Name: Philip Pham

AMATH 515 Homework Set 3

Due: Monday, March 9th by midnight..

(1) Compute the conjugates of the following functions.

(a) $f(x) = \delta_{\mathbb{B}_\infty}(x)$.

Proof. We have that

$$\boxed{f^\star(x) = \delta^\star_{\mathbb{B}_\infty}(x) = \max_{y \in \mathbb{B}_\infty} \langle x, y\rangle = \sum_i \left|x_i\right|.}$$

By definition,

$$f^\star(x) = \sup_y \left\{\langle x, y\rangle - f(y)\right\}.$$

If $y \not\in \mathbb{B}_\infty$, then $\langle x, y\rangle - f(y) = -\infty$ for all x. For $y \in \mathbb{B}_\infty$, $\langle x, y\rangle - f(y) = \langle x, y\rangle$, so to maximize $\langle x, y\rangle - f(y)$, we should choose y such that $\langle x, y\rangle$ is maximized. □

(b) $f(x) = \delta_{\mathbb{B}_2}(x)$.

Proof. We have that

$$\boxed{f^\star(x) = \delta^\star_{\mathbb{B}_2}(x) = \max_{y \in \mathbb{B}_2} \langle x, y\rangle = \left\|x\right\|_2.}$$

The proof is identical to the previous one with $\mathbb{B}_\infty$ replaced by $\mathbb{B}_2$. □

(c) $f(x) = \exp(x)$.

Proof. We have that

$$\boxed{f^\star(x) = \begin{cases} x\log x - x = x\left(\log x - 1\right), & x > 0; \\ 0, & x = 0; \\ \infty, & x < 0. \end{cases}}$$

To see this, we can maximize $\langle x, y\rangle - \exp\left(y\right)$ with respect to y by taking the derivative, setting it to 0, and solving for y. In doing so, we find that $y = \log x$, which is only defined when $x > 0$. When $x \leq 0$, we see that we can maximize $xy - \exp\left(y\right)$ by sending y to $-\infty$. □

(d) $f(x) = \log(1 + \exp(x))$

Proof. We have that

$$\boxed{f^\star(x) = \begin{cases} \infty, & x > 1; \\ 0, & x = 1; \\ x \log x + (1 - x) \log(1 - x), & 0 < x < 1; \\ 0, & x = 0; \\ \infty, & x < 0. \end{cases}}$$

Consider $xy - \log(1 + \exp(y))$. The derivative with respect to y is $x - \frac{\exp(y)}{1+\exp(y)}$. When $x \in (0, 1)$, we can solve for $y = \log\left(\frac{x}{1-x}\right)$. When $x \geq 1$, the derivative is always positive, so we have that the max is obtained when $y \to \infty$. We note that

$$xy - \log(1 + \exp(y)) = xy - \left[y + \exp(-y) - \frac{\exp(-2y)}{2} + \frac{\exp(-3y)}{3} - \frac{\exp(-4y)}{4} + \cdots\right]$$

by a Taylor series expansion, so as $y \to \infty$, we have ∞ if $x > 1$ and 0 if $x = 1$.

When $x \leq 0$, the derivative is always negative, so we maximize the expression with $y \to -\infty$. When $x = 0$, only the second term remains, so we have 0. If $x < 0$, we have that the first term tends to ∞ and the second term tends to 0. □

(e) $f(x) = x \log(x)$

Proof. We have that

$$\boxed{f^\star(x) = \exp(x - 1).}$$

We can take the derivative of $xy - y \log y$ with respect to y, set this equal to 0, and solve for y. We find that $y = \exp(x - 1)$. Substituting, we obtained the desired result. □

(2) Let g be any convex function; f is formed using g. Compute $f^\star$ in terms of $g^\star$.

(a) $f(x) = \lambda g(x)$.

