Infinite-dimensional Log-Determinant divergences between positive definite trace class operators

Hà Quang Minh, Istituto Italiano di Tecnologia (IIT), Genova, ITALY

ISTITUTO ITALIANO DI TECNOLOGIA

Finite-dimensional Alpha Log-Determinant divergences

For $A, B \in \mathrm{Sym}^{++}(n)$, the set of $n \times n$ symmetric, positive definite matrices

$$d^{\alpha}_{\mathrm{logdet}}(A,B) = \frac{4}{1-\alpha^2}\log\frac{\det(\frac{1-\alpha}{2}A + \frac{1+\alpha}{2}B)}{\det(A)^{\frac{1-\alpha}{2}}\det(B)^{\frac{1+\alpha}{2}}}, \quad -1<\alpha<1, \qquad (1)$$

$$d^{1}_{\mathrm{logdet}}(A,B) = \lim_{\alpha\to 1^-} d^{\alpha}_{\mathrm{logdet}}(A,B) = \mathrm{tr}(B^{-1}A - I) - \log\det(B^{-1}A), \qquad (2)$$

$$d^{-1}_{\mathrm{logdet}}(A,B) = \lim_{\alpha\to -1^+} d^{\alpha}_{\mathrm{logdet}}(A,B) = \mathrm{tr}(A^{-1}B - I) - \log\det(A^{-1}B). \qquad (3)$$

The case $\alpha = 0$ gives the symmetric divergence, $d^{0}_{\mathrm{logdet}}(A,B) = 4[\log\det(\frac{A+B}{2}) - \frac{1}{2}\log\det(AB)] = 4d^2_{\mathrm{stein}}$, with d_{stein} being a metric.

Extended trace class operators and extended Fredholm determinant

For $A \in \mathrm{Sym}^{++}(n)$, $\det(A) = \prod_{k=1}^{n}\lambda_k = \exp(\mathrm{tr}[\log(A)])$ and $\log\det(A) = \mathrm{tr}[\log(A)]$

Let $\mathcal{H}$ be a separable Hilbert space, $\dim(\mathcal{H}) = \infty$. For a positive, self-adjoint compact operator $A : \mathcal{H} \to \mathcal{H}$ with eigenvalues $\{\lambda_k\}_{k=1}^{\infty}$ and corresponding orthonormal eigenvectors $\{\phi_k\}_{k=1}^{\infty}$

1. $\log(A) = \sum_{k=1}^{\infty}\log(\lambda_k)\phi_k \otimes \phi_k$ is unbounded since $\lim_{k\to\infty}\log(\lambda_k) = -\infty \Rightarrow$ Consider the bounded regularization $\log(A+\gamma I) = \sum_{k=1}^{\infty}\log(\lambda_k+\gamma)\phi_k\otimes\phi_k$, $\gamma > 0$
2. $\log(A+\gamma I)$, $\gamma \neq 1$, is *not* trace class, thus $\mathrm{tr}[\log(A+\gamma I)]$ is infinite $\Rightarrow$ Consider the extended trace

Extended (unitized) trace class operators $\boxed{\mathrm{Tr}_X(\mathcal{H}) = \{A+\gamma I \ : \ A \in \mathrm{Tr}(\mathcal{H}), \gamma \in \mathbb{R}\}}$ which is a Banach algebra under the extended trace norm $||A+\gamma I||_{\mathrm{tr}_X} = ||A||_{\mathrm{tr}} + |\gamma| = \mathrm{tr}|A| + |\gamma|$

Extended trace: $\boxed{\mathrm{tr}_X(A+\gamma I) = \mathrm{tr}(A) + \gamma}$ for $(A+\gamma I) \in \mathrm{Tr}_X(\mathcal{H})$

Positive definite (unitized) trace class operators $\boxed{\mathrm{PTr}(\mathcal{H}) = \{A+\gamma I > 0 \ : \ A \in \mathrm{Tr}(\mathcal{H}), \gamma \in \mathbb{R}\}}$

$$(A+\gamma I) \in \mathrm{PTr}(\mathcal{H}) \Longleftrightarrow \log(A+\gamma I) \in \mathrm{Tr}_X(\mathcal{H}) \Rightarrow \mathrm{tr}_X[\log(A+\gamma I)] \text{ is finite}$$

