Riemannian and information geometry in signal processing and machine learning

Part I: Riemannian Geometry and Optimization

Florent Bouchard, Arnaud Breloy and Ammar Mian

cnrs

Université Paris Nanterre

UNIVERSITÉ SAVOIE MONT BLANC

Outline

Outline

Matrix function differentiation – reference

Nicholas J Higham. *Functions of matrices: theory and computation.* SIAM, 2008

Matrix function differentiation – definition

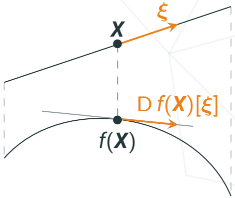
directional derivative (Fréchet)

$f:\mathbb{R}^{p\times q}\to\mathbb{R}^{m\times n}$ differentiable at $\boldsymbol{X}$ if $\exists\,\mathrm{D}\,f(\boldsymbol{X}):\mathbb{R}^{p\times q}\to\mathbb{R}^{m\times n}$ linear mapping such that

$$\lim_{\|\xi\|_2\to 0}\frac{\|f(\boldsymbol{X}+\xi-f(\boldsymbol{X})-\mathrm{D}\,f(\boldsymbol{X})[\xi]\|_2}{\|\xi\|_2}=0 \qquad \text{exists}$$

If $\mathrm{D}\,f(\boldsymbol{X})$ exists, it is unique

Equivalently, $\mathrm{D}\,f(\boldsymbol{X})[\xi]=f(\boldsymbol{X}+\xi)-f(\boldsymbol{X})+o(\|\xi\|)$



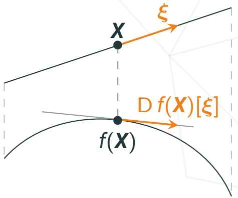
Matrix function differentiation – definition

directional derivative (Gateaux)

if f differentiable at $\boldsymbol{X}$, then $\forall \xi$,

$$\mathrm{D}\, f(\boldsymbol{X})[\xi] = \left.\frac{d}{dt}\right|_{t=0} f(\boldsymbol{X} + t\xi) = \lim_{t\to 0} \frac{f(\boldsymbol{X}+t\xi) - f(\boldsymbol{X})}{t}$$

⚠ $\left.\frac{d}{dt}\right|_{t=0} f(\boldsymbol{X} + t\xi)$ not linear $\Rightarrow$ f not differentiable



Matrix function differentiation – definition

Examples

$f(\boldsymbol{X}) = \boldsymbol{X}^2$

$$f(\boldsymbol{X}+\xi) = (\boldsymbol{X}+\xi)^2 = \boldsymbol{X}^2 + \boldsymbol{X}\xi + \xi\boldsymbol{X} + \xi^2 = f(\boldsymbol{X}) + \boldsymbol{X}\xi + \xi\boldsymbol{X} + o(\|\xi\|)$$

$$\mathrm{D}\, f(\boldsymbol{X})[\xi] = \boldsymbol{X}\xi + \xi\boldsymbol{X}$$

$f(\boldsymbol{X}) = \log\det(\boldsymbol{X})$

$$f(\boldsymbol{X}+\xi) = \log\det(\boldsymbol{X}+\xi) = \log\det(\boldsymbol{X}) + \log\det(\boldsymbol{I}+\boldsymbol{X}^{-1}\xi)$$

$$\det(\boldsymbol{I}+\boldsymbol{Y}) = 1 + \mathrm{tr}(\boldsymbol{Y}) + o(\|\boldsymbol{Y}\|)$$

$$\log(1+t) = t + o(t)$$

$$\mathrm{D}\, f(\boldsymbol{X})[\xi] = \mathrm{tr}(\boldsymbol{X}^{-1}\xi)$$

Matrix function differentiation – Vectorization

Vectorization

$\mathrm{D}\, f(\boldsymbol{X}) : \mathbb{R}^{p\times q} \to \mathbb{R}^{m\times n}$ linear mapping. Thus, $\exists\, \boldsymbol{M_X} \in \mathbb{R}^{pq\times mn}$ such that

$$\mathrm{vec}(\mathrm{D}\, f(\boldsymbol{X})[\xi]) = \boldsymbol{M_X}\, \mathrm{vec}(\xi)$$

$\boldsymbol{M_X}$ can usually be found with:

$$\begin{aligned}
\mathrm{vec}(\boldsymbol{ABC}) &= (\boldsymbol{C}^T \otimes \boldsymbol{A})\, \mathrm{vec}(\boldsymbol{B}) \\
\mathrm{tr}(\boldsymbol{AB}) &= \mathrm{vec}(\boldsymbol{A})^T\, \mathrm{vec}(\boldsymbol{B})
\end{aligned}$$

Matrix function differentiation – Vectorization

Special case – $f : \mathbb{R}^{p\times q} \to \mathbb{R}$

$$\exists\, \boldsymbol{G_X}, \quad \mathrm{D}\, f(\boldsymbol{X})[\xi] = \mathrm{tr}(\boldsymbol{G_X}^T \xi)$$

since $\mathrm{tr}(\boldsymbol{AB}) = \mathrm{vec}(\boldsymbol{A})^T \mathrm{vec}(\boldsymbol{B})$

Special case – $f : \mathbb{R}^{p} \to \mathbb{R}^n$

Jacobian at $\boldsymbol{x}$: $\boldsymbol{J_x} \in \mathbb{R}^{n\times p}$ such that $(\boldsymbol{J_x})_{ij} = \frac{\partial f_i}{\partial x_j}$

