Math Probset 4
Optimization and Convex Analysis
Jan Ertl
Kendra Robbins

6.6

critical points: $(0,0), (\frac{-1}{3}, 0), (0, \frac{-1}{4}), (\frac{-1}{9}, \frac{-1}{12})$

Hessian: $D^2 f(x,y) = \begin{bmatrix} 6y & 6x+8y+1 \\ 6x+8y+1 & 8x \end{bmatrix}$

saddle points: $(0,0), (\frac{-1}{3}, 0), (0, \frac{-1}{4})$ because the Hessian evaluated at each point has eigenvalues of opposite sign.

maximum: $(\frac{-1}{9}, \frac{-1}{12})$ because the hessian evaluated there has 2 negative eigenvalues.

6.7

i) Let $f : \mathbb{R}^n \to \mathbb{R}$ and let $< \cdot, \cdot >$ be the usual dot product.

$Q^T = (A^T + A)^T = A + A^T = Q$

$\Rightarrow Q$ is symmetric

Since the domain is real, we have $< Ax, x >=< x, Ax >$, so:

$x^T Q x = x^T (A^T + A) x = x^T A^T x + x^T A x = (Ax)^T x + x^T A x =< Ax, x > + < x, Ax >$
$= 2 < x, Ax >= 2x^T A x$

$\Rightarrow f(x) = \frac{1}{2} x^T Q x - b^T x + c$

ii) Suppose $x*$ is a minimizer of f.

$f'(x) = x^T Q - b^T$

Since the domain of f is $\mathbb{R}^n$, $x*$ must be an interior point. So we must have $f'(x*) = 0$.

$\Rightarrow (x*)^T Q = b^T$

By taking the transpose of both sides: $Q^T x* = b$

iii) Suppose $x*$ is the unique solution of the minimization problem of $f(x) = \frac{1}{2} x^T Q x - b^T x + c$.

Then $f''(x*) = Q$ is positive semi definite.

But if Q has a zero eigenvalue, $x*$ would not be unique. So Q must be positive definite.

Suppose Q is positive definite.

$f''(x) = Q$ so $D^2 f(x) > 0 \ \forall x$

$\Rightarrow f$ is convex by 7.2.10

$\Rightarrow$the minimizer of f is unique according to Albi

From ii) we have $Q^T x* = b$ is a solution of f.

Let λ_1, λ_2 be the eigenvalues of Q. Since Q is positive definite, $\lambda_1, \lambda_2 > 0$.

So $det(Q) = \lambda_1 \lambda_2 > 0$

$\Rightarrow Q = Q^T$ is invertible.

So $x* = Q^{-1} b$ is a solution to $f(x) = \frac{1}{2} x^T Q x - b^T x + c = 0$

By second order sufficient conditions, $x*$ is a unique minimizer of f.

6.11
Define $f(x) = a^2x_0^2 + bx_0 + c$.
With $f'(x) = 2ax + b$ and $f''(x) = 2a$ we find that the degree 2 taylor approximation of f at x_0 is $q(x) = a^2x_0^2 + bx_0 + c = f(x)$.
Then $x_1 = x_0 - (\frac{2ax_0+b}{2a}) = \frac{-b}{2a}$.
So $f'(x_1) = -b + b = 0$
$\Rightarrow x_1$ is a minimizer of f. And since f is quadratic x_1 must be the unique minimizer.

