CONVEX OPTIMIZATION HOMEWORK 3

TOMMENIX YU
ID: 12370130
STAT 31015
DUE WED JAN 25, 2023, 3PM

Exercise 1.

Proof.

$\underline{\nabla^2_{xx} f \succeq 0 \Rightarrow f \text{ is convex:}}$ (same as in notes)

For $x, y \in \mathrm{dom}(f)$, define

$$\Gamma := f(y) - f(x) - \nabla f(x)^T (y - x)$$

and

$$g(t) = f(x + t(y - x)).$$

Note that by the first order condition of convexity, we only need to check that $\Gamma \geq 0$ for convexity. Moreover, we know that $g(0) = f(x)$, $g(1) = f(y)$ and $g'(0) = \nabla f(x)^T (y - x)$, which means that

$$\Gamma = g(1) - g(0) - g'(0) = \int_0^1 g'(t)dt - g'(0) = \int_0^1 g'(t) - g'(0)dt = \int_0^1 \int_0^t g''(\tau)d\tau dt$$

where we can just compute

$$\begin{aligned} g(t) &= f(x + t(y - x)) \\ \Rightarrow g'(t) &= \nabla f(x + t(y - x))^T (y - x) \\ \Rightarrow g''(t) &= (y - x)^T \nabla^2_{xx} f(x + t(y - x))(y - x) \end{aligned}$$

where by our assumptions we get that $g''(t) \geq 0$ for any t that makes $x + t(y - x) \in \mathrm{dom}(f)$, i.e. $t \in [0, 1]$. But this means

$$\Gamma = \int_0^1 \int_0^t g''(\tau)d\tau dt \geq 0$$

which by our argument above implies that f is convex.

$\underline{f \text{ is convex} \Rightarrow \nabla^2_{xx} f \succeq 0:}$

Assume, for the purpose of contradiction, that $v'^T \nabla^2_{xx} f(x) v' < 0$ for $x \in \mathrm{dom}(f)$. Then, since $\mathrm{dom}(f)$ is open we can find a corresponding $v = cv'$ such that $x + v \in \mathrm{dom}(f)$. Now by the first order condition we have that

$$f(x + v) \geq f(x) + \nabla f(x)^T \cdot v.$$

And by Taylor we have

$$f(x+v) = f(x) + \nabla f(x)^T \cdot v + v^T \nabla^2_{xx} f(x) v + o(v^2).$$

Now we can choose $|v|$ small so that

$$|v^T \nabla^2_{xx} f(x) v| - |v^T \nabla^2_{xx} f(x) v + o(v^2)| \leq \varepsilon$$

for any fixed ε. Thus we can conclude that

$$v^T \nabla^2_{xx} f(x) v + o(v^2) < 0$$

since

$$v^T \nabla^2_{xx} f(x) v < 0.$$

But then Taylor directly means that

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

contradiction to the first order condition. Thus $v'^T \nabla^2_{xx} f(x) v' \geq 0$ for any x, v, which means that $\nabla^2_{xx} f(x) \succeq 0$. □

Exercise 2.

Proof.

$$\phi(\varepsilon) = \frac{f(x_0+\varepsilon) - f(x_0)}{\varepsilon}.$$

(a): Show that $\phi(\varepsilon)$ is non-decreasing:

First, we only need to prove that $\phi_{x_0}(\varepsilon)$ is non-decreasing for $\varepsilon > 0$ for all x_0, since for any fixed x_1, the non-decreasing of ε_{x_1} at $\varepsilon < 0$ can be written as:

$$\varepsilon_{x_1}(\varepsilon) = \frac{f(x_1+\varepsilon) - f(x_1)}{\varepsilon} = \frac{f(x_1+\varepsilon-\varepsilon) - f(x_1+\varepsilon)}{-\varepsilon} = \phi_{x_0+\varepsilon}(-\varepsilon)$$

thus since our argument below works for any x_0, we can only deal with one side of the function.

