Homework Assignments for ORIE 6300: Mathematical Programming I

Damek Davis[*]

[*]School of Operations Research and Information Engineering, Cornell University, Ithaca, NY 14850, USA; `people.orie.cornell.edu/dsd95/`.

1 Homework 1

1. Prove the following basic consequences of convexity:

   (a) The set of optimal solutions to a convex program is convex.
   (b) Intersections of convex sets are convex.
   (c) Cartesian products of convex sets are convex.
   (d) If $\mathcal{X}_1$ and $\mathcal{X}_2$ are convex, then so is $\mathcal{X}_1 + \mathcal{X}_2 = \{x_1 + x_2 \colon x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2\}$.
   (e) If $\mathcal{X} \subseteq \mathbb{R}^d$ is a convex set and A is a matrix, then $\{Ax \colon x \in \mathcal{X}\}$ is convex.
   (f) If $\mathcal{Y} \subseteq \mathbb{R}^m$ is a convex set and A is a matrix, then $\{x \in \mathbb{R}^d \colon Ax \in \mathcal{Y}\}$ is convex.
   (g) The set $\{Ax \colon x \in \mathcal{X}\}$ is not necessarily closed, even when $\mathcal{X}$ is closed.
   (h) A convex set $\mathcal{X} \subseteq \mathbb{R}^d$ has a convex closure.
   (i) Let $\mathcal{X}$ be a closed convex set and let $x \in \mathcal{X}$. Show that $\mathcal{N}_{\mathcal{X}}(x)$ is a closed convex cone, meaning $\mathcal{N}_{\mathcal{X}}(x)$ is closed and convex and for all $v \in \mathcal{N}_{\mathcal{X}}(x)$ and $t \geq 0$, the inclusion $tv \in \mathcal{N}_{\mathcal{X}}(x)$ holds.

2. Consider the ℓ_1 ball:

$$\mathcal{X} := \left\{ x \in \mathbb{R}^d \colon \sum_{i=1}^{d} |x_i| \leq 1 \right\}.$$

   (a) Prove that $\mathcal{X}$ is a *polyhedron* (i.e., the intersection of finitely many linear inequalities, meaning $\mathcal{X} = \{x \in \mathbb{R}^d \colon a_i^T x \leq b_i \text{ for } i = 1, \ldots, n\}$ for a set of vectors a_i and scalars b_i). How many inequalities are needed to describe $\mathcal{X}$ (how large is n)?

   (b) A *lifting* of a polyhedron $\mathcal{P}_1 \subseteq \mathbb{R}^d$ is a description of the form $\mathcal{P}_1 = \{Ax \colon x \in \mathcal{P}_2\}$ where $\mathcal{P}_2 \subseteq \mathbb{R}^m$ is a polyhedron and $A \in \mathbb{R}^{d \times m}$ is a matrix.

   Find a lifting of $\mathcal{X}$ to $\mathbb{R}^{2d}$, where the associate polyhedron in $\mathbb{R}^{2d}$ is defined by at most $2d + 1$ inequalities.

3. Calculate the normal cones of the following sets:

   (a) $\mathcal{X}$ = a subspace of $\mathbb{R}^d$.
   (b) $\mathcal{X} = B_1(0)$ (closed unit ball in $\mathbb{R}^d$)
   (c) $\mathcal{X} = \mathbb{R}^d_+ = \{x \in \mathbb{R}^d \colon x_i \geq 0 \text{ for } i = 1, \ldots, d\}$.
   (d) $\mathcal{X} = \{x \in \mathbb{R}^d \colon Ax = b\}$ where $b \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m \times d}$ is a matrix.

4. Let $f \colon \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ be a convex function. Prove that any local minimum of f is a global minimum.

5. (**Weierstrass**) Let $f \colon \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ be a function that has a closed epigraph and bounded sublevel sets. Show that f has a minimizer. (Hint: consider the epigraphical form from Section 2.2.1 of the course lecture notes.)

6. (**The Rayleigh Quotient**; see Exercise 6 of Chapter 2.1 in Borwein and Lewis.)

   (a) Let $f \colon \mathbb{R}^d \backslash \{0\} \to \mathbb{R} \cup \{+\infty\}$ be continuous, satisfying $f(\lambda x) = f(x)$ for all $\lambda > 0$ in $\mathbb{R}$ and nonzero x in $\mathbb{R}^d$. Prove f has a minimizer.

(b) Given a symmetric matrix $A \in \mathbb{R}^{d \times d}$, define a function $g(x) = x^T Ax/\|x\|^2$ for nonzero $x \in \mathbb{R}^d$. Prove that g has a minimizer.

(c) Calculate $\nabla g(x)$ for nonzero x.

(d) Deduce that minimizers of g must be eigenvectors, and calculate the minimum value.

