This file contains the supplementary material for the article “Review of Riemannian distances and divergences, applied to SSVEP-based BCI”.

A Appendix: Distances, divergences and means

The literature on the geometry of symmetric and positive-definite (SPD) matrices describes several distances and divergence that are well suited for the classification of covariance matrices. We present here the main results and the formal definitions of these distances and divergences, along with their associated means.

For I covariance matrices $\{\Sigma_i\}_{i=1\dots I}$, the mean $\bar{\Sigma}$ is the covariance matrix that minimizes the dispersion of the matrices Σ_i. It is written as:

$$\bar{\Sigma} = \mu(\{\Sigma_1, \dots, \Sigma_I\}) = \arg\min_{\Sigma \in \mathcal{M}} \sum_{i=1}^{I} d^p(\Sigma_i, \Sigma) \ , \tag{1}$$

with $p = 2$, $d^p(\cdot,\cdot)$ is a distance and it is a divergence for $p = 1$.

A.1 Euclidean distance

The Euclidean distance is defined as the Frobenius norm of the difference of two matrices:

$$d_{\mathrm{E}}(\Sigma_1, \Sigma_2) = \|\Sigma_1 - \Sigma_2\|_F \ . \tag{2}$$

Using this distance in (1), this defines the arithmetic mean:

$$\bar{\Sigma}_{\mathrm{E}} = \frac{1}{I} \sum_{i=1}^{I} \Sigma_i \ . \tag{3}$$

The arithmetic mean is drawn from a family of power means (Lim and Pálfia 2012; Congedo et al. 2017), defined as the unique solution of:

$$\bar{\Sigma}_t = \frac{1}{I} \sum_{i=1}^{I} \Sigma_i^{1/2} \left(\Sigma_i^{-1/2} \bar{\Sigma}_t \Sigma_i^{-1/2} \right)^t \Sigma_i^{1/2} \ , \ t \in [-1, +1] \ . \tag{4}$$

It could be thus expressed as $\bar{\Sigma}_{\mathrm{E}} = \bar{\Sigma}_{t|t=1}$. From the same family, one can define the *geometric* mean as $\bar{\Sigma}_{t|t\to 0}$, and the *harmonic* mean as $\bar{\Sigma}_{\mathrm{H}} = \bar{\Sigma}_{t|t=-1}$, also derived from the harmonic distance $d_{\mathrm{H}}(\Sigma_1, \Sigma_2) = \|\Sigma_1^{-1} - \Sigma_2^{-1}\|_F = d_{\mathrm{E}}(\Sigma_1^{-1}, \Sigma_2^{-1})$.

The arithmetic mean is a possible choice, as the mean of SPD matrices is a SPD matrix, but it is not suited for two reasons. Firstly, the Euclidean distance (hence the arithmetic mean) is not invariant under inversion, meaning that a matrix and its inverse are not at the same distance from the identity matrix. Secondly, the arithmetic mean is plagued with the *swelling effect*, that affects its determinant. The determinant is a measure of the dispersion of the multivariate variable and it is proportional to the volume of the column space. The arithmetic mean induces a distortion, as the determinant of the mean could be larger the determinants of the Σ_i (Arsigny et al. 2007). It is thus more appropriate to rely on a mean that adapt to the geometry of the SPD matrices.

A.2 Affine-invariant Riemannian distance

Relying on a Riemannian metric, a manifold could be defined for the space of SPD matrices, that is called a Riemannian manifold.

The affine-invariant Riemannian (AIR) distance is defined as the curve length connecting two points on a Riemannian manifold (Pennec et al. 2006) and is defined as:

$$d_{\mathrm{AIR}}(\Sigma_1, \Sigma_2) = \|\mathrm{Log}(\Sigma_1^{-1/2} \Sigma_2 \Sigma_1^{-1/2})\|_F = \left(\sum_{c=1}^{C} \log^2 \lambda_c \right)^{1/2} , \tag{5}$$

where Log is the matrix logarithm and the eigenvalues of $\Sigma_1^{-1/2} \Sigma_2 \Sigma_1^{-1/2}$ are λ_c, $c = 1, \dots, C$. This definition comes from the geodesic equations expressed in the space of SPD matrices.

