Review

Log-Determinant Divergences Revisited: Alpha-Beta and Gamma Log-Det Divergences

Andrzej Cichocki 1,2,*, Sergio Cruces 3,* and Shun-ichi Amari 4

1 Laboratory for Advanced Brain Signal Processing, Brain Science Institute, RIKEN, 2-1 Hirosawa, Wako, 351-0198 Saitama, Japan
2 Systems Research Institute, Intelligent Systems Laboratory, Newelska 6, 01-447 Warsaw, Poland
3 Dpto de Teoría de la Señal y Comunicaciones, University of Seville, Camino de los Descubrimientos s/n, 41092 Seville, Spain
4 Laboratory for Mathematical Neuroscience, RIKEN BSI, Wako, 351-0198 Saitama, Japan; E-Mail: amari@brain.riken.jp

* Authors to whom correspondence should be addressed; E-Mails: a.cichocki@riken.jp (A.C.); sergio@us.es (S.C.).

Academic Editor: Raúl Alcaraz Martínez

Received: 19 December 2014 / Accepted: 5 May 2015 / Published: 8 May 2015

Abstract: This work reviews and extends a family of log-determinant (log-det) divergences for symmetric positive definite (SPD) matrices and discusses their fundamental properties. We show how to use parameterized Alpha-Beta (AB) and Gamma log-det divergences to generate many well-known divergences; in particular, we consider the Stein’s loss, the S-divergence, also called Jensen-Bregman LogDet (JBLD) divergence, Logdet Zero (Bhattacharyya) divergence, Affine Invariant Riemannian Metric (AIRM), and other divergences. Moreover, we establish links and correspondences between log-det divergences and visualise them on an alpha-beta plane for various sets of parameters. We use this unifying framework to interpret and extend existing similarity measures for semidefinite covariance matrices in finite-dimensional Reproducing Kernel Hilbert Spaces (RKHS). This paper also shows how the Alpha-Beta family of log-det divergences relates to the divergences of multivariate and multilinear normal distributions. Closed form formulas are derived for Gamma divergences of two multivariate Gaussian densities; the special cases of the Kullback-Leibler, Bhattacharyya, Rényi, and Cauchy-Schwartz divergences are discussed. Symmetrized versions of log-det divergences are also considered and briefly reviewed. Finally, a class of divergences is extended to multiway divergences for separable covariance (or precision) matrices.
**Keywords:** Similarity measures; generalized divergences for symmetric positive definite (covariance) matrices; Stein’s loss; Burg’s matrix divergence; Affine Invariant Riemannian Metric (AIRM); Riemannian metric; geodesic distance; Jensen-Bregman LogDet (JBLD); S-divergence; LogDet Zero divergence; Jeffrey’s KL divergence; symmetrized KL Divergence Metric (KLDM); Alpha-Beta Log-Det divergences; Gamma divergences; Hilbert projective metric and their extensions

1. Introduction

Divergences or (dis)similarity measures between symmetric positive definite (SPD) matrices underpin many applications, including: Diffusion Tensor Imaging (DTI) segmentation, classification, clustering, pattern recognition, model selection, statistical inference, and data processing problems [1–3]. Furthermore, there is a close connection between divergence and the notions of entropy, information geometry, and statistical mean [2,4–7], while matrix divergences are closely related to the invariant geometrical properties of the manifold of probability distributions [4,8–10]. A wide class of parameterized divergences for positive measures are already well understood and a unification and generalization of their properties can be found in [11–13].

The class of SPD matrices, especially covariance matrices, play a key role in many areas of statistics, signal/image processing, DTI, pattern recognition, and biological and social sciences [14–16]. For example, medical data produced by diffusion tensor magnetic resonance imaging (DTI-MRI) represent the covariance in a Brownian motion model of water diffusion. The diffusion tensors can be represented as SPD matrices, which are used to track the diffusion of water molecules in the human brain, with applications such as the diagnosis of mental disorders [14]. In array processing, covariance matrices capture both the variance and correlation of multidimensional data; this data is often used to estimate (dis)similarity measures, i.e., divergences. This all has led to an increasing interest in divergences for SPD (covariance) matrices [1,5,6,14,17–20].

The main aim of this paper is to review and extend log-determinant (log-det) divergences and to establish a link between log-det divergences and standard divergences, especially the Alpha, Beta, and Gamma divergences. Several forms of the log-det divergence exist in the literature, including the log–determinant $\alpha$ divergence, Riemannian metric, Stein’s loss, S-divergence, also called the Jensen-Bregman LogDet (JBLD) divergence, and the symmetrized Kullback-Leibler Density Metric (KLDM) or Jeffrey’s KL divergence [5,6,14,17–20]. Despite their numerous applications, common theoretical properties and the relationships between these divergences have not been established. To this end, we propose and parameterize a wide class of log-det divergences that provide robust solutions and/or even improve the accuracy for a noisy data. We next review fundamental properties and provide relationships among the members of this class. The advantages of some selected log-det divergences are also discussed; in particular, we consider the efficiency, simplicity, and resilience to noise or outliers, in addition to simplicity of calculations [14]. The log-det divergences between two SPD matrices have
also been shown to be robust to biases in composition, which can cause problems for other similarity measures.

The divergences discussed in this paper are flexible enough to facilitate the generation of several established divergences (for specific values of the tuning parameters). In addition, by adjusting the adaptive tuning parameters, we optimize the cost functions of learning algorithms and estimate desired model parameters in the presence of noise and outliers. In other words, the divergences discussed in this paper are robust with respect to outliers and noise if the tuning parameters, $\alpha$, $\beta$, and $\gamma$, are chosen properly.

1.1. Preliminaries

We adopt the following notation: SPD matrices will be denoted by $P \in \mathbb{R}^{n \times n}$ and $Q \in \mathbb{R}^{n \times n}$, and have positive eigenvalues $\lambda_i$ (sorted in descending order); by $\log(P)$, $\det(P) = |P|$, $\text{tr}(P)$ we denote the logarithm, determinant, and trace of $P$, respectively.

For any real parameter $\alpha \in \mathbb{R}$ and for a positive definite matrix $P$, the matrix $P^\alpha$ is defined using symmetric eigenvalue decomposition as follows:

$$P^\alpha = (V \Lambda V^T)^\alpha = V(\Lambda^\alpha) V^T,$$

where $\Lambda$ is a diagonal matrix of the eigenvalues of $P$, and $V \in \mathbb{R}^{n \times n}$ is the orthogonal matrix of the corresponding eigenvectors. Similarly, we define

$$\log(P^\alpha) = \log((V \Lambda V^T)^\alpha) = V \log(\Lambda^\alpha) V^T,$$

where $\log(\Lambda)$ is a diagonal matrix of logarithms of the eigenvalues of $P$. The basic operations for positive definite matrices are provided in Appendix A.

The dissimilarity between two SPD matrices is called a metric if the following conditions hold:

1. $D(P\parallel Q) \geq 0$, where the equality holds if and only if $P = Q$ (nonnegativity and positive definiteness).
2. $D(P\parallel Q) = D(Q\parallel P)$ (symmetry).
3. $D(P\parallel Z) \leq D(P\parallel Q) + D(Q\parallel Z)$ (subadditivity/triangle inequality).

Dissimilarities that only satisfy condition (1) are not metrics and are referred to as (asymmetric) divergences.

2. Basic Alpha-Beta Log-Determinant Divergence

For SPD matrices $P \in \mathbb{R}^{n \times n}$ and $Q \in \mathbb{R}^{n \times n}$, consider a new dissimilarity measure, namely, the AB log-det divergence, given by

$$D_{AB}^{(\alpha, \beta)}(P\parallel Q) = \frac{1}{\alpha \beta} \log \det \frac{\alpha(PQ^{-1})^\beta + \beta(PQ^{-1})^{-\alpha}}{\alpha + \beta} \quad \text{for } \alpha \neq 0, \beta \neq 0, \alpha + \beta \neq 0,$$

for $\alpha \neq 0, \beta \neq 0, \alpha + \beta \neq 0,$
where the values of the parameters $\alpha$ and $\beta$ can be chosen so as to guarantee the non-negativity of the divergence and it vanishes to zero if and only if $P = Q$ (this issue is addressed later by Theorems 1 and 2). Observe that this is not a symmetric divergence with respect to $P$ and $Q$, except when $\alpha = \beta$. Note that using the identity $\log \det(P) = \tr \log(P)$, the divergence in (3) can be expressed as

$$D_{AB}^{(\alpha,\beta)}(P\|Q) = \frac{1}{\alpha \beta} \tr \left[ \log \left( \frac{\alpha (PQ^{-1})^\beta + \beta (PQ^{-1})^{-\alpha}}{\alpha + \beta} \right) \right]$$

(4)

for $\alpha \neq 0$, $\beta \neq 0$, $\alpha + \beta \neq 0$.

This divergence is related to the Alpha, Beta, and AB divergences discussed in our previous work, especially Gamma divergences [11–13,21]. Furthermore, the divergence in (4) is related to the AB divergence for SPD matrices [1,12], which is defined by

$$\bar{D}_{AB}^{(\alpha,\beta)}(P\|Q) = \frac{1}{\alpha \beta} \tr \left( \frac{\alpha P^{\alpha+\beta} + \beta Q^{\alpha+\beta}}{\alpha + \beta} - P^{\alpha}Q^{\beta} \right)$$

(5)

for $\alpha \neq 0$, $\beta \neq 0$. $\alpha + \beta \neq 0$.

Note that $\alpha$ and $\beta$ are chosen so that $D_{AB}^{(\alpha,\beta)}(P\|Q)$ is nonnegative and equal to zero if $P = Q$. Moreover, such divergence functions can be evaluated without computing the inverses of the SPD matrices; instead, they can be evaluated easily by computing (positive) eigenvalues of the matrix $PQ^{-1}$ or its inverse. Since both matrices $P$ and $Q$ (and their inverses) are SPD matrices, their eigenvalues are positive. In general, it can be shown that even though $PQ^{-1}$ is nonsymmetric, its eigenvalues are the same as those of the SPD matrix $Q^{-1/2}PQ^{-1/2}$; hence, its eigenvalues are always positive.

Next, consider the eigenvalue decomposition:

$$(PQ^{-1})^\beta = V \Lambda^\beta V^{-1},$$

(6)

where $V$ is a nonsingular matrix, and $\Lambda^\beta = \text{diag}\{\lambda_1^\beta, \lambda_2^\beta, \ldots, \lambda_n^\beta\}$ is the diagonal matrix with the positive eigenvalues $\lambda_i > 0$, $i = 1, 2, \ldots, n$, of $PQ^{-1}$. Then, we can write

$$D_{AB}^{(\alpha,\beta)}(P\|Q) = \frac{1}{\alpha \beta} \log \det \frac{\alpha V \Lambda^\beta V^{-1} + \beta V \Lambda^{-\alpha} V^{-1}}{\alpha + \beta}$$

$$= \frac{1}{\alpha \beta} \log \left[ \det V \det \frac{\alpha \Lambda^\beta + \beta \Lambda^{-\alpha}}{\alpha + \beta} \det V^{-1} \right]$$

$$= \frac{1}{\alpha \beta} \log \det \frac{\Lambda^\beta + \beta \Lambda^{-\alpha}}{\alpha + \beta},$$

(7)

which allows us to use simple algebraic manipulations to obtain

$$D_{AB}^{(\alpha,\beta)}(P\|Q) = \frac{1}{\alpha \beta} \log \prod_{i=1}^{n} \frac{\alpha \lambda_i^\beta + \beta \lambda_i^{-\alpha}}{\alpha + \beta}$$

$$= \frac{1}{\alpha \beta} \sum_{i=1}^{n} \log \left( \frac{\alpha \lambda_i^\beta + \beta \lambda_i^{-\alpha}}{\alpha + \beta} \right), \quad \alpha, \beta, \alpha + \beta \neq 0.$$  

(8)

It is straightforward to verify that $D_{AB}^{(\alpha,\beta)}(P\|Q) = 0$ if $P = Q$. We will show later that this function is nonnegative for any SPD matrices if the $\alpha$ and $\beta$ parameters take both positive or negative values.
For the singular values $\alpha = 0$ and/or $\beta = 0$ (also $\alpha = -\beta$), the AB log-det divergence in (3) is defined as a limit for $\alpha \to 0$ and/or $\beta \to 0$. In other words, to avoid indeterminacy or singularity for specific parameter values, the AB log-det divergence can be reformulated or extended by continuity and by applying L'Hôpital’s formula to cover the singular values of $\alpha$ and $\beta$. Using L'Hôpital’s rule, the AB log-det divergence can be defined explicitly by

$$D_{AB}^{(\alpha,\beta)}(\parallel{P||Q}) = \begin{cases} 
\frac{1}{\alpha\beta} \log \det \frac{\alpha(PQ^{-1})^\beta + \beta(QP^{-1})^\alpha}{\alpha + \beta} & \text{for } \alpha, \beta \neq 0, \ \alpha + \beta \neq 0 \\
\frac{1}{\alpha^2} \left[ \text{tr} \left( (QP^{-1})^\alpha - I \right) - \alpha \log \det(QP^{-1}) \right] & \text{for } \alpha \neq 0, \ \beta = 0 \\
\frac{1}{\beta^2} \left[ \text{tr} \left( (PQ^{-1})^\beta - I \right) - \beta \log \det(PQ^{-1}) \right] & \text{for } \alpha = 0, \ \beta \neq 0 \\
\frac{1}{\alpha^2} \log \frac{\det(PQ^{-1})^\alpha}{\det(I + \log(PQ^{-1})^\alpha)} & \text{for } \alpha = -\beta \neq 0 \\
\frac{1}{2} \text{tr} \log^2(PQ^{-1}) = \frac{1}{2} \left\| \log(Q^{-1/2}PQ^{-1/2}) \right\|_F^2 & \text{for } \alpha, \ \beta = 0.
\end{cases}$$

Equivalently, using standard matrix manipulations, the above formula can be expressed in terms of the eigenvalues of $PQ^{-1}$, i.e., the generalized eigenvalues computed from $\lambda_i Q v_i = P v_i$ (where $v_i (i = 1, 2, \ldots, n)$ are corresponding generalized eigenvectors) as follows:

$$D_{AB}^{(\alpha,\beta)}(P\parallel{Q}) = \begin{cases} 
\frac{1}{\alpha\beta} \sum_{i=1}^{n} \log \left( \frac{\alpha\lambda_i^\beta + \beta\lambda_i^{-\alpha}}{\alpha + \beta} \right) & \text{for } \alpha, \beta \neq 0, \ \alpha + \beta \neq 0 \\
\frac{1}{\alpha^2} \left[ \sum_{i=1}^{n} \left( \lambda_i^{-\alpha} - \log(\lambda_i^{-\alpha}) \right) - n \right] & \text{for } \alpha \neq 0, \ \beta = 0 \\
\frac{1}{\beta^2} \left[ \sum_{i=1}^{n} \left( \lambda_i^\beta - \log(\lambda_i^\beta) \right) - n \right] & \text{for } \alpha = 0, \ \beta \neq 0 \\
\frac{1}{\alpha^2} \sum_{i=1}^{n} \log \left( \frac{\lambda_i^\alpha}{1 + \log \lambda_i^\alpha} \right) & \text{for } \alpha = -\beta \neq 0 \\
\frac{1}{2} \sum_{i=1}^{n} \log^2(\lambda_i) & \text{for } \alpha, \ \beta = 0.
\end{cases}$$

**Theorem 1.** The function $D_{AB}^{(\alpha,\beta)}(P\parallel{Q}) \geq 0$ given by (3) is nonnegative for any SPD matrices with arbitrary positive eigenvalues if $\alpha \geq 0$ and $\beta \geq 0$ or if $\alpha < 0$ and $\beta < 0$. It is equal to zero if and only if $P = Q$.

Equivalently, if the values of $\alpha$ and $\beta$ have the same sign, the AB log-det divergence is positive independent of the distribution of the eigenvalues of $PQ^{-1}$ and goes to zero if and only if all the
The eigenvalues are equal to one. However, if the eigenvalues are sufficiently close to one, the AB log-det divergence is also positive for different signs of $\alpha$ and $\beta$. The conditions for positive definiteness are given by the following theorem.

**Theorem 2.** The function $D_{AB}^{(\alpha,\beta)}(P\|Q)$ given by (9) is nonnegative if $\alpha > 0$ and $\beta < 0$ or if $\alpha < 0$ and $\beta > 0$ and if all the eigenvalues of $PQ^{-1}$ satisfy the following conditions:

$$
\lambda_i > \left| \frac{\beta}{\alpha} \right|^{-\frac{1}{\alpha+\beta}} \quad \forall i, \text{ for } \alpha > 0 \text{ and } \beta < 0,
$$

**(11)**

and

$$
\lambda_i < \left| \frac{\beta}{\alpha} \right|^{-\frac{1}{\alpha+\beta}} \quad \forall i, \text{ for } \alpha < 0 \text{ and } \beta > 0.
$$

**(12)**

If any of the eigenvalues violate these bounds, the value of the divergence, by definition, is infinite. Moreover, when $\alpha \to -\beta$ these bounds simplify to

$$
\lambda_i > e^{-1/\alpha} \quad \forall i, \alpha = -\beta > 0,
$$

**(13)**

$$
\lambda_i < e^{-1/\alpha} \quad \forall i, \alpha = -\beta < 0.
$$

**(14)**

In the limit, when $\alpha \to 0$ or $\beta \to 0$, the bounds disappear. A visual presentation of these bounds for different values of $\alpha$ and $\beta$ is shown in Figure 1.

Additionally, $D_{AB}^{(\alpha,\beta)}(P\|Q) = 0$ only if $\lambda_i = 1$ for all $i = 1, \ldots, n$, i.e., when $P = Q$.

![Figure 1](image-url)

**Figure 1.** Shaded-contour plots of the bounds of $\lambda_i$ that prevent $D_{AB}^{(\alpha,\beta)}(P\|Q)$ from diverging to $\infty$. The positive lower-bounds are shown in the lower-right quadrant of (a). The finite upper-bounds are shown in the upper-left quadrant of (b).

The proofs of these theorems are given in Appendices B, C and D.

Figure 2 illustrates the typical shapes of the AB log-det divergence for different values of the eigenvalues for various choices of $\alpha$ and $\beta$. 

![Figure 2](image-url)
In general, the AB log-det divergence is not a metric distance since the triangle inequality is not satisfied for all parameter values. Therefore, we can define the metric distance as the square root of the AB log-det divergence in the special case when $\alpha = \beta$ as follows:

$$d_{\alpha,\alpha}^{AB}(P\|Q) = \sqrt{D_{\alpha,\alpha}^{AB}(P\|Q)}.$$  \hspace{1cm} (15)

This follows from the fact that $D_{\alpha,\alpha}^{AB}(P\|Q)$ is symmetric with respect to $P$ and $Q$.

