Homework 3 solutions

Submit answers for problems 1–6. Problems 7–10 are included as practice problems.

1. *Generalized mean inequalities.* Suppose $w \in \mathbf{R}^n_{++}$ is a positive weight vector, normalized to satisfy $\mathbf{1}^T w = 1$. For positive $x \in \mathbf{R}^n_{++}$, define the generalized mean $f_r(x)$, parametrized by $r \in \mathbf{R}$, as

$$f_0(x) = \prod_{k=1}^n x_k^{w_k}, \qquad f_r(x) = (\sum_{k=1}^n w_k x_k^r)^{1/r} \quad \text{if } r \neq 0.$$

For $r = 0$ this is the weighted geometric mean. Examples for $r \neq 0$ are the weighted arithmetic mean ($r = 1$) and the weighted harmonic mean ($r = -1$).

We show that $f_r(x) \leq f_s(x)$ if $r < s$. This inequality includes as special cases the arithmetic–geometric mean inequality ($r = 0$, $s = 1$) and the arithmetic–harmonic mean inequality ($r = -1$, $s = 1$).

(a) First consider $r = 0 < s$. If we take the logarithm of both sides of the inequality $f_0(x) \leq f_s(x)$, we get the equivalent inequality

$$\sum_{k=1}^n w_k \log x_k \leq \frac{1}{s} \log \sum_{k=1}^n w_k x_k^s.$$

Show that this follows from concavity of the logarithm.

(b) Next consider $0 < r < s$. Show that the inequality $f_r(x) \leq f_s(x)$ follows from convexity of the function $t^{s/r}$ for $t > 0$.

The remaining cases ($r < s = 0$ and $r < s < 0$) follow from (a), (b) and the fact that $f_r(x) = 1/f_{-r}(y)$ where $y_i = 1/x_i$ for $i = 1, \ldots, n$.

(c) What are the conditions on x to have equality $f_r(x) = f_s(x)$ when $r \neq s$?

Solution.

(a) If we rewrite the inequality as

$$\sum_{k=1}^n w_k \log y_k \leq \log \sum_{k=1}^n w_k y_k,$$

where $y_k = x_k^s$, we recognize Jensen's inequality for the logarithm (a concave function).

(b) This inequality can be written as

$$(\sum_{k=1}^{n} w_k y_k)^{s/r} \leq \sum_{k=1} w_k y_k^{s/r},$$

where $y_k = x_k^s$. This is Jensen's inequality for the convex function $y^{s/r}$.

(c) The logarithm is strictly concave, and the function $y^{s/r}$ with $s > r$ is strictly convex. Equality therefore requires that all the components x_k are equal, *i.e.*, x is a constant vector.

2. Suppose the functions $g : \mathbf{R} \to \mathbf{R}$ and $f : \mathbf{R} \to \mathbf{R}$ are convex. Show that the function $h : \mathbf{R}^n \to \mathbf{R}$ defined as

$$h(x) = \inf_z \left(g(z) + \sum_{i=1}^{n} p_i f(x_i - z) \right)$$

is convex, where $p_i > 0$ for $i = 1, \ldots, n$. The variable z in the minimization is a scalar.

Remark. The function h arises in stochastic optimization. Suppose X is a discrete random variable which takes the value x_i with probability p_i. The random variable X is the quantity of a product you will purchase in the future. The price for purchasing the quantity X is $f(X)$. You have the option of preordering an amount z at a price $g(z)$. As a rational consumer, you choose z by minimizing

$$g(z) + \mathbf{E}\, f(X - z) = g(z) + \sum_{i=1}^{n} p_i f(x_i - z).$$

Solution. The function $g(z) + \sum_i p_i f(x_i - z)$ is jointly convex in $x = (x_1, \ldots, x_n)$ and z. Partial minimization over z gives a convex function of x.

3. Exercise A3.5 (a,b).

Solution.

