Non-Euclidean Computational Geometry and Its Applications in Machine Learning

Frank Nielsen
Sony Computer Science Laboratories, Inc
Tokyo, Japan

2026

Abstract

This note provides a concise overview of my work in computational geometry and geometric computing on non-Euclidean and information-geometric spaces. It also highlights applications to machine learning, including unsupervised clustering and supervised classification, across a range of domains such as computational statistics, computer vision, medical imaging, and finance.

Since its inception, Artificial Intelligence (AI) builds upon concepts from computational geometry [1] (CG), e.g., the celebrated book "Perceptrons: An Introduction to Computational Geometry" [2] by Marvin Minsky and Seymour Papert (1969). Key supervised learning methods like k-Nearest Neighbor (k-NN rule) classifications [3][4] and support vector machines [5] (SVMs) rely on geometric k-order Voronoi partitions [6] and smallest enclosing ball procedures [7][8], respectively. Computational geometry in Euclidean spaces and more generally in normed ℓ_p-spaces have been thoroughly investigated[1] [9] but geometric computing in other spaces has received comparatively little consideration.

A main focus of my research is to consider computational geometry and geometric computing primitives in non-Euclidean spaces, specially in "information spaces" where points denote *typed objects* (e.g., model parameters like vectors or matrices, probability measures, functions, etc) and consider applications ranging from computational statistics [11] to computer vision [12] or medical imaging [13], just to name a few. In particular, information geometry [14][15] (IG) originally started by studying families of statistical models (parametric or not) with main applications in statistics and mathematical neuroscience [16]. Information geometry highlights a fundamental *dual geometric structure*[2] which can be studied purely from the mathematical point of view. In fact, the mechanism of the information-geometric dual structures was independently found in the specific cases of Hessian manifolds [17] and in affine differential geometry [18] by pure geometers (e.g., Professor Koszul and Professor Nomizu). There is a rich interplay between smooth dissimilarities (called divergences, yokes, or contrast functions) and underlying information-geometric structures [19]. Some information-geometric structures yield canonical divergences: For example, Bregman divergences are canonical divergences of dually flat spaces and logarithmic divergences [20] yield dualistic structures with constant section curvatures, and vice-versa. Thus in our work, we have also proposed new dissimilarity measures (e.g., total Bregman divergence [12], geometric Jensen-Shannon

[1] The lead author of the textbook [9] is Professor Jean-Daniel Boissonnat who advised my PhD [10] (defended in 1996).

[2] Namely, a pair of torsion-free affine connections coupled to a Riemannian metric.

divergence [21, 22], Hölder projective divergence [23], duo Bregman pseudo-divergence [24], statistical Minkowski distances [25], etc), studied their properties, and considered applications in machine learning (ML). Some statistical problems like simplifying statistical mixtures can be tackled as clustering problems [26] (unsupervised learning) of corresponding weighted parameter point sets sitting in information spaces. Also interestingly, some methods to approximate the calculations of dissimilarities rely on tasks of computational geometry: For example, we show how computing upper and lower envelopes [27] of parametric-density function graphs is useful to bound the Kullback-Leibler divergence, α-divergences [28, 29], and more generally Ali-Silvey-Csiszár f-divergences in [30, 31].

I shall report concisely my main contributions to non-Euclidean computational geometry. The contributions are categorized by types of geometry as follows:

Hyperbolic geometry. Hyperbolic geometry is frequently encountered in many scientific areas like (a) in machine learning since it can embed discrete hierarchical structures into continuous hyperbolic spaces with low distortions [32], (b) in information geometry as the Fisher-Rao geometry of location-scale families [14, 33] for example, (c) in computer vision/medical imaging as the geometry of symmetric positive-definite (SPD) matrices equipped with the Riemannian trace metric [34]. We described a practical robust method to calculate the hyperbolic Voronoi diagram of a finite point set as a clipped power diagram in Klein non-conformal ball model [35] which can then be converted and visualized in other conformal models [36] like the Poincaré ball or upper half-space, or the hyperboloid models. The method extends to k-order hyperbolic Voronoi diagrams and we reported the closed-form formula for the constant speed velocity parametrizations of geodesics in Klein model[3] [37]. The hyperbolic Delaunay complex as the dual of the hyperbolic Voronoi diagram was studied in [38]. There is no known closed-form formula for calculating the center of mass in hyperbolic geometry, so we considered alternatively the closed-form formula of Galperin model centroid in [39] and used this primitive to simplify kernel density estimators (KDEs) by clustering methods. Next, we designed a fast iterative geodesic-walk algorithm extending [40] for computing the smallest enclosing ball of finite point sets and theoretically guaranteed its performance [41]. We investigated a novel class of so-called Poincaré probability distributions in the Poincaré disk, study their invariance properties, and provided various formula for measuring their dissimilarities [42]. Computational hyperbolic geometry is very useful in machine learning because hierarchical graph structures can be embedded with low distortions as Delaunay trees [43] but embedding techniques face robustness and numerical problems in practice. We considered the novel problem of hyperbolic embeddings of supervised models in [44] and studied a robust tempered calculus for hyperbolic embeddings [45].

Hilbert/Birkhoff geometry. The Klein model of hyperbolic geometry can be interpreted as a Riemannian manifold equipped with the Klein Riemannian metric [46] or as a Hilbert geometry induced by an open bounded convex domain[4] [47]. When the considered domain is an open ellipsoid, the corresponding Hilbert geometry yields the Cayley-Klein geometry [48, 49]. Hilbert geometry is very-well suited to computational geometry because straight lines are pregeodesics (i.e., unparameterized geodesics) but geodesics may not be unique depending on the domain. Geodesics are unique when the Hilbert domain boundary is smooth. We extended Klein Hilbert geometry to the Siegel matrix ball model [50] and termed this space

[3]Pregeodesics in Klein model of hyperbolic geometry are straight lines in the Cartesian coordinate system.

[4]Thus Hilbert geometry are never Hilbert spaces because vector spaces cannot be Hilbert geometry domains!

the Siegel-Klein model[5] [51]. Many statistical models encountered in machine learning have open convex parameter domains. For example, let us mention (i) the probability simplex of categorical distributions [52] whose Hilbert simplex balls have hexagonal shapes [53], (ii) the elliptope of correlation matrices [52], (iii) the double cone model [54] of symmetric-positive definite matrices [55]. We considered the Hilbert geometry for those open bounded convex domains with various applications in machine learning. The case of simplicial domains (like the probability simplex) is very particular (and unique) in as sense that the Hilbert geometry is isomorphic to a vector space equipped with a polyhedral norm [52]. We compare several geometric structures of the probability simplex for dimension reduction tasks [56] including the Aitchison geometry [57] which is used in compositional data analysis (CoDA, e.g., often used in hyperspectral imaging [58]). Birkhoff geometry is related to Hilbert geometry [59] and the contraction theorem of the Birkhoff distance is often used in machine learning to prove convergence of Sinkhorn-type algorithms in optimal transport [60][61]. Recently, Hilbert geometry gained attention from the computational geometry community: For example, refer to the recent works on Hilbert Voronoi/Delaunay studies [62] and classifiers in Hilbert geometry [63, 64].

Riemannian and Finslerian geometry. Riemannian geometry [65] generalizes Euclidean geometry by introducing a metric tensor field smoothly varying on tangent planes. The metric tensor provides the notion of lengths of vectors and angles between vectors (e.g., thus allowing one to check for orthogonality between vectors). To move a vector from one tangent plane to another tangent plane, one needs to specify explicitly the parallel transport along curves connecting tangents planes. Local notions of curvature and torsions are derived from the connection in general, and geodesic lengths with respect to metric connections provide a metric distance called the Riemannian distance. In Riemannian geometry, the default connection is built from the metric tensor and is called the Levi-Civita (LC) connection. The LC connection is unique and torsion-free: This is the fundamental theorem of Riemannian geometry [65]. In information geometry [14], regular statistical models (i.e., with symmetric positive-definite Fisher information matrix) induce Fisher metrics [66, 67, 68][69] and the domains of statistical models are then interpreted as Fisher-Rao manifolds with the geodesic distances called Fisher-Rao distances (or Rao distances [70]). We designed and analyzed an interactive approximation algorithm for calculating the circumcenter of the smallest enclosing ball on arbitrary Riemannian manifolds [71] with sectional curvature constraints. Although very attractive in statistical applications, the Fisher-Rao distances can be difficult to calculate for some statistical models. For example, there is no known closed-form formula for the Fisher-Rao distance between any two multivariate normal distributions [72] (MVNs) although we can approximate arbitrarily finely this distance [73, 74]. Hilbert geometry with smooth ellipsoid boundaries can be analyzed using Riemannian metrics. In general, Hilbert geometry can be studied from the viewpoint of differential geometry using a Finsler metric [47][75]. We studied the uniqueness of p-means (e.g., medians when $p = 1$, centroids when $p = 2$, and circumcenters when $p \to \infty$) on Finsler manifolds in [76].

