From Covariance Matrices to Covariance Operators

Data Representation from Finite to Infinite-Dimensional Settings

Hà Quang Minh

Pattern Analysis and Computer Vision (PAVIS)
Istituto Italiano di Tecnologia, ITALY

ISTITUTO ITALIANO
DI TECNOLOGIA

November 29, 2017

Collaborators

- Marco San Biagio
- Vittorio Murino

ISTITUTO ITALIANO
DI TECNOLOGIA

Symmetric Positive Definite (SPD) matrices

$\mathrm{Sym}^{++}(n)$ = set of $n \times n$ SPD matrices

- Have been studied extensively mathematically
- Numerous practical applications
  - Brain imaging (Arsigny et al 2005, Dryden et al 2009, Qiu et al 2015)
  - Computer vision: object detection (Tuzel et al 2008, Tosato et al 2013), image retrieval (Cherian et al 2013), visual recognition (Jayasumana et al 2015), many more
  - Radar signal processing: Barbaresco (2013), Formont et al 2013
  - Machine learning: kernel learning (Kulis et al 2009)

Objectives

- Overview of SPD, especially covariance matrices and their applications in computer vision
- Generalization to infinite-dimensional setting, namely covariance operators, via kernels
  - Nonlinear generalizations of covariance matrices
  - Can achieve substantial gains in practical performance compared to the finite-dimensional setting

Outline

1. Part I: Finite-dimensional setting

   Covariance Matrices and Applications

2. Part II: Infinite-dimensional setting

   Covariance Operators and Applications

Part I - Outline

Finite-dimensional setting

Covariance Matrices and Applications

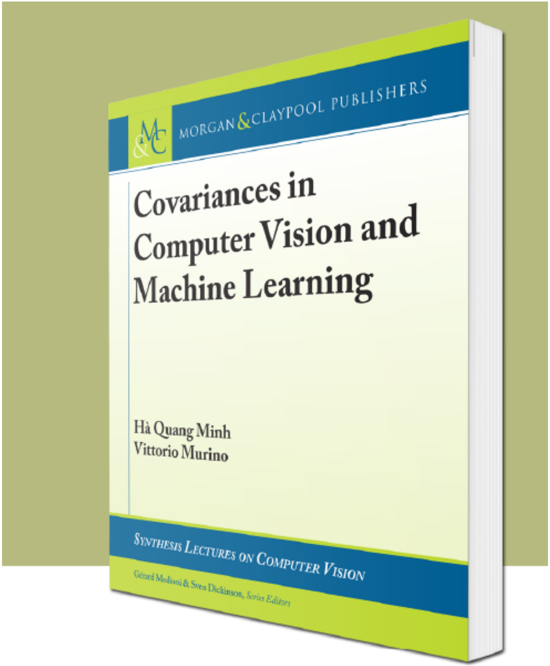
1. Data Representation by Covariance Matrices
2. Geometry of SPD matrices
3. Machine Learning Methods on Covariance Matrices and Applications in Computer Vision

Part II - Outline

Infinite-dimensional setting

Covariance Operators and Applications

1. Data Representation by Covariance Operators
2. Geometry of Covariance Operators
3. Machine Learning Methods on Covariance Operators and Applications in Computer Vision


MORGAN & CLAYPOOL PUBLISHERS
Covariances in Computer Vision and Machine Learning
Hà Quang Minh
Vittorio Murino
Synthesis Lectures on Computer Vision
Gérard Medioni & Sven Dickinson, Series Editors

Part I

Finite-dimensional setting

Part I - Outline

Finite-dimensional setting

Covariance Matrices and Applications

1. Data Representation by Covariance Matrices
2. Geometry of SPD matrices
3. Machine Learning Methods on Covariance Matrices and Applications in Computer Vision

Part I - Outline

Finite-dimensional setting

Covariance Matrices and Applications

1. Data Representation by Covariance Matrices
2. Geometry of SPD matrices
3. Machine Learning Methods on Covariance Matrices and Applications in Computer Vision

Covariance matrix representation of images

- Tuzel, Porikli, Meer (ECCV 2006, CVPR 2006): covariance matrices as region descriptors for images (covariance descriptors)
- Given an image F (or a patch in F), at each pixel, extract a feature vector (e.g. intensity, colors, filter responses etc)
- Each image corresponds to a data matrix $\mathbf{X}$

$$\mathbf{X} = [x_1, \ldots, x_m] = n \times m \quad \text{matrix}$$

where

- m = number of pixels
- n = number of features at each pixel

Covariance matrix representation of images

$\mathbf{X} = [x_1, \ldots, x_m]$ = data matrix of size $n \times m$, with m observations

- Empirical mean vector

$$\mu_{\mathbf{X}} = \frac{1}{m}\sum_{i=1}^{m} x_i = \frac{1}{m}\mathbf{X}\mathbf{1}_m, \quad \mathbf{1}_m = (1, \ldots, 1)^T \in \mathbb{R}^m$$

- Empirical covariance matrix

$$C_{\mathbf{X}} = \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu_{\mathbf{X}})(x_i - \mu_{\mathbf{X}})^T = \frac{1}{m}\mathbf{X}J_m\mathbf{X}^T$$

$$J_m = I_m - \frac{1}{m}\mathbf{1}_m\mathbf{1}_m^T \ = \ \text{centering matrix}$$

Covariance matrix representation of images

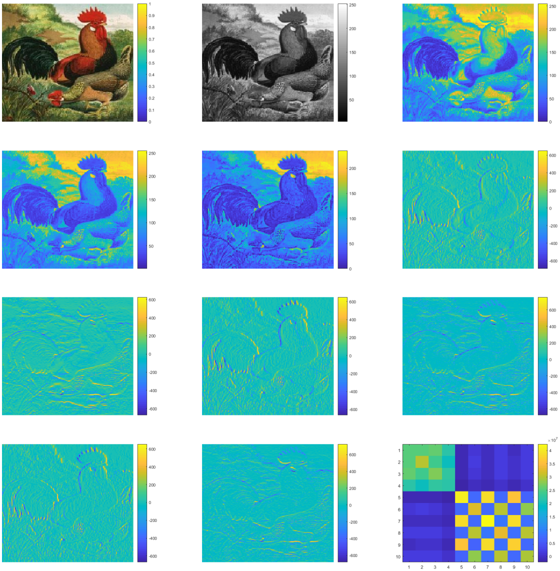
Image F $\Rightarrow$ Data matrix $\mathbf{X}$ $\Rightarrow$ Covariance matrix $C_{\mathbf{X}}$

- Each image is represented by a covariance matrix
- Example of image features

$$\mathbf{f}(x,y) = \left[I(x,y), R(x,y), G(x,y), B(x,y), \left|\frac{\partial R}{\partial x}\right|, \left|\frac{\partial R}{\partial y}\right|, \left|\frac{\partial G}{\partial x}\right|, \left|\frac{\partial G}{\partial y}\right|, \left|\frac{\partial B}{\partial x}\right|, \left|\frac{\partial B}{\partial y}\right|\right]$$

at pixel location (x,y)

Example



Covariance matrix representation - Properties

- Encode linear correlations (second order statistics) between image features
- Flexible, allowing the fusion of multiple and different features
  - Handcrafted features, e.g. colors and SIFT
  - Convolutional features
- Compact
- Robust to noise

Covariance matrix representation - generalization

- Covariance representation for video: e.g. Guo et al (AVSS 2010), Sanin et al (WACV 2013)
  - Employ features that capture temporal information, e.g. optical flow
- Covariance representation for 3D point clouds and 3D shapes: e.g. Fehr et al (ICRA 2012, ICRA 2014), Tabias et al (CVPR 2014), Hariri et al (Pattern Recognition Letters 2016)
  - Employ geometric features e.g. curvature, surface normal vectors