2 Homework 2

Your homework partly relies on the following definition:

Definition 2.1 (Dual Cone). *Let $\mathcal{K} \subseteq \mathbb{R}^d$ be a cone. Then the* dual cone *of $\mathcal{K}$ is the set*

$$\mathcal{K}^* := \{s \in \mathbb{R}^d \colon \langle x, s\rangle \geq 0 \quad \forall x \in \mathcal{K}\}.$$

Please complete the following exercises.

1. Prove the following:
   (a) The closure of any cone must contain the origin.
   (b) The intersection of two cones is a cone.
   (c) The Cartesian product of two cones is a cone.
   (d) If $\mathcal{K}_1, \mathcal{K}_2 \subseteq \mathbb{R}^d$ are cones, then $\mathcal{K}_1 + \mathcal{K}_2$ is a cone.
   (e) A cone $\mathcal{K} \subseteq \mathbb{R}^d$ is convex if and only if $\mathcal{K} + \mathcal{K} = \mathcal{K}$.

   Suppose $A \in \mathbb{R}^{m\times d}$ is a matrix.

   (f) If $\mathcal{K} \subseteq \mathbb{R}^d$ is a cone, then $\{Ax \colon x \in \mathcal{K}\}$ is a cone in $\mathbb{R}^m$.
   (g) If $\mathcal{K}' \subseteq \mathbb{R}^m$ is a cone, then $\{x \colon Ax \in \mathcal{K}'\}$ is a cone in $\mathbb{R}^d$.
   (h) Give an example of a closed convex cone $\mathcal{K} \subseteq \mathbb{R}^d$ and a matrix $A \in \mathbb{R}^{m\times d}$ such that the set $\{Ax \colon x \in \mathcal{K}\}$ is not closed.

2. (a) Suppose $\mathcal{X}$ is a closed convex set. Prove that

   $$\mathcal{K}_{\mathcal{X}} = \{(x, t) \colon t > 0 \text{ and } x/t \in \mathcal{X}\}$$

   is a convex cone.

   (b) If $\mathcal{X}$ is bounded, show that $\overline{\mathcal{K}}_{\mathcal{X}} = \mathcal{K}_{\mathcal{X}} \cup \{(0, 0)\}$.
   (c) Give an example of a closed convex set $\mathcal{X}$ for which $\overline{\mathcal{K}}_{\mathcal{X}} \neq \mathcal{K}_{\mathcal{X}} \cup \{(0, 0)\}$.

3. Let $\mathcal{K}$ be a polyhedral cone.[1] Prove that $\mathcal{K}^*$ is also polyhedral.

4. Prove that each of the following cones $\mathcal{K}$ are *self-dual*, meaning $\mathcal{K} = \mathcal{K}^*$.
   (a) $\mathbb{R}^d_+$
   (b) $\mathrm{SOC}(d + 1)$
   (c) $\mathbb{S}^{d\times d}_+$

5. Let $\mathcal{X} \subseteq \mathbb{R}^d$ be a closed convex set. For any $x \in \mathcal{X}$, define the *proximal normal cone*

   $$\mathcal{N}^P_{\mathcal{X}}(x) = \left\{v \in \mathbb{R}^d \colon x = \mathrm{proj}_{\mathcal{X}}(x + v)\right\}.$$

   Prove that $\mathcal{N}_{\mathcal{X}}(x) = \mathcal{N}^P_{\mathcal{X}}(x)$.

[1] The term polyhedral means the cone is defined by finitely many linear inequalities.

3 Homework 3

1. **(Normal Cone to A Cone.)** Let $\mathcal{K} \subseteq \mathbb{R}^m$ be a convex cone. Prove that

$$\mathcal{N}_{\mathcal{K}}(x) = -\mathcal{K}^* \cap \{x\}^\perp \qquad \forall x \in \mathcal{K}.$$

2. **(A Compressive Sensing Problem.)** Consider the following optimization problem

$$\begin{aligned} &\text{minimize } \|x\|_1 \\ &\text{subject to: } Ax = b. \end{aligned}$$

(The symbol $\|x\|_1$ denotes the ℓ_1 *norm* on $\mathbb{R}^d$, a particular member of the the family of ℓ_p norms defined as follows: for any $p \in [1, \infty)$, we define

$$\|x\|_p^p := \sum_{i=1}^d |x_i|^p \qquad \forall x \in \mathbb{R}^d.$$

If $p = \infty$, we define $\|x\|_\infty := \max_{i=1,\ldots,d} |x_i|$ for all $x \in \mathbb{R}^d$.)

   (a) Write an equivalent linear programming formulation of this problem.
   (b) Take the dual of the linear program from part 2a.
   (c) Prove that the linear program from part 2b is equivalent to the following problem

$$\begin{aligned} &\text{maximize } \langle y, b\rangle \\ &\text{subject to: } \|A^T y\|_\infty \leq 1. \end{aligned}$$

3. **(Failure Cases.)**
   (a) Give an example of a linear program where `val` $= +\infty$ and `val`$^* = -\infty$.
   (b) Give an example of a conic program where `val` is finite but not attained.
   (c) Give an example of a conic program where `val` $= +\infty$, but `val`* is finite.
   (d) Give an example of a conic program where `val`, `val`$^* \in \mathbb{R}$ and `val` $\neq$ `val`*.

