EE563 Convex Optimization

Assignment 02 Solution

Problem 1

a) $f(x) = \sqrt{x}$, $\mathrm{dom} f = \mathbb{R}_+$

Since $f''(x) = -\frac{1}{4} x^{-3/2} \leq 0$ for $x \geq 0$,

$f(x) = \sqrt{x}$ is concave & quasi-linear ie; both quasi-convex & quasiconcave because its α-sublevel & α-superlevel sets are intervals on Real line & hence, convex.

b) $f(x_1, x_2) = \frac{1}{x_1 x_2}$, $\mathrm{dom} f = \mathbb{R}^2_{++}$

Since, $\log f = -\log(x_1 x_2)$ which is convex on $\mathbb{R}^2_{++}$,

$f(x_1, x_2) = \frac{1}{x_1 x_2}$ is log-convex & hence convex (from composition Rules). Now $g = -x_1 x_2$ is quasiconvex (see Example 3.31) & $h(x) = \frac{-1}{x}$ is non-decreasing on R_{++} $\Rightarrow$ $f(x_1, x_2) = \frac{1}{x_1 x_2} = h \circ g$ is quasi-convex. (see composition rules)

c) $f(x, Q) = x^T Q x$, $\mathrm{dom} f = \mathbb{R}^n_{++} \times S^n_{++}$

Now $\nabla^2 f = 2Q \succ 0 \Rightarrow f(x, Q)$ is convex in x.

Also $\nabla^2 f = 0 \Rightarrow f(x, Q)$ is linear in Q.

Hence, $f(x, Q)$ is convex.

Since $f(x, Q)$ & $-f(x, Q)$ are non-decreasing & non-increasing respectively, both are quasiconvex $\Rightarrow f(x, Q)$ is quasilinear.

d) $f(x) = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$, $0 \le p \le 1$, $\text{dom} f = \mathbb{R}^n$

Now, $f(x) = g(x) = \left(\sum_{i=1}^{n} x_i^p \right)^{1/p}$ for $\text{dom} f = \mathbb{R}^n_{++}$

$$\Rightarrow \frac{\partial g}{\partial x_i} = \frac{1}{p}\left(\sum_{i=1}^{n} x_i^p\right)^{\frac{1}{p}-1} \cdot p x_i^{p-1} = \left(\frac{g(x)}{x_i}\right)^{1-p}$$

$$\frac{\partial^2 g}{\partial x_i^2} = (1-p)\left(\frac{g(x)}{x_i}\right)^{1-p-1} \cdot \left[\frac{1}{x_i}\cdot\left(\frac{g(x)}{x_i}\right)^{1-p} - \frac{g(x)}{x_i^2}\right]$$

$$= \frac{1-p}{g(x)}\left(\frac{g(x)^2}{x_i^2}\right)^{1-p} - \frac{1-p}{x_i}\left(\frac{g(x)}{x_i}\right)^{1-p}$$

Also, $$\frac{\partial^2 g}{\partial x_i \partial x_j} = (1-p)\left(\frac{g(x)}{x_j}\right)^{1-p-1} \cdot \frac{1}{x_j}\left(\frac{g(x)}{x_i}\right)^{1-p} = \left(\frac{1-p}{g(x)}\right)\left(\frac{g^2(x)}{x_i x_j}\right)^{1-p}$$

Now,
$$u^T \nabla^2 g(x) u = \frac{1-p}{g(x)}\left[\left(\sum_{i=1}^{n} u_i\left(\frac{g(x)}{x_i}\right)^{1-p}\right)^2 - \sum_{i=1}^{n} u_i^2 \left(\frac{g(x)}{x_i}\right)^{2-p}\right]$$

Letting $a_i = \left(\frac{g(x)}{x_i}\right)^{-\frac{p}{2}}$ & $b_i = u_i\left(\frac{g(x)}{x_i}\right)^{1-\frac{p}{2}}$

$$u^T \nabla^2 g(x) u = \frac{1-p}{g(x)}\left[(a^T b)^2 - (\|a_2\| \|b\|_2)^2\right] \quad \left(\because \|a_2\| = \sum_{i=1}^{n} a_i^2 = 1\right)$$

By Cauchy-Schwarz inequality, $a^T b \le \|a\|_2 \|b\|_2$,

$u^T \nabla^2 g(x) u \le 0 \Rightarrow g(x)$ is concave.