Assume that it is not non-decreasing, then there exists $\varepsilon_0 > 0$ and $\delta > 0, \delta \neq -\varepsilon_0$ (since ϕ does not take value at 0) such that

$$\phi(\varepsilon_0+\delta) < \phi(\varepsilon_0)$$

which implies

$$\frac{f(x_0+\varepsilon_0+\delta) - f(x_0)}{\varepsilon_0+\delta} < \frac{f(x_0+\varepsilon_0) - f(x_0)}{\varepsilon_0}$$
$$\Rightarrow \varepsilon_0 f(x_0+\varepsilon_0+\delta) < (\varepsilon_0+\delta) f(x_0+\varepsilon_0) - \delta f(x_0) \tag{2.1}$$

by cross multiplication.

Now we know the order of the points: $x_0 < x_0+\varepsilon_0 < x_0+\varepsilon_0+\delta$. This is due to our assumption that we only need to deal with $\varepsilon > 0$.

Let $\theta = \dfrac{\delta}{\varepsilon_0+\delta}$, we have by convexity

$$\frac{\delta}{\varepsilon_0+\delta} f(x_0) + \frac{\varepsilon_0}{\varepsilon_0+\delta} f(x_0) f(x_0+\varepsilon_0+\delta) \geq f\left(\frac{\delta x_0}{\varepsilon_0+\delta} + \frac{\varepsilon_0(x_0+\varepsilon_0+\delta)}{\varepsilon_0+\delta}\right) = f(x_0+\varepsilon_0)$$
$$\Rightarrow \delta f(x_0) + \varepsilon_0 f(x_0+\varepsilon_0+\delta) \geq (\varepsilon_0+\delta) f(x_0+\varepsilon_0)$$

contradiction to 2.1!

Thus, $\phi(\varepsilon)$ is non-decreasing.

(b) Conclude that the left and right derivatives of $f(x)$ exist:

For right limit, we know that $\lim_{\varepsilon\to 0^+} \dfrac{f(x+\varepsilon) - f(x)}{\varepsilon}$ is non-increasing as $\varepsilon \to 0^+$ by (a). Moreover, the limit is bounded by $[\phi_x(-1), \phi_x(1)]$. Hence by monotone convergence theorem the limit exists.

For left limit the argument is exactly the same with left and write flipped. Thus again by monotone convergence theorem the limit exists.

(c): <u>What about the following limit:</u>

$$\lim_{t\to 0^+} \frac{f(x+tv)-f(x)}{t}.$$

We follow the same line of argument as above. First define

$$g(t) = \frac{f(x+tv)-f(x)}{t}$$

for some arbitrary x, v, and t such that $f(x+tv)$ is in the domain so that we won't mension this later. Now since x, v arbitrary, if we can prove that g is increasing on $t > 0$, we can show that g is increasing for all $t \neq 0$ and x, v.

Now assume that g is not non-decreasing by fixing $t, s > 0$ such that

$$\frac{f(x+(t+s)v)-f(x)}{t+s} < \frac{f(x+tv)-f(x)}{t}$$
$$\Rightarrow tf(x+(t+s)v) + sf(x) < (t+s)f(x+tv).$$

Now again, since f is convex and $x+tv$ lies on the line $[x, x+(t+s)v]$, we can let $\theta = \dfrac{s}{t+s}$ and get

$$\frac{s}{t+s}f(x) + \frac{t}{t+s}f(x+(t+s)v) \geq f\left(\frac{s}{t+s}x + \frac{t}{t+s}(x+(t+s)v)\right) = f(x+tv)$$
$$\Rightarrow sf(x) + tf(x+(t+s)v) \geq (t+s)f(x+tv)$$

contradiction to above inequality. Thus g is increasing.

And again, by monotone convergence theorem $\lim_{t\to 0^+} g(t)$ exists.

$\square$

Exercise 3.

Proof.

(a): Define

$$g(t) := f(X+tA, y+tb) = (y+tb)^T (X+tA)^{-1} (y+tb)$$

by the dimension reduction argument in class, we know that f is convex $\iff$ g is convex.