4. **(Closed Functions.)** Let $f : \mathbb{R}^d \to [-\infty, +\infty]$ be an extended valued function.
   (a) Prove there exists a unique function $\operatorname{cl} f : \mathbb{R}^d \to [-\infty, +\infty]$, called the *closure* of f, satisfying

$$\operatorname{epi}(\operatorname{cl} f) = \overline{\operatorname{epi}(f)}.$$

   Moreover, prove the closure satisfies the following limiting formula:

$$\operatorname{cl} f(x) = \lim_{\varepsilon \to 0} \inf_{y \in B_\varepsilon(x)} f(y). \tag{3.1}$$

   (b) Suppose f is convex. Prove that $\operatorname{cl} f$ is convex.

Def. An extended-valued function is *closed* if $\operatorname{epi}(f)$ is closed.

   (c) Prove that $\operatorname{cl} f$ is closed.
   (d) Prove that $\operatorname{cl} f(x) \leq f(x)$ for all $x \in \mathbb{R}^d$.

(e) Suppose f is continuous. Prove that f closed.

(f) Suppose that f is continuous at a point $x \in \mathbb{R}^d$. Prove that $f(x) = \operatorname{cl} f(x)$. (In other words,

$$f(x) = \lim_{\varepsilon \to 0} \inf_{y \in B_\varepsilon(x)} f(y).)$$

(g) Suppose that for all $x \in \mathbb{R}^d$, we have

$$f(x) = \lim_{\varepsilon \to 0} \inf_{y \in B_\varepsilon(x)} f(y).$$

Prove that f is closed. (Such functions are called lower semicontinuous.)

(h) Give an example of a closed extended valued function such that $\operatorname{dom}(f) = \{x \colon f(x) < +\infty\}$ is open.

5. **(Strong Duality.)** Let $A \in \mathbb{R}^{m \times d}$, let $c \in \mathbb{R}^d$, and let $\mathcal{K} \subseteq \mathbb{R}^d$ be a closed convex cone. Consider the family of primal and dual conic problems, which both depend on a parameter $b \in \mathbb{R}^m$:

$$\underbrace{\left\{ \begin{array}{r} \text{minimize } c^T x \\ \text{subject to: } Ax = b \\ x \in \mathcal{K} \end{array} \right\}}_{\mathcal{P}(b)} \qquad \underbrace{\left\{ \begin{array}{c} \text{maximize } b^T x \\ \text{subject to: } c - A^T y \in \mathcal{K}^* \\ \end{array} \right\}}_{\mathcal{D}(b)} \tag{3.2}$$

Recall the value function $\mathtt{val} \colon \mathbb{R}^m \to [-\infty, \infty]$

$$\mathtt{val}(b) = \inf\{c^T x \colon Ax = b, x \in \mathcal{K}\} \qquad \forall b \in \mathbb{R}^m,$$

and the asymptotic value function $\mathtt{a\text{-}val} \colon \mathbb{R}^m \to [-\infty, \infty]$

$$\mathtt{a\text{-}val} = \operatorname{cl} \mathtt{val}.$$

(a) Suppose there is a point $b \in \mathbb{R}^m$ such that $\mathtt{val}(b) = \mathtt{a\text{-}val}(b) \in \mathbb{R}$. Prove that $\mathtt{val}(b') > -\infty$ for all $b' \in \mathbb{R}^m$.

(b) Give an example of a conic program and a vector b such that the $\mathtt{val}(b) = +\infty$ and $\mathtt{a\text{-}val}(b) < +\infty$.

(c) Suppose that $\mathtt{val}$ is continuous at a point $b \in \mathbb{R}^m$. Prove that strong duality holds:

$$\mathtt{val}(b) = \sup\{b^T y \colon c - A^T y \in \mathcal{K}^*\}.$$

(d) Prove that $\mathtt{val}$ is convex. Is $\mathtt{a\text{-}val}$ convex?

Consider the following basic property of convex functions:

Theorem 3.1 (Borwein and Lewis Theorem 4.1.3)**.** *Let $f \colon \mathbb{R}^d \to (-\infty, +\infty]$ be a convex function. Then f is continuous on the interior of its domain.*[2]

Notice that the function f in the above theorem never takes value $-\infty$.

[2]Recall that $\operatorname{dom}(f) = \{x \colon f(x) < +\infty\}$.

(e) We say that $\mathcal{P}(b)$ is *strongly feasible* if there exists an $\varepsilon > 0$ such that for all $b' \in B_\varepsilon(b)$ the perturbed problem $\mathcal{P}(b')$ is feasible.

