DSA3102: Solutions to Tutorial Set 5
Assigned: 07/09/23

1. BV Problem 3.21
   Solution:

   (a) f is the pointwise maximum of k functions $\|A^{(i)}x - b^{(i)}\|$. Each of those functions is convex because it is the composition of an affine transformation and a norm.

   (b) Write f as
   $$f(x) = \sum_{i=1}^{r} |x|_{[i]} = \max_{1 \leq i_1 < i_2 < \ldots < i_r \leq n} |x_{i_1}| + \ldots + |x_{i_n}|$$
   which is the pointwise maximum of $\binom{n}{r}$ convex functions.

2. BV Problem 3.31
   Solution:

   (a) If $t > 0$,
   $$g(tx) = \inf_{\alpha > 0} \frac{f(\alpha t x)}{\alpha} = t \inf_{\alpha > 0} \frac{f(\alpha t x)}{t\alpha} = tg(x)$$
   For $t = 0$, we have $g(tx) = g(0) = 0$.

   (b) If h is a homogeneous underestimator, then
   $$h(x) = \frac{h(\alpha x)}{\alpha} \leq \frac{f(\alpha x)}{\alpha}$$
   for all $\alpha > 0$. Taking the infimum over α gives $h(x) \leq g(x)$.

   (c) We can express g as
   $$g(x) = \inf_{t>0} t f(x/t) = \inf_{t>0} h(x, t)$$
   where h is the perspective function of f. We know h is convex, jointly in x and t, so g is convex.

3. BV Problem 3.36
   Solutions:

   (a) Max function $f(x) = \max_i x_i$ on $\mathbb{R}^n$. We show that
   $$f^*(y) = \begin{cases} 0 & y \succeq 0, \mathbf{1}^T y = 1 \\ \infty & \text{otherwise} \end{cases}$$
   We first verify the domain of f^*. First suppose y has a negative component, say $y_k < 0$. If we choose a vector x with $x_k = -t$ and $x_i = 0$ for all $i \neq k$ and let $t \to \infty$, we see that
   $$x^T y - \max_i x_i = -t y_k \to \infty,$$

so $y \notin \mathbf{dom} f^*$. Next $y \succeq 0$ but $\mathbf{1}^T y > 1$. We choose $x = t\mathbf{1}$ and let t go to infinity, to show that

$$x^T y - \max_i x_i = t\mathbf{1}^T y - t$$

is unbounded above. The same argument goes for $y \succeq 0$ but $\mathbf{1}^T y < 1$.
The remaining case is $y \succeq 0$ but $\mathbf{1}^T y = 1$. In this case, we have that

$$x^T y \leq \max_i x_i$$

for all x and therefore $x^T y - \max_i x_i \leq 0$ for all x with equality when $x = 0$. Therefore $f^*(y) = 0$.

(b) Sum of largest elements $f(x) = \sum_{i=1}^r x_{[i]}$ on $\mathbb{R}^n$. The conjugate is

$$f^*(y) = \begin{cases} 0 & 0 \preceq y \preceq \mathbf{1}, \mathbf{1}^T y = r \\ \infty & \text{otherwise} \end{cases}$$

We first verify the domain of f^*. Suppose y has a negative component, say $y_k < 0$. If we choose a vector x with $x_k = -t$, $x_i = 0$ for all $i \neq k$ and let $t \to \infty$, we obtain

$$x^T y - f(x) = -ty_k \to \infty$$

so $y \notin \mathbf{dom} f^*$. Next suppose y has a component greater than one, say $y_k > 1$. If we choose a vector x with $x_k = t$ and $x_i = 0$ for all $i \neq k$, and let $t \to \infty$, we have

$$x^T y - f(x) = ty_k - t \to \infty$$

so $y \notin \mathbf{dom} f^*$. Finally assume that $\mathbf{1}^T x \neq r$. We choose $x = t\mathbf{1}$ and find that

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

is unbounded above as $t \to \infty$ or $t \to -\infty$.
If y satisfies all the conditions, we have

$$x^T y \leq f(x)$$

for all x with equality if $x = 0$. Therefore $f^*(y) = 0$.