Directional derivative: $\mathrm{D}\, f(\boldsymbol{x})[\xi] = \boldsymbol{J_x} \cdot \xi$

Matrix function differentiation – properties

Sum property

$g : \mathbb{R}^{p\times q} \to \mathbb{R}^{m\times n}$, $h : \mathbb{R}^{p\times q} \to \mathbb{R}^{m\times n}$ differentiable functions

$f = \alpha g + \beta h$ differentiable and

$$\mathrm{D}\, f(\boldsymbol{X})[\xi] = \alpha\, \mathrm{D}\, g(\boldsymbol{X})[\xi] + \beta\, \mathrm{D}\, h(\boldsymbol{X})[\xi]$$

Matrix function differentiation – properties

Product property

$g : \mathbb{R}^{p\times q} \to \mathbb{R}^{n\times n}$, $h : \mathbb{R}^{p\times q} \to \mathbb{R}^{n\times n}$ differentiable functions

$f = g \cdot h$ differentiable and

$$\mathrm{D}\, f(\boldsymbol{X})[\xi] = g(\boldsymbol{X}) \cdot \mathrm{D}\, h(\boldsymbol{X})[\xi] + \mathrm{D}\, g(\boldsymbol{X})[\xi] \cdot h(\boldsymbol{X})$$

Matrix function differentiation – properties

Examples

$$g(\boldsymbol{X}) = h(\boldsymbol{X}) = \boldsymbol{X} \qquad f(\boldsymbol{X}) = \boldsymbol{X}^2 \qquad \mathrm{D}\, g(\boldsymbol{X})[\xi] = \mathrm{D}\, h(\boldsymbol{X})[\xi] = \xi$$

$$\mathrm{D}\, f(\boldsymbol{X})[\xi] = \boldsymbol{X}\xi + \xi\boldsymbol{X}$$

$$g(\boldsymbol{X}) = \boldsymbol{X} \qquad h(\boldsymbol{X}) = \boldsymbol{X}^{-1} \qquad f(\boldsymbol{X}) = \boldsymbol{I} \qquad \mathrm{D}\, g(\boldsymbol{X})[\xi] = \xi \qquad \mathrm{D}\, f(\boldsymbol{X})[\xi] = \boldsymbol{0}$$

$$\xi\boldsymbol{X}^{-1} + \boldsymbol{X}\,\mathrm{D}\, h(\boldsymbol{X})[\xi] = \boldsymbol{0}$$

$$\mathrm{D}\, h(\boldsymbol{X})[\xi] = -\boldsymbol{X}^{-1}\xi\boldsymbol{X}^{-1}$$

Matrix function differentiation – properties

Composition property

$g : \mathbb{R}^{p\times q} \to \mathbb{R}^{m\times n}$, $h : \mathbb{R}^{k\times \ell} \to \mathbb{R}^{p\times q}$ differentiable functions

$f = g \circ h$ differentiable and

$$\mathrm{D}\, f(\boldsymbol{X})[\xi] = \mathrm{D}\, g(h(\boldsymbol{X}))[\mathrm{D}\, h(\boldsymbol{X})[\xi]]$$

Matrix function differentiation – properties

Example

$$g(\boldsymbol{X}) = \boldsymbol{X}^2 \qquad h(\boldsymbol{X}) = \boldsymbol{X}^{1/2} \qquad f(\boldsymbol{X}) = \boldsymbol{X} \qquad \mathrm{D}\, g(\boldsymbol{X})[\xi] = \boldsymbol{X}\xi + \xi\boldsymbol{X} \qquad \mathrm{D}\, f(\boldsymbol{X})[\xi] = \xi$$

$$\boldsymbol{X}^{1/2}\, \mathrm{D}\, h(\boldsymbol{X})[\xi] + \mathrm{D}\, h(\boldsymbol{X})[\xi]\boldsymbol{X}^{1/2} = \xi$$

Thus, $\mathrm{D}\, h(\boldsymbol{X})[\xi]$ solution to a Sylvester equation

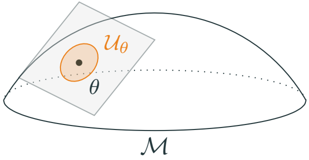
Outline

Riemannian geometry – manifold

Manifold $\mathcal{M}$

space locally diffeomorphic to $\mathbb{R}^d$, with $\dim(\mathcal{M}) = d$, *i.e.*

$$\forall \theta \in \mathcal{M},\ \exists \mathcal{U}_\theta \subset \mathcal{M} \text{ and } \varphi_\theta : \mathcal{U}_\theta \to \mathbb{R}^d,\ \text{diffeomorphism}$$



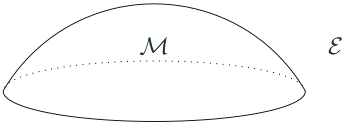
Riemannian geometry – manifold embedded in Euclidean space

manifold $\mathcal{M}$ embedded in Euclidean space $\mathcal{E}$

$\mathcal{M}$ defined through set of constraints in $\mathcal{E}$

$$\mathcal{M} = \{\theta \in \mathcal{E} : F(\theta) = 0_{\hat{\mathcal{E}}}\}$$

$F : \mathcal{E} \to \hat{\mathcal{E}}$ submersion, $\hat{\mathcal{E}}$ Euclidean space, $0_{\hat{\mathcal{E}}}$ zero element of $\hat{\mathcal{E}}$