Suppose that $\mathcal{P}(b)$ is strongly feasible. Then show that strong duality holds:

$$\mathtt{val}(b) = \sup\{b^T y \colon c - A^T y \in \mathcal{K}^*\}.$$

(f) **(Slater's Condition.)** If $\mathcal{P}(b)$ has a feasible point x lying in the interior of $\mathcal{K}$ and if $\operatorname{rank}(A) = m$, prove that strong duality holds:

$$\mathtt{val}(b) = \sup\{b^T y \colon c - A^T y \in \mathcal{K}^*\}.$$

(Note that it is very common to assume $\operatorname{rank}(A) = m$ in the optimization literature, and we can assume this without loss of generality. Indeed, if A is not full rank, we can simply row reduce the system and form a new problem with the row reduced matrix. Clearly, any solution to the new problem still solves the original one.)

4 Homework 4

1. **(Extreme Points.)**

   (a) Give an example of a polyhedron with no extreme points.

   (b) Prove that any nonempty polyhedron in standard form $P = \{x \colon Ax = b, x \geq 0\}$ has at least one extreme point.

2. **(Polyhedral Functions.)** We call a function $f : \mathbb{R}^d \to (-\infty, \infty]$ *polyhedral* if $\mathrm{epi}(f)$ is polyhedral. Prove that any polyhedral function f admits the representation:

$$f(x) = \max_{i=1,\ldots,n} \{a_i^T x + b_i\} + \delta_{\mathcal{X}}(x), \qquad \forall x \in \mathbb{R}^d$$

   where $n \geq 0$, $\mathcal{X} \subseteq \mathbb{R}^d$ is a polyhedral set, and for $i = 1, \ldots, n$, we have $a_i \in \mathbb{R}^d$ and $b_i \in \mathbb{R}$. (Hint: write $f(x) = \inf\{t \colon (x, t) \in \mathrm{epi}(f)\}$.) Does the value function of a polyhedral program admit such a representation? Justify your answer.

3. **(Strict Complementary Slackness.)** In this exercise, we examine the strict complementary slackness condition. To that end consider the following primal-dual pair of linear programs:

$$\begin{aligned} &\text{minimize } c^T x & &\text{maximize } b^T x \\ &\text{subject to: } Ax = b & &\text{subject to: } A^T y + s - c = 0 \\ &x \in \mathbb{R}_+ & &s \geq 0 \end{aligned} \tag{4.1}$$

   Throughout this exercise, we suppose that optimal solutions exist. Consider the following condition.

   Condition. Suppose that there is some $j \in \{1, \ldots, d\}$ so that every optimal solution x^* satisfies $x_j^* = 0$.

   In the next three parts, suppose the above condition holds. Under this condition, we will prove there is a dual optimal pair (y, s) with $s_j > 0$.

   (a) Consider the following linear program:

$$\begin{aligned} \text{minimize } & -x_j \\ \text{subject to: } & Ax = b \\ & c^T x \leq \texttt{val} \\ & x \geq 0. \end{aligned}$$

   Show that its dual is

$$\begin{aligned} \text{maximize } & b^T y - t\texttt{val} \\ \text{subject to: } & A^T y - tc + s = -e_j \\ & s, t \geq 0, \end{aligned}$$

where e_j denotes the jth standard basis vector. Prove that this dual has an optimal solution $(\bar{y}, \bar{t}, \bar{s})$ and show that $b^T\bar{y} = \bar{t}\texttt{val}$.

(b) Suppose $\bar{t} > 0$ and let $y = \bar{y}/t$ and $s = (\bar{s} + e_j)/\bar{t}$. Prove that $s_j > 0$ (obvious) and (y, s) solves the original dual problem.

(c) Suppose that $\bar{t} = 0$. Find an optimal solution (y, s) to the original dual problem with $s_j > 0$.

Using the above results, we can construct a primal-dual pair satisfying the strict complementary slackness condition. To that end, define a subset of indices $J \subseteq \{1, \ldots, d\}$ by the following formula

$$J := \{j : \exists \text{ primal optimal } x \text{ with } x_j > 0\}.$$

Using J, we will construct a sequence $(x^1, y^1), \ldots, (x^d, y^d)$ of primal-dual optimal pairs with the following properties: For each $j \in J$, we let y^j be an arbitrary dual optimal solution and let x^j be a primal optimal solution with $x^j_j > 0$. On the other hand, for each $j \notin J$, we let x^j be an arbitrary primal optimal solution and let y^j be a dual optimal optimal solution with $(c - A^T y^j)_j > 0$ (exists by Parts 1-3). Given these primal-dual optimal pairs, define

$$x^* := \frac{1}{d}\sum_{j=1}^{d} x^j \qquad \text{and} \qquad y^* := \frac{1}{d}\sum_{j=1}^{d} y^j.$$

(d) **Bonus.** Show that the pair (x^*, y^*) is primal-dual optimal and in addition satisfies strict complementary slackness, namely,

$$x^*_j > 0 \text{ if and only if } (c - A^T y^*)_j = 0, \qquad \forall j$$

4. (**A Closed Value Function.**) Prove that $\texttt{val} : \mathbb{R}^d \to (-\infty, +\infty]$ is closed if for every $\gamma, \tau \in \mathbb{R}$, the set
$$\{x : c^T x \leq \gamma, \|Ax\| \leq \tau, x \in \mathcal{K}\}$$
is bounded. Under this condition, prove that whenever $\texttt{val}(b)$ is finite, strong duality holds ($\texttt{val} = \texttt{val}^*$) and there exists a primal optimal solution.