The mean $\bar{\Sigma}_{\mathrm{AIR}}$ associated to the affine-invariant Riemannian metric is obtained by combining (5) and (1). This could be written as:

$$\sum_{i=1}^{I} \mathrm{Log}(\bar{\Sigma}_{\mathrm{AIR}}^{-1/2} \Sigma_i \bar{\Sigma}_{\mathrm{AIR}}^{-1/2}) = 0 \ .$$

There is no closed form for this equation and it is efficiently solved by a gradient descent algorithm (Fletcher et al. 2004). Even if AIR mean is commonly referred as the geometric mean, it is still an active field of research (Congedo et al. 2017).

This distance is indeed invariant to affine transforms and has also other invariances:

1. *Invariance under congruent transformation, for any invertible matrix W*

   $$d_{\mathrm{AIR}}(\Sigma_1, \Sigma_2) = d_{\mathrm{AIR}}(W^T \Sigma_1 W, W^T \Sigma_2 W) \ ;$$

2. *Invariance under inversion*

   $$d_{\mathrm{AIR}}(\Sigma, \mathbf{I}) = d_{\mathrm{AIR}}(\Sigma^{-1}, \mathbf{I})$$

   implying

   $$d_{\mathrm{AIR}}(\Sigma_1, \Sigma_2) = d_{\mathrm{AIR}}(\Sigma_1^{-1}, \Sigma_2^{-1}) \ .$$

A.3 Log-Euclidean distance

The Log-Euclidean was introduced to lower the computational complexity of affine-invariant metric, while keeping some of its properties. The distance between two SPD matrices is expressed as:

$$\begin{aligned} d_{\mathrm{LE}}(\Sigma_1, \Sigma_2) &= \|\mathrm{Log}(\Sigma_1) - \mathrm{Log}(\Sigma_2)\|_F \\ &= d_{\mathrm{E}}\left(\mathrm{Log}(\Sigma_1), \mathrm{Log}(\Sigma_2)\right) \ , \end{aligned} \tag{6}$$

and $d_{\mathrm{LE}}(\mathbf{I}, \Sigma_1^{-1}\Sigma_2) = d_{\mathrm{AIR}}(\Sigma_1, \Sigma_2)$. The mean associated to the Log-Euclidean distance is defined explicitly as the arithmetic mean in the domain of matrix logarithms:

$$\bar{\Sigma}_{\mathrm{LE}} = \mathrm{Exp}\left(\frac{1}{I} \sum_{i=1}^{I} \mathrm{Log}(\Sigma_i) \right) \ . \tag{7}$$

There is a closed form expression for the Log-Euclidean mean, unlike the AIR mean, yielding an obvious computational advantage. Another advantage is that the Log-Euclidean mean is usually close or equivalent to the AIR mean, with $\operatorname{tr}(\bar{\Sigma}_{\mathrm{LE}}) > \operatorname{tr}(\bar{\Sigma}_{\mathrm{AIR}})$, and share similar properties. The Log-Euclidean and AIR means have the same determinants, that are the equivalent to the geometric mean of the determinants of their matrices (Arsigny et al. 2007):

$$\det \bar{\Sigma}_{\mathrm{LE}} = \det \bar{\Sigma}_{\mathrm{AIR}} = \prod_{i=1}^{I} (\det \Sigma_i)^{1/I} = \exp\left(\frac{1}{I}\sum_{i=1}^{I} \log(\det \Sigma_i)\right) .$$

Also, the Log-Euclidean mean enjoys the same properties than AIR mean with the noticeable difference that Log-Euclidean mean has a similarity-invariance instead of a invariance by congruence.

