Class #8

Friday, February 14, 2025 10:17 AM

$f \in C^1_L(\mathbb{R}^n) \Rightarrow f$ is L-Smooth $\rightarrow$ f cont.

Function is Smooth (or non-smooth) $\neq$

Quadratic Function $f(x) = \frac{1}{2} x^T Q x + p^T x + r$

Generally, we focus on $f(x)$ being convex

$\Rightarrow$ Q is Positive Semi definite or Positive definite

$f(x) = a x^2$

$\hookrightarrow$ Convex when $a > 0$ $\rightarrow$ otherwise concave.

Special case: $f(x) = \frac{1}{2} x^T Q x$; Q is P.D.

$Q \in S^n_{++}$

$f(x)$ in this case is L-Smooth

Linear algebra review (Appendix A.1.5)

Operator norm of a matrix

Eigenvalue Decomposition: only for "symmetric" "square" matrix.

$\hookrightarrow$ Singular Value Decomposition of a matrix

Every matrix $A \in \mathbb{R}^{m \times n}$ has a Singular value decomposition $\rightarrow$ for all mat

$A = U \Sigma V^T \rightarrow$ V is $n \times n$ matrix with orthonormal columns

$\hookrightarrow$ right singular vectors.

U: $m \times m$

Σ: $\hookrightarrow$ diagonal matrix $m \times n$

For rectangular matrix: only singular vaule decomposition
For Square symmetric matrix: both svd and eigenvale decomposition

m×m with orthonormal columns — left singular vectors of A

diagonal matrix m×n — singular values ≥ 0

r.s.v. singular vectors.

SVD is related to eigenvale decomposition!

$$A v_i = \sigma_i u_i$$

$$2\times 3 \ \Sigma = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \end{bmatrix}$$

$$\Sigma^T \Sigma = \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

one eigenvalue is '0'

$$AA^T = (U\Sigma \underbrace{V^T)(V}_{=I} \Sigma^T U^T)$$

$$= \underbrace{U\Sigma\Sigma^T U^T}_{(m\times n)(n\times m)} = \underbrace{U \underset{(m\times m)}{\Lambda} U^T}_{\text{Eigenvalue decomposition of } AA^T}$$

entries are σ_i^2

AA^T: symmetric & S.P.

① Singular values of A are the squareroot eigenvalues of AA^T / A^TA

② Left singular vectors of A are the eigenvectors of AA^T

$$A^TA = \underset{n\times n}{V} \underset{n\times n}{\Lambda} \underset{n\times n}{V^T}$$

entries are σ_i^2 — $\Sigma^T\Sigma$

eigenbevctor are directions that are invariants under the action of the matrix

Singular vector are the direction which get the most inflation.

③ Right singular vectors of A are the eigenvectors of A^TA

The operator norm of a matrix A with respect to the Euclidean norm is defined as:

operator norm : when it operates on a vector, how much it inflate the size of the vector ?

$$\|A\|_2 = \sup\{\underbrace{\|Ax\|_2}\ : \|x\|_2 \le 1\}$$

$$\|A\|_2 = \sigma_{max}(A) \Rightarrow \text{maximum singular value of } A$$

$$= \sqrt{\lambda_{max}(A^TA)} = \sqrt{\lambda_{max}(AA^T)}$$

maximum eigen value of A^TA

Submultiplicative property

$$\|Az\|_2 \le \|A\|_2 \|z\|_2$$

operator norm is the worst case elongation.

General Definition of Operator Norm

$$\|A\|_{a,b} = \boxed{\sup\{\|Ax\|_a : \|x\|_b \le 1\}}$$

this is the worst elongation that A will cause to X in the a norm, when X is bounded by 1 in the b norm

Special Cases: $a = b$

① $a = 2$

② $a = 1$

③ $a = \infty$

the euclueudian op. norm

$$\|A\|_1 = \max_{j=1,\dots,n} \sum_{i=1}^{m} |A_{ij}|$$

max L1 norm of the columns

$$\|A\|_\infty = \max_{i=1,\dots,m} \sum_{j=1}^{n} |A_{ij}|$$

.... rows

Problem: Let $\underbrace{f(x) = \frac{1}{2}x^TQx}$ with $Q \in S^n$.

Might not be convex → if $Q \to PD$

Prove that $f(x)$ is L-smooth and derive the

OPT, PSD

Prove that $f(x)$ is $\underbrace{\text{L-smooth}}_{f \in C^1_L}$ and define the value of L.