Later, we will show that measures defined in this manner lead to many important and well-known divergences and metric distances such as the Logdet Zero divergence, Affine Invariant Riemannian metric (AIRM), and square root of Stein’s loss [5,6]. Moreover, new divergences can be generated; specifically, generalized Stein’s loss, the Beta-log-det divergence, and extended Hilbert metrics.

From the divergence $D_{\alpha,\beta}^{AB}(P\|Q)$, a Riemannian metric and a pair of dually coupled affine connections are introduced in the manifold of positive definite matrices. Let $dP$ be a small deviation.
of $P$, which belongs to the tangent space of the manifold at $P$. Calculating $D_{AB}^{(\alpha,\beta)}(P + dP\|P)$ and neglecting higher-order terms yields (see Appendix E)

$$D_{AB}^{(\alpha,\beta)}(P + dP\|P) = \frac{1}{2} \text{tr}[dP \, P^{-1} \, dP \, P^{-1}].$$

(16)

This gives a Riemannian metric that is common for all $(\alpha, \beta)$. Therefore, the Riemannian metric is the same for all AB log-det divergences, although the dual affine connections depend on $\alpha$ and $\beta$. The Riemannian metric is also the same as the Fisher information matrix of the manifold of multivariate Gaussian distributions of mean zero and covariance matrix $P$.

Interestingly, note that the Riemannian metric or geodesic distance is obtained from (3) for $\alpha = \beta = 0$:

$$d_R(P\|Q) = \sqrt{2 \, D_{AB}^{(0,0)}(P\|Q)} = \sqrt{\text{tr} \, \log^2(PQ^{-1})}$$

(17)

$$= \| \log(PQ^{-1}) \|_F = \| \log(Q^{-1/2}PQ^{-1/2}) \|_F$$

(18)

$$= \sqrt{\sum_{i=1}^{n} \log^2(\lambda_i)},$$

(19)

Figure 3. Graphical illustration of the fundamental, nonsymmetric, AB log-det divergences. On the $\alpha$-$\beta$-plane, important divergences are indicated by points and lines, especially the Stein’s loss and its generalization, the AIRM (Riemannian) distance, S-divergence (JBLD), Alpha-log-det divergence $D_A^{(\alpha)}$, and Beta-log-det divergence $D_B^{(\beta)}$. 

The Riemannian metric is also the same as the Fisher information matrix of the manifold of multivariate Gaussian distributions of mean zero and covariance matrix $P$.
where $\lambda_i$ are the eigenvalues of $PQ^{-1}$.

This is also known as the AIRM. AIRM takes advantage of several important and useful theoretical properties and is probably one of the most widely used (dis)similarity measure for SPD (covariance) matrices [14,15].

For $\alpha = \beta = 0.5$ (and for $\alpha = \beta = -0.5$), the recently defined and deeply analyzed S-divergence (JBLD) [6,14,15,17] is obtained:

$$D_S(P\|Q) = D_{AB}^{(0.5,0.5)}(P\|Q) = 4 \log \det \left( \frac{1}{2} \left[ \left( PQ^{-1}\right)^{1/2} + \left( PQ^{-1}\right)^{-1/2} \right] \right)$$

$$= \frac{4 \log \det \frac{\left( PQ^{-1}\right)^{1/2} + \left( PQ^{-1}\right)^{-1/2}}{2}}{\det(P)^{1/2} \det(Q)^{1/2}}$$

$$= \frac{4 \log \frac{\det \frac{1}{2}(P + Q)}{\sqrt{\det(P) \det(Q)}}}{\det(P)^{1/2} \det(Q)^{1/2}}$$

$$= 4 \left( \log \det \left( \frac{P + Q}{2} \right) - \frac{1}{2} \log \det(PQ) \right) = 4 \sum_{i=1}^{n} \log \left( \frac{\lambda_i + 1}{2\sqrt{\lambda_i}} \right). \quad (20)$$

The S-divergence is not a metric distance. To make it a metric, we take its square root and obtain the LogDet Zero divergence, or Bhattacharyya distance [5,7,18]:

$$d_{Bh}(P\|Q) = \sqrt{D_{AB}^{(0.5,0.5)}(P\|Q)}$$

$$= 2 \sqrt{\log \det \left( \frac{P + Q}{2} \right) - \frac{1}{2} \log \det(PQ)}$$

$$= 2 \sqrt{\log \frac{\det \frac{1}{2}(P + Q)}{\sqrt{\det(P) \det(Q)}}}. \quad (21)$$

Moreover, for $\alpha = 0$, $\beta \neq 0$ and $\alpha \neq 0$, $\beta = 0$, we obtain divergences which are generalizations of Stein’s loss (called also Burg matrix divergence or simply LogDet divergence):

$$D_{AB}^{(0,\beta)}(P\|Q) = \frac{1}{\beta^2} \left[ \text{tr} \left( (PQ^{-1})^\beta - I \right) - \beta \log \det(PQ^{-1}) \right], \ \beta \neq 0. \quad (22)$$

$$D_{AB}^{(\alpha,0)}(P\|Q) = \frac{1}{\alpha^2} \left[ \text{tr} \left( (QP^{-1})^\alpha - I \right) - \alpha \log \det(QP^{-1}) \right], \ \alpha \neq 0 \quad (23)$$

The divergences in (22) and (23) simplify, respectively, to the standard Stein’s loss if $\beta = 1$ and to its dual loss if $\alpha = 1$.

3. Special Cases of the AB Log-Det Divergence

We now illustrate how a suitable choice of the $(\alpha, \beta)$ parameters simplify the AB log-det divergence into other known divergences such as the Alpha- and Beta-log-det divergences [5,11,18,23] (see Figure 3 and Table 1).
### Table 1. Fundamental Log-det Divergences and Distances

<table>
<thead>
<tr>
<th>Divergence Type</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geodesic Distance (AIRM) ((\alpha = \beta = 0))</td>
<td>(\frac{1}{2} d_{\text{AR}}^2(P \parallel Q) = \frac{1}{2} \text{tr} \log^2(PQ^{-1}) = \frac{1}{2} \sum_{i=1}^{n} \log^2 \lambda_i)</td>
</tr>
<tr>
<td>S-divergence (Squared Bhattacharyya Distance) ((\alpha = \beta = 0.5))</td>
<td>(D_S(P \parallel Q) = d_{\text{bh}}^2(P \parallel Q) = 4 \log \frac{\det \frac{P + Q}{2}}{\det \frac{PQ}{2}} = 4 \sum_{i=1}^{n} \log \frac{\lambda_i + 1}{2\sqrt{\lambda_i}})</td>
</tr>
<tr>
<td>Power divergence ((\alpha = \beta \neq 0))</td>
<td>(\frac{1}{\alpha^2} \log \det \frac{(PQ^{-1})^\alpha - (PQ^{-1})^{-\alpha}}{2} = \frac{1}{\alpha^2} \sum_{i=1}^{n} \log \frac{\lambda_i^\alpha + \lambda_i^{-\alpha}}{2})</td>
</tr>
<tr>
<td>Generalized Burg divergence (Stein’s Loss) ((\alpha = 0, \beta \neq 0))</td>
<td>(\frac{1}{\beta^2} \text{tr} \left[ (PQ^{-1})^\beta - I \right] - \log \det (PQ^{-1})^\beta = \frac{1}{\beta^2} \left( \sum_{i=1}^{n} \left( \lambda_i^\beta - \log \lambda_i^\beta \right) - n \right))</td>
</tr>
<tr>
<td>Generalized Itakura-Saito log-det divergence ((\alpha = -\beta \neq 0))</td>
<td>(\frac{1}{\alpha^2} \log \det \frac{\det (PQ^{-1})^\alpha}{1 + \log (PQ^{-1})^\alpha} = \frac{1}{\alpha^2} \sum_{i=1}^{n} \log \frac{\lambda_i^\alpha}{1 + \log^2 \lambda_i})</td>
</tr>
<tr>
<td>Alpha log-det divergence ((0 &lt; \alpha &lt; 1, \beta = 1 - \alpha))</td>
<td>(D_A^{(\alpha)}(P\parallel Q) = \frac{1}{\alpha(1 - \alpha)} \log \frac{\det (\alpha P + (1 - \alpha)Q)}{\det (P^{\alpha} Q^{1-\alpha})} = \frac{1}{\alpha(1 - \alpha)} \sum_{i=1}^{n} \log \left( \frac{\alpha(\lambda_i - 1) + 1}{\lambda_i^\alpha} \right))</td>
</tr>
<tr>
<td>Beta log-det divergence ((\alpha = 1, \beta \geq 0))</td>
<td>(D_B^{(\beta)}(P\parallel Q) = \frac{1}{\beta} \log \det \frac{(PQ^{-1})^\beta + \beta(PQ^{-1})}{1 + \beta} = \frac{1}{\beta} \sum_{i=1}^{n} \log \frac{\lambda_i^\beta + \beta \lambda_i^{-1}}{1 + \beta})</td>
</tr>
<tr>
<td>Symmetric Jeffrey KL divergence ((\alpha = 1, \beta = 0))</td>
<td>(D_J(P\parallel Q) = \frac{1}{2} \text{tr} (PQ^{-1} + QP^{-1} - 2I) = \frac{1}{2} \sum_{i=1}^{n} \left( \sqrt{\lambda_i} - \frac{1}{\sqrt{\lambda_i}} \right)^2)</td>
</tr>
<tr>
<td>Generalized Hilbert metrics</td>
<td>(D_{\text{CCA}}^{(\gamma_2, \gamma_1)}(P\parallel Q) = \log \frac{M_{\gamma_2}({\lambda_i})}{M_{\gamma_1}({\lambda_i})}, \quad d_H(P\parallel Q) = \log \frac{M_{\infty}({\lambda_i})}{M_{-\infty}({\lambda_i})} = \log \frac{\lambda_{\text{max}}}{\lambda_{\text{min}}})</td>
</tr>
</tbody>
</table>
When $\alpha + \beta = 1$, the AB log-det divergence reduces to the Alpha-log-det divergence \cite{5}:

$$D_{\alpha}^{(\alpha,1-\alpha)}(P\|Q) = D_{\alpha}^{(\alpha)}(P\|Q)$$

$$=\begin{cases} \frac{1}{\alpha(1-\alpha)} \log \det \left( (1-\alpha)P + \alpha Q^{-1} \right) & \\
\frac{1}{\alpha(1-\alpha)} \log \frac{\det \left( \alpha P + (1-\alpha)Q \right)}{\det \left( \alpha Q^{1-\alpha} \right)} & \text{for } 0 < \alpha < 1,
\end{cases}$$

$$=\begin{cases} \frac{1}{\alpha(1-\alpha)} \sum_{i=1}^{n} \log \left( \left( \lambda_i - 1 \right) \right) & \\
\text{tr}(QP^{-1}) - \log \det(QP^{-1}) - n & \text{for } \alpha = 1,
\end{cases}$$

$$=\begin{cases} \frac{1}{\alpha(1-\alpha)} \sum_{i=1}^{n} \log \left( \frac{\alpha(\lambda_i - 1) + 1}{\lambda_i} \right) & \\
\text{tr}(PQ^{-1}) - \log \det(PQ^{-1}) - n & \text{for } \alpha = 0.
\end{cases}$$

On the other hand, when $\alpha = 1$ and $\beta \geq 0$, the AB log-det divergence reduces to the Beta-log-det divergence:

$$D_{\beta}^{(1,\beta)}(P\|Q) = D_{\beta}^{(\beta)}(P\|Q)$$

$$=\begin{cases} \frac{1}{\beta} \log \det \left( \frac{PQ^{-1}}{1+\beta} + \beta (QP^{-1}) \right) = \frac{1}{\beta} \sum_{i=1}^{n} \log \left( \frac{\lambda_i^\beta + \beta \lambda_i^{-1}}{1+\beta} \right) \text{ for } \beta > 0,
\end{cases}$$

$$=\begin{cases} \text{tr}(QP^{-1} - I) - \log \det(QP^{-1}) = \sum_{i=1}^{n} \left( \lambda_i^{-1} + \log(\lambda_i) \right) - n \text{ for } \beta = 0,
\end{cases}$$

$$=\begin{cases} \log \frac{\det(PQ^{-1})}{\det(I + \log(PQ^{-1}))} = \sum_{i=1}^{n} \log \frac{\lambda_i}{1 + \log(\lambda_i)} \text{ for } \beta = -1, \lambda_i > e^{-1} \forall i.
\end{cases}$$

Note that $\det(I + \log(PQ^{-1}) = \prod_{i=1}^{n}[1 + \log(\lambda_i)]$, and the Beta-log-det divergence is well defined for $\beta = -1$ and if all the eigenvalues are larger than $\lambda_i > e^{-1} \approx 0.367 (e \approx 2.72)$.

It is interesting to note that the Beta-log-det divergence for $\beta \to \infty$ leads to a new divergence that is robust with respect to noise. This new divergence is given by

$$\lim_{\beta \to \infty} D_{\beta}^{(\beta)}(P\|Q) = D_{\beta}^{(\infty)}(P\|Q) = \log \left( \prod_{i=1}^{k} \lambda_i \right) \text{ for all } \lambda_i \geq 1.$$  

This can be easily shown by applying the L’Hôpital’s formula. Assuming that the set $\Omega = \{i : \lambda_i > 1\}$ gathers the indices of those eigenvalues greater than one, we can more formally express this divergence as

$$D_{\beta}^{(\infty)}(P\|Q) = \begin{cases} \sum_{i \in \Omega} \log \lambda_i & \text{for } \Omega \neq \phi, \\
0 & \text{for } \Omega = \phi.
\end{cases}$$

The Alpha-log-det divergence gives the standard Stein’s losses (Burg matrix divergences) for $\alpha = 1$ and $\alpha = 0$, and the Beta-log-det divergence is equivalent to Stein’s loss for $\beta = 0$. 

\cite{17}
Another important class of divergences is Power log-det divergences for any \( \alpha = \beta \in \mathbb{R} \):

\[
D_{\text{AB}}^{(\alpha,\alpha)}(P\|Q) = D_{\text{P}}^{(\alpha)}(P\|Q) = \begin{cases}
\frac{1}{\alpha^2} \log \det \left( \frac{(PQ^{-1})^\alpha + (PQ^{-1})^{-\alpha}}{2} \right) = \frac{1}{\alpha^2} \sum_{i=1}^{n} \log \frac{\lambda_i^\alpha + \lambda_i^{-\alpha}}{2} & \text{for } \alpha \neq 0, \\
\frac{1}{2} \text{tr} \log^2(PQ^{-1}) = \frac{1}{2} \text{tr} \log^2(QP^{-1}) = \frac{1}{2} \sum_{i=1}^{n} \log^2(\lambda_i) & \text{for } \alpha = 0.
\end{cases}
\] (28)

4. Properties of the AB Log-Det Divergence

The AB log-det divergence has several important and useful theoretical properties for SPD matrices.

1. Nonnegativity; given by

\[
D_{\text{AB}}^{(\alpha,\beta)}(P\|Q) \geq 0, \quad \forall \alpha, \beta \in \mathbb{R}. \tag{29}
\]

2. Identity of indiscernibles (see Theorems 1 and 2); given by

\[
D_{\text{AB}}^{(\alpha,\beta)}(P\|Q) = 0 \text{ if and only if } P = Q. \tag{30}
\]

3. Continuity and smoothness of \( D_{\text{AB}}^{(\alpha,\beta)}(P\|Q) \) as a function of \( \alpha \in \mathbb{R} \) and \( \beta \in \mathbb{R} \), including the singular cases when \( \alpha = 0 \) or \( \beta = 0 \), and when \( \alpha = -\beta \) (see Figure 2).

4. The divergence can be expressed in terms of the diagonal matrix \( \Lambda = \text{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_n\} \) with the eigenvalues of \( PQ^{-1} \), in the form

\[
D_{\text{AB}}^{(\alpha,\beta)}(P\|Q) = D_{\text{AB}}^{(\alpha,\beta)}(\Lambda\|I). \tag{31}
\]

5. Scaling invariance; given by

\[
D_{\text{AB}}^{(\alpha,\beta)}(cP\|cQ) = D_{\text{AB}}^{(\alpha,\beta)}(P\|Q), \tag{32}
\]

for any \( c > 0 \).

6. Relative invariance for scale transformation: For given \( \alpha \) and \( \beta \) and nonzero scaling factor \( \omega \neq 0 \), we have

\[
D_{\text{AB}}^{(\omega \alpha, \omega \beta)}(P\|Q) = \frac{1}{\omega^2} D_{\text{AB}}^{(\alpha,\beta)}((Q^{-1/2}PQ^{-1/2})^\omega\|I). \tag{33}
\]

7. Dual-invariance under inversion (for \( \omega = -1 \)); given by

\[
D_{\text{AB}}^{(-\alpha,-\beta)}(P\|Q) = D_{\text{AB}}^{(\alpha,\beta)}(P^{-1}\|Q^{-1}). \tag{34}
\]

8. Dual symmetry; given by

\[
D_{\text{AB}}^{(\alpha,\beta)}(P\|Q) = D_{\text{AB}}^{(\beta,\alpha)}(Q\|P). \tag{35}
\]
9. Affine invariance (invariance under congruence transformations); given by

$$D_{AB}^{(\alpha,\beta)}(APA^T\|AQ\!) = D_{AB}^{(\alpha,\beta)}(P\|Q),$$

(36)

for any nonsingular matrix $A \in \mathbb{R}^{n\times n}$

10. Divergence lower-bound; given by

$$D_{AB}^{(\alpha,\beta)}(X^TP\|X^TQ) \leq D_{AB}^{(\alpha,\beta)}(P\|Q),$$

(37)

for any full-column rank matrix $X \in \mathbb{R}^{n\times m}$ with $n \leq m$.

11. Scaling invariance under the Kronecker product; given by

$$D_{AB}^{(\alpha,\beta)}(Z \otimes P\|Z \otimes Q) = n D_{AB}^{(\alpha,\beta)}(P\|Q),$$

(38)

for any symmetric and positive definite matrix $Z$ of rank $n$.