Neuromanifold geometry. Stochastic (deep) neural networks (NNs) can be analyzed from the viewpoint of information geometry [77]. Stochasticity can be obtained by adding some Gaus-

[5]The Siegel upper space [50] is a matrix extension of Poincaré upper plane which consists of symmetric complex matrices with SPD imaginary parts.

sian noise in the output or inside the activation functions of NN layers themselves [78]. Training neural networks can be visualized as trajectories on the (usually huge) parameter spaces of NNs which exhibit symmetries and Fisher metric degeneracies. Indeed, the Fisher informatiom matrix (FIM) is only guaranteed to be symmetric semi-positive definite (as a covariance matrix) and the its rank may vary locally. Intrinsic learning is often perform using natural gradient descents [79] (NGDs) or some of its variants but are often not practical because of their time complexity or numerical instability. We proposed several improved methods of NGDs like using local reparameterizations or adding momentum, and carried out experiments in [80, 81]. Natural gradient is related to mirror descent [82][83]. We also introduced a modular view of NN architectures where geometry is defined relatively (and dynamically) using the concept of relative Fisher information matrix [84]. One striking property of deep learning is that despite the huge numbers of parameters of NNs, the stochastic gradient descent (SGD) learning algorithm generalizes well with several phenomena observed during the training stage (e.g., double descent [85]). We studied why overparameterization in NNs work for learning from the viewpoint of minimum description length principle [86] (MDL) and reported an Occam's razor based on a geometric modeling of the parameter spaces as lightlike manifolds [87]. In another work, we considered geodesic latent space regularization for variational autoencoder architectures (VAEs) in [88].

Information geometry/Hessian geometry. At the core of information geometry [14] is the dualistic differential-geometric structure which consists of two torsion-free affine connections coupled to the metric tensor which ensures that the metric is preserved under dual parallel transports [33, 15]. When considering parametric statistical models as manifolds, we can equip those manifolds with Amari dual $\pm\alpha$-connections [14] which are coupled to the Fisher-Rao metric. The dual structures can also be built from smooth divergences [19] (divergence-based information geometry) or convex functions [17] (Bregman manifolds[6] [89] which are Hessian manifolds with global charts, commonly called dually flat spaces in information geometry). Since convex functions remain convex after affine reparameterizations and additional affine terms, there is a gauge freedom [90] on Bregman manifolds which manifests itself by affine invariance of the (geometric) Legendre transform.

From a dualistic information-geometric structure, we can recover the Amari-Chentsov cubic tensor from which we can then build a family of dual geometries called the α-geometries [33]. I considered algorithmic geometry on Bregman manifolds because they admit dual canonical affine coordinate systems where geodesics are straight in the proper coordinate systems and thus well-suited for computational geometry. Moreover, dual parallel transports on Bregman manifolds is path independent because the dual connections have zero curvature. A fundamental property on Bregman manifolds is the dual Pythagoras' theorem [91] which allows one to prove the uniqueness of Bregman information projections [92] met in applications as minimization problems of Bregman divergences.

Let us describe the computational toolbox we developed for Bregman manifolds: First, we considered computing exactly [93] or approximating [94] the smallest enclosing Bregman balls [95]. The Bregman Voronoi diagrams [96] and its dual Bregman Delaunay complexes where studied in [97] where we generalize the paraboloid lifting technique [98, 99] of computational geometry (related to the polarity duality [100] with respect to the paraboloid). See [101]

[6]See the software library named `pyBregMan`, `https://franknielsen.github.io/pyBregMan/index.html`

for a pedagogical video which illustrates the space of Bregman balls and its boolean algebra of union and intersection operations manipulated as intersections of halfspaces. In particular, we proved that the Vapnik-Chervonenkis dimensions[7] of d-variate Bregman balls [102] is $d+1$. The reduction technique of Bregman Voronoi diagrams to power diagrams can further incorporate representational functions to compute representational Voronoi diagrams [103] which allows one to further consider α-divergences and β-divergences as particular examples. Voronoi diagrams with respect to skew Jensen-Voronoi diagrams (including Bregman Voronoi diagrams in limit skewing cases) were studied in [104]. We described Bregman proximity query data-structures (useful for supervised classification with respect to Bregman divergences): Bregman vantage point trees [105] and Bregman ball trees [106].

Soft/hard clustering with Bregman divergences [107] allows one to unify k-means (vector quantization) and expectation-maximization (EM) algorithms for learning locally statistical mixtures. In particular, Bregman k-means can also be interpreted as a classification EM for mixtures of densities of a given exponential family[8], termed k-MLE [111]. We may extend k-MLE to learning mixtures with different exponential families for components (e.g., generalized Gaussians [112]) and reported our implementation for Wishart mixtures in [113]. We studied Bregman sided and Bregman symmetrized centroids in [114]: Symmetrized Bregman centroids are geometrically characterized as the intersection of a mixed bisector with a geodesic yielding a fast numerical approximation scheme. The centroids with respect to conformal divergences is addressed in [115]. We further considered k-means clustering with respect to α-divergences [116] and total Bregman/Jensen divergences [117]. We can also reinterpret symmetrized Bregman divergences as curved Bregman divergences [118]. The particular case of Jeffreys centroid (i.e., symmetrized Kullback-Leibler centroid interpreted as a symmetrized Bregman centroid) is expressed using Lambert W function [119] and the Jeffreys' centroids can be approximated fast [120] using proxy centroids (e.g., Gauss-Bregman inductive mean [121]). Jeffreys divergence between Gaussian mixture models (GMMs) are not known in closed-form but can be approximated efficiently [122]. Information-theoretic weighted clustering unifying various clustering techniques was reported in [123]. More generally, we studied the centroids with respect to skew Jensen divergences [124] (including Bregman sided centroids in limit skewing cases). In general, k-means type algorithms are NP-hard as soon as the dimension is greater or equal than 2. In 1D, we can use dynamic programming (DP) to solve for the optimal interval clustering [125].

Chernoff information [126] (CI) between two densities of an exponential family (CI is commonly used for example when bounding the exponent error in Bayesian hypothesis testing [127] or in information fusion [128]) can be explained from the information-geometric viewpoint as the intersection of a bisector with a dual geodesic [129]. For arbitrary two probability measures, their geometric mixture arc forms an exponential family called a Likelihood ratio exponential family (LREF) [130] and the CI analysis of [129] is generalized in [131]. As a byproduct, this CI study proves that the intersection of a primal bisector with a dual

[7] VC-dimension is used to characterize performance of algorithms in machine learning.

[8] Exponential families [108] are families which maximize entropies given moments constraints. For example, the Gaussian family maximizes the differential entropy under the first and second moments constraints (i.e., a continuous exponential family). Similarly, lattice Gaussian families are discrete exponential families under the first two moment constraints for samples spaces being lattices [109]. Thus exponential families are also called maximal entropy distributions [110].

geodesic is unique: This is a pure information geometric result. More general paths between probability/unnormalized distributions than the arithmetic and geometric mixtures can be considered: For example, the q-power paths was shown to yield useful results in simulated annealing and thermodynamic integration methods [132, 133] often used in statistics and ML.

