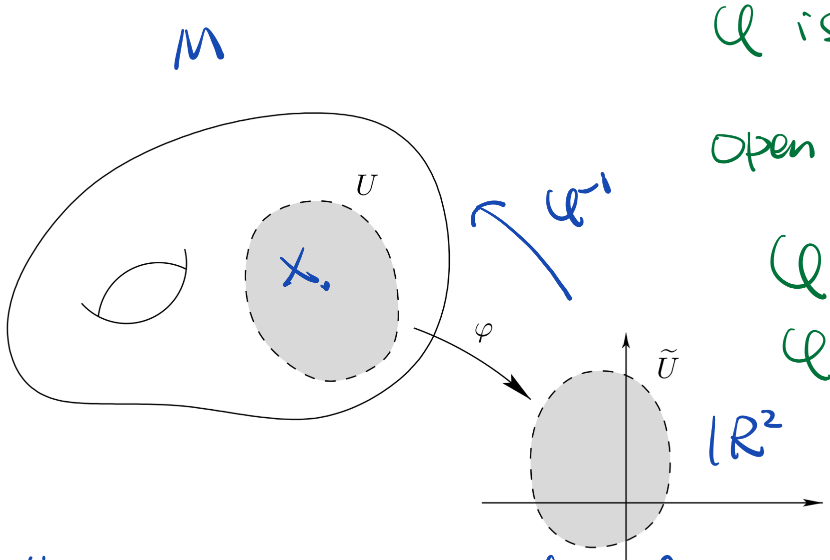
Definition of topological manifold, charts, smooth manifold



φ is homeomorphic $\begin{cases} \varphi \text{ is cont.} \\ \varphi^{-1} \text{ is cont.} \end{cases}$

open set U is called neighborhood

$\varphi : U \to \tilde{U}$

$\varphi^{-1} : \tilde{U} \to U$

<u>Definition</u> M is a topological manifold of dimension n

if M is a second countable Hausdorff topological space

and any $x \in M$ has a neighborhood U that is homeomorphic to

an open set $\tilde{U}$ in $\mathbb{R}^n$

<u>Def</u> (U, φ) is called a (coordinate) chart

<u>Def</u> A topological manifold is smooth if

any two charts are smoothly compatible:

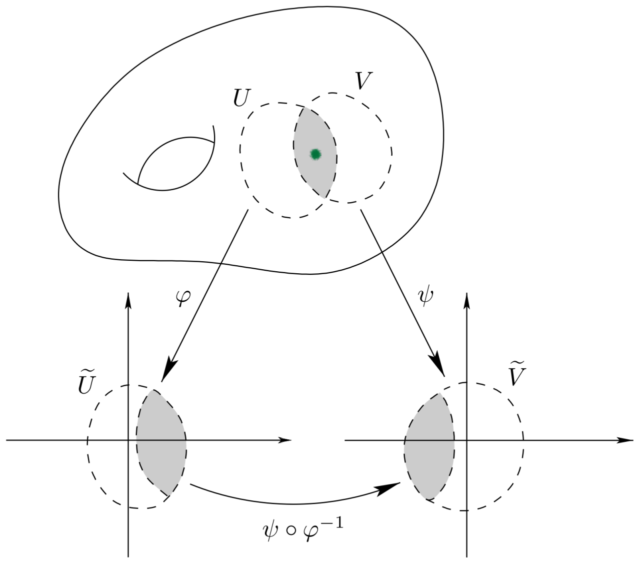
(U, φ) & (V, ψ) are smoothly compatible if

either $U \cap V = \emptyset$

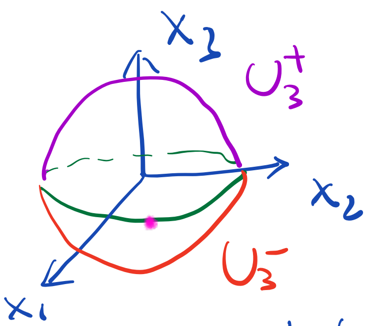
or $\psi \circ \varphi^{-1}$ is <u>smooth</u> derivatives of any order exists

$\underline{\psi \circ \varphi^{-1}} : \varphi(U \cap V) \to \psi(U \cap V)$

transition map



Example: $M = S^2 = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{R}^3 : x_1^2 + x_2^2 + x_3^2 = 1 \right\}$