12. Double Sided Orthogonal Procrustes property. Consider an orthogonal matrix $\Omega \in O(n)$ and two symmetric positive definite matrices $P$ and $Q$, with respective eigenvalue matrices $\Lambda_P$ and $\Lambda_Q$ which elements are sorted in descending order. The AB log-det divergence between $\Omega^TP\Omega$ and $Q$ is globally minimized when their eigenspaces are aligned, i.e.,

$$\min_{\Omega \in O(n)} D_{AB}^{(\alpha,\beta)}(\Omega^TP\Omega\|Q) = D_{AB}^{(\alpha,\beta)}(\Lambda_P\|\Lambda_Q).$$

(39)

13. Triangle Inequality-Metric Distance Condition, for $\alpha = \beta \in \mathbb{R}$. The previous property implies the validity of the triangle inequality for arbitrary positive definite matrices, i.e.,

$$\sqrt{D_{AB}^{(\alpha,\alpha)}(P\|Q)} \leq \sqrt{D_{AB}^{(\alpha,\alpha)}(P\|Z)} + \sqrt{D_{AB}^{(\alpha,\alpha)}(Z\|Q)}.$$  

(40)

The proof of this property exploits the metric characterization of the square root of the S-divergence proposed first by S. Sra in [6,17] for arbitrary SPD matrices.

Several of these properties have been already proved for the specific cases of $\alpha$ and $\beta$ that lead to the S-divergence ($\alpha, \beta = 1/2$) [6], the Alpha log-det divergence ($0 \leq \alpha \leq 1, \beta = 1 - \alpha$) [5] and the Riemannian metric ($\alpha, \beta = 0$) [28, Chapter 6]. We refer the reader to Appendix F for their proofs when $\alpha, \beta \in \mathbb{R}$.

5. Symmetrized AB Log-Det Divergences

The basic AB log-det divergence is asymmetric; that is, $D_{AB}^{(\alpha,\beta)}(P\|Q) \neq D_{AB}^{(\alpha,\beta)}(Q\|P)$, except the spacial case of $\alpha = \beta$.

In general, there are several ways to symmetrize a divergence; for example, Type-1,

$$D_{AB^{(1)}}^{(\alpha,\beta)}(P\|Q) = \frac{1}{2} \left[ D_{AB}^{(\alpha,\beta)}(P\|Q) + D_{AB}^{(\alpha,\beta)}(Q\|P) \right],$$

(41)
and Type-2, based on the Jensen-Shannon symmetrization (which is too complex for log-det divergences),

\[ D^{(\alpha,\beta)}_{\text{ABS2}}(P \parallel Q) = \frac{1}{2} \left[ D^{(\alpha,\beta)}_{\text{AB}} \left( P \parallel \frac{P+Q}{2} \right) + D^{(\alpha,\beta)}_{\text{AB}} \left( Q \parallel \frac{P+Q}{2} \right) \right] . \]  

(42)

The Type-1 symmetric AB log-det divergence is defined as

\[
D^{(\alpha,\beta)}_{\text{ABS1}}(P \parallel Q) = \begin{cases} 
\frac{1}{2\alpha\beta} \log \det \left[ I + \frac{\alpha\beta}{(\alpha + \beta)^2} ((PQ^{-1})^{\alpha+\beta} + (QP^{-1})^{\alpha+\beta} - 2I) \right] & \text{for } \alpha\beta > 0, \\
\frac{1}{2\alpha^2} \left[ \text{tr} \left( (PQ^{-1})^\alpha + (QP^{-1})^\alpha - 2I \right) \right] & \text{for } \alpha \neq 0, \beta = 0, \\
\frac{1}{2\beta^2} \left[ \text{tr} \left( (PQ^{-1})^\beta + (QP^{-1})^\beta - 2I \right) \right] & \text{for } \alpha = 0, \beta \neq 0, \\
\frac{1}{2} \text{tr} \log(I - \log^2(PQ^{-1})^\alpha)^{-1} & \text{for } \alpha = -\beta \neq 0, \\
\frac{1}{2} \text{tr} \log^2(PQ^{-1}) = \frac{1}{2} \| \log(Q^{-1/2}PQ^{-1/2}) \|_F^2 & \text{for } \alpha, \beta = 0. 
\end{cases}
\]

(43)

Equivalently, this can be expressed by the eigenvalues of \( PQ^{-1} \) in the form

\[
D^{(\alpha,\beta)}_{\text{ABS1}}(P \parallel Q) = \begin{cases} 
\frac{1}{2\alpha\beta} \sum_{i=1}^{n} \log \left( 1 + \frac{\alpha\beta}{(\alpha + \beta)^2} (\lambda_i^{\alpha+\beta} - \lambda_i^{-\alpha+\beta})^2 \right) & \text{for } \alpha\beta > 0, \\
\frac{1}{2\alpha^2} \sum_{i=1}^{n} (\lambda_i^\alpha + \lambda_i^{-\alpha} - 2) = \frac{1}{2\alpha^2} \sum_{i=1}^{n} (\lambda_i^\alpha - \lambda_i^{-\alpha})^2 & \text{for } \alpha \neq 0, \beta = 0, \\
\frac{1}{2\beta^2} \sum_{i=1}^{n} (\lambda_i^\beta + \lambda_i^{-\beta} - 2) = \frac{1}{2\beta^2} \sum_{i=1}^{n} (\lambda_i^\beta - \lambda_i^{-\beta})^2 & \text{for } \alpha = 0, \beta \neq 0, \\
\frac{1}{2\alpha^2} \sum_{i=1}^{n} \log \frac{1}{1 - \log^2(\lambda_i^\alpha)} & \text{for } \alpha = -\beta \neq 0, \\
\frac{1}{2} \sum_{i=1}^{n} \log^2(\lambda_i) & \text{for } \alpha, \beta = 0. 
\end{cases}
\]

(44)

We consider several well-known symmetric log-det divergences (see Figure 4); in particular, we consider the following:

1. For \( \alpha = \beta = \pm 0.5 \), we obtain the S-divergence or JBLD divergence (20).
2. For \( \alpha = \beta = 0 \), we obtain the square of the AIRM (Riemannian metric) (19).
For $\alpha = 0$ and $\beta = \pm 1$ or for $\beta = 0$ and $\alpha = \pm 1$, we obtain the KLDM (symmetrized KL Density Metric), also known as the symmetric Stein’s loss or Jeffreys KL divergence [3]:

$$D_J(P\|Q) = \frac{1}{2} \text{tr} \left( PQ^{-1} + QP^{-1} - 2I \right)$$

$$= \frac{1}{2} \sum_{i=1}^{n} \left( \sqrt{\lambda_i} - \frac{1}{\sqrt{\lambda_i}} \right)^2. \quad (45)$$

$$\begin{align*}
\text{(Generalized KLDM or}\ & \text{Jeffreys KL divergence)} \\
\frac{1}{2\alpha^2} [\text{tr}(PQ^{-1})^\alpha + (PQ^{-1})^\alpha - 2I] \\
\text{(S-divergence)} \\
\text{(Squared AIRM)} \\
\text{(Jeffreys KL divergence)} \\
\frac{1}{2\beta^2} [\text{tr}(PQ^{-1})^\beta + (PQ^{-1})^\beta - 2I]
\end{align*}$$

**Figure 4.** Graphical illustration of the fundamental symmetric AB log-det divergences. On the $(\alpha, \beta)$-plane, the special cases of particular divergences are indicated by points (Jeffreys KL divergence (KLDM) or symmetric Stein’s loss and its generalization, S-divergence (JBLD), and the Power log-det divergence.

One important potential application of the AB log-det divergence is to generate conditionally positive definite kernels, which are widely applied to classification and clustering. For a specific set of parameters, the AB log-det divergence gives rise to a Hilbert space embedding in the form of a Radial Basis Function (RBF) kernel [22]; more specifically, the AB log-det kernel is defined by

$$K_{AB}^{(\alpha, \beta)}(P\|Q) = \exp \left( -\gamma D_{\text{ABS1}}^{(\alpha, \beta)}(P\|Q) \right)$$

$$= \left( \det \left[ I + \frac{\alpha \beta}{(\alpha + \beta)^2} (PQ^{-1})^{\alpha + \beta} + (Q^{-1})^{\alpha + \beta} - 2I \right] \right)^{-\frac{\gamma}{2\alpha\beta}} \quad (46)$$
for some selected values of $\gamma > 0$ and $\alpha, \beta > 0$ or $\alpha, \beta < 0$ that can make the kernel positive definite.

6. Similarity Measures for Semidefinite Covariance Matrices in Reproducing Kernel Hilbert Spaces

There are many practical applications for which the underlying covariance matrices are symmetric but only positive semidefinite, i.e., their columns do not span the whole space. For instance, in classification problems, assume two classes and a set of observation vectors $\{x_1, \ldots, x_T\}$ and $\{y_1, \ldots, y_T\}$ in $\mathbb{R}^m$ for each class, then we may wish to find a principled way to evaluate the ensemble similarity of the data from their sample similarity. The problem of the modeling of similarity between two ensembles was studied by Zhou and Chellappa in [32]. For this purpose, they proposed several probabilistic divergence measures between positive semidefinite covariance matrices in a Reproducing kernel Hilbert space (RKHS) of finite dimensionality. Their strategy was later extended for image classification problems [33] and formalized for the Log-Hilbert-Schmidt metric between infinite-dimensional RKHS covariance operators [34].

In this section, we propose the unifying framework of the AB log-det divergences to reinterpret and extend the similarity measures obtained in [32,33] for semidefinite covariance matrices in the finite-dimensional RKHS.

We shall assume that the nonlinear functions $\Phi_x : \mathbb{R}^m \to \mathbb{R}^n$ and $\Phi_y : \mathbb{R}^m \to \mathbb{R}^n$ (where $n > m$) respectively map the data from each of the classes into their higher dimensional feature spaces. We implicitly define the feature matrices as

$$\Phi_x = [\Phi_x(x_1), \ldots, \Phi_x(x_T)], \quad \Phi_y = [\Phi_y(y_1), \ldots, \Phi_y(y_T)], \quad (47)$$

and the sample covariance matrices of the observations in the feature space as: $C_x = \Phi_x J \Phi_x^T / T \in \mathbb{R}^{n \times n}$ and $C_y = \Phi_y J \Phi_y^T / T \in \mathbb{R}^{n \times n}$, where $J = I_T - \frac{1}{T} 11^T$ denotes the $T \times T$ centering matrix.

In practice, it is common to consider low-rank approximations of sample covariance matrices. For a given basis $V_x = (v_1, \ldots, v_r) \in \mathbb{R}^{T \times r}$ of the principal subspace of $J \Phi_x^T \Phi_x J$, we can define the projection matrix $\Pi_x = V_x V_x^T$ and redefine the covariance matrices as

$$C_x = \frac{1}{T} \Phi_x V_x^T \Phi_x^T \quad \text{and} \quad C_y = \frac{1}{T} \Phi_y V_y^T \Phi_y^T. \quad (48)$$

Assuming the Gaussianity of the data in the feature space, the mean vector and covariance matrix are sufficient statistics and a natural measure of dissimilarity between $\Phi_x$ and $\Phi_y$ should be a function of the first and second order statistics of the features. Furthermore, in most practical problems the mean value should be ignored due to robustness considerations, and then the comparison reduces to the evaluation of a suitable dissimilarity measure between $C_x$ and $C_y$.

The dimensionality of the feature space $n$ is typically much larger than $r$, so the rank of the covariance matrices in (48) will be $r \ll n$ and, therefore, both matrices are positive semidefinite. The AB log-det divergence is infinite when the range spaces of the covariance matrices $C_x$ and $C_y$ differ. This property is useful in applications which require an automatic constraint in the range of the estimates [22], but it will prohibit the practical use of the comparison when the ranges of the covariance matrices differ. The next subsections present two different strategies to address this challenging problem.
6.1. Measuring the Dissimilarity with a Divergence Lower-Bound

One possible strategy is to use dissimilarity measures which ignore the contribution to the divergence caused by the rank deficiency of the covariance matrices. This is useful when performing one comparison of the covariances matrices after applying a congruence transformation that aligns their range spaces, and can be implemented by retaining only the finite and non-zero eigenvalues of the matrix pencil \((C_x, C_y)\).

Let \(I_r\) denote the identity matrix of size \(r\) and \((\cdot)^+\) the Moore-Penrose pseudoinverse operator. Consider the eigenvalue decomposition of the symmetric matrix

\[
(C_y^+)\frac{1}{2}C_x(C_y^+)\frac{1}{2} = U\Lambda U^T
\]

where \(U\) is a semi-orthogonal matrix for which the columns are the eigenvectors associated with the positive eigenvalues of the matrix pencil and

\[
\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_r) \equiv \text{diag} \ E_{ig+} \{(C_y^+)\frac{1}{2}C_x(C_y^+)\frac{1}{2}\}.
\]

is a diagonal matrix with the eigenvalues sorted in a descending order.

Note that the tall matrix \(W = (C_y^+)\frac{1}{2}U \in \mathbb{R}^{n \times r}\) diagonalizes the covariance matrices of the two classes

\[
W^T C_x W = \Lambda,
\]

\[
W^T C_y W = I_r
\]

and compress them to a common range space. The compression automatically discards the singular and infinite eigenvalues of the matrix pencil \((C_x, C_y)\), while it retains the finite and positive eigenvalues. In this way, the following dissimilarity measures can be obtained:

\[
L_{AB}^{(\alpha, \beta)}(C_x, C_y) \equiv D_{AB}^{(\alpha, \beta)}(W^T C_x W \| W^T C_y W) = D_{AB}^{(\alpha, \beta)}(\Lambda \| I_r),
\]

\[
L_{ABS1}^{(\alpha, \beta)}(C_x, C_y) \equiv D_{ABS1}^{(\alpha, \beta)}(W^T C_x W \| W^T C_y W) = D_{ABS1}^{(\alpha, \beta)}(\Lambda \| I_r).
\]

Note, however, that these measures should not be understood as a strict comparison of the original covariance matrices, but rather as an indirect comparison through their respective compressed versions \(W^T C_x W\) and \(W^T C_y W\).

With the help of the kernel trick, the next lemma shows that the evaluation of the dissimilarity measures \(L_{AB}^{(\alpha, \beta)}(C_x, C_y)\) and \(L_{ABS1}^{(\alpha, \beta)}(C_x, C_y)\), does not require the explicit computation of the covariance matrices or of the feature vectors.

Lemma 1. Given the Gram matrix or kernel matrix of the input vectors

\[
\begin{pmatrix}
K_{xx} & K_{xy} \\
K_{yx} & K_{yy}
\end{pmatrix}
= \begin{pmatrix}
\Phi_x^T \Phi_x & \Phi_x^T \Phi_y \\
\Phi_y^T \Phi_x & \Phi_y^T \Phi_y
\end{pmatrix}
\]

and the matrices \(V_x\) and \(V_y\) which respectively span the principal subspaces of \(K_{xx}\) and \(K_{yy}\), the positive and finite eigenvalues of the matrix pencil can be expressed by

\[
\Lambda = \text{diag} \ E_{ig+} \{(V_x^T K_{xy} K_{yy}^{-1} V_y)(V_x^T K_{xy} K_{yy}^{-1} V_y)^T\}.
\]
Proof. The proof of the lemma relies on the property that for any pair of \( m \times n \) matrices \( A \) and \( B \), the non-zero eigenvalues of \( AB^T \) and of \( B^TA \) are the same (see [30, pag. 11]). Then, there is an equality between the following matrices of positive eigenvalues

\[
\Lambda \equiv \text{diag} \, E_{+} \left\{ (C_y^+) \frac{1}{2} C_x (C_y^+) \frac{1}{2} \right\} = \text{diag} \, E_{+} \left\{ C_x C_y^+ \right\}.
\]

(57)

Taking into account the structure of the covariance matrices in (48), such eigenvalues can be explicitly obtained in terms of the kernel matrices

\[
E_{+} \left\{ C_x C_y^+ \right\} = E_{+} \left\{ (\Phi_x V_x V^T_x \Phi_x) + (\Phi_y V_y V^T_y + \Phi_y) \right\} = E_{+} \left\{ (V_x^T (K_x - 1) (K_{xy} y y)^{-1} V_y) (V_x^T K_{xy} K_{yy} y y V^T_y) \right\}.
\]

(58)

Then, the dissimilarity measure of the data in the feature space can be obtained just by measuring a divergence between the SPD matrices \( \tilde{C}_x \) and \( \tilde{C}_y \). Again, the idea is to compute the value of the divergence without requiring the evaluation of the feature vectors but by using the available kernels.

Using the properties of the trace and the determinants, a practical formula for the log-det Alpha-divergence has been obtained in [32,33] for \( 0 < \alpha < 1 \). The resulting expression

\[
D_{AB}^{(\alpha, 1-\alpha)}(\tilde{C}_x || \tilde{C}_y) = \frac{1}{\alpha (1-\alpha)} \log \det (H_2 + \rho^{-1} H) - \frac{1}{1-\alpha} \log \det (\rho^{-1} \Lambda_x) - \frac{1}{\alpha} \log \det (\rho^{-1} \Lambda_y)
\]

is a function of the principal eigenvalues of the kernels

\[
\Lambda_x = V_x^T K_{xx} V_x, \quad \Lambda_y = V_y^T K_{yy} V_y,
\]

and the matrix

\[
H = \begin{pmatrix}
(\alpha \frac{1}{2} W_x & 0 \\
0 & (1-\alpha \frac{1}{2} W_y
\end{pmatrix}
\begin{pmatrix}
K_{xx} & K_{xy} \\
K_{yx} & K_{yy}
\end{pmatrix}
\begin{pmatrix}
(\alpha \frac{1}{2} W_x & 0 \\
0 & (1-\alpha \frac{1}{2} W_y
\end{pmatrix}
\]
where
\[
W_x = V_x (I_r - \rho \Lambda_x^{-1})^{\frac{1}{2}} \quad \text{and} \quad W_y = V_y (I_r - \rho \Lambda_y^{-1})^{\frac{1}{2}}.
\] (66)

The evaluation of the divergence outside the interval \(0 < \alpha < 1\), or when \(\beta \neq 1 - \alpha\), is not covered by this formula and, in general, requires knowledge of the eigenvalues of the matrix \(C_y^{-\frac{1}{2}} C_x C_y^{-\frac{1}{2}}\). However, different analyses are necessary depending on the dimension of the intersection of the range space of both covariance matrices \(C_x\) and \(C_y\). In the following, we study the two more general scenarios.