Statistical interpretation

Representing an image by a covariance matrix

is essentially equivalent to

Representing an image by a Gaussian probability density ρ in $\mathbb{R}^n$ with mean zero

Features extracted are random observations of a n-dimensional random vector with probability density ρ

Statistical interpretation

$X = (X_1, \ldots, X_n)$ = random vector in $\mathbb{R}^n$ with probability density function ρ

- Mean vector

$$\mu = \mathbb{E}(X)$$

- Covariance matrix

$$C = \mathbb{E}[(X - \mu)(X - \mu)^T]$$

Statistical interpretation

- Variance of the random variable X_i

$$C_{ii} = \mathbb{E}[(X_i - \mu_i)^2]$$

- Covariance between the random variables X_i and X_j

$$C_{ij} = \mathrm{cov}(X_i, X_j) = \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)]$$

- Correlation between X_i and X_j

$$\mathrm{corr}(X_i, X_j) = \begin{cases} \frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}}, & i \neq j \\ 1, & i = j \end{cases}$$

Statistical interpretation

- Multivariate Gaussian probability density $\rho = \mathcal{N}(\mu, C)$, with covariance matrix $C \in \mathrm{Sym}^{++}(n)$

$$\rho(x) = \frac{1}{\sqrt{(2\pi)^n \det(C)}} \exp\left(-\frac{1}{2}(x-\mu)^T C^{-1}(x-\mu)\right)$$

- If $\mu = 0$, then ρ is completely determined by the covariance matrix C

Empirical covariance matrix - MLE estimate

$\mathbf{X} = [x_1, \dots, x_m]$, x_i's = IID samples drawn from $\mathcal{N}(\mu, C)$

- Log-likelihood function

$$\log \mathcal{L}(\mu, C|\mathbf{X}) = -\frac{mn}{2}\log(2\pi) - \frac{m}{2}\log\det(C) - \frac{1}{2}\sum_{i=1}^{m}(x_i - \mu)^T C^{-1}(x_i - \mu).$$

- MLE estimates of μ and C

$$\mu_{\mathbf{X}} = \frac{1}{m}\sum_{i=1}^{m} x_i, \quad C_{\mathbf{X}} = \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu_{\mathbf{X}})(x_i - \mu_{\mathbf{X}})^T$$

Empirical covariance matrix - Unbiased estimate

- MLE estimate of μ is unbiased

$$\mathbb{E}(\mu_{\mathbf{x}}) = \mu$$

- MLE estimate of C is biased

$$\mathbb{E}(C_{\mathbf{x}}) = \frac{m-1}{m} C \neq C$$

- Unbiased estimate

$$\tilde{C}_{\mathbf{x}} = \frac{m}{m-1} C_{\mathbf{x}} = \frac{1}{m-1} \mathbf{X} J_m \mathbf{X}^T$$

Empirical covariance matrix - MLE vs. unbiased

$C_{\mathbf{X}} = \frac{1}{m}\mathbf{X}J_m\mathbf{X}^T$ = MLE estimate, $\tilde{C}_{\mathbf{X}} = \frac{1}{m-1}\mathbf{X}J_m\mathbf{X}^T$ = unbiased estimate

- In the Gaussian case, in terms of MSE (mean square error)

$$MSE(C_{\mathbf{X}}) = \mathbb{E}||C_{\mathbf{X}} - C||_F^2 < MSE(\tilde{C}_{\mathbf{X}}) = \mathbb{E}||\tilde{C}_{\mathbf{X}} - C||_F^2$$

  since $C_{\mathbf{X}}$ has smaller variance
- Both estimates are consistent: MSE $\to 0$ as $m \to \infty$
- The difference (by a factor $\frac{m-1}{m}$) diminishes as m becomes large
- Both estimates have been used in practice

Empirical covariance matrices - Regularization

- $C_{\mathbf{X}}$ is only guaranteed to be positive semi-definite
- $C_{\mathbf{X}}$ can be ill-conditioned
- For $C_{\mathbf{X}}$ to be positive definite/well-conditioned, need to use regularization in general
- Diagonal loading, widely used, readily generalizable to infinite-dimensional setting

$$C_{\mathbf{X}} + \gamma I \quad \gamma > 0$$

- More generally, shrinkage estimators, Ledoit and Wolf (Journal of Multivariate Analysis, 2004)

$$\hat{C}_{\mathbf{X}} = (1-\rho)C_{\mathbf{X}} + \rho\nu I, \quad 0 \leq \rho \leq 1, \nu > 0$$

Encoding the mean

- Some recent work also employ the mean μ in the representation, e.g. Wang et al (CVPR 2016), Li et al (PAMI 2017)
- Current presentation: assuming zero mean

Data Representation by Covariance Matrices

- Representation of images by covariance matrices and its generalizations
- Statistical interpretation

After Data Representation by Covariance Matrices

For practical applications, e.g.
classification, clustering, regression etc

- We need notions of distances/similarities between covariance matrices

Part I - Outline

Finite-dimensional setting

Covariance Matrices and Applications

1. Data Representation by Covariance Matrices
2. Geometry of SPD matrices
3. Machine Learning Methods on Covariance Matrices and Applications in Computer Vision

Geometry of SPD matrices - Outline

Three different views of the set of SPD matrices

1. Subset of Euclidean space
   - Euclidean metric

- Invariances and Interpretations

2. Riemannian manifold
   - Affine-invariant Riemannian metric
   - Log-Euclidean metric
3. Convex cone
   - Bregman divergences

Three different views of the set of SPD matrices

1. Subset of Euclidean space
    - Euclidean metric

- Invariances and Interpretations

2. Riemannian manifold
    - Affine-invariant Riemannian metric
    - Log-Euclidean metric
3. Convex cone
    - Bregman divergences

$$\mathrm{Mat}(n) \text{ = set of real } n \times n \text{ matrices}$$

- $(\mathrm{Mat}(n), +, \cdot)$ is a vector space, under standard matrix addition $(+)$ and scalar multiplication $(\cdot)$

$$\dim(\mathrm{Mat}(n)) = n^2$$

Euclidean metric

- Frobenius inner product between $A = (a_{ij})_{i,j=1}^n$, $B = (b_{ij})_{i,j=1}^n$

$$\langle A, B \rangle_F = \mathrm{tr}(A^T B) = \sum_{i,j=1^n} a_{ij} b_{ij}$$

- Frobenius norm of an $n \times n$ matrix $A = (a_{ij})_{i,j=1}^n$

$$||A||_F^2 = \mathrm{tr}(A^T A) = \sum_{i,j=1}^n a_{ij}^2$$

- Euclidean (Frobenius) distance between two $n \times n$ matrices A, B

$$d_E(A, B) = ||A - B||_F$$

Euclidean metric

In the Euclidean metric, each $n \times n$ matrix A is equivalent to its vectorized version $\mathrm{vec}(A) \in \mathbb{R}^{n^2}$

$$\langle A, B \rangle_F = \langle \mathrm{vec}(A), \mathrm{vec}(B) \rangle$$

$$||A||_F = ||\mathrm{vec}(A))||$$

$$d_E(A, B) = ||A - B||_F = ||\mathrm{vec}(A) - \mathrm{vec}(B)||$$

Euclidean metric

$\mathrm{Sym}(n)$ = set of real, symmetric $n \times n$ matrices

- $\mathrm{Sym}(n)$ is a vector subspace of $\mathrm{Mat}(n)$

$$\dim(\mathrm{Sym}(n)) = \frac{n(n+1)}{2}$$

$\mathrm{Sym}^{++}(n)$ = set of real SPD $n \times n$ matrices

- $\mathrm{Sym}^{++}(n)$ is an open subset in $\mathrm{Sym}(n) \subset \mathrm{Mat}(n)$
- $\mathrm{Sym}^{++}(n)$ automatically inherits the Euclidean metric on $\mathrm{Mat}(n)$