A.4 Bregman divergences

Divergences have been considered for the computation of mean in applications of clustering and classification of SPD matrices due to the fact that they induce a Riemannian metric (Amari and Cichocki 2010). Consider a strictly convex and differentiable function $f : \mathbb{R} \to \mathbb{R}$; then $f(x) > f(y) + f'(y)(x-y)$ and $f(x) = f(y) + f'(y)(x-y) \Leftrightarrow x = y$ for all $x, y \in \mathbb{R}$. The Bregman divergence (Bregman 1967) is the first-order Taylor expansion of the so-called *seed function* f :

$$D_{\mathrm{f}}(x, y) = f(x) - f(y) - f'(y)(x-y) ,$$

This divergence D_{f} verifies the non-negativity and the identity properties. When the seed function is quadratic, it can also be symmetric. The other properties of D_{f} are reported in Bregman (1967). Geometrically, the Bregman divergence can be seen as the measure of the difference between $f(x)$ and its representation on the plane tangent to f at y, see (Dhillon and Tropp 2007; Nielsen and Nock 2009) for illustrations.

The scalar divergence can be directly adapted to SPD matrices with a seed function f depending on eigenvalues, such as the trace or as the determinant. Depending on the chosen seed function, various divergences can be defined from the Bregman divergence.

The asymmetry of divergences results in a right- and left-sided mean:

$$D_{\mathrm{f}}(\Sigma_1, \Sigma_2) \neq D_{\mathrm{f}}(\Sigma_2, \Sigma_1)$$
$$\Rightarrow \arg\min_{\Sigma\in\mathcal{M}} \sum_{i=1}^{I} D_{\mathrm{f}}(\Sigma_i, \Sigma) \neq \arg\min_{\Sigma\in\mathcal{M}} \sum_{i=1}^{I} D_{\mathrm{f}}(\Sigma, \Sigma_i) .$$

It is usually sufficient to consider a single sided divergence: in this work, only the right-sided divergence and mean are used.

A.4.1 Euclidean divergence

A Bregman divergence naturally follows from the Frobenius norm (Dhillon and Tropp 2007), with $f(x) = x^2$:

$$D_{\mathrm{E}}(\Sigma_1, \Sigma_2) = \frac{1}{2}\|\Sigma_1 - \Sigma_2\|_F^2 . \quad (8)$$

This divergence is equivalent to the square distance in the Euclidean case. The arithmetic mean of SPD matrices is thus equivalent to the Euclidean divergence-based mean, as it could be seen in Eq. (3).

A.4.2 Kullback-Leibler divergence

Using the *Shannon entropy* $f(x) = x \log x$ yields the *Kullback-Leibler* divergence (Kullback and Leibler 1951; Nielsen and Nock 2009; Duchi 2007; Chebbi and Moakher 2012). This divergence is also called the *discrimination information* or the *relative entropy*. Considering two multivariate Gaussian distributions $\mathcal{N}(\mu_1, \Sigma_1)$ and $\mathcal{N}(\mu_2, \Sigma_2)$, with means μ_1 and μ_2 and covariance matrices Σ_1 and Σ_2 is given by Duchi (2007); Chebbi and Moakher (2012):

$$D_{\mathrm{KL}}(\mathcal{N}(\mu_1, \Sigma_1), \mathcal{N}(\mu_2, \Sigma_2)) = \frac{1}{2}\left(-\log\det(\Sigma_2^{-1}\Sigma_1) + \operatorname{tr}(\Sigma_2^{-1}\Sigma_1) - C\right) + \frac{1}{2}\left((\mu_1 - \mu_2)^T \Sigma_2^{-1}(\mu_1 - \mu_2)\right) .$$

But considering that $\mu_1 = \mu_2 = 0$, the Kullback-Leibler divergence between two SPD matrices Σ_1, Σ_2 becomes:

$$D_{\mathrm{KL}}(\Sigma_1, \Sigma_2) = \frac{1}{2}\left(-\log\det(\Sigma_2^{-1}\Sigma_1) + \operatorname{tr}(\Sigma_2^{-1}\Sigma_1) - C\right) . \quad (9)$$