(a) Suppose $t_2 > t_1 > 0$ and $x/t_1 \in \mathbf{dom}\, f$. We write x/t_2 as a convex combination of x/t_1 and 0:

$$x/t_2 = \theta(x/t_1) + (1 - \theta)\, 0$$

where $\theta = t_1/t_2$. Since $0 \in \mathbf{dom}\, f$, this shows that $x/t_2 \in \mathbf{dom}\, f$. Applying Jensen's inequality and the fact that $f(0) \leq 0$, we get

$$\begin{aligned} f(x/t_2) &\leq \frac{t_1}{t_2} f(x/t_1) + (1 - \frac{t_1}{t_2}) f(0) \\ &\leq \frac{t_1}{t_2} f(x/t_1). \end{aligned}$$

Therefore

$$t_2 f(x/t_2) \leq t_1 f(x/t_1).$$

If we assume that f is differentiable, we can also verify that the derivative of the perspective with respect to t is less than or equal to zero. We have

$$\frac{\partial}{\partial t}(tf(x/t)) = f(x/t) - \nabla f(x/t)^T(x/t).$$

This is less than or equal to zero, because, from convexity of f,

$$0 \geq f(0) \geq f(x/t) + \nabla f(x/t)^T(0 - x/t).$$

(b) This follows from the composition theorem on page 3.23 of the slides. The perspective function $tf(y/t)$ is convex in $(y_1, \ldots, y_n, t)$ and nonincreasing in t. Therefore if we substitute linear functions x_i for the arguments y_i and a concave positive function $g(x)$ for the argument t, the result $g(x)f(x/g(x))$ is convex.

We can also establish convexity of h directly from the definition. Consider two points u, $v \in \mathbf{dom}\, g$, with $u/g(u) \in \mathbf{dom}\, f$ and $v/g(v) \in \mathbf{dom}\, f$. We take a convex combination $x = \theta u + (1-\theta)v$. This point lies in $\mathbf{dom}\, g$ because $\mathbf{dom}\, g$ is convex. Also,

$$\begin{aligned} \frac{x}{g(x)} &= \frac{\theta g(u)}{g(x)}\frac{u}{g(u)} + \frac{(1-\theta)g(v)}{g(x)}\frac{v}{g(v)} \\ &= \mu_1\frac{u}{g(u)} + \mu_2\frac{v}{g(v)} + \mu_3 \cdot 0 \end{aligned}$$

where

$$\mu_1 = \frac{\theta g(u)}{g(x)}, \qquad \mu_2 = \frac{(1-\theta)g(v)}{g(x)}, \qquad \mu_3 = 1 - \mu_1 - \mu_2.$$

These coefficients are nonnegative and add up to one because $g(x) > 0$ on its domain, and $\theta g(u) + (1-\theta)g(v) \leq g(x)$ by concavity of g. Therefore $x/g(x)$ is a convex combination of three points in $\mathbf{dom}\, f$. Since $\mathbf{dom}\, f$ is a convex set, $x/g(x) \in \mathbf{dom}\, f$. This shows that $\mathbf{dom}\, h$ is convex.

Next we apply Jensen's inequality for f:

$$\begin{aligned} f(x/g(x)) &\leq \mu_1 f(u/g(u)) + \mu_2 f(v/g(v)) + \mu_3 f(0) \\ &\leq \mu_1 f(u/g(u)) + \mu_2 f(v/g(v)) \end{aligned}$$

because $f(0) \leq 0$. Substituting the expressions for μ_1 and μ_2 we get

$$g(x)f(x/g(x)) \leq \theta g(u)f(u/g(u)) + (1-\theta)g(v)f(v/g(v)).$$

This is Jensen's inequality $h(x) \leq \theta h(u) + (1-\theta)h(v)$.

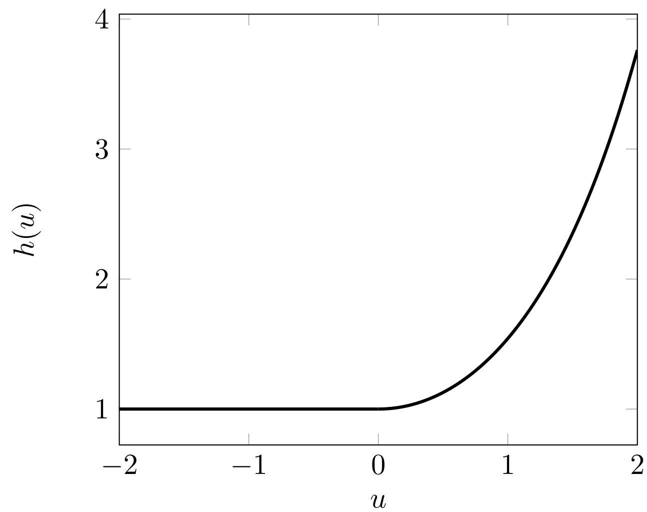
4. Exercise A3.48 (a,b). The function in part (b) can be treated as an application of the perspective composition rule of A3.5(b). Another short proof is to show that the epigraph of f is a convex set.