Euclidean metric - Properties

- Simple to implement and efficient to compute
- Essentially treats matrices as vectors, without taking into account their inherent structures
- Not intrinsic to $\mathrm{Sym}^{++}(n)$
- May lead to swelling effect (mean of a set of SPD matrices might have larger determinant than the component matrices, Arsigny et al 2007)
- $(\mathrm{Sym}^{++}(n), d_E)$ is an incomplete metric space: Cauchy sequences, $\{A_n\}$, $||A_n - A_m||_F$ arbitrarily small, may not converge

Euclidean metric - Properties

- Unitary (orthogonal) invariance: $CC^T = I \Longleftrightarrow C^{-1} = C^T$

$$d_E(CAC^{-1}, CBC^{-1}) = d_E(A, B)$$

- Corresponds to e.g. rotation invariance of Euclidean distance in $\mathbb{R}^n$

Three different views of the set of SPD matrices

1. Subset of Euclidean space
    - Euclidean metric

- Invariances and Interpretations

2. Riemannian manifold
    - Affine-invariant Riemannian metric
    - Log-Euclidean metric
3. Convex cone
    - Bregman divergences

X = random vector in $\mathbb{R}^n$

- Affine transformation

$$X \to \Gamma_{A,b}(X) = AX + b, \quad A \in \mathrm{Mat}(n) \text{ invertible }, b \in \mathbb{R}^n$$

- Transformed mean vector

$$\mu_{\Gamma_{A,b}(X)} = A\mu_X + b$$

- Transformed covariance matrix

$$C_{\Gamma_{A,b}(X)} = AC_XA^T$$

Invariances and Interpretations

d = distance or divergence on $\mathrm{Sym}^{++}(n)$
X, Y = random vectors in $\mathbb{R}^n$

$$X \to AX + b, \quad Y \to AY + b$$

- Affine invariance

$$d(AC_X A^T, AC_Y A^T) = d(C_X, C_Y)$$

- Scale invariance: $A = \sqrt{s}I$, $s > 0$, $b = 0$

$$d(sC_X, sC_Y) = d(C_X, C_Y)$$

Invariances and Interpretations

d = distance or divergence on $\mathrm{Sym}^{++}(n)$
X, Y = random vectors in $\mathbb{R}^n$

$$X \to AX + b, \quad Y \to AY + b$$

- Unitary (orthogonal) invariance: $AA^T = I \Longleftrightarrow A^T = A^{-1}$, $b = 0$

$$d(AC_XA^{-1}, AC_YA^{-1}) = d(C_X, C_Y)$$

e.g. A is a rotation matrix

Invariances and Interpretations

d = distance or divergence on $\mathrm{Sym}^{++}(n)$
X, Y = random vectors in $\mathbb{R}^n$

- Invariance under inversion

$$d(C_X^{-1}, C_Y^{-1}) = d(C_X, C_Y)$$

- For two Gaussian densities $\rho_1 = \mathcal{N}(0, C_X)$ and $\rho_2 = \mathcal{N}(0, C_Y)$

$$d(\rho_1, \rho_2) = d(C_X^{-1}, C_Y^{-1}) = d(C_X, C_Y)$$

- $d(\rho_1, \rho_2)$ can be equally measured via either the distance/divergence between the corresponding covariance matrices or precision matrices.

Summary - Invariances and Interpretations

- Affine invariance
- Scale invariance
- Unitary (orthogonal) invariance
- Invariance under inversion
- The corresponding data transformations

Geometry of SPD matrices - Outline

Three different views of the set of SPD matrices

1. Subset of Euclidean space
   - Euclidean metric

- Invariances and Interpretations

2. Riemannian manifold
   - Affine-invariant Riemannian metric
   - Log-Euclidean metric
3. Convex cone
   - Bregman divergences

Smooth manifold

$\mathcal{M}$ = n-dimensional manifold

- Locally Euclidean $(\mathbb{R}^n)$
- Generalizations of two-dimensional regular surfaces in $\mathbb{R}^3$
- Examples
  - $\mathbb{R}^n$
  - Open subsets of $\mathbb{R}^n$
  - n-dimensional unit sphere

$$\mathbb{S}^{n-1} = \{x \in \mathbb{R}^n \ : \ x_1^2 + \cdots + x_n^2 = 1\}$$

Smooth manifold - tangent vectors and tangent spaces

- Smooth curve $\gamma : (a,b) \to \mathbb{R}^3$, $\gamma'(t) \neq 0$

$$\text{Tangent vector at } t_0 : V = \gamma'(t_0)$$

- Regular surface $\mathcal{S} \subset \mathbb{R}^3$:
    - Tangent vector at a point $P \in \mathcal{S}$

$$V = \gamma'(0), \quad \gamma : (-\epsilon, \epsilon) \to \mathcal{S}, \gamma(0) = P$$

    - Tangent plane $T_P(\mathcal{S})$

$$T_P(\mathcal{S}) = \{\text{all tangent vectors at } P\}$$
$$\dim(T_P(\mathcal{S})) = 2 \quad \text{(2-dimensional plane)}$$

Smooth manifold - tangent vectors and tangent spaces

$\mathcal{M}$ = n-dimensional smooth manifold

- Tangent vector: generalization of tangent vectors on regular surfaces (e.g. Do Carmo 1992, Jost 2008)
- Tangent space $T_P(\mathcal{M})$ at a point $P \in \mathcal{M}$

$$T_P(\mathcal{M}) = \{\text{all tangent vectors at } P\}$$
$$\dim(T_P(\mathcal{M})) = n \quad (n\text{-dimensional vector space})$$

Riemannian metric and Riemannian manifold

Riemannian manifold $\mathcal{M}$ = smooth manifold $\mathcal{M}$ + Riemannian metric

- Riemannian metric: a way to define distances on manifolds
- Riemannian metric: a family of inner products on the tangent spaces $T_P(\mathcal{M})$

$$\langle V, W \rangle_P, \quad V, W \in T_P(\mathcal{M})$$

  that varies smoothly with P

- Length of a vector V in the tangent space $T_P(\mathcal{M})$

$$||V||_P = \sqrt{\langle V, V \rangle_P}$$

Riemannian metric and Riemannian distance

A, B = two points on the manifold $\mathcal{M}$
$\gamma : [a, b] \to \mathcal{M}$ = smooth curve joining A,B, $\gamma(a) = A$, $\gamma(b) = B$

- $\gamma'(t)$ is a tangent vector on the tangent space $T_{\gamma(t)}(\mathcal{M})$
- Length of the curve γ

$$L(\gamma) = \int_a^b ||\gamma'(t)||_{\gamma(t)} dt$$

- Riemannian distance between two points A and B on the manifold

$$d(A, B) = \inf\{L(\gamma) : \gamma(a) = A, \gamma(b) = B\}$$

- $(\mathcal{M}, d)$ is a metric space (satisfying positivity, symmetry, and triangle inequality)

Geodesics

- Curves of zero acceleration (e.g. Do Carmo 1992)
- Generalization of straight lines in $\mathbb{R}^n$
- Constant speed

$$\frac{d}{dt}||\gamma'(t)||_{\gamma(t)} = 0$$

- Geodesically complete manifold:
  - $(\mathcal{M}, d)$ is a complete metric space
  - Geodesic of shortest length: $\exists$ a geodesic γ_{AB} joining any two points A, B, with