(c) Piecewise linear function on $\mathbb{R}$: $f(x) = \max_{i=1,\ldots,m}(a_i x + b_i)$ on $\mathbb{R}$. You can assume that the a_i are sorted in increasing order, and that none of the functions $a_i x + b_i$ is redundant.
Under the assumption, the graph of f is a piecewise-linear, with break-points

$$(b_i - b_{i+1})/(a_{i+1} - a_i), \qquad i = 1, \ldots, m-1.$$

We can write f^* as

$$f^*(y) = \sup_x \left\{ xy - \max_{i=1,\ldots,m} (a_i x + b_i) \right\}$$

We see that $\mathbf{dom}\, f = [a_1, a_m]$, since for y outside that range, the expression inside the supremum is unbounded above. For $a_i \leq y \leq a_{i+1}$, the supremum in the definition of f^* is reached at the breakpoint between the segments i and $i+1$, i.e., at the point $(b_i - b_{i+1})/(a_{i+1} - a_i)$ so we obtain

$$f^*(y) = -b_i - (b_{i+1} - b_i)\frac{y - a_i}{a_{i+1} - a_i}$$

where i is defined by $a_i \leq y \leq a_{i+1}$. Hence the graph of r^* is also a piecewise-linear curve connecting the points $(a_i, -b_i)$ for $i = 1, \ldots, m$. Geometrically, the epigraph of f^* is the epigraphical hull of the points $(a_i, -b_i)$.

(d) Power function: $f(x) = x^p$ on $\mathbb{R}_{++}$ where $p > 1$. Repeat for $p < 0$.

We let q be the conjugate of p, i.e., $1/p + 1/q = 1$.

We start with the case $p > 1$. Then x^p is strictly convex on $\mathbb{R}_+$. For $y < 0$ the function $yx - x^p$ achieves its maximum for $x > 0$ at $x = 0$ so $f^*(y) = 0$. For $y > 0$ the function achieves its maximum at $x = (y/p)^{1/(p-1)}$, where it has value

$$y(y/p)^{1/(p-1)} - (y/p)^{p/(p-1)} = (p-1)(y/p)^q$$

Therefore we have

$$f^*(y) = \begin{cases} 0 & y \le 0 \\ (p-1)(y/p)^q & y > 0 \end{cases}$$

For $p < 0$, similar arguments show that $\mathbf{dom}\, f^* = -\mathbb{R}_{++}$ and

$$f^*(y) = -\frac{p}{q}(-y/p)^q.$$

(e) Geometric mean: $f(x) = -(\prod_i x_i)^{1/n}$ on $\mathbb{R}^n_{++}$. The conjugate function is

$$f^*(y) = \begin{cases} 0 & y \preceq 0, (\prod_i(-y_i))^{1/n} \ge 1/n \\ \infty & \text{otherwise} \end{cases}$$

We first verify the domain of f^*. Assume y has a positive component, say $y_k > 0$. Then we can choose $x_k = t$ and $x_i = 1$ for all $i \ne k$, to show that

$$x^T y - f(x) = ty_k + \sum_{i \ne k} y_i + t^{1/n}$$

is unbounded above as a function of $t > 0$. Hence the condition $y \preceq 0$ is indeed required.

Next assume that $y \preceq 0$ but $(\prod_i(-y_i))^{1/n} < 1/n$. Then we choose $x_i = -t/y_i$ and obtain

$$x^T y - f(x) = -tn - t\left(\prod_i(-1/y_i)\right)^{1/n} \to \infty.$$

as $t \to \infty$. This shows that $(\prod_i(-y_i))^{1/n} \ge 1/n$ is needed.

Now assume both conditions are satisfied and $x \succeq 0$. The arithmetic mean-geometric mean (AM-GM) inequality states that

$$\frac{x^T y}{n} \ge \left(\prod_i(-y_i x_i)\right)^{1/n} \ge \frac{1}{n}\left(\prod_i x_i\right)^{1/n}$$

That is $x^T y \ge f(x)$ with equality iff $x_i = -1/y_i$. Hence $f^*(y) = 0$.