$U_i^+ = \{ (x_1, x_2, x_3) \in S^2, \ x_i > 0 \}$

$U_i^- = \{ (x_1, x_2, x_3) \in S^2, \ x_i < 0 \}$

Define $\varphi_3^+ (x_1, x_2, x_3) = (x_1, x_2)$

$\varphi_3^- (x_1, x_2, x_3) = (x_1, x_2)$

Then $\varphi_3^\pm : S^2 \longrightarrow B^2 = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 : x^2 + y^2 \leq 1 \right\}$

is homeomorphism

Define $\varphi_i^\pm$ for $i = 1, 2$ similarly

Examples & Applications of Minimization over a manifold Constraint

I. Examples of manifolds

① Euclidean space $\mathbb{R}^n$ is a manifold of dim n

② Any surface such as paraboloid is a manifold of dim 2

③ Sphere in $\mathbb{R}^n$

$$S^{n-1} = \{x \in \mathbb{R}^n : x_1^2 + x_2^2 + \cdots + x_n^2 = 1\} = \{x \in \mathbb{R}^n : x^T x = 1\}$$

$A \in \mathbb{R}^{n\times n}$, $A^T = A$, $\min\limits_{\substack{x \in \mathbb{R}^n \\ x \neq 0}} f(x) = \dfrac{x^T A x}{x^T x}$ $\begin{cases} \text{minimizer is eigenvector } \vec{v}_n \\ \text{minimum is eigenvalue } \lambda_n \end{cases}$

$\lambda_n \leq \lambda_{n-1} \leq \cdots \leq \lambda_1$

$\forall a \in \mathbb{R}$, $f(x) = f(ax) \Rightarrow$ minimizers of $f(x)$ are not isolated

Instead, we should consider $\min\limits_{x \in S^{n-1}} f(x)$

④ Noncompact Stiefel manifold $n \geq p$

$$\mathbb{R}_*^{n\times p} = \{U \in \mathbb{R}^{n\times p} : \text{columns of } U \text{ are linearly independent}\}$$
$$= \{U \in \mathbb{R}^{n\times p} : UU^T \text{ is a rank } p \text{ matrix}\}$$
$$= \{U \in \mathbb{R}^{n\times p} : U^T U \text{ is invertible}\}$$

Guest Lecture on Nov 1st (Last Friday):

Low-rank approximation to Semi-Definite Programming

for K-means

⑤ (Compact) Stiefel manifold

$St(n,p) = \{ U \in \mathbb{R}^{n\times p} :$ columns of U are orthonormal $\}$

$= \{ U \in \mathbb{R}^{n\times p} : U^T U = I_p \}$

Theorem A.1 (Courant-Fischer-Weyl min-max principle). *Let λ_1 and λ_n be the largest and the smallest eigenvalues of a Hermitian matrix A, then for any vector $x \in \mathbb{C}^n$,*

$$\lambda_n \le \frac{x^* A x}{x^* x} \le \lambda_1.$$

$A\vec{v}_i = \lambda_i \vec{v}_i$

Facts for a real symmetric A:

$$\left.\begin{aligned} \vec{v}_n &= \underset{x\in\mathbb{R}^n\ \|x\|=1}{\arg\min} \frac{x^T A x}{x^T x} \\ \vec{v}_{n-1} &= \underset{x \perp \vec{v}_n,\ \|x\|=1}{\arg\min} \frac{x^T A x}{x^T x} \\ \vec{v}_{n-2} &= \underset{\substack{x \perp \vec{v}_n \\ x \perp \vec{v}_{n-1} \\ \|x\|=1}}{\arg\min} \frac{x^T A x}{x^T x} \end{aligned}\right\} \Rightarrow \min_{\substack{U\in\mathbb{R}^{n\times p} \\ U^T U = I_p}} tr(U^T A U)$$

$$= \lambda_n + \lambda_{n-1} + \cdots + \lambda_{n-p+1}$$

Eigenvalue problems are everywhere, see Monday's notes

⑥ 3D Rotation Group

$SO_3 = \{ R \in \mathbb{R}^{3\times 3} : R^T R = I,\ \det(R) = 1 \}$

e.g., $R = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Cryo-EM (Cryogenic electron microscopy)