The cumulant functions of exponential families [134] are Legendre-type convex analytic functions. Their corresponding partition functions are log-convex and thus also convex (because the geometric mean is dominated by the arithmetic mean). The statistical divergences corresponding to the Bregman divergences induced by the cumulants and partition functions of exponential families are the reverse Kullback-Leibler divergence (KLD) and reverse extended KLD between normalized and unnormalized densities [135], respectively.

Another class of statistical models admitting dually flat connections are mixture families [136], i.e., linear combinations of linearly independent (probability) densities. It can be shown that the neg-entropy of such statistical mixtures with prescribed components (called w-mixtures in [137]) are Legendre-type generators which yields Hessian/Bregman manifolds. In general, the neg-entropy of continuous mixtures is not available in closed form: But let us note in passing that there is an interesting case of mixtures of two Cauchy distributions (i.e., mixture family of order 1 parameterized by a scalar weight $w \in (0, 1)$) for which the differential entropy and the Jensen-Shannon divergence admit closed-form formula [138]. Cauchy distributions enjoy the property that their Kullback-Leibler divergence and any other f-divergences [139] is symmetric and can be expressed as scalar functions of their closed-form chi-square distances [140]. The Jensen–Shannon centroid [141] of a set of categorical distributions amount to a Jensen divergence on the mixture parameters (because mixtures are closed by linear combinations) and can be calculated using the convex-concave procedure [142] (i.e., difference of convex function algorithm, DCA [143] for short).

On a Bregman manifold, the Hessian metric can be expressed equivalently as the matrix Hessians of the dual convex potential functions in the respective primal/dual parameterizations. Remarkably, the squared Hessian matrices of the convex function encodes the Euclidean metric in the dual coordinate systems [144], and thus the Riemannian geodesic distances for the squared Hessian matrices are the Euclidean distance expressed in the dual coordinate systems (see [145] for the case of separable potential functions).

A novel dualistic structure on the James' SPD double cone model equipped with a mix of logdet and identity-complementary logdet barrier functions was studied in [75].

In practice, the generator F and/or its inverse gradient $(\nabla F)^{-1}$ may not be available in closed form or is computationally intractable when handling log-normalizers of exponential families (e.g., discrete exponential families of some learning machines like Boltzmann machines). We proposed Monte Carlo information geometry which samples the log-normalizer integrals and yields with high probability a proper Bregman tractable generator [146] (random information geometry). We also consider using NNs to estimate Legendre transforms [147] using the affine gauge freedom [90] of Legendre transform. We currently work on the framework of neural information geometry (NIG) which uses NNs as function approximators to implement information-geometric structures in practice.

Information geometry of f-divergences are studied in [148]. When not available in closed-form, f-divergences can be estimated using Monte Carlo techniques [149]. Invariance and spectral structures of f-divergences between multivariate location-scale families are reported in [150].

Optimal transport geometry. Optimal transport [60] (OT) is a mathematical field that gained a lot of attention in machine learning and computer vision because it allows to flexibly measures dissimilarities between discrete (empirical) distributions, or between continuous and discrete distributions. We proposed the earth mover distance [151] (i.e., with respect to ground distance ℓ_1) between superpixels obtained by oversegmentation for computer vision tasks [152]. Sinkhorn algorithm [153] is one of the most important computational methods for solving OTs through entropic regularization. We considered Tsallis regularized OT (instead of Shannon entropy [154]) with applications to ecological inference tasks [155]. We compared several distances for clustering multivariate time series [156, 157] based on Sklar's theorem, and proposed optimal copula transport in [158]. We gave an approximation method based on the chain rule and marginalization of probabilities in [159]. We proposed a Sinkhorn autoencoder architecture [61]. Kantorovich unregularized OT yields a sparse plan but is time consuming while Sinkhorn entropic-regularized OT is fast but yields dense transport plans. We built a middle ground technique for fast OT approximation algorithm which allows one to control the degree of sparsity based on tempered exponential measures [160] (TEMs) in [161]. Otto proposed a Riemannian geometry by means of a regular Riemannian metric [162] (Otto metric), and information-geometric dualistic structures [163] are studied in the Wasserstein information geometry.

Cone geometry. A domain K of a vector space is a cone if for any $\lambda > 0$, we have $\lambda K \subseteq K$. Cones need not necessarily to be convex and may potentially contain lines. We consider regular cones defined as open convex pointed cones. The most studied cone is the cone of symmetric positive-definite (SPD) matrices. The geometry of cones was studied by Vinberg [164] where he defined a characteristic function χ_K associated to each regular cone K. Since the characteristic function is Legendre-type, we recover a dual Hessian structure [17]. We considered the diffeomorphic embeddings of d-variate normal distributions into $(d+1)$-SPD cone in [165] to define a fast distance between normal distributions. We considered also an equivariant mean on irreducible symmetric cones [166] with an application to k-mean clustering of SPD matrices in [167].

Infinite-dimensional (RKHS) Hilbert space. Bregman already considered (potentially infinite-dimensional) Hilbert metrics in his celebrated paper [168] that gave rise to the so-called (finite-dimensional vector) Bregman divergences. Matrix Bregman divergences with respect to inner products were investigated in [169] and in [170] with an application to portfolio theory. Gaussian measures in infinite-dimensional Hilbert spaces behave very differently from the finite-dimensional cases (e.g., there is no infinite-dimensional Lebesgue measure and the usual notion of covariance matrices is replaced by trace-class operators, etc.). We extend the geometric Jensen-Shannon divergence [21] which admits a closed-form[9] between finite-dimensional Gaussians to the Gaussian measures in a general Hilbert space in [172]. We generalize the maximum mean discrepancy (MMD) in [173] and proposed kernelised functional Bregman divergences using reproducible kernel Hilbert spaces [174] (RKHSs). Functional Bregman divergences have proven useful in Bayesian statistics [175]

[9]The ordinary Jensen-Shannon divergence between Gaussians is not even analytic [171].

and pointwise Bregman divergences [176] are commonly considered, e.g., in rate-distortion view of uncertainty quantification [177].

Euclidean geometry. At infinitesimal scale, manifolds can be locally considered as Euclidean spaces (i.e., extrinsic geometry on tangent planes). Euclidean geometry can also be handled as a particular Riemannian manifold equipped with the Euclidean (Riemannian) metric. My PhD work focused on adaptive computational geometry [10] which generalizes the notion of output-sensitive algorithms: For example, I considered output-sensitive algorithms for computing a given number of maximal layers (Pareto fronts [178]) or convex layers [179] (onion peelings) which finds useful applications in database applications for computing skylines [180]. I designed a paradigm called grouping-and-querying [181] to get output-sensitive algorithms and used this paradigm to calculate the convex hulls of arbitrary objects (i.e., ellipses) on the plane [27]. The concept of adaptive algorithms vs output-sensitive algorithms is well illustrated when calculating the union of n intervals on a line [182]: We report a $O(n \log p)$-time algorithm for calculating the union of intervals where $p \geq c$ is the minimum number of piercing points and c is the number of connected components (in worst-case scenario, $c = 1$ but $p = n - 1$). Thus given a problem, it is open to find new measures varying with instances and design algorithms taking into account those adaptive parameters. For example, we considered data structures for point location using the notion of fatness of objects in [183], etc. Combinatorial piercing/covering problems are such kinds of NP-hard problems with applications in telecommunications where adaptive heuristics can be designed (e.g., finding locations of base transceiver stations [184]).

My Publications

[4] Frank Nielsen. *Introduction to HPC with MPI for Data Science*. Springer, 2016 (cit. on p. 1).

[5] Corinna Cortes and Vladimir Vapnik. "Support-vector networks". In: *Machine learning* 20.3 (1995), pp. 273–297 (cit. on p. 1).

[8] Frank Nielsen and Richard Nock. "Approximating smallest enclosing balls with applications to machine learning". In: *International Journal of Computational Geometry & Applications* 19.05 (2009), pp. 389–414 (cit. on p. 1).

[10] Frank Nielsen. "Algorithmes géométriques adaptatifs". PhD thesis. Université Nice Sophia Antipolis, 1996 (cit. on pp. 1, 8).

