Math Problem Set 4

Matthew Brown
OSM Boot Camp 2018

July 16, 2018

Problem 6.6.

$$Df(x,y) = \begin{bmatrix} 6xy + 4y^2 + y & 3x^2 + 8xy + x \end{bmatrix}$$
$$D^2f(x,y) = \begin{bmatrix} 6y & 6x + 8y + 1 \\ 6x + 8y + 1 & 8x \end{bmatrix}$$

The critical points are $(0,0)$ and $(-\frac{1}{3}, -\frac{2}{3})$. $D^2f(0,0)$ is indefinite, so the point is not a local extremum. $D^2f(-\frac{1}{3}, -\frac{2}{3})$ is actually "negative definite" (I think?) so the point is a local maximum.

Problem 6.7. Two parts.

- (i)

  Proof. To see that $Q = A^T + A$ is symetric, examine the coordinates.

  $$(A^T + A)_{ij} + A^T_{ij} + A_{ij} = A_{ji} + A^T_{ji} = (A^T + A)_{ji}$$

  Then, see since $x^T Ax \in \mathbb{F}$ that

  $$(x^T Ax) = (x^T Ax)^T = xA^T x$$

  and so

  $$x^T Qx = x^T(A^T + A)x = x^T A^T x + x^T Ax = 2x^T Ax$$

  □

  which shows that 6.17 is equivalent to 6.18.

- (ii)

  Proof. Take the derivative of f and see that the FOC is:

  $$f'(x) = 0 \iff \frac{1}{2}(Q + Q^T)x - b = 0 \iff Qx = b$$

  □

- (iii)

  Proof. Note that $f''(x) = Q$, so the condition $f''(x) > 0$ will always be satisfied.
  $\Leftarrow$
  Suppose Q is positive definite. Then Q is invertible and if $x^* = Q^{-1}b$, then x^* satisfies the second order sufficient condition for being a minimizer of f.
  $\Rightarrow$ The only way that a minimizer will exist is if the second derivative is positive definite, and this only happens if Q is positive definite. $\square$

Problem 6.11. The minimizer of f is $-\frac{b}{2a}$. If we apply Newton's method to any $x_0 \in \mathbb{R}$,

$$x_1 = x_0 - \frac{f'(x_0)}{f''(x_0)} = x_0 - \frac{2ax_0 + b}{2a} = \frac{2ax_0 - (2a_x0 + b)}{2a} = -\frac{b}{2a}$$

Problem 6.15. See jupyter notebook

Problem 7.1. *Proof.* Consider two elements $a, b \in C = \text{Conv}(S)$ Then by the definition of C, we can represent a, b as linear combinations

$$\begin{aligned} a &= \lambda_1 x_1 + ... + \lambda_n x_n \\ b &= \Lambda_1 y_1 + ... + \Lambda_n y_m \end{aligned}$$

where $\{x_i\}_{i=1}^n, \{y_i\}_{j=1}^m \in S$ and $\sum_{i=1}^n \lambda_i = \sum_{j=1}^m \Lambda_j = 1$ with all $\lambda_i, \lambda_j \in [0, 1]$.
Now consider a combination $ka + (1-k)b$, with $k \in [0, 1]$. We want to show that this is in C. We have

$$\begin{aligned} ka + (1-k)b &= k(\lambda_1 x_1 + ... + \lambda_n x_n) + (1-k)(\Lambda_1 y_1 + ... + \Lambda_n y_m) \\ &= k\lambda_1 x_1 + ... + k\lambda_n x_n + ... + (1-k)\Lambda_1 y_1 + ... + (1-k)\Lambda_m y_m \end{aligned}$$

Define

$$\{r_i\}_{i=1}^{m+n} = \{k\lambda_1, ...k\lambda_n, (1-k)\Lambda_1, ..., (1-k)\Lambda_m\}$$

Then

$$\sum_{i=1}^{m+n} r_i = k \sum_{i=1}^{n} x_i + (1-k) \sum_{j=1}^{m} y_j = k \cdot 1 + (1-k) \cdot 1 = 1$$

Also, $r_i \in [0, 1]$ for each i. And we showed above that we can write $ka + (1-k)b$ as a linear combination of elements in S with coefficients $r_1, ...r_{m+n}$. so $ka + (1-k)b \in C$, and this shows that C is convex $\square$

Problem 7.2. *Proof.* Two Parts:

- (i) Suppose H is a hyperplane. Let $x, y \in H, \lambda \in [0, 1]$. We know $\langle a, x \rangle = b$ and $\langle a, y \rangle = b$, so

  $$\langle a, \lambda x + (1-\lambda)y \rangle = \langle a, \lambda x \rangle + \langle a, (1-\lambda)y \rangle = \lambda b + (1-\lambda)b = b$$

  And so $\lambda x + (1-\lambda)y \in H$.