5. Prove the following Lemma:

Lemma 4.1 (Fréchet Subgradients of Convex Functions)**.** *Let $f : \mathbb{R}^d \to (-\infty, \infty]$ be a convex function. Then*

$$\partial_F f(x) = \left\{v : f(y) \geq f(x) + \langle v, y - x\rangle, \quad \forall y \in \mathbb{R}^d\right\}, \qquad \forall x \in \mathrm{dom}(f).$$

Equivalently, $v \in \partial_F f(x)$ if and only if $f(y) - \langle v, y\rangle$ is minimized at x.

6. **(Fermat's Rule.)** Prove the following Theorem

Theorem 4.2 (Fermat's Rule). *Let $f : \mathbb{R}^d \to (-\infty, \infty]$ be a proper function and suppose that $\bar{x}$ is a local minimizer of f. Then*

$$0 \in \partial_F f(\bar{x}).$$

If moreover f is convex, the condition $0 \in \partial f(x)$ is both necessary and sufficient for x to be a global minimum.

7. **(Mean Value Theorem.)** Suppose $f : \mathbb{R}^d \to \mathbb{R}$ is a closed convex function and let $x, y \in \mathbb{R}^d$. Show that there exists $t \in [0, 1]$ such that

$$f(x) - f(y) \in \langle x - y, \partial f((1-t)x + ty) \rangle$$

(Hint: consider the convex function $t \mapsto f((1-t)x+ty)+t(f(x)-f(y))$ on the compact interval $[0, 1]$.)

The next exercise relies on the following definition.

Definition 4.3 (Lipschitz Continuity). *A function $f : \mathbb{R}^d \to \mathbb{R}$ is called* Lipschitz continuous *if*

$$|f(x) - f(y)| \leq L\|x - y\|, \qquad \forall x, y \in \mathbb{R}^d.$$

for some $L > 0$. The constant L is called a Lipschitz constant *of f.*

8. (**Lipschitz Continuity.**) Let $f : \mathbb{R}^d \to \mathbb{R}$ be a closed convex function. Show that f is Lipschitz continuous with Lipschitz constant L if and only if for all $x \in \mathbb{R}^d$, it holds

$$v \in \partial f(x) \implies \|v\| \leq L$$

5 Homework 5

1. Consider the simplex method applied to a standard form problem. Assume that the rows of the matrix A are linearly independent. Prove or disprove the following.

   (a) A variable that just left the basis cannot reenter in the very next iteration (under any choice of pivoting rule).

   (b) A variable that just entered the basis cannot leave in the very next iteration (under any choice of pivoting rule).

   (c) If there is a nondegenerate optimal solution, then there exists a unique optimal basis.

   (d) If x is an optimal solution, no more than m of its components can be positive, where m is the number of equality constraints.

2. Consider a polyhedron in standard form $\{x\colon Ax = b, x \geq 0\}$ and let x, y be two different basic feasible solutions. If we are allowed to move from any basic feasible solution to an adjacent one in a single step, show that we can go from x to y in a finite number of steps.

3. **Convex Hulls.** Let $\mathcal{X} \subseteq \mathbb{R}^d$. We define the convex hull to be the smallest convex set containing $\mathcal{X}$ and denote this set by $\operatorname{conv}(\mathcal{X})$. Here, the word "smallest" means that whenever a convex set $\mathcal{Y} \subseteq \mathbb{R}^d$ contains $\mathcal{X}$, it must be the case that $\mathcal{Y}$ contains $\operatorname{conv}(\mathcal{X})$ as well. Prove that

$$\operatorname{conv}(\mathcal{X}) = \left\{ x \in \mathbb{R}^d \colon x = \sum_{i=1}^{n_x} \alpha_i x_i \text{ for some } n_x > 0,\ x_i \in \mathcal{X}, \text{ and } \alpha_i \in [0,1] \text{ with } \sum_{i=1}^{n_x} \alpha_i = 1 \right\}.$$

4. **Easy Subdifferential Facts.**

   (a) Let $f\colon \mathbb{R}^d \to (-\infty, \infty]$ be a closed, proper, convex function. Show that for all $x \in \operatorname{dom}(f)$, the set $\partial f(x)$ is closed and convex.

