Part IV : Riemannian Optimization

- A rough example:

  Consider constrained minimization $\min_{x \in C} f(x)$

  When C is a Riemannian manifold,

  e.g. a surface like sphere

  the minimization algorith defined via geometry is called

  Riemannian Optimization.

- What is a manifold?

  The concept of manifold is an extension of surfaces

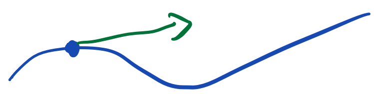
  to higher dimension and also more abstract sets

  ① Curve : $x(t) : \mathbb{R} \to \mathbb{R}^3$

  is a smooth mapping

  $x'(t)$ is tangent to

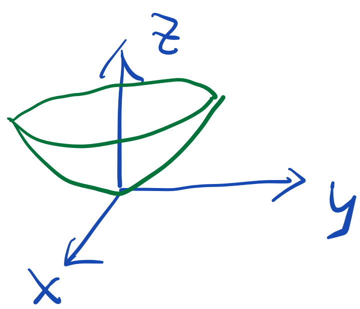
  the curve



  ② Surface :

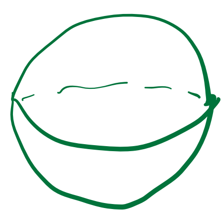
  a) $z = f(x, y)$

  $z = x^2 + y^2$



  b) $F(x, y, z) = C$

  $x^2 + y^2 + z^2 = 1$



③ Definition of a topological manifold M of dimension n:

1) M is a set   can be any set, such as set of $\begin{cases} \text{matrices} \\ \text{functions} \\ \text{mappings} \end{cases}$

2) M is a topological space (M has a topology $\mathcal{T}$)

a topological space is a set X with $\underbrace{\text{a collection}}_{\mathcal{T}}$ of its subsets

satisfying: ① $\phi \in \mathcal{T}$

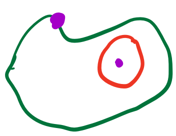
② $X \in \mathcal{T}$

③ Intersection of finitely many sets in $\mathcal{T}$ is in $\mathcal{T}$

④ Union of arbitrary number of sets in $\mathcal{T}$ is in $\mathcal{T}$

and sets in $\mathcal{T}$ are called open sets

Example: 1) $X = \mathbb{R}$, $\mathcal{T} = \{\text{any open interval in } \mathbb{R}\} \cup \{\phi\}$

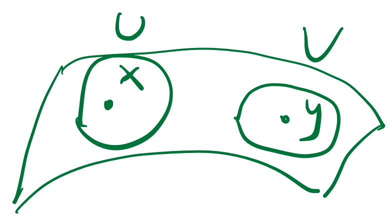


2) $X = \mathbb{R}^2$, $\mathcal{T} = \{\text{any open sets in } \mathbb{R}^2\} \cup \{\phi\}$

3) M is a Hausdorff space:

$\forall x, y \in M$, $\exists$ open subsets $U, V \subset M$ s.t.

$x \in U$, $y \in V$, $U \cap V = \phi$



S is countable if $S = \{x_1, x_2, x_3, \cdots, x_n, \cdots\}$

4) M is second countable:

there is a countable basis for topology $\mathcal{T}$

meaning that there are open sets $U_1, U_2, U_3, \cdots, U_n, \cdots$

s.t. any open set V can be generated by U_i via

either intersection or union.

Example: For $X = \mathbb{R}$, basis consists of open balls $B(x_0, r)$ centered at rational number x_0 with rational radius r

rational numbers are countable $\Rightarrow$ the basis is countable

an open interval (a,b) can be generated by

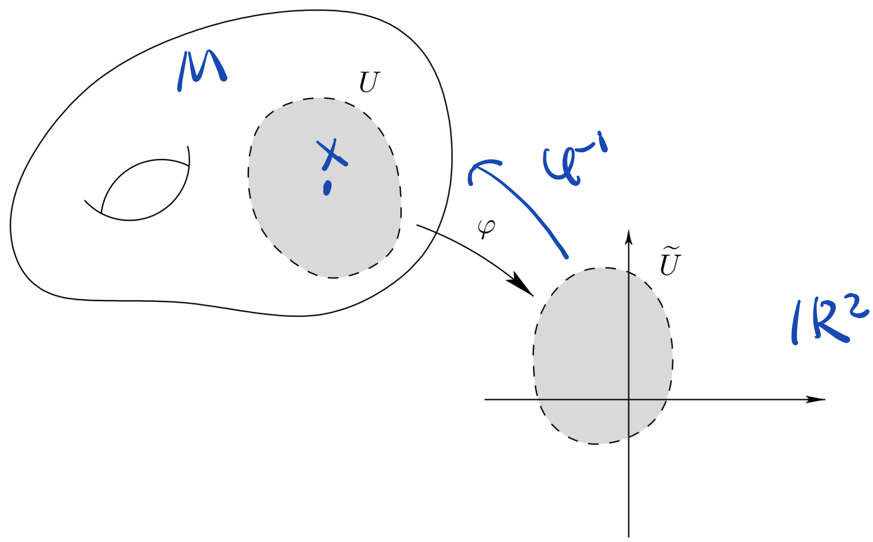
$$(a,b) = \bigcup_{\substack{x_0 \in \mathbb{Q} \\ r \in \mathbb{Q}, r>0 \\ B(x_0,r) \subset (a,b)}} B(x_0, r) \qquad B(x_0,r) = (x_0 - r, x_0 + r)$$

