Optimization on flag manifolds

Ke Ye

Department of Statistics, UChicago

July 10, 2017

work in progress with Lek-Heng Lim and Ken Sze-Wai Wong

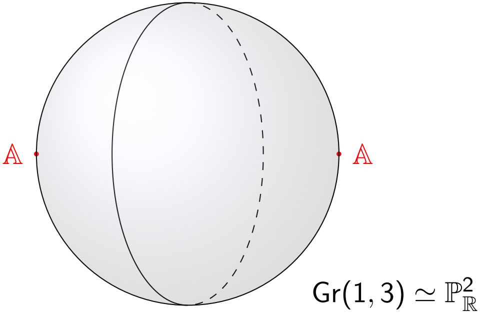
outline

motivation: data analysis

- objects: butterfly wings, movies, websites
- features:
  - butterfly wings: color, texture, shape, enviroment, etc.
  - movies: ratings, box office, cast, profit, etc.
  - websites: content relavence, updating frequency, metadata, click rates, etc.
- object: described by n features $\rightsquigarrow$ column vector $a \in \mathbb{R}^n$
- data set: k objects
- $a_1, \ldots, a_k \in \mathbb{R}^n \rightsquigarrow A = [a_1, \ldots, a_k] \in \mathbb{R}^{n \times k}$
- naive idea: analyze data sets on $\mathbb{R}^{n \times k}$
- bad news: A depends on choice of coordinates :(

motivation: data analysis

- good news: can use geometry !
- $\mathbb{A}$: span of $a_1, \ldots, a_k \in \mathbb{R}^{n \times k} \rightsquigarrow \mathbb{A} \in \mathrm{Gr}(k, n)$



- new idea: optimization on $\mathrm{Gr}(k, n)$ (Absil-Mahony-Sepulchre)

optimization algorithms on manifolds

- (M, g): Riemannian manifold
- $F : M \to \mathbb{R}$ smooth
- goal: a critical point $x \in M$ of F, i.e., $\nabla F(x) = 0$
- Newton's method on manifolds
- input: initial value x_0
- step 1: for $k = 0, 1, \ldots$ solve for $Y_k \in T_{x_k} M$

$$H(F)_{x_k}(Y_k, Y_k) = -\nabla F(x_k)$$

- step 2: minimize F along the geodesic $\gamma(x_k, Y_k)(t)$
- step 3: set $x_{k+1} = \gamma(x_k, Y_k)(t)$

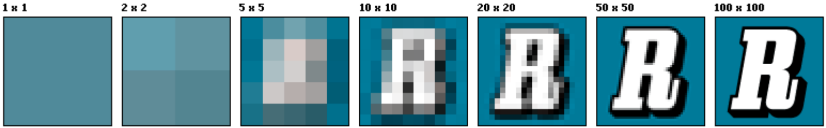
optimization on manifolds

- (M, g): Riemannian manifold
- ingredients: tangent spaces, geodesics, parallel transport, gradient and Hessian
- general M: no explicit formula!
- special M: homogeneous spaces G/H
- examples:
  - Stiefel: $V_k(\mathbb{R}^n) = \mathrm{O}(n)/\,\mathrm{O}(k)$
  - Grassmannian: $\mathrm{Gr}(k, n) = \mathrm{O}(n)/(\mathrm{O}(k) \times \mathrm{O}(n-k))$
  - flags: $\mathrm{Flag}_{n_1,\dots,n_d} = \mathrm{O}(n)/(\mathrm{O}(n_1) \times \cdots \times \mathrm{O}(n_d))$, $n_1 + \cdots + n_d = n$
  - fixed rank psd: $S_+(n, k) = GL_+(n)/\,\mathrm{Stab}_E$
  - shape: $(\mathbb{R}^{k\times n} \setminus \{0\})/\,\mathrm{O}(n)$
- this talk: flag manifolds $\mathrm{Flag}_{n_1,\dots,n_d}$

real world applications

- Stiefel: computer vision, principal component analysis and independent component analysis
- Grassmannian: data comparison, dimension reduction
- flags: quantum information, nested data analysis
- fixed rank psd: low rank approximation
- shape: shape analysis

- nested data analysis



- quantum information
  - simulate quantum systems
  - density matrix: state of a quantum system
  - orbit of a density matrix: all states of a quantum system
  - orbital optimization: complex flag manifold

review: flag manifolds

- $\mathbb{V}$: n-dimensional real vector space
- $d \leq n$
- flag of type $(n_1, \dots, n_d)$: $\{0\} \subsetneq \mathbb{V}_1 \subsetneq \cdots \subsetneq \mathbb{V}_d = \mathbb{V}$
- $\dim \mathbb{V}_j = \sum_{s=1}^{j} n_s$
- $\mathrm{Flag}_{n_1,\dots,n_d}$ = set of all flags of type $(n_1, \dots, n_d)$
- $Flag_{n_1,\dots,n_d}$ is a smooth manifold
- $Flag_{n_1,\dots,n_d}$ is a homogeneous space: $O(n)/(O(n_1) \times \cdots \times O(n_d))$
- $d = 2$: $\mathrm{Gr}(n_1, n)$
- over $\mathbb{C}$: complex flag manifolds $U(n)/(U(n_1) \times \cdots \times U(n_d))$