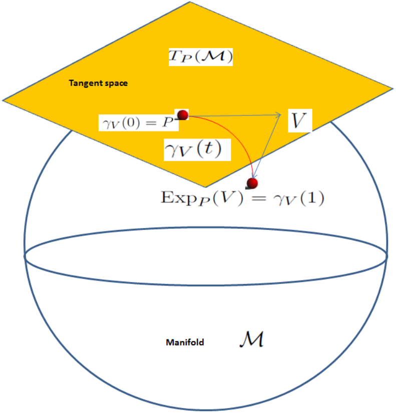
$$L(\gamma_{AB}) = d(A, B)$$

Geodesics - Example

Riemannian exponential map

$$\mathrm{Exp}_P : T_P(\mathcal{M}) \to \mathcal{M}$$

Moving along the manifold, starting from point $P \in \mathcal{M}$, in the direction $V \in T_P(\mathcal{M})$



Geodesics - Example

Geodesics may not be unique

Curvature

- Sectional curvature: For a two-dimensional subspace $\Sigma_P \subset T_P(\mathcal{M})$ and a basis $\{X, Y\}$ in Σ_P

$$K(\Sigma_P) = K_P(X, Y) = \frac{\langle R(X, Y)X, Y\rangle_P}{||X||_P^2||Y||_P^2 - \langle X, Y\rangle_P^2}$$

- R: Riemannian curvature tensor
- Measures how much a manifold deviates from Euclidean space
- Generalization of Gaussian curvature on regular surfaces in $\mathbb{R}^3$
- Euclidean space $\mathbb{R}^n$: $K = 0$ (flat)
- Sphere $\mathbb{S}^{n-1}$: $K = 1$

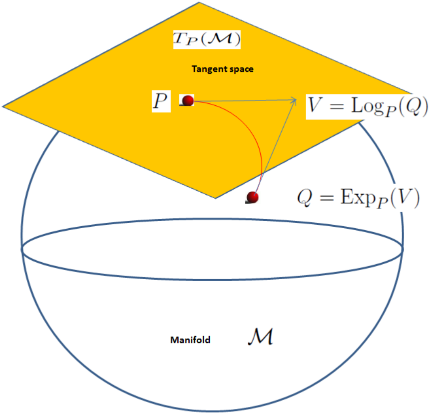
Curvature - Nonpositive curvature

- Cartan-Hadamard manifold: simply-connected, geodesically complete, nonpositive curvature
- Geodesics between any two points are unique
- $\mathrm{Sym}^{++}(n)$ under both the affine-invariant Riemannian metric and Log-Euclidean metric
- The exponential map $\mathrm{Exp}_P : T_P(\mathcal{M}) \to \mathcal{M}$ is a diffeomorphism, i.e. it is a bijection, smooth, with smooth inverse

$$\mathrm{Log}_P : \mathcal{M} \to T_P(\mathcal{M})$$

Riemannian logarithm map

$$\mathrm{Log}_P : \mathcal{M} \to T_P(\mathcal{M})$$



Riemannian metric and Riemannian manifold

Riemannian manifold $\mathcal{M}$ = smooth manifold $\mathcal{M}$ + Riemannian metric

- Every smooth manifold $\mathcal{M}$ admits a Riemannian metric
- Riemannian metrics are not unique
- Each Riemannian metric defines a different distance on $\mathcal{M}$

Geometry of SPD matrices - Outline

Three different views of the set of SPD matrices

1. Subset of Euclidean space
   - Euclidean metric

- Invariances and Interpretations

2. Riemannian manifold
   - Affine-invariant Riemannian metric
   - Log-Euclidean metric
3. Convex cone
   - Bregman divergences

Smooth manifold of SPD matrices

$\mathrm{Sym}^{++}(n)$ = symmetric, positive definite $n \times n$ matrices

- Smooth manifold of dimension $\frac{n(n+1)}{2}$
- Tangent space $T_P(\mathrm{Sym}^{++})(n) \cong \mathrm{Sym}(n)$ = vector space of symmetric matrices

Affine-invariant Riemannian metric

- Has been studied extensively in mathematics
- Siegel (1943), Mostow (1955), Pennec et al (2006), Bhatia (2007), Moakher and Zéraï (2011), Bini and Iannazzo (2013)

Affine-invariant Riemannian metric

- Riemannian metric: On the tangent space $T_P(\mathrm{Sym}^{++}(n)) \cong \mathrm{Sym}(n)$, the inner product $\langle\ ,\ \rangle_P$ is

$$\langle V, W\rangle_P = \langle P^{-1/2}VP^{-1/2}, P^{-1/2}WP^{-1/2}\rangle_F$$
$$= \mathrm{tr}(P^{-1}VP^{-1}W)$$
$$P \in \mathrm{Sym}^{++}(n), \quad V, W \in \mathrm{Sym}(n)$$

- Affine-invariance

$$\langle CAC^T, CBC^T\rangle_{CPC^T} = \langle A, B\rangle_P$$

for any invertible $n \times n$ matrix C

Affine-invariant Riemannian metric

- Geodesically complete Riemannian manifold, nonpositive curvature
- Unique geodesic joining $A, B \in \mathrm{Sym}^{++}(n)$

$$\gamma_{AB}(t) = A^{1/2}(A^{-1/2}BA^{-1/2})^t A^{1/2}$$
$$\gamma_{AB}(0) = A, \quad \gamma_{AB}(1) = B$$

- Riemannian (geodesic) distance

$$d_{\mathrm{aiE}}(A, B) = ||\log(A^{-1/2}BA^{-1/2})||_F$$

where $\log(A)$ is the principal logarithm of A

$$A = UDU^T = U\mathrm{diag}(\lambda_1, \ldots, \lambda_n)U^T$$
$$\log(A) = U\log(D)U^T = U\mathrm{diag}(\log\lambda_1, \ldots, \log\lambda_n)U^T$$

Affine-invariant Riemannian distance - Properties

- Affine-invariance

$$d_{\mathrm{aiE}}(CAC^T, CBC^T) = d_{\mathrm{aiE}}(A, B), \quad \text{any } C \text{ invertible}$$

- Scale invariance: $C = \sqrt{s}I$, $s > 0$,

$$d_{\mathrm{aiE}}(sA, sB) = d_{\mathrm{aiE}}(A, B)$$

- Unitary (orthogonal) invariance: $CC^T = I \Longleftrightarrow C^{-1} = C^T$

$$d_{\mathrm{aiE}}(CAC^{-1}, CBC^{-1}) = d_{\mathrm{aiE}}(A, B)$$

Affine-invariant Riemannian distance - Properties

- Invariance under inversion

$$d_{\mathrm{aiE}}(A^{-1}, B^{-1}) = d_{\mathrm{aiE}}(A, B)$$

- $(\mathrm{Sym}^{++}(n), d_{\mathrm{aiE}})$ is a complete metric space

Connection with Fisher-Rao metric

- Close connection with Fisher-Rao metric in information geometry (e.g. Amari 1985, 2016)
- For two multivariate Gaussian probability densities $\rho_1 \sim \mathcal{N}(\mu, C_1)$, $\rho_2 \sim \mathcal{N}(\mu, C_2)$

$$d_{\mathrm{aiE}}(C_1, C_2) = 2(\text{Fisher-Rao distance between } \rho_1 \text{ and } \rho_2)$$

Affine-invariant Riemannian distance - Complexity

- For two matrices $A, B \in \mathrm{Sym}^{++}(n)$

$$d^2_{\mathrm{aiE}}(A, B) = ||\log(A^{-1/2}BA^{-1/2})||_F^2 = \sum_{k=1}^{n}(\log \lambda_k)^2$$

where $\{\lambda_k\}_{k=1}^n$ are the eigenvalues of

$$A^{-1/2}BA^{-1/2} \quad \text{or equivalently} \quad A^{-1}B$$

- Matrix inversion, SVD, eigenvalue computation all have computational complexity $O(n^3)$
- Therefore $d_{\mathrm{aiE}}(A, B)$ has computational complexity $O(n^3)$