uses electron microscope to study samples to find (bio) molecular structures

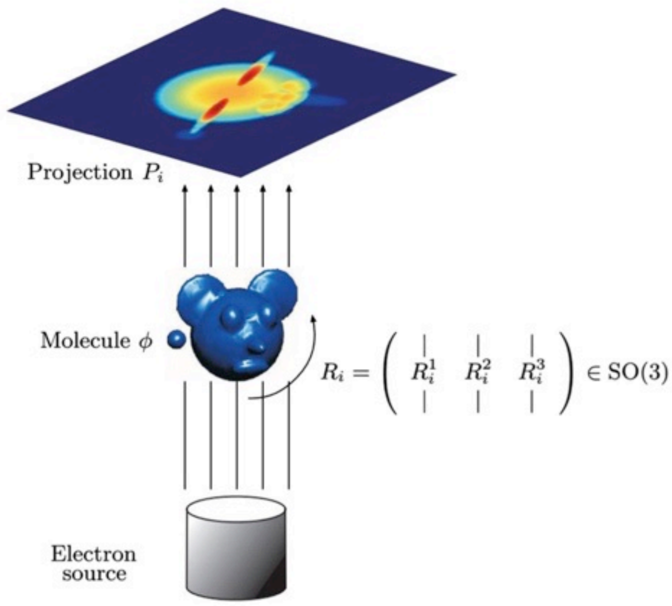


Fig. 2 Recover the 3-D structure from 2-D projections [8]

$$\min_{R} \|AR-B\|^2$$

Next, will introduce Riemannian metric and first order geometry such as Riemannian gradient.

Let us first consider the eigenvalue example

$$\min_{x \in S^{n-1}} \frac{1}{2} \frac{x^T A x}{x^T x} = \min_{x \in S^{n-1}} \underbrace{\frac{1}{2} x^T A x}_{f(x)}$$

$$\Longleftrightarrow \min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad x^T x = 1$$

Define Lagrangian as

$$L(x, \lambda) = f(x) - \frac{\lambda}{2}(x^T x - 1)$$

KKT system for saddle points

$$\begin{cases} \frac{\partial L}{\partial x} = \nabla f(x) - \lambda x = 0 & ① \\ \frac{\partial L}{\partial \lambda} = \frac{1}{2}(x^T x - 1) = 0 & ② \end{cases}$$

$\nabla f(x) = Ax$

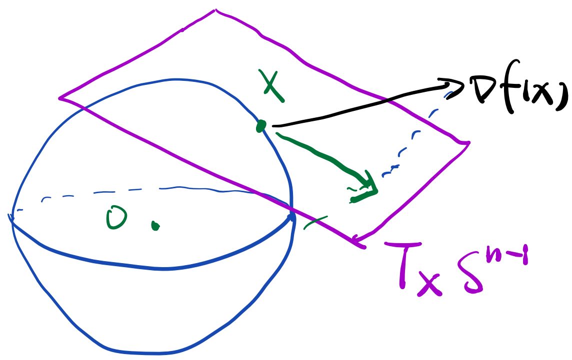
Multiply ① by x^T $\Rightarrow$ $x^TAx - \lambda x^Tx = 0 \Rightarrow \lambda = x^TAx$
$= x^T\nabla f(x)$

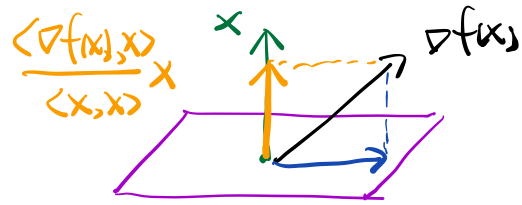
$$\Rightarrow \frac{\partial L}{\partial x} = \nabla f(x) - \langle x, \nabla f(x)\rangle x$$

To use ① & ②, we get an algorithm

$$\begin{cases} \tilde{x}_{k+1} = x_k - \eta\left[\nabla f(x_k) - \langle x_k, \nabla f(x_k)\rangle x_k\right] \\ x_{k+1} = \dfrac{\tilde{x}_{k+1}}{\|\tilde{x}_{k+1}\|} \end{cases}$$

$P_{S^{n-1}}(x_k - \eta \nabla f(x_k))$





Remark:

① $\nabla f(x) - \langle x, \nabla f(x)\rangle x$ $\begin{cases}\text{is the projection to tangent plane} \\ \text{is the Riemannian gradient on } S^{n-1}\end{cases}$

② For general surfaces, Lagrangian may not give Riemannian gradient

③ For surfaces, Riemannian gradient can be thought as projection to tangent plane

④ Advantage of manifold: a solid framework for convergence analysis