6.15 See jupyter notebook

7.1
Suppose $S \subseteq V$ is nonempty and $x, y \in conv(S)$.
Then $x = \sum_{i=1}^{I} c_i x_i$ and $y = \sum_{j=1}^{J} k_j y_j$ where $\{x_i\}_{i=1}^{I}, \{y_i\}_{j=1}^{J} \subseteq S$ and $\sum_{i=1}^{I} c_i = 1$ and $\sum_{j=1}^{J} k_j = 1$.
Let $\lambda \in [0, 1]$ and consider $\lambda x + (1 - \lambda)y$:
$\lambda x + (1 - \lambda)y = \lambda \sum_{i=1}^{I} c_i x_i + (1 - \lambda) \sum_{j=1}^{J} k_j y_j$
$= \lambda \sum_{i=1}^{I} c_i x_i + \sum_{j=1}^{J} k_j y_j - \lambda \sum_{j=1}^{J} k_j y_j$
$= \sum_{i=1}^{I} \lambda c_i x_i + \sum_{j=1}^{J} k_j y_j - \sum_{j=1}^{J} \lambda k_j y_j$
Since $\sum_{i=1}^{I} c_i = 1$, $\sum_{i=1}^{I} \lambda c_i x_i = \lambda \sum_{i=1}^{I} c_i = \lambda$
Since $\sum_{j=1}^{J} k_j = 1$, $\sum_{j=1}^{J} \lambda k_j = \lambda \sum_{j=1}^{J} k_j = \lambda$.
So the sum of the coefficients on $\sum_{i=1}^{I} \lambda c_i x_i + \sum_{j=1}^{J} k_j y_j - \sum_{j=1}^{J} \lambda k_j y_j$ are $\lambda + 1 - \lambda = 1$, and all x_i, y_i are elements of S, so $\lambda x + (1 - \lambda)y \in conv(S)$
$\Rightarrow conv(S)$ is convex.

7.2
i) Let P be a hyperplane such that $P = \{x \in V : < a, x >= b\}$ for some vector a and some real number b.
Let $x, y \in P$ and let $\lambda \in [0, 1]$. Then we have $< a, x >= b$ and $< a, y >= b$.
Then $< a, \lambda x + (1 - \lambda)y >=< a, \lambda x > + < a,(1 - \lambda)y >$
$= \lambda < a, x > +(1 - \lambda) < a, y >= \lambda b + (1 - \lambda)b = b$
$\Rightarrow \lambda x + (1 - \lambda)y \in P$
$\Rightarrow P$ is convex.

ii) Let H be a half space such that $H = \{x \in V : < a, x > \leq b\}$ for some vector a and some real number b.
Let $x, y \in H$ and let $\lambda \in [0, 1]$. Then we have $< a, x > \leq b$ and $< a, y > \leq b$.

Then $< a, \lambda x + (1-\lambda)y > = < a, \lambda x > + < a, (1-\lambda)y >$
$= \lambda < a, x > + (1-\lambda) < a, y > \leq \lambda b + (1-\lambda)b = b$
$\Rightarrow \lambda x + (1-\lambda)y \in H$
$\Rightarrow H$ is convex.

7.4

i) $||x-y||^2 = ||(x-p)+(p-y)||^2 = < (x-p)+(p-y), (x-p)+(p-y) >$
$= < x-p, x-p > + < x-p, p-y > + < p-y, x-p > + < p-y, p-y >$
$= ||x-p||^2 + ||p-y||^2 + 2 < x-p, p-y >$

ii) Suppose $y \neq p$ and $< x-p, p-y > \geq 0 \ \forall y \in C$
By i) we have $||x-y||^2 - ||x-p||^2 = ||p-y||^2 + 2 < x-p, p-y >$
Since $y \neq p$, $||p-y||^2 > 0$, so $||x-y||^2 - ||x-p||^2 > 0$
$\Rightarrow ||x-y||^2 > ||x-p||^2$
$\Rightarrow ||x-y|| > ||x-p||$

iii) Suppose $\lambda \in [0,1]$ and $z = \lambda y + (1-\lambda)p$.
By i), $||x-z||^2 = ||x-p||^2 + ||p-z||^2 + 2 < x-p, p-z >$
Further, $||p-z||^2 = < p-z, p-z > = < p - \lambda y - (1-\lambda), p - \lambda y - (1-\lambda) >$
$= < \lambda(p-y), \lambda(p-y) = \lambda^2 < p-y, p-y > = \lambda^2 ||p-y||^2 = \lambda^2 ||y-p||^2$
Additionally, $2 < x-p, p-z > = 2 < 2 < x-p, p - \lambda y - (1-\lambda)p >$
$= 2 < x-p, \lambda(p-y) > = 2\lambda < x-p, p-y >$
So $||x-z||^2 = ||x-p||^2 + ||p-z||^2 + 2 < x-p, p-z >$
$= ||x-p||^2 + 2\lambda < x-p, p-y > + \lambda^2 ||y-p||^2$