Solution.

(a) The hyperbolic cosine $\cosh(u)$ is convex (a sum of two convex functions) but not increasing (its derivative is $\cosh(u)' = (\exp(u) - \exp(-u))/2$, so it is increasing for $u > 0$ and decreasing for $u < 0$). To apply a composition rule we write f as $f(x) = h(g(x))$ where

$$h(u) = \cosh(\max\{u, 0\}), \qquad g(x) = \|x\|,$$

The function h is convex and increasing (see figure) and g is convex.



(b) The function $h(x, y) = (x^T Ax)/y$, with domain $\{(x, y) \mid y > 0\}$ is the perspective of $x^T Ax$. It is jointly convex in x and y, and nonincreasing in y. By the perspective composition rule, the composition $f(x) = h(x, g(x))$ is convex if g is concave and positive.
We can also note that the epigraph is convex. The inequality $(x^T Ax)/g(x) \le t$ with $t > 0$ and $g(x) > 0$ is equivalent to $(x^T Ax)/t \le g(x)$. Therefore the epigraph is a sublevel set of the function $(x^T Ax)/t - g(x)$ which is jointly convex in (x, t).

5. Exercise A3.44.

Solution.

(a) This follows from the composition rules: $f(x) = h(g_1(x), \ldots, g_n(x))$, where h is nondecreasing in each argument and $g_k(x) = |x_k|$, a convex function.

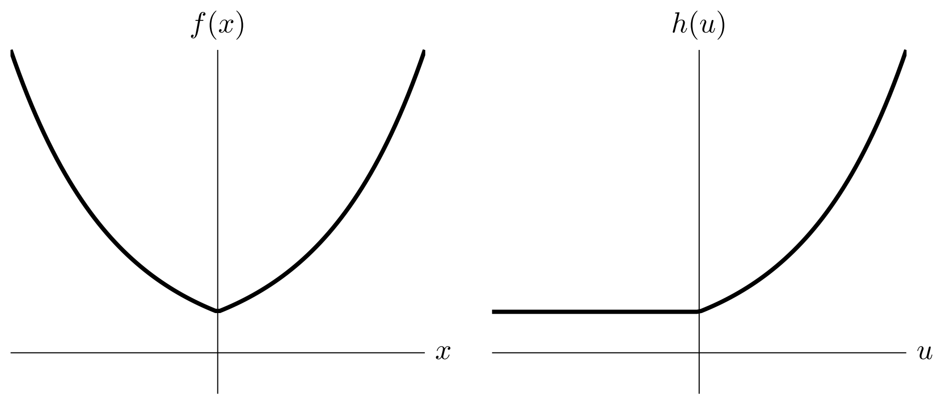
(b) The conjugate is

$$\begin{aligned} f^*(y) &= \sup_x \,(y^T x - h(|x_1|, \ldots, |x_n|)) \\ &= \sup_{u \succeq 0} \max_{s \in \{-1,1\}^n} \,(\sum_k y_k s_k u_k - h(u)) \\ &= \sup_{u \succeq 0} \,(\sum_k |y_k| u_k - h(u)) \end{aligned}$$

$$
\begin{aligned}
&= \sup_u \left(\sum_k |y_k| u_k - h(u)\right) \\
&= f^*(|y_1|, \ldots, |y_n|).
\end{aligned}
$$

On line 2 we write x as $x = \mathbf{diag}(s)u$ with u nonnegative and s a sign vector. Line 3 follows because the optimal choice of s in the maximization problem is $s_k = \operatorname{sign}(y_k)$. On line 4 we use the property $h(u) = h((u_1)_+, \ldots, (u_n)_+)$. Because of this property, allowing negative components in u does not increase the maximum value of $|y|^T u - h(u)$.

(c) The figure shows the functions $f(x) = \exp(|x|)$ and $h(u) = \exp(u_+)$.