Since, $f(x)$ is symmetric about origin, across the origin it cannot be concave. Hence, $f(x)$ is neither convex nor concave. However, its α-sublevel sets are convex,

$\Rightarrow f(x)$ is quasiconvex.

e) $f(x,y,z) = \ln(xyz)$, $\text{dom} f = \{(x,y,z) \mid x>0, y>0, z>0\}$

$f: \mathbb{R}^3 \to \mathbb{R}$, $f(x,y,z)$ is concave on dom f & non-negative as well. It is also non-decreasing over dom f $\Rightarrow$ α-sublevel & α-super-level sets are convex sets $\Rightarrow$ it is quasilinear.

f) $f(x) = 1/x^p$, $p>0$, $\text{dom} f = R_{++}$

$$f''(x) = \underbrace{\frac{p(p+1)}{x^2}}_{\geq 0} \cdot \underbrace{\frac{1}{x^p}}_{\geq 0} \geq 0 \Rightarrow f(x) \text{ is convex.}$$

Also, $f'(x) = -\frac{p}{x} \cdot \frac{1}{x^p} \leq 0 \Rightarrow f(x)$ is non-increasing.

& $-f'(x) = \frac{p}{x} \cdot \frac{1}{x^p} \geq 0 \Rightarrow -f(x)$ is non-decreasing.

$\Rightarrow f(x)$ is quasiconvex & & $-f(x)$ is quasiconvex.

$\Rightarrow f(x)$ is quasilinear.

g) $f(x) = -\sqrt[n]{\prod_{i=1}^{n} x_i}$, $\text{dom} f = \mathbb{R}^n_+$

$f(x) = h(g(x))$; $h(x) = x$, $g(x) = -\sqrt[n]{\prod_{i=1}^{n} x_i}$ $\leftarrow$ Convex

$\uparrow$ non-decreasing affine

$\Rightarrow f(x)$ is convex & quasiconvex.

See Textbook Section 3.1 (Page 74) for the concavity of the geometric mean.

h) $f(x_1,x_2) = x_1^{\alpha} x_2^{(1-\alpha)}$, $\text{dom} f = \mathbb{R}^2_{++}$; $0 \leq \alpha \leq 1$

$\log f(x_1,x_2) = \alpha \log x_1 + (1-\alpha)\log x_2$ = non-negative weighted sum of concave functions.

$\Rightarrow f(x_1,x_2)$ is log-concave & hence, quasi concave. ($\because$ log is monotonically increasing function)

$\rightarrow$ Since, $\log f(x_1,x_2)$ is concave & log is concave & non-decreasing,

$\Rightarrow f(x_1,x_2)$ is concave (composition rule).

Problem 2

$$f = g \circ U \Rightarrow f(x) = g(U(x))$$

$g : \mathbb{R} \to \mathbb{R}$ is a non-decreasing function $\Rightarrow g(r) \geq g(s)$ for $r \geq s$

& $U : \mathbb{R}^K \to \mathbb{R}$

$$f[\theta x + (1-\theta)y] = g[U\{\theta x + (1-\theta)y\}]$$

Since, $U\{\theta x + (1-\theta)y\} \geq [\theta U(x) + (1-\theta)U(y)]$ concavity of U

& Since, $g(r) \geq g(s)$ for $r \geq s$,

$$f[\theta x + (1-\theta)y] = g[U\{\theta x + (1-\theta)y\}] \geq g[\theta U(x) + (1-\theta)U(y)]$$

$$\geq \theta\, g[U(x)] + (1-\theta)\, g[U(y)] = \theta f(x) + (1-\theta) f(y)$$

Hence, f is concave.