ingredients for optimization: tangent spaces

- $\text{Flag}_{n_1,n_2,n_3} = O(n)/(O(n_1) \times O(n_2) \times O(n_3))$
- $\mathbb{V}_1 \subsetneq \mathbb{V}_2 \subsetneq \mathbb{V}_3 = \mathbb{V} \longleftrightarrow [Q] = Q \cdot (O(n_1) \times O(n_2) \times O(n_3))$
- $Q = [Q_1, Q_2, Q_3] \in O(n)$: $Q_i \in \mathbb{R}^{n \times n_i}$, $Q_i^\mathsf{T} Q_i = I_{n_i}$, $i = 1, 2, 3$
- $T_{I_n} \text{Flag}_{n_1,n_2,n_3} = \mathfrak{so}(n)/(\mathfrak{so}(n_1) \times \mathfrak{so}(n_2) \times \mathfrak{so}(n_3))$
- $\mathfrak{so}(n) = n \times n$ skew-symmetric matrices

ingredients for optimization: tangent spaces

- $T_{[I_n]}\,\mathrm{Flag}_{n_1,n_2,n_3} = \left\{ \begin{bmatrix} 0 & B_{12} & B_{13} \\ -B_{12}^{\mathsf{T}} & 0 & B_{23} \\ -B_{13}^{\mathsf{T}} & -B_{23}^{\mathsf{T}} & 0 \end{bmatrix} : B_{ij} \in \mathbb{R}^{n_i \times n_j}, 1 \le i \ne j \le 3 \right\}$
- $T_{[Q]}\,\mathrm{Flag}_{n_1,n_2,n_3} = \left\{ Q \begin{bmatrix} 0 & B_{12} & B_{13} \\ -B_{12}^{\mathsf{T}} & 0 & B_{23} \\ -B_{13}^{\mathsf{T}} & -B_{23}^{\mathsf{T}} & 0 \end{bmatrix} : B_{ij} \in \mathbb{R}^{n_i \times n_j}, 1 \le i \ne j \le 3 \right\}$

ingredients for optimization: Riemannian metric and geodesics

- metric $\tilde{g}$ on $O(n)$: $\tilde{g}(Q_1, Q_2) = \text{tr}(Q_1^{\mathsf{T}} Q_2)$
- geodesics in $O(n)$ through $Q \in O(n)$: $Q\exp(tX), X \in \mathfrak{so}(n)$
- $\tilde{g}$ induces a metric g on $\text{Flag}_{n_1,n_2,n_3}$
- why g?
- $\text{Gr}(n_1, n)$: symmetric space
- $\text{Flag}_{n_1,n_2,n_3}$: not symmetric
- $(\text{Flag}_{n_1,n_2,n_3}, g)$ is a GO-space: every geodesic is obtained by a geodesic in $O(n)$
- geodesics in $\text{Flag}_{n_1,n_2,n_3}$: $Q\exp(tX) \cdot (O(n_1) \times O(n_2) \times O(n_3))$

another formula for geodesics

- need more computable formula
- $\gamma(t) = [Q \exp(tX)]$: geodesic
- $\gamma(0) = [Q], \gamma'(t) = QX$
- $X = VDV^{\mathsf{T}}$: spectral decomposition
- $D = \operatorname{diag}\left\{ \begin{bmatrix} 0 & \lambda_1 \\ -\lambda_1 & 0 \end{bmatrix}, \dots, \begin{bmatrix} 0 & \lambda_r \\ -\lambda_r & 0 \end{bmatrix}, 0_{n-2r} \right\}, V \in O(n)$
- $\gamma(t) = [U\Sigma(t)V^{\mathsf{T}}]$
- $\Sigma(t) = \operatorname{diag}\left\{ \begin{bmatrix} \cos t\lambda_1 & -\sin t\lambda_1 \\ \sin t\lambda_1 & \cos t\lambda_1 \end{bmatrix}, \dots, \begin{bmatrix} \cos t\lambda_r & -\sin t\lambda_r \\ \sin t\lambda_r & \cos t\lambda_r \end{bmatrix}, I_{n-2r} \right\}$
- $U, V \in O(n)$

distance on flag manifolds

- $[Q_1], [Q_2] \in \text{Flag}_{n_1, n_2, n_3}$
- Hopf-Rinow Thm $\implies$ $\exists$ geodesic $\gamma(t)$, $\gamma(0) = [Q_1], \gamma(1) = [Q_2]$
- $Q_1^{\mathsf{T}} Q_2 = V \Sigma V^{\mathsf{T}}$
- $\Sigma = \text{diag}\left\{ \begin{bmatrix} \cos\lambda_1 & -\sin\lambda_1 \\ \sin\lambda_1 & \cos\lambda_1 \end{bmatrix}, \dots, \begin{bmatrix} \cos\lambda_r & -\sin\lambda_r \\ \sin\lambda_r & \cos\lambda_r \end{bmatrix}, I_{n-2r} \right\}$
- $d([Q_1], [Q_2]) = \sqrt{\sum_{i=1}^{r} \lambda_i^2}$