   (b) Let $d = d_1 + \ldots + d_n$ for integers d_i and let $f_i\colon \mathbb{R}^{d_i} \to (-\infty, +\infty]$ be proper convex functions. Then

$$\partial(f_1 + \ldots + f_n)(x_1, \ldots, x_n) = \partial f_1(x_1) \times \ldots \times \partial f_n(x_n) \qquad \forall x_i \in \operatorname{dom}(f_i)$$

   (c) Let $f\colon \mathbb{R}^d \to (-\infty, \infty]$ be a closed, proper, convex function and let $\lambda > 0$. Then prove that the function $g = \lambda f$ is satisfies

$$\partial g(x) = \lambda \partial f(x), \qquad \forall x \in \operatorname{dom}(f).$$

   (d) Let $f\colon \mathbb{R}^d \to (-\infty, \infty]$ be a closed, proper, convex function and let $b \in \mathbb{R}^d$. Then prove that the shifted function $g(\cdot) = f((\cdot) + b)$ satisfies

$$\partial g(x) = \partial f(x + b), \qquad \forall x \in \operatorname{dom}(f) - \{b\}.$$

5. Compute the subdifferentials of the following functions on $\mathbb{R}^d$ (some are differentiable, others are easy applications of the chain rule/the easy from subdifferential facts in Exercise 4):

   (a) ℓ_1 **norm.** $f(x) = \|x\|_1 = \sum_{i=1}^d |x_i|$.
   (b) **Hinge loss.** $f(x) = \max\{0, x\}$ (where $d = 1$).
   (c) **Hybrid Norm.** $f(x) = \sqrt{1 + x^2}$ (where $d = 1$).
   (d) **Logistic function.** $f(x) = \log(1 + \exp(x))$ (where $d = 1$).
   (e) **Indicator of ℓ_p ball.** $f(x) = \delta_{\mathcal{X}}(x)$ where for $p \in [1, \infty]$ and $\tau > 0$, we have $\mathcal{X} = \{x \colon \|x\|_p \leq \tau\}$.
   (f) **Max of coordinates.** $f(x) = \max\{x_1, \ldots, x_d\}$.
   (g) **Polyhedral Function.** $f(x) = \max_{i \leq m}\{\langle a_i, x\rangle + b_i\}$ where $a_1, \ldots, a_m \in \mathbb{R}^d$ are vectors and $b_1, \ldots, b_m \in \mathbb{R}$
   (h) **Quadratic.** $f(x) = \frac{1}{2}\langle Ax, x\rangle$ for some symmetric matrix $A \in \mathbb{R}^{d\times d}$.
   (i) **Least Squares.** $f(x) = \frac{1}{2}\|Ax - b\|_2^2$ where $A \in \mathbb{R}^{m\times d}$ and $b \in \mathbb{R}^m$.
   (j) **Least Absolute Deviations.** $f(x) = \|Ax - b\|_1$ where $A \in \mathbb{R}^{m\times d}$ and $b \in \mathbb{R}^m$.

6. **Descent Directions**

   (a) Suppose that f is Fréchet differentiable on $\mathbb{R}^d$ and that $\nabla f(x)$ is a continuous function of x. Show that for all $x \in \mathbb{R}^d$ with $\nabla f(x) \neq 0$, there exists $\gamma > 0$ such that
   $$f(x - \gamma\nabla f(x)) < f(x).$$
   (**Hint:** Consider the derivative of the one variable function $g(\gamma) = f(x-\gamma\nabla f(x))$.)

   (b) Consider a convex function $f(x, y) = a|x| + b|y|$ for scalars $a, b > 0$. Find a point $(x_0, y_0) \in \mathbb{R}^2$, coefficients $a, b > 0$, and a subgradient $v \in \partial f(x, y)$ so that
   $$f((x_0, y_0) - \gamma v) > f(x_0, y_0) \qquad \forall\gamma > 0.$$

   (c) Let f be a continuous convex function. Let $x \in \mathbb{R}^d$ and suppose that $0 \notin \partial f(x)$. In this exercise, we will show that the minimal norm subgradient of f at x
   $$v := \mathrm{proj}_{\partial f(x)}(0).$$
   is a descent direction.

      i. Show that
      $$\langle w, -v\rangle \leq -\|v\|^2 \qquad \forall w \in \partial f(x).$$
      ii. Next, define the one variable continuous convex function $g(\gamma) = f(x - \gamma v)$. Show that
      $$\eta \in \partial g(0) \implies \eta < -\|v\|^2.$$
      Can 0 be a minimizer of g?
      iii. Show that for all $\gamma < 0$, we have $g(\gamma) > g(0)$.
      iv. Use parts (b) and (c) to show that for $g(\gamma) < g(0)$ for all sufficiently small $\gamma > 0$.

7. Prove the following propositions.

   (a) **Clipped/Bundle Models.** Let $x \in \mathbb{R}^d$ and suppose that f_x is an (l, q) model of f at x. Moreover, assume that $g\colon \mathbb{R}^d \to (-\infty, \infty]$ is closed, proper, convex, and dominated by f: $g(y) \leq f(y)$ for all $y \in \mathbb{R}^d$. Then

   $$\max\{f_x, g\}$$

   is an (l, q)-model of f at x.