Riemannian geometry – manifold

Examples

Manifold of symmetric positive definite matrices

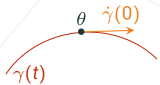
$$\mathcal{S}_p^{++} = \{\boldsymbol{\Sigma} \in \mathcal{S}_p : \ \forall \boldsymbol{x} \in \mathbb{R}^p, \ \boldsymbol{x}^T \boldsymbol{\Sigma} \boldsymbol{x} > 0\}$$

Orthogonal group

$$\mathcal{O}_p = \{\boldsymbol{O} \in \mathbb{R}^{p \times p} : \ \boldsymbol{O}^T \boldsymbol{O} = \boldsymbol{I}_p\}$$

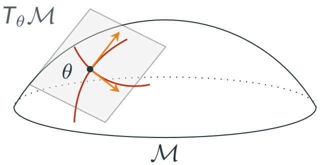
Riemannian geometry – tangent space

Curve $\gamma : \mathbb{R} \to \mathcal{M}$, $\gamma(0) = \theta$, derivative: $\dot{\gamma}(0) = \lim_{t \to 0} \frac{\gamma(t) - \gamma(0)}{t}$



Tangent space $T_\theta \mathcal{M}$

$$T_\theta \mathcal{M} = \{\dot{\gamma}(0) : \gamma : \mathbb{R} \to \mathcal{M}, \ \gamma(0) = \theta\}$$



Riemannian geometry – manifold embedded in Euclidean space

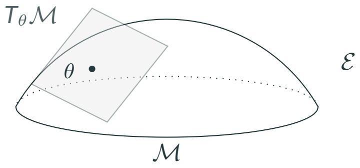
Manifold $\mathcal{M}$ embedded in Euclidean space $\mathcal{E}$

$$\mathcal{M} = \{\theta \in \mathcal{E} : F(\theta) = 0_{\hat{\mathcal{E}}}\}$$

$F : \mathcal{E} \to \hat{\mathcal{E}}$, $\hat{\mathcal{E}}$ Euclidean space, $0_{\hat{\mathcal{E}}}$ zero of $\hat{\mathcal{E}}$

Tangent space $T_\theta\mathcal{M}$ of embedded manifold

$$T_\theta\mathcal{M} = \{\xi \in \mathcal{E} : \mathrm{D}\, F(\theta)[\xi] = 0_{\hat{\mathcal{E}}}\}$$



Riemannian geometry – tangent space

Examples

Manifold of symmetric positive definite matrices $\mathcal{S}_p^{++}$

$$\mathcal{S}_p^{++} \text{ open in } \mathcal{S}_p \quad \Rightarrow \quad \forall \boldsymbol{\Sigma} \in \mathcal{S}_p^{++}, \;\; T_{\boldsymbol{\Sigma}}\mathcal{S}_p^{++} \simeq \mathcal{S}_p$$

Orthogonal group $\mathcal{O}_p$

$$f(\boldsymbol{O}) = \boldsymbol{O}^T\boldsymbol{O} \qquad\qquad \mathrm{D}\, f(\boldsymbol{O})[\xi] = \boldsymbol{O}^T\xi + \xi^T\boldsymbol{O}$$

$$\boldsymbol{O}^T\boldsymbol{O} = \boldsymbol{I}_p \quad \Rightarrow \quad \boldsymbol{O}^T\xi + \xi^T\boldsymbol{O} = \boldsymbol{0}_p$$

$$T_{\boldsymbol{O}}\mathcal{O}_p = \{\xi \in \mathbb{R}^{p\times p} : \; \boldsymbol{O}^T\xi + \xi^T\boldsymbol{O} = \boldsymbol{0}_p\}$$
$$= \{\xi = \boldsymbol{O}\boldsymbol{\Omega} : \; \boldsymbol{\Omega} \in \mathbb{R}^{p\times p}, \; \boldsymbol{\Omega}^T = -\boldsymbol{\Omega}\}$$

Riemannian geometry – Riemannian metric

Riemannian metric $\langle \cdot, \cdot \rangle$.

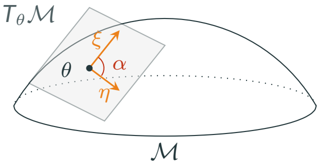
$\forall \theta \in \mathcal{M}$, $\langle \cdot, \cdot \rangle_\theta : T_\theta\mathcal{M} \times T_\theta\mathcal{M} \to \mathbb{R}$ inner product on $T_\theta\mathcal{M}$

i.e., bilinear, symmetric, positive definite mapping

$\langle \cdot, \cdot \rangle_\theta$ varies smoothly in θ on $\mathcal{M}$

Riemannian metric defines length and relative positions of tangent vectors

$$\|\xi\|_\theta^2 = \langle \xi, \eta \rangle_\theta \qquad \alpha(\xi, \eta) = \frac{\langle \xi, \xi \rangle_\theta}{\|\xi\|_\theta \|\eta\|_\theta}$$



Riemannian geometry – Riemannian metric

Examples

Manifold of symmetric positive definite matrices $\mathcal{S}_p^{++}$

$$\langle \xi, \eta \rangle_{\mathbf{\Sigma}} = \text{tr}(\mathbf{\Sigma}^{-1} \xi \mathbf{\Sigma}^{-1} \eta)$$

$\mathcal{S}_p^{++}$ open in $\mathcal{S}_p \Rightarrow$ boundary at the infinite through metric

Orthogonal group $\mathcal{O}_p$

$$\langle \xi, \eta \rangle_{\mathbf{O}} = \text{tr}(\xi^T \eta)$$

restriction to $\mathcal{O}_p$ of the Euclidean metric on $\mathbb{R}^{p \times p}$