[11] Frank Nielsen, Arni SR Srinivasa Rao, and Calyampudi Radhakrishna Rao. *Geometry and statistics*. Vol. 46. Academic Press, 2022 (cit. on p. 1).

[12] Meizhu Liu, Baba C Vemuri, Shun-ichi Amari, and Frank Nielsen. "Shape retrieval using hierarchical total Bregman soft clustering". In: *IEEE Transactions on pattern analysis and machine intelligence* 34.12 (2012), pp. 2407–2419 (cit. on p. 1).

[13] Baba C Vemuri, Meizhu Liu, Shun-Ichi Amari, and Frank Nielsen. "Total Bregman divergence and its applications to DTI analysis". In: *IEEE Transactions on medical imaging* 30.2 (2010), pp. 475–483 (cit. on p. 1).

[15] Frank Nielsen. "The many faces of information geometry". In: *Not. Am. Math. Soc* 69.1 (2022), pp. 36–45 (cit. on pp. 1, 4).

[21] Frank Nielsen. "On the Jensen–Shannon symmetrization of distances relying on abstract means". In: *Entropy* 21.5 (2019), p. 485 (cit. on pp. 2, 7).

[22] Frank Nielsen. "Two Types of Geometric Jensen–Shannon Divergences". In: *Entropy* 27.9 (2025), p. 947 (cit. on p. 2).

[23] Frank Nielsen, Ke Sun, and Stéphane Marchand-Maillet. "On Hölder projective divergences". In: *Entropy* 19.3 (2017), p. 122 (cit. on p. 2).

[24] Frank Nielsen. "Statistical divergences between densities of truncated exponential families with nested supports: Duo Bregman and duo Jensen divergences". In: *Entropy* 24.3 (2022), p. 421 (cit. on p. 2).

[25] Frank Nielsen. "The statistical Minkowski distances: Closed-form formula for Gaussian mixture models". In: *International Conference on Geometric Science of Information*. Springer. 2019, pp. 359–367 (cit. on p. 2).

[26] Vincent Garcia and Frank Nielsen. "Simplification and hierarchical representations of mixtures of exponential families". In: *Signal Processing* 90.12 (2010), pp. 3197–3212 (cit. on p. 2).

[27] Frank Nielsen and Mariette Yvinec. "An output-sensitive convex hull algorithm for planar objects". In: *International Journal of Computational Geometry & Applications* 8.01 (1998), pp. 39–65 (cit. on pp. 2, 8).

[28] Frank Nielsen and Ke Sun. "Combinatorial bounds on the α-divergence of univariate mixture models". In: *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2017, pp. 4476–4480 (cit. on p. 2).

[29] Frank Nielsen. "Generalizing the Alpha-Divergences and the Oriented Kullback–Leibler Divergences with Quasi-Arithmetic Means". In: *Algorithms* 15.11 (2022), p. 435 (cit. on p. 2).

[30] Frank Nielsen and Ke Sun. "Guaranteed bounds on the Kullback–Leibler divergence of univariate mixtures". In: *IEEE Signal Processing Letters* 23.11 (2016), pp. 1543–1546 (cit. on p. 2).

[31] Frank Nielsen and Ke Sun. "Guaranteed bounds on information-theoretic measures of univariate mixtures using piecewise log-sum-exp inequalities". In: *Entropy* 18.12 (2016), p. 442 (cit. on p. 2).

[33] Frank Nielsen. "An elementary introduction to information geometry". In: *Entropy* 22.10 (2020), p. 1100 (cit. on pp. 2, 4).

[35] Frank Nielsen and Richard Nock. "Hyperbolic Voronoi diagrams made easy". In: *International Conference on Computational Science and Its Applications*. IEEE. 2010, pp. 74–80 (cit. on p. 2).

[36] Frank Nielsen and Richard Nock. "Visualizing hyperbolic Voronoi diagrams". In: *Proceedings of the thirtieth annual symposium on Computational geometry*. 2014, pp. 90–91 (cit. on p. 2).

[37] Frank Nielsen and Richard Nock. "The hyperbolic Voronoi diagram in arbitrary dimension". In: *arXiv preprint arXiv:1210.8234* (2012) (cit. on p. 2).

[38] Frank Nielsen. "On Voronoi diagrams on the information-geometric Cauchy manifolds". In: *Entropy* 22.7 (2020), p. 713 (cit. on p. 2).

[39] Olivier Schwander and Frank Nielsen. "Model centroids for the simplification of kernel density estimators". In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2012, pp. 737–740 (cit. on p. 2).

[41] Frank Nielsen and Gaëtan Hadjeres. "Approximating covering and minimum enclosing balls in hyperbolic geometry". In: *International Conference on Geometric Science of Information*. Springer. 2015, pp. 586–594 (cit. on p. 2).

[42] Frank Nielsen and Kazuki Okamura. "Information measures and geometry of the hyperbolic exponential families of Poincaré and hyperboloid distributions". In: *Information Geometry* 7.Suppl 2 (2024), pp. 943–989 (cit. on p. 2).

[44] Richard Nock, Ehsan Amid, Frank Nielsen, Alexander Soen, and Manfred K Warmuth. "Hyperbolic embeddings of supervised models". In: *Advances in Neural Information Processing Systems* 37 (2024), pp. 140865–140902 (cit. on p. 2).

[45] Richard Nock, Ehsan Amid, Frank Nielsen, Alexander Soen, and Manfred K Warmuth. "Tempered Calculus for ML: Application to Hyperbolic Model Embedding". In: *arXiv preprint arXiv:2402.04163* (2024) (cit. on p. 2).

[48] Frank Nielsen, Boris Muzellec, and Richard Nock. "Classification with mixtures of curved Mahalanobis metrics". In: *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2016, pp. 241–245 (cit. on p. 2).

[49] Laëtitia Shao and Frank Nielsen. *Machine Learning with Cayley-Klein metrics*. Tech. rep. Technical report, Ecole Polytechnique, France, 2017 (cit. on p. 2).

[50] Frank Nielsen and Rajendra Bhatia. *Matrix information geometry*. Springer, 2013 (cit. on pp. 2, 3).

[51] Frank Nielsen. "The Siegel–Klein disk: Hilbert geometry of the Siegel disk domain". In: *Entropy* 22.9 (2020), p. 1019 (cit. on p. 3).

[52] Frank Nielsen and Ke Sun. "Clustering in Hilbert's projective geometry: The case studies of the probability simplex and the elliptope of correlation matrices". In: *Geometric Structures of Information*. Springer, 2018, pp. 297–331 (cit. on p. 3).

[53] Frank Nielsen and Laetitia Shao. "On balls in a Hilbert polygonal geometry (multimedia contribution)". In: *33rd International Symposium on Computational Geometry (SoCG 2017)*. Schloss Dagstuhl–Leibniz-Zentrum für Informatik. 2017, pp. 67–1 (cit. on p. 3).

[55] Jacek Karwowski and Frank Nielsen. "Hilbert geometry of the symmetric positive-definite bicone". In: *NeurIPS 2025 Workshop on Symmetry and Geometry in Neural Representations*. 2025 (cit. on p. 3).

[56] Frank Nielsen and Ke Sun. "Non-linear embeddings in Hilbert simplex geometry". In: *Topological, Algebraic and Geometric Learning Workshops 2023*. PMLR. 2023, pp. 254–266 (cit. on p. 3).

[61] Giorgio Patrini, Rianne Van den Berg, Patrick Forre, Marcello Carioni, Samarth Bhargav, Max Welling, Tim Genewein, and Frank Nielsen. "Sinkhorn autoencoders". In: *Uncertainty in Artificial Intelligence*. PMLR. 2020, pp. 733–743 (cit. on pp. 3, 7).

[69] Frank Nielsen. "Cramér-Rao lower bound and information geometry". In: *Connected at Infinity II: A Selection of Mathematics by Indians*. Springer, 2013, pp. 18–37 (cit. on p. 3).

[71] Marc Arnaudon and Frank Nielsen. "On approximating the Riemannian 1-center". In: *Computational Geometry* 46.1 (2013), pp. 93–104 (cit. on p. 3).