Affine-invariant Riemannian distance - Complexity

For a set $\{A_i\}_{i=1}^N$ of N SPD matrices, consider computing all the pairwise distances

$$d_{\text{aiE}}(A_i, A_j) = ||\log(A_i^{-1/2} A_j A_i^{-1/2})||_F, \quad 1 \leq i, j \leq N$$

- The matrices A_i, A_j are all coupled together
- The computational complexity required is $O(N^2 n^3)$
- This is very large when N is large

Three different views of the set of SPD matrices

1. Subset of Euclidean space
   - Euclidean metric
2. Riemannian manifold
   - Affine-invariant Riemannian metric
   - Log-Euclidean metric
3. Convex cone
   - Bregman divergences

Log-Euclidean metric

- Log-Euclidean metric as a Riemannian metric
- Log-Euclidean distance as an approximation of the affine-invariant Riemannian distance
- Log-Euclidean vs. Euclidean

Log-Euclidean metric

- Arsigny, Fillard, Pennec, Ayache (SIAM Journal on Matrix Analysis and Applications 2007)
- Another Riemannian metric on $\mathrm{Sym}^{++}(n)$
- Much faster to compute than the affine-invariant Riemannian distance on large sets of matrices
- Can be used to define many positive definite kernels on $\mathrm{Sym}^{++}(n)$

Log-Euclidean metric

Riemannian metric: On the tangent space $T_P(\mathrm{Sym}^{++}(n))$

$$\langle V, W\rangle_P = \langle D\log(P)(V), D\log(P)(W)\rangle_F$$
$$P \in \mathrm{Sym}^{++}(n), \quad V, W \in \mathrm{Sym}(n)$$

where

- $D\log$ is the Fréchet derivative of the function $\log : \mathrm{Sym}^{++}(n) \to \mathrm{Sym}(n)$
- $D\log(P) : \mathrm{Sym}(n) \to \mathrm{Sym}(n)$ is a linear map
- Explicit knowledge of $\langle\ ,\ \rangle_P$ is not necessary for computing geodesics and Riemannian distances

Log-Euclidean metric

- Unique geodesic joining $A, B \in \mathrm{Sym}^{++}(n)$

$$\gamma_{AB}(t) = \exp[(1-t)\log(A) + t\log(B)]$$

- Riemannian (geodesic) distance

$$d_{\mathrm{logE}}(A, B) = ||\log(A) - \log(B)||_F$$

Log-Euclidean distance - Complexity

- For two matrices $A, B \in \mathrm{Sym}^{++}(n)$

$$d_{\mathrm{logE}}(A, B) = ||\log(A) - \log(B)||_F$$

- The computation of the $\log$ function, requiring an SVD, has computational complexity $O(n^3)$
- Therefore $d_{\mathrm{logE}}(A, B)$ has computational complexity $O(n^3)$

Log-Euclidean distance - Complexity

For a set $\{A_i\}_{i=1}^N$ of N SPD matrices, consider computing all the pairwise distances

$$d_{\mathrm{logE}}(A_i, A_j) = ||\log(A_i) - \log(A_j)||_F, \quad 1 \leq i, j \leq N$$

- The matrices A_i, A_j are all uncoupled
- The computational complexity required is $O(Nn^3)$
- This is much faster than the affine-invariant Riemannian distance d_{aiE} when N is large

Log-Euclidean vector space

- Arsigny et al (2007): Log-Euclidean metric is a bi-invariant Riemannian metric associated with the Lie group operation

$$\odot : \mathrm{Sym}^{++}(n) \times \mathrm{Sym}^{++}(n) \to \mathrm{Sym}^{++}(n)$$

$$A \odot B = \exp(\log(A) + \log(B)) = B \odot A$$

- Bi-invariance: for any $C \in \mathrm{Sym}^{++}(n)$

$$d_{\mathrm{logE}}[(A \odot C), (B \odot C)] = d_{\mathrm{logE}}[(C \odot A), (C \odot B)] = d_{\mathrm{logE}}(A, B)$$

Log-Euclidean vector space

- Arsigny et al (2007): scalar multiplication operation

$$\circledast : \mathbb{R} \times \mathrm{Sym}^{++}(n) \to \mathrm{Sym}^{++}(n)$$

$$\lambda \circledast A = \exp(\lambda \log(A)) = A^{\lambda}$$

- $(\mathrm{Sym}^{++}(n), \odot, \circledast)$ is a vector space, with $\odot$ acting as vector addition and $\circledast$ acting as scalar multiplication
- $\mathrm{Sym}^{++}(n)$ under the Log-Euclidean metric is a Riemannian manifold with zero curvature

- Vector space isomorphism

$$\log : (\mathrm{Sym}^{++}(n), \odot, \circledast) \to (\mathrm{Sym}(n), +, \cdot)$$
$$A \to \log(A)$$

- The vector space $(\mathrm{Sym}^{++}(n), \odot, \circledast)$ is not a subspace of the Euclidean vector space $(\mathrm{Sym}(n), +, \cdot)$

Log-Euclidean inner product space

- Log-Euclidean inner product (Li, Wang, Zuo, Zhang, ICCV 2013)

$$\langle A, B\rangle_{\mathrm{logE}} = \langle \log(A), \log(B)\rangle_F$$
$$||A||_{\mathrm{logE}} = ||\log(A)||_F$$

- Log-Euclidean inner product space

$$(\mathrm{Sym}^{++}(n), \odot, \circledast, \langle\ ,\ \rangle_{\mathrm{logE}})$$

- Log-Euclidean distance

$$d_{\mathrm{logE}}(A, B) = ||\log(A) - \log(B)||_F = ||(A \odot B^{-1})||_{\mathrm{logE}}$$

Kernels with Log-Euclidean metric

- Positive definite kernels on $\mathrm{Sym}^{++}(n)$ defined with the Log-Euclidean inner product $\langle\ ,\ \rangle_{\mathrm{logE}}$ and norm $||\ ||_{\mathrm{logE}}$
- Polynomial kernels

$$\begin{aligned} K(A,B) &= (\langle A,B\rangle_{\mathrm{logE}} + c\rangle)^d \\ &= (\langle \log(A), \log(B)\rangle_F + c)^d, \quad d \in \mathbb{N}, c \geq 0 \end{aligned}$$

- Gaussian and Gaussian-like kernels

$$\begin{aligned} K(A,B) &= \exp(-\frac{1}{\sigma^2}||(A \odot B^{-1})||^p_{\mathrm{logE}}), \quad 0 < p \leq 2 \\ &= \exp(-\frac{1}{\sigma^2}||\log(A) - \log(B)||^p_F) \end{aligned}$$

Log-Euclidean metric

- Log-Euclidean metric as a Riemannian metric
- Log-Euclidean distance as an approximation of the affine-invariant Riemannian distance
- Log-Euclidean vs. Euclidean

Log-Euclidean distance as an approximation of the affine-invariant Riemannian distance

Affine-invariant Riemannian metric

- Riemannian exponential map: $\mathrm{Exp}_P : T_P(\mathrm{Sym}^{++}(n)) \to \mathrm{Sym}^{++}(n)$

$$\mathrm{Exp}_P(V) = P^{1/2}\exp(P^{-1/2}VP^{-1/2})P^{1/2}$$

- Riemannian logarithm map: $\mathrm{Log}_P : \mathrm{Sym}^{++}(n) \to T_P(\mathrm{Sym}^{++}(n))$

$$\mathrm{Log}_P(A) = P^{1/2}\log(P^{-1/2}AP^{-1/2})P^{1/2} \quad A \in Sym^{++}(n)$$

