Optimization HW 4

er2978

October 13, 2022

1 Question 1

Proof. Take the function $f(x,y) =: x^2 - y^2$, Then

$$\nabla f = \begin{bmatrix} 2x \\ -2y \end{bmatrix}$$

and

$$\nabla^2 f = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}.$$

Consider also the point $(x^*, y^*) = (0,0)$. The value of f at this point is 0. As we see, both second order necessary conditions are fulfilled since $\nabla f = 0$ and $\nabla^2 f$ is positive semidefinite. But, this is a saddlepoint so it is not a local minimum. $\square$

2 Question 2

Proof. Suppose FTSOC that the statement is not true: $\implies \exists x^* \in \mathbb{R}^n$ s.t. $f(x^*) \leq y$ but $\rho = (1-\alpha)x^* + \alpha x \notin L(y)$.[1] Note that $(1-\alpha)x^* \leq x^*$ and $\alpha x \leq x$. In particular if $(1-\alpha)x^* = x^*$ then $\alpha x = 0$, and vice versa. $\implies \min(x^*, x) \leq \rho \leq \max(x^*, x)$ given α's range.[2] $\therefore$ for $f(\rho) > y$, f cannot be a convex function, since $f(x) \leq f(\rho) \leq f(x^*)$ on a convex curve.[3]

$\Rightarrow\Leftarrow \quad f(\rho) \notin L \implies$ f is not convex is a contradiction. So $L(y)$ is a convex set. $\square$

[1] WLOG we assign $1-\alpha$ to x^*.

[2] min and max refer to the position on the surface described by f: in particular x and x^*'s closeness to y.

[3] WLOG we can say this.

3 Question 3

Proof. WTS: $\forall x' \in L(y), x' \in H(x)$; i.e. every $x' \in L(y)$ satisfies: $\nabla f(x)^T(x' - x) \leq 0$.

Take $x^* \in L(f(x))$. Then $f(x^*) \leq f(x)$ for some fixed x. $\implies f(x^*) - f(x) \leq 0$.

Let's Taylor expand: $f(x^*) \approx g(x^*) = f(x) + \nabla f(x)^T(x^* - x) + O(||x^* - x||_2^2)$. We drop the $O(||x^* - x||_2^2)$ for a second order Taylor approximation of $f(x^*)$. Note that f is convex, so the Taylor approximation is an underestimate. $\implies g(x^*) = f(x) + \nabla f(x)^T(x^* - x) \leq f(x^*) \implies \nabla f(x)^T(x^* - x) \leq f(x^*) - f(x) \leq 0$.

Therefore we have that if $x^* \in L(f(x))$ then $\nabla f(x)^T(x^* - x) \leq 0 \implies x^* \in H(x')$. $\square$

4 Question 4

Proof. Take $x' \in L(f(x)) \cap \{x' \in \mathbb{R}^n : \nabla f(x)^T(x' - x) = 0\}$ for some fixed x.

x' satisfies: $f(x') \leq f(x)$ and $\nabla f(x)^T(x' - x) = 0$

Suppose FTSOC $x' \neq x$. Then $x' - x \neq 0$ meaning that in order for $\nabla f(x)^T(x' - x) = 0$ to be satisfied, we need $\nabla f(x)^T = 0 \implies f(x)$ is a unique global minimum on the surface of f (given that it is strictly convex). Therefore $f(x') \neq f(x) \implies f(x') \geq f(x)$ which directly contradicts $f(x') \leq f(x)$.

$\Rightarrow\Leftarrow$ we have a contradiction so our supposition is false, and the original statement is true. $\square$

5 Question 5

Ehhh

Proof. Let's Taylor Expand: $f(x+\alpha p) \approx f(x)+\nabla f(x)^T(\alpha p)+\frac{1}{2}(\alpha p)^T\nabla^2 f(x)(\alpha p)+O(||\alpha p||_2^3)$ We know that $\nabla f(x)^T(\alpha p) < 0$ since p points away from $\nabla f(x)$. We also know that $(\alpha p)^T\nabla^2 f(x)(\alpha p) > 0$ since $\nabla^2 f$ is positive definite.

$\Rightarrow$ We need α such that $\frac{1}{2}(\alpha p)^T\nabla^2 f(x)(\alpha p) + O(||\alpha p||_2^3) < |\nabla f(x)^T(\alpha p)|$ in order for $f(x+\alpha p) < f(x)$.

$\frac{1}{2}(\alpha p)^T\nabla^2 f(x)(\alpha p) + O(||\alpha p||_2^3) < |\nabla f(x)^T(\alpha p)|$

$\Rightarrow \frac{1}{2}\alpha p^T\nabla^2 f(x)p + O(||p||_2^3) < |\nabla f(x)^T p|$

$\Rightarrow \frac{1}{2}\alpha p^T\nabla^2 f(x)p < |\nabla f(x)^T p| + O(||p||_2^3)$

$\Rightarrow \alpha p^T\nabla^2 f(x)p < 2[|\nabla f(x)^T p| + O(||p||_2^3)]$

$\Rightarrow \alpha < \dfrac{2[|\nabla f(x)^T p| + O(||p||_2^3)]}{p^T\nabla^2 f(x)p}$.

pick α such that this is satisfied. $\square$

6 Question 6

Proof.