There are two means induced by the Kullback-Leibler divergence: the right-type mean is:

$$\arg\min_{\Sigma\in\mathcal{M}} \sum_{i=1}^{I} D_{\mathrm{KL}}(\Sigma_i, \Sigma) ,$$

and coincides with the arithmetic mean $\bar{\Sigma}_{\mathrm{E}}$ (Nielsen and Nock 2009); and the left-type mean $\operatorname{argmin}_{\Sigma\in\mathcal{M}} \sum_{i=1}^{I} D_{\mathrm{KL}}(\Sigma, \Sigma_i)$ which coincides with the harmonic mean $\bar{\Sigma}_{\mathrm{H}}$ (Moakher and Batchelor 2006).

A.4.3 Log-det divergence

Another function often used in Bregman divergences of symmetric matrices is the *logarithmic barrier* using the seed function $f(x) = -\log(x)$ (Dhillon and Tropp 2007; Cherian et al. 2011; Sra 2016):

$$f(\Sigma) = -\log\det(\Sigma) .$$

The corresponding divergence is called the *log-det* divergence and is given by Dhillon and Tropp (2007):

$$D_{\mathrm{ld}}(\Sigma_1, \Sigma_2) = -\log\det(\Sigma_1\Sigma_2^{-1}) + \operatorname{tr}(\Sigma_1\Sigma_2^{-1}) - C . \quad (10)$$

Interestingly, for centered multivariate Gaussian distributions, log-det divergence is equivalent to the Kullback-Leibler of Eq. 9, since $\det(AB) = \det(BA)$ and $\operatorname{tr}(AB) = \operatorname{tr}(BA)$ for squared matrices A and B.

A.5 Symmetrized Bregman divergences

In some cases, the asymmetry of divergences can be undesirable, and there are two ways to symmetrize the Bregman divergences (Nielsen and Boltz 2011). The Jeffreys-Bregman divergences consider half of the double-sided divergences (Jeffreys 1946; Nielsen and Boltz 2011):

$$D_{\mathrm{JfB}}(\Sigma_1, \Sigma_2) = \frac{1}{2}\left(D_{\mathrm{f}}(\Sigma_1, \Sigma_2) + D_{\mathrm{f}}(\Sigma_2, \Sigma_1)\right) , \quad (11)$$

and the Jensen-Bregman divergences, considering the divergences to the averaged matrix (Nielsen and Boltz 2011):

$$D_{\mathrm{JnB}}(\Sigma_1,\Sigma_2)=\frac{1}{2}\left(D_{\mathrm{f}}(\Sigma_1,\frac{\Sigma_1+\Sigma_2}{2})+D_{\mathrm{f}}(\Sigma_2,\frac{\Sigma_1+\Sigma_2}{2})\right). \tag{12}$$

Various symmetric divergences can be defined from these two ways.

A.5.1 Jeffreys divergence

The *Jeffreys divergence* (Jeffreys 1946), sometimes called J-divergence, is a symmetrized Kullback-Leibler divergence (Jeffreys 1946; Sra 2016):

$$\begin{aligned} D_{\mathrm{J}}(\Sigma_1,\Sigma_2)&=\frac{1}{2}D_{\mathrm{KL}}(\Sigma_1,\Sigma_2)+\frac{1}{2}D_{\mathrm{KL}}(\Sigma_2,\Sigma_1)\\ &=\frac{1}{2}\left(\mathrm{tr}(\Sigma_2^{-1}\Sigma_1)+\mathrm{tr}(\Sigma_1^{-1}\Sigma_2)\right)-C\,. \end{aligned} \tag{13}$$

