Class #5

Tuesday, February 4, 2025 10:28 AM

Example

① $f(x) = \frac{1}{2} \underline{x^T P x} + q^T x + r$; $f: \mathbb{R}^n \to \mathbb{R}$

$P \in S^n$ and $q \in \mathbb{R}^n$ and $r \in \mathbb{R}$

$g(x) = ax^2 + bx + c$

$g'(x) = 2ax + b$

$Df(x)$?

$\nabla f(x) = Df(x)^T$

$Df(x) = 2 \cdot \frac{1}{2} x^T P + q^T$

$\nabla f(x) = Px + q$

② $f(X) = \log(\det X)$; $\det X > 0$

↳ $\text{dom } f = S^n_{++}$ ⇒ Positive definite matrices

$Df(X)$ or $\nabla f(X)$?

Make use of the fact that $Df(X)$ provides a linear approximation of $f(Z)$, when Z is 'close' to X.

Consider $Z \in S^n_{++}$ such that Z is close to X

$\Rightarrow Z = X + \Delta X$, where $\Delta X \to \underline{0}$ as $Z \to X$

__Note__: We cannot claim that $\Delta X \in S^n_{++}$

Goal: show that $f(Z) = f(X) + \underbrace{Df(X)}(Z-X)$

as $Z \to X$ $\Rightarrow$ $Df(X)$ would be our derivative

$$Z = X + \Delta X$$

$$= \left(X^{1/2} \left(I + X^{-1/2} \Delta X X^{-1/2} \right) X^{1/2} \right)$$

$$\underbrace{\log\det(Z)}_{f(Z)} = \log\det\left(X^{1/2} \left(I + X^{-1/2} \Delta X X^{-1/2} \right) X^{1/2} \right)$$

$$= \log\left[\left(\det X^{1/2}\right) \left(\det\left(I + X^{-1/2}\Delta X X^{-1/2}\right)\right) \left(\det X^{1/2}\right) \right]$$

$$= \log\left[\underbrace{\det X^{1/2} \cdot \det X^{1/2}}_{\det X} \cdot \det\left(I + X^{-1/2}\Delta X X^{-1/2}\right) \right]$$

$$= \underbrace{\log\det X}_{f(X)} + \underbrace{\log\det\left(I + X^{-1/2}\Delta X X^{-1/2}\right)}_{\quad (Z-X)}$$

$I + \underbrace{X^{-1/2}\Delta X X^{-1/2}}$ $\Rightarrow$ This has eigenvalue decomposition

$\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of $X^{-1/2}\Delta X X^{-1/2}$

eigenvalues of $(I + A)$ = $1 +$ eigenvalues(A)

$1 + \lambda_i$, $i = 1, \ldots, n$ are the eigenvalues of

$\rightarrow$ $1+\lambda_i$, $i=1,\dots,n$ are the eigenvalues of

$$\det\left(I + X^{-1/2}\Delta X X^{-1/2}\right)$$

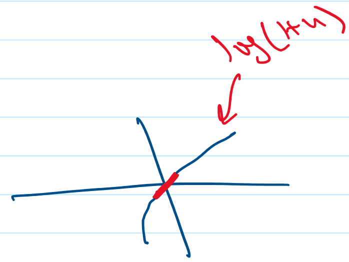
$$\hookrightarrow = \prod_{i=1}^{n}(1+\lambda_i)$$

$$f(Z) = f(X) + \log\left(\prod_{i=1}^{n}(1+\lambda_i)\right)$$

$$= f(X) + \sum_{i=1}^{n}\log(1+\lambda_i)$$

Since $\Delta X \to 0 \Rightarrow X^{-1/2}\Delta X X^{-1/2} \to 0$

$$\Rightarrow \lambda_i \to 0 \ \forall i$$



$\Rightarrow$ As $Z \to X$, $\lambda_i \to 0$

$\log(1+u) \simeq u$ for u very small

$\Rightarrow \sum_{i=1}^{n}\log(1+\lambda_i) \simeq \sum_{i=1}^{n}\lambda_i$ when $Z \to X$

$f(Z) = f(X) + \sum_{i=1}^{n}\lambda_i$ as $Z \to X$

$\downarrow$

eigenvalues of $X^{-1/2}\Delta X X^{-1/2}$

Σ eigenvalues of $A = tr(A)$

$$= f(X) + tr\left(X^{-1/2}\Delta X X^{-1/2}\right)$$

$$tr(ABC) = tr(CAB)$$

$$= f(X) + \mathrm{tr}\left(X^{-1/2} X^{-1/2} \Delta X\right)$$

$$= f(X) + \mathrm{tr}\left(X^{-1} \Delta X\right)$$

$$\hookrightarrow (Z - X)$$

$$f(Z) \doteq f(X) + \langle X^{-1}, Z - X \rangle$$

$$\nabla f(X) = X^{-1}$$

Chain rule

It applies in higher dimensions also.