- (ii) Very similar to (i)...
  Suppose H is the half-plane $\{x \in \mathbb{R}^n | \langle a, x\rangle \leq b\}$. Let $x, y \in H, \lambda \in [0,1]$. We know $\langle a, x\rangle = c$ and $\langle a, y\rangle = d$ for some $c, d \leq b$, so

  $$\langle a, \lambda x + (1-\lambda)y\rangle = \langle a, \lambda x\rangle + \langle a, (1-\lambda)y\rangle = \lambda c + (1-\lambda)c \leq \lambda b + (1-\lambda)b = b$$

  And so $\lambda x + (1-\lambda)y \in H$.

$\square$

Problem 7.4. *Proof.* I'll first show the four facts.

- (i)

  $$\begin{aligned} ||x-y||^2 &= \langle x, x\rangle + \langle y, y\rangle - 2\langle x, y\rangle \\ &= \langle x, x\rangle - 2\langle x, p\rangle + \langle p, p\rangle + \langle y, y\rangle - 2\langle p, y\rangle + \langle p, p\rangle + 2(-\langle x, y\rangle + \langle x, p\rangle - \langle p, p\rangle + \langle p, y\rangle) \\ &= ||x-p||^2 + ||p-y||^2 + 2\langle x-p, p-y\rangle \end{aligned}$$

- (ii) We can use the identity from (i). See that $||p-y||^2$ is always strictly positive for $y \neq p$, and if 7.14 holds then the term $\langle x-p, p-y\rangle$ is also nonnegative and the result follows.

- (iii) This will just be more manipulation of $\langle\rangle$ (RETURN TO THIS PROBLEM)

- (iv)

$\square$

Problem 7.8. *Proof.* Let $\lambda \in [0,1]$, $x, y \in \mathbb{R}^m$. Then

$$\begin{aligned} g(\lambda x + (1-\lambda)y) &= f(A(\lambda x + (1-\lambda)y) + b) \\ &= f(\lambda Ax + (1-\lambda)Ay + b) \\ &= f(\lambda(Ax+b) + (1-\lambda)(Ay+b)) \\ &\leq \lambda f(Ax+b) + (1-\lambda)f(Ay+b) \\ &= g(x) + g(y) \end{aligned}$$

and g is convex. $\square$

Problem 7.12. Two parts

- (i) Let A, $B \in PD_n(\mathbb{R})$. Then for any $x \in \mathbb{R}^n, \lambda \in [0,1]$

  $$\langle x, (\lambda A + (1-\lambda)B)x\rangle = \langle x, \lambda Ax\rangle + \langle x, (1-\lambda)By\rangle \geq 0$$

  and we see that $\lambda A + (1-\lambda)B \in PD_n(\mathbb{R})$.

- (ii) There are a few intermediate results. (many thanks to Albi here!) First I'll show the following:

Lemma 0.1. A function $f : PD_n(\mathbb{R}) \to \mathbb{R}$ is convex if for every $A, B \in PD_n(\mathbb{R})$, the function $g : [0,1] \to \mathbb{R}$ given by $g(t) = f(tA + (1-t)B)$ is convex.

Proof. Suppose that for every $A, B \in PD_n(\mathbb{R})$, the function $g : [0,1] \to \mathbb{R}$ given by $g(t) = f(tA + (1-t)B)$ is convex. Name the function g associated to matrices X, Y as g_{XY} Then for A, $B \in PD_n(\mathbb{R}), t_1, t_2 \in \mathbb{R}, \lambda \in [0,1]$, we have left-hand side

$$\lambda g_{AB}(t_1) + (1-\lambda)g_{AB}(t_2) = \lambda(f(t_1A + (1-t_1)B) + (1-\lambda)(f(t_2A + (1-t_2)B)$$

and right-hand side

$$\begin{aligned} g_{AB}(\lambda t_1 + (1-\lambda)t_2) &= f((\lambda t_1 + (1-\lambda)t_2)A + (1 - (\lambda t_1 - (1-\lambda)t_2))B) \\ &= f(\lambda(t_1A + (1-t_1)B) + (1-\lambda)(t_2A + (1-t_2)B)) \end{aligned}$$

By the convexity of g, we know that the right hand side is greater than or equal to the left hand side. And since the choice of t_1, t_2 was arbitrary, we can set $t_1 = 1, t_2 = 0$ and get the result

$$\lambda f(A) + (1-\lambda)f(B) \leq f(\lambda A + (1-\lambda)B)$$

for all matrices A, B, which proves the lemma. □

Next, I'll show a fact about G.
First of all, I know that A positive definite $\implies$ there exists a matrix S such that $S^HS = A$ by a previous exercise. Now, $tA + (1-t)B = S^H(tI + (1-t)(S^H)^{-1}BS^{-1})S$, and we see that:

$$g(t) = -\log(\det(tA + (1-t)B)) = -\log(\det(S^H(tI + (1-t)(S^H)^{-1}BS^{-1})S)).$$