[73] Frank Nielsen. "A simple approximation method for the Fisher–Rao distance between multivariate normal distributions". In: *Entropy* 25.4 (2023), p. 654 (cit. on p. 3).

[74] Frank Nielsen. "Approximation and bounding techniques for the Fisher–Rao distances between parametric statistical models". In: *Handbook of Statistics*. Vol. 51. Elsevier, 2024, pp. 67–116 (cit. on p. 3).

[75] Jacek Karwowski and Frank Nielsen. "Geometric structures and deviations on James' symmetric positive-definite matrix bicone domain". In: *arXiv preprint arXiv:2603.02483* (2026) (cit. on pp. 3, 6).

[76] Marc Arnaudon and Frank Nielsen. "Medians and means in Finsler geometry". In: *LMS Journal of Computation and Mathematics* 15 (2012), pp. 23–37 (cit. on p. 3).

[78] Frank Nielsen and Ke Sun. "q-neurons: Neuron activations based on stochastic Jackson's derivative operators". In: *IEEE Transactions on Neural Networks and Learning Systems* 32.6 (2020), pp. 2782–2789 (cit. on p. 4).

[80] Wu Lin, Frank Nielsen, Khan Mohammad Emtiyaz, and Mark Schmidt. "Tractable structured natural-gradient descent using local parameterizations". In: *International Conference on Machine Learning*. PMLR. 2021, pp. 6680–6691 (cit. on p. 4).

[81] Wu Lin, Valentin Duruisseaux, Melvin Leok, Frank Nielsen, Mohammad Emtiyaz Khan, and Mark Schmidt. "Simplifying momentum-based positive-definite submanifold optimization with applications to deep learning". In: *International Conference on Machine Learning*. PMLR. 2023, pp. 21026–21050 (cit. on p. 4).

[83] Andrzej Cichocki, Toshihisa Tanaka, Frank Nielsen, and Sergio Cruces. "Mirror Descent and Exponentiated Gradient Algorithms Using Trace-Form Entropies". In: *Entropy* 27.12 (2025), p. 1243 (cit. on p. 4).

[84] Ke Sun and Frank Nielsen. "Relative Fisher information and natural gradient for learning large modular models". In: *International Conference on Machine Learning*. PMLR. 2017, pp. 3289–3298 (cit. on p. 4).

[87] Ke Sun and Frank Nielsen. "A geometric modeling of Occam's razor in deep learning". In: *Information Geometry* (2025), pp. 1–41 (cit. on p. 4).

[88] Gaëtan Hadjeres, Frank Nielsen, and François Pachet. "GLSR-VAE: Geodesic latent space regularization for variational autoencoder architectures". In: *2017 IEEE symposium series on computational intelligence (SSCI)*. IEEE. 2017, pp. 1–7 (cit. on p. 4).

[89] Frank Nielsen and Alexander Soen. "pyBregMan: A Python library for Bregman Manifolds". In: *arXiv preprint arXiv:2408.04175* (2024) (cit. on p. 4).

[90] Frank Nielsen. "Generalized Legendre Transforms Have Roots in Information Geometry". In: *Entropy* 28.1 (2025), p. 44 (cit. on pp. 4, 6).

[91] Frank Nielsen. "On geodesic triangles with right angles in a dually flat space". In: *Progress in information geometry: theory and applications*. Springer, 2021, pp. 153–190 (cit. on p. 4).

[93] Frank Nielsen and Richard Nock. "On the smallest enclosing information disk". In: *Information Processing Letters* 105.3 (2008), pp. 93–97 (cit. on p. 4).

[94] Richard Nock and Frank Nielsen. "Fitting the smallest enclosing Bregman ball". In: *European Conference on Machine Learning*. Springer. 2005, pp. 649–656 (cit. on p. 4).

[95] Frank Nielsen and Richard Nock. "On approximating the smallest enclosing Bregman balls". In: *Proceedings of the twenty-second annual symposium on Computational geometry*. 2006, pp. 485–486 (cit. on p. 4).

[96] Jean-Daniel Boissonnat, Frank Nielsen, and Richard Nock. "Bregman Voronoi diagrams". In: *Discrete & Computational Geometry* 44.2 (2010), pp. 281–307 (cit. on p. 4).

[97] Frank Nielsen, Jean-Daniel Boissonnat, and Richard Nock. "On Bregman Voronoi diagrams". In: *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '07. New Orleans, Louisiana: Society for Industrial and Applied Mathematics, 2007, pp. 746–755. ISBN: 9780898716245 (cit. on p. 4).

[101] Frank Nielsen, Jean-Daniel Boissonnat, and Richard Nock. "Visualizing Bregman Voronoi diagrams". In: *Proceedings of the twenty-third annual symposium on Computational geometry*. 2007, pp. 121–122 (cit. on p. 4).

[102] Frank Nielsen, Jean-Daniel Boissonnat, and Richard Nock. "Bregman Voronoi diagrams: Properties, algorithms and applications". In: *arXiv preprint arXiv:0709.2196* (2007) (cit. on p. 5).

[103] Frank Nielsen and Richard Nock. "The dual Voronoi diagrams with respect to representational Bregman divergences". In: *2009 Sixth International Symposium on Voronoi Diagrams*. IEEE. 2009, pp. 71–78 (cit. on p. 5).

[104] Frank Nielsen and Richard Nock. "Skew Jensen-Bregman Voronoi diagrams". In: *Transactions on Computational Science XIV: Special Issue on Voronoi Diagrams and Delaunay Triangulation*. Springer, 2011, pp. 102–128 (cit. on p. 5).

[105] Frank Nielsen, Paolo Piro, and Michel Barlaud. "Bregman vantage point trees for efficient nearest neighbor queries". In: *2009 IEEE International Conference on Multimedia and Expo*. IEEE. 2009, pp. 878–881 (cit. on p. 5).

[106] Frank Nielsen, Paolo Piro, and Michel Barlaud. "Tailored Bregman ball trees for effective nearest neighbors". In: *Proceedings of the 25th European Workshop on Computational Geometry (EuroCG)*. 2009, pp. 29–32 (cit. on p. 5).

[108] Frank Nielsen and Vincent Garcia. "Statistical exponential families: A digest with flash cards". In: *arXiv preprint arXiv:0911.4863* (2009) (cit. on p. 5).

[109] Frank Nielsen. "The Kullback–Leibler divergence between lattice Gaussian distributions". In: *Journal of the Indian Institute of Science* 102.4 (2022), pp. 1177–1188 (cit. on p. 5).

[110] Frank Nielsen and Richard Nock. "MaxEnt upper bounds for the differential entropy of univariate continuous distributions". In: *IEEE Signal Processing Letters* 24.4 (2017), pp. 402–406 (cit. on p. 5).

[111] Frank Nielsen. "k-MLE: A fast algorithm for learning statistical mixture models". In: *2012 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE. 2012, pp. 869–872 (cit. on p. 5).

[112] Olivier Schwander, Aurélien J Schutz, Frank Nielsen, and Yannick Berthoumieu. "k-MLE for mixtures of generalized Gaussians". In: *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*. IEEE. 2012, pp. 2825–2828 (cit. on p. 5).

[113] Christophe Saint-Jean and Frank Nielsen. “A new implementation of k-MLE for mixture modeling of Wishart distributions”. In: *International conference on geometric science of information*. Springer. 2013, pp. 249–256 (cit. on p. 5).

[114] Frank Nielsen and Richard Nock. “Sided and symmetrized Bregman centroids”. In: *IEEE transactions on Information Theory* 55.6 (2009), pp. 2882–2904 (cit. on p. 5).

[115] Richard Nock, Frank Nielsen, and Shun-ichi Amari. “On conformal divergences and their population minimizers”. In: *IEEE Transactions on Information Theory* 62.1 (2015), pp. 527–538 (cit. on p. 5).

[116] Frank Nielsen, Richard Nock, and Shun-ichi Amari. “On clustering histograms with k-means by using mixed α-divergences”. In: *Entropy* 16.6 (2014), pp. 3273–3301 (cit. on p. 5).