5) M is Locally Euclidean of dimension n:

$\forall x \in M$, $\exists$ $\underbrace{\text{open set } U \subset M \text{ s.t. } x \in U}_{\text{neighborhood of } x}$ and

U is homeomorphic to an open set in $\mathbb{R}^n$

a bijective continuous function φ and its inverse function is also continuous

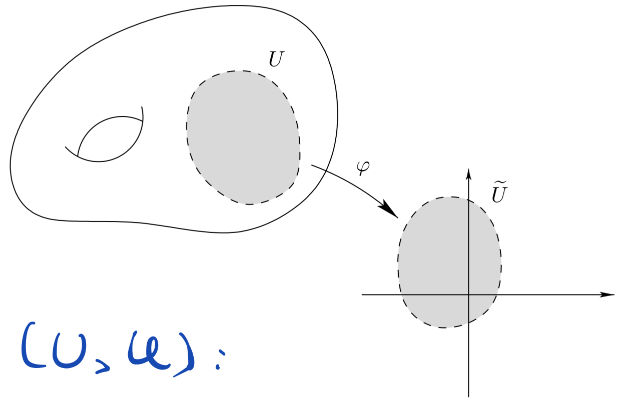


Definition M is a topological manifold of dimension n

if M is a second countable Hausdorff topological space

and any $x \in M$ has a neighborhood that is homeomorphic to an open set in $\mathbb{R}^n$

④ Coordinate Charts of M



Let M be a topological manifold of dim n

A Coordinate Chart on M is a pair (U, φ):

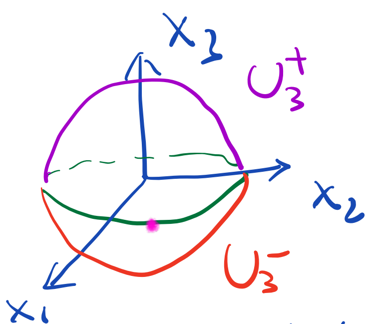
1) U is an open set

2) $\varphi: U \to \tilde{U} \subset \mathbb{R}^n$ is a homeomorphism

U is called coordinate domain/neighborhood

φ is a (local) coordinate 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

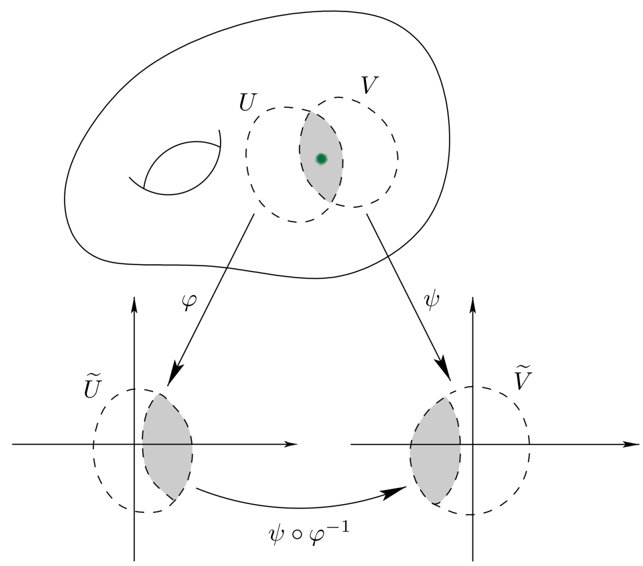
$\forall \vec{p} \in S^2$, $\exists$ U_i^+ or U_i^- s.t. $\vec{p} \in U_i^+$ or U_i^-

⑤ Smooth Manifolds

Let M be a topological manifold.