The conjugate of $f(x) = \exp(|x|)$ is

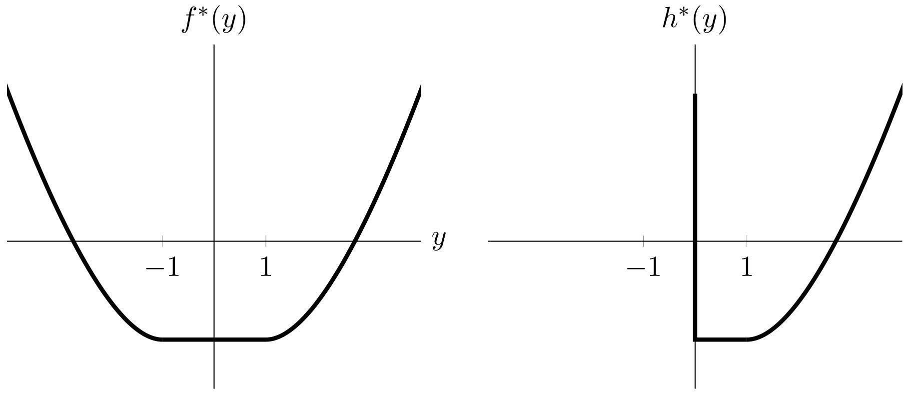
$$
f^*(y) = \sup_x (yx - \exp(|x|)) = \begin{cases} |y| \log |y| - |y| & |y| \geq 1 \\ -1 & |y| \leq 1. \end{cases}
$$

If $|y| \leq 1$, the maximum over x is at $x = 0$. If $y > 1$ we differentiate $yx - e^x$ to find that $x = \log y$ is optimal and the optimal value is $y \log y - y$. If $y < -1$ we differentiate $yx - e^{-x}$ to find that $x = -\log(-y)$ is optimal and the optimal value is $-y \log(-y) + y$.

The conjugate of $h(x) = \exp(x_+)$ is

$$
h^*(y) = \sup_x (yx - \exp(x_+)) = \begin{cases} y \log y - y & y \geq 1 \\ -1 & 0 \leq y \leq 1 \\ +\infty & y < 0. \end{cases}
$$

If $y > 1$, we differentiate $xy - e^x$ to find that $x = \log y$ is optimal. If $0 \leq y \leq 1$, the maximum over x is at $x = 0$. For $y < 0$, the function $xy - \exp(x_+)$ goes to $+\infty$ as $x \to -\infty$.



6. Exercise A4.17.

Solution. We solve the problem in CVX, using the formulation

$$\begin{array}{ll} \text{minimize} & \sum_{t=0}^{N-1} \max\{|u(t)|, 2|u(t)|-1\} \\ \text{subject to} & x(t+1) = Ax(t) + bu(t),\ t = 0, \ldots, n-1 \\ & x(0) = 0 \\ & x(N) = x_{\text{des}}, \end{array}$$

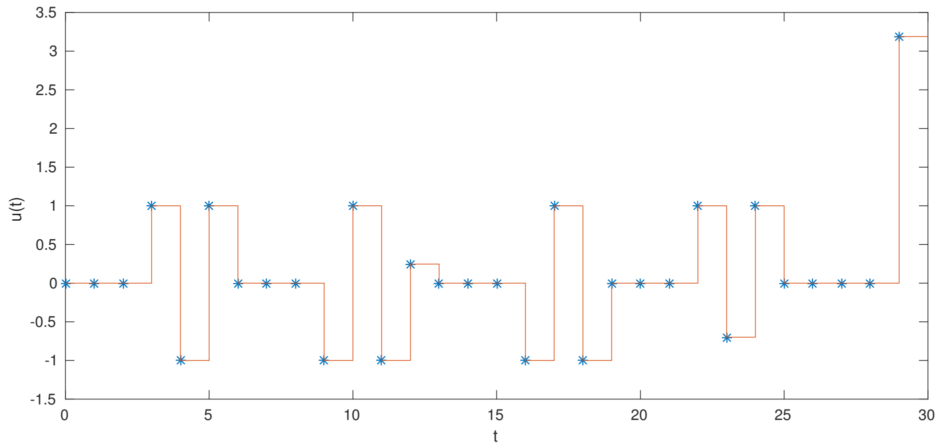
with variables $u(0), \ldots, u(N-1)$, and $x(0), \ldots, x(N)$. The CVX code defines two matrix variables

$$X = \left[\begin{array}{cccc} x(0) & x(1) & \cdots & x(N) \end{array}\right], \qquad u = \left[\begin{array}{cccc} u(0) & u(1) & \cdots & u(N-1) \end{array}\right].$$

```
n = 3;
N = 30;
A = [ -1, 0.4, 0.8; 1, 0, 0; 0, 1, 0];
b = [ 1; 0; 0.3];
x0 = zeros(n,1);
xdes = [ 7; 2; -6];
cvx_begin
    variable X(n,N+1);
    variable u(1,N);
    minimize (sum(max(abs(u), 2*abs(u)-1)))
    subject to
        X(:,2:N+1) == A*X(:,1:N) + b*u;
        X(:,1) == x0;
        X(:,N+1) == xdes;
cvx_end
plot(0:N-1, u, '*');
hold on
stairs(0:N, [u, u(end)]);
axis([0, 30, -1.5, 3.5])
xlabel('t')
```

```
ylabel('u(t)')
hold off
```

The optimal actuator signal is shown in the figure.



