Not a great week for me. Copied some problems from Albi.

Problem 6.6. Find the critical points of $f(x, y) = 3x^2y + 4xy^2 + xy$

Proof. We want $f' = 0$

$$\Delta f = (6xy + 4y^2 + y, 8xy + 3x^2 + x)$$

This becomes a system of linear equations

$$\begin{cases} +0 = 6xy + 4y^2 + y \\ +0 = 8xy + 3x^2 + x \end{cases}$$

Case 1: x=0

$$\begin{aligned} &\implies 4y^2 + y = 0 \\ &\implies y(4y+1) = 0 \\ &\implies y = 0, \frac{-1}{4} \end{aligned}$$

Case 2: y=0

$$\begin{aligned} &\implies 3x^2 + x = 0 \\ &\implies x(3x+1) = 0 \\ &\implies x = 0, \frac{-1}{3} \end{aligned}$$

Case 3: $y \neq 0, \quad x \neq 0$

$$\begin{cases} +0 = y(6x + 4y + 1) \\ +0 = x(3x + 8y + 1) \end{cases} \implies \begin{cases} +0 = 6x + 4y + 1 \\ +0 = 3x + 8y + 1 \end{cases}$$

$$\begin{aligned} &\implies -12y - 1 = 0 \\ &\implies y = \frac{-1}{12} \\ &\implies x = \frac{-1}{9} \end{aligned}$$

x	y	Name
0	0	A
0	-1/4	B
-1/3	0	C
-1/9	-1/12	D

Eigs of the hessian for A are mixed, so A is a saddle point. Eigs of the hessian for B are mixed, so B is a saddle point. Eigs of the hessian for C are mixed, so C is a saddle point. Eigs of the hessian for D are negative, so D is a local maximizer.

□

Problem 6.7. [1]

Proof. (i) Notice that $Q^T = (A^T+A)^T = A^T+A = A+A^T = Q$. Also, $x^T Ax = \sum_{i=1}^{n} a_{ij}x_ix_j = \sum_{i=1}^{n} a_{ji}x_ix_j = x^T A^T x$. Therefore $x^T Qx = 2x^T Ax$ and (6.17) is equivalent to

$$f(x) = x^T Qx/2 - b^T x + c.$$

(ii)
The first order necessary conditions for a minimizer imply $Q^T x^* = b$, since $f'(x) = Q^T x - b$.
(iii)
If Q is positive definite, then $f''(x) > 0$ for any x. Also, Q is invertible and by (6.19) we have that $x^* = Q^{-1}b$ is such that $f'(x^*) = 0$. Then by the second order sufficient condition, x^* is the unique minimizer of f. Now assume x^* is the unique minimizer of f. Then by the second order necessary condition, Q is positive semi-definite. Also, x^* is a solution to $Q^T x^* = b$. If Q has at least one zero eigenvalue, then x^* is not unique. Therefore Q must be positive definite. □

Problem 6.11. $f(x) = ax^2 + bx + c$. Show that one iteration of newton's method will give you a unique solution.

Proof.

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

□

Problem 7.1. WTS: if $S \subset V, s \neq \emptyset \quad \text{then} \quad conv(S)$ is convex.

Proof. **WTS:**

$$\lambda x + (1-\lambda)y \in conv(S)$$
$$\lambda a_1x_1 + \ldots + \lambda a_kx_k + (1-\lambda)b_1y_1 + \ldots + (1-\lambda)b_ky_k$$

As $0 \leq \lambda \leq 1$,

$$\lambda \sum a_i + (1-\lambda) \sum b_i = \lambda + (1-\lambda) = 1$$

□

[1] Thank you to Jayhyung and Albi for much of these notes

Problem 7.2. (i)

Proof. Let $P = \{x \in V| < a, x >= b\}$, a hyperplane in V. Then, pick arbitrary $x, y \in P$, satisfying $< a, x >= b$ and $< a, y >= b$. Then, for arbitrary scalar $\lambda \in [0, 1]$, the following is satisfied;

$$< a, \lambda x + (1-\lambda)y >= \lambda < a, x > +(1-\lambda) < a, y >= b$$

Thus, $\lambda x + (1-\lambda)y \in P$. Q.E.D

(ii)

□

Proof. The argument is the same as above. □

Problem 7.4. (i)

Proof.

$$\begin{aligned}\|x-y\|^2 =& \|x-p+p-y\|^2 \\ =& < x-p+p-y, x-p+p-y > \\ =& \|x-p\|^2 + \|p-y\|^2 + 2 < x-p, p-y >\end{aligned}$$

(ii)

□

Proof. By the assumption that $p \neq y$, $\|p-y\|^2 > 0$. If we have the assumption that $< x-p, p-y >\geq 0$, using (i), the staement trivially holds. Q.E.D

(iii) Using (i),

$$\begin{aligned}\|x-z\|^2 =& \|x-p\|^2 + \|\lambda y - \lambda p\|^2 + < x-p, \lambda p - \lambda y > \\ =& \|x-p\|^2 + 2\lambda < x-p, p-y > + \lambda^2 \|y-p\|^2\end{aligned}$$

(iv)
Using (7.15), and setting $\lambda = 1$, thus $z = y$. Then, using (7.15),

$$0 \leq \|x-y\|^2 - \|x-p\|^2 = 2\lambda < x-p, p-y > + \lambda^2 \|y-p\|^2$$

If you divide by λ , then $0 \leq 2 < x-p, p-y > + \lambda \|y-p\|^2$
This holds for every $y \in C$, so $< x-p, p-y >\geq 0$ □

Problem 7.8.

