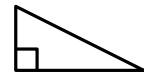


# Taxonomy of principal distances and divergences



## Euclidean geometry



Euclidean distance  
 $d_2(\mathbf{p}, \mathbf{q}) = \sqrt{\sum_i (p_i - q_i)^2}$   
(Pythagoras' theorem circa 500 BC)



## Hyperbolic/spherical geometry

$\rho_{\text{Poincaré}}(z_1, z_2) = \text{arctanh}(\frac{|z_1 - z_2|}{|z_1 - \bar{z}_2|})$   
 $\rho_{\text{sphere}}(p, q) = \arccos(p^T q)$

Hamming distance  
 $(|\{i : p_i \neq q_i\}|)$

Manhattan distance  
 $d_1(\mathbf{p}, \mathbf{q}) = \sum_i |p_i - q_i|$   
(city block-taxi cab)

Minkowski distance ( $L_k$ -norm)  
 $d_k(\mathbf{p}, \mathbf{q}) = \sqrt[k]{\sum_i |p_i - q_i|^k}$   
(H. Minkowski 1864-1909)

Lévy-Prokhorov distance  
 $LP_\rho(p, q) = \inf_{\epsilon > 0} \{p(A) \leq q(A^\epsilon) + \epsilon \forall A \in \mathcal{B}(\mathcal{X})\}$   
 $A^\epsilon = \{y \in \mathcal{X}, \exists x \in A : \rho(x, y) < \epsilon\}$

Quadratic distance  
 $d_Q = \sqrt{(\mathbf{p} - \mathbf{q})^T \mathbf{Q} (\mathbf{p} - \mathbf{q})}$

Fisher information (local entropy)  
 $\mathbf{I}(\theta) = E[(\frac{\partial}{\partial \theta} \ln p(X|\theta))^2]$   
(R. A. Fisher 1890-1962)

Finsler metric tensor  
 $g_{ij} = \frac{1}{2} \partial^2 F^2(x, y) / \partial y^i \partial y^j$

Aitchison distance  
Probability simplex

Chernoff divergence (1952)  
 $C_\alpha(p||q) = -\ln \int p^\alpha q^{1-\alpha} d\mu$   
 $C(p, q) = \max_{\alpha \in (0,1)} C_\alpha(p||q)$

Rényi divergence (1961)  
 $H_\alpha = \frac{1}{\alpha(1-\alpha)} \log \int f^\alpha d\mu$   
 $R_\alpha(p||q) = \frac{1}{\alpha(\alpha-1)} \ln \int p^\alpha q^{1-\alpha} d\mu$   
(additive entropy)

Csiszár'  $f$ -divergence  
 $D_f(p||q) = \int p f(\frac{p}{q}) d\mu$   
(Ali& Silvey 1966, Csiszár 1967)

Dual div. \*-conjugate ( $f^*(y) = yf(1/y)$ )  
 $D_{f^*}(p||q) = D_f(q||p)$

Generalized  $f$ -means duality...  
 $\beta = 1$

Sharma-Mittal entropies  
 $h_{\alpha, \beta}(p) = \frac{1}{1-\beta} \left( (\int p^\alpha d\mu)^{\frac{1-\beta}{1-\alpha}} - 1 \right)$

Non-extensive entropy

Tsallis entropy (1998)  
(Non-additive entropy)  
 $T_\alpha(\mathbf{p}) = \frac{1}{1-\alpha} (\int p^\alpha d\mu - 1)$   
 $T_\alpha(p||q) = \frac{1}{1-\alpha} (1 - \int \frac{p^\alpha}{q^{\alpha-1}} d\mu)$

Wasserstein distances  
 $W_{\alpha, \rho}(p, q) = (\inf_{\gamma \in \Gamma(p, q)} \int \rho(p, q)^\alpha d\gamma(x, y))^\frac{1}{\alpha}$

Regularized optimal transport

Sinkhorn divergence ( $h$ -regularized OT)

Maximum Mean Discrepancy

Integral probability metrics

IPMs

MMD

Log Det divergence  
 $D(\mathbf{P}||\mathbf{Q}) = \langle \mathbf{P}, \mathbf{Q}^{-1} \rangle - \log \det \mathbf{P} \mathbf{Q}^{-1} - \dim \mathbf{P}$