We then use the properties of logs and determinants to see that:

$$\begin{aligned} -\log(\det(S^H(tI + (1-t)(S^H)^{-1}BS^{-1})S)) &= -\log(\det(S^H)) \\ &\quad - \log(\det(tI + (1-t)(S^H)^{-1}BS^{-1})) - \log(\det(S)) \\ &= -\log(\det(S^H)\det(S)) \\ &\quad - \log(\det(tI + (1-t)(S^H)^{-1}BS^{-1})) \\ &= -\log(\det(A)) - \log(\det(tI + (1-t)(S^H)^{-1}BS^{-1})). \end{aligned}$$

Remember that all this is equal to g(t) and call this fact (b).
Now for a positive definite matrix B, I consider the $\lambda_1, ..., \lambda_n$ eigenvalues of the matrix $(S^H)^{-1}BS^{-1}$ and corresponding eigenvectors $x_1, ..., x_n$, where S is as defined above for some positive definite matrix A (as in right hand side of fact (b)).
For any x_i,

$$(tI + (1-t)(S^H)^{-1}BS^{-1})x_i = tx_i + (1-t)(\lambda_i)x_i = (t + (1-t)\lambda_i)x_i$$

Which gives us an expression for the eigenvalues of $tI + (1-t)(S^H)^{-1}BS^{-1}$. We see now that:

$$\det tI + (1-t)(S^H)^{-1}BS^{-1} = \Pi_{i=1}^n(t + (1-t)\lambda_i))$$

so by fact (b) and applying the properties of log:

$$g(t) == -\log(\det(A)) - \sum_{i=1}^{n} \log((t + (1-t)\lambda_i))$$

Using this expression, we see that $g''(t) = \sum_{i=1}^{n}(1-\lambda_i)^2/(t+(1-t)\lambda_i)^2$, which is nonnegative for all $t \in [0,1]$. By this fact, we can say that $g(t)$ is convex for all t, and it follows from lemma 0.1 that f is convex.

Problem 7.13. I'll argue by contradiction. Suppose that f is convex and bounded above, but f is not a constant function. Then there exist points $x_1, x_2 \in \mathbb{R}^n$ where $f(x_1) \neq f(x_2)$. Let M be the upper bound for f.
Suppose WLOG that $f(x_2) \geq f(x_1)$ Now consider the G graph of f,

$$G \in \mathbb{R}^{n+1}, G = \{(x, f(x)) | x \in \mathbb{R}^n\}$$

I wasn't able to formally show the rest of the proof, but drawing a 2-dimensional picture helps, and I'm pretty sure it generalizes.
The line passing through $f(x_1)$ and $f(x_2)$ does not lie on any hyperplane which a translate of the hyperplane $f(x) = M$ Therefore, the line intersects M at some point. However, $f(x_2)$ must lie on or above this line in the n+1st dimension. Therefore, there also is a point $f(y)$ s.t. $f(y) >$ the value of the line in the n+1st dimension $> M$ and we have a contradiction.

Problem 7.20. The first thing to note is that that

$$\begin{aligned} -f \text{ is convex } &\iff -f(\lambda x_1 + (1-\lambda)x_2) \leq -(\lambda f(x_1) + (1-\lambda) f(x_2)) \\ &\iff f(\lambda x_1 + (1-\lambda)x_2) \geq (\lambda f(x_1) + (1-\lambda) f(x_2)) \end{aligned}$$

for $\lambda \in [0,1]$. And if we combine with the fact that f is convex we see that

$$f(\lambda x_1 + (1-\lambda)x_2) = \lambda f(x_1) + (1-\lambda) f(x_2)$$

for $\lambda \in [0,1]$, which is quite a handy fact. Indeed, this looks a LOT like the conditions we need for linearity - I just need a way to pass from λ to other scalars... I do not think that this is sufficient, but I'm fairly certain that this is the first part of the proof.

Problem 7.21. *Proof.* I'll show both implications.

- $\Rightarrow$ Suppose x^* minimizes f. Then there exists an open neighborhood U of x^* such that for $x \in U, f(x*) < f(x)$. Because ϕ is increasing, this implies that for $x \in U$, $\phi(f(x^*)) < \phi(f(x))$ and x^* is a local minimizer for $\phi \circ f$.

- $\Leftarrow$ I'll show the contrapositive (even though that's probably needlessly complicating it, sorry). Suppose x^* does not minimize f. Then for any open neighborhood U of x^*, there exists $x_0 \in U$ such that $f(x*) \geq f(x_0)$. Because ϕ is increasing, this implies that for $\phi(f(x^*)) \geq \phi(f(x_0))$ and x^* is not a local minimizer for $\phi \circ f$.

□