Riemannian inner product

1. Euclidean gradient

dot product!

$$f(x+\delta) \overset{\circledast}{\approx} f(x) + \underbrace{\nabla f(x)^T \delta}$$

Cauchy-Schwarz: $\nabla f(x)^T \delta \leq |\nabla f(x)| |\delta|$
w/ equality when $\nabla f = \delta$

2. Riemann

$$\circledast = f(x) + \nabla f(x)^T g^{-1}(x) \cdot g(x) \delta$$

$$= f(x) + \langle (g^{-1}(x) \nabla f(x)), \delta \rangle$$

Stiefel Manifold

1. Tangent space.

Trace out a curve $X(t) \in V_k(\mathbb{R}^n)$. Then

$$\forall t, \quad X(t)^T X(t) \equiv I_{k \times k}$$

$$\xrightarrow{\frac{d}{dt}} \quad X(t)^T X'(t) + X'(t)^T X(t) \equiv 0$$

$$\Rightarrow T_X V_k(\mathbb{R}^n) = \{ \xi : X^T \xi + \xi^T X = 0 \}$$

2. Retraction.

$$R_X(\xi) = (X + \xi)(I + \xi^T \xi)^{-1/2}$$

a. Always a m'fld.

$$R_X(\xi)^T R_X(\xi) = (I + \xi^T \xi)^{-1/2} \underbrace{(X+\xi)^T (X+\xi)}_{\underbrace{X^T X}_{I} + \underbrace{X^T \xi + \xi^T X}_{0 \text{ by } ①} + \xi^T \xi} (I + \xi^T \xi)^{-1/2}$$

$$= (I + \xi^T \xi)^{-1/2} (I + \xi^T \xi)(I + \xi^T \xi)^{-1/2}$$

$$= I$$

b. Differentially agrees w/ exp.

Take $\xi(t)$ with $\xi(0) = 0$

$$\frac{d}{dt} R_X(\xi) = (X + \xi')(I + \xi^T \xi)^{-1/2} + (X + \xi) \underbrace{\frac{d}{dt} (I + \xi^T \xi)^{-1/2}}_{= \frac{d}{dt}(I + \frac{1}{2}\xi^T \xi) = \frac{1}{2}\xi'^T \xi + \frac{1}{2}\xi^T \xi'}$$

$$\boxed{\sqrt{1+t^2} = 1 + \frac{t^2}{2} + O(t^4)}$$

$$\left. \frac{d}{dt} R_X(\xi) \right|_{t=0} = X + \xi'$$

$$\Rightarrow dR_X(0) = id \checkmark$$

$$\min_{x \in S^{n-1}} x^T A x \simeq \min_{\|x\|_2^2 = 1} \frac{1}{2} x^T A x$$

1 Eigenproblem

$$\Lambda(x;\lambda) = \frac{1}{2} x^T A x + \lambda \left(\frac{1}{2} - \frac{1}{2}\|x\|^2 \right)$$

$$\Rightarrow 0 = Ax - \lambda x$$

2 Intrinsic gradient.

$$P_x = I - xx^T$$

$$\Rightarrow \nabla_{S^{n-1}} f(x) = (I - xx^T) \nabla f = 2(I - xx^T) A x$$

3 First order algorithm

$$v_k \leftarrow 2(I - xx^T) A x$$

$$x_{k+1} \leftarrow \exp_{x_k}(-v_k \cdot \alpha_k)$$

$\boxed{\text{or}}$

$$x_{k+1} \leftarrow R_{x_k}(-v_k \cdot \alpha_k)$$

4 $\tilde{f}(x) = \|\nabla_{S^{n-1}} f(x)\|_2^2$ ← Turns all saddle points into local mins

$$\simeq \|\underbrace{(I - xx^T)}_{P} A x\|_2^2 = x^T A^T P A x = x^T A^2 x - (x^T A x)^2$$

$$\Rightarrow \nabla_{S^{n-1}} \tilde{f}(x) = 2P_x (A^2 x - 2A x x^T A x)$$

Same carries through...

$$\max_{X \in V_k(\mathbb{R}^n)} \|X^T A\|_F^2$$

① Relationship to PCA.

Write $A = U \Sigma V^T$ orthogonal

$$\|X^T A\|_F^2 = \|X^T U \Sigma V^T\|_F^2 = \|\underbrace{(U^T X)^T}_{\bar{X}} \Sigma\|_F^2$$

$$\Rightarrow \max_{\bar{X} \in V_k(\mathbb{R}^n)} \left\|\bar{X}^T \begin{pmatrix} \sigma_1 & & \\ & \sigma_2 & \\ & & \ddots \end{pmatrix}\right\|_F^2 \quad \text{with } \sigma_1 \geq \sigma_2 \geq \cdots$$

Take $\bar{X} = (e_1 \cdots e_k) \Rightarrow X = U\bar{X} = k$ dominant singular vecs.

② Robust PCA.

Write $A = \begin{pmatrix} | & | & \\ a_1 & a_2 & \cdots \\ | & | & \end{pmatrix}$. Want set of columns w/ bad error to be sparse.

$$\tilde{f}(X) := \sum_i \|X^T a_i\|_2 \leftarrow \text{not squared!}$$

③ Regularized PCA.

Want X to be sparse

$$\bar{f}(X) = \|X^T A\|_F^2 + \sum_{ij} |X_{ij}|$$

④ Intrinsic gradient. (see Absil p81)

Recall $T_X V_k(\mathbb{R}^n) = \{\xi : X^T \xi + \xi^T X = 0\} \subseteq \mathbb{R}^{n \times k}$

Can check: $[T_X V_k(\mathbb{R}^n)]^\perp = \{XS : S^T = S\}$

$\Rightarrow$ projection at X takes Z to

$$Z - X\,\mathrm{sym}(X^T Z) = (I - XX^T)Z + X\,\mathrm{skew}(X^T Z) \quad (*)$$

$$\mathrm{sym}(M) = \tfrac{1}{2}(M + M^T) \qquad \mathrm{skew}(M) = \tfrac{1}{2}(M - M^T)$$

We have $\nabla_{\mathbb{R}^{n \times k}} \|X^T A\|_F^2 = 2AA^T X$

$$\overset{(*)}{\Rightarrow} \nabla_{V_k(\mathbb{R}^n)} \|X^T A\|_F^2 = 2AA^T X - 2X\,\mathrm{sym}(X^T A A^T X)$$

Plug into retraction for 1st order optimization.