Earth mover distance  
(EMD 1998)

G. Monge

L. Kantorovich

Earth mover distance

Wasserstein distances

Regularized optimal transport

## Statistical geometry

Physics entropy  $JK^{-1}$   
 $-k \int p \log p d\mu$   
(Boltzmann-Gibbs 1878)

Mahalanobis metric (1936)  
 $\mathbf{I}_\Sigma = \sqrt{(\mathbf{p} - \mathbf{q})^T \Sigma^{-1} (\mathbf{p} - \mathbf{q})}$

Haussdorf set distance  
 $d_H(X, Y) = \max\{\sup_x \rho(x, Y), \sup_y \rho(X, y)\}$

Information entropy  
 $H(p) = -\int p \log p d\mu$   
(C. Shannon 1948)

Kullback-Leibler divergence  
 $KL(\mathbf{p}||\mathbf{q}) = \int p \log \frac{p}{q} d\mu = E_p[\log \frac{P}{Q}]$   
(relative entropy, 1951)

Bhattacharya distance (1967)  
 $d(p, q) = -\log \sqrt{\int \sqrt{p} \sqrt{q} d\mu}$

Kolmogorov  
 $K(p||q) = \int |q - p| d\mu$   
(Kolmogorov-Smirnoff max  $|p - q|$ )

Matushita distance (1956)  
 $M_\alpha(p, q) = \sqrt[\alpha]{\int |q^\frac{1}{\alpha} - p^\frac{1}{\alpha}| d\mu}$

Hellinger  
 $H(p||q) = \sqrt{\int (\sqrt{p} - \sqrt{q})^2 d\mu}$   
 $= \sqrt{2(1 - \int \sqrt{pq} d\mu)}$

$\chi^2$  test  
 $\chi^2(p||q) = \int \frac{(q-p)^2}{p} d\mu$   
(K. Pearson, 1857-1936)

Vajda Neyman L. LeCam

Amari  $\alpha$ -divergence (1985)  
 $f_\alpha(x) = \begin{cases} x \log x & \alpha = 1 \\ -\log x & \alpha = -1 \\ \frac{4}{1-\alpha^2} (1-x)^{\frac{1+\alpha}{2}} & -1 < \alpha < 1 \end{cases}$

Fröbenius & Hilbert-Schmidt norm

Quantum entropy  
 $S(\rho) = -k \text{Tr}(\rho \log \rho)$   
(Von Neumann 1927)

Quantum  $f$ -divergences  
(Dénes Petz)

Von Neumann divergence  
 $D(\mathbf{P}||\mathbf{Q}) = \text{Tr}(\mathbf{P}(\log \mathbf{P} - \log \mathbf{Q}) - \mathbf{P} + \mathbf{Q})$

Stein discrepancies

Gromov-Hausdorff distance  
(between compact metric spaces)  
 $d_{GH}(X, Y) = \inf_{\phi_X: X \rightarrow Z, \phi_Y: Y \rightarrow Z} \{\rho_H^Z(\phi_X(X), \phi_Y(Y))\}$   
 $\phi_X, \phi_Y$ : isometric embeddings

Information geometries  
(dually flat)

Quantum & matrix geometry

Fröbenius & Hilbert-Schmidt norm

Quantum entropy

Quantum  $f$ -divergences

Von Neumann divergence

Stein discrepancies

Gromov-Hausdorff distance

Log Det divergence

Earth mover distance

Wasserstein distances

Regularized optimal transport

## Riemannian geometry

Riemannian metric tensor  
 $\int \sqrt{g_{ij} \frac{dx_i}{ds} \frac{dx_j}{ds}} ds$   
(B. Riemann 1826-1866.)