ingredients for optimization: parellel transport

- $B, X \in \mathfrak{so}(n)$
- $\varphi_B(X) = \frac{1}{2}[B, X]_{\mathfrak{so}(n)} = \begin{bmatrix} 0 & -B_{12}X_{23}^{\mathsf{T}} + X_{12}B_{23}^{\mathsf{T}} & B_{11}X_{23} - X_{11}B_{23} \\ X_{23}B_{12}^{\mathsf{T}} - B_{23}X_{12}^{\mathsf{T}} & 0 & -B_{11}X_{12}^{\mathsf{T}} + X_{11}B_{12}^{\mathsf{T}} \\ -X_{23}^{\mathsf{T}}B_{11}^{\mathsf{T}} + B_{23}^{\mathsf{T}}X_{11}^{\mathsf{T}} & X_{12}B_{11}^{\mathsf{T}} - B_{12}X_{11}^{\mathsf{T}} & 0 \end{bmatrix}$
- $e^{-\varphi_B} = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \varphi_B^k$
- parallel transport of QX along $[Q\exp(tB)]$: $[Q\exp(tB)e^{-\varphi_{tB}}(X)]$
- $\mathrm{Gr}(n_1, n)$: $[Q\exp(tB)X]$ (Absil-Mahony-Sepulchre)

another description of $\text{Flag}_{n_1,n_2,n_3}$

- as projections:

$$\text{Flag}_{n_1,n_2,n_3} \simeq \{(P_1, P_2, P_3) : P_i \in \mathbb{R}^{n\times n}, P_i^2 = P_i, P_i^\mathsf{T} = P_i, \\ \text{tr}(P_i) = \sum_{s=1}^{i} n_s, P_j P_i = P_i, 1 \le i < j \le 3\}$$

- as quotient space: $\text{Flag}_{n_1,n_2,n_3} \simeq \text{St}(n, n_3)/(O(n_1) \times O(n_2))$
- equivalence class:

$$[A] = \left\{ A \begin{bmatrix} Q_1 & 0 & 0 \\ 0 & Q_2 & 0 \\ 0 & 0 & Q_3 \end{bmatrix} : A \in \mathbb{R}^{n\times(n_1+n_2)}, A^\mathsf{T} A = I_{n_1+n_2}, \\ Q_i \in O(n_i), i = 1, 2, 3 \right\}$$

ingredients for optimization: gradient

- $F : \mathrm{Flag}_{n_1,n_2,n_3} \to \mathbb{R}$ smooth
- gradient: $g(\nabla F, X) = X(F)$
- $[A] \in \mathrm{Flag}_{n_1,n_2,n_3}$, $A \in \mathbb{R}^{n\times(n_1+n_2)}, A^{\mathsf{T}}A = I_{n_1+n_2}$
- $A = [A_1, A_2]$
- $D = (\frac{\partial F}{\partial x_{ij}}) = [D_1, D_2], D_i \in \mathbb{R}^{n\times n_i}, i = 1, 2$
- gradient: $\nabla F([A]) = [\Delta_1, \Delta_2]$
- $\Delta_i = D_i - (A_i A_i^{\mathsf{T}} D_i + A_j D_j^{\mathsf{T}} A_i), 1 \leq i \neq j \leq 2$
- $\mathrm{Gr}(n_1, n) : D - AA^{\mathsf{T}}D$(Absil-Mahony-Sepulchre)

ingredients for optimization: Hessian

- $F : \mathrm{Flag}_{n_1,n_2,n_3} \to \mathbb{R}$ smooth
- Hessian $H(F)_{[A]}$:
  - symmetric bilinear form on $T_{[A]}\,\mathrm{Flag}_{n_1,n_2,n_3}$
  - $H(F)_{[A]}(\gamma'(0), \gamma'(0)) = \frac{d^2}{dt^2}\,|_{t=0}\, F(\gamma(t))$, $\gamma(t)$ geodesic
- $F_{A,A} = \left(\frac{\partial^2 F}{\partial x_{ij} \partial x_{kl}}\right)$
- Hessian of F:

$$H(F)_{[A]}(X,Y) = F_{A,A}(X,Y) + \frac{1}{2}\big(\mathrm{tr}(F_A^{\mathsf{T}} Q B^{\mathsf{T}} Q^{\mathsf{T}} X) + \mathrm{tr}(F_A^{\mathsf{T}} Q C^{\mathsf{T}} Q^{\mathsf{T}} Y) - \mathrm{tr}(F_A^{\mathsf{T}} Q (B+C)^{\mathsf{T}} Q^{\mathsf{T}} (X+Y))\big)$$

- $A = Q I_{n,n_1+n_2}, X = B I_{n,n_1+n_2}, Y = C I_{n,n_1+n_2}$
- $Q \in O(n), B, C \in \mathfrak{so}(n)$

summary

- assemble ingredients: Newton's method on $\text{Flag}_{n_1,\ldots,n_d}$
- numerical experiments: available soon
- numerical properties: convergence, stability?
- other metrics?
- other manifolds?

Thank you for your attention!