant metric defined by Eq. (13)

$$d[(A + \gamma I), (B + \mu I)] = \|\log[(A + \gamma I)^{-1/2}(B + \mu I)(A + \gamma I)^{-1/2}]\|_{\text{eHS}}$$

is finite. Since  $(A + \gamma I)$  and  $(B + \mu I)$  are both positive definite, the operator  $(A + \gamma I)^{-1/2}(B + \mu I)(A + \gamma I)^{-1/2}$  is self-adjoint and positive. Furthermore, it is invertible, with bounded inverse  $(A + \gamma I)^{1/2}(B + \mu I)^{-1}(A + \gamma I)^{1/2}$ . Thus, it is positive definite, that is it belongs to  $\Sigma(\mathcal{H})$ . By Lemma 1, there exists a unique operator  $C \in \text{HS}(\mathcal{H}) \cap \text{Sym}(\mathcal{H})$  and a unique scalar  $\alpha \in \mathbb{R}$  such that

$$(A + \gamma I)^{-1/2}(B + \mu I)(A + \gamma I)^{-1/2} = \exp(C + \alpha I),$$

with

$$\log[(A + \gamma I)^{-1/2}(B + \mu I)(A + \gamma I)^{-1/2}] = C + \alpha I \in \mathcal{H}_{\mathbb{R}}.$$

By definition of the extended Hilbert-Schmidt norm, we have

$$d^2[(A + \gamma I), (B + \mu I)] = \|C + \alpha I\|_{\text{eHS}}^2 = \|C\|_{\text{HS}}^2 + \alpha^2 < \infty,$$

as we claimed.

### D.2 Further discussions on the experimental framework

**Two-layer kernel machine interpretation:** In the kernel setting, our framework can be viewed as a kernel machine with two nonlinear layers. The kernel in the first layer defines a covariance operator for each input sample, which is assumed to be generated by its own probability distribution. In our setting, covariance operators for samples belonging to the same class should be close to each other under the Log-HS metric, and those corresponding to samples belonging to different classes should be far apart. After computing the Log-HS distances between the covariance operators, we enter the second layer where a new kernel is defined using these distances. With this kernel, we can perform any kernel method, such as KernelPCA, for all the input samples, just like we currently perform SVM on top of the Log-HS metric computation. Our current experiments clearly demonstrate the substantial gain in performance of this 2-layer kernel machine over the 1-layer kernel machine, that is SVM with the Log-Euclidean metric on directCOVs. We will report further numerical experiments using other kernel methods in a longer version of the paper and in future work.

**Comparison with the Support Measure Machine (SMM) of [8]:** It would be interesting to compare the performance of our framework in the kernel setting to that of the SMM of [8]. In the SMM approach, for the present context, each input sample would be represented by a mean vector in an RKHS, instead of by a covariance operator as in our framework. We note that while the mean vector

in the feature space may theoretically characterize a distribution, in practice this might not necessarily hold true since we only have finite samples, see e.g. [10]. Thus the performance of the SMM will also depend on the particular application at hand. Experimental results comparing the SMM approach and our framework will be reported in the longer version of the current paper.

**Computational complexity:** Let  $n$  be the number of features and  $m$  be the number of observations for each input sample. Computing the Log-Euclidean metric between two  $n \times n$  covariance matrices takes time  $O(n^2m + n^3)$ , since it takes  $O(n^2m)$  time to compute the matrices themselves and the SVD of an  $n \times n$  matrix takes time  $O(n^3)$ . On the other hand, computing the Log-HS metric between two covariance operators in an infinite-dimensional RKHS, where we deal with Gram matrices of size  $m \times m$ , takes time  $O(m^3)$ . This is of the same order as the computational complexity for computing the Stein and Jeffreys divergences in RKHS in [4].

## References

- [1] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM J. on Matrix An. and App.*, 29(1):328–347, 2007.
- [2] B. J. Boom, J. He, S. Palazzo, P. X. Huang, C. Beyan, H.-M. Chou, F.-P. Lin, C. Spampinato, and R. B. Fisher. A research tool for long-term and continuous analysis of fish assemblage in coral-reefs using underwater camera footage. *Ecological Informatics*, in press, 2013.
- [3] B. Caputo, E. Hayman, and P. Mallikarjuna. Class-specific material categorisation. In *ICCV*, pages 1597–1604, 2005.
- [4] M. Harandi, M. Salzmann, and F. Porikli. Bregman divergences for infinite dimensional covariance matrices. In *CVPR*, 2014.
- [5] R.V. Kadison and J.R. Ringrose. *Fundamentals of the theory of operator algebras. Volume I: Elementary Theory*. Pure and Applied Mathematics. Academic Press, 1983.
- [6] G. Kylberg. The Kylberg texture dataset v. 1.0. External report (Blue series) 35, Centre for Image Analysis, Swedish University of Agricultural Sciences and Uppsala University, 2011.
- [7] G. Laroitonda. Nonpositive curvature: A geometrical approach to Hilbert–Schmidt operators. *Differential Geometry and its Applications*, 25:679–700, 2007.
- [8] K. Muandet, K. Fukumizu, F. Dinuzzo, and B. Schölkopf. Learning from distributions via support measure machines. In *Advances in neural information processing systems (NIPS)*, 2012.
- [9] I. J. Schoenberg. Metric spaces and positive definite functions. *Transactions of the American Mathematical Society*, 44:522–536, 1938.
- [10] B.K. Sriperumbudur, A. Gretton, K. Fukumizu, B. Schölkopf, and G. Lanckriet. Hilbert space embeddings and metrics on probability measures. *The Journal of Machine Learning Research*, 11:1517–1561, 2010.