Proof.

$$\boxed{f^\star(x) = \lambda g^\star\left(\frac{x}{\lambda}\right)}$$

To see this,

$$\begin{aligned} f^\star(x) &= \sup_y \{\langle x, y\rangle - f(y)\} \\ &= \sup_y \{\langle x, y\rangle - \lambda g(y)\} \\ &= \lambda \sup_y \left\{\left\langle \frac{x}{\lambda}, y\right\rangle - g(y)\right\} \\ &= \lambda g^\star\left(\frac{x}{\lambda}\right). \end{aligned}$$

□

(b) $f(x) = g(x-a) + \langle x, b\rangle$.

Proof.

$$\boxed{f^\star(x) = \langle x-b, a\rangle + g^\star(x-b)}$$

$$\begin{aligned} f^\star(x) &= \sup_y \{\langle x, y\rangle - f(y)\} \\ &= \sup_y \{\langle x, y\rangle - g(y-a)\} \\ &= \sup_y \{\langle x, y\rangle - \langle y, b\rangle - g(y-a)\} \\ &= \sup_y \{\langle x-b, y\rangle - g(y-a)\} \\ &= \sup_y \{\langle x-b, a\rangle + \langle x-b, y-a\rangle - g(y-a)\} \\ &= \langle x-b, a\rangle + \sup_y \{\langle x-b, y-a\rangle - g(y-a)\} \\ &= \langle x-b, a\rangle + g^\star(x-b) \end{aligned}$$

since shiftfing by a does not change the supremum. □

(c) $f(x) = \inf_z \{g(x, z)\}$.

Proof. Rewrite $g(x, z) = g_z(x)$ to emphasize that we're conjugating on the x argument. Then, we have that

$$\boxed{f^\star(x) = \sup_z g_z^\star(x).}$$

Since

$$\sup_y \left\{ \langle x, y\rangle - \inf_z g(y,z) \right\} \geq \sup_y \left\{ \langle x, y\rangle - g(y,z) \right\}$$

for all z, and we can choose z so that the right-hand size is arbitrarily close to the left-hand size, we have that

$$f^\star(x) = \sup_y \left\{ \langle x, y\rangle - \inf_z g(y,z) \right\} = \sup_z \sup_y \left\{ \langle x, y\rangle - g(y,z) \right\}.$$

Since $g_z^\star(x) = \sup_y \left\{ \langle x, y\rangle - g(y,z) \right\}$, we have our desired result. □

(d) $f(x) = \inf_z \left\{ \frac{1}{2}\|x-z\|^2 + g(z) \right\}$

Proof. Let $h_z(x) = \frac{1}{2}\|x-z\|^2 + g(z)$.

Noting that the conjugate of $\|x\|^2/2$ is itself, we can apply part (b) to obtain $h_z^\star(x) = \frac{1}{2}\|x\|^2 + \langle x, z\rangle - g(z)$.

Applying the results of part (c), we have that

$$f^\star(x) = \sup_z \left\{ \frac{1}{2}\|x\|^2 + \langle x, z\rangle - g(z) \right\} = \boxed{\frac{1}{2}\|x\|^2 + g^\star(x).}$$

□

(3) Moreau Identities.

(a) Derive the Moreau Identity:

$$\operatorname{prox}_f(z) + \operatorname{prox}_{f^\star}(z) = z.$$

Proof. It follows from Corollary 4.41 of the text:

$$y \in \partial f(x) \Leftrightarrow x \in \partial f^\star(y).$$

We also use Proposition 4.39 to note that

$$\begin{aligned}\operatorname{prox}_f(z) = \arg\min_x f(x) + \frac{1}{2}\|x-z\|^2 &\Leftrightarrow 0 \in \partial \left(f\left(\operatorname{prox}_f(z)\right) + \frac{1}{2}\left\|\operatorname{prox}_f(z) - z\right\|^2 \right) \\ &\Leftrightarrow z - \operatorname{prox}_f(z) \in \partial f\left(\operatorname{prox}_f(z)\right).\end{aligned}$$

So we use these two facts to find that

$$\begin{aligned} z - \operatorname{prox}_f(z) \in \partial f\left(\operatorname{prox}_f(z)\right) &\Leftrightarrow \operatorname{prox}_f(z) \in \partial f^\star\left(z - \operatorname{prox}_f(z)\right) \\ &\Leftrightarrow z - \left(z - \operatorname{prox}_f(z)\right) \in \partial f^\star\left(z - \operatorname{prox}_f(z)\right) \\ &\Leftrightarrow \operatorname{prox}_{f^\star}(z) = z - \operatorname{prox}_f(z) \\ &\Leftrightarrow z = \operatorname{prox}_f(z) + \operatorname{prox}_{f^\star}(z), \end{aligned}$$

which is our desired result.