Riemannian geometry – orthogonal projection

Manifold $\mathcal{M}$ embedded in Euclidean space $\mathcal{E}$

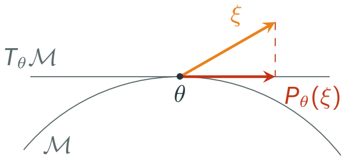
Normal space: $(T_\theta\mathcal{M})^\perp = \{\nu \in \mathcal{E} : \langle \xi, \nu \rangle_\theta = 0, \ \forall \xi \in T_\theta\mathcal{M}\}$

Orthogonal projection P

Every $\xi \in \mathcal{E}$ can be uniquely decomposed into

$$\xi = P_\theta(\xi) + P_\theta^\perp(\xi)$$

$P_\theta, \ P_\theta^\perp$ orthogonal projections onto $T_\theta\mathcal{M}$ and $(T_\theta\mathcal{M})^\perp$



Riemannian geometry – orthogonal projection

Examples

Manifold of symmetric positive definite matrices $\mathcal{S}_p^{++}$

$$\begin{array}{rccl} P_{\boldsymbol{\Sigma}}: & \mathbb{R}^{p\times p} & \to & T_{\boldsymbol{\Sigma}}\mathcal{S}_p^{++} \simeq \mathcal{S}_p \\ & \boldsymbol{\xi} & \mapsto & \operatorname{sym}(\boldsymbol{\xi}) \end{array}$$

Orthogonal group $\mathcal{O}_p$

$$\begin{array}{rccl} P_{\boldsymbol{O}}: & \mathbb{R}^{p\times p} & \to & T_{\boldsymbol{O}}\mathcal{O}_p \\ & \boldsymbol{\xi} & \mapsto & \boldsymbol{\xi} - \boldsymbol{O}\operatorname{sym}(\boldsymbol{O}^T\boldsymbol{\xi}) \end{array}$$

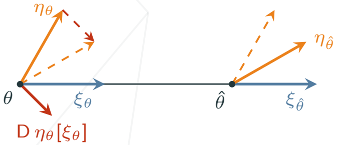
$$T_{\boldsymbol{O}}\mathcal{O}_p = \{\boldsymbol{\xi} = \boldsymbol{O}\boldsymbol{\Omega} :\ \boldsymbol{\Omega} \in \mathbb{R}^{p\times p},\ \boldsymbol{\Omega}^T = -\boldsymbol{\Omega}\}$$

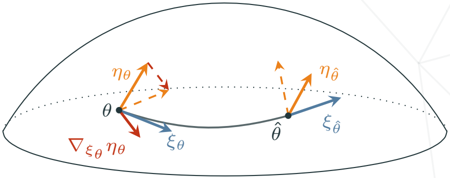
Riemannian geometry – Levi-Civita connection

Vector field: function $\xi : \theta \in \mathcal{M} \mapsto \xi_\theta \in T_\theta\mathcal{M}$ $\mathcal{X}(\mathcal{M})$: set of vector fields of $\mathcal{M}$

Levi-Civita connection: $\nabla : \mathcal{X}(\mathcal{M}) \times \mathcal{X}(\mathcal{M}) \to \mathcal{X}(\mathcal{M})$

generalizes notion of directional derivatives for vector fields





Riemannian geometry – Levi-Civita connection

Levi-Civita connection ∇

$\nabla : \mathcal{X}(\mathcal{M}) \times \mathcal{X}(\mathcal{M}) \to \mathcal{X}(\mathcal{M})$ such that

$\mathcal{X}(\mathcal{M})$: set of vector fields of $\mathcal{M}$

- $\nabla_{f(\theta)\xi_\theta + g(\theta)\nu_\theta} \eta_\theta = f(\theta)\nabla_{\xi_\theta}\eta_\theta + g(\theta)\nabla_{\nu_\theta}\eta_\theta$ — $f, g : \mathcal{M} \to \mathbb{R}$
- $\nabla_{\xi_\theta}(a\eta_\theta + b\nu_\theta) = a\nabla_{\xi_\theta}\eta_\theta + b\nabla_{\xi_\theta}\nu_\theta$
- $\nabla_{\xi_\theta}(f(\theta)\eta_\theta) = \mathrm{D}\, f(\theta)[\xi_\theta]\eta_\theta + f(\theta)\nabla_{\xi_\theta}\eta_\theta$

∇ associated to Riemannian metric $\langle \cdot, \cdot \rangle_\cdot$, characterized by Koszul formula

$$\langle 2\nabla_{\xi_\theta}\eta_\theta, \nu_\theta \rangle_\theta = \mathrm{D}\langle \xi_\theta, \nu_\theta\rangle_\theta[\eta_\theta] + \mathrm{D}\langle \eta_\theta, \nu_\theta\rangle_\theta[\xi_\theta] - \mathrm{D}\langle \xi_\theta, \eta_\theta\rangle_\theta[\nu_\theta] - \langle \xi_\theta, [\eta_\theta, \nu_\theta]\rangle_\theta + \langle \eta_\theta, [\nu_\theta, \xi_\theta]\rangle_\theta + \langle \nu_\theta, [\xi_\theta, \eta_\theta]\rangle_\theta$$

Riemannian geometry – Levi-Civita connection

Examples

Manifold of symmetric positive definite matrices $\mathcal{S}_p^{++}$

$$\nabla_{\xi_{\mathbf{\Sigma}}} \eta_{\mathbf{\Sigma}} = \mathrm{D}\, \eta_{\mathbf{\Sigma}}[\xi_{\mathbf{\Sigma}}] - \mathrm{sym}(\eta_{\mathbf{\Sigma}} \mathbf{\Sigma}^{-1} \xi_{\mathbf{\Sigma}})$$

