

Problem 1 Score: _____. **The Riemannian Newton method**

In this problem, you will implement Newton's method for optimization on the rotation matrices

$$\mathcal{M} = \text{SO}(d) = \{Q \in \mathbb{R}^{d \times d} : Q^T Q = I, \det(Q) = 1\}.$$

As usual, we consider $\mathcal{M} = \text{SO}(d)$ as a Riemannian submanifold of $\mathbb{R}^{d \times d}$ with the usual Frobenius inner product. Recall that orthogonal projection onto a tangent space of $\text{SO}(d)$ at Q ,

$$T_Q \text{SO}(d) = \{U \in \mathbb{R}^{d \times d} : Q^T U + U^T Q = 0\} = \{Q\Omega : \Omega \in \mathbb{R}^{d \times d}, \Omega + \Omega^T = 0\},$$

is given by

$$\text{Proj}_Q(U) = Q \frac{Q^T U - U^T Q}{2} = Q \text{skew}(Q^T U).$$

Consider the following optimization problem

$$\min_{Q \in \text{SO}(d)} f(Q), \quad f : \text{SO}(d) \rightarrow \mathbb{R}, \quad f(Q) = \|AQ - QB\|^2,$$

for symmetric $A, B \in \mathbb{R}^{d \times d}$.

- (1) Compute the Riemannian gradient $\text{grad } f(Q)$.
- (2) Compute the Riemannian Hessian $\text{Hess } f(Q)$.
- (3) Choose a retraction R on $\text{SO}(d)$.

Newton's method is given by $Q_{k+1} = R_{Q_k}(U_k)$, where the step U_k is the solution to the "Newton system":

$$\text{Hess } f(Q_k)[U_k] = -\text{grad } f(Q_k).$$

To solve the Newton system, you may need to write $\text{Hess } f(Q_k)$ and $\text{grad } f(Q_k)$ as a matrix and a vector, respectively. That is, you should choose a basis for $T_{Q_k} \text{SO}(d)$. There are two ways to do this.

- (4) Generate a random set of matrices in the embedding space $\mathbb{R}^{d \times d}$, and then project them onto the tangent space $T_{Q_k} \text{SO}(d)$. Will this form a linearly independent set of tangent vectors? How many random matrices do you need to obtain a basis for the tangent space?
- (5) Propose an explicit orthonormal basis for each tangent space $T_{Q_k} \text{SO}(d)$.

With a basis for each tangent space in hand, let us solve the Newton system. You will try three different methods. Attention, in the following ways to solve the Newton system we suppose that the Riemannian Hessian is positive definite. To be sure that this is the case you should first run Riemannian Gradient Descent to be close to the a minimizer of f .

- (6) Write $\text{Hess } f(Q_k)$ and $\text{grad } f(Q_k)$ in terms of a basis for $T_{Q_k} \text{SO}(d)$.
- (7) Solve the Newton system using the backslash symbol in MATLAB.
- (8) Solve the Newton system using the gradient descent method on

$$g : T_{Q_k} \text{SO}(d) \rightarrow \mathbb{R}, \quad g(U) = \frac{1}{2} \langle H[U], U \rangle - \langle b, U \rangle,$$

where $H = \text{Hess } f(Q_k)$ and $b = -\text{grad } f(Q_k)$. Determine an explicit expression for the optimal step size α_k , supposing that H is positive definite.

- (9) Solve the Newton system using the conjugate gradient method on g .
- (10) Compare the three methods in terms of the number of iterations and the computational cost.

Solution: (1) f can be smoothly extended by

$$\begin{aligned} \bar{f}(Q) &= \|AQ - QB\|^2 \\ &= \|A\|^2 + \|B\|^2 - 2\langle AQ, QB \rangle, \end{aligned}$$

whose derivative along $Z \in T_Q \text{SO}(d)$ at Q is

$$D\bar{f}(Q)[Z] = -2\langle AQ, ZB \rangle - 2\langle AZ, QB \rangle$$

$$\begin{aligned}
&= -2\langle Z, AQB^\top \rangle - 2\langle Z, A^\top QB \rangle \\
&= \langle Z, -2AQB - 2A^\top QB \rangle \\
&= \langle Z, -4AQB \rangle,
\end{aligned}$$

which implies that the gradient of \bar{f} at Q is

$$\text{grad } \bar{f}(Q) = -4AQB.$$

Then, the Riemannian gradient of f at Q is

$$\begin{aligned}
\text{grad } f(Q) &= \text{Proj}_Q(\text{grad } \bar{f}(Q)) \\
&= Q\text{skew}(Q^\top \text{grad } \bar{f}(Q)) \\
&= Q\text{skew}(Q^\top (-4AQB)) \\
&= -2Q(Q^\top AQB - BQ^\top AQ) \\
&= 2(QBQ^\top AQ - AQB)
\end{aligned}$$