- At the identity matrix I

$$\mathrm{Exp}_I(V) = \exp(V), \quad \mathrm{Log}_I(A) = \log(A)$$

Log-Euclidean distance as an approximation of the affine-invariant Riemannian distance

- Affine-invariant Riemannian metric on the tangent space $T_P(\mathrm{Sym}^{++}(n))$

$$\langle V, W\rangle_P = \langle P^{-1/2}VP^{-1/2}, P^{-1/2}WP^{-1/2}\rangle_F$$

$$V, W \in \mathrm{Sym}(n)$$

- On the tangent space $T_I(\mathrm{Sym}^{++}(n))$

$$\langle V, W\rangle_I = \langle V, W\rangle_F, \quad ||V - W||_I = ||V - W||_F$$

Log-Euclidean distance as an approximation of the affine-invariant Riemannian distance

- At the identity matrix I

$$||\log(A) - \log(B)||_F = ||\mathrm{Log}_I(A) - \mathrm{Log}_I(B)||_I$$

- Log-Euclidean distance is obtained by projecting A, B onto the tangent space at the identity $T_I(\mathrm{Sym}^{++}(n))$
- This viewpoint does not capture the intrinsic nature of the Log-Euclidean metric

Log-Euclidean metric

- Log-Euclidean metric as a Riemannian metric
- Log-Euclidean distance as an approximation of the affine-invariant Riemannian distance
- Log-Euclidean vs. Euclidean

Log-Euclidean vs. Euclidean

- Euclidean: $\mathrm{Sym}^{++}(n)$ is an open subset in the Euclidean vector space $(\mathrm{Sym}(n), +, \cdot)$
- Log-Euclidean: $(\mathrm{Sym}^{++}(n), \odot, \circledast)$ is a vector space
- The two vector space structures are fundamentally different
- The differences between the metrics can be seen via the invariance properties

Log-Euclidean vs. Euclidean

Unitary (orthogonal) invariance $CC^T = I \Longleftrightarrow C^T = C^{-1}$

- Euclidean distance

$$\begin{aligned} d_E(CAC^{-1}, CBC^{-1}) &= ||CAC^{-1} - CBC^{-1}||_F \\ &= ||A - B||_F = d_E(A, B) \end{aligned}$$

- Log-Euclidean distance

$$\begin{aligned} d_{\mathrm{logE}}(CAC^{-1}, CBC^{-1}) &= ||\log(CAC^{-1}) - \log(CBC^{-1})||_F \\ &= ||\log(A) - \log(B)||_F = d_{\mathrm{logE}}(A, B) \end{aligned}$$

Log-Euclidean vs. Euclidean

- Log-Euclidean distance is scale-invariant

$$\begin{aligned} d_{\mathrm{logE}}(sA, sB) &= ||\log(sA) - \log(sB)||_F \\ &= ||\log(A) - \log(B)||_F = d_{\mathrm{logE}}(A, B) \end{aligned}$$

- Euclidean distance is not scale-invariant

$$d_E(sA, sB) = s||A - B||_F = sd_E(A, B)$$

Log-Euclidean vs. Euclidean

- Log-Euclidean distance is inversion-invariant

$$
\begin{aligned}
d_{\mathrm{logE}}(A^{-1}, B^{-1}) &= ||\log(A^{-1}) - \log(B^{-1})|| \\
&= ||-\log(A) + \log(B)||_F = d_{\mathrm{logE}}(A, B)
\end{aligned}
$$

- Euclidean distance is not inversion-invariant

$$
\begin{aligned}
d_E(A^{-1}, B^{-1}) &= ||A^{-1} - B^{-1}||_F \\
&\neq ||A - B||_F = d_E(A, B)
\end{aligned}
$$

Log-Euclidean vs. Euclidean

As metric spaces

- $(\mathrm{Sym}^{++}(n), d_E)$ is incomplete
- $(\mathrm{Sym}^{++}(n), d_{\mathrm{logE}})$ is complete

Log-Euclidean vs. Euclidean

Summary of comparison

- The two metrics are fundamentally different
- Euclidean metric is extrinsic to $\mathrm{Sym}^{++}(n)$
- Log-Euclidean metric is intrinsic to $\mathrm{Sym}^{++}(n)$
- The vector space structures are fundamentally different
- They have different invariance properties

Geometry of SPD matrices - Outline

Three different views of the set of SPD matrices

1. Subset of Euclidean space
   - Euclidean metric
2. Riemannian manifold
   - Affine-invariant Riemannian metric
   - Log-Euclidean metric
3. Convex cone
   - Bregman divergences

Bregman divergences

- Ω = convex subset in $\mathbb{R}^n$
- $\phi : \Omega \to \mathbb{R}$ = differentiable, strictly convex function
- Bregman divergence on Ω (Bregman, 1967)

$$B_\phi(x, y) = \phi(x) - \phi(y) - \langle \nabla \phi(y), x - y \rangle$$

- Divergence properties

$$B_\phi(x, y) \geq 0,$$
$$B_\phi(x, y) = 0 \Longleftrightarrow x = y$$

Bregman divergences

More generally, ϕ induces a parametrized family of divergences $\{d_\phi^\alpha\}$ (Zhang, Neural Computation, 2004)

$$d_\phi^\alpha(x,y) = \frac{4}{1-\alpha^2}\left[\frac{1-\alpha}{2}\phi(x) + \frac{1+\alpha}{2}\phi(y) - \phi\left(\frac{1-\alpha}{2}x + \frac{1+\alpha}{2}y\right)\right]$$
$$-1 < \alpha < 1$$

Limiting cases

$$d_\phi^1(x,y) = \lim_{\alpha\to 1} d_\phi^\alpha(x,y) = B_\phi(x,y)$$
$$d_\phi^{-1}(x,y) = \lim_{\alpha\to -1} d_\phi^\alpha(x,y) = B_\phi(y,x)$$

Alpha Log-Determinant divergences

Chebbi and Moakher (Linear Algebra and Its Applications 2012)
$\Omega = \mathrm{Sym}^{++}(n)$, $\phi(X) = -\log\det(X)$

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

$$-1 < \alpha < 1$$

Limiting cases

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

(Burg divergence)

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

Alpha Log-Determinant divergences

- $\alpha = 0$: Symmetric Stein divergence (also called S-divergence)

$$d^0_{\mathrm{logdet}}(A, B) = 4\left[\log\det\left(\frac{A+B}{2}\right) - \frac{1}{2}\log\det(AB)\right] = 4d^2_{\mathrm{stein}}(A, B)$$

- Sra (NIPS 2012):

$$d_{\mathrm{stein}}(A, B) = \sqrt{\log\det\left(\frac{A+B}{2}\right) - \frac{1}{2}\log\det(AB)}$$

is a metric (satisfying positivity, symmetry, and triangle inequality)

Alpha Log-Determinant divergences - Properties

- Positivity

$$d^{\alpha}_{\mathrm{logdet}}(A, B) \geq 0$$
$$d^{\alpha}_{\mathrm{logdet}}(A, B) = 0 \Longleftrightarrow A = B$$

- In general, they are not symmetric and are not metrics.