Orthogonal group $\mathcal{O}_p$

$$\nabla_{\xi_{\mathbf{O}}} \eta_{\mathbf{O}} = P_{\mathbf{O}}(\mathrm{D}\, \eta_{\mathbf{O}}[\xi_{\mathbf{O}}])$$

$$P_{\mathbf{O}}(\xi) = \xi - \mathbf{O}\, \mathrm{sym}(\mathbf{O}^T \xi)$$

Riemannian geometry – geodesics

Geodesics γ

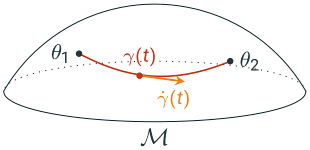
$\gamma : [0,1] \to \mathcal{M}$ solution to initial value problem

$$\nabla_{\dot{\gamma}(t)}\, \dot{\gamma}(t) = 0_{\gamma(t)}$$

given $(\gamma(0), \dot{\gamma}(0))$ or $(\gamma(0), \gamma(1))$

$0_{\gamma(t)}$, zero element of $T_{\gamma(t)}\mathcal{M}$

Geodesics generalize straight lines to manifolds: curves with no acceleration



Riemannian geometry – geodesics

Examples

Manifold of symmetric positive definite matrices $\mathcal{S}_p^{++}$

$$\nabla_{\dot{\gamma}(t)}\dot{\gamma}(t) = \mathbf{0} \quad \Rightarrow \quad \ddot{\gamma}(t) - \dot{\gamma}(t)\gamma(t)^{-1}\dot{\gamma}(t) = \mathbf{0}$$

$$\gamma(0) = \mathbf{\Sigma}, \quad \dot{\gamma}(0) = \xi : \quad \gamma(t) = \mathbf{\Sigma}\exp(t\mathbf{\Sigma}^{-1}\xi)$$

$$\gamma(0) = \mathbf{\Sigma}_1, \quad \gamma(1) = \mathbf{\Sigma}_2 : \quad \gamma(t) = \mathbf{\Sigma}_1^{1/2}(\mathbf{\Sigma}_1^{-1/2}\mathbf{\Sigma}_2\mathbf{\Sigma}_1^{-1/2})^t\mathbf{\Sigma}_1^{1/2}$$

Orthogonal group $\mathcal{O}_p$

$$\nabla_{\dot{\gamma}(t)}\dot{\gamma}(t) = \mathbf{0} \quad \Rightarrow \quad \ddot{\gamma}(t) - \gamma(t)\ddot{\gamma}(t)^T\gamma(t) = \mathbf{0}$$

$$\gamma(0) = \mathbf{O}, \quad \dot{\gamma}(0) = \xi : \quad \gamma(t) = \mathbf{O}\exp(t\mathbf{O}^T\xi)$$

$$\gamma(0) = \mathbf{O}_1, \quad \gamma(1) = \mathbf{O}_2 : \quad \gamma(t) = \mathbf{O}_1(\mathbf{O}_1^T\mathbf{O}_2)^t$$

Riemannian geometry – exponential and logarithm mappings

Riemannian exponential

$\forall \theta \in \mathcal{M}$, $\exp_\theta : T_\theta\mathcal{M} \to \mathcal{M}$ such that

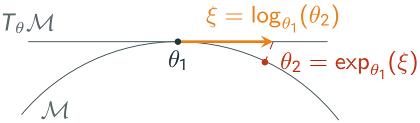
$$\exp_\theta(\xi) = \gamma(1),$$

$\gamma : [0,1] \to \mathcal{M}$ geodesic with $\gamma(0) = \theta$, $\dot{\gamma}(0) = \xi$

Riemannian logarithm

$\forall \theta_1 \in \mathcal{M}$, $\log_{\theta_1} : \mathcal{M} \to T_{\theta_1}\mathcal{M}$ such that

$$\exp_{\theta_1}(\log_{\theta_1}(\theta_2)) = \theta_2$$



Riemannian geometry – exponential and logarithm mappings

Examples

Manifold of symmetric positive definite matrices $\mathcal{S}_p^{++}$

$$\exp_{\boldsymbol{\Sigma}}(\xi) = \boldsymbol{\Sigma} \exp(\boldsymbol{\Sigma}^{-1}\xi)$$

$$\log_{\boldsymbol{\Sigma}_1}(\boldsymbol{\Sigma}_2) = \boldsymbol{\Sigma}_1 \log(\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_2)$$

Orthogonal group $\mathcal{O}_p$

$$\exp_{\boldsymbol{O}}(\xi) = \boldsymbol{O} \exp(\boldsymbol{O}^T\xi)$$

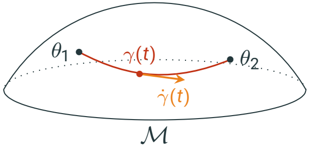
$$\log_{\boldsymbol{O}_1}(\boldsymbol{O}_2) = \boldsymbol{O}_1 \log(\boldsymbol{O}_1^T\boldsymbol{O}_2)$$

Riemannian geometry – distance

Riemannian distance $d : \mathcal{M} \times \mathcal{M} \to \mathbb{R}^+$ associated to $\langle \cdot, \cdot \rangle$.