The mean associated with the symmetrized Kullback-Leibler is easily computed as it coincides with the middle of the geodesic going from the arithmetic mean $\bar{\Sigma}_{\mathrm{E}}$ to the harmonic mean $\bar{\Sigma}_{\mathrm{H}}$ (Moakher and Batchelor 2006):

$$\begin{aligned} \bar{\Sigma}_{\mathrm{J}}&=\bar{\Sigma}_{\mathrm{H}}^{1/2}\left(\bar{\Sigma}_{\mathrm{H}}^{-1/2}\bar{\Sigma}_{\mathrm{E}}\bar{\Sigma}_{\mathrm{H}}^{-1/2}\right)^{1/2}\bar{\Sigma}_{\mathrm{H}}^{1/2}\\ &=\bar{\Sigma}_{\mathrm{E}}^{1/2}\left(\bar{\Sigma}_{\mathrm{E}}^{-1/2}\bar{\Sigma}_{\mathrm{H}}\bar{\Sigma}_{\mathrm{E}}^{-1/2}\right)^{1/2}\bar{\Sigma}_{\mathrm{E}}^{1/2}\,. \end{aligned} \tag{14}$$

A.5.2 S-divergence

The other way to symmetrize the Kullback-Leibler divergence gives the *Jensen-Shannon* divergence (Lin 1991; Briët and Harremoës 2009; Nielsen and Boltz 2011), also called S-divergence (Sra 2016):

$$D_{\mathrm{S}}(\Sigma_1,\Sigma_2)=\frac{1}{2}\left(D_{\mathrm{KL}}(\Sigma_1,\frac{\Sigma_1+\Sigma_2}{2})+D_{\mathrm{KL}}(\Sigma_2,\frac{\Sigma_1+\Sigma_2}{2})\right).$$

The S-divergence is obtained by using the logarithmic barrier function for the positive-definite cone $f(\Sigma)=-\log\det(\Sigma)$:

$$D_{\mathrm{S}}(\Sigma_1,\Sigma_2)=\log\det(\frac{\Sigma_1+\Sigma_2}{2})-\frac{1}{2}\log\det(\Sigma_1\Sigma_2)\,. \tag{15}$$

Despite its symmetry, S-divergence is not a metric: it does not satisfy the triangular inequality criterion. However, its squared root has been shown to be a distance (Sra 2016).

The S-divergence between two SPD matrices corresponds to the Bhattacharyya divergence between them (Bhattacharyya 1943; Sra 2016), and it is also equivalent to a symmetrized log-det divergence (Cherian et al. 2011).

A.6 Weighted divergences

Another family of divergence is defined when the right- and left-sided divergences are mixed in a weighted manner.

A.6.1 Log-det α-divergence

One such family is the α*-divergence* (Nielsen et al. 2014), and D_{f}^{α} can be expressed in terms of Bregman divergence for $\alpha^2\neq 1$ (Chebbi and Moakher 2012):

$$\begin{aligned} D_{\mathrm{f}}^{\alpha}(\Sigma_1,\Sigma_2)=\frac{4}{1-\alpha^2}\Big(&\frac{1-\alpha}{2}D_{\mathrm{f}}\left(\Sigma_1,\frac{1-\alpha}{2}\Sigma_1+\frac{1+\alpha}{2}\Sigma_2\right)\\ &+\frac{1+\alpha}{2}D_{\mathrm{f}}\left(\Sigma_2,\frac{1-\alpha}{2}\Sigma_1+\frac{1+\alpha}{2}\Sigma_2\right)\Big). \end{aligned} \tag{16}$$