[117] Frank Nielsen and Richard Nock. “Total Jensen divergences: definition, properties and clustering”. In: *2015 ieee international conference on acoustics, speech and signal processing (icassp)*. IEEE. 2015, pp. 2016–2020 (cit. on p. 5).

[118] Frank Nielsen. “Curved representational Bregman divergences and their applications”. In: *International Conference on Geometric Science of Information*. Springer. 2025, pp. 131–140 (cit. on p. 5).

[119] Frank Nielsen. “Jeffreys centroids: A closed-form expression for positive histograms and a guaranteed tight approximation for frequency histograms”. In: *IEEE Signal Processing Letters* 20.7 (2013), pp. 657–660 (cit. on p. 5).

[120] Frank Nielsen. “Fast Proxy Centers for the Jeffreys Centroid: The Jeffreys–Fisher–Rao Center and the Gauss–Bregman Inductive Center”. In: *Entropy* 26.12 (2024), p. 1008 (cit. on p. 5).

[121] Frank Nielsen. “What is... an inductive mean?” In: *Not. Am. Math. Soc* 70.11 (2023), pp. 1851–1855 (cit. on p. 5).

[122] Frank Nielsen. “Fast approximations of the Jeffreys divergence between univariate Gaussian mixtures via mixture conversions to exponential-polynomial distributions”. In: *Entropy* 23.11 (2021), p. 1417 (cit. on p. 5).

[123] Richard Nock and Frank Nielsen. “On weighting clustering”. In: *IEEE Transactions on Pattern Analysis & Machine Intelligence* 28.08 (2006), pp. 1223–1235 (cit. on p. 5).

[124] Frank Nielsen and Sylvain Boltz. “The Burbea-Rao and Bhattacharyya centroids”. In: *IEEE Transactions on Information Theory* 57.8 (2011), pp. 5455–5466 (cit. on p. 5).

[125] Frank Nielsen and Richard Nock. “Optimal interval clustering: Application to Bregman clustering and statistical mixture learning”. In: *IEEE Signal Processing Letters* 21.10 (2014), pp. 1289–1292 (cit. on p. 5).

[127] Frank Nielsen. “Generalized Bhattacharyya and Chernoff upper bounds on Bayes error using quasi-arithmetic means”. In: *Pattern Recognition Letters* 42 (2014), pp. 25–34 (cit. on p. 5).

[129] Frank Nielsen. “An information-geometric characterization of Chernoff information”. In: *IEEE Signal Processing Letters* 20.3 (2013), pp. 269–272 (cit. on p. 5).

[130] Rob Brekelmans, Frank Nielsen, Alireza Makhzani, Aram Galstyan, and Greg Ver Steeg. “Likelihood ratio exponential families”. In: *arXiv preprint arXiv:2012.15480* (2020) (cit. on p. 5).

[131] Frank Nielsen. “Revisiting Chernoff information with likelihood ratio exponential families”. In: *Entropy* 24.10 (2022), p. 1400 (cit. on p. 5).

[132] Vaden Masrani, Rob Brekelmans, Thang Bui, Frank Nielsen, Aram Galstyan, Greg Ver Steeg, and Frank Wood. “q-paths: Generalizing the geometric annealing path using power means”. In: *Uncertainty in Artificial Intelligence*. PMLR. 2021, pp. 1938–1947 (cit. on p. 6).

[133] Rob Brekelmans and Frank Nielsen. “Variational representations of annealing paths: Bregman information under monotonic embedding”. In: *Information Geometry* 7.1 (2024), pp. 193–228 (cit. on p. 6).

[135] Frank Nielsen. “Divergences Induced by the Cumulant and Partition Functions of Exponential Families and Their Deformations Induced by Comparative Convexity”. In: *Entropy* 26.3 (2024), p. 193. DOI: `10.3390/E26030193`. URL: `https://doi.org/10.3390/e26030193` (cit. on p. 6).

[137] Frank Nielsen and Richard Nock. “On the geometry of mixtures of prescribed distributions”. In: *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2018, pp. 2861–2865 (cit. on p. 6).

[138] Frank Nielsen. “The analytic dually flat space of the mixture family of two prescribed distinct Cauchy distributions”. In: *Handbook of Statistics*. Vol. 46. Elsevier, 2022, pp. 225–256 (cit. on p. 6).

[139] Frank Nielsen and Richard Nock. “On the chi square and higher-order chi distances for approximating f-divergences”. In: *IEEE Signal Processing Letters* 21.1 (2013), pp. 10–13 (cit. on p. 6).

[140] Frank Nielsen and Kazuki Okamura. “On f-divergences between Cauchy distributions”. In: *IEEE Transactions on Information Theory* 69.5 (2022), pp. 3150–3171 (cit. on p. 6).

[141] Frank Nielsen. “On a generalization of the Jensen–Shannon divergence and the Jensen–Shannon centroid”. In: *Entropy* 22.2 (2020), p. 221 (cit. on p. 6).

[144] Frank Nielsen. “Any interior point of a finite interval on the real line can be interpreted as dual Fréchet means”. In: *arXiv preprint arXiv:2511.21173* (2025) (cit. on p. 6).

[145] Erika Gomes-Gonçalves, Henryk Gzyl, and Frank Nielsen. “Geometry and fixed-rate quantization in Riemannian metric spaces induced by separable Bregman divergences”. In: *International Conference on Geometric Science of Information*. Springer. 2019, pp. 351–358 (cit. on p. 6).

[146] Frank Nielsen and Gaëtan Hadjeres. “Monte Carlo information-geometric structures”. In: *Geometric Structures of Information*. Springer, 2018, pp. 69–103 (cit. on p. 6).

[147] Basile Plus-Gourdon and Frank Nielsen. “Neural Legendre-Fenchel transform with Hessian Preconditioning”. In: *arXiv preprint arXiv:2606.09077* (2026) (cit. on p. 6).

[149] Frank Nielsen. *Non-negative Monte Carlo estimation of f-divergences*. 2020 (cit. on p. 6).

[150] Frank Nielsen and Kazuki Okamura. “On the f-divergences between densities of a multivariate location or scale family”. In: *Statistics and Computing* 34.1 (2024), p. 60 (cit. on p. 7).

[152] Sylvain Boltz, Frank Nielsen, and Stefano Soatto. "Earth mover distance on superpixels". In: *2010 IEEE International Conference on Image Processing*. IEEE. 2010, pp. 4597–4600 (cit. on p. 7).

[154] Marco Cuturi. "Sinkhorn distances: Lightspeed computation of optimal transport". In: *Advances in neural information processing systems* 26 (2013) (cit. on p. 7).

[155] Boris Muzellec, Richard Nock, Giorgio Patrini, and Frank Nielsen. "Tsallis regularized optimal transport and ecological inference". In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 31. 1. 2017 (cit. on p. 7).

[156] Gautier Marti, Sébastien Andler, Frank Nielsen, and Philippe Donnat. "Optimal transport vs. Fisher-Rao distance between copulas for clustering multivariate time series". In: *IEEE statistical signal processing workshop (SSP)*. IEEE. 2016, pp. 1–5 (cit. on p. 7).

[157] Gautier Marti, Frank Nielsen, Mikołaj Bińkowski, and Philippe Donnat. "A review of two decades of correlations, hierarchies, networks and clustering in financial markets". In: *Progress in information geometry: Theory and applications* (2021), pp. 245–274 (cit. on p. 7).

[158] Gautier Marti, Frank Nielsen, and Philippe Donnat. "Optimal copula transport for clustering multivariate time series". In: *2016 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE. 2016, pp. 2379–2383 (cit. on p. 7).

[159] Frank Nielsen and Ke Sun. "Chain rule optimal transport". In: *Progress in Information Geometry: Theory and Applications*. Springer, 2021, pp. 191–217 (cit. on p. 7).

[161] Ehsan Amid, Frank Nielsen, Richard Nock, and Manfred K Warmuth. "Optimal transport with tempered exponential measures". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 38. 10. 2024, pp. 10838–10846 (cit. on p. 7).

[162] Felix Otto. "The Geometry of Dissipative Evolution Equations: the Porous Medium Equation". In: *Communications in Partial Differential Equations* 26.1-2 (2001), pp. 101–174 (cit. on p. 7).