- $Aq_i = \lambda_i q_i \implies$

  $$\text{When } i=1,\ i \to n: \quad Aq_i = \begin{bmatrix} \lambda_1 q_{11} & \lambda_2 q_{21} & \cdots & \lambda_n q_{n1} \\ \lambda_1 q_{12} & \lambda_2 q_{22} & \cdots & \lambda_n q_{n2} \\ \vdots & & & \\ \lambda_1 q_{1n} & \lambda_2 q_{2n} & \cdots & \lambda_n q_{nn} \end{bmatrix}$$

  We can pull out:

  $$\Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

  S.t. $AQ = \Lambda Q$.

  (Note that since Λ is a matrix of unique scalars per each q_i it can be multiplied on either side of Q.)

  $AQ = Q\Lambda \implies (AQ)Q^T = \Lambda Q Q^T \implies (A)^{-1} = A = (QQ^T\Lambda)^{-1} = \Lambda Q Q^{-1} = Q\Lambda Q^T$.

- Take the ith entry in each matrix from $Q\Lambda Q^T$. the product we get is $q_i \lambda_i q_i^T$ and since λ_i is a scalar we can rearrange as follows: $\lambda_i q_i q_i^T$. As $i \to n$ beginning at $i = 1$, we get $\sum_{i=1}^{n} \lambda_i q_i q_i^T$. Both the matrix product and the sum count the same information, so they are equivalent statements.

- 1. $\rightarrow$ If $\forall \lambda \in \Lambda$, we have $\lambda > 0$, then A is positive definite. We need $\forall q_i \in Q, \quad q_i A q_i^T > 0$.

     By (1) we know that $A = Q\Lambda Q^T$. We know $\forall \lambda \in \Lambda$, we have $\lambda > 0$, so let's test if A is positive definite.

     Take arbitrary vector x: $\quad x^T A x = x^T(Q^T \Lambda Q)x = (x^T Q^T)\Lambda(Qx) = (Qx)^T \Lambda (Qx)$.Let $Qx = v$. Then we have $\sum_{i=1}^{n} \lambda_i (v)^2$ where $v^2 > 0$. Therefore $\forall \lambda_i, \quad \lambda_i > 0$.

  2. $\leftarrow$ If A is positive definite then $\lambda_i > 0$.

     For A to be positive definite: $x^T A x > 0$ for any x. Let x be an eigenvector. Then $x^T A x = x^T \lambda x = \lambda x^T x = \lambda > 0$. Therefore $\forall \lambda \in \Lambda, \quad \lambda > 0$.

$\square$

7 Question 7

Proof. Let $l = x^* + th$ where $t = [0, 1]$ be a parameterization between 2 x's in S, where x^* is a global minimum. Then we have:

- $F(t) = f(l) = f(x^* + th)$
- $F'(t) = \nabla f(x^* + th)^T h$
- $F''(t) = h^T \nabla^2 f(x^* + th)h$

$$F'(1) = F'(0) + \int_0^1 F''(t)dt = \nabla f(x^*)^T h + \int_0^1 h^T \nabla^2 f(x^* + th)hdt = 0 + \int_0^1 h^T \nabla^2 f(x^* + th)hdt > 0$$

Since the Hessian is positive definite by assumption, then $h^T \nabla^2 f(x^* + th)h > 0$ so the integral is positive. $\implies$ the gradients of all points emanating from x^* in the set S will be positive.

$$F(1) = F(0) + \int_0^1 F'(t)dt = f(x^*) + \int_0^1 \nabla f(x^* + th)^T hdt = f(x^*) + C$$

where $C > 0$. $\implies$ $F(1) > F(0)$ i.e. every point in S generates a value through f that is larger than $f(x^*)$.

Therefore, $f(x^*)$ is a global minimum on a convex set S when $\nabla^2 f$ is positive definite.

□

8 Question 8

Proof. We have $\nabla g(x) = \nabla f(Ax+b) = A\nabla f(Ax+b)$ and $\nabla^2(g) = A^T\nabla^2 f(Ax+b)A$.

Newton:

$x_{n+1} = x_n - (\nabla^2(g(x_n))^{-1}\nabla g(x_n))$

$= x_n - (A^T\nabla^2 f(Ax+b)A)^{-1}A\nabla f(Ax+b)$

$= x_n - A^{-1}\nabla^2 f(Ax+b)^{-1}[(A^T)^{-1}A]\nabla f(Ax+b)$

Multiply both sides by A on the left, then add b:

$Ax_{n+1} = Ax_n - \nabla^2 f(Ax+b)^{-1}\nabla f(Ax+b)$

$Ax_{n+1} + b = Ax_n + b - \nabla^2 f(Ax+b)^{-1}\nabla f(Ax+b)$

Set $Ax_n + b = y \implies y_{n+1} = y_n - \nabla^2 f(y_n)^{-1}\nabla f(y_n)$ is the same as the Newton Step for f. $\square$