**Case (A)** The range spaces of \(C_x\) and \(C_y\) are the same.

In this case \(U_y^\perp (U_y^\perp)^T = U_x^\perp (U_x^\perp)^T\), and the eigenvalues of the matrix
\[
\hat{C}_x \hat{C}_y^{-1} = (C_x + \rho \, U_x^\perp (U_x^\perp)^T) (C_y^+ + \rho^{-1} U_x^\perp (U_x^\perp)^T)
\] (67)
\[
= C_x C_y^+ + U_x^\perp (U_x^\perp)^T
\] (68)
coincide with the nonzero eigenvalues of \(C_x C_y^+\) except for \((n - r)\) additional eigenvalues which are equal to 1. Then, using the equivalence between (57) and (60), the divergence reduces to the following form
\[
D_{AB}^{(\alpha,\beta)} (\hat{C}_x \| \hat{C}_y) = L_{AB}^{(\alpha,\beta)} (C_x, C_y)
\] (69)
\[
= D_{AB}^{(\alpha,\beta)} \left( (V_x^T K_{xy} K_{yy}^{-1} V_y) (V_x^T K_{xy} K_{yy}^{-1} V_y)^T \parallel I_r \right).
\] (70)

**Case (B)** The range spaces of \(C_x\) and \(C_y\) are disjoint.

In practice, for \(n \gg r\) this is the most probable scenario. In such a case, the \(r\) largest eigenvalues of the matrix \(\hat{C}_x \hat{C}_y^{-1}\) diverge as \(\rho\) tends to zero. Hence, we can not bound above these eigenvalues and, for this reason, it makes no sense to study the case of \(\text{sign}(\alpha) \neq \text{sign}(\beta)\), so in this section we assume that \(\text{sign}(\alpha) = \text{sign}(\beta)\).

**Theorem 3.** When range spaces of \(C_x\) and \(C_y\) are disjoint and for a sufficiently small value of \(\rho > 0\), the AB log-det divergence is closely approximated by the formula
\[
D_{AB}^{(\alpha,\beta)} (\hat{C}_x \parallel \hat{C}_y) \approx D_{AB}^{(\alpha,\beta)} (C_{x|y}^{(\rho)} \parallel \rho I_r) + D_{AB}^{(\beta,\alpha)} (C_{y|x}^{(\rho)} \parallel \rho I_r),
\] (71)
where \(C_{x|y}^{(\rho)}\) (and respectively \(C_{y|x}^{(\rho)}\) by interchanging \(x\) and \(y\)) denotes the matrix
\[
C_{x|y}^{(\rho)} = \Lambda_x - \rho I_r - \rho^2 \Lambda_y^{-1} - W_x^T K_{xy} W_y \Lambda_y^{-1} W_y^T K_{yx} W_x.
\] (72)

The proof of the theorem is presented in the Appendix G. The eigenvalues of the matrices \(C_{x|y}^{(\rho)}\) and \(C_{y|x}^{(\rho)}\), estimate the \(r\) largest eigenvalues of \(C_y^{-\frac{1}{2}} C_x C_y^{-\frac{1}{2}}\) and of its inverse \(C_x^{-\frac{1}{2}} C_y^{-\frac{1}{2}}\), respectively. The relative error in the estimation of these eigenvalues is of order \(O(\rho)\), i.e., it gradually improves as \(\rho\)
tend to zero. The approximation is asymptotically exact, and $C_{x|y}^{(p)}$ and $C_{x|y}^{(p)}$ converge respectively to the conditional covariance matrices

$$C_{x|y} = \lim_{\rho \to 0} C_{x|y}^{(p)} = V_x^T K_{xx} V_x - (V_x^T K_{xy} V_y)(V_y^T K_{yy} V_y)^{-1}(V_y^T K_{xy} V_x)^T, \quad (73)$$

$$C_{y|x} = \lim_{\rho \to 0} C_{y|x}^{(p)} = V_y K_{yy} V_y - (V_y^T K_{yx} V_x)(V_x^T K_{xx} V_x)^{-1}(V_x^T K_{yx} V_y)^T, \quad (74)$$

while $\rho$ converges to the zero matrix.

In the limit, the value of the divergence is not very useful because

$$\lim_{\rho \to 0} D_{AB}^{(\alpha,\beta)}(\tilde{C}_x || \tilde{C}_y) = \infty, \quad (75)$$

though there are some practical ways to circumvent this limitation. For example, when $\alpha = 0$ or $\beta = 0$, the divergence can be scaled by a suitable power of $\rho$ to make it finite (see Section 3.3.1 in [32]). The scaled form of the divergence between the regularized covariance matrices is

$$SD_{AB}^{(\alpha,\beta)}(\tilde{C}_x || \tilde{C}_y) = \lim_{\rho \to 0} \rho^{\max{\{\alpha,\beta\}}} D_{AB}^{(\alpha,\beta)}(\tilde{C}_x || \tilde{C}_y). \quad (76)$$

Examples of scaled divergences are the following versions of Stein’s losses

$$SD_{AB}^{(0,\beta)}(\tilde{C}_x || \tilde{C}_y) = \lim_{\rho \to 0} \rho^\beta D_{AB}^{(0,\beta)}(\tilde{C}_x || \tilde{C}_y) = \frac{1}{\beta^2} \text{tr}((C_{x|y})^\beta) \geq 0, \quad \beta > 0, \quad (77)$$

$$SD_{AB}^{(\alpha,0)}(\tilde{C}_x || \tilde{C}_y) = \lim_{\rho \to 0} \rho^\alpha D_{AB}^{(\alpha,0)}(\tilde{C}_x || \tilde{C}_y) = \frac{1}{\alpha^2} \text{tr}((C_{y|x})^\alpha) \geq 0, \quad \alpha > 0, \quad (78)$$

as well as the Jeffrey’s KL family of symmetric divergences (cf. Equation (23) in [33])

$$SD_{AB}^{(0,0)}(\tilde{C}_x || \tilde{C}_y) = \lim_{\rho \to 0} \rho^\alpha D_{AB}^{(0,0)}(\tilde{C}_x || \tilde{C}_y) = \frac{1}{2\alpha^2} \left(\text{tr}((C_{x|y})^\alpha) + \text{tr}((C_{y|x})^\alpha)\right), \quad \alpha > 0. \quad (79)$$

In other cases, when the scaling is not sufficient to obtain a finite and practical dissimilarity measure, an affine transformation may be used. The idea is to identify the divergent part of $D_{AB}^{(\alpha,\beta)}(\tilde{C}_x || \tilde{C}_y)$ as $\rho \to 0$ and use its value as a reference for the evaluation the dissimilarity. For $\alpha, \beta \geq 0$, the relative AB log-det dissimilarity measure is the limiting value of the affine transformation

$$RD_{AB}^{(\alpha,\beta)}(\tilde{C}_x || \tilde{C}_y) \equiv \lim_{\rho \to 0} \min\{\alpha, \beta\} \left(D_{AB}^{(\alpha,\beta)}(\tilde{C}_x || \tilde{C}_y) - \frac{r}{\alpha\beta} \log \frac{\alpha\beta \rho^{-(\alpha+\beta)}}{\alpha + \beta})^2\right), \quad \alpha, \beta > 0. \quad (80)$$

After its extension by continuity (including as special cases $\alpha = 0$ or $\beta = 0$), the function

$$RD_{AB}^{(\alpha,\beta)}(\tilde{C}_x || \tilde{C}_y) = \left\{ \begin{array}{ll}
\log \text{det}(C_{x|y}) + \frac{\alpha}{\beta} \log \text{det}(C_{y|x}) & \beta > \alpha \geq 0 \\
\log \text{det}(C_{x|y}) + \log \text{det}(C_{y|x}) & \alpha = \beta \geq 0 \\
\log \text{det}(C_{y|x}) + \frac{\beta}{\alpha} \log \text{det}(C_{x|y}) & \alpha > \beta \geq 0 
\end{array} \right. \quad (81)$$

provides simple formulas to measure the relative dissimilarity between symmetric positive semidefinite matrices $C_x$ and $C_y$. However, it should be taken into account that, as a consequence of its relative character, this function is not bounded below and can achieve negative values.
7. Modifications and Generalizations of AB Log-Det Divergences and Gamma Matrix Divergences

The divergence (3) discussed in the previous sections can be extended and modified in several ways. It is interesting to note that the positive eigenvalues of $PQ^{-1}$ play a similar role as the ratios $(p_i/q_i)$ and $(q_i/p_i)$ when used in the wide class of standard discrete divergences, see for example, [11,12]; hence, we can apply such divergences to formulate a modified log-det divergence as a function of the eigenvalues $\lambda_i$.

For example, consider the Itakura-Saito distance defined by

$$D_{IS}(p \parallel q) = \sum_i \left( \frac{p_i}{q_i} + \log \frac{q_i}{p_i} - 1 \right).$$

It is worth noting that we can generate the large class of divergences or cost functions using Csiszár $f$-functions [13,24,25]. By replacing $p_i/q_i$ with $\lambda_i$ and $q_i/p_i$ with $\lambda_i^{-1}$, we obtain the log-det divergence for SPD matrices:

$$D_{IS}(P \parallel Q) = \sum_{i=1}^n \left( \lambda_i - \log(\lambda_i) \right) - n,$$

which is consistent with (24) and (26).

As another example, consider the discrete Gamma divergence [11,12] defined by

$$D_{AC}^{(\alpha,\beta)}(p \parallel q) = \frac{1}{\beta(\alpha + \beta)} \log \left( \sum_i p_i^{\alpha+\beta} \right) + \frac{1}{\alpha(\alpha + \beta)} \log \left( \sum_i q_i^{\alpha+\beta} \right) - \frac{1}{\alpha\beta} \ln \left( \sum_i p_i^\alpha q_i^\beta \right),$$

for $\alpha \neq 0$, $\beta \neq 0$, $\alpha + \beta \neq 0$,

which when $\alpha = 1$ and $\beta \to -1$, simplifies to the following form [11]:

$$\lim_{\beta \to -1} D_{AC}^{(1,\beta)}(p \parallel q) = \frac{1}{n} \sum_{i=1}^n \left( \log \frac{q_i}{p_i} \right) + \log \left( \sum_{i=1}^n \frac{p_i}{q_i} \right) - \log(n) = \log \left( \frac{\prod_{i=1}^n p_i}{\prod_{i=1}^n q_i} \right)^{1/n}. \quad (85)$$

Hence, by substituting $p_i/q_i$ with $\lambda_i$, we derive a new Gamma matrix divergence for SPD matrices:

$$D_{CCA}^{(1,0)}(P \parallel Q) = D_{AC}^{(1,-1)}(P \parallel Q) = \frac{1}{n} \sum_{i=1}^n \left( \log \frac{\lambda_i^{-1}}{\lambda_i} \right) + \log \left( \sum_{i=1}^n \lambda_i \right) - \log(n)$$

$$= \log \left( \frac{\prod_{i=1}^n \lambda_i}{\prod_{i=1}^n \lambda_i} \right)^{1/n} = \log \frac{M_1\{\lambda_i\}}{M_0\{\lambda_i\}}, \quad (86)$$
where \( M_1 \) denotes the arithmetic mean, and \( M_0 \) denotes the geometric mean.

Interestingly, (86) can be expressed equivalently as

\[
D_{\text{CCA}}^{(1,0)}(P \parallel Q) = \log(\text{tr}(PQ^{-1})) - \frac{1}{n} \log \det(PQ^{-1}) - \log(n). \tag{87}
\]

Similarly, using the symmetric Gamma divergence defined in [11,12],

\[
D_{\text{ACS}}^{(\alpha,\beta)}(P\parallel Q) = \frac{1}{\alpha \beta} \log \left( \frac{\sum_i p_i^{\alpha+\beta} q_i^{\alpha+\beta}}{\left( \sum_i p_i^\alpha q_i^\beta \right) \left( \sum_i p_i^\beta q_i^\alpha \right)} \right), \tag{88}
\]

for \( \alpha \neq 0, \beta \neq 0, \alpha + \beta \neq 0 \),

for \( \alpha = 1 \) and \( \beta \to -1 \) and by substituting the ratios \( p_i/q_i \) with \( \lambda_i \), we obtain a new Gamma matrix divergence as follows:

\[
D_{\text{ACS}}^{(1,-1)}(P \parallel Q) = \log \left( \frac{\sum_i \lambda_i}{\left( \frac{1}{n} \sum_i \lambda_i \right)^{-1}} \right) - \log(n)^2
= \log \left( \frac{1}{n} \sum_i \lambda_i \left( \frac{1}{n} \sum_i \lambda_i^{-1} \right) \right)
= \log \left( M_1 \{\lambda_i\} M_1 \{\lambda_i^{-1}\} \right)
= \log \frac{M_1 \{\lambda_i\}}{M_{-1} \{\lambda_i\}}, \tag{89}
\]

where \( M_{-1} \{\lambda_i\} \) denotes the harmonic mean.

Note that for \( n \to \infty \), this formulated divergence can be expressed compactly as

\[
D_{\text{ACS}}^{(1,-1)}(P \parallel Q) = \log(E\{u\} E\{u^{-1}\}), \tag{91}
\]

where \( u_i = \{\lambda_i\} \) and \( u_i^{-1} = \{\lambda_i^{-1}\} \).

The basic means are defined as follows:

\[
M_\gamma(\lambda) = \begin{cases} 
M_{-\infty} = \min\{\lambda_1, \ldots, \lambda_n\}, & \gamma \to -\infty, \\
M_{-1} = n \left( \sum_{i=1}^n \frac{1}{\lambda_i} \right)^{-1}, & \gamma = -1, \\
M_0 = \left( \prod_{i=1}^n \lambda_i \right)^{1/n}, & \gamma = 0, \\
M_1 = \frac{1}{n} \sum_{i=1}^n \lambda_i, & \gamma = 1, \\
M_2 = \left( \frac{1}{n} \sum_{i=1}^n \lambda_i^2 \right)^{1/2}, & \gamma = 2, \\
M_{\infty} = \max\{\lambda_1, \ldots, \lambda_n\}, & \gamma \to \infty,
\end{cases} \tag{92}
\]

with

\[
M_{-\infty} \leq M_{-1} \leq M_0 \leq M_1 \leq M_2 \leq M_{\infty}, \tag{93}
\]
where equality holds only if all $\lambda_i$ are equal. By increasing the values of $\gamma$, more emphasis is put on large relative errors, i.e., on $\lambda_i$ whose values are far from one. Depending on the value of $\gamma$, we obtain the minimum entry of the vector $\lambda$ (for $\gamma \to -\infty$), its harmonic mean ($\gamma = -1$), the geometric mean ($\gamma = 0$), the arithmetic mean ($\gamma = 1$), the quadratic mean ($\gamma = 2$), and the maximum entry of the vector ($\gamma \to \infty$).

Exploiting the above inequalities for the means, the divergences in (86) and (90) can be heuristically generalized (defined) as follows:

$$D_{CCA}^{(\gamma_2, \gamma_1)}(P \parallel Q) = \log \frac{M_{\gamma_2}\{\lambda_i\}}{M_{\gamma_1}\{\lambda_i\}},$$  
(94)

for $\gamma_2 > \gamma_1$.

The new divergence in (94) is quite general and flexible, and in extreme cases, it takes the following form:

$$D_{CCA}^{(\infty, -\infty)}(P \parallel Q) = d_H(P \parallel Q) = \log \frac{M_{\infty}\{\lambda_i\}}{M_{-\infty}\{\lambda_i\}} = \log \frac{\lambda_{\max}}{\lambda_{\min}},$$  
(95)

which is, in fact, a well-known Hilbert projective metric [6,26].

The Hilbert projective metric is extremely simple and suitable for big data because it requires only two (minimum and maximum) eigenvalue computations of the matrix $PQ^{-1}$.

The Hilbert projective metric satisfies the following important properties [6,27]:

1. Nonnegativity, $d_H(P \parallel Q) \geq 0$, and definiteness, $d_H(P \parallel Q) = 0$, if and only if there exists a $c > 0$ such that $Q = cP$.
2. Invariance to scaling:

$$d_H(c_1P \parallel c_2Q) = d_H(P \parallel Q),$$  
(96)

for any $c_1, c_2 > 0$.
3. Symmetry:

$$d_H(P \parallel Q) = d_H(Q \parallel P).$$  
(97)

4. Invariance under inversion:

$$d_H(P \parallel Q) = d_H(P^{-1} \parallel Q^{-1}).$$  
(98)

5. Invariance under congruence transformations:

$$d_H(APA^T \parallel AQA^T) = d_H(P \parallel Q),$$  
(99)

for any invertible matrix $A$.
6. Invariance under geodesic (Riemannian) transformations (by taking $A = P^{-1/2}$ in (99)):

$$d_H(I \parallel P^{-1/2}QP^{-1/2}) = d_H(P \parallel Q).$$  
(100)

7. Separability of divergence for the Kronecker product of SPD matrices:

$$d_H(P_1 \otimes P_2 \parallel Q_1 \otimes Q_2) = d_H(P_1 \parallel Q_1) + d_H(P_2 \parallel Q_2).$$  
(101)
8. Scaling of power of SPD matrices:

\[ d_H(P^\omega \parallel Q^\omega) = |\omega| \ d_H(P \parallel Q), \]  

(102)

for any \( \omega \neq 0 \).

Hence, for \( 0 < |\omega_1| \leq 1 \leq |\omega_2| \) we have

\[ d_H(P^{\omega_1} \parallel Q^{\omega_1}) \leq d_H(P \parallel Q) \leq d_H(P^{\omega_2} \parallel Q^{\omega_2}). \]  

(103)

9. Scaling under the weighted geometric mean:

\[ d_H(P^{#s}Q \parallel P^{#u}Q) = |s-u| \ d_H(P \parallel Q), \]  

(104)

for any \( u, s \neq 0 \), where

\[ P^{#u}Q = P^{1/2}(P^{-1/2}QP^{-1/2})^u P^{1/2}. \]  

(105)

10. Triangular inequality:

\[ d_H(P \parallel Q) \leq d_H(P \parallel Z) + d_H(Z \parallel Q). \]

These properties can easily be derived and verified. For example, property (9) can easily be derived as follows [6,27]:

\[
\begin{align*}
    d_H(P^{#s}Q \parallel P^{#u}Q) &= d_H(\frac{P^{1/2}(P^{-1/2}QP^{-1/2})^s P^{1/2}}{(P^{-1/2}QP^{-1/2})^u} \parallel \frac{P^{1/2}(P^{-1/2}QP^{-1/2})^u P^{1/2}}{(P^{-1/2}QP^{-1/2})^u}) \\
    &= d_H((P^{-1/2}QP^{-1/2})^s \parallel (P^{-1/2}QP^{-1/2})^u) \\
    &= d_H((P^{-1/2}QP^{-1/2})^s \parallel I) \\
    &= |s-u| \ d_H(P \parallel Q). 
\end{align*}
\]

(106)

In Table 2, we summarize and compare some fundamental properties of three important metric distances: the Hilbert projective metric, Riemannian metric, and LogDet Zero (Bhattacharyya) distance. Since some of these properties are new, we refer to [6,27,28].