iv) Let's use iii) with $\lambda = 1$. So $z = y$ and:
$||x-y||^2 = ||x-p||^2 + 2 < x-p, p-y > + ||y-p||^2$
From ii) we have $||x-y|| > ||x-p||$ for $y \neq p$. Without the condition $y \neq p$ we have $||x-y|| \geq ||x-p||$
So $||x-y||^2 - ||x-p||^2 \geq 0$
So $||x-y||^2 - ||x-p||^2 = 2 < x-p, p-y > + ||y-p||^2 \geq 0$

7.8

Suppose $f : \mathbb{R}^m \to \mathbb{R}$ is convex, $A \in M_{mxn}, b \in \mathbb{R}^m$.
Define $g(x) = f(Ax+b)$. Then:
$g(\lambda x + (1-\lambda)y) = f(\lambda Ax + (1-\lambda)Ay + b)$
$= f(\lambda(Ax+b) + (1-\lambda)(Ay+b)) \leq \lambda f(Ax+b) + (1-\lambda)f(Ay+b)$
$= \lambda g(x) + (1-\lambda)g(y)$
$\Rightarrow g$ is convex.

7.12

i) Suppose $A, B \in PD_n(\mathbb{R})$, $\lambda \in [0,1]$, and $x \in \mathbb{R}^n$.

Then $A = A^T$ and $B = B^T$ so we have:
$(\lambda A + (1 - \lambda B))^T = \lambda A^T + (a - \lambda)B^T = \lambda A + (1 - \lambda B)$
Further, using $< x, Ax > > 0, < x, Bx > > 0$ we have:
$< x, (\lambda A + (1 - \lambda B))x >= x^T(\lambda A + (1 - \lambda)B)x = \lambda(x^T Ax) + (1 - \lambda)(x^T Bx) > 0$
$\Rightarrow \lambda A + (1 - \lambda B) \in PD_n(\mathbb{R})$
$\Rightarrow PD_n(\mathbb{R})$ is convex.

iia) Suppose that $\forall A, B \in PD_n(\mathbb{R})$, $g(t) : [0, 1] \to \mathbb{R}$ defined by $g(t) = f(tA+(1-t)B)$ is convex.
Let $t_1, t_2 \in [0, 1]$ and $\lambda \in [0, 1]$. Then we have:
$\lambda g(t_1) + (1 - \lambda)g(t_2) = \lambda f(t_1A + (1 - t_1)B) + (1 - \lambda)f(t_2A + (1 - t_2)B)$.
Also:
$g(\lambda t_1 + (1 - \lambda)t_2) = f((\lambda t_1 + (1 - \lambda)t_2)A + (1 - \lambda t_1 + (1 - \lambda)t_2)B)$
$= f(\lambda(t_1A + (1 - t_1)B) + (1 - \lambda)(t_2A + (1 - t_2)B))$.
Since g is convex we have: $g(\lambda t_1 + (1 - \lambda)t_2) \le \lambda g(t_1) + (1 - \lambda)g(t_2)$
Using the substitutions $X = t_1A + (1 - t_1)B$ and $Y = t_2A + (1 - t_2)B$ we get:
$f(\lambda X + (1 - \lambda)Y) \le \lambda f(X) + (1 - \lambda)f(Y)$. $\Rightarrow f$ is convex.

iib) Since A is positive definite by 4.5.7 there exits a nonsingular matrix S such that $A = S^H S$. Then, $tA + (1 - t)B = S^H(tI + (1 - t)(S^H)^{-1}BS^{-1})S$, so:
$g(t) = -\log(\det(tA + (1 - t)B)) = -\log(\det(S^H(tI + (1 - t)(S^H)^{-1}BS^{-1})S))$. Then:
$-\log(\det(S^H(tI + (1 - t)(S^H)^{-1}BS^{-1})S))$
$= -\log(\det(S^H)) - \log(\det(tI + (1 - t)(S^H)^{-1}BS^{-1})) - \log(\det(S))$
$= -\log(\det(S^H)\det(S)) - \log(\det(tI + (1 - t)(S^H)^{-1}BS^{-1}))$
$= -\log(\det(A)) - \log(\det(tI + (1 - t)(S^H)^{-1}BS^{-1}))$.