<u>scalar:</u> $h(u) = g(f(u))$

$$h'(u) = g'(f(u))\, f'(u)$$

If $h(x) = g(f(x))$; $x \in \mathbb{R}^n$

$$Dh(x) = D_{f(x)}\, g(f(x)) \cdot D_x f(x)$$

<u>Example:</u> Say $g(x) = f(Ax+b)$

$$Dg(x) = Df(Ax+b) \cdot D(Ax+b)$$

$$= Df(Ax+b) \cdot A$$

$$\nabla g(x) = A^T \nabla f(Ax+b)$$

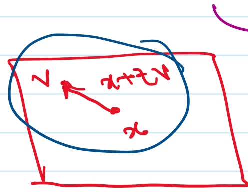
Directional derivative of a function

Let $f : \mathbb{R}^n \to \mathbb{R}$, $x \in \mathbb{R}^n$

Directional derivative along a vector $v \in \mathbb{R}^n$ at $x \in \mathbb{R}^n$ is defined as follows:

$$\tilde{f}(t) = f(x + tv) : \mathbb{R} \to \mathbb{R}$$

$$\longrightarrow f\left(x + t \frac{v}{\|v\|}\right)$$

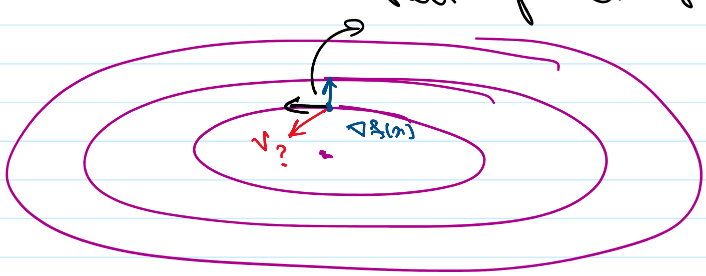


$$\tilde{f}'(t) = Df(x+tv) \cdot \underbrace{D(x+tv)}$$

$$= \nabla f(x+tv)^T \cdot v$$

Directional derivative of $f(x)$ in the direction v is defined as $\underline{\tilde{f}'(0) = \nabla f(x)^T v}$

rate of change $= 0$



Algorithms for unconstrained Optimization

$f : \mathbb{R}^n \to \mathbb{R}$; dom f

$$\min_{x \in \text{dom} f} f(x)$$

$$\min_{x \in \text{dom} f} f(x)$$

Assume the minimum value is attained by $f(x)$.

Define $p^* = \min_x f(x)$

An optimization algorithm is an iterative method that produces a sequence of points $x^{(0)}, x^{(1)}, \ldots, \in \text{dom} f$ such that

$$f(x^{(k)}) \longrightarrow p^* \quad \text{as } k \to \infty$$

In the case when $\arg\min_x f(x)$ is unique ($= x^*$) then we also hope that $x^{(k)} \longrightarrow x^*$ as $k \to \infty$

Suppose $f(\cdot)$ is continuous

$$f(x^{(k)}) \longrightarrow f(x^*) = p^*$$

Question: When to terminate the algorithm?

maybe: $f(x^{(k)}) - p^* \leq \epsilon$ for ϵ small $\Rightarrow$ terminate

↳ we do not know p^* $\Rightarrow$ not practical solution.

Search Direction-based Iterative Optimization Algorithms

Psuedo Code

Initialize: $x^{(0)} \in \text{dom} f$

$k \leftarrow 0$

while Stopping criterion not Satisfied

$$x^{(k+1)} \leftarrow x^{(k)} + t^{(k)} \Delta x^{(k)}$$

$$k \leftarrow k+1$$

do

$t^{(k)} \in \underbrace{\mathbb{R}_{++}}_{>0} \rightarrow$ Step size $\rightarrow$ (learning rate in stochastic optimization / machine learning)

$\Delta x^{(k)} \longrightarrow$ Search direction at time k

$x^{(k)} \Rightarrow$ Iterate

Main challenge: ① How to pick $\Delta x^{(k)}$?
② How to pick $t^{(k)}$?