Optimization and Machine Learning, Spring 2020
Homework 6

(Due Tuesday, June 2 at 11:59pm (CST))

1. Which of the following sets are convex?

   (a) A *wedge, i.e.,* $\{x \in \mathbb{R}^n | a_1^T x \leq b_1, a_2^T x \leq b_2\}$. (5 points)
   Solution:
   A wedge is an intersection of two halfspaces, so it is convex set.

   (b) The set of points closer to a given points than a given set, *i.e.,*

   $$\{x | ||x - x_0||_2 \leq ||x - y||_2 \text{ for all } y \in S\}$$

   where $S \subseteq \mathbb{R}^n$. (5 points)
   Solution:
   This set is convex because it can be expressed as

   $$\bigcap_{y \in S} \{x | ||x - x_0||_2 \leq ||x - y||_2\}$$

   , which is an intersection of halfspaces

   (c) The set of points closer to one set than another, *i.e.,*

   $$\{x | \mathbf{dist}(x, S) \leq \mathbf{dist}(x, T)\}$$

   where $S, T \subseteq \mathbb{R}^n$, and

   $$\mathbf{dist}(x, S) = \inf\{||x - z||_2 \in S\}.$$

   (5 points)
   Solution:
   In general this set is not convex, as the following example in R shows. With $S = \{-1, 1\}$ and $T = \{0\}$,we have

   $$\{x | \mathbf{dist}(x, S) \leq \mathbf{dist}(x, T)\} = \{x \in R | x \leq -1/2 \text{ or } x \geq 1/2\}$$

   , which is not convex.

   (d) The set $\{x | x + S_2 \subseteq S_1\}$, where $S_1, S_2 \subseteq \mathbb{R}^n$ with S_1 convex. (5 points)
   Solution:
   This set is convex. $x + S_2 \subseteq S_1$ if $x + y \in S_1$, for all $y \in S_2$. Therefore $\{x | x + S_2 \subseteq S_1\} = \bigcap y \in S_2 \{x | x + y \in S_1\} = \bigcap_{y \in S_2} (S_1 - y)$, which is an intersection of convex sets $S_1 - y$

   (e) The set of multiplication

   $$\{x \in \mathbb{R}_+^n | \prod_{i=1}^n x_i \geq 1\}.$$

   (5 points)
   Solution:
   Assume that $\prod_i x_i \geq 1$ and $\prod_i y_i \geq 1$

   $$\prod_i (\theta x_i + (1 - \theta) y_i) \geq \prod_i x_i^\theta y_i^{(1-\theta)} \geq 1$$

   So this is convex.

2. Determine whether the following functions are convex, strictly convex, concave, strictly concave, both or neither. (5 points)

(a) $f(x) = e^x - 1$ on $\mathbb{R}$. (5 points)
Solution:

$$f''(x) = e^x > 0$$

So it's strictly convex.

(b) $f(x_1, x_2) = x_1 x_2$ on $\mathbb{R}^2_{++}$. (5 points)
Solution:

$$H(f) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

So it's neither.

(c) $f(x) = \log(\sum_{i=1}^n \exp(x_i))$ on $\mathbb{R}^n$, use the second-order condition. (5 points)
Solution:

$$\nabla^2 f(x) = \frac{1}{\mathbf{1}^T z}\mathbf{diag}(z) - \frac{1}{(\mathbf{1}^T z)^2} z z^T$$

to show $\nabla^2 f(x) \succeq 0$, we must verify $v^T \nabla^2 f(x) v \geq 0$ for all v

$$v^T \nabla^2 f(x) v = \frac{(\sum_k z_k v_k^2)(\sum_k z_k) - (\sum_k v_k z_k)^2}{(\sum_k z_k)^2} \geq 0$$

since $(\sum_k v_k z_k)^2 \leq (\sum_k z_k v_k^2)(\sum_k z_k)$ (from Cauchy-Schwarz inequality)

(d) $f(w) = ||Xw - y||_2^2 + \lambda ||w||_2^2$ for $\lambda > 0$. (5 points)
Solution:

$$\nabla f(x) = 2(Xw - y)^T X + 2\lambda w$$
$$\nabla^2 f(x) = 2X^T X + 2\lambda$$