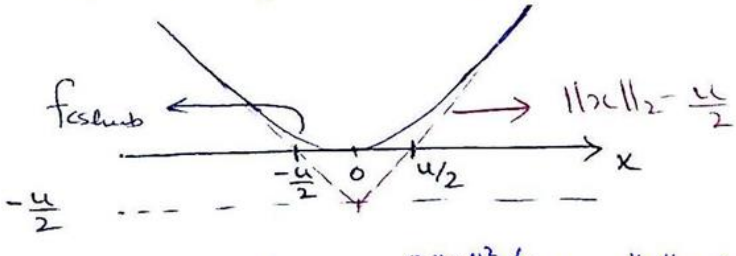
Problem 3

$$f_{hub}(x) = \begin{cases} \frac{1}{2}x^2 , & |x| \leq 1 \\ |x| - \frac{1}{2} , & |x| > 1 \end{cases}$$

$$f_{cshub}(x) = f_{hub}(\|x\|) = \begin{cases} \frac{1}{2}\|x\|_2^2 , & \|x\|_2 \leq 1 \\ \|x\|_2 - \frac{1}{2} , & \|x\|_2 > 1 \end{cases}$$

In general, $f_{cshub}(x) = \begin{cases} \|x\|_2^2 / 2u ; & \|x\|_2 \leq u \\ \|x\|_2 - \frac{u}{2} ; & \|x\|_2 > u \end{cases}$

In 1D, its graph looks like this $\longrightarrow$



Now $f_{cshub}(x) = \inf_y \left(\|y\|_2 + \frac{1}{2}\|x-y\|_2^2\right) = \begin{cases} \|x\|_2^2/2u , & \|x\|_2 \leq u \\ \|x\|_2 - \frac{1}{2} , & \|x\|_2 > u \end{cases}$

Since $g(x,y)$ is convex in (x,y) & minimization preserves convexity, $f_{cshub}(x)$ is convex.

Problem 4

a) Since 'f' is convex & symmetric,

$$f(Px) = f(x) \quad \& \quad f\left(\sum_{i=1}^{n} \lambda_i x_i\right) \leq \sum_{i=1}^{n} \lambda_i f(x_i)$$

Since, $S = \sum_{i=1}^{n} \lambda_i P_i$ where $\lambda_i \geq 0$ & $\sum_{i=1}^{n} \lambda_i = 1$

& P_i are the permutation matrices,

$$\Rightarrow f(Sx) = f\left(\sum_{i=1}^{n} \lambda_i P_i x\right) \leq \sum_{i=1}^{n} \lambda_i f(P_i x)$$

$$f(P_i x) = f(x)$$

$$\Rightarrow f(Sx) \leq \sum_{i=1}^{n} \lambda_i f(x) = f(x) \sum_{i=1}^{n} \lambda_i$$

$$\Rightarrow f(Sx) \leq f(x).$$

b) $Y = Q \operatorname{diag}(\lambda) Q^T$

$QQ^T = Q^T Q = I \Rightarrow \sum_{i=1}^{n} Q_{ij}^2 = 1 \quad \forall j = 1, 2, \ldots, n$

$\Rightarrow$ if $S_{ij} = Q_{ij}^2$ then

$$\sum_{i=1}^{n} S_{ij} = 1 \quad \forall j = 1, 2, \ldots, n$$

Hence $S_{n \times n}$ is doubly stochastic if $S_{ij} = Q_{ij}^2$.

Now, $Q \operatorname{diag}(\lambda) = \begin{bmatrix} | & | & & | \\ \lambda_1 Q_{i1} & \lambda_2 Q_{i2} & \cdots & \lambda_n Q_{in} \\ | & | & & | \end{bmatrix}$; $i = 1, \ldots, n$

$$\Rightarrow Q \operatorname{diag}(\lambda) Q^T = \begin{bmatrix} | & | & & | \\ \lambda_1 Q_{i1} & \lambda_2 Q_{i2} & \cdots & \lambda_n Q_{in} \\ | & | & & | \end{bmatrix} \begin{bmatrix} - & Q_{i1} & - \\ - & Q_{i2} & - \\ & \vdots & \\ - & Q_{in} & - \end{bmatrix}$$

$$\Rightarrow \operatorname{diag}(Y) = \operatorname{diag}(Q \operatorname{diag}(\lambda) Q^T)$$

$$= \sum_{j=1}^{n} \lambda_j Q_{ij}^2, \quad i = 1, 2, \ldots, n$$

$$= \sum_{j=1}^{n} \lambda_j S_{ij}, \quad i = 1, 2, \ldots, n$$

$$= S\lambda, \quad \lambda = [\lambda_1 \; \lambda_2 \; \cdots \; \lambda_n]^T.$$

c) $\operatorname{diag}(V^T X V) = S\lambda(X)$; $V^T V = VV^T = I$ &

from part b) $\lambda(X)$ are the eigenvalues of X

Now from a),

$$f(\mathrm{diag}(V^T X V)) = f(S\lambda(X)) \leq f(\lambda(X)) \quad \forall S;\ S_{ij} = V_{ij}^2$$

$$\Rightarrow f(\lambda(X)) = \sup_{V \in \mathcal{V}} f(\mathrm{diag}(V^T X V))$$