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

Problem 7.12. (i)

Take $X, Y \in PD_n(\mathbb{R})$ and $\lambda \in [0,1]$. Then for every $v \in \mathbb{R}^n$ we have that

$$v^T(\lambda X + (1-\lambda)Y)v = \lambda(v^T X v) + (1-\lambda)(v^T Y v) > 0,$$

because X and Y are positive definite.

(ii)

(a) Take $t_1, t_2 \in \mathbb{R}$ and $\lambda \in [0,1]$. On the one hand,

$$\lambda g(t_1) + (1-\lambda)g(t_2) = \lambda f(t_1 A + (1-t_1)B) + (1-\lambda)f(t_2 A + (1-t_2)B).$$

On the other,

$$\begin{aligned} g(\lambda t_1 + (1-\lambda)t_2) &= f((\lambda t_1 + (1-\lambda)t_2)A + (1 - \lambda t_1 + (1-\lambda)t_2)B) \\ &= f(\lambda(t_1 A + (1-t_1)B) + (1-\lambda)(t_2 A + (1-t_2)B)). \end{aligned}$$

Since g is convex we get

$$f(\lambda X + (1-\lambda)Y) \leq \lambda f(X) + (1-\lambda)f(Y),$$

with $X = t_1 A + (1-t_1)B$ and $Y = t_2 A + (1-t_2)B$. Since the choice of t was arbitrary and this holds for any $A, B \in PD_n(\mathbb{R})$, we conclude that f is convex.

(b) By Proposition (4.5.7), we know that if A is posititve definite, then there exits a nonsingular matrix S such that $A = S^H S$. Then, $tA + (1-t)B = S^H(tI + (1-t)(S^H)^{-1}BS^{-1})S$, and so

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

By the fact that $\det(AB) = \det(A)\det(B)$ and the properties of logarithms, we obtain

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

(c)

Since $A, B \in PD_n(\mathbb{R})$, then $B^{-1} \in PD_n(\mathbb{R})$ and $((S^H)^{-1}BS^{-1})^{-1} = SB^{-1}S^H$ is positive definite since

$$x^H S B^{-1} S^H x = (S^H x)^H B^{-1}(xS) > 0.$$

Therefore $(S^H)^{-1}BS^{-1}$ is positive definite. Now let $\{\lambda_i\}_i$ be the collection of eigenvalues of $((S^H)^{-1}BS^{-1})$ and $\{x_i\}_i$ the corresponding collection of eigenvectors. Then for every i:

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

Thus, $\{t + (1-t)\lambda_i\}_i$ are the eigenvalues of $(tI + (1-t)(S^H)^{-1}BS^{-1})$ corresponding to the $\{x_i\}_i$, and we can conclude that

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

(d)

By using the expression of $g(t)$ in part (c) we can see that $g'(t) \sum_{i=1}^{n} (1-\lambda_i)/(t+(1-t)\lambda_i)$ and $g''(t) = \sum_{i=1}^{n} (1-\lambda_i)^2/(t+(1-t)\lambda_i)^2$, which is clearly nonnegative for all $t \in [0,1]$.

Problem 7.13. Suppose $f(x) < M$ for all x for some real M and f is convex and not constant. Then, there exist $x, y \in \mathbb{R}^n$ such that $f(x) \neq f(y)$. But then the line between $(x, f(x))$ and $(y, f(y))$ intersects $f(\cdot) = M$. Since f must lie on or above this line, at some point it must cross $f(\cdot) = M$ as well, which is a contraddiction.

Problem 7.20. Take $x, y \in \mathbb{R}^n$, with $x \neq y$, and $\lambda \in [0,1]$. Since f is convex we have $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$. Since $-f$ is convex, the opposite hold. Therefore we must have $f(\lambda x + (1-\lambda)y) = \lambda f(x) + (1-\lambda) f(y)$. Therefore f is affine.

Problem 7.21. Let $x^* \in \mathbb{R}^n$ be a local minimizer of f. Then $f(x^*) \leq f(x)$ for all $x \in \mathcal{N}_r(x^*)$, where $\mathcal{N}_r(x^*)$ is an open ball around x^* of radius $r > 0$. Since ϕ is monothonically increasing, $\phi(f(x^*)) \leq \phi(f(x))$ for all $x \in \mathcal{N}_r(x^*)$. Thus, x^* is a local minimizer of $\phi \circ f$. Now let x^* be a local minimizer of $\phi \circ f$. Then $\phi(f(x^*)) \leq \phi(f(x))$ for all $x \in \mathcal{N}_r(x^*)$, and since ϕ is monothonically increasing, this implies that $f(x^*) \leq f(x)$ for all $x \in \mathcal{N}_r(x^*)$. Thus, x^* is a local minimizer of f.