Solution:

$$\nabla f(x) = Qx$$

Let x and $y \in \mathbb{R}^n$

$$\|\nabla f(x) - \nabla f(y)\|_2 = \|Qx - Qy\|_2$$

the smoothness parameter (L)
is defined by eignevale

the growth of the gradient is a
function of the max eigenvaule of Q

$$= \|Q(x-y)\|_2$$

the bigger the eigenavlue ,
the bigger the L,
the faster the gradienet are growing,
the smaller step size to be taken (1/L) ,
cause now the function is changing fast !

$$\le \underbrace{\|Q\|_2}_{L} \|x-y\|_2$$

$$L = \|Q\|_2 = \sqrt{\lambda_{max}(Q^T Q)}$$

$$= \sqrt{\lambda_{max}(Q^2)} = \sqrt{\lambda_{max}(Q)^2}$$

$$= \lambda_{max}(Q)$$

Q is sym.

$Q \in S^n$

so eigenv exists

A with EVD $A = U \Lambda U^T$

$$A^n = U \Lambda^n U^T$$

46 min

Does Gradient Descent Converge?

If it does, at what rate?

Assume $f \in C^1_L(\mathbb{R}^n)$ and $p^* = \arg\min f(x)$ exists.

$\Downarrow$

$\Downarrow$

$$f(y) \le f(x) + \nabla f(x)^T (y-x) + \frac{L}{2}\|y-x\|_2^2$$

Put $y = x^{(k+1)} = x^{(k)} - t^{(k)}\nabla f(x^{(k)})$

$x = x^{(k)}$

$t^{(k)} = \frac{1}{L}$

$$f(x^{(k+1)}) \le f(x^{(k)}) - \frac{1}{L}\|\nabla f(x^{(k)})\|_2^2 + \frac{1}{2L}\|\nabla f(x^{(k)})\|_2^2$$

$$f(x^{(k+1)}) \le f(x^{(k)}) - \frac{1}{2L}\|\nabla f(x^{(k)})\|_2^2$$

↪ Descent lemma with $t^{(k)} = \frac{1}{L}$ and gradient descent

★ $$f(x^{(k+1)}) - f(x^{(k)}) \le -\frac{1}{2L}\|\nabla f(x^{(k)})\|_2^2$$

Add the above expression for $k = 0, 1, \ldots, \infty$

$$\underbrace{\sum_{k=0}^{\infty}\left[f(x^{(k+1)}) - f(x^{(k)})\right]}_{\text{telescoping sum}} \le -\frac{1}{2L}\sum_{k=0}^{\infty}\|\nabla f(x^{(k)})\|_2^2$$

↪ $\lim_{k\to\infty} f(x^{(k+1)}) = f(x^{(\infty)})$

$f(x^{(\infty)}) - f(x^{(0)})$

$$\sum_{n=0}^{\infty}(a_{n+1} - a_n) \Longrightarrow \text{telescoping sum}$$

$$= \lim_{N\to\infty} \underbrace{\sum_{n=0}^{N}(a_{n+1} - a_n)}$$

$$(a_{N+1} - \cancel{a_N}) + (\cancel{a_N} - \cancel{a_{N-1}}) + (\cancel{a_{N-1}} - \cancel{a_{N-2}}) + \cdots + (\cancel{a_1} - a_0)$$

$$= \lim_{N\to\infty} (a_{N+1} - a_0) = a_\infty - a_0$$

if $\lim_{N\to} a_N = a_\infty$

$$\Rightarrow f(x^{(\infty)}) - f(x^{(0)}) \le -\frac{1}{2L} \sum_{k=0}^{\infty} \|\nabla f(x^{(k)})\|_2^2$$

$$\sum_{k=0}^{\infty} \|\nabla f(x^{(k)})\|_2^2 \le 2L \cdot [f(x^{(0)}) - \underbrace{f(x^{(\infty)})}_{\ge p^*}]$$

$$\le 2L(f(x^{(0)}) - p^*)$$

B/C $\sum_{k=0}^{\infty} \|\nabla f(x^{(k)})\|_2^2 < \infty$

$$\Rightarrow \lim_{k\to\infty} \|\nabla f(x^{(k)})\|_2^2 = 0 \iff \lim_{k\to\infty} \nabla f(x^{(k)}) = 0$$

↪ we have converged to a local minimum or a saddle point.

↪ Basically a first-order stationary point.