7. Exercise A3.21.

Solution.

(a) Using the result mentioned in the problem, we can express a doubly stochastic matrix S as a convex combination $S = \sum_k \theta_k P_k$ of permutation matrices P_k. From convexity and symmetry of f,

$$f(Sx) = f(\sum_k \theta_k P_k x) \leq \sum_k \theta_k f(P_k x) = \sum_k \theta_k f(x) = f(x).$$

The inequality is Jensen's inequality (convexity of f). The second equality follows from the symmetry of f.

(b) The diagonal elements of $Y = Q\,\mathbf{diag}(\lambda)Q^T$ are given by

$$Y_{ii} = \sum_{j=1}^n Q_{ij}^2 \lambda_j.$$

A matrix Q is orthogonal if $QQ^T = Q^TQ = I$. From $QQ^T = I$, we see that $\sum_{j=1}^n Q_{ij}^2 = 1$ for all i. From $Q^TQ = I$, we have $\sum_{i=1}^n Q_{ij}^2 = 1$ for all j. Therefore the matrix with elements $S_{ij} = Q_{ij}^2$ is doubly stochastic.

(c) Combining the results in parts (a) and (b), we conclude that for any symmetric X, the inequality

$$f(\mathbf{diag}(X)) \leq f(\lambda(X))$$

holds. Also, if V is orthogonal, $\lambda(X) = \lambda(V^TXV)$. Therefore

$$f(\mathbf{diag}(V^TXV)) \leq f(\lambda(X))$$

for all orthogonal V. Moreover, this inequality holds with equality if $V = Q$, where Q is the matrix of eigenvectors of X. Hence

$$f(\lambda(X)) = \sup_{V \in \mathcal{V}} f(\mathbf{diag}(V^TXV)).$$

For fixed V, the function $f(\mathbf{diag}(V^T XV))$ is the composition of f with a linear function of X, and therefore convex. We have shown that $f(\lambda(X))$ is the supremum of a family of convex functions of X. Therefore $f(\lambda(X))$ is a convex function of X.

8. Exercise A3.31.

Solution.

(a) Follows from the composition rules.

(b)

$$\begin{aligned} f^*(y) = \sup_x (y^T x - h(\|x\|_2) &= \sup_{t\geq 0} \sup_{\|x\|_2=t} (y^T x - h(t)) \\ &= \sup_{t\geq 0} (t\|y\|_2 - h(t)) \\ &= \sup_{t} (t\|y\|_2 - h(t)) \\ &= h^*(\|y\|_2). \end{aligned}$$

The third line follows because $h(t) = h(0)$ for $t \leq 0$.

(c) The conjugate of $h(t)$ is

$$h^*(s) = \sup_t (st - h(t)) = \begin{cases} (1/q)s^q & s \geq 0 \\ +\infty & s < 0, \end{cases}$$

where $q = p/(p-1)$. To see this, we distinguish the two cases. If $s < 0$, the function $st - h(t)$ goes to infinity for $t \to -\infty$. If $s \geq 0$, it reaches a maximum where

$$s - h'(t) = s - t^{p-1} = 0.$$

Substituting the optimal $t = s^{1/(p-1)}$ in $st - h(t)$ gives $(1/q)s^q$.

9. Exercise T3.23 (a).

Solution. This is the perspective function of the convex function $\|x\|_p^p = |x_1|^p + \cdots + |x_n|^p$.

10. Exercise A3.20 (c).

Solution. By making a change of variables $t = 1/\alpha$, we can write this as

$$f(x) = \inf_{t>0} (t(g(y + x/t) - g(y))).$$

This is the infimum over t of the perspective of the convex function

$$h(x) = g(y + x) - g(y)).$$

The perspective is convex, jointly in x, t. Partial minimization over t results in a convex function of x.