Since we will try to differentiate the term $(X+tA)^{-1}$ during our test, let's do it first. We know by definition that the derivative at 0 is

$$\frac{d}{dt}(X+tA)^{-1} = \lim_{t\to 0} \frac{(X+tA)^{-1} - X^{-1}}{t}$$

and we write the nominator by the following way:

$$\begin{aligned}
(X+tA)^{-1} - X^{-1} &= X^{-1}(X(X+tA)^{-1} X - X)X^{-1} \\
&= X^{-1}(X(X+tA)^{-1}(X+tA) - X(X+tA)^{-1}(tA) - X)X^{-1} \\
&= X^{-1}(X - X - X(X+tA)^{-1}(tA))X^{-1} \\
&= -X^{-1}X(X+tA)^{-1}(tA)X^{-1}
\end{aligned}$$

and plugging back to the derivative quotient we get

$$\begin{aligned}
\frac{d}{dt}(X+tA)^{-1} &= \lim_{t\to 0} \frac{-X^{-1}X(X+tA)^{-1}(tA)X^{-1}}{t} \\
&= \lim_{t\to 0} -X^{-1}X(X+tA)^{-1} AX^{-1} \\
&= -X^{-1}AX^{-1}.
\end{aligned}$$

To extend the whole thing to derivative at any point s, we simply do it by noting that if we let $X_0 := X(s) := (X+sA)$, then the derivative at s will be

$$\frac{d}{dt}(X+tA)^{-1}(s) = \lim_{t\to 0} \frac{(X+sA+tA)^{-1} - (X+sA)^{-1}}{t} = X_0^{-1}AX_0^{-1}.$$

Hence taking the derivative by product rule, and denoting things by

$$X_1 = (X+tA)^{-1}; X_2 = X_1AX_1; X_3 = X_1AX_1AX_1$$

we obtain the following:

$$g'(t) = b^T X_1(y+tb) - (y+tb)^T X_2^{-1}(y+tb) + (y+tb)^T X_1 b$$

and

$$g''(t) = 2b^T X_1 b - 2b^T X_2(y+tb) - 2(y+tb)^T X_2 b + 2(y+tb)^T X_3(y+tb)$$

where if we let $X_1^{\frac{1}{2}}b = b'$, $X_1^{\frac{1}{2}}AX_1(y+tb) = y'$, (which is valid since $X_1 > 0$) we have by expanding

$$\begin{aligned} g''(t) &= 2b^T X_1 b - 2b^T X_1 A X_1 (y+tb) - 2(y+tb)^T X_1 A X_1 b \\ &\quad + 2(y+tb)^T X_1 A X_1 A X_1 (y+tb) \\ &= 2(b'^T b' - b'^T y' - y'^T b' + y'^T y') = (y'-b')^2 \geq 0 \end{aligned}$$

and we are done.

(b): It is not true since we just take $X = A = -Id$ (so that it's negative definite, thus the function is concave). Then let

$$y = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad b = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

we have

$$\frac{1}{2}f(A,b) + \frac{1}{2}f(X,y) = -1 < -\frac{1}{2} = f\left(\frac{1}{2}(A+X), \frac{1}{2}(b+y)\right)$$

so it's a counter example.

(c): It is not true since we can pick

$$X = \begin{pmatrix} 1 & 10 \\ -10 & 1 \end{pmatrix}, A = \begin{pmatrix} 1 & -10 \\ 10 & 1 \end{pmatrix}$$

then they satisfy the requirement since $X^T + X = A^T + A = 2I > 0$.

Now, let

$$y = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad b = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

then

$$f(X,y) = y^T \frac{1}{101}\begin{pmatrix} 1 & -10 \\ 10 & 1 \end{pmatrix} y = \frac{1}{101}$$

and

$$f(A,b) = b^T \frac{1}{101}\begin{pmatrix} 1 & 10 \\ -10 & 1 \end{pmatrix} b = \frac{1}{101}$$

and

$$f\left(\frac{1}{2}(A+X), \frac{1}{2}(b+y)\right) = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}^T \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix} = \frac{1}{2}.$$

Hence

$$\frac{1}{2}f(A,b) + \frac{1}{2}f(X,y) = \frac{2}{101} < \frac{1}{2} = f\left(\frac{1}{2}(A+X), \frac{1}{2}(b+y)\right)$$

thus the function is not convex.