To obtain the α-divergences at $\alpha=\pm 1$, we could consider the limit values $\lim_{\alpha\to\pm 1}D_{\mathrm{f}}^{\alpha}$ that yield for the logarithmic-barrier function:

$$\begin{aligned} D_{\mathrm{LD}}^{\alpha}(\Sigma_1,\Sigma_2)=&\frac{4}{1-\alpha^2}\log\det\Big(\frac{1-\alpha}{2}\left(\Sigma_1\Sigma_2^{-1}\right)^{\frac{1+\alpha}{2}}\\ &+\frac{1+\alpha}{2}\left(\Sigma_2\Sigma_1^{-1}\right)^{\frac{1-\alpha}{2}}\Big),\quad \alpha\neq -1,1 \end{aligned} \tag{17}$$

$$D_{\mathrm{LD}}^{-1}(\Sigma_1,\Sigma_2)=\mathrm{tr}\left(\Sigma_1^{-1}\Sigma_2-\mathbf{I}\right)-\log\det\left(\Sigma_1^{-1}\Sigma_2\right)$$
$$D_{\mathrm{LD}}^{1}(\Sigma_1,\Sigma_2)=\mathrm{tr}\left(\Sigma_2^{-1}\Sigma_1-\mathbf{I}\right)-\log\det\left(\Sigma_2^{-1}\Sigma_1\right).$$

With $-1\leq\alpha\leq 1$ the log-det α divergence smoothly changes from the left-sided Kullback-Leibler D_{LD}^{-1} to the right-sided Kullback-Leibler D_{LD}^{1} (Chebbi and Moakher 2012). A specific case is the so-called Bhattacharyya divergence D_{B}, corresponding to D_{LD}^{0} (Bhattacharyya 1943; Chebbi and Moakher 2012; Sra 2016).

The family of α-divergence is a generalization of the Kullback-Leibler divergence obtained by substituting the logarithm function in the Kullback-Leibler divergence with the generalized logarithm function or the α-logarithm (Cichocki and Amari 2010) :

$$\log_{\alpha}=\frac{2}{1-\alpha}\left(x^{\frac{\alpha-1}{2}}-1\right).$$

Varying the value of α in the α-divergence yields various divergences (Cichocki and Amari 2010).

A.7 Wasserstein distance

The Wasserstein distance, also called Earth's mover distance, is a distance between two probability distributions. This is the optimal cost for the transport of one distribution onto the other (Monge 1781; Kantorovitch 1958; Villani 2008).

The ℓ_2-Wasserstein distance between multivariate Gaussian distributions, with means μ_1 and μ_2 and covariance matrices Σ_1 and Σ_2, which are noted $\mathcal{N}(\mu_1,\Sigma_1)$ and $\mathcal{N}(\mu_2,\Sigma_2)$, is reduced to (Bures 1969; Givens and Shortt 1984):

$$d_{W,2}^2(\mathcal{N}(\mu_1,\Sigma_1),\mathcal{N}(\mu_2,\Sigma_2))=$$
$$\|\mu_1-\mu_2\|_2^2+\mathrm{tr}\,\Sigma_1+\mathrm{tr}\,\Sigma_2-2\,\mathrm{tr}\left(\left(\Sigma_1^{1/2}\Sigma_2\Sigma_1^{1/2}\right)^{1/2}\right).$$

Considering that $\mu_1=\mu_2=0$, the Wasserstein distance between two covariance matrices is:

$$d_W(\Sigma_1,\Sigma_2)=\left(\mathrm{tr}\left(\Sigma_1+\Sigma_2-2\left(\Sigma_1^{1/2}\Sigma_2\Sigma_1^{1/2}\right)^{1/2}\right)\right)^{1/2}. \tag{18}$$

The associated mean is computed iteratively, as described in Barbaresco (2011); Agueh and Carlier (2011) and improved in Álvarez-Esteban et al. (2016).