7.1. The AB Log-Det Divergence for Noisy and Ill-Conditioned Covariance Matrices

In real-world signal processing and machine learning applications, the SPD sampled matrices can be strongly corrupted by noise and extremely ill conditioned. In such cases, the eigenvalues of the generalized eigenvalue (GEVD) problem \( P v_i = \lambda_i Q v_i \) can be divided into a signal subspace and noise subspace. The signal subspace is usually represented by the largest eigenvalues (and corresponding eigenvectors), and the noise subspace is usually represented by the smallest eigenvalues (and corresponding eigenvectors), which should be rejected; in other words, in the evaluation of log-det divergences, only the eigenvalues that represent the signal subspace should be taken into account. The simplest approach is to find the truncated dominant eigenvalues by applying the suitable threshold \( \tau > 0 \); equivalently, find an index \( r \leq n \) for which \( \lambda_{r+1} \leq \tau \) and perform a summation. For example, truncation reduces the summation in (8) from 1 to \( r \) (instead of 1 to \( n \)) [22]. The threshold parameter \( \tau \) can be selected via cross validation.
Table 2. Comparison of the fundamental properties of three basic metric distances: the Riemannian (geodesic) metric (19), LogDet Zero (Bhattacharyya) divergence (21), and the Hilbert projective metric (95). Matrices \( P, Q, P_1, P_2, Q_1, Q_2, Z \in \mathbb{R}^{n \times n} \) are SPD matrices, \( A \in \mathbb{R}^{n \times n} \) is nonsingular, and the matrix \( X \in \mathbb{R}^{n \times r} \) with \( r < n \) is full column rank. The scalars satisfy the following conditions: \( c > 0, c_1, c_2 > 0; 0 < \omega \leq 1, s, u \neq 0, \psi = |s - u| \). The geometric means are defined by \( P^\parallel u Q = P^{1/2}(P^{-1/2}QP^{-1/2})^u P^{1/2} \) and \( P^\# Q = P^\#_1 Q = P^{1/2}(P^{-1/2}QP^{-1/2})^{1/2} P^{1/2} \). The Hadamard product of \( P \) and \( Q \) is denoted by \( P \circ Q \) (cf. with [6]).

<table>
<thead>
<tr>
<th>Riemannian (geodesic) metric</th>
<th>LogDet Zero (Bhattacharyya) div.</th>
<th>Hilbert projective metric</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d_R(P|Q) = | \log(Q^{-1/2}PQ^{-1/2})|_F )</td>
<td>( d_{bh}(P|Q) = 2\sqrt{\log \det \frac{1}{2}(P + Q) / \sqrt{\det(P) \det(Q)} } )</td>
<td>( d_H(P|Q) = \log \frac{\lambda_{max}{PQ^{-1}}}{\lambda_{min}{PQ^{-1}}} )</td>
</tr>
<tr>
<td>( d_R(P \parallel Q) = d_R(Q \parallel P) )</td>
<td>( d_{bh}(P \parallel Q) = d_{bh}(Q \parallel P) )</td>
<td>( d_H(P \parallel Q) = d_H(Q \parallel P) )</td>
</tr>
<tr>
<td>( d_R(cP \parallel cQ) = d_R(P \parallel Q) )</td>
<td>( d_{bh}(cP \parallel cQ) = d_{bh}(P \parallel Q) )</td>
<td>( d_H(cP \parallel cQ) = d_H(P \parallel Q) )</td>
</tr>
<tr>
<td>( d_R(APA^T \parallel AQA^T) = d_R(P \parallel Q) )</td>
<td>( d_{bh}(APA^T \parallel AQA^T) = d_{bh}(P \parallel Q) )</td>
<td>( d_H(APA^T \parallel AQA^T) = d_H(P \parallel Q) )</td>
</tr>
<tr>
<td>( d_R(P^{-1} \parallel Q^{-1}) = d_R(P \parallel Q) )</td>
<td>( d_{bh}(P^{-1} \parallel Q^{-1}) = d_{bh}(P \parallel Q) )</td>
<td>( d_H(P^{-1} \parallel Q^{-1}) = d_H(P \parallel Q) )</td>
</tr>
<tr>
<td>( d_R(P^{\omega} \parallel Q^{\omega}) \leq \omega \ d_R(P \parallel Q) )</td>
<td>( d_{bh}(P^{\omega} \parallel Q^{\omega}) \leq \sqrt{\omega} \ d_{bh}(P \parallel Q) )</td>
<td>( d_H(P^{\omega} \parallel Q^{\omega}) \leq \omega \ d_H(P \parallel Q) )</td>
</tr>
<tr>
<td>( d_R(P \parallel P^# u Q) = \omega \ d_R(P \parallel Q) )</td>
<td>( d_{bh}(P \parallel P^# u Q) \leq \sqrt{\omega} \ d_{bh}(P \parallel Q) )</td>
<td>( d_H(P \parallel P^# u Q) = \omega \ d_H(P \parallel Q) )</td>
</tr>
<tr>
<td>( d_R(Z#<em>\omega P \parallel Z#</em>\omega Q) \leq \omega \ d_R(P \parallel Q) )</td>
<td>( d_{bh}(Z#<em>\omega P \parallel Z#</em>\omega Q) \leq \sqrt{\omega} \ d_{bh}(P \parallel Q) )</td>
<td>( d_H(Z#<em>\omega P \parallel Z#</em>\omega Q) \leq \omega \ d_H(P \parallel Q) )</td>
</tr>
<tr>
<td>( d_R(P # s Q \parallel P # u Q) = \psi \ d_R(P \parallel Q) )</td>
<td>( d_{bh}(P # s Q \parallel P # u Q) \leq \sqrt{\psi} \ d_{bh}(P \parallel Q) )</td>
<td>( d_H(P # s Q \parallel P # u Q) = \psi \ d_H(P \parallel Q) )</td>
</tr>
<tr>
<td>( d_R(P \parallel P # Q) = d_R(Q \parallel P # Q) )</td>
<td>( d_{bh}(P \parallel P # Q) = d_{bh}(Q \parallel P # Q) )</td>
<td>( d_H(P \parallel P # Q) = d_H(Q \parallel P # Q) )</td>
</tr>
<tr>
<td>( d_R(X^T PX \parallel X^T QX) \leq d_R(P \parallel Q) )</td>
<td>( d_{bh}(X^T PX \parallel X^T QX) \leq d_{bh}(P \parallel Q) )</td>
<td>( d_H(X^T PX \parallel X^T QX) \leq d_H(P \parallel Q) )</td>
</tr>
<tr>
<td>( d_R(Z \otimes P \parallel Z \otimes Q) = \sqrt{n} \ d_R(P \parallel Q) )</td>
<td>( d_{bh}(Z \otimes P \parallel Z \otimes Q) = \sqrt{n} \ d_{bh}(P \parallel Q) )</td>
<td>( d_H(Z \otimes P \parallel Z \otimes Q) = d_H(P \parallel Q) )</td>
</tr>
<tr>
<td>( d_H(P_1 \otimes P_2 \parallel Q_1 \otimes Q_2) = )</td>
<td>( d_{bh}(P_1 \otimes P_2 \parallel Q_1 \otimes Q_2) )</td>
<td>( d_H(P_1 \otimes P_2 \parallel Q_1 \otimes Q_2) )</td>
</tr>
<tr>
<td>( = n \ d_H^2(P_1 \parallel Q_1) + n \ d_H^2(P_2 \parallel Q_2) + \geq d_{bh}(P_1 \circ P_2 \parallel Q_1 \circ Q_2) )</td>
<td>( = d_{bh}(P_1 \parallel Q_1) + d_{bh}(P_2 \parallel Q_2) )</td>
<td>( = d_H(P_1 \parallel Q_1) + d_H(P_2 \parallel Q_2) )</td>
</tr>
<tr>
<td>( 2 \log \det(P_1 Q_1^{-1}) \log \det(P_2 Q_2^{-1}) )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Recent studies suggest that the real signal subspace covariance matrices can be better represented by truncating the eigenvalues. A popular and relatively simple method applies a thresholding and shrinkage rule to the eigenvalues [35]:

$$\tilde{\lambda}_i = \lambda_i \max\{1 - \frac{\gamma}{\lambda_i}, 0\},$$  \hspace{1cm} (107)

where any eigenvalue smaller than the specific threshold is set to zero, and the remaining eigenvalues are shrunk. Note that the smallest eigenvalues are shrunk more than the largest one. For $\gamma = 1$, we obtain a standard soft thresholding, and for $\gamma \to \infty$ a standard hard thresholding is obtained [36].

The optimal threshold $\tau > 0$ can be estimated along with the parameter $\gamma > 0$ using cross validation. However, a more practical and efficient method is to apply the Generalized Stein Unbiased Risk Estimate (GSURE) method even if the variance of the noise is unknown (for more details, we refer to [35] and the references therein).

In this paper, we propose an alternative approach in which the bias generated by noise is reduced by suitable choices of $\alpha$ and $\beta$ [12]. Instead of using the eigenvalues $\lambda_i$ of $PQ^{-1}$ or its inverse, we use regularized or shrinked eigenvalues [35–37]. For example, in light of (8), we can use the following shrinked eigenvalues:

$$\tilde{\lambda}_i = \left(\frac{\alpha \lambda_i^\beta + \beta \lambda_i^{-\alpha}}{\alpha + \beta}\right)^{\frac{1}{\alpha + \beta}} \geq 1, \ \text{for} \ \alpha, \beta \neq 0, \ \alpha, \beta > 0 \ \text{or} \ \alpha, \beta < 0,$$  \hspace{1cm} (108)

which play a similar role as the ratios $(p_i/q_i)$ ($p_i \geq q_i$), which are used in the standard discrete divergences [11,12]. It should be noted that equalities $\tilde{\lambda}_i = 1$, $\forall i$ hold only if all $\lambda_i$ of $PQ^{-1}$ are equal to one, which occurs only if $P = Q$. For example, the new Gamma divergence in (94) can be formulated even more generally as

$$D_{\text{CCA}}^{(\gamma_2, \gamma_1)}(P \parallel Q) = \log \frac{M_{\gamma_2}\{\tilde{\lambda}_i\}}{M_{\gamma_1}\{\lambda_i\}},$$  \hspace{1cm} (109)

where $\gamma_2 > \gamma_1$, and $\tilde{\lambda}_i$ are the regularized or optimally shrinked eigenvalues.

8. Divergences of Multivariate Gaussian Densities and Differential Relative Entropies of Multivariate Normal Distributions

In this section, we show the links or relationships between a family of continuous Gamma divergences and AB log-det divergences for multivariate Gaussian densities.

Consider the two multivariate Gaussian (normal) distributions:

$$p(x) = \frac{1}{\sqrt{(2\pi)^n \det P}} \exp\left(-\frac{1}{2}(x - \mu_1)^T P^{-1} (x - \mu_1)\right),$$  \hspace{1cm} (110)

$$q(x) = \frac{1}{\sqrt{(2\pi)^n \det Q}} \exp\left(-\frac{1}{2}(x - \mu_2)^T Q^{-1} (x - \mu_2)\right), \ x \in \mathbb{R}^n,$$  \hspace{1cm} (111)

where $\mu_1 \in \mathbb{R}^n$ and $\mu_2 \in \mathbb{R}^n$ are mean vectors, and $P = \Sigma_1 \in \mathbb{R}^{n \times n}$ and $Q = \Sigma_2 \in \mathbb{R}^{n \times n}$ are the covariance matrices of $p(x)$ and $q(x)$, respectively.
Furthermore, consider the Gamma divergence for these distributions:

\[
D_{AC}^{(\alpha, \beta)}(p(x)||q(x)) = \frac{1}{\alpha \beta (\alpha + \beta)} \log \left( \frac{\int p^{\alpha+\beta}(x) \, dx}{\int q^{\alpha+\beta}(x) \, dx} \right) \tag{112}
\]

for \( \alpha \neq 0, \beta \neq 0, \alpha + \beta \neq 0 \),

which generalizes a family of Gamma divergences [11,12].

**Theorem 4.** The Gamma divergence in (112) for multivariate Gaussian densities (110) and (111) can be expressed in closed form as follows:

\[
D_{AC}^{(\alpha, \beta)}(p(x)||q(x)) = \frac{1}{2} D_{AB}^{(\beta, \alpha)} \left( \left( Q^{-1/2} P Q^{-1/2} \right)^{1/\alpha} \right) + \frac{1}{2} (\mu_1 - \mu_2)^T (\alpha Q + \beta P)^{-1} (\mu_1 - \mu_2),
\]

\[
= \frac{1}{2 \alpha \beta} \log \frac{\det \left( \frac{\alpha}{\alpha + \beta} Q + \frac{\beta}{\alpha + \beta} P \right)}{\det(Q)^{\alpha/\alpha + \beta} \det(P)^{\beta/\alpha + \beta}} + \frac{1}{2(\alpha + \beta)} (\mu_1 - \mu_2)^T \left( \frac{\alpha}{\alpha + \beta} Q + \frac{\beta}{\alpha + \beta} P \right)^{-1} (\mu_1 - \mu_2),
\tag{113}
\]

for \( \alpha > 0 \) and \( \beta > 0 \).

The proof is provided in Appendix H. Note that for \( \alpha + \beta = 1 \), the first term in the right-hand-side of (113) also simplifies as

\[
\frac{1}{2} D_{AB}^{(\beta, \alpha)} \left( \left( Q^{-1/2} P Q^{-1/2} \right)^{1/\alpha} \right) \bigg|_{\beta = 1 - \alpha} = \frac{1}{2} D_{AB}^{(1-\alpha, \alpha)} (P || Q) = \frac{1}{2} D_{A}^{(1-\alpha)} (P || Q). \tag{114}
\]

Observe that Formula (113) consists of two terms: the first term is expressed via the AB log-det divergence, which measures the similarity between two covariance or precision matrices and is independent from the mean vectors, while the second term is a quadratic form expressed by the Mahalanobis distance, which represents the distance between the means (weighted by the covariance matrices) of multivariate Gaussian distributions. Note that the second term is zero when the mean values \( \mu_1 \) and \( \mu_2 \) coincide.

Theorem 4 is a generalization of the following well-known results:

1. For \( \alpha = 1 \) and \( \beta = 0 \) and as \( \beta \to 0 \), the Kullback-Leibler divergence can be expressed as \([5,38]\)

\[
\lim_{\beta \to 0} D_{AC}^{(1, \beta)}(p(x)||q(x)) = D_{KL}(p(x)||q(x)) = \int_{\Omega} p(x) \log \frac{p(x)}{q(x)} \, dx \tag{115}
\]

\[
= \frac{1}{2} \left( \text{tr}(PQ^{-1}) - \log \det(PQ^{-1}) - n \right) + \frac{1}{2} (\mu_1 - \mu_2)^T Q^{-1} (\mu_1 - \mu_2),
\]

where the last term represents the Mahalanobis distance, which becomes zero for zero-mean distributions \( \mu_1 = \mu_2 = 0 \).
2. For \( \alpha = \beta = 0.5 \) we have the Bhattacharyya distance [5,39]

\[
D_{AC}(0.5,0.5) = \frac{1}{2}d_{Bh}^2(p(\mathbf{x})||q(\mathbf{x})) = -4 \log \int_{\Omega} \sqrt{p(\mathbf{x})q(\mathbf{x})} d\mathbf{x}
\]

\[
= 2 \log \frac{\det P + Q}{\sqrt{\det P \det Q}} + \frac{1}{2}(\mathbf{\mu}_1 - \mathbf{\mu}_2)^T \left[ \frac{P + Q}{2} \right]^{-1}(\mathbf{\mu}_1 - \mathbf{\mu}_2),
\]

3. For \( \alpha + \beta = 1 \) and \( 0 < \alpha < 1 \), the closed form expression for the Rényi divergence is obtained [5,32,40]:

\[
D_A(p||q) = -\frac{1}{\alpha(1-\alpha)} \log \int_{\Omega} p^\alpha(\mathbf{x}) q^{1-\alpha}(\mathbf{x}) d\mathbf{x}
\]

\[
= \frac{1}{2\alpha(1-\alpha)} \log \frac{\det(\alpha Q + (1-\alpha)P)}{\det(Q^\alpha P^{1-\alpha})} + \frac{1}{2}(\mathbf{\mu}_1 - \mathbf{\mu}_2)^T \left[ \alpha Q + (1-\alpha)P \right]^{-1}(\mathbf{\mu}_1 - \mathbf{\mu}_2).
\]

4. For \( \alpha = \beta = 1 \), the Gamma-divergences reduce to the Cauchy-Schwartz divergence:

\[
D_{CS}(p(\mathbf{x})||q(\mathbf{x})) = -\log \frac{\int p(\mathbf{x}) q(\mathbf{x}) d\mu(\mathbf{x})}{\left( \int p^2(\mathbf{x}) d\mu(\mathbf{x}) \right)^{1/2} \left( \int q^2(\mathbf{x}) d\mu(\mathbf{x}) \right)^{1/2}}
\]

\[
= \frac{1}{2} \log \frac{\det P + Q}{\sqrt{\det Q \det P}} + \frac{1}{4}(\mathbf{\mu}_1 - \mathbf{\mu}_2)^T \left( \frac{P + Q}{2} \right)^{-1}(\mathbf{\mu}_1 - \mathbf{\mu}_2).
\]

Similar formulas can be derived for the symmetric Gamma divergence for two multivariate Gaussian distributions. Furthermore, analogous expressions can be derived for Elliptical Gamma distributions (EGD) [41], which facilitate more flexible modeling than standard multivariate Gaussian distributions.

8.1. Multiway Divergences for Multivariate Normal Distributions with Separable Covariance Matrices

Recently, there has been growing interest in the analysis of tensors or multiway arrays [42–45]. One of the most important applications of multiway tensor analysis and multilinear distributions, is magnetic resonance imaging (MRI) (we refer to [46] and the references therein). For multiway arrays, we often use multilinear (array or tensor) normal distributions that correspond to the multivariate normal (Gaussian) distributions in (110) and (111) with common means \( \mathbf{\mu}_1 = \mathbf{\mu}_2 \) and separable (Kronecker structured) covariance matrices:

\[
\bar{P} = \sigma^2_P (P_1 \otimes P_2 \otimes \cdots \otimes P_K) \in \mathbb{R}^{N \times N}
\]

\[
\bar{Q} = \sigma^2_Q (Q_1 \otimes Q_2 \otimes \cdots \otimes Q_K) \in \mathbb{R}^{N \times N},
\]

where \( P_k \in \mathbb{R}^{n_k \times n_k} \) and \( Q_k \in \mathbb{R}^{n_k \times n_k} \) for \( k = 1, 2, \ldots, K \) are SPD matrices, usually normalized so that \( \det P_k = \det Q_k = 1 \) for each \( k \) and \( N = \prod_{k=1}^{K} n_k \) [45].