Let (U,φ) & (V,ψ) be two charts s.t. $U \cap V \neq \emptyset$

- $\underbrace{\psi \circ \varphi^{-1}}_{\text{transition map}} : \varphi(U \cap V) \to \psi(U \cap V)$ is also homeomorphic.



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

  either $U \cap V = \emptyset$

  or $\psi \circ \varphi^{-1}$ is $\underline{\text{smooth}}$ derivatives of any order exists

- An atlas of M is a collection of all charts

- A smooth atlas means that any two charts are smoothly compatible.

Some quick examples of Riemannian optimization

① Let $A \in \mathbb{R}^{n \times n}$

Want to find eigenvalues & eigenvectors λ & $\vec{v}$ s.t. $A\vec{v} = \lambda \vec{v}$

- Your Linear Algebra Class $A = \begin{pmatrix} 5 & 2 & 0 \\ 2 & 5 & 0 \\ -3 & 4 & 6 \end{pmatrix}$

$$|A - \lambda I| = -(\lambda-6)(\lambda-7)(\lambda-3) = 0$$

$$\Rightarrow \lambda_1 = 6, \quad \lambda_2 = 7, \quad \lambda_3 = 3$$

- $A \in \mathbb{R}^{n \times n}$, $|A - \lambda I|$ is a polynomial of degree n

  just find all its roots

$$\boxed{\text{Fundamental Theorem of Calculus} \qquad \int_a^b f'(x)\,dx = f(b) - f(a)}$$

Theorem (Fundamental Theorem of Algebra)

A polynomial of degree n with real or complex coefficients always has n complex roots

- But how exactly can we find all the roots?

1) [Abel Theorem] No root formula for polynomial of degree 5 and higher

2) "roots" function in MATLAB can easily find accurate approximations to roots.

3) For any polynomial

$$p(t) = t^n + a_{n-1}t^{n-1} + a_{n-2}t^{n-2} + \cdots + a_1 t + a_0,$$

it has a companion matrix

$$A = \begin{bmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & \cdots & 0 & -a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -a_{n-1} \end{bmatrix}$$

s.t. $|A - tI| = p(t)$

The "roots" function obtains approximation to roots by finding approximations to eigenvalues of its companion matrix via numerical linear algebra solvers, e.g., direct or iterative methods.

Remark: Eigenvalues are defined as roots of $|A - \lambda I|$

But roots of single variable polynomial are Numerically computed as eigenvalues of the companion matrix.

② Consider a real symmetric $A \in \mathbb{R}^{n\times n}$

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 $x \in \mathbb{C}^n$,*

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

$$\min_{x \in \mathbb{R}^n} \underbrace{\frac{x^T A x}{x^T x}}_{\text{Rayleigh Quotient}} \quad \begin{cases} \text{minimizer is } \vec{v}_n \\ \text{minimum value is } \lambda_n \end{cases}$$

- Let $f(x) = \dfrac{x^T A x}{x^T x}$, then $\nabla f(x) = \dfrac{2}{x^T x}(Ax - f(x)x)$

$$\nabla^2 f(x) z = \frac{2}{x^T x}(Az - f(x)z) - \frac{4}{(x^T x)^2}\left[x^T A z x + x^T z A x - 2f(x) x^T z x\right]$$

$\Rightarrow$ Newton's method produces $x_{k+1} = 2x_k$ as long as x_k is NOT an eigenvector

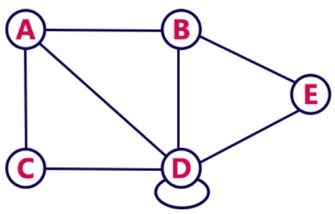
So always diverges.

- The key issue is $f(x) = f(ax)$, $\forall a \in \mathbb{R}$
- A remedy is to constrain $f(x)$ to $S^{n-1} = \{x \in \mathbb{R}^n : x^T x = 1\}$

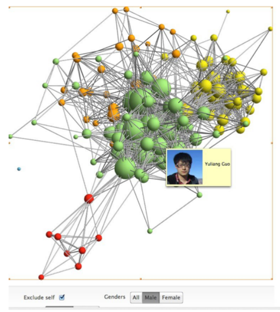
The Rayleigh quotient $f(x)$ on S^{n-1} is well-behaved and has isolated minimizers

③ Examples of Eigenvalues

Graph Adjacency Matrix



$$\begin{array}{c|ccccc}
 & A & B & C & D & E \\
\hline
A & 0 & 1 & 1 & 1 & 0 \\
B & 1 & 0 & 0 & 1 & 1 \\
C & 1 & 0 & 0 & 1 & 0 \\
D & 1 & 1 & 1 & 1 & 1 \\
E & 0 & 1 & 0 & 1 & 0
\end{array}$$



- A graph with n nodes can be represented as a $n \times n$ matrix A. If there is any edge from node i to node j, then $A_{ij} = 1$. Otherwise $A_{ij} = 0$.
- This matrix is called Graph Adjacency Matrix, which can quantify everything about graph. This is Spectral Graph Theory.
- Example: if graph is connected (there is a path between any two nodes), then the smallest eigenvalue of Graph Laplacian Matrix has algebraic multiplicity one.
- Example: in our social media network, is anyone a friend of friends of any other one?