Fisher-Rao distance:  
 $ds^2 = g_{ij} d\theta^i d\theta^j = d\theta^T \mathbf{I}(\theta) d\theta$   
 $\rho_{FR}(p, q) = \min_\gamma \int_0^1 \sqrt{\dot{\gamma}(t) \mathbf{I}(\dot{\gamma}(t))} dt$

Logarithmic divergence  
 $L_{G, \alpha}(\theta_1 : \theta_2) = \frac{1}{\alpha} \log(1 + \alpha \nabla G(\theta_2)^T (\theta_1 - \theta_2)) + G(\theta_2) - G(\theta_1)$

Bregman divergences (1967):  
 $B_F(\theta_1||\theta_2) = F(\theta_1) - F(\theta_2) - (\theta_1 - \theta_2)^T \nabla F(\theta_2)$

Dual div. (Legendre)  $D_{F^*}(\nabla F(\theta_1)||\nabla F(\theta_2)) = D_F(\theta_2||\theta_1)$

Itakura-Saito divergence  
 $IS(\mathbf{p}||\mathbf{q}) = \sum_i (\frac{p_i}{q_i} - \log \frac{p_i}{q_i} - 1)$   
(Burg entropy)

Bregman-Csiszár divergence (1991)  
 $F_\alpha(x) = \begin{cases} x - \log x - 1 & \alpha = 0 \\ x \log x - x + 1 & \alpha = 1 \\ \frac{1}{\alpha(1-\alpha)} (-x^\alpha + \alpha x - \alpha + 1) & 0 < \alpha < 1 \end{cases}$

Generalized Pythagoras' theorem  
(Generalized projection)

Sharma-Mittal entropies

Non-extensive entropy

Tsallis entropy (1998)

Wasserstein distances

Regularized optimal transport

Sinkhorn divergence ( $h$ -regularized OT)

Maximum Mean Discrepancy

Integral probability metrics

IPMs

MMD

Log Det divergence

Earth mover distance

Wasserstein distances

Regularized optimal transport

## Non-Euclidean geometries

Fisher information (local entropy)

Finsler metric tensor

Aitchison distance

Chernoff divergence (1952)

Rényi divergence (1961)

Csiszár'  $f$ -divergence

Dual div. \*-conjugate

Generalized  $f$ -means duality

Sharma-Mittal entropies

Non-extensive entropy

Tsallis entropy (1998)

Wasserstein distances

Regularized optimal transport

Sinkhorn divergence ( $h$ -regularized OT)

Maximum Mean Discrepancy

Integral probability metrics

IPMs

MMD

Log Det divergence

Earth mover distance

Wasserstein distances

Regularized optimal transport

Sinkhorn divergence ( $h$ -regularized OT)

## Cone geometry

Bhattacharya distance (1967)

Kolmogorov

Matushita distance (1956)

Hellinger

$\chi^2$  test

Vajda Neyman L. LeCam

Amari  $\alpha$ -divergence (1985)

Fröbenius & Hilbert-Schmidt norm

Quantum entropy

Quantum  $f$ -divergences

Von Neumann divergence

Stein discrepancies

Gromov-Hausdorff distance

Log Det divergence

Earth mover distance

Wasserstein distances

Regularized optimal transport

Sinkhorn divergence ( $h$ -regularized OT)

Maximum Mean Discrepancy

Integral probability metrics

IPMs

MMD

Log Det divergence

## Symplectic geometry

Bhattacharya distance (1967)

Kolmogorov

Matushita distance (1956)

Hellinger

$\chi^2$  test

Vajda Neyman L. LeCam

Amari  $\alpha$ -divergence (1985)

Fröbenius & Hilbert-Schmidt norm

Quantum entropy

Quantum  $f$ -divergences

Von Neumann divergence

Stein discrepancies

Gromov-Hausdorff distance

Log Det divergence

Earth mover distance

Wasserstein distances

Regularized optimal transport

Sinkhorn divergence ( $h$ -regularized OT)

Maximum Mean Discrepancy

Integral probability metrics

IPMs

MMD

Log Det divergence

## Information geometries

Bhattacharya distance (1967)

Kolmogorov

Matushita distance (1956)

Hellinger

$\chi^2$  test

Vajda Neyman L. LeCam

Amari  $\alpha$ -divergence (1985)

Fröbenius & Hilbert-Schmidt norm

Quantum entropy

Quantum  $f$ -divergences

Von Neumann divergence

Stein discrepancies

Gromov-Hausdorff distance

Log Det divergence

Earth mover distance

Wasserstein distances

Regularized optimal transport

Sinkhorn divergence ( $h$ -regularized OT)

Maximum Mean Discrepancy

Integral probability metrics

IPMs

MMD

Log Det divergence

## Optimal transport geometry

Regularized optimal transport