And it's positive definite matrix. So it's strictly convex

(e) The log-likelihood of a set of points $\{x_1, \cdots, x_n\}$ that are normally distributed with mean μ and finite variance $\sigma > 0$ is given by:

$$f(\mu, \sigma) = n \log(\frac{1}{\sqrt{2\pi}\sigma}) - \frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu)^2$$

Show that if we view the log likelihood for fixed σ as a function of the mean, *i.e.*,

$$g(\mu) = n \log(\frac{1}{\sqrt{2\pi}\sigma}) - \frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu)^2$$

then g is strictly concave.
Show that if we view the log likelihood for fixed μ as a function of the mean, *i.e.*,

$$h(z) = n \log(\frac{\sqrt{z}}{\sqrt{2\pi}}) - \frac{z}{2}\sum_{i=1}^n (x_i - \mu)^2$$

then h is strictly concave (equivalently, we say f is strictly concave in $z = \frac{1}{\sigma^2}$).
We say $f(x, y)$ with $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^\kappa$ is jointly convex if

$$f(\lambda(x_1, y_1) + (1 - \lambda)(x_2, y_2)) \leq \lambda f((x_1, y_1)) + (1 - \lambda) f((x_2, y_2)).$$

Show that f is not jointly concave in $\mu, \frac{1}{\sigma^2}$. (5 points)
Solution:

$$\nabla g(\mu) = \frac{1}{2\sigma^2}\sum_{i=1}^n 2(x_i - \mu)$$
$$\nabla^2 g(\mu) = \frac{-n}{\sigma^2} < 0$$

So it's strictly concave

$$\nabla^2 h(z) = -\frac{n}{2z^2} < 0$$

So it's strictly concave

$$\nabla^2 f(\mu, \frac{1}{\sigma^2}) = \begin{bmatrix} \frac{-n}{\sigma^2} & \sum_{i=1}^{n}(x_i - \mu) \\ \sum_{i=1}^{n}(x_i - \mu) & -\frac{n\sigma^4}{2} \end{bmatrix}$$

The determinant of the Hessian is given by,

$$det(\nabla^2 f) = \frac{n^2\sigma^2}{2} - (\sum_{i=1}^{n}(x_i - \mu))^2$$

and the trace of the Hessian is given by,

$$tr(\nabla^2 f) = -\frac{n}{\sigma^2} - \frac{n\sigma^4}{2} < 0$$

Note that the trace is the sum of the eigenvalues, and the determinant is the product of the eigenvalues. Since the trace is always negative, if the determinant is negative it must imply that one eigenvalue is positive and another is negative; that is, we have f is neither convex nor concave. It is easy to see that $det(\nabla^2 f)$ can sometimes be negative – for example, if we choose σ^2 to be close to zero and μ away from x_i , the second negative term dominates and make $det(\nabla^2 f) \leq 0$.

3. Consider the problem

$$\begin{array}{ll} \text{minimize} & \|Ax - b\|_1 / \left(c^T x + d\right) \\ \text{subject to} & \|x\|_\infty \leq 1 \end{array}$$

where $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m, c \in \mathbb{R}^n$ and $d \in \mathbb{R}$. We assume that $d > \|c\|_1$, which implies that $c^{\mathrm{T}}x + d > 0$ for all feasible x.

(a) Show that this is a quasiconvex optimization problem. (5 points)

(b) Show that it is equivalent to the convex optimization problem

$$\begin{array}{ll} \text{minimize} & \|Ay - bt\|_1 \\ \text{subject to} & \|y\|_\infty \leq t \\ & c^T y + dt = 1 \end{array}$$

with variables $y \in \mathbb{R}^n, t \in \mathbb{R}$. (10 points)

Solution:

(a) $f_0(x) \leq \alpha$ if and only if

$$\|Ax - b\|_1 - \alpha(c^T x + d) \leq 0,$$

which is a convex constraint.