One of the main advantages of the separable Kronecker model is the significant reduction in the number of variance-covariance parameters [42]. Usually, such separable covariance matrices are sparse
and very large-scale. The challenge is to design an efficient and relatively simple dissimilarity measure for big data between two zero-mean multivariate (or multilinear) normal distributions ((110) and (111)). Because of its unique properties, the Hilbert projective metric is a good candidate; in particular, for separable Kronecker structured covariances, it can be expressed very simply as

$$ D_H(P \parallel Q) = \sum_{k=1}^{K} D_H(P_k \parallel Q_k) = \sum_{k=1}^{K} \log \frac{\tilde{\lambda}_{\max}^{(k)}}{\tilde{\lambda}_{\min}^{(k)}} = \log \prod_{k=1}^{K} \left( \frac{\tilde{\lambda}_{\max}^{(k)}}{\tilde{\lambda}_{\min}^{(k)}} \right), \quad (121) $$

where $\tilde{\lambda}_{\max}^{(k)}$ and $\tilde{\lambda}_{\min}^{(k)}$ are the (shrinked) maximum and minimum eigenvalues of the (relatively small) matrices $P_k Q_k^{-1}$ for $k = 1, 2, \ldots, K$, respectively. We refer to this divergence as the multiway Hilbert metric. This metric has many attractive properties, especially invariance under multilinear transformations.

Using the fundamental properties of divergence and SPD matrices, we derive other multiway log-det divergences. For example, the multiway Stein’s loss can be obtained:

$$ D_{MSL}(P, Q) = 2 D_{KL}(p(x) \parallel q(x)) = D_{0,1}^{(0,1)}(P \parallel Q) = \text{tr} \left( \frac{P Q^{-1}}{Q} \right) - \log \det \left( \frac{P Q^{-1}}{Q} \right) - N $$

$$ = \frac{\sigma_P^2}{\sigma_Q^2} \left( \prod_{k=1}^{K} \text{tr} \left( P_k Q_k^{-1} \right) \right) - \sum_{k=1}^{K} \frac{N}{n_k} \log \det \left( P_k Q_k^{-1} \right) - N \log \left( \frac{\sigma_P^2}{\sigma_Q^2} \right) - N. \quad (123) $$

Note that under the constraint that $\det P_k = \det Q_k = 1$, this simplifies to

$$ D_{MSL}(P \parallel Q) = \text{tr} \left( \frac{P Q^{-1}}{Q} \right) - \log \det \left( \frac{P Q^{-1}}{Q} \right) - N $$

$$ = \frac{\sigma_P^2}{\sigma_Q^2} \left( \prod_{k=1}^{K} \text{tr} \left( P_k Q_k^{-1} \right) \right) - N \log \left( \frac{\sigma_P^2}{\sigma_Q^2} \right) - N, $$

which is different from the multiway Stein’s loss recently proposed by Gerard and Hoff [45].

Similarly, if $\det P_k = \det Q_k = 1$ for each $k = 1, 2, \ldots, K$, we can derive the multiway Riemannian metric as follows:

$$ d_R^2(P \parallel Q) = N \log \frac{\sigma_P^2}{\sigma_Q^2} + \sum_{k=1}^{K} \frac{N}{n_k} d_R^2(P_k \parallel Q_k). \quad (125) $$

The above multiway divergences are derived using the following properties:

$$ \tilde{P} Q^{-1} = (P_1 \otimes P_2 \otimes \cdots \otimes P_K)(Q_1^{-1} \otimes Q_2^{-1} \otimes \cdots \otimes Q_K^{-1}) $$

$$ = P_1 Q_1^{-1} \otimes P_2 Q_2^{-1} \otimes \cdots \otimes P_K Q_K^{-1}, \quad (126) $$

$$ \text{tr}(P Q^{-1}) = \text{tr}(P_1 Q_1^{-1} \otimes P_2 Q_2^{-1} \otimes \cdots \otimes P_K Q_K^{-1}) = \prod_{k=1}^{K} \text{tr}(P_k Q_k^{-1}), \quad (127) $$

$$ \det(P Q^{-1}) = \det(P_1 Q_1^{-1} \otimes P_2 Q_2^{-1} \otimes \cdots \otimes P_K Q_K^{-1}) = \prod_{k=1}^{K} \det(P_k Q_k^{-1})^{N/n_k}. \quad (128) $$

and the basic property: If the eigenvalues $\{\lambda_i\}$ and $\{\theta_j\}$ are eigenvalues with corresponding eigenvectors $\{v_i\}$ and $\{u_j\}$ for SPD matrices A and B, respectively, then $A \otimes B$ has eigenvalues $\{\lambda_i \theta_j\}$ with corresponding eigenvectors $\{v_i \otimes u_j\}$. 
Other possible extensions of the AB and Gamma matrix divergences to separable multiway divergences for multilinear normal distributions under additional constraints and normalization conditions will be discussed in future works.

9. Conclusions

In this paper, we presented novel (dis)similarity measures; in particular, we considered the Alpha-Beta and Gamma log-det divergences (and/or their square-roots) that smoothly connect or unify a wide class of existing divergences for SPD matrices. We derived numerous results that uncovered or unified theoretic properties and qualitative similarities between well-known divergences and new divergences. The scope of the results presented in this paper is vast, especially since the parameterized Alpha-Beta and Gamma log-det divergence functions include several efficient and useful divergences, including those based on relative entropies, the Riemannian metric (AIRM), S-divergence, generalized Jeffreys KL (KLDM), Stein’s loss, and Hilbert projective metric. Various links and relationships between divergences were also established. Furthermore, we proposed several multiway log-det divergences for tensor (array) normal distributions.

Acknowledgments

Part of this work was supported by the Spanish Government under MICINN projects TEC2014-53103, TEC2011-23559, and by the Regional Government of Andalusia under Grant TIC-7869.

Appendices

A. Basic operations for positive definite matrices

Functions of positive definite matrices frequently appear in many research areas, for an introduction we refer the reader to Chapter 11 in [31]. Consider a positive definite matrix $P$ of rank $n$ with eigendecomposition $V \Lambda V^T$. The matrix function $f(P)$ is defined as

$$f(P) = V f(\Lambda)V^T,$$

where $f(\Lambda) \equiv \text{diag}(f(\lambda_1), \ldots, f(\lambda_n))$. With the help of this definition, the following list of well-known properties can be easily obtained:
### B. Extension of $D_{AB}^{(\alpha,\beta)}(P\|Q)$ for $(\alpha,\beta) \in \mathbb{R}^2$

**Remark 1.** Equation (3) is only well defined in the first and third quadrants of the $(\alpha,\beta)$-plane. Outside these regions, where $\alpha$ and $\beta$ have opposite signs (i.e., $\alpha > 0$ and $\beta < 0$ or $\alpha < 0$ and $\beta > 0$), the divergence can be complex valued.

This undesirable behavior can be avoided with the help of the truncation operator

$$[x]_+ = \begin{cases} 
  x & x \geq 0, \\
  0 & x < 0,
\end{cases}$$

which prevents the arguments of the logarithms from being negative.

The new definition of the AB log-det divergence is

$$D_{AB}^{(\alpha,\beta)}(P\|Q) = \frac{1}{\alpha \beta} \log \left[ \det \frac{\alpha(PQ^{-1})^\beta + \beta(PQ^{-1})^{-\alpha}}{\alpha + \beta} \right]_+$$

for $\alpha \neq 0$, $\beta \neq 0$, $\alpha + \beta \neq 0$,

which is compatible with the previous definition in the first and third quadrants of the $(\alpha,\beta)$-plane. It is also well defined in the second and fourth quadrants except for the special cases when $\alpha = 0$, $\beta = 0$, $\alpha + \beta = 0$. 

---

\[ \log(\det P) = \text{tr} \log(P), \quad (\det P)^{\alpha} = \det(P^\alpha), \]
\[ \det(P^\alpha) = \det(V\Lambda V^T)^\alpha = \det(V) \det(\Lambda^\alpha) \det(V^T) = \prod_{i=1}^n \lambda_i^\alpha, \]
\[ \text{tr}(P^\alpha) = \text{tr}(V\Lambda V^T)^\alpha = \text{tr}(VV^T\Lambda^\alpha) = \sum_{i=1}^n \lambda_i^\alpha, \]
\[ P^{\alpha+\beta} = P^\alpha P^\beta, \quad (P^\alpha)^\beta = P^{\alpha \beta}, \quad P^0 = I, \]
\[ (\det P)^{\alpha+\beta} = \det(P^\alpha) \det(P^\beta), \]
\[ \det((PQ^{-1})^\alpha) = [\det(P) \det(Q^{-1})]^\alpha = \det(P^\alpha) \det(Q^{-\alpha}), \]
\[ \frac{\partial}{\partial \alpha} (P^\alpha) = P^\alpha \log(P), \]
\[ \frac{\partial}{\partial \alpha} \log[\det(P(\alpha))] = \text{tr} \left( P^{-1} \frac{\partial P}{\partial \alpha} \right), \]
\[ \log(\det(P \otimes Q)) = n \log(\det P) + n \log(\det Q), \]
\[ \text{tr}(P) - \log(\det P) \geq n. \]
and $\alpha + \beta = 0$, which is where the formula is undefined. By enforcing continuity, we can explicitly define the AB log-det divergence on the entire $(\alpha,\beta)$-plane as follows:

$$D_{AB}^{(\alpha,\beta)}(P\parallel Q) = \begin{cases} 
\frac{1}{\alpha\beta} \log \det \left[ \frac{\alpha(PQ^{-1})^\beta + \beta(QP^{-1})^\alpha}{\alpha + \beta} \right]_+ & \text{for } \alpha, \beta \neq 0, \ \alpha + \beta \neq 0, \\
\frac{1}{\alpha^2} \left[ \text{tr} \left( (QP^{-1})^\alpha - I \right) - \alpha \log \det(QP^{-1}) \right] & \text{for } \alpha \neq 0, \beta = 0, \\
\frac{1}{\beta^2} \left[ \text{tr} \left( (QP^{-1})^\beta - I \right) - \beta \log \det(QP^{-1}) \right] & \text{for } \alpha = 0, \beta \neq 0, \\
\frac{1}{\alpha^2} \log \det[ (PQ^{-1})^{-\alpha}(I + \log(PQ^{-1}))^{-\alpha}]^{-1} & \text{for } \alpha = -\beta, \\
\frac{1}{2} \text{tr} \log^2(PQ^{-1}) = \frac{1}{2} \| \log(Q^{-1/2}PQ^{-1/2}) \|_F^2 & \text{for } \alpha, \beta = 0.
\end{cases}$$

(145)

C. Eigenvalues Domain for Finite $D_{AB}^{(\alpha,\beta)}(P\parallel Q)$

In this section, we assume that $\lambda_i$, an eigenvalue of $PQ^{-1}$, satisfies $0 \leq \lambda_i \leq \infty$ for all $i = 1, \ldots, n$. We will determine the bounds of the eigenvalues of $PQ^{-1}$ that prevent the AB log-det divergence from being infinite. First, recall that

$$D_{AB}^{(\alpha,\beta)}(P\parallel Q) = \frac{1}{\alpha\beta} \sum_{i=1}^n \log \left[ \frac{\alpha\lambda_i^\beta + \beta\lambda_i^{-\alpha}}{\alpha + \beta} \right]_+ , \ \alpha, \beta, \alpha + \beta \neq 0. \quad (146)$$

We assume that $0 \leq \lambda_i \leq \infty$ for all $i$. For the divergence to be finite, the arguments of the logarithms in the previous expression must be positive. This happens when

$$\frac{\alpha\lambda_i^\beta + \beta\lambda_i^{-\alpha}}{\alpha + \beta} > 0 \quad \forall i,$$

which is always true when $\alpha, \beta > 0$ or when $\alpha, \beta < 0$. On the contrary, when $\text{sign}(\alpha\beta) = -1$, we have the following two cases. In the first case when $\alpha > 0$, we initially solve for $\lambda_i^{\alpha+\beta}$ and later for $\lambda_i$ to obtain

$$\frac{\lambda_i^{\alpha+\beta}}{\alpha + \beta} > \frac{-\beta}{\alpha(\alpha + \beta)} = \frac{\beta}{\alpha} \frac{1}{\alpha + \beta} \quad \Rightarrow \quad \lambda_i > \left| \frac{\beta}{\alpha} \right|^{\frac{1}{\alpha+\beta}} \quad \forall i, \text{ for } \alpha > 0 \text{ and } \beta < 0. \quad (148)$$

In the second case when $\alpha < 0$, we obtain

$$\frac{\lambda_i^{\alpha+\beta}}{\alpha + \beta} < \frac{-\beta}{\alpha(\alpha + \beta)} = \frac{\beta}{\alpha} \frac{1}{\alpha + \beta} \quad \Rightarrow \quad \lambda_i < \left| \frac{\beta}{\alpha} \right|^{\frac{1}{\alpha+\beta}} \quad \forall i, \text{ for } \alpha < 0 \text{ and } \beta > 0. \quad (149)$$
Using \(\text{sign}(\alpha \beta) = -1\), we can solve for \(\lambda_i^{\alpha + \beta}\), which yields

\[
\frac{\lambda_i^{\alpha + \beta}}{\alpha + \beta} > \left| \frac{\beta}{\alpha} \right| \frac{1}{\alpha + \beta} \quad \forall i. \tag{150}
\]

Solving again for \(\lambda_i\), we see that

\[
\lambda_i > \left| \frac{\beta}{\alpha} \right| ^{\frac{1}{\alpha + \beta}} \quad \forall i, \text{ for } \alpha > 0 \text{ and } \beta < 0, \tag{151}
\]

and

\[
\lambda_i < \left| \frac{\beta}{\alpha} \right| ^{\frac{1}{\alpha + \beta}} \quad \forall i, \text{ for } \alpha < 0 \text{ and } \beta > 0. \tag{152}
\]

In the limit, when \(\alpha \to -\beta \neq 0\), these bounds simplify to

\[
\lim_{\alpha \to -\beta} \left| \frac{\beta}{\alpha} \right| ^{\frac{1}{\alpha + \beta}} = e^{-1/\alpha} \quad \forall i, \text{ for } \beta \neq 0. \tag{153}
\]

On the other hand, when \(\alpha \to 0\) or when \(\beta \to 0\), the bounds disappear. The lower-bounds converge to 0, while the upper-bounds converge to \(\infty\), leading to the trivial inequalities \(0 < \lambda_i < \infty\).

This concludes the determination of the domain of the eigenvalues that result in a finite divergence. Outside this domain, we expect \(D_{AB}^{(\alpha, \beta)}(P\|Q) = \infty\). A complete picture of bounds for different values of \(\alpha\) and \(\beta\) is shown in Figure 1.

### D. Proof of the Nonnegativity of \(D_{AB}^{(\alpha, \beta)}(P\|Q)\)

The AB log-det divergence is separable; it is the sum of the individual divergences of the eigenvalues from unity, \textit{i.e.,}

\[
D_{AB}^{(\alpha, \beta)}(P\|Q) = \sum_{i=1}^{n} D_{AB}^{(\alpha, \beta)}(\lambda_i\|1), \tag{154}
\]

where

\[
D_{AB}^{(\alpha, \beta)}(\lambda_i\|1) = \frac{1}{\alpha \beta} \log \left[ \frac{\alpha \lambda_i^\beta + \beta \lambda_i^{-\alpha}}{\alpha + \beta} \right]_+, \quad \alpha, \beta, \alpha + \beta \neq 0. \tag{155}
\]

We prove the nonnegativity of \(D_{AB}^{(\alpha, \beta)}(P\|Q)\) by showing that the divergence of each of the eigenvalues \(D_{AB}^{(\alpha, \beta)}(\lambda_i\|1)\) is nonnegative and minimal at \(\lambda_i = 1\).

First, note that the only critical point of the criterion is obtained when \(\lambda_i = 1\). This can be shown by setting the derivative of the criterion equal to zero, \textit{i.e.,}

\[
\frac{\partial D_{AB}^{(\alpha, \beta)}(\lambda_i\|1)}{\partial \lambda_i} = \frac{\lambda_i^{\alpha + \beta} - 1}{\alpha \lambda_i^{\alpha + \beta + 1} + \beta \lambda_i} = 0, \tag{156}
\]

and solving for \(\lambda_i\).
Next, we show that the sign of the derivative only changes at the critical point $\lambda_i = 1$. If we rewrite
\[
\frac{\partial D^{(\alpha,\beta)}_{AB}(\lambda_i\|1)}{\partial \lambda_i} = \left( \frac{\lambda_i^{\alpha + \beta} - 1}{\alpha + \beta} \right) \left( \lambda_i \frac{\alpha \lambda_i^{\alpha + \beta} + \beta}{\alpha + \beta} \right)^{-1},
\]
and observe that the condition for the divergence to be finite enforces $\alpha \lambda_i^{\alpha + \beta} + \beta > 0$, then it follows that
\[
\text{sign} \left\{ \frac{\partial D^{(\alpha,\beta)}_{AB}(\lambda_i\|1)}{\partial \lambda_i} \right\} \equiv \text{sign} \left\{ \frac{\lambda_i^{\alpha + \beta} - 1}{\alpha + \beta} \right\} = \begin{cases} -1 & \text{for } \lambda_i < 1, \\ 0 & \text{for } \lambda_i = 1, \\ +1 & \text{for } \lambda_i > 1. \end{cases}
\]
Since the derivative is strictly negative for $\lambda_i < 1$ and strictly positive for $\lambda_i > 1$, the critical point at $\lambda_i = 1$ is the global minimum of $D^{(\alpha,\beta)}_{AB}(\lambda_i\|1)$. From this result, the nonnegativity of the divergence $D^{(\alpha,\beta)}_{AB}(P\|Q) \geq 0$ easily follows. Moreover, $D^{(\alpha,\beta)}_{AB}(P\|Q) = 0$ only when $\lambda_i = 1$ for $i = 1, \ldots, n$, which concludes the proof of the Theorems 1 and 2.

E. Derivation of the Riemannian Metric

We calculate $D^{(\alpha,\beta)}_{AB}(P + dP \| P)$ using the Taylor expansion when $dP$ is small, i.e.,
\[
(P + dP)P^{-1} = I + dZ,
\]
where
\[
dZ = dPP^{-1}, \quad \alpha [(P + dP)P^{-1}]^\beta = \alpha I + \alpha \beta dZ + \frac{\alpha \beta (\beta - 1)}{2} dZ dZ + O(|dZ|^3).
\]
Similar calculations hold for $\beta[(P + dP)P^{-1}]^{-\alpha}$, and
\[
\alpha [(P + dP)P^{-1}]^\beta + \beta[(P + dP)P^{-1}]^{-\alpha} = (\alpha + \beta) \left( I + \frac{\alpha \beta}{2} dZ dZ \right),
\]
where the first-order term of $dZ$ disappears and the higher-order terms are neglected. Since
\[
\det \left( I + \frac{\alpha \beta}{2} dZ dZ \right) = 1 + \frac{\alpha \beta}{2} \text{tr}(dZ dZ),
\]
by taking its logarithm, we have
\[
D^{(\alpha,\beta)}_{AB}(P + dP \| P) = \frac{1}{2} \text{tr}(dP P^{-1} dP P^{-1}),
\]
for any $\alpha$ and $\beta$. 
F. Proof of the Properties of the AB Log-Det Divergence

Next we provide a proof of the properties of the AB log-det divergence. The proof will only be omitted for those properties which can be readily verified from the definition of the divergence.