(2) The Riemannian gradient can be smoothly extended by

$$\overline{\text{grad } f}(Q) = 2(QBQ^\top AQ - AQB),$$

whose derivative along $Z \in T_Q\text{SO}(d)$ at Q is

$$D\overline{\text{grad } f}(Q)[Z] = 2(ZBQ^\top AQ + QBZ^\top AQ + QBQ^\top AZ - AZB).$$

Then, the Riemannian Hessian of f at Q is

$$\text{Hess } f(Q)[Z] = \text{Proj}_Q(D\overline{\text{grad } f}(Q)[Z]).$$

(3) $R_Q(Z) = Q \exp(Q^\top Z)$.

(4) Since $\dim \text{SO}(d) = d(d-1)/2$, we need $d(d-1)/2$ random matrices to obtain a basis for the tangent space.

(5) Let e_1, \dots, e_d be the standard unit basis of \mathbb{R}^d . Then,

$$\left\{ \frac{e_i e_j^\top - e_j e_i^\top}{\sqrt{2}} : 1 \leq i < j \leq d \right\}$$

forms an orthonormal basis for skew-symmetric matrices. Thus,

$$\left\{ Q \left(\frac{e_i e_j^\top - e_j e_i^\top}{\sqrt{2}} \right) : 1 \leq i < j \leq d \right\}$$

forms an orthonormal basis for $T_Q\text{SO}(d)$.

(6) Let $n = \dim \text{SO}(d) = d(d-1)/2$ and Z_1, \dots, Z_n be the orthonormal basis for $T_Q\text{SO}(d)$. Then the Newton system

$$\text{Hess } f(Q)[U] = -\text{grad } f(Q)$$

can be written as

$$H\alpha = b,$$

where $H_{ij} = \langle \text{Hess } f(Q)[Z_i], Z_j \rangle$, $b_i = \langle -\text{grad } f(Q), Z_i \rangle$, and $\alpha = (\alpha_1, \dots, \alpha_n)^\top$. Then the solution to the Newton system is

$$U = \sum_{i=1}^n \alpha_i Z_i.$$

(7) $\alpha = H \setminus b$.

(8) Let $U \in T_{Q^k}\text{SO}(d)$. Then

$$\begin{aligned} Dg(U)[V] &= \frac{1}{2}\langle H[V], U \rangle + \frac{1}{2}\langle H[U], V \rangle - \langle b, V \rangle \\ &= \langle H[U] - b, V \rangle \end{aligned}$$

which implies that the gradient of g at U , $\text{grad } g(U)$ is $H[U] - b$.

For $\alpha \in \mathbb{R}$, Let

$$\begin{aligned} h(\alpha) &= g(U + \alpha V) \\ &= \frac{1}{2}\langle H[U + \alpha V], U + \alpha V \rangle - \langle b, U + \alpha V \rangle \\ &= \frac{1}{2}\langle H[U], U \rangle + \alpha\langle H[U], V \rangle + \frac{1}{2}\alpha^2\langle H[V], V \rangle - \langle b, U \rangle - \alpha\langle b, V \rangle \\ &= g(U) + \alpha\langle H[U] - b, V \rangle + \frac{1}{2}\alpha^2\langle H[V], V \rangle, \end{aligned}$$

which is minimized at

$$\alpha = -\frac{\langle H[U] - b, V \rangle}{\langle H[V], V \rangle}.$$

For the gradient descent method, $V = -\text{grad } g(U) = b - H[U]$, and thus

$$\begin{aligned} \alpha &= -\frac{\langle H[U] - b, b - H[U] \rangle}{\langle H[b - H[U]], b - H[U] \rangle} \\ &= \frac{\|V\|^2}{\langle H[V], V \rangle}. \end{aligned}$$

□