$\theta_1, \theta_2 \in \mathcal{M}$, $d(\theta_1, \theta_2)$: length of the geodesic connecting θ_1 and θ_2

$$d(\theta_1, \theta_2) = \int_0^1 \sqrt{\langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle_{\gamma(t)}} \, dt$$



Riemannian geometry – distance

Examples

Manifold of symmetric positive definite matrices $\mathcal{S}_p^{++}$

$$d(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2) = \left\| \log(\boldsymbol{\Sigma}_1^{-1/2} \boldsymbol{\Sigma}_2 \boldsymbol{\Sigma}_1^{-1/2}) \right\|_2$$

Orthogonal group $\mathcal{O}_p$

$$d(\boldsymbol{O}_1, \boldsymbol{O}_2) = \left\| \log(\boldsymbol{O}_1^T \boldsymbol{O}_2) \right\|_2$$

Riemannian geometry – parallel transport

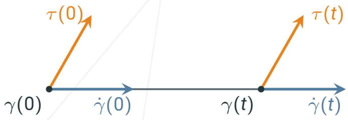
Parallel transport τ

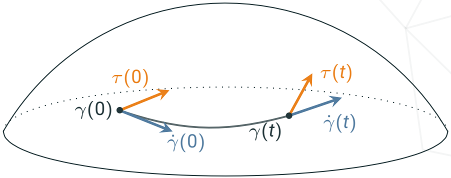
$\tau : [0,1] \to T\mathcal{M}$, solution to

$$\nabla_{\dot{\gamma}(t)}\tau(t) = 0_{\gamma(t)},$$

given curve $\gamma : [0,1] \to \mathcal{M}$ and $\tau(0)$

$0_{\gamma(t)}$, zero element of $T_{\gamma(t)}\mathcal{M}$





Riemannian geometry – parallel transport

Examples

Manifold of symmetric positive definite matrices $\mathcal{S}_p^{++}$

transport along geodesic $\gamma(t) = \mathbf{\Sigma} \exp(t\mathbf{\Sigma}^{-1}\xi)$, $\quad \tau(0) = \eta$

$$\nabla_{\dot{\gamma}(t)}\tau(t) = \mathbf{0} \quad \Rightarrow \quad \dot{\tau}(t) - \mathrm{sym}(\dot{\gamma}(t)\gamma(t)^{-1}\tau(t)) = \mathbf{0}$$

$$\tau(t) = \exp(t\xi\mathbf{\Sigma}^{-1}/2)\eta \exp(t\mathbf{\Sigma}^{-1}\xi/2)$$

Orthogonal group $\mathcal{O}_p$

transport along geodesic $\gamma(t) = \mathbf{O} \exp(t\mathbf{O}^T\xi)$, $\quad \tau(0) = \eta$

$$\nabla_{\dot{\gamma}(t)}\tau(t) = \mathbf{0} \quad \Rightarrow \quad \dot{\tau}(t) - \gamma(t)\dot{\tau}(t)^T\gamma(t) = \mathbf{0}$$

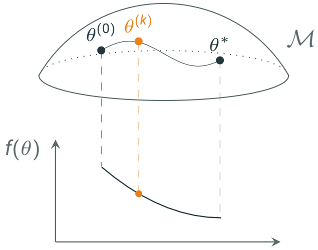
$$\tau(t) = \exp(t\xi\mathbf{O}^T/2)\eta \exp(t\mathbf{O}^T\xi/2)$$

Outline

Riemannian optimization

$$\theta^* = \operatorname*{argmin}_{\theta \in \mathcal{M}} \quad f(\theta)$$

from $\theta^{(0)}$, sequence of iterates $\{\theta^{(k)}\}$ converging to θ^*



Riemannian optimization

Examples

Fréchet mean of $\{\theta_i\}$ on $\mathcal{M}$

$$f(\theta) = \frac{1}{2n} \sum_{i=1}^{n} d^2(\theta, \theta_i)$$

$d(\cdot,\cdot)$ Riemannian distance on $\mathcal{M}$ associated to $\langle\cdot,\cdot\rangle$.

Tyler estimator for samples $\{\boldsymbol{x}_i\}$ on $\mathcal{S}_p^{++}$

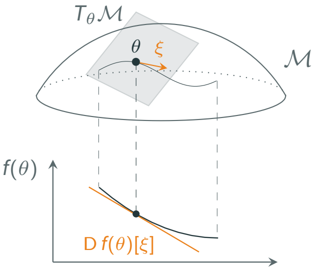
$$f(\boldsymbol{\Sigma}) = p \sum_{i=1}^{n} \log(\boldsymbol{x}_i^T \boldsymbol{\Sigma}^{-1} \boldsymbol{x}_i) + n \log\det(\boldsymbol{\Sigma})$$

Riemannian optimization – descent direction

Descent direction

$\theta \in \mathcal{M}$, descent direction $\xi \in T_\theta\mathcal{M}$ of f such that

$$\mathrm{D}\, f(\theta)[\xi] < 0$$



Riemannian optimization – gradient

Riemannian gradient $\operatorname{grad} f$

$\theta \in \mathcal{M}$, $\operatorname{grad} f(\theta) \in T_\theta\mathcal{M}$, unique tangent vector such that $\forall \xi \in T_\theta\mathcal{M}$

$$\langle \operatorname{grad} f(\theta), \xi \rangle_\theta = \mathrm{D}\, f(\theta)[\xi]$$

Riemannian optimization – gradient

Riemannian gradient in $\mathcal{M}$ can usually be obtained from Euclidean gradient in $\mathcal{E}$