Alpha Log-Determinant divergences - Properties

- Affine-invariance

$$d^{\alpha}_{\mathrm{logdet}}(CAC^T, CBC^T) = d^{\alpha}_{\mathrm{logdet}}(A, B), \quad \text{any } C \text{ invertible}$$

- Scale invariance: $C = \sqrt{s}I$, $s > 0$,

$$d^{\alpha}_{\mathrm{logdet}}(sA, sB) = d^{\alpha}_{\mathrm{logdet}}(A, B)$$

- Unitary (orthogonal) invariance: $CC^T = I \Longleftrightarrow C^{-1} = C^T$

$$d^{\alpha}_{\mathrm{logdet}}(CAC^{-1}, CBC^{-1}) = d^{\alpha}_{\mathrm{logdet}}(A, B)$$

Alpha Log-Determinant divergences - Properties

- Dual-symmetry

$$d^{\alpha}_{\mathrm{logdet}}(A, B) = d^{-\alpha}_{\mathrm{logdet}}(B, A)$$

$\alpha = 0$: symmetry

$$d^{0}_{\mathrm{logdet}}(A, B) = d^{0}_{\mathrm{logdet}}(B, A)$$

- Dual-invariance under inversion

$$d^{\alpha}_{\mathrm{logdet}}(A^{-1}, B^{-1}) = d^{-\alpha}_{\mathrm{logdet}}(A, B)$$

$\alpha = 0$: invariance under inversion

$$d^{0}_{\mathrm{logdet}}(A^{-1}, B^{-1}) = d^{0}_{\mathrm{logdet}}(A, B)$$

Connection with the Rényi divergences

P_1, P_2 = Borel probability density functions on $\mathbb{R}^n$

- For $0 < r < 1$, the Rényi divergence (Rényi, 1961) of order r between P_1 and P_2 is

$$d_R^r(P_1, P_2) = -\frac{1}{1-r} \log \int_{\mathbb{R}^n} P_1(x)^r P_2(x)^{1-r} dx$$

- As $r \to 1$, the Rényi divergence becomes the Kullback-Leibler divergence (Kullback, 1951)

$$d_{KL}(P_1, P_2) = \int_{\mathbb{R}^n} P_1(x) \log \frac{P_1(x)}{P_2(x)} dx$$

Connection with the Rényi divergences

For two multivariate Gaussian density functions $P_1 \sim \mathcal{N}(\mu, C_1)$ and $P_2 \sim \mathcal{N}(\mu, C_2)$ (same mean)

$$
\begin{aligned}
d_R^r(P_1, P_2) &= \frac{1}{2(1-r)} \log \left[\frac{\det[(1-r)C_1 + rC_2]}{\det(C_1)^{1-r} \det(C_2)^r} \right] \\
&= \frac{r}{2} d_{\text{logdet}}^{(2r-1)}(C_1, C_2)
\end{aligned}
$$

$$
\begin{aligned}
d_{KL}(P_1, P_2) &= \frac{1}{2}[\mathrm{tr}(C_2^{-1}C_1 - I) - \log\det(C_2^{-1}C_1)] \\
&= \frac{1}{2} d_{\text{logdet}}^{1}(C_1, C_2)
\end{aligned}
$$

Alpha Log-Determinant divergences - Computational complexity

- Can compute the $\log\det$ function using the Cholesky decomposition. The computational complexity $O(n^3)$.
- For a set of N SPD matrices, the computational complexity required for computing all pairwise divergences $d^{\alpha}_{\mathrm{logdet}}(A_i, A_j)$ is $O(N^2 n^3)$.
- No need to carry out matrix inversion and multiplication
- Much faster than affine-invariant Riemannian metric

Alpha-Beta Log-Determinant divergences

Cichocki, Cruces, Amari (Entropy 2015)

$$D^{(\alpha,\beta)}(A,B) = \frac{1}{\alpha\beta}\log\det\left[\frac{\alpha(AB^{-1})^{\beta} + \beta(AB^{-1})^{-\alpha}}{\alpha+\beta}\right],$$

$$\alpha > 0, \beta > 0$$

- A highly general family of divergences, encompassing many divergences and distances on $\mathrm{Sym}^{++}(n)$
- Affine-invariant divergences

Alpha-Beta Log-Determinant divergences

- Alpha Log-Determinant divergences

$$d^{\alpha}_{\mathrm{logdet}}(A, B) = D^{(\frac{1-\alpha}{2}, \frac{1+\alpha}{2})}(A, B)$$

- $\sqrt{D^{(\alpha,\alpha)}(A, B)}$ is a metric on $\mathrm{Sym}^{++}(n)$

$$D^{(1/2,1/2)}(A, B) = 4d^2_{\mathrm{stein}}(A, B)$$

- Squared affine-invariant Riemannian distance

$$\lim_{\alpha \to 0} D^{(\alpha,\alpha)}(A, B) = \frac{1}{2}||\log(A^{-1/2}BA^{-1/2})||^2_F$$

Alpha-Beta Log-Determinant divergences

$\{\lambda_k\}_{k=1}^n$ = eigenvalues of AB^{-1}

$$D^{(\alpha,\beta)}(A,B) = \frac{1}{\alpha\beta}\log\left(\frac{\alpha\lambda_k^{\beta} + \beta\lambda_k^{-\alpha}}{\alpha+\beta}\right)$$

- Require the computation of the eigenvalues of AB^{-1}
- Generally have the same computational complexity as the affine-invariant Riemannian distance

Power Euclidean distances

Dryden, Koloydenko, Zhou (Annals of Applied Statistics 2009)

$$d_{E,\alpha}(A, B) = \frac{1}{\alpha}||A^{\alpha} - B^{\alpha}||_F, \quad \alpha > 0$$

- Euclidean distance $d_{E,1}(A, B) = d_E(A, B)$
- Log-Euclidean distance as a limiting case

$$\lim_{\alpha \to 0} d_{E,\alpha}(A, B) = ||\log(A) - \log(B)||_F$$

Power Euclidean distances

$$d_{E,\alpha}(A, B) = \frac{1}{\alpha}||A^\alpha - B^\alpha||_F, \quad \alpha > 0$$

- Do not share the same invariance properties as the Log-Euclidean distance (neither scale-invariant nor inversion-invariant)
- For $\alpha \notin \mathbb{N}$, these divergences have the same computational complexity as the Log-Euclidean distance

Summary of distances and divergences

- Euclidean (Frobenius) distance and inner product
- Affine-invariant Riemannian distance
- Log-Euclidean distance and inner product
- Alpha Log-Determinant divergences, including symmetric Stein divergence
- Alpha-Beta Log-Determinant divergences
- Power-Euclidean distances

Properties of common distances and divergences

	Euclidean	Log-E	Affine-invariant	Stein
geodesic distance	Yes	Yes	Yes	No
affine invariance	No	No	Yes	Yes
scale invariance	No	Yes	Yes	Yes
unitary invariance	Yes	Yes	Yes	Yes
inversion invariance	No	Yes	Yes	Yes
inner product distance	Yes	Yes	No	No

Part I - Outline

- Data representation by covariance matrices
- Geometry of SPD matrices
- Machine learning methods on covariance matrices

Machine learning methods on covariance matrices

- Sparse coding and dictionary learning: Cherian and Sra (Trans. Neural Net. Learning Systems 2016), Harandi et al (Trans. Neural Net. Learning Systems 2016), Li et al (ICCV 2013), Sivalingam et al (PAMI 2016), Wu et al (Trans. Image Processing 2015)
- Metric learning: Huang et al (ICML 2015), Matsuzawa et al (ECCV 2016)
- Dimensionality reduction: Harandi et al (PAMI 2017)
- Kernel methods on covariance matrices

- Positive definite kernels defined with the Frobenius inner product $\langle\ ,\ \rangle_F$ and norm $||\ ||_F$
- Polynomial kernels

$$K(A,B) = (\langle A, B\rangle_F + c\rangle)^d = [\mathrm{tr}(A^TB) + c]^d, \quad d \in \mathbb{N}, c \geq 0$$