[165] Frank Nielsen. "Fisher-Rao and pullback Hilbert cone distances on the multivariate Gaussian manifold with applications to simplification and quantization of mixtures". In: *Topological, Algebraic and Geometric Learning Workshops 2023*. PMLR. 2023, pp. 488–504 (cit. on p. 7).

[167] Gabriel Trindade, Emmanuel Chevallier, André Nicolet, and Frank Nielsen. "Fast Equivariant K-Means on SPD Matrices Using Log-Extrinsic Means". In: *International Conference on Geometric Science of Information*. Springer. 2025, pp. 263–270 (cit. on p. 7).

[170] Richard Nock, Brice Magdalou, Eric Briys, and Frank Nielsen. "Mining matrix data with Bregman matrix divergences for portfolio selection". In: *Matrix Information Geometry*. Springer, 2012, pp. 373–402 (cit. on p. 7).

[172] Hà Quang Minh and Frank Nielsen. "Geometric Jensen-Shannon divergence between Gaussian measures on Hilbert space". In: *International Conference on Geometric Science of Information*. Springer. 2025, pp. 69–79 (cit. on p. 7).

[173] Russell Tsuchida and Frank Nielsen. "Generalising maximum mean discrepancy: kernelised functional Bregman divergences". In: *arXiv preprint arXiv:2604.24047* (2026) (cit. on p. 7).

[177] Ifigeneia Apostolopoulou, Benjamin Eysenbach, Frank Nielsen, and Artur Dubrawski. "A rate-distortion view of uncertainty quantification". In: *Proceedings of the 41st International Conference on Machine Learning*. 2024, pp. 1631–1654 (cit. on p. 8).

[178] Nicolas Dupin, Frank Nielsen, and El-Ghazali Talbi. "Unified polynomial dynamic programming algorithms for p-center variants in a 2D Pareto front". In: *Mathematics* 9.4 (2021), p. 453 (cit. on p. 8).

[179] Frank Nielsen. "Output-sensitive peeling of convex and maximal layers". In: *Information processing letters* 59.5 (1996), pp. 255–259 (cit. on p. 8).

[181] Frank Nielsen. "Grouping and querying: A paradigm to get output-sensitive algorithms". In: *Japanese Conference on Discrete and Computational Geometry*. Springer. 1998, pp. 250–257 (cit. on p. 8).

[182] Frank Nielsen. "Fast stabbing of boxes in high dimensions". In: *Theoretical Computer Science* 246.1-2 (2000), pp. 53–72 (cit. on p. 8).

[183] Alon Efrat, Matthew J Katz, Frank Nielsen, and Micha Sharir. "Dynamic data structures for fat objects and their applications". In: *Computational Geometry* 15.4 (2000), pp. 215–227 (cit. on p. 8).

[184] Patrice Calégari, Frédéric Guidec, Pierre Kuonen, and Frank Nielsen. "Combinatorial optimization algorithms for radio network planning". In: *Theoretical Computer Science* 263.1-2 (2001), pp. 235–245 (cit. on p. 8).

References

[1] Franco P Preparata and Michael I Shamos. *Computational geometry: an introduction.* Springer Science & Business Media, 1983 (cit. on p. 1).

[2] Marvin Minsky and Seymour Papert. *Perceptrons: An Introduction to Computational Geometry.* Cambridge, MA, USA: MIT Press, 1969 (cit. on p. 1).

[3] Thomas Cover and Peter Hart. "Nearest neighbor pattern classification". In: *IEEE transactions on information theory* 13.1 (1967), pp. 21–27 (cit. on p. 1).

[6] Der-Tsai Lee. "On k-nearest neighbor Voronoi diagrams in the plane". In: *IEEE transactions on computers* 100.6 (1982), pp. 478–487 (cit. on p. 1).

[7] Ivor W Tsang, James T Kwok, and Nello Cristianini. "Core vector machines: Fast SVM training on very large data sets". In: *Journal of machine Learning research* 6.4 (2005) (cit. on p. 1).

[9] Jean-Daniel Boissonnat and Mariette Yvinec. *Algorithmic geometry.* Cambridge university press, 1998 (cit. on p. 1).

[14] Shun-ichi Amari. *Differential-geometrical methods in statistics.* Springer Science & Business Media, 1985 (cit. on pp. 1–4).

[16] Shun-ichi Amari. "Dreaming of mathematical neuroscience for half a century". In: *Neural Networks* 37 (2013), pp. 48–51 (cit. on p. 1).

[17] Hirohiko Shima. *The geometry of Hessian structures.* World Scientific, 2007 (cit. on pp. 1, 4, 7).

[18] Katsumi Nomizu and Takeshi Sasaki. *Affine differential geometry: geometry of affine immersions.* Cambridge university press, 1994 (cit. on p. 1).

[19] Shinto Eguchi. "Geometry of minimum contrast". In: *Hiroshima Mathematical Journal* 22.3 (1992), pp. 631–647 (cit. on pp. 1, 4).

[20] Ting-Kam Leonard Wong. "Logarithmic divergences from optimal transport and Rényi geometry". In: *Information Geometry* 1.1 (2018), pp. 39–78 (cit. on p. 1).

[32] Frederic Sala, Chris De Sa, Albert Gu, and Christopher Ré. "Representation tradeoffs for hyperbolic embeddings". In: *International conference on machine learning*. PMLR. 2018, pp. 4460–4469 (cit. on p. 2).

[33] Frank Nielsen. "An elementary introduction to information geometry". In: *Entropy* 22.10 (2020), p. 1100 (cit. on pp. 2, 4).

[34] Maher Moakher and Philipp G Batchelor. "Symmetric positive-definite matrices: From geometry to applications and visualization". In: *Visualization and processing of tensor fields*. Springer, 2006, pp. 285–298 (cit. on p. 2).

[40] Mihai Badoiu and Kenneth L Clarkson. "Smaller core-sets for balls". In: *SODA*. Vol. 3. 2003, pp. 801–802 (cit. on p. 2).

[43] Rik Sarkar. "Low distortion Delaunay embedding of trees in hyperbolic plane". In: *International symposium on graph drawing*. Springer. 2011, pp. 355–366 (cit. on p. 2).

[46] HS Bear. "Part metric and hyperbolic metric". In: *The American Mathematical Monthly* 98.2 (1991), pp. 109–123 (cit. on p. 2).

[47] Athanase Papadopoulos and Marc Troyanov. *Handbook of Hilbert geometry*. European Mathematical Society, 2014 (cit. on pp. 2, 3).

[54] Alan Treleven James. "The variance information manifold and the functions on it". In: *Multivariate Analysis–III*. Elsevier, 1973, pp. 157–169 (cit. on p. 3).

[57] John Aitchison. "The statistical analysis of compositional data". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 44.2 (1982), pp. 139–160 (cit. on p. 3).

[58] Hector Blondel, Lucas Drumetz, and Thierry Chonavel. "Aitchison Geometry on the Simplex for Uncertainty Quantification in Bayesian Hyperspectral Image Unmixing". In: *arXiv preprint arXiv:2603.24108* (2026) (cit. on p. 3).

[59] Bas Lemmens and Roger D Nussbaum. "Birkhoff's version of Hilbert's metric and its applications in analysis". In: *Handbook of Hilbert geometry*. European Mathematical Society-EMS-Publishing House GmbH, 2014, pp. 275–303 (cit. on p. 3).

[60] Peyré Gabriel and Cuturi Marco. "Computational optimal transport with applications to data sciences". In: *Foundations and Trends® in Machine Learning* 11.5-6 (2019), pp. 355–607 (cit. on pp. 3, 7).

[62] Auguste H. Gezalyan, Soo Kim, Carlos Lopez, Daniel Skora, Zofia Stefankovic, and David M. Mount. "Delaunay Triangulations in the Hilbert Metric". In: *19th Scandinavian Symposium and Workshops on Algorithm Theory, SWAT 2024, Helsinki, Finland, June 12-14, 2024*. Ed. by Hans L. Bodlaender. Vol. 294. LIPIcs. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2024, 25:1–25:17. DOI: `10.4230/LIPICS.SWAT.2024.25` (cit. on p. 3).