Since, $h(X,V) = \mathrm{diag}(V^T X V)$ is linear in X, it is convex in X. So $f(h(X,V))$ is composition of a convex function 'f' with affine mapping $h(X,V)$.

$$\Rightarrow g(X,V) = f(h(X,V)) \text{ is convex.}$$

& since, $f(\lambda(X)) = \sup_{V \in \mathcal{V}} g(X,V)$ is the pointwise supremum of convex functions in X, $f(\lambda(X))$ is convex in X.

Problem 5

$$f(x,t) = -\log t(t - x^T x/t) = -\log t - \log(t - x^T x/t)$$

The first term is convex and the second term is the composition of a decreasing convex function and a concave function and is therefore convex.

Problem 6

α-sub-level set of a function is given by

$$\begin{aligned} S_\alpha &= \{x \mid c^T x + d > 0,\ (a^T x + b)/(c^T x + d) \leq \alpha\} \\ &= \{x \mid c^T x + d > 0,\ a^T x + b \leq \alpha(c^T x + d)\}, \end{aligned}$$

which is convex since it is the intersection of an open halfspace and a closed halfspace.

To analyze convexity, we form the Hessian:

$$\nabla^2 f(x) = -(c^T x + d)^{-2}(ac^T + ca^T) + (a^T x + b)(c^T x + d)^{-3} cc^T.$$

First assume that a and c are not colinear. In this case, we can find x with $c^T x + d = 1$ (so, $x \in \mathbf{dom}\, f$) with $a^T x + b$ taking any desired value. By taking it as a large and negative, we see that the Hessian is not positive semidefinite, so f is not convex.

So for f to be convex, we must have a and c colinear. If c is zero, then f is affine (hence convex). Assume now that c is nonzero, and that $a = \alpha c$ for some $\alpha \in \mathbf{R}$. In this case, f reduces to

$$f(x) = \frac{\alpha c^T x + b}{c^T x + d} = \alpha + \frac{b - \alpha d}{c^T x + d},$$

which is convex if and only if $b \geq \alpha d$.

So a linear fractional function is convex only in some very special cases: it is affine, or a constant plus a nonnegative constant times the inverse of $c^T x + d$.

Problem 7

For $f(x) = \log(\sum_{i=1}^n e^{x_i})$, we first determine the values of y for which the maximum over x of $y^T x - f(x)$ is attained. By setting the gradient with respect to x equal to zero, we obtain the condition

$$y_i = \frac{e^{x_i}}{\sum_{j=1}^n e^{x_j}}, \qquad i = 1, \ldots, n.$$

These equations are solvable for x if and only if $y \succ 0$ and $\mathbf{1}^T y = 1$. By substituting the expression for y_i into $y^T x - f(x)$ we obtain $f^*(y) = \sum_{i=1}^n y_i \log y_i$. This expression for f^* is still correct if some components of y are zero, as long as $y \succeq 0$ and $\mathbf{1}^T y = 1$, and we interpret $0 \log 0$ as 0.

In fact the domain of f^* is exactly given by $\mathbf{1}^T y = 1$, $y \succeq 0$. To show this, suppose that a component of y is negative, say, $y_k < 0$. Then we can show that $y^T x - f(x)$ is unbounded above by choosing $x_k = -t$, and $x_i = 0$, $i \neq k$, and letting t go to infinity.

If $y \succeq 0$ but $\mathbf{1}^T y \neq 1$, we choose $x = t\mathbf{1}$, so that

$$y^T x - f(x) = t\mathbf{1}^T y - t - \log n.$$

If $\mathbf{1}^T y > 1$, this grows unboundedly as $t \to \infty$; if $\mathbf{1}^T y < 1$, it grows unboundedly as $t \to -\infty$.