1. Nonnegativity; given by

\[ D^{(\alpha, \beta)}_{AB}(P \parallel Q) \geq 0, \quad \forall \alpha, \beta \in \mathbb{R}. \]  

(162)

The proof of this property is presented in Appendix D.

2. Identity of indiscernibles; given by

\[ D^{(\alpha, \beta)}_{AB}(P \parallel Q) = 0 \quad \text{if and only if} \quad P = Q. \]  

(163)

See Appendix D for its proof.

3. Continuity and smoothness of \( D^{(\alpha, \beta)}_{AB}(P \parallel Q) \) as a function of \( \alpha \in \mathbb{R} \) and \( \beta \in \mathbb{R} \), including the singular cases when \( \alpha = 0 \) or \( \beta = 0 \), and when \( \alpha = -\beta \) (see Figure 2).

4. The divergence can be explicitly expressed in terms of \( \Lambda = \text{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_n\} \), the diagonal matrix with the eigenvalues of \( Q^{-1}P \); in the form

\[ D^{(\alpha, \beta)}_{AB}(P \parallel Q) = D^{(\alpha, \beta)}_{AB}(\Lambda \parallel I). \]  

(164)

**Proof.** From the definition of divergence and taking into account the eigenvalue decomposition \( PQ^{-1} = V\Lambda V^{-1} \), we can write

\[
D^{(\alpha, \beta)}_{AB}(P \parallel Q) = \frac{1}{\alpha \beta} \log \det \frac{\alpha V\Lambda^\beta V^{-1} + \beta V\Lambda^{-\alpha} V^{-1}}{\alpha + \beta} \\
= \frac{1}{\alpha \beta} \log \left[ \det V \left( \frac{\alpha \Lambda^\beta + \beta \Lambda^{-\alpha}}{\alpha + \beta} \right) \det V^{-1} \right] \\
= \frac{1}{\alpha \beta} \log \det \frac{\alpha \Lambda^\beta + \beta \Lambda^{-\alpha}}{\alpha + \beta} \]  

(165)

\[ = D^{(\alpha, \beta)}_{AB}(\Lambda \parallel I). \]  

(166)

\[ \square \]

5. Scaling invariance; given by

\[ D^{(\alpha, \beta)}_{AB}(cP \parallel cQ) = D^{(\alpha, \beta)}_{AB}(P \parallel Q), \]  

(167)

for any \( c > 0 \).

6. For a given \( \alpha \) and \( \beta \) and nonzero scaling factor \( \omega \neq 0 \), we have

\[ D^{(\omega \alpha, \omega \beta)}_{AB}(P \parallel Q) = \frac{1}{\omega^2} D^{(\alpha, \beta)}_{AB}((Q^{-1/2}PQ^{-1/2})^\omega \parallel I). \]  

(168)
Proof. From the definition of divergence, we write
\[
D_{AB}^{(\omega \alpha, \omega \beta)}(P \parallel Q) = \frac{1}{(\omega \alpha)(\omega \beta)} \log \det \frac{\omega \alpha \Lambda^{\omega \beta} + \omega \beta \Lambda^{-\omega \alpha}}{(\omega \alpha + \omega \beta)} \tag{169}
\]
\[
= \frac{1}{\omega^2} \frac{1}{\alpha \beta} \log \det \frac{\alpha (\Lambda^{\omega \beta})^\beta + \beta (\Lambda^{\omega})^{-\alpha}}{(\alpha + \beta)} \tag{170}
\]
\[
= \frac{1}{\omega^2} D_{AB}^{(\alpha, \beta)}((Q^{-1/2}PQ^{-1/2})^\omega \parallel I) \tag{171}
\]
Hence, the additional inequality
\[
D_{AB}^{(\alpha, \beta)}((Q^{-1/2}PQ^{-1/2})^\omega \parallel I) \leq D_{AB}^{(\omega \alpha, \omega \beta)}(P \parallel Q) \tag{172}
\]
is obtained for \(|\omega| \leq 1\). □

7. Dual-invariance under inversion (for \(\omega = -1\)); given by
\[
D_{AB}^{(-\alpha, -\beta)}(P \parallel Q) = D_{AB}^{(\alpha, \beta)}(P^{-1} \parallel Q^{-1}) \tag{173}
\]

8. Dual symmetry; given by
\[
D_{AB}^{(\alpha, \beta)}(P \parallel Q) = D_{AB}^{(\beta, \alpha)}(Q \parallel P) \tag{174}
\]

9. Affine invariance (invariance under congruence transformations); given by
\[
D_{AB}^{(\alpha, \beta)}(APA^T \parallel AQ^T) = D_{AB}^{(\alpha, \beta)}(P \parallel Q) \tag{175}
\]
for any nonsingular matrix \(A \in \mathbb{R}^{n \times n}\).

Proof.
\[
D_{AB}^{(\alpha, \beta)}(APA^T \parallel AQ^T) = \frac{1}{\alpha \beta} \log \det \frac{\alpha ((APA^T)(AQ^T)^{-1})^\beta + \beta ((APA^T)(AQ^T)^{-1})^{-\alpha}}{\alpha + \beta} \\
= \frac{1}{\alpha \beta} \log \det \frac{\alpha (A(PQ^{-1})A^{-1})^\beta + \beta (A(PQ^{-1})A^{-1})^{-\alpha}}{\alpha + \beta} \\
= \frac{1}{\alpha \beta} \log \left[ \det(\Lambda^\alpha) \det \frac{\alpha \Lambda^\beta + \beta \Lambda^{-\alpha}}{\alpha + \beta} \det(\Lambda^{-1}) \right] \\
= \frac{1}{\alpha \beta} \log \frac{\alpha \Lambda^\beta + \beta \Lambda^{-\alpha}}{\alpha + \beta} \\
= D_{AB}^{(\alpha, \beta)}(P \parallel Q) \tag{176}
\]

□

10. Divergence lower-bound; given by
\[
D_{AB}^{(\alpha, \beta)}(X^TPX \parallel X^TQX) \leq D_{AB}^{(\alpha, \beta)}(P \parallel Q), \tag{177}
\]
for any full-column rank matrix \(X \in \mathbb{R}^{n \times m}\) with \(n \leq m\).

This result has been already proved for some special cases of \(\alpha\) and \(\beta\), especially these that lead to the S-divergence and the Riemannian metric [6]. Next, we present a different argument to prove it for any \(\alpha, \beta \in \mathbb{R}\).
Proof. As already discussed, the divergence \( D_{AB}^{(\alpha,\beta)}(P\|Q) \) depends on the generalized eigenvalues of the matrix pencil \((P, Q)\), which have been denoted by \( \lambda_i, i = 1, \ldots, n \). Similarly, the presumed lower-bound \( D_{AB}^{(\alpha,\beta)}(X^T PX\|X^T QX) \) is determined by \( \mu_i, i = 1, \ldots, m \), the eigenvalues of the matrix pencil \((X^T PX, X^T QX)\). Assuming that both sets of eigenvalues are arranged in decreasing order, the Cauchy interlacing inequalities \([29]\) provide the following upper and lower-bounds for \( \mu_j \) in terms of the eigenvalues of the first matrix pencil,

\[
\lambda_j \leq \mu_j \leq \lambda_{n-m+j}.
\] (178)

We classify the eigenvalues \( \mu_j \) on three sets \( S^-\mu, S^0\mu \) and \( S^+\mu \), according to the sign of \( (\mu_j - 1) \). By the affine invariance we can write

\[
D_{AB}^{(\alpha,\beta)}(X^T PX\|X^T QX) = D_{AB}^{(\alpha,\beta)}(((X^T QX)^{-1/2}X^T PX(X^T QX)^{-1/2})\|I) \quad (179)
\]

\[
= \sum_{\mu_j \in S^-\mu} D_{AB}^{(\alpha,\beta)}(\mu_j\|1) + \sum_{\mu_j \in S^+\mu} D_{AB}^{(\alpha,\beta)}(\mu_j\|1), \quad (180)
\]

where the eigenvalues \( \mu_j \in S^0\mu \) have been excluded since for them \( D_{AB}^{(\alpha,\beta)}(\mu_j\|1) = 0 \).

With the help of (178), the first group of eigenvalues \( \mu_j \in S^-\mu \) (which are smaller than one) are one-to-one mapped with their lower-bounds \( \lambda_j \), which we include in the set \( S^-\lambda \). Also those \( \mu_j \in S^+\mu \) (which are greater than one) are mapped with their upper-bounds \( \lambda_{n-m+j} \), which we group in \( S^+\lambda \). It is shown in Appendix D that the scalar divergence \( D_{AB}^{(\alpha,\alpha)}(\lambda\|1) \) is strictly monotone descending for \( \lambda < 1 \), zero for \( \lambda = 1 \) and strictly monotone ascending for \( \lambda > 1 \). This allows one to upperbound (180) as follows

\[
\sum_{\mu_j \in S^-\mu} D_{AB}^{(\alpha,\beta)}(\mu_j\|1) + \sum_{\mu_j \in S^+\mu} D_{AB}^{(\alpha,\beta)}(\mu_j\|1) \leq \sum_{\lambda_j \in S^-\lambda} D_{AB}^{(\alpha,\beta)}(\lambda_j\|1) + \sum_{\lambda_j \in S^+\lambda} D_{AB}^{(\alpha,\beta)}(\lambda_j\|1)
\]

\[
\leq \sum_{j=1}^{n} D_{AB}^{(\alpha,\beta)}(\lambda_j\|1), \quad (181)
\]

\[
= D_{AB}^{(\alpha,\beta)}(P\|Q), \quad (182)
\]

obtaining the desired property. \( \square \)

11. Scaling invariance under the Kronecker product; given by

\[
D_{AB}^{(\alpha,\beta)}(Z \otimes P\|Z \otimes Q) = n D_{AB}^{(\alpha,\beta)}(P\|Q), \quad (183)
\]

for any symmetric and positive definite matrix \( Z \) of rank \( n \).
Double Sided Orthogonal Procrustes property. Consider an orthogonal matrix \( \Omega \in \mathcal{O}(n) \) and two symmetric positive definite matrices \( P \) and \( Q \), with respective eigenvalue matrices \( \Lambda_P \) and \( \Lambda_Q \), which elements are sorted in descending order. The AB log-det divergence between \( \Omega^T P \Omega \) and \( Q \) is globally minimized when their eigenspaces are aligned, i.e.,

\[
\min_{\Omega \in \mathcal{O}(n)} D^{(\alpha,\beta)}_{AB}(\Omega^T P \Omega||Q) = D^{(\alpha,\beta)}_{AB}(\Lambda_P||\Lambda_Q).
\]

Proof. Let \( \Lambda \) denote the matrix of eigenvalues of \( \Omega^T P \Omega Q^{-1} \) with its elements sorted in descending order. We start showing that for \( \Delta = \log \Lambda \), the function \( D^{(\alpha,\beta)}_{AB}(\exp \Delta || I) \) is convex. Its Hessian matrix is diagonal and positive definite, i.e., with non-negative diagonal elements

\[
\frac{\partial^2 D^{(\alpha,\beta)}_{AB}(e^\Delta || 1)}{\partial \Delta_{ii}^2} > 0,
\]

where

\[
\frac{\partial^2 D^{(\alpha,\beta)}_{AB}(e^\Delta || 1)}{\partial \Delta_{ii}^2} = \begin{cases} 
\left( \frac{\beta}{\alpha+\beta} e^{-\frac{\alpha+\beta}{2} \Delta_{ii}} + \frac{\alpha}{\alpha+\beta} e^{\frac{\alpha+\beta}{2} \Delta_{ii}} \right)^{-2} & \text{for } \alpha, \beta, \alpha + \beta \neq 0 \\
0 & \text{for } \alpha = 0 \\
(1 + \alpha \Delta_{ii})^{-2} & \text{for } \alpha + \beta = 0 \\
en^\alpha \Delta_{ii} & \text{for } \beta = 0.
\end{cases}
\]

Since \( f(e^\Delta_{ii}) = D^{(\alpha,\beta)}_{AB}(e^\Delta_{ii} || 1) \) is strictly convex and non-negative, we are in the conditions of the Corollary 6.15 in [47]. This result states that for two symmetric positive definite matrices \( A \) and \( B \), which vectors of eigenvalues are respectively denoted by \( \vec{\lambda}_A \) (when sorted in descending order) and \( \vec{\lambda}_B \) (when sorted in ascending order), the function \( f(\vec{\lambda}_A \vec{\lambda}_B^\top) \) is submajorized by \( f(\vec{\lambda}_{AB}^\top) \). By choosing \( A = \Omega^T P \Omega, B = Q^{-1} \), and applying the corollary, we obtain

\[
D^{(\alpha,\beta)}_{AB}(\Lambda_P||\Lambda_Q) = D^{(\alpha,\beta)}_{AB}(\Lambda_P \Lambda_Q^\top || I) \leq D^{(\alpha,\beta)}_{AB}(\Lambda || I) = D^{(\alpha,\beta)}_{AB}(\Omega^T P \Omega || Q),
\]

12.
where the equality is only reached when the eigendecompositions of the matrices $\Omega^T P \Omega = V \Lambda_P V^T$ and $Q = V \Lambda_Q V^T$, share the same matrix of eigenvectors $V$.  

13. Triangle Inequality-Metric Distance Condition, for $\alpha = \beta \in \mathbb{R}$; given by

$$\sqrt{D^{(\alpha, \alpha)}_{AB} (P \| Q)} \leq \sqrt{D^{(\alpha, \alpha)}_{AB} (P \| Z)} + \sqrt{D^{(\alpha, \alpha)}_{AB} (Z \| Q)}.$$  \hspace{1cm} (192)

**Proof.** The proof of this property exploits the recent result that the square root of the S-divergence $\sqrt{D_{Bh}(P \| Q)} = \sqrt{D_{S}(P \| Q)} = 2 \sqrt{\log \frac{\det \frac{1}{2} (P + Q)}{\sqrt{\det P \det Q}}}$.

is a metric [17]. Given three arbitrary symmetric positive definite matrices $P, Q, Z$, with common dimensions, consider the following eigenvalue decompositions

$$Q^{-\frac{1}{2}} P Q^{-\frac{1}{2}} = V_1 \Lambda_1 V_1^T$$  \hspace{1cm} (194)

$$Q^{-\frac{1}{2}} Z Q^{-\frac{1}{2}} = V_2 \Lambda_2 V_2^T,$$  \hspace{1cm} (195)

and assume that the diagonal matrices $\Lambda_1$ and $\Lambda_2$ have the eigenvalues sorted in a descending order.

For a given value of $\alpha$ in the divergence, we define $\omega = 2\alpha \neq 0$ and use properties 6 and 9 (see Equations (168) and (175)) to obtain the equivalence

$$\sqrt{D^{(\alpha, \alpha)}_{AB} (P \| Q)} = \sqrt{D^{(\omega, \omega)}_{AB} (P \| Q)} \leq \sqrt{D^{(\omega, \omega)}_{AB} (\Lambda_1 \| \Lambda_2)} + \sqrt{D^{(\omega, \omega)}_{AB} (\Lambda_2 \| I)}.$$  

Since the S-divergence satisfies the triangle inequality for diagonal matrices [5,6,17]

$$d_{Bh}(\Lambda_1^{2\alpha} \| I) \leq d_{Bh}(\Lambda_1^{2\alpha} \| \Lambda_2^{2\alpha}) + d_{Bh}(\Lambda_2^{2\alpha} \| I),$$  

from (196), this implies that

$$\sqrt{D^{(\alpha, \alpha)}_{AB} (P \| Q)} \leq \sqrt{D^{(\alpha, \alpha)}_{AB} (\Lambda_1 \| \Lambda_2)} + \sqrt{D^{(\alpha, \alpha)}_{AB} (\Lambda_2 \| I)}.$$  

In similarity with the proof of the metric condition for S-divergence [6], we can use property 12 to bound above the first term in the right-hand-side of the equation by

$$\sqrt{D^{(\alpha, \alpha)}_{AB} (\Lambda_1 \| \Lambda_2)} \leq \sqrt{D^{(\alpha, \alpha)}_{AB} (V_1 \Lambda_1 V_1^T \| V_2 \Lambda_2 V_2^T)} = \sqrt{D^{(\alpha, \alpha)}_{AB} (Q^{-\frac{1}{2}} P Q^{-\frac{1}{2}} \| Q^{-\frac{1}{2}} Z Q^{-\frac{1}{2}})} = \sqrt{D^{(\alpha, \alpha)}_{AB} (P \| Z)},$$  \hspace{1cm} (199)
whereas the second term satisfies
\[
\sqrt{D_{AB}^{(\alpha,\alpha)}(A_2\|I)} = \sqrt{D_{AB}^{(\alpha,\alpha)}(V_2\Lambda_2 V_2^T\|I)} = \sqrt{D_{AB}^{(\alpha,\alpha)}(Q^{-\frac{1}{2}}PQ^{-\frac{1}{2}}\|I)} = \sqrt{D_{AB}^{(\alpha,\alpha)}(Z\|Q)}. 
\]

(200)

After bounding the right-hand-side of (198) with the help of (199) and (200), the divergence satisfies the desired triangle inequality (192) for \( \alpha \neq 0 \).

On the other hand, \( \sqrt{D_{AB}^{(\alpha,\alpha)}(P\|Q)} \) as \( \alpha \to 0 \) converges to the Riemannian metric
\[
\sqrt{D_{AB}^{(0,0)}(P\|Q)} = \lim_{\alpha \to 0} \sqrt{D_{AB}^{(\alpha,\alpha)}(P\|Q)} = \| \log(Q^{-1/2}PQ^{-1/2}) \|_F = d_R(P\|Q),
\]
which concludes the proof of the metric condition of \( \sqrt{D_{AB}^{(\alpha,\alpha)}(P\|Q)} \) for any \( \alpha \in \mathbb{R} \).