[63] Aditya Acharya, Auguste H. Gezalyan, Julian Vanecek, David M. Mount, and Sunil Arya. "Support Vector Machines in the Hilbert Geometry". In: *19th International Symposium on Algorithms and Data Structures, WADS 2025, York University, Toronto, Canada, August 11-15, 2025*. Ed. by Pat Morin and Eunjin Oh. Vol. 349. LIPIcs. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2025, 3:1–3:20. DOI: 10.4230/LIPICS.WADS.2025.3 (cit. on p. 3).

[64] Aditya Acharya, Auguste H. Gezalyan, and David M. Mount. "Classifiers in High Dimensional Hilbert Metrics". In: *20th Scandinavian Symposium on Algorithm Theory, SWAT 2026, Copenhagen, Denmark, June 17-19, 2026*. Ed. by Pierre Fraigniaud. Vol. 370. LIPIcs. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2026, 1:1–1:16. DOI: 10.4230/LIPICS.SWAT.2026.1 (cit. on p. 3).

[65] John M Lee. *Introduction to Riemannian manifolds*. Vol. 2. Springer, 2018 (cit. on p. 3).

[66] Harold Hotelling. "Spaces of statistical parameters". In: *Bull. Amer. Math. Soc* 36 (1930). First mention hyperbolic geometry for Fisher-Rao metric of location-scale family, p. 191 (cit. on p. 3).

[67] Calyampudi Radhakrishna Rao. "Information and the Accuracy Attainable in the Estimation of Statistical Parameters". In: *Bulletin of the Calcutta Mathematical Society* 37.3 (1945), pp. 81–91 (cit. on p. 3).

[68] Calyampudi Radhakrishna Rao. "Information and the accuracy attainable in the estimation of statistical parameters". In: *Breakthroughs in statistics*. Springer, 1992, pp. 235–247 (cit. on p. 3).

[70] Colin Atkinson and Ann FS Mitchell. "Rao's distance measure". In: *Sankhyā: The Indian Journal of Statistics, Series A* (1981), pp. 345–365 (cit. on p. 3).

[72] Julianna Pinele, João E Strapasson, and Sueli IR Costa. "The Fisher–Rao distance between multivariate normal distributions: Special cases, bounds and applications". In: *Entropy* 22.4 (2020), p. 404 (cit. on p. 3).

[77] Ovidiu Calin. *Deep learning architectures*. Springer, 2020 (cit. on p. 3).

[79] Shun-ichi Amari. "Natural gradient works efficiently in learning". In: *Neural computation* 10.2 (1998), pp. 251–276 (cit. on p. 4).

[82] Garvesh Raskutti and Sayan Mukherjee. "The information geometry of mirror descent". In: *IEEE Transactions on Information Theory* 61.3 (2015), pp. 1451–1457 (cit. on p. 4).

[85] Marco Loog, Tom Viering, Alexander Mey, Jesse H Krijthe, and David MJ Tax. "A brief prehistory of double descent". In: *Proceedings of the National Academy of Sciences* 117.20 (2020), pp. 10625–10626 (cit. on p. 4).

[86] Peter D Grünwald. *The minimum description length principle*. MIT press, 2007 (cit. on p. 4).

[98] Kevin Q Brown. "Voronoi diagrams from convex hulls". In: *Information processing letters* 9.5 (1979), pp. 223–228 (cit. on p. 4).

[99] Bernard Chazelle, Leo J Guibas, and Der-Tsai Lee. "The power of geometric duality". In: *BIT Numerical Mathematics* 25.1 (1985), pp. 76–90 (cit. on p. 4).

[100] S Chynoweth, MJ Sewell, and Douglas Samuel Jones. "Mesh duality and Legendre duality". In: *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* 428.1875 (1990), pp. 351–377 (cit. on p. 4).

[107] Arindam Banerjee, Srujana Merugu, Inderjit S Dhillon, and Joydeep Ghosh. "Clustering with Bregman divergences". In: *Journal of machine learning research* 6.Oct (2005), pp. 1705–1749 (cit. on p. 5).

[126] Herman Chernoff. "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations". In: *The Annals of Mathematical Statistics* (1952), pp. 493–507 (cit. on p. 5).

[128] Simon J Julier. "An empirical study into the use of Chernoff information for robust, distributed fusion of Gaussian mixture models". In: *2006 9th International Conference on Information Fusion*. IEEE. 2006, pp. 1–8 (cit. on p. 5).

[134] Ole Barndorff-Nielsen. *Information and exponential families*. John Wiley & Sons, 2014 (cit. on p. 6).

[136] Shun-ichi Amari. *Information Geometry and Its Applications*. Applied Mathematical Sciences. Springer Japan, 2016. ISBN: 9784431559771 (cit. on p. 6).

[142] Alan L Yuille and Anand Rangarajan. "The concave-convex procedure". In: *Neural computation* 15.4 (2003), pp. 915–936 (cit. on p. 6).

[143] Pham Dinh Tao and LT Hoai An. "Convex analysis approach to DC programming: theory, algorithms and applications". In: *Acta mathematica vietnamica* 22.1 (1997), pp. 289–355 (cit. on p. 6).

[148] Paul W Vos. "Geometry of f-divergence". In: *Annals of the Institute of Statistical Mathematics* 43.3 (1991), pp. 515–537 (cit. on p. 6).

[151] Yossi Rubner, Carlo Tomasi, and Leonidas J Guibas. "The earth mover's distance as a metric for image retrieval". In: *International journal of computer vision* 40.2 (2000), pp. 99–121 (cit. on p. 7).

[153] Richard Sinkhorn and Paul Knopp. "Concerning nonnegative matrices and doubly stochastic matrices". In: *Pacific Journal of Mathematics* 21.2 (1967), pp. 343–348 (cit. on p. 7).

[160] Ehsan Amid, Richard Nock, and Manfred K Warmuth. "Clustering above exponential families with tempered exponential measures". In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2023, pp. 2994–3017 (cit. on p. 7).

[163] Nihat Ay. "Information geometry of the Otto metric". In: *Information geometry* 8.Suppl 1 (2025), pp. 209–232 (cit. on p. 7).

[164] Ernest B Vinberg. "The structure of the group of automorphisms of a homogeneous convex cone". In: *Trans. Moscow Math. Soc* 13 (1965), pp. 63–93 (cit. on p. 7).

[166] Jacques Faraut and Adam Korányi. *Analysis on symmetric cones*. Oxford university press, 1994 (cit. on p. 7).

[168] Lev M. Bregman. "The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming". In: *USSR computational mathematics and mathematical physics* 7.3 (1967), pp. 200–217 (cit. on p. 7).

[169] Brian Kulis, Mátyás A Sustik, and Inderjit S Dhillon. "Low-rank kernel learning with Bregman matrix divergences". In: *Journal of Machine Learning Research* 10.2 (2009) (cit. on p. 7).

[171] Sumio Watanabe, Keisuke Yamazaki, and Miki Aoyagi. “Kullback information of normal mixture is not an analytic function”. In: *IEICE Technical Report* 2004 (2004), pp. 41–46 (cit. on p. 7).

[174] Nachman Aronszajn. “Theory of reproducing kernels”. In: *Transactions of the American mathematical society* 68.3 (1950), pp. 337–404 (cit. on p. 7).

[175] Béla A Frigyik, Santosh Srivastava, and Maya R Gupta. “Functional Bregman divergence and Bayesian estimation of distributions”. In: *IEEE Transactions on Information Theory* 54.11 (2008), pp. 5130–5139 (cit. on p. 7).

[176] Noboru Murata, Takashi Takenouchi, Takafumi Kanamori, and Shinto Eguchi. “Information geometry of U-Boost and Bregman divergence”. In: *Neural Computation* 16.7 (2004), pp. 1437–1481 (cit. on p. 8).

[180] Wen Jin, Jiawei Han, and Martin Ester. “Mining thick skylines over large databases”. In: *European Conference on Principles of Data Mining and Knowledge Discovery*. Springer. 2004, pp. 255–266 (cit. on p. 8).