(f) Negative generalized logarithm for second-order cone. $f(x,t) = -\log(t^2 - x^T x)$ on $\{(x,t) : \|x\|_2 \le t\}$. The conjugate is

$$f^*(y,u) = -2 + \log 4 - \log(u^2 - y^T y), \qquad \mathbf{dom}\, f^* = \{(y,u) : \|y\|_2 < -u\}.$$

We first verify the domain. Suppose $\|y\|_2 \ge -u$. Choose any $x = sy$, $t = s(\|x\|_2 + 1) > s\|y\|_2 \ge -su$ with $s \ge 0$. Then

$$y^T x + tu > sy^T y - su^2 \ge 0,$$

so $y^T x + tu$ goes to infinity at a linear rate, while the function $-\log(t^2 - x^T x)$ goes to $-\infty$ as $-\log s$. Therefore

$$y^T x + tu + \log(t^2 - x^T x)$$

is unbounded above.

Next assume that $\|y\|_2 < -u$. Setting the derivative of

$$y^T x + tu + \log(t^2 - x^T x)$$

with respect to x and t equal to zero, and solving for t and x, we see that the maximizer is

$$x = \frac{2y}{u^2 - y^T y}, \qquad t = -\frac{2u}{u^2 - y^T y}$$

This gives

$$f^*(y,u) = ut + y^T x + \log(t^2 - x^T x) = -2 + \log 4 - \log(y^2 - u^T u).$$

4. BV Problem 3.37

Solution: We first verify the domain of f^*. Suppose Y has eigenvalue decomposition

$$Y = Q\Lambda Q^T = \sum_i \lambda q_i q_i^T$$

with $\lambda_1 > 0$. Let $X = Q\mathbf{diag}(t, 1, \ldots, 1)Q^T = tq_1q_1^T + \sum_{i=2}^n q_i q_i^T$. We have

$$\mathbf{tr}(XY) - \mathbf{tr}(X^{-1}) = t\lambda_1 + \sum_{i=2}^n \lambda_i - 1/t - (n-1)$$

which grows unboundedly as $t \to \infty$. Therefore $Y \notin \mathbf{dom} f^*$. Next assume that $Y \preceq 0$. If $Y \prec 0$, we can find the maximum of

$$\mathbf{tr}(XY) - \mathbf{tr}(X^{-1})$$

by setting the gradient to zero. We obtain $Y = -X^{-2}$ and

$$f^*(Y) = -2\mathbf{tr}(-Y)^{1/2}$$

Finally we verify that this expression remains valid when $Y \preceq 0$, but Y is singular. This follows from the fact that conjugate functions are always closed, i.e., have closed epigraphs.

5. BV Problem 3.42

Solution: To show that W is quasiconcave we show that the sets $\{x : W(x) \geq \alpha\}$ are convex for all α. We have $W(x) \geq \alpha$ if and only if

$$-\epsilon \leq x_1 f_1(t) + \ldots + x_n f_n(t) - f_0(t) \leq \epsilon$$

for all $t \in [0, \alpha)$. Therefore the set $\{x : W(x) \geq \alpha\}$ is an intersection of infinitely many halfspaces (two for each t), hence a convex set.

6. BV Problem 3.45

Solution: The first and second derivatives of f are

$$\nabla f(x) = \begin{bmatrix} -x_2 & -x_1 \end{bmatrix} \qquad \nabla^2 f(x) = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$$

We start with the first-order condition

$$f(x) \leq f(y) \Rightarrow \nabla f(x)^T (y - x) \leq 0$$

which in this case reduces to

$$-y_1 y_2 \leq -x_1 x_2 \Rightarrow -x_2(y_1 - x_1) - x_1(y_2 - x_2) \leq 0$$

for all $x, y \succ 0$. Simplifying each side,

$$y_1 y_2 \geq x_1 x_2 \Rightarrow 2x_1 x_2 \leq x_1 y_2 + x_2 y_1,$$

and dividing by $x_1 x_2$ (which is positive) we get the equivalent statement

$$(y_1/x_1)(y_2/x_2) \geq 1 \Rightarrow 1 \leq ((y_2/x_2) + (y_1/x_1))/2$$

which is true (it is the arithmetic-geometric mean inequality).