Google's PageRank (Brin & Page, 98) is a "random walk on graph" method to approximate eigenvectors "Locally"

The "importance" of a website is proportional to the sum of the importance of all the sites that link to it.

A solution for the "importance" of a website

$$x_1 = \rho(x_{14} + x_{79} + x_{785})$$

$$x_2 = \rho(x_{1002} + x_{3225} + x_{9883} + x_{30027})$$

$$\cdots = \cdots$$

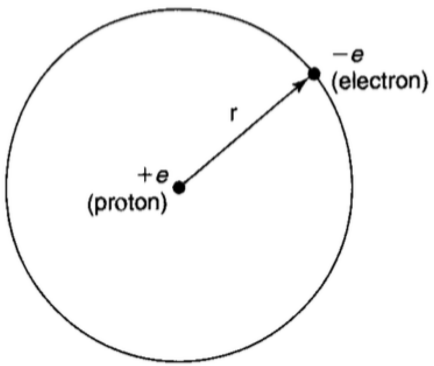
Solve $\quad x_i = \rho \sum_{j=1}^{n} a_{ij} x_j \quad$ for $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$

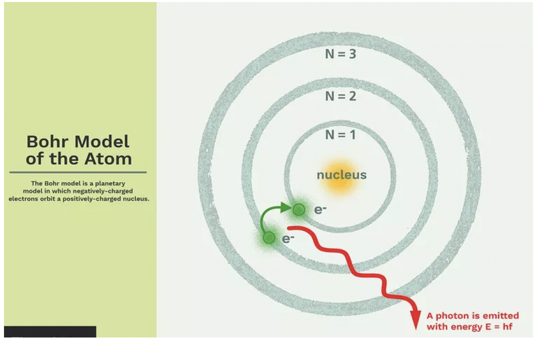
$$\boldsymbol{x} = \rho A \boldsymbol{x} \qquad A = \left[a_{ij} \right]_{n \times n}$$

Two-Slide Course of Quantum Mechanics

- Everything in physics is model but not necessarily truth. Example: $F = ma$, general relativity (for very large scales), quantum mechanics (for very small scales) are successful models/theories because they fit our observations well.
- Classical/Newtonian Mechanics Model: a particle (e.g., electron in Hydrogen atom), is located at some coordinate $\mathbf{x} = (x, y, z)$ at a given time t.
- Quantum Mechanics Model: we look for a complex valued wave function $\Psi(\mathbf{x}, t)$ for this particle and $|\Psi(\mathbf{x}, t)|^2$ is the probability of finding the particle at location $\mathbf{x}$ and time t.
- In 1D, wave function satisfies Schrödinger equation $i\frac{\partial \Psi}{\partial t} = -\frac{\partial^2 \Psi}{\partial x^2} + V(x)\Psi$ where $V(x)$ is an external potential (e.g., from proton).
- When solving this Schrödinger equation, one step is to find eigenvectors for the linear operator $\frac{d^2}{dx^2}$: we want to find $u(r)$ and λ satisfying

$$-\frac{d^2 u}{dr^2} = \lambda u$$





- We want to find $u(r)$ and λ satisfying $-\frac{d^2u}{dr^2} = \lambda u$.
- Here r is the distance between electron and proton: so $u(r)$ should also satisfy $u(0) = 0$.
- If we only want eigenvectors of $\frac{d^2}{dx^2}$, then eigenvectors include $e^{ax}, \cos(ax), \sin(ax), ax + b$ for any $a, b \in \mathbb{C}$: the eigenvalues can be any number since a, b are arbitrary.
- Here we also need eigenvector to satisfy $u(0) = 0$: eigenvectors can only be $u(r) = \cos[(n + \frac{1}{2})\pi r]$ with eigenvalue $\lambda = (n + \frac{1}{2})^2\pi^2$ (or $u(r) = \sin(n\pi r)$ with $\lambda = n^2\pi^2$), where n is integer.
- In Quantum Mechanics, eigenvalue λ is defined as the energy of particle.
- This perfectly explains Rutherford–Bohr Atom Model: electron has only discretized energy.