- Gaussian and Gaussian-like kernels

$$K(A,B) = \exp(-\frac{1}{\sigma^2}||A-B||_F^p), \quad 0 < p \leq 2$$

Kernels on SPD matrices with Log-Euclidean metric

- Positive definite kernels on $\mathrm{Sym}^{++}(n)$ defined with the Log-Euclidean inner product $\langle\ ,\ \rangle_{\mathrm{logE}}$ and norm $||\ ||_{\mathrm{logE}}$
- Polynomial kernels

$$
\begin{aligned}
K(A,B) &= (\langle A, B\rangle_{\mathrm{logE}} + c\rangle)^d \\
&= (\langle \log(A), \log(B)\rangle_F + c)^d, \quad d \in \mathbb{N}, c \geq 0
\end{aligned}
$$

- Gaussian and Gaussian-like kernels

$$
\begin{aligned}
K(A,B) &= \exp(-\frac{1}{\sigma^2}||(A \odot B^{-1})||^p_{\mathrm{logE}}), \quad 0 < p \leq 2 \\
&= \exp(-\frac{1}{\sigma^2}||\log(A) - \log(B)||^p_F)
\end{aligned}
$$

Kernels on SPD matrices - Stein divergence

- Sra (NIPS 2012)

$$K(A,B) = \exp[-\sigma d^2_{\text{stein}}(A,B)]$$

  is positive definite if and only if

$$\sigma \in \left\{\frac{1}{2}, 1 \dots, \frac{n-1}{2}\right\} \cup \left\{\sigma \in \mathbb{R} : \sigma > \frac{n-1}{2}\right\}$$

- One needs to be careful when fine-tuning σ to ensure the kernel remains positive definite.

Kernels on SPD matrices with Affine-invariant Riemannian distance?

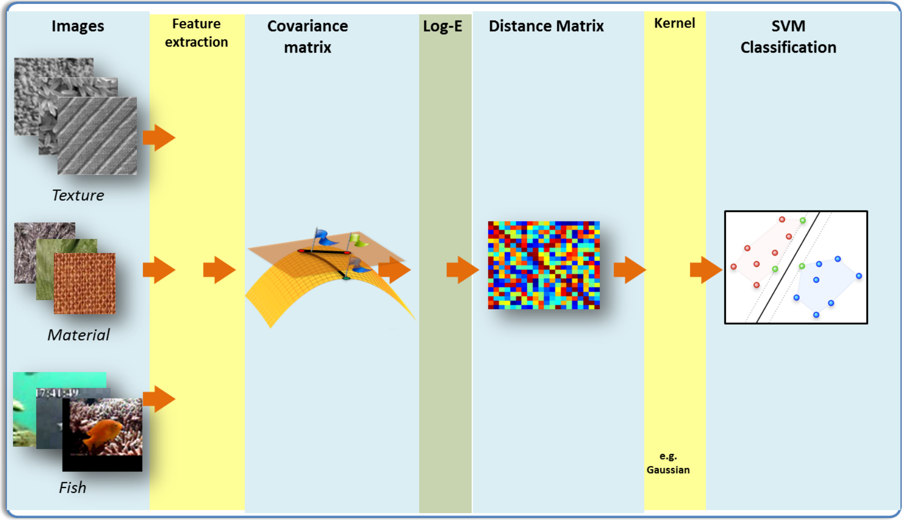
$$K(A, B) = \exp[-\sigma d^2_{\text{aiE}}(A, B)]$$

- K cannot be positive definite for all $\sigma > 0$, since the Riemannian manifold has nonpositive curvature (Feragen et al, CVPR 2015)
- Open question: Can K be positive definite for some specific choices of σ?

Kernel methods with Log-Euclidean metric

- S. Jayasumana, R. Hartley, M. Salzmann, H. Li, and M. Harandi. Kernel methods on the Riemannian manifold of symmetric positive definite matrices. CVPR 2013.
- S. Jayasumana, R. Hartley, M. Salzmann, H. Li, and M. Harandi. Kernel methods on Riemannian manifolds with Gaussian RBF kernels, PAMI 2015.
- P. Li, Q. Wang, W. Zuo, and L. Zhang. Log-Euclidean kernels for sparse representation and dictionary learning, ICCV 2013
- D. Tosato, M. Spera, M. Cristani, and V. Murino. Characterizing humans on Riemannian manifolds, PAMI 2013

Kernel methods with Log-Euclidean metric for image classification



Material classification

Example: KTH-TIPS2b data set

$$\mathbf{f}(x,y) = \left[R(x,y), G(x,y), B(x,y), \left|G^{0,0}(x,y)\right|, \ldots \left|G^{3,4}(x,y)\right|\right]$$

Object recognition

Example: ETH-80 data set

$$\mathbf{f}(x, y) = [x, y, I(x, y), |I_x|, |I_y|]$$

Numerical results

Better results with covariance operators (Part II)!

Method	KTH-TIPS2b	ETH-80
E	55.3% ($\pm$7.6%)	64.4% ($\pm$0.9%)
Stein	73.1% ($\pm$8.0%)	67.5% ($\pm$0.4%)
Log-E	74.1 % ($\pm$7.4%)	71.1% ($\pm$1.0%)

Comparison of metrics

Results from Cherian et al (PAMI 2013) using Nearest Neighbor

Method	Texture	Activity
Affine-invariant	85.5%	99.5%
Stein	85.5%	99.5%
Log-E	82.0%	96.5%

Texture: images from Brodatz and CURET datasets
Activity: videos from Weizmann, KTH, and UT Tower datasets

Summary of Part I - Finite-dimensional setting

Covariance Matrices and Applications

1. Data Representation by Covariance Matrices
2. Geometry of SPD matrices
3. Machine Learning Methods on Covariance Matrices and Applications in Computer Vision

References

- A. Cherian, S. Sra, A. Banerjee, and N. Papanikolopoulos. Jensen-Bregman LogDet divergence with application to efficient similarity search for covariance matrices, PAMI 2013.
- S. Jayasumana, R. Hartley, M. Salzmann, H. Li, and M. Harandi. Kernel methods on the Riemannian manifold of symmetric positive definite matrices. CVPR 2013.
- S. Jayasumana, R. Hartley, M. Salzmann, H. Li, and M. Harandi. Kernel methods on Riemannian manifolds with Gaussian RBF kernels, PAMI 2015.
- P. Li, Q. Wang, W. Zuo, and L. Zhang. Log-Euclidean kernels for sparse representation and dictionary learning, ICCV 2013
- D. Tosato, M. Spera, M. Cristani, and V. Murino. Characterizing humans on Riemannian manifolds, PAMI 2013
- V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Geometric means in a novel vector space structure on symmetric positive-definite matrices. SIAM J. on Matrix An. and App.,, 2007.

References

- A. Cichocki, S. Cruces, and S. Amari. Log-Determinant divergences revisited: Alpha- Beta and Gamma Log-Det divergences. Entropy, 2015.
- I. Dryden, A. Koloydenko, and D. Zhou. Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. Annals of Applied Statistics, , 2009.
- O. Ledoit and M. Wolf. A well-conditioned estimator for large-dimensional covariance matrices. Journal of multivariate analysis, 2004.
- M. Moakher and M. Zerai. The Riemannian geometry of the space of positive-definite matrices and its application to the regularization of positive-definite matrix-valued data. Journal of Mathematical Imaging and Vision, 2011.
- X. Pennec, P. Fillard, and N. Ayache. A Riemannian framework for tensor computing. International Journal of Computer Vision, 2006.

References

- R. Bhatia. Positive Definite Matrices. Princeton University Press, 2007.
- S. Sra. A new metric on the manifold of kernel matrices with application to matrix geometric means. In Advances in Neural Information Processing Systems (NIPS), pages 144-152, 2012.

Thank you for listening!

Questions, comments, suggestions?