$\square$

Exercise 4.

Proof.

(a) If the sets are convex, closed, bounded and disjoint, they can be strictly separated:

(Follow path in book) Let the sets be called C, D, then define

$$\mathrm{Dist}(C, D) = \inf\{||u - v||_2|, u \in C, v \in D\}.$$

Then we show that this distance is positive and there exists points $c \in C$, $d \in D$ such that $||c - d||_2 = \mathrm{Dist}(C, D)$. The reason is that the inf in the definition of set distance is equivalent to the limit of some sequences $\{x_n\} \subset C$ and $\{y_n\} \subset D$ for which

$$\lim_{n\to\infty} d(x_n - y_n) = \mathrm{Dist}(C, D).$$

But now C, D are closed and bounded in $\mathbb{R}^n$, so they are compact and thus there exists subsequences of x_n and y_n that converges to a point in C and D. Denote the subsequence by the same name x_n and y_n for convenience and thus we have $x_n \to x$ and $y_n \to y$ for $x \in C$ and $y \in D$. Now since the limit of the distance function, a continuous function, exists, we can pass the limit and know that

$$\mathrm{Dist}(C, D) = \lim_{n\to\infty} d(x_n, y_n) = d(x, y).$$

From which we've constructed explicitly the two points. Since the two sets are disjoint, $d(x, y) > 0$, $x = c$ and $y = d$.

Now we know by the proof in book that let $a = d - c$, $b = \dfrac{||d||_2^2 - ||c||_2^2}{2}$, and define the affine function

$$f(x) = a^T x - b = (d - c)^T (x - (1/2)(d + c)).$$

Then we show that the function f is negative on C and positive on D, i.e. the hyperplane $\{a^x = b\}$ strictly separates C and D.

I only show that f is positive on D. As for why it's negative on C, we just multiply by -1 and we do the exactly same with c, d switch places.

Suppose there exists a point $u \in D$ for which

$$f(u) = (d - c)^T (u - (1/2)(d + c)) \leq 0.$$

Then we can re-express $f(u)$ as

$$f(u) = (d - c)^T (u - d + (1/2)(d - c)) = (d - c)^T (u - d) + (1/2)||d - c||^2 \leq 0$$

where since we've proven that $d(c, d) > 0$, so the part $(d - c)^T (u - d) < 0$ strictly. Now, following the rest in book, we have

$$\frac{d}{dt}||d + t(u - d) - c||^2 \Big| t = 0 = 2(d - c)^T (u - d) < 0$$

and hence for some small $0 < t \leq 1$ we have

$$||d + t(u - d) - c|| < ||d - c||$$

which means that the point $d+t(u-d)$ is closer to c than d, contradiction! Thus, the separation is strict.

(b) Counter example:

Let $A, B \subset \mathbb{R}^2$ such that

$$A := \left\{(x, y)|y \geq \frac{1}{x}, x > 0\right\}$$

and

$$B := \{(x, y)|y \leq 0, x \geq 0\}$$

i.e. A is the epigraph of $y = \frac{1}{x}$ and B is the closed 4th quadrant. Thus A is the epigraph of a convex function, hence convex (page 75 bottom part in textbook), and B is an orthant, thus convex. Both are closed since it's defined so, and they are disjoint since $\frac{1}{x} > 0$ for all $x \geq 0$.

Now, to show that they cannot be strictly separated, assume that they can be strictly separated by the hyperplane $y = ax + b$. This is because they cannot be separated by a vertical line since for all t, $x \geq t$ intersects both A and B.

Now by our assumption, since all the lower half y-axis is in B, $b > 0$ and we know for all $(x, y) \in B$, $y < ax + b$ since $0 < 0 + b$ and $(0, 0) \in B$. But if $a \geq 0$ then $y = ax + b$ intersects A, while if $a < 0$, $y = ax + b$ intersects B. Hence $y = ax + b$ cannot strictly separates them (since the equal sign cannot be reached for any points in both sets), thus the counter example holds.

□