The second-order condition is

$$y^T \nabla f(x) = 0, y \neq 0 \Rightarrow y^T \nabla^2 f(x) y > 0$$

which reduces to

$$-y_1 x_2 - y_2 x_1 = 0, y \neq 0 \Rightarrow -2y_1 y_2 > 0$$

for all $x \succ 0$, i.e.,

$$y_2 = -y_1 x_2 / x_1 \Rightarrow -2y_1 y_2 > 0$$

which is correct if $x \succ 0$.

7. **Solution**: We can compute

$$\frac{\partial^2 f}{\partial x_i^2} = \frac{\alpha_i(\alpha_i - 1)}{x_i^2} \prod_i x_i^{\alpha_i}, \qquad \frac{\partial^2 f}{\partial x_i x_j} = \frac{\alpha_i \alpha_j}{x_i x_j} \prod_i x_i^{\alpha_i}$$

Hence, we may write the Hessian as

$$\nabla^2 f(x) = \left(\prod_i x_i^{\alpha_i}\right) \left[\mathbf{diag}(-\alpha_1/x_1^2, \ldots, -\alpha_n/x_n^2) + qq^T\right]$$

where

$$q_i = \alpha_i / x_i$$

Now, we need to show that $v^T \nabla^2 f(x) v \geq 0$ for all $v \in \mathbb{R}^n$. We have

$$v^T \left[\mathbf{diag}(-\alpha_1/x_1^2, \ldots, -\alpha_n/x_n^2) + qq^T\right] v = -\sum_i \alpha_i \frac{v_i^2}{x_i^2} + \left(\sum_i \frac{\alpha_i v_i}{x_i}\right)^2$$

This is non-positive by the Cauchy–Schwarz inequality $((\langle a, b\rangle^2 \leq \|a\| \|b\|)$. Take $a_i = \alpha_i^{1/2}$ and $b_i = \alpha_i^{1/2} v_i / x_i$ so we have

$$\left(\sum_i a_i b_i\right)^2 \leq \|a\|^2 \|b\|^2 \qquad \Leftrightarrow \qquad \left(\sum_i \frac{\alpha_i v_i}{x_i}\right)^2 \leq \left(\sum_i \alpha_i\right) \left(\sum_i \alpha_i \frac{v_i^2}{x_i^2}\right)$$

Note that $\sum_i a_i = 1$. So the weighted geometric mean is concave.

8. (Convexity and Majorization)

Here is yet another characterization of convexity due to Hardy, Littlewood, and Pólya. For any vector $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$, denote by

$$x_1^{\downarrow} \geq x_2^{\downarrow} \geq \ldots \geq x_n^{\downarrow}$$

the components of x in decreasing order. We say that $x \in \mathbb{R}^n$ is *majorized* by $y \in \mathbb{R}^n$, denoted as $x \prec_{\mathrm{w}} y$ if

$$\sum_{i=1}^k x_i^{\downarrow} \leq \sum_{i=1}^k y_i^{\downarrow} \quad \forall\, k = 1, \ldots, n \quad \text{and} \quad \sum_{i=1}^n x_i^{\downarrow} = \sum_{i=1}^n y_i^{\downarrow}.$$

(a) Show that if x and y are length-n vectors and $x \prec_{\mathrm{w}} y$, then

$$\sum_{k=1}^{n} f(x_k) \leq \sum_{k=1}^{n} f(y_k) \tag{1}$$

for every convex function f whose domain $\mathbf{dom}\, f$ contains the components of x and y.

(b) Conversely, if (1) holds for every convex function f such that $x_i, y_i \in \mathbf{dom}\, f$ for all $i = 1, \ldots, n$.

For part (a), use the result in Problem 3.1(b) of BV and Abel's summation formula which says that

$$\sum_{k=1}^{n} c_k g(k) = C(n)g(n) - \sum_{k=1}^{n-1} C(k)(g(k+1) - g(k)), \quad \textit{where} \quad C(n) = \sum_{k=1}^{n} c_k$$

This is nothing but integration by parts.