$\mathcal{M}$ with ambient space $\mathcal{E}$

Examples

Manifold of symmetric positive definite matrices $\mathcal{S}_p^{++}$

$$\operatorname{grad} f(\boldsymbol{\Sigma}) = \boldsymbol{\Sigma} \operatorname{sym}(\operatorname{grad}_{\mathcal{E}} f(\boldsymbol{\Sigma}))\boldsymbol{\Sigma}$$

Orthogonal group $\mathcal{O}_p$

$$\operatorname{grad} f(\boldsymbol{O}) = P_{\boldsymbol{O}}(\operatorname{grad}_{\mathcal{E}} f(\boldsymbol{O}))$$

Riemannian optimization – gradient

Examples

Fréchet mean of $\{\theta_i\}$ on $\mathcal{M}$

$$\operatorname{grad} f(\theta) = -\frac{1}{n}\sum_{i=1}^{n} \log_\theta(\theta_i)$$

Tyler estimator for samples $\{\boldsymbol{x}_i\}$ on $\mathcal{S}_p^{++}$

$$\operatorname{grad} f(\boldsymbol{\Sigma}) = n\boldsymbol{\Sigma} - p\Psi(\boldsymbol{\Sigma}) \qquad \Psi(\boldsymbol{\Sigma}) = \sum_{i=1}^{n} \frac{\boldsymbol{x}_i\boldsymbol{x}_i^T}{\boldsymbol{x}_i^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x}_i}$$

Riemannian optimization – retraction

retraction R

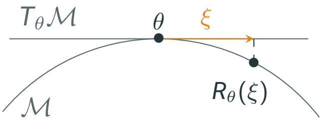
$\theta \in \mathcal{M}$, $R_\theta : T_\theta\mathcal{M} \to \mathcal{M}$ such that

$$R_\theta(0_\theta) = \theta \qquad \mathrm{D}\, R_\theta(0_\theta)[\xi] = \xi, \ \forall \xi \in T_\theta\mathcal{M}$$

Most natural retraction: Riemannian exponential mapping

But: might be complicated, numerically expensive or unstable

$\Rightarrow$ Other retractions might be advantageous



Riemannian optimization – retraction

Examples

Manifold of symmetric positive definite matrices $\mathcal{S}_p^{++}$

$$R_{\mathbf{\Sigma}}(\xi) = \mathbf{\Sigma} + \xi + \frac{1}{2}\xi\mathbf{\Sigma}^{-1}\xi$$

Orthogonal group $\mathcal{O}_p$

$$R_{\mathbf{O}}(\xi) = \mathrm{uf}(\mathbf{O} + \xi)$$

$\mathrm{uf}(\mathbf{M}) = \mathbf{U}\mathbf{V}^T$ from svd $\mathbf{M} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T$

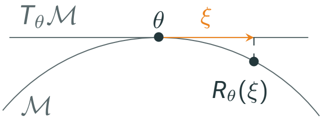
Riemannian optimization – optimization scheme

Minimize f on $\mathcal{M}$ from θ:

- descent direction $\xi \in T_\theta\mathcal{M}$

$$\langle \operatorname{grad} f(\theta), \xi \rangle_\theta < 0$$

- retraction of ξ on $\mathcal{M}$



- reiterate until critical point

$$\operatorname{grad} f(\theta) = 0_\theta$$

Riemannian optimization – gradient descent

descent direction

$$\xi^{(k)} = -\operatorname{grad} f(\theta^{(k)})$$

update

$$\theta^{(k+1)} = R_{\theta^{(k)}}(-t_k \operatorname{grad} f(\theta^{(k)}))$$

t_k: stepsize, can be computed with linesearch

Riemannian optimization – vector transport

Vector transport $\mathcal{T}$

$\theta_1,\ \theta_2 \in \mathcal{M}$, $\mathcal{T}_{\theta_1 \to \theta_2} : T_{\theta_1}\mathcal{M} \to T_{\theta_2}\mathcal{M}$ such that

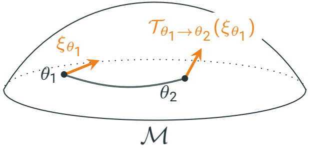
$$\mathcal{T}_{\theta_1 \to \theta_1}(\xi_{\theta_1}) = \xi_{\theta_1}$$

$$\mathcal{T}_{\theta_1 \to \theta_2}(a\xi_{\theta_1} + b\nu_{\theta_1}) = a\mathcal{T}_{\theta_1 \to \theta_2}(\xi_{\theta_1}) + b\mathcal{T}_{\theta_1 \to \theta_2}(\nu_{\theta_1})$$

Most natural vector transport: from parallel transport on $\mathcal{M}$

But: might be complicated, numerically expensive or unstable

$\Rightarrow$ Other vector transports might be advantageous



Riemannian optimization – vector transport

Examples

Manifold of symmetric positive definite matrices $\mathcal{S}_p^{++}$

from parallel transport: $\mathcal{T}_{\mathbf{\Sigma}_1 \to \mathbf{\Sigma}_2}(\xi_1) = (\mathbf{\Sigma}_2 \mathbf{\Sigma}_1^{-1})^{1/2} \xi_1 (\mathbf{\Sigma}_1^{-1} \mathbf{\Sigma}_2)^{1/2}$

alternative ones: $\mathcal{T}_{\mathbf{\Sigma}_1 \to \mathbf{\Sigma}_2}(\xi_1) = \xi_1$ $\quad$ $\mathcal{T}_{\mathbf{\Sigma}_1 \to \mathbf{\Sigma}_2}(\xi_1) = \mathbf{\Sigma}_2^{1/2} \mathbf{\Sigma}_1^{-1/2} \xi_1 \mathbf{\Sigma}_1^{-1/2} \mathbf{\Sigma}_2^{1/2}$