□

(b) Use either of the Moreau identities and 1a, 1b to check your formulas for

$$\operatorname{prox}_{\|\cdot\|_1}, \quad \operatorname{prox}_{\|\cdot\|_2}$$

from last week's homework.

Proof. For $\operatorname{prox}_{\|\cdot\|_1}$, we note that

$$\begin{aligned} (\|x\|_1)^\star &= \sup_y \{\langle x, y\rangle - |y|_1\} \\ &= \sup_y \left\{ \sum_i x_i y_i - |y_i| \right\} \\ &= \begin{cases} \infty, & \text{if any } |x_i| > 1; \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

Thus, the proximal operator is

$$\begin{aligned} \operatorname{prox}_{(\|\cdot\|_1)^\star}(x) &= \arg\min_y \frac{1}{2}\|y - x\|^2 + (\|y\|_1)^\star \\ &= \min(|x|, 1)\operatorname{sign}(x), \end{aligned}$$

where I have abused notation and all operations are done element-wise.

Then, applying the Moreau identity, we obtain

$$\begin{aligned} \operatorname{prox}_{\|\cdot\|_1}(x) &= x - \operatorname{prox}_{(\|\cdot\|_1)^\star}(x) \\ &= x - \min(|x|, 1)\operatorname{sign}(x) \\ &= \operatorname{sign}(x)\max(|x| - 1, 0), \end{aligned}$$

which agrees with my previous result.

We have that

$$\begin{aligned}(\|x\|_2)^\star &= \sup_y \{\langle x, y\rangle - \| y\|_2\} \\ &\geq \| x\|_2\| y\|_2 - \| y\|_2 \\ &\geq \| y\|_2 (\| x\|_2 - 1),\end{aligned}$$

where we have equality if we choose y in the same direction as x. Thus, we'll have that $(\|x\|_2)^* = \infty$ if $\| x\|_2 > 1$ and $(\|x\|_2)^* = 0$ if $\| x\|_2 \leq 1$.

Then, we have that

$$\begin{aligned}\operatorname{prox}_{(\|\cdot\|_2)^\star}(x) &= \arg\min_y \frac{1}{2}\|y - x\|^2 + (\|y\|_2)^\star \\ &= \begin{cases} x, & \|x\|_2 \leq 1; \\ \frac{x}{\|x\|_2}, & \|x\|_2 > 1. \end{cases}\end{aligned}$$

Then, applying the Moreau identity, we obtain

$$\begin{aligned}\operatorname{prox}_{\|\cdot\|_2}(x) &= x - \operatorname{prox}_{(\|\cdot\|_2)^\star}(x) \\ &= \begin{cases} 0, & \|x\|_2 \leq 1; \\ x\left(1 - \frac{1}{\|x\|_2}\right), & \|x\|_2 > 1. \end{cases}\end{aligned}$$

This agrees with my previous result for $\operatorname{prox}_{\|\cdot\|_2}$.

□

(4) Duals of regularized GLM. Consider the Generalized Linear Model family:

$$\min_x \sum_{i=1}^n g(\langle a_i, x\rangle) - b^T Ax + R(x),$$

Where g is convex and R is any regularizer.

(a) Write down the general dual obtained from the perturbation

$$p(u) = \min_x \sum_{i=1}^n g(\langle a_i, x\rangle + u_i) - b^T Ax + R(x).$$

Proof. Let

$$f_x(u) = \sum_{i=1}^n g(\langle a_i, x\rangle + u_i) - b^T Ax + R(x).$$

By Problem (2)(c),

$$p^\star(u) = \sup_x \{f_x^\star(u)\}.$$

By Problem (2)(b),

$$\begin{aligned}
f_x^{\star}(v) &= \sup_u \left\{ \langle v, u \rangle