In summary,

$$f^*(y) = \begin{cases} \sum_{i=1}^n y_i \log y_i & \text{if } y \succeq 0 \text{ and } \mathbf{1}^T y = 1 \\ \infty & \text{otherwise.} \end{cases}$$

In other words, the conjugate of the log-sum-exp function is the negative entropy function, restricted to the probability simplex.

Problem 8

a) Γ function is log-convex since $u^{x-1}e^{-u}$ is log-convex in x for each $u > 0$.

b)

We prove that

$$h(X) = \log f(X) = \log\det X - \log \mathbf{tr}\, X$$

is concave. Consider the restriction on a line $X = Z + tV$ with $Z \succ 0$, and use the eigenvalue decomposition $Z^{-1/2}VZ^{-1/2} = Q\Lambda Q^T = \sum_{i=1}^n \lambda_i q_i q_i^T$:

$$\begin{aligned}
h(Z+tV) &= \log\det(Z+tV) - \log\mathbf{tr}(Z+tV) \\
&= \log\det Z - \log\det(I + tZ^{-1/2}VZ^{-1/2}) - \log\mathbf{tr}\, Z(I + tZ^{-1/2}VZ^{1/2}) \\
&= \log\det Z - \sum_{i=1}^n \log(1+t\lambda_i) - \log\sum_{i=1}^n (q_i^T Z q_i)(1+t\lambda_i)) \\
&= \log\det Z + \sum_{i=1}^n \log(q_i^T Z q_i) - \sum_{i=1}^n \log((q_i^T Z q_i)(1+t\lambda_i)) \\
&\quad - \log\sum_{i=1}^n ((q_i^T Z q_i)(1+t\lambda_i)),
\end{aligned}$$

which is a constant, plus the function $\displaystyle\sum_{i=1}^n \log y_i - \log\sum_{i=1}^n y_i$

evaluated at $y_i = (q_i^T Z q_i)(1+t\lambda_i)$.

This is concave since product over sum is log-concave.

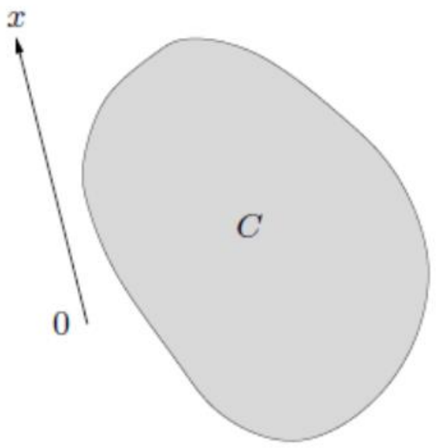
c)

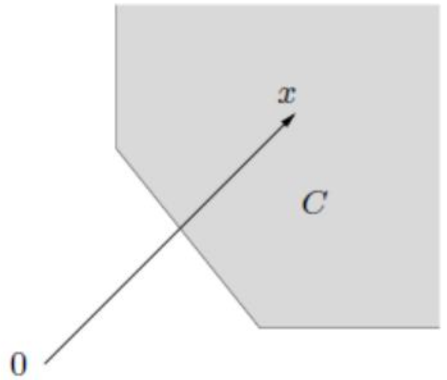
$$\log(e^x/(1+e^x)) = x - \log(1+e^x).$$

The first term is linear, hence concave. Since the function $\log(1+e^x)$ is convex (it is the log-sum-exp function, evaluated at $x_1 = 0$, $x_2 = x$), the second term above is concave. Thus, $e^x/(1+e^x)$ is log-concave.

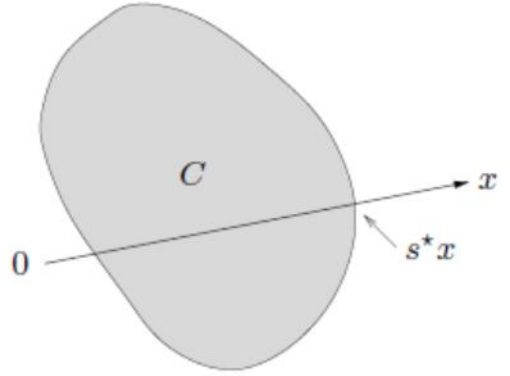
Problem 9

a) Consider the ray, excluding 0, generated by x, *i.e.*, sx for $s > 0$. The intersection of this ray and C is either empty (meaning, the ray doesn't intersect C), a finite interval, or another ray (meaning, the ray enters C and stays in C).
In the first case, the set $\{t > 0 \mid t^{-1}x \in C\}$ is empty, so the infimum is ∞. This means $M_C(x) = \infty$. This case is illustrated in the figure below, on the left.
In the third case, the set $\{s > 0 \mid sx \in C\}$ has the form $[a, \infty)$ or (a, ∞), so the set $\{t > 0 \mid t^{-1}x \in C\}$ has the form $(0, 1/a]$ or $(0, 1/a)$. In this case we have $M_C(x) = 0$. That is illustrated in the figure below to the right.