References

Agueh M, Carlier G (2011) Barycenters in the Wasserstein space. SIAM J Math Anal 43(2):904–924

Álvarez-Esteban P, del Barrio E, Cuesta-Albertos J, Matrán C (2016) A fixed-point approach to barycenters in Wasserstein space. Journal of Mathematical Analysis and Applications 441:744–762

Amari SI, Cichocki A (2010) Information geometry of divergence functions. Bulletin of the Polish Academy of Sciences: Technical Sciences 58(1):183–195

Arsigny V, Fillard P, Pennec X, Ayache N (2007) Geometric means in a novel vector space structure on symmetric positive-definite matrices. SIAM J Matrix Anal Appl 29(1):328–347

Barbaresco F (2011) Geometric radar processing based on Fréchet distance: information geometry versus optimal transport theory. In: 2011 12th International Radar Symposium (IRS), pp 663–668

Bhattacharyya A (1943) On a measure of divergence between two statistical populations defined by their probability distribution. Bull Calcutta Math Soc 35:99–109

Bregman LM (1967) The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. USSR computational mathematics and mathematical physics 7(3):200–217

Briët J, Harremoës P (2009) Properties of classical and quantum Jensen-Shannon divergence. Physical review A 79:052311, DOI 10.1103/PhysRevA.79.052311

Bures D (1969) An extension of Kakutani's theorem on infinite product measures to the tensor product of semifinite w*-algebras. Trans Am Math Soc 135:199–212

Chebbi Z, Moakher M (2012) Means of Hermitian positive-definite matrices based on the log-determinant α-divergence function. Linear Algebra Appl 436(7):1872–1889

Cherian A, Sra S, Banerjee A, Papanikolopoulos N (2011) Efficient similarity search for covariance matrices via the Jensen-Bregman LogDet divergence. In: Int Conf Computer Vision, IEEE, pp 2399–2406

Cichocki A, Amari S (2010) Families of alpha-beta-and gamma-divergences: Flexible and robust measures of similarities. Entropy 12(6):1532–1568

Congedo M, Barachant A, Koopaei E (2017) Fixed point algorithms for estimating power means of positive definite matrices. IEEE Trans Signal Process 65:2211–2220

Dhillon IS, Tropp JA (2007) Matrix nearness problems with Bregman divergences. SIAM J Matrix Anal Appl 29(4):1120–1146

Duchi J (2007) Derivations for linear algebra and optimization. Tech. rep., Berkeley

Fletcher PT, Lu C, Pizer SM, Joshi S (2004) Principal geodesic analysis for the study of nonlinear statistics of shape. IEEE Trans Med Imaging 23(8):995–1005

Givens C, Shortt R (1984) A class of Wasserstein metrics for probability distributions. The Michigan Mathematical Journal 31:231–240

Jeffreys H (1946) An invariant form for the prior probability in estimation problems. Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences 186:453–461

Kantorovitch L (1958) On the translocation of masses. Management Science 5:1–4

Kullback S, Leibler R (1951) On information and sufficiency. The Annals of Mathematical Statistics 22:79–86

Lim Y, Pálfia M (2012) Matrix power means and the Karcher mean. J Funct Anal 262(4):1498–1514

Lin J (1991) Divergence measures based on the Shannon entropy. IEEE Trans Inf Theory 37:145–151

Moakher M, Batchelor PG (2006) Symmetric Positive-Definite Matrices: From Geometry to Applications and Visualization, Springer Berlin Heidelberg, Berlin, Heidelberg, chap 17, pp 285–298

Monge G (1781) Mémoire sur la théorie des déblais et des remblais. PhD thesis, De l'Imprimerie Royale

Nielsen F, Boltz S (2011) The Burbea-Rao and Bhattacharyya centroids. IEEE Trans Inf Theory 57:5455–5466

Nielsen F, Nock R (2009) Sided and symmetrized Bregman centroids. IEEE Trans Inf Theory 55(6):2882–2904

Nielsen F, Nock R, Amari S (2014) On clustering histograms with k-means by using mixed α-divergences. Entropy 16(6):3273–3301

Pennec X, Fillard P, Ayache N (2006) A Riemannian Framework for Tensor Computing. Int J Comput Vis 66(1):41–66

Sra S (2016) Positive definite matrices and the S-divergence. Proceedings of the American Mathematical Society 144(7):2787–2797

Villani C (2008) Optimal transport: old and new, vol 338. Springer Science & Business Media