G. Proof of Theorem 3

This theorem assumes that the range spaces of the symmetric positive semidefinite matrices \( C_x \) and \( C_y \) are disjoint, in the sense that they only intersect at the origin, which is the most probable situation for \( n \gg r \) (where \( n \) is the size of the matrices while \( r \) is their common rank). For \( \rho > 0 \) the regularized versions \( \tilde{C}_x \) and \( \tilde{C}_y \) of these matrices are full rank.

Let \( \tilde{\Lambda} = \text{diag}(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_n) \) denote the diagonal matrix representing the \( n \) eigenvalues of the matrix pencil \( (\tilde{C}_x, \tilde{C}_y) \). The AB log-det divergence between the regularized matrices is equal to the divergence between \( \tilde{\Lambda} \) and the identity matrix of size \( n \), i.e.,
\[
D_{AB}^{(\alpha,\beta)}(\tilde{C}_x \| \tilde{C}_y) = D_{AB}^{(\alpha,\beta)}(\tilde{C}_x^{-\frac{1}{2}}\tilde{C}_x\tilde{C}_y^{-\frac{1}{2}} \| I_n) = D_{AB}^{(\alpha,\beta)}(\tilde{\Lambda} \| I_n).
\]

(204)

The positive eigenvalues of the matrix pencil satisfy
\[
\tilde{\Lambda} \equiv \text{diag} E ig_+ \left\{ (\tilde{C}_y)^{-\frac{1}{2}} \tilde{C}_x (\tilde{C}_y)^{-\frac{1}{2}} \right\} = \text{diag} E ig_+ \left\{ \tilde{C}_x \tilde{C}_y^{-1} \right\},
\]

(205)

therefore, the divergence can be directly estimated from the eigenvalues of \( \tilde{C}_x \tilde{C}_y^{-1} \). In order to simplify this matrix product, we first express \( \tilde{C}_x \) and \( \tilde{C}_y^{-1} \) in term of the auxiliary matrices
\[
T_x = U_x(\Lambda_x - \rho I_r)^\frac{1}{2} \quad \text{and} \quad T_y = U_y(\Lambda_y - \rho I_r)^\frac{1}{2}.
\]

(206)

In this way, they are written as a scaled version of the identity matrix plus a symmetric term:
\[
\tilde{C}_x = C_x + \rho U_x^T (U_x^T)^T \quad \text{and} \quad \tilde{C}_y^{-1} = C_y^{-1} - \rho I_r.
\]

(207)
and

\[
\tilde{C}_y^{-1} = C_y^* + \rho^{-1}U_y^T(U_y^T)^T \\
= U_yA_y^{-1}U_y^T + \rho^{-1}(I_n - U_yU_y^T) \\
= \rho^{-1}I_n - \rho^{-1}U_y(A_y + \rho I_r)A_y^{-1}U_y^T \\
= \rho^{-1}I_n - \rho^{-1}T_yA_y^{-1}T_y^T.
\]

(208)

Next, using (207) and (208), we expand the product

\[
\tilde{C}_x\tilde{C}_y^{-1} = I_n + \rho^{-1}T_xT_x^T(I_n - T_yA_y^{-1}T_y^T) + R
\]

and approximate the eigenvectors \( U_y \to U_x \) of the residual matrix \( R \) to obtain the estimate

\[
R \equiv -U_y(I_r + \rho A_y^{-1})U_y^T \approx -U_x(I_r + \rho A_y^{-1})U_x^T \equiv \tilde{R}.
\]

(210)

Hence, it is not difficult to see that the estimated residual is equal to

\[
\tilde{R} = -T_x(I_r + \rho A_y^{-1})T_x^+.
\]

(211)

After substituting (211) in (209) and collecting common terms, we obtain the expansion

\[
\tilde{C}_x\tilde{C}_y^{-1} = \underbrace{I_n + T_x\left(\rho^{-1}T_x^T - \rho^{-1}T_xT_yA_y^{-1}T_y^T - (I_r + \rho A_y^{-1})T_x^+\right)}_{\tilde{C}_x\tilde{C}_y^{-1}} + O(\rho^0).
\]

(212)

Let \( Eig_{\leq 1}\{\cdot\} \) denote the arrangement of the ordered eigenvalues of the matrix argument after excluding those that are equal to 1. For convenience, we reformulate the property proved in [30] that for any pair of matrices \( A, B \in \mathbb{R}^{m \times n} \), the non-zero eigenvalues of \( AB^T \) and of \( B^TA \) are the same, into the following proposition.

**Proposition 1.** For any pair of \( m \times n \) matrices \( A \) and \( B \), the eigenvalues of the matrices \( I_n + AB^T \) and \( I_n + B^TA \), which are not equal to 1, coincide.

\[
Eig_{\leq 1}\{I_n + AB^T\} = Eig_{\leq 1}\{I_n + B^TA\}
\]

(213)

Since range spaces of \( C_x \) and of \( C_y \) only intersect at the origin, the approximation matrix \( \tilde{C}_x\tilde{C}_y^{-1} \) has \( r \) dominant eigenvalues of order \( O(\rho^{-1}) \) and \( (n - r) \) remaining eigenvalues equal to 1. Using Proposition 1, these \( r \) dominant eigenvalues are given by

\[
Eig_{\leq 1}\left\{\tilde{C}_x\tilde{C}_y^{-1}\right\} = Eig_{\leq 1}\left\{I_r + \left(\rho^{-1}T_x^T - \rho^{-1}T_xT_yA_y^{-1}T_y^T - (I_r + \rho A_y^{-1})T_x^+\right)T_x\right\}
\]

\[
= Eig_{\leq 1}\left\{\rho^{-1}T_x^T - \rho^{-1}T_xT_yA_y^{-1}T_y^T - \rho A_y^{-1}\right\}.
\]

(214)

Let \( \tilde{\Lambda}_{max} \) and \( \tilde{\Lambda}_{min} \), respectively denote the diagonal submatrices of \( \tilde{\Lambda} \) with the \( r \) largest and with the \( r \) smallest eigenvalues. From the definitions in (66) and (206), one can recognize that \( T_x^T T_x = \Lambda_x - \rho I_r \),
while $T_y^T T_y = W_y^T K_{x,y} W_y$, and substituting them in (214) we obtain the estimate of the $r$ largest eigenvalues

$$
\hat{\Lambda}_{\text{max}} = \text{diag} \ E\text{ig}_{\geq 1}\left\{ \hat{C}_y \hat{C}_x^{-1} \right\} 
$$

(215)

$$
= \text{diag} \ E\text{ig}_{\leq 1}\left\{ \rho^{-1} \Lambda_x - I - \rho \Lambda_y^{-1} - \rho^{-1} W_x^T K_{x,y} W_y \Lambda_y^{-1} W_y^T K_{x,y} W_x \right\}
$$

(216)

The relative error between these eigenvalues and the $r$ largest eigenvalues of $\hat{C}_x \hat{C}_y^{-1}$ is of order $O(\rho)$. This is a consequence of the fact that these eigenvalues are $O(\rho)$. Then, the relative error between the dominant eigenvalues of the two matrices can be bounded above by

$$
\left( \sum_{i=1}^r \left( \lambda_i - \hat{\lambda}_i \right)^2 \right)^{\frac{1}{4}} \leq \left\| \hat{C}_x \hat{C}_y^{-1} - \hat{C}_x \hat{C}_y^{-1} \right\|_F \approx O(\rho^0) \approx O(\rho). 
$$

On the other hand, the $r$ smallest eigenvalues of $\hat{\Lambda}$ are the reciprocal of the $r$ dominant eigenvalues of the inverse matrix $(\hat{C}_y^{-\frac{1}{2}} \hat{C}_x \hat{C}_y^{-\frac{1}{2}})^{-1}$, so we can estimate them using essentially the same procedure

$$
\hat{\Lambda}_{\text{min}}^{-1} = \text{diag} \ E\text{ig}_{\leq 1}\left\{ \hat{C}_y \hat{C}_x^{-1} \right\} 
$$

(218)

$$
= \text{diag} \ E\text{ig}_{\leq 1}\left\{ \rho^{-1} C_{y|x}^{(\rho)} \right\}. 
$$

(219)

For a sufficient small value of $\rho > 0$, the dominant contribution to the AB log-det divergence comes from the $r$ largest and $r$ smallest eigenvalues of the matrix pencil $(\hat{C}_x, \hat{C}_y)$, so we obtain the desired approximation

$$
D_{AB}^{(a,b)} (\hat{\Lambda} \parallel I_n) \approx D_{AB}^{(a,b)} (\hat{\Lambda}_{\text{max}} \parallel I_r) + D_{AB}^{(a,b)} (\hat{\Lambda}_{\text{min}} \parallel I_r) 
$$

(220)

$$
= D_{AB}^{(a,b)} (\rho \hat{\Lambda}_{\text{max}} \parallel \rho I_r) + D_{AB}^{(b,a)} (\rho \hat{\Lambda}_{\text{min}}^{-1} \parallel \rho I_r) 
$$

(221)

$$
\approx D_{AB}^{(a,b)} (\rho \hat{\Lambda}_{\text{max}} \parallel \rho I_r) + D_{AB}^{(b,a)} (\rho \hat{\Lambda}_{\text{min}}^{-1} \parallel \rho I_r) 
$$

(222)

$$
= D_{AB}^{(a,b)} (C_{x|y}^{(\rho)} \parallel \rho I_r) + D_{AB}^{(b,a)} (C_{y|x}^{(\rho)} \parallel \rho I_r). 
$$

(223)

Moreover, as $\rho \to 0$, the relative error of this approximation also tends to zero.

**H. Gamma Divergence for Multivariate Gaussian Densities**

Recall that for a given quadratic function $f(x) = -c + b^T x - \frac{1}{2} x^T A x$, where $A$ is an SPD matrix, the integral of $\exp\{ f(x) \}$ with respect to $x$ is given by

$$
\int_{\Omega} e^{-\frac{1}{2} x^T A x + b^T x - c} dx = (2\pi)^{\frac{N}{2}} \det(A)^{-\frac{1}{2}} e^{\frac{1}{2} b^T A^{-1} b - c}. 
$$

(224)

This formula is obtained by evaluating the integral as follows:

$$
\int_{\Omega} e^{-\frac{1}{2} x^T A x + b^T x - c} dx = e^{\frac{1}{2} b^T A^{-1} b - c} \int e^{-\frac{1}{2} x^T A x + b^T x - \frac{1}{2} b^T A^{-1} b} dx
$$

(225)

$$
= e^{\frac{1}{2} b^T A^{-1} b - c} \int_{\Omega} e^{(x - A^{-1} b)^T A (x - A^{-1} b)} dx 
$$

(226)

$$
= e^{\frac{1}{2} b^T A^{-1} b - c} (2\pi)^{\frac{N}{2}} \det(A)^{-\frac{1}{2}}, 
$$

(227)
assuming that $A$ is an SPD matrix, which assures the convergence of the integral and the validity of (224).

The Gamma divergence involves the a product of densities. In the multivariate Gaussian case, this simplifies as

$$
p^\alpha(x)q^\beta(x) = (2\pi)^{-\frac{N}{2}(\alpha+\beta)} \det(P)^{-\frac{\alpha}{2}} \det(Q)^{-\frac{\beta}{2}} \times 
\exp\left\{-\frac{\alpha}{2}(x-\mu_1)^T P^{-1}(x-\mu_1) - \frac{\beta}{2}(x-\mu_2)^T Q^{-1}(x-\mu_2)\right\}
= d \exp\left\{-c + b^T x - \frac{1}{2} x^T A x\right\},
$$

where

$$
A = \alpha P^{-1} + \beta Q^{-1},
$$

$$
b = (\mu_1^T \alpha P^{-1} + \mu_2^T \beta Q^{-1})^T,
$$

$$
c = \frac{1}{2} \mu_1 (\alpha P^{-1}) \mu_1 + \frac{1}{2} \mu_2 (\beta Q^{-1}) \mu_2,
$$

$$
d = (2\pi)^{-\frac{N}{2}(\alpha+\beta)} \det(P)^{-\frac{\alpha}{2}} \det(Q)^{-\frac{\beta}{2}}.
$$

Integrating this product with the help of (224), we obtain

$$
\int_{\Omega} p^\alpha(x)q^\beta(x) dx = d (2\pi)^{\frac{N}{2}} e^{\frac{1}{2} b^T A^{-1} b - c}
= (2\pi)^{\frac{N}{2}(1-(\alpha+\beta))} \det(P)^{-\frac{\alpha}{2}} \det(Q)^{-\frac{\beta}{2}} \det(\alpha P^{-1} + \beta Q^{-1})^{-\frac{1}{2}} \times 
\frac{1}{2} \mu_1^T (\alpha P^{-1} + \beta Q^{-1}) (\alpha P^{-1} + \beta Q^{-1})^{-1} (\mu_1^T \alpha P^{-1} + \mu_2^T \beta Q^{-1})^T \times 
\frac{1}{2} \mu_1 \mu_1 - \frac{1}{2} \mu_2 (\beta Q^{-1}) \mu_2,
$$

provided that $\alpha P^{-1} + \beta Q^{-1}$ is positive definite.

Rearranging the expression in terms of $\mu_1$ and $\mu_2$ yields

$$
\int_{\Omega} p^\alpha(x)q^\beta(x) dx = (2\pi)^{\frac{N}{2}(1-(\alpha+\beta))} \det(P)^{-\frac{\alpha}{2}} \det(Q)^{-\frac{\beta}{2}} \det(\alpha P^{-1} + \beta Q^{-1})^{-\frac{1}{2}} \times 
\frac{1}{2} \mu_1^T [\alpha P^{-1} (\alpha P^{-1} + \beta Q^{-1})^{-1} \alpha P^{-1} - \alpha P^{-1}] \mu_1 \times 
\frac{1}{2} \mu_2^T [\beta Q^{-1} (\alpha P^{-1} + \beta Q^{-1})^{-1} \beta Q^{-1} - \beta Q^{-1}] \mu_2 \times 
\frac{1}{2} \mu_1 (\alpha P^{-1} + \beta Q^{-1})^{-1} \beta Q^{-1} \mu_2.
$$

With the help of the Woodbury matrix identity, we simplify

$$
e^{\frac{1}{2} \mu_1^T [\alpha P^{-1} (\alpha P^{-1} + \beta Q^{-1})^{-1} \alpha P^{-1} - \alpha P^{-1}] \mu_1} = e^{-\frac{1}{2} \mu_1^T (\alpha^{-1} P + \beta^{-1} Q) \mu_1},
$$

$$
e^{\frac{1}{2} \mu_2^T [\beta Q^{-1} (\alpha P^{-1} + \beta Q^{-1})^{-1} \beta Q^{-1} - \beta Q^{-1}] \mu_2} = e^{-\frac{1}{2} \mu_2^T (\alpha^{-1} P + \beta^{-1} Q) \mu_2},
$$

$$
e^{\frac{1}{2} \mu_1^T (\alpha P^{-1} + \beta Q^{-1})^{-1} \beta Q^{-1} \mu_2} = e^{\mu_1^T (\alpha^{-1} P + \beta^{-1} Q) \mu_2},
$$

and hence, arriving at the desired result:
\[
\int_{\Omega} p^{\alpha}(x)q^{\beta}(x)dx = \left(2\pi\right)^{\frac{N}{2}(1-(\alpha+\beta))}det(P)^{-\frac{\alpha}{2}}det(Q)^{-\frac{\beta}{2}}(\alpha+\beta)^{-\frac{N}{2}} \times \\
\det\left(\frac{\alpha}{\alpha+\beta}P^{-1} + \frac{\beta}{\alpha+\beta}Q^{-1}\right)^{-\frac{1}{2}} \times \\
e^{-\frac{\alpha\beta}{2(\alpha+\beta)}(\mu_1-\mu_2)^T\left(\frac{\beta}{\alpha+\beta}P + \frac{\alpha}{\alpha+\beta}Q\right)^{-1}(\mu_1-\mu_2)}. \tag{240}
\]

This formula can be easily particularized to evaluate the integrals
\[
\int_{\Omega} p^{\alpha+\beta}(x)dx = \int_{\Omega} p^{\alpha}(x)p^{\beta}(x)dx \\
= \left(2\pi\right)^{\frac{N}{2}(1-(\alpha+\beta))}det(P)^{-\frac{\alpha}{2}}det(P)^{-\frac{\beta}{2}}det(\alpha P^{-1} + \beta P^{-1})^{-\frac{1}{2}} \times \\
e^{-\frac{\alpha\beta}{2(\alpha+\beta)}(\mu_1-\mu_1)^T\left(\frac{\beta}{\alpha+\beta}P + \frac{\alpha}{\alpha+\beta}P\right)^{-1}(\mu_1-\mu_1)} = \left(2\pi\right)^{\frac{N}{2}(1-(\alpha+\beta))}(\alpha + \beta)^{-\frac{N}{2}}det(\frac{\alpha}{\alpha+\beta}P) \tag{241}
\]
and
\[
\int_{\Omega} q^{\alpha+\beta}(x)dx = \left(2\pi\right)^{\frac{N}{2}(1-(\alpha+\beta))}(\alpha + \beta)^{-\frac{N}{2}}det(\frac{\beta}{\alpha+\beta}Q) \tag{242}
\]

By substituting these integrals into the definition of the Gamma divergence and simplifying, we obtain a generalized closed form formula:
\[
D^{(\alpha,\beta)}_{AC}(p(x)||q(x)) = \frac{1}{\alpha\beta} \log \left( \int_{\Omega} p^{\alpha+\beta}(x)dx \right)^{\frac{\alpha}{\alpha+\beta}} \left( \int_{\Omega} q^{\alpha+\beta}(x)dx \right)^{\frac{\beta}{\alpha+\beta}} \\
= \frac{1}{2\alpha\beta} \log \frac{\det\left(\frac{\alpha}{\alpha+\beta}Q + \frac{\beta}{\alpha+\beta}P\right)}{\det(Q)^{\frac{\alpha}{\alpha+\beta}} \det(P)^{\frac{\beta}{\alpha+\beta}}} \\
+ \frac{1}{2(\alpha+\beta)}(\mu_1-\mu_2)^T\left(\frac{\alpha}{\alpha+\beta}Q + \frac{\beta}{\alpha+\beta}P\right)^{-1}(\mu_1-\mu_2), \tag{243}
\]
which concludes the proof of Theorem 4.

**Conflicts of Interest**

The authors declare no conflict of interest.

**References**


29. Li, R.-C. *Rayleigh Quotient Based Optimization Methods For Eigenvalue Problems*; Summary of Lectures Delivered at Gene Golub SIAM Summer School 2013; Fudan University: Shanghai, China, 2013.


44. PHoff, P.D. Separable covariance arrays via the Tucker product, with applications to multivariate relational data. *Bayesian Anal.* 2011, 6, 179–196.

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).