In the second case, the set $\{s > 0 \mid sx \in C\}$ is a bounded , interval with endpoints $a \le b$, so we have $M_C(x) = 1/b$. That is shown below. In this example, the optimal scale factor is around $s^\star \approx 3/4$, so $M_C(x) \approx 4/3$.



In any case, if $x = 0 \in C$ then $M_C(0) = 0$.

b) If $\alpha > 0$, then

$$\begin{aligned} M_C(\alpha x) &= \inf\{t > 0 \mid t^{-1}\alpha x \in C\} \\ &= \alpha \inf\{t/\alpha > 0 \mid t^{-1}\alpha x \in C\} \\ &= \alpha M_C(x). \end{aligned}$$

If $\alpha = 0$, then

$$M_C(\alpha x) = M_C(0) = \begin{cases} 0 & 0 \in C \\ \infty & 0 \notin C. \end{cases}$$

c) $\mathbf{dom}\, M_C = \{x \mid x/t \in C \text{ for some } t > 0\}$. This is also known as the conic hull of C, except that $0 \in \mathbf{dom}\, M_C$ only if $0 \in C$.

d) We have already seen that $\mathbf{dom}\, M_C$ is a convex set. Suppose $x, y \in \mathbf{dom}\, M_C$, and let $\theta \in [0,1]$. Consider any $t_x, t_y > 0$ for which $x/t_x \in C$, $y/t_y \in C$. (There exists at least one such pair, because $x, y \in \mathbf{dom}\, M_C$.) It follows from convexity of C that

$$\frac{\theta x + (1-\theta)y}{\theta t_x + (1-\theta)t_y)} = \frac{\theta t_x(x/t_x) + (1-\theta)t_y(y/t_y)}{\theta t_x + (1-\theta)t_y} \in C$$

and therefore

$$M_C(\theta x + (1-\theta)y) \le \theta t_x + (1-\theta)t_y.$$

This is true for any $t_x, t_y > 0$ that satisfy $x/t_x \in C$, $y/t_y \in C$. Therefore

$$\begin{aligned} M_C(\theta x + (1-\theta)y) &\le \theta \inf\{t_x > 0 \mid x/t_x \in C\} + (1-\theta)\inf\{t_y > 0 \mid y/t_y \in C\} \\ &= \theta M_C(x) + (1-\theta)M_C(y). \end{aligned}$$

Here is an alternative snappy, modern style proof:

- The indicator function of C, *i.e.*, I_C, is convex.
- The perspective function, $tI_C(x/t)$ is convex in (x,t). But this is the same as $I_C(x/t)$, so $I_C(x/t)$ is convex in (x,t).
- The function $t + I_C(x/t)$ is convex in (x,t).
- Now let's minimize over t, to obtain $\inf_t(t + I_C(x/t)) = M_C(x)$, which is convex by the minimization rule.

e) It is the norm with unit ball C.

(a) Since by assumption, $0 \in \mathbf{int}\, C$, $M_C(x) > 0$ for $x \neq 0$. By definition $M_C(0) = 0$.

(b) Homogeneity: for $\lambda > 0$,

$$\begin{aligned} M_C(\lambda x) &= \inf\{t > 0 \mid (t\lambda)^{-1}x \in C\} \\ &= \lambda \inf\{u > 0 \mid u^{-1}x \in C\} \\ &= \lambda M_C(x). \end{aligned}$$

By symmetry of C, we also have $M_C(-x) = -M_C(x)$.

(c) Triangle inequality. By convexity (part d), and homogeneity,

$$M_C(x+y) = 2M_C((1/2)x + (1/2)y) \le M_C(x) + M_C(y).$$