(b) Suppose $\|x\|_\infty \leq 1$. We have $c^T x + d > 0$, because $d > \|c\|_1$. Define

$$y = x/(c^T x + d), \quad t = 1/(c^T x + d).$$

Then y and t are feasible in the convex problem with objective value

$$\|Ay - bt\|_1 = \|Ax - b\|_1/(c^T x + d).$$

Conversely, suppose y, t are feasible for the convex problem. We must have $t > 0$, since $t = 0$ would imply $y = 0$, which contradicts $c^T y + dt = 1$. Define

$$x = y/t.$$

Then $\|x\|_\infty \leq 1$, and $c^T x + d = 1/t$, and hence

$$\|Ax - b\|_1/(c^T x + d) = \|Ay - bt\|_1.$$

4. Consider the QCQP

$$\begin{array}{ll} \text{minimize} & (1/2)x^TPx + q^Tx + r \\ \text{subject to} & x^Tx \leq 1, \end{array}$$

with $P \in \mathbf{S}^n_{++}$. Show that $x^\star = -(P + \lambda I)^{-1}q$ where $\lambda = \max\{0, \bar{\lambda}\}$ and $\bar{\lambda}$ is the largest solution of the nonlinear equation

$$q^{\mathrm{T}}(P + \lambda I)^{-2}q = 1.$$

(15 points)

Solution:

x is optimal if and only if

$$x^Tx < 1, \quad Px + q = 0$$

or

$$x^x = 1, \quad Px + q = -\lambda x$$

for some $\lambda \geq 0$. (Geometrically, either x is in the interior of the ball and the gradient vanishes, or x is on the boundary, and the negative gradient is parallel to the outward pointing normal.)

The algorithm goes as follows. First solve $Px = -q$. If the solution has norm less than or equal to one ($\|P^{-1}q\|_2 \leq 1$), it is optimal. Otherwise, from the optimally conditions, x must satisfy $\|x\|_2 = 1$ and $(P + \lambda)x = -q$ for some $\lambda \geq 0$. Define

$$f(\lambda) = \left\|(P + \lambda)^{-1}q\right\|_2^2 = \sum_{i=1}^n \frac{q_i^2}{(\lambda + \lambda_i)^2},$$

where $\lambda_i > 0$ are the eigenvalues of P. (Note that $P + \lambda I \succ 0$ for all $\lambda \geq 0$ because $P \succ 0$.) We have $f(0) = \|P^{-1}q\|_2^2 > 1$. Also f monotonically decreases to zero as $\lambda \to \infty$. Therefore the nonlinear equation $f(\lambda) = 1$ has exactly one nonnegative solution $\bar{\lambda}$. Solve for $\bar{\lambda}$. The optimal solution is $x^\star = -(P + \bar{\lambda}I)^{-1}q$.

5. Consider the inequality form LP

$$\begin{array}{ll} \min\limits_x & c^Tx \\ s.t. & Ax \preceq b, \end{array}$$

with $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$. Let $w \in \mathbb{R}^m_+$. If x is feasible for the LP, *i.e.*, satisfies $Ax \preceq b$, then it also satisfies the inequality

$$w^TAx \leq w^Tb.$$

Geometrically, for any $w \succeq 0$, the halfspace $H_w = \{x \mid w^TAx \leq w^Tb\}$ contains the feasible set for the LP. Therefore if we minimize the objective c^Tx over the halfspace Hw we get a lower bound on p^*.

(a) Derive an expression for the minimum value of c^Tx over the halfspace H_w (which will depend on the choice of $w \succeq 0$). (5 points)

(b) Formulate the problem of finding the best such bound, by maximizing the lower bound over $w \succeq 0$. (5 points)

(c) Relate the results of (a) and (b) to the Lagrange dual of the LP. (10 points)

Solution:

(a) The optimal value is

$$\inf_{x\in H_w} c^Tx = \begin{cases} \lambda w^Tb & c = \lambda A^Tw \text{ for some } \lambda \leq 0 \\ -\infty & \text{otherwise.} \end{cases}$$

(b) We maximize the lower bound by solving

$$\begin{array}{ll} \max\limits_{\lambda,w} & \lambda w^Tb \\ s.t. & c = \lambda A^Tw \\ & \lambda \leq 0, \ w \succeq 0 \end{array}$$

with variables λ and w. Note that, as posed, this is not a convex problem.

(c) Defining $z = -\lambda w$, we obtain the equivalent problem

$$\begin{aligned} \max_{z} \quad & -b^T z \\ s.t. \quad & A^T z + c = 0 \\ & z \succeq 0. \end{aligned}$$

This is the dual of the original LP.