   (b) **Projected/Proximal Models.** Suppose that f admits the decomposition

   $$f = g + h,$$

   where $g, h : \mathbb{R}^d \to (-\infty, \infty]$ are closed, proper, convex functions. Let $x \in \mathbb{R}^d$ and suppose that g_x is an (l, q) model of g at x. Then

   $$g_x + h$$

   is an (l, q)-model of f at x.

   (c) **Max-Linear Models.** Suppose that f admits the decomposition

   $$f = \max(f_1, \ldots, f_n),$$

   where for each i, the function $f_i : \mathbb{R}^d \to (-\infty, \infty]$ is closed, proper, and convex. Let $x \in \mathbb{R}^d$ and suppose for each i, the function $(f_i)_x$ is an (l, q) model of f_i at x. Then

   $$\max\{(f_1)_x, \ldots, (f_n)_x\}$$

   is an (l, q)-model of f at x.

8. **Clipping subproblem.** ☕☕ Let $a, x \in \mathbb{R}^d$, let $\texttt{lb} \in \mathbb{R}$, let $\rho > 0$, and let $b \in \mathbb{R}$. Prove that the point

   $$x_+ = \operatorname*{argmin}_{y \in \mathbb{R}^d} \left\{ \max\{\langle a, y\rangle + b, \texttt{lb}\} + \frac{\rho}{2}\|y - x\|^2 \right\}$$

   satisfies

   $$x_+ = x - \texttt{clip}\left(\frac{\rho}{\|a\|^2}(\langle a, x\rangle + b - \texttt{lb})\right)\frac{a}{\rho} \qquad \text{where} \qquad \texttt{clip}(t) = \max\{\min\{t, 1\}, 0\}.$$

   (**Hint:** use first order optimality conditions.)

6 Homework 6

In this homework we study the core algorithmic subproblem in *proximal algorithms.* For motivation recall the *proximal subgradient method* from lecture. This is perhaps the most common algorithm one encounters in first-order methods, so you should at least have a working knowledge of how to implement its steps, when possible. In general it can be quite hard to implement these steps. Indeed, the subproblem includes as a special case the projection of a vector onto a convex set, a generally difficult task. Still for a few useful functions we can implement these steps, even with simple closed form expressions.

1. Let $f\colon \mathbb{R}^d \to (-\infty, \infty]$ be a closed, proper, convex function. Let $\gamma > 0$ and define the *proximal operator* $\mathrm{prox}_{\gamma f}\colon \mathbb{R}^d \to \mathbb{R}^d$:

$$\mathrm{prox}_{\gamma f}(x) = \operatorname*{argmin}_{y \in \mathbb{R}^d} \left\{ f(y) + \frac{1}{2\gamma}\|y - x\|^2 \right\}.$$

   (a) Prove that for all $x \in \mathbb{R}^d$, we have

$$x_+ = \mathrm{prox}_{\gamma f}(x) \iff (x - x_+) \in \gamma \partial f(x_+)$$

   (**Hint:** use strong convexity.)

   (b) Prove that $x \in \mathbb{R}^d$ is minimizes f if and only if $x = \mathrm{prox}_{\gamma f}(x)$.

   (c) **(Minty's Theorem.)** Prove that

$$\mathrm{range}(I + \partial f) = \{x + v \colon v \in \partial f(x)\} = \mathbb{R}^d.$$

   (**Hint:** use part (a).)

   (d) Prove that $\mathrm{prox}_{\gamma f}$ is 1-Lipschitz, i.e.,

$$\|\mathrm{prox}_{\gamma f}(x) - \mathrm{prox}_{\gamma f}(y)\| \le \|x - y\|, \qquad \forall x, y \in \mathbb{R}^d.$$

   (**Hint:** use strong convexity.)

   Notice the relation between proximal and projection operators: If $f(x) = \delta_{\mathcal{X}}$ for a closed convex set $\mathcal{X}$, then $\mathrm{prox}_{\gamma f} = \mathrm{proj}_{\mathcal{X}}$ for all $\gamma > 0$.

2. **Calculus of Proximal Operators.**

   (a) **(Linear Perturbation.)** Suppose that $f\colon \mathbb{R}^d \to (-\infty, \infty]$ is closed, proper, and convex, let $\gamma > 0$, and let $b, v \in \mathbb{R}^d$. Define a function

$$g(x) = f(x + b) + v^T x, \qquad \forall x \in \mathbb{R}^d$$

   Prove that

$$\mathrm{prox}_{\gamma g}(x) = \mathrm{prox}_{\gamma f}(x - \gamma v + b) - b, \qquad \forall x \in \mathbb{R}^d$$

   (**Hint:** First try the cases where $b = 0$ or $v = 0$.)