Orthogonal group $\mathcal{O}_p$

from parallel transport: $\mathcal{T}_{\boldsymbol{O}_1 \to \boldsymbol{O}_2}(\xi_1) = (\boldsymbol{O}_2 \boldsymbol{O}_1^T)^{1/2} \xi_1 (\boldsymbol{O}_1^T \boldsymbol{O}_2)^{1/2}$

alternative one: $\mathcal{T}_{\boldsymbol{O}_1 \to \boldsymbol{O}_2}(\xi_1) = P_{\boldsymbol{O}_2}(\xi_1)$

Riemannian optimization – conjugate gradient

descent direction

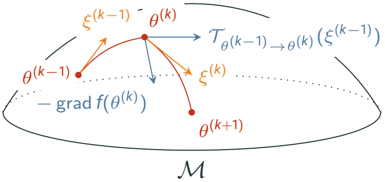
$$\xi^{(k)} = -\operatorname{grad} f(\theta^{(k)}) + \beta_k \mathcal{T}_{\theta^{(k-1)}\to\theta^{(k)}}(\xi^{(k-1)})$$

β_k: several rules – Fletcher-Reeves, Polak-Ribière,...

update

$$\theta^{(k+1)} = R_{\theta^{(k)}}(t_k \xi^{(k)})$$

t_k: stepsize, can be computed with linesearch



Riemannian optimization – Hessian

Riemannian Hessian $\operatorname{Hess} f$

$\theta \in \mathcal{M}$, $\operatorname{Hess} f(\theta) : T_\theta\mathcal{M} \to T_\theta\mathcal{M}$ such that $\forall \xi \in T_\theta\mathcal{M}$

$$\operatorname{Hess} f(\theta)[\xi] = \nabla_\xi \operatorname{grad} f(\theta)$$

Riemannian optimization – Hessian

Riemannian Hessian in $\mathcal{M}$ can be obtained from Euclidean Hessian and gradient in $\mathcal{E}$

$\mathcal{M}$ with ambient space $\mathcal{E}$

Examples

Manifold of symmetric positive definite matrices $\mathcal{S}_p^{++}$

$$\operatorname{Hess} f(\mathbf{\Sigma})[\xi] = \mathbf{\Sigma} \operatorname{sym}(\operatorname{Hess}_{\mathcal{E}} f(\mathbf{\Sigma})[\xi])\mathbf{\Sigma} + \operatorname{sym}(\xi \operatorname{sym}(\operatorname{grad}_{\mathcal{E}} f(\mathbf{\Sigma}))\mathbf{\Sigma})$$

Orthogonal group $\mathcal{O}_p$

$$\operatorname{Hess} f(\mathbf{O})[\xi] = P_{\mathbf{O}}(\operatorname{Hess}_{\mathcal{E}} f(\mathbf{O})[\xi] - \xi \operatorname{sym}(\mathbf{O}^T \operatorname{grad}_{\mathcal{E}} f(\mathbf{O})))$$

Riemannian optimization – Hessian

Examples

Fréchet mean of $\{\theta_i\}$ on $\mathcal{M}$

$$\operatorname{Hess} f(\theta)[\xi] = -\frac{1}{n}\sum_{i=1}^{n} \nabla_\xi \log_\theta(\theta_i)$$

Tyler estimator for samples $\{\boldsymbol{x}_i\}$ on $\mathcal{S}_p^{++}$

$$\operatorname{Hess} f(\boldsymbol{\Sigma})[\xi] = p \operatorname{D}\Psi(\boldsymbol{\Sigma})[\xi] + p \operatorname{sym}(\xi\boldsymbol{\Sigma}^{-1}\Psi(\boldsymbol{\Sigma}))$$

$$\Psi(\boldsymbol{\Sigma}) = \sum_{i=1}^{n} \frac{\boldsymbol{x}_i\boldsymbol{x}_i^T}{\boldsymbol{x}_i^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x}_i} \qquad \operatorname{D}\Psi(\boldsymbol{\Sigma})[\xi] = \sum_{i=1}^{n} \frac{\boldsymbol{x}_i^T\boldsymbol{\Sigma}^{-1}\xi\boldsymbol{\Sigma}^{-1}\boldsymbol{x}_i}{(\boldsymbol{x}_i^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x}_i)^2}\boldsymbol{x}_i\boldsymbol{x}_i^T$$

Riemannian optimization – Newton method

descent direction

$\xi^{(k)}$ solution to

$$\operatorname{Hess} f(\theta^{(k)})[\xi^{(k)}] = -\operatorname{grad} f(\theta^{(k)})$$

update

$$\theta^{(k+1)} = R_{\theta^{(k)}}(\xi^{(k)})$$

Outline

Numerical ressources

- Matlab: https://www.manopt.org
- Python:

  Riemannian geometry: https://geomstats.github.io/

  Optimization:

  https://pymanopt.org
  https://geoopt.readthedocs.io/en/latest/
  https://github.com/mctorch/mctorch

  Autodifferentiation:

  pytorch, tensorflow
  https://github.com/HIPS/autograd
  https://jax.readthedocs.io/en/latest/

Julia: https://manoptjl.org/

Example with pymanopt

Task

Optimizing the negative log-likelihood of a Gaussian distribution over the manifold $\mathcal{M} = \mathbb{R}^d \times \mathcal{S}_d^+$.

Code: `https://replit.com/@fallingtree/Riemannian-Optimization-Gaussian-Likelihood?v=1`

Outline

Conclusion