Fredholm determinant $\det(A+I) = \prod_{k=1}^{\infty}(\lambda_k+1)$, $A \in \mathrm{Tr}(\mathcal{H})$, $\gamma = 1$

Extended Fredholm determinant for positive definite trace class operators. For $(A+\gamma I) \in \mathrm{PTr}(\mathcal{H})$, define

$$\boxed{\det{}_X(A+\gamma I) = \exp(\mathrm{tr}_X[\log(A+\gamma I)])} \quad \text{satisfying} \quad \det{}_X(A+\gamma I) = \gamma\det[(A/\gamma)+I]$$

Extended Fredholm determinant for extended trace class operators (motivated by the previous property)

$$\boxed{\det{}_X(A+\gamma I) = \gamma\det[(A/\gamma)+I]} \quad (A+\gamma I) \in \mathrm{Tr}_X(\mathcal{H}), \gamma \neq 0 \qquad (4)$$

Generalization of Ky Fan's inequality. For $(A+\gamma I), (B+\mu I) \in \mathrm{PTr}(\mathcal{H})$, for $0 \leq \alpha \leq 1$,

$$\boxed{\det{}_X[\alpha(A+\gamma I) + (1-\alpha)(B+\mu I)] \geq (\gamma/\mu)^{\alpha-\delta}\det{}_X(A+\gamma I)^{\delta}\det{}_X(B+\mu I)^{1-\delta}} \qquad (5)$$

where $\delta = \frac{\alpha\gamma}{(\alpha\gamma+(1-\alpha)\mu)}$. For $0<\alpha<1$, equality occurs if and only if $A = B$ and $\gamma = \mu$. For $\gamma = \mu$,

$$\boxed{\det{}_X[\alpha(A+\gamma I) + (1-\alpha)(B+\gamma I)] \geq \det{}_X(A+\gamma I)^{\alpha}\det{}_X(B+\gamma I)^{1-\alpha}} \qquad (6)$$

Infinite-dimensional Alpha Log-Determinant divergences

For $(A+\gamma I), (B+\mu I) \in \mathrm{PTr}(\mathcal{H})$, $-1 < \alpha < 1$, define

$$\boxed{d^{\alpha}_{\mathrm{logdet}}[(A+\gamma I),(B+\mu I)] = \frac{4}{1-\alpha^2}\log\left[\frac{\det_X\left(\frac{1-\alpha}{2}(A+\gamma I) + \frac{1+\alpha}{2}(B+\mu I)\right)}{\det_X(A+\gamma I)^{\beta}\det_X(B+\mu I)^{1-\beta}}\left(\frac{\gamma}{\mu}\right)^{\beta-\frac{1-\alpha}{2}}\right]} \qquad (7)$$

where $\beta = \frac{(1-\alpha)\gamma}{(1-\alpha)\gamma+(1+\alpha)\mu}$. Defining $d^{1}_{\mathrm{logdet}}[(A+\gamma I),(B+\mu I)] = \lim_{\alpha\to 1^-} d^{\alpha}_{\mathrm{logdet}}[(A+\gamma I),(B+\mu I)]$, then

$$d^{1}_{\mathrm{logdet}}[(A+\gamma I),(B+\mu I)] = (\frac{\gamma}{\mu}-1)\log\frac{\gamma}{\mu} + \mathrm{tr}_X[(B+\mu I)^{-1}(A+\gamma I) - I]$$
$$-\frac{\gamma}{\mu}\log\det{}_X[(B+\mu I)^{-1}(A+\gamma I)]. \qquad (8)$$