If you're interested in this (the interplay between majorization and convexity), please read page 75 of G. M. Hardy, J. E. Littlewood, G. Pólya, "Inequalities", (second ed.), Cambridge University Press, Cambridge (1952).

Solution:

(a) For the first part, let us, first assume that $x_k^{\downarrow} \neq y_k^{\downarrow}$ for all k. Then,

$$\sum_{k=1}^{n} f(y_k^{\downarrow}) - f(x_k^{\downarrow}) = \sum_{k=1}^{n} \left[(y_k^{\downarrow} - x_k^{\downarrow}) \cdot \frac{f(y_k^{\downarrow}) - f(x_k^{\downarrow})}{y_k^{\downarrow} - x_k^{\downarrow}} \right]$$

Take $c_k = y_k^{\downarrow} - x_k^{\downarrow}$ and $g(k) = \frac{f(y_k^{\downarrow}) - f(x_k^{\downarrow})}{y_k^{\downarrow} - x_k^{\downarrow}}$. Then according to Abel's partial summation formula, we have

$$\sum_{k=1}^{n} f(y_k^{\downarrow}) - f(x_k^{\downarrow}) = \sum_{k=1}^{n-1} \left(\frac{f(y_k^{\downarrow}) - f(x_k^{\downarrow})}{y_k^{\downarrow} - x_k^{\downarrow}} - \frac{f(y_{k+1}^{\downarrow}) - f(x_{k+1}^{\downarrow})}{y_{k+1}^{\downarrow} - x_{k+1}^{\downarrow}} \right) \cdot \left(\sum_{i=1}^{k} y_i^{\downarrow} - \sum_{i=1}^{k} x_i^{\downarrow} \right).$$

Note that $C(n) = 0$ because $\sum_{i=1}^{n} x_i^{\downarrow} = \sum_{i=1}^{n} y_i^{\downarrow}$. Now, by the fact $x \prec_{\mathrm{w}} y$, $\sum_{i=1}^{k} y_i^{\downarrow} - \sum_{i=1}^{k} x_i^{\downarrow} \geq 0$ for all k. By the convexity of f and the facts that $x_k^{\downarrow} \geq x_{k+1}^{\downarrow}$ and $y_k^{\downarrow} \geq y_{k+1}^{\downarrow}$, we have $\frac{f(y_k^{\downarrow}) - f(x_k^{\downarrow})}{y_k^{\downarrow} - x_k^{\downarrow}} - \frac{f(y_{k+1}^{\downarrow}) - f(x_{k+1}^{\downarrow})}{y_{k+1}^{\downarrow} - x_{k+1}^{\downarrow}} \geq 0$. This follows from Problem 3.1(a) of BV and intuitively, the slope does not decrease for convex functions. Hence, the sum in the displayed equation above is non-negative. In the case that there exists k such that $x_k^{\downarrow} = y_k^{\downarrow}$, the term $f(y_k^{\downarrow}) - f(x_k^{\downarrow}) = 0$. Together, these show that $\sum_{k=1}^{n} f(x_k^{\downarrow}) \leq \sum_{k=1}^{n} f(y_k^{\downarrow})$ and hence

$$\sum_{k=1}^{n} f(x_k) \leq \sum_{k=1}^{n} f(y_k)$$

as desired.

(b) For the converse, since the identity function is convex, we infer from (1) that $\sum_{i=1}^{n} x_i^{\downarrow} = \sum_{i=1}^{n} y_i^{\downarrow}$. Also, for fixed $1 \leq k \leq n$, using the convexity of $f(x) = \max\{0, x - y_k\}$, we obtain

$$x_1^{\downarrow} + \ldots + x_k^{\downarrow} - k y_k^{\downarrow} \leq \sum_{j=1}^{n} f(x_j^{\downarrow}) \leq \sum_{j=1}^{n} f(y_j^{\downarrow}) \leq y_1^{\downarrow} + \ldots + y_k^{\downarrow} - k y_k^{\downarrow}$$

which yields $x_1^{\downarrow} + \ldots + x_k^{\downarrow} \leq y_1^{\downarrow} + \ldots + y_k^{\downarrow}$ for any $1 \leq k \leq n$. This completes the proof.