(b) **(Separability.)** Let $d = d_1 + \ldots + d_n$ for integers d_i and let $f_i \colon \mathbb{R}^{d_i} \to (-\infty, +\infty]$ be proper convex functions. Let $\gamma > 0$ and for all $x = (x_1, \ldots, x_n) \in \mathbb{R}^d$, define $f(x_1, \ldots, x_n) := \sum_{i=1}^n f(x_i)$. Prove that

$$\mathrm{prox}_{\gamma f}(x_1, \ldots, x_n) = (\mathrm{prox}_{\gamma f}(x_1), \ldots, \mathrm{prox}_{\gamma f_n}(x_n)), \qquad \forall x \in \mathbb{R}^d.$$

(c) **(Scalarization.)** Let $f \colon \mathbb{R} \to (-\infty, \infty]$ be a scalar function, let $\gamma > 0$, and let $a \in \mathbb{R}^d \backslash \{0\}$. Define

$$g(x) = f(a^T x), \qquad \forall x \in \mathbb{R}^d$$

Prove that for all $x \in \mathbb{R}^d$, we have

$$\mathrm{prox}_{\gamma g}(x) = x - \rho a \qquad \text{where } \rho = \frac{1}{\|a\|^2}(a^T x - \mathrm{prox}_{(\gamma \|a\|^2) f}(a^T x)).$$

(**Hint:** Be careful: the chain rule $\partial g(y) = a \partial f(a^T y)$ may not hold. Instead, use the inclusion $a \partial f(a^T y) \subseteq \partial g(y)$.)

3. **Proximal Operator Examples.** Compute the proximal operators of the following functions

(a) $f(x) := \|x\|_1 = \sum_{i=1}^d |x_i|$.

(b) $f(x) = \max\{0, x\}$ for a scalar variable $x \in \mathbb{R}$.

(c) $f(x) = \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle$, where $b \in \mathbb{R}^d$ and $A \in \mathbb{R}^{d \times d}$ is a symmetric positive semidefinite matrix.

(d) $f(x) = \|x\|_2$.
(**Hint:** First compute the subdifferential of f, keeping in mind that f is differentiable everywhere except the origin.)

(e) $f(x) = \delta_{\mathcal{X}}$, where $\mathcal{X} = \{x \in \mathbb{R}^d \colon x \geq 0\}$ is the nonnegative orthant.

(f) $f(x) = \delta_{\mathcal{X}}$, where $\mathcal{X} = \{x \colon Ax = b\}$ is an affine space defined by matrix $A \in \mathbb{R}^{m \times d}$ and $b \in \mathbb{R}^m$.

(g) $f(x) = \delta_{\mathcal{X}}$, where $\mathcal{X} = \{x \colon \|x\|_\infty \leq 1\}$
(**Hint:** You already computed $\partial f(x)$ on a previous homework assignment.)

4. **(Projection onto $\mathbb{S}_+^{d \times d}$).** Recall that any symmetric matrix $A \in \mathbb{S}^{d \times d}$ (not necessarily positive semidefinite) has an eigenvalue decomposition

$$A = Q \Lambda Q^T \qquad \text{where} \left\{ \begin{array}{r} Q^T Q = I \\ \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_d) \\ \lambda_1 \geq \ldots \geq \lambda_d. \end{array} \right\}.$$

For any such matrix, prove that

$$\mathrm{proj}_{\mathbb{S}_+^{d \times d}}(A) = Q \max\{\Lambda, 0\} Q^T.$$

(**Hint:** Verify the first order optimality conditions $A - \mathrm{proj}_{\mathbb{S}_+^{d \times d}}(A) \in \mathcal{N}_{\mathbb{S}_+^{d \times d}}(\mathrm{proj}_{\mathbb{S}_+^{d \times d}}(A))$)

Finally consider the following problem on sensitivity analysis for linear programs.

5. Consider the linear program $\min(c^T x : x \geq 0, Ax = b)$. Let B denote an optimal basis. Assume that the problem is generic in that each vertex has a unique basis for which it is the corresponding basic solution. Suppose now that you want to solve a parametric problem, i.e., a set of problems of the form $\min((c+\lambda d)^T x : x \geq 0, Ax = b)$, for each possible value of $\lambda \geq 0$. Assume that for any $\lambda \geq 0$ the problem has an optimal solution and that the basis B is a solution for the problem when $\lambda = 0$.

   (a) Prove that the set of values of λ for which basis B is optimal forms an interval $[0, a_1]$. Explain how to compute a_1.

   (b) Show that there is a finite set $a_0 = 0 \leq a_1 \leq ... \leq a_k$ and corresponding bases B_i for $i = 0, ..., k$ such that $B_0 = B$ and B_i (for $i = 0, ..., k$) is the optimal basis if and only if $\lambda \in [a_i, a_{i+1}]$, and B_k is optimal if $\lambda \geq a_k$.