Defining $d^{-1}_{\mathrm{logdet}}[(A+\gamma I),(B+\mu I)] = \lim_{\alpha\to -1^+} d^{\alpha}_{\mathrm{logdet}}[(A+\gamma I),(B+\mu I)]$, then

$$d^{-1}_{\mathrm{logdet}}[(A+\gamma I),(B+\mu I)] = (\frac{\mu}{\gamma}-1)\log\frac{\mu}{\gamma} + \mathrm{tr}_X\left[(A+\gamma I)^{-1}(B+\mu I) - I\right]$$
$$-\frac{\mu}{\gamma}\log\det{}_X[(A+\gamma I)^{-1}(B+\mu I)]. \qquad (9)$$

Special cases. For $\gamma = \mu$, $\beta = (1-\alpha)$ and

$$d^{\alpha}_{\mathrm{logdet}}[(A+\gamma I),(B+\gamma I)] = \frac{4}{1-\alpha^2}\log\left[\frac{\det_X\left(\frac{1-\alpha}{2}(A+\gamma I) + \frac{1+\alpha}{2}(B+\gamma I)\right)}{\det_X(A+\gamma I)^{\frac{1-\alpha}{2}}\det_X(B+\gamma I)^{\frac{1+\alpha}{2}}}\right], \quad -1<\alpha<1 \qquad (10)$$

$$d^{1}_{\mathrm{logdet}}[(A+\gamma I),(B+\gamma I)] = \mathrm{tr}_X[(B+\gamma I)^{-1}(A+\gamma I) - I] - \log\det{}_X[(B+\gamma I)^{-1}(A+\gamma I)] \qquad (11)$$

$$d^{-1}_{\mathrm{logdet}}[(A+\gamma I),(B+\gamma I)] = \mathrm{tr}_X\left[(A+\gamma I)^{-1}(B+\gamma I) - I\right] - \log\det{}_X[(A+\gamma I)^{-1}(B+\gamma I)]. \qquad (12)$$

Finite-dimensional case: $A, B \in \mathrm{Sym}^{++}(n)$, $\gamma = 0$.

Log-Determinant divergences between RKHS covariance operators

Reproducing kernel Hilbert spaces (RKHS) and feature maps. Let $\mathcal{X}$ be a separable topological space, K a continuous positive definite kernel on $\mathcal{X}\times\mathcal{X}$, with corresponding RKHS $\mathcal{H}_K$, which is a separable Hilbert space. Then $\exists$ a corresponding feature map $\Phi : \mathcal{X} \to \mathcal{H}_K$, so that $K(x,y) = \langle\Phi(x),\Phi(y)\rangle_{\mathcal{H}_K}$ $\forall (x,y) \in \mathcal{X}\times\mathcal{X}$.
Covariance operators. Let $\mathbf{X} = [x_1,\ldots,x_m]$ be a data matrix randomly sampled from $\mathcal{X}$ according to some probability distribution. Define the bounded linear operator $\Phi(\mathbf{X}) : \mathbb{R}^m \to \mathcal{H}_K$ by $\Phi(\mathbf{X})\mathbf{b} = \sum_{j=1}^{m} b_j\Phi(x_j)$, $\mathbf{b} \in \mathbb{R}^m$. Informally, $\Phi(\mathbf{X})$ can be viewed as a (infinite) data matrix $\Phi(\mathbf{X}) = [\Phi(x_1),\ldots,\Phi(x_m)]$ in $\mathcal{H}_K$ with

$$\text{Covariance operator} \quad C_{\Phi(\mathbf{X})} = \frac{1}{m}\Phi(\mathbf{X})J_m\Phi(\mathbf{X})^* : \mathcal{H}_K \to \mathcal{H}_K, \quad J_m = I_m - \frac{1}{m}\mathbf{1}_m\mathbf{1}_m^T. \qquad (13)$$

For two covariance operators $C_{\Phi(\mathbf{X})}$ and $C_{\Phi(\mathbf{Y})}$, the Log-Determinant divergence

$$d^{\alpha}_{\mathrm{logdet}}[(C_{\Phi(\mathbf{X})} + \gamma I_{\mathcal{H}_K}), (C_{\Phi(\mathbf{Y})} + \mu I_{\mathcal{H}_K})], \quad \gamma > 0, \mu > 0 \qquad (14)$$

has a closed form expressed in terms of the corresponding Gram matrices.

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