iic) Since $A, B \in PD_n(\mathbb{R})$, then $B^{-1} \in PD_n(\mathbb{R})$ and $x^H SB^{-1}S^H x = (S^H x)^H B^{-1}(xS) > 0$ so $((S^H)^{-1}BS^{-1})^{-1} = SB^{-1}S^H$ is positive definite.
Further, $(S^H)^{-1}BS^{-1}$ is positive definite.
Let $\{\lambda_i\}_i$ be the eigenvalues of $((S^H)^{-1}BS^{-1})$ and $\{x_i\}_i$ the corresponding eigenvectors. Then for every i: $(tI + (1 - t)(S^H)^{-1}BS^{-1})x_i = tx_i + (1 - t)\lambda_i x_i = (t + (1 - t)\lambda_i)x_i$.
So $\{t + (1 - t)\lambda_i\}_i$ are the eigenvalues of $(tI + (1 - t)(S^H)^{-1}BS^{-1})$ corresponding to the $\{x_i\}_i$
Now:
$-\log(\det(A)) - \log(\det(tI + (1 - t)(S^H)^{-1}BS^{-1}))$
$= -\log(\det(A)) - \log(\Pi_{i=1}^n(t + (1 - t)\lambda_i))$
$= -\log(\det(A)) - \sum_{i=1}^n \log((t + (1 - t)\lambda_i))$.

iid)Using iic) we get $g'(t) \sum_{i=1}^n (1 - \lambda_i)/(t + (1 - t)\lambda_i)$ and $g''(t) = \sum_{i=1}^n (1 - \lambda_i)^2/(t + (1 - t)\lambda_i)^2$, which is nonnegative for all $t \in [0, 1]$.

7.13
Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is convex and bounded above.

Then $\exists M \in \mathbb{R}$ such that $f(x) \leq M \; \forall x \in \mathbb{R}^n$.
For contradiction suppose that f is not a constant function. Then $\exists x, y \in \mathbb{R}R^n$ such that $f(x) \neq f(y)$. The line connecting $(x, f(x))$ and $(y, f(y))$ is above f for values between x and y and is below f for values not between x and y. Since the line connecting $(x, f(x))$ and $(y, f(y))$ is not constant, it intersects and exceeds M for some values in $\mathbb{R}^n$ that are not between $(x, f(x))$ and $(y, f(y))$. Since f is above the line connecting $(x, f(x))$ and $(y, f(y))$ at those values, this is a contradiction.

7.20

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ and $-f$ are convex.
Let $x, y \in \mathbb{R}^n$, with $x \neq y$, and let $\lambda \in [0, 1]$.
f convex $\Rightarrow f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$.
$-f$ convex $\Rightarrow -f(\lambda x + (1 - \lambda)y) \leq -\lambda f(x) - (1 - \lambda)f(y)$
So we must have $f(\lambda x + (1 - \lambda)y) = \lambda f(x) + (1 - \lambda)f(y)$. $\Rightarrow f$ is affine.

7.21

$\Rightarrow$
Let $x^* \in \mathbb{R}^n$ be a local minimizer of f.
Then there exists an open neighborhood V of x^* such that $f(x^*) \leq f(x)$ for all $x \in V$.
Since ϕ is strictly increasing, $\phi(f(x^*)) \leq \phi(f(x))$ for all $x \in V$.
Thus, x^* is a local minimizer of $\phi \circ f$.
$\Leftarrow$
Suppose x^* is a local minimizer of $\phi \circ f$.
Then there exists an open neighborhood U of x^* such that $\phi(f(x^*)) \leq \phi(f(x))$ for all $x \in U$.
Since ϕ is strictly increasing, $f(x^*) \leq f(x)$ for all $x \in U$.
$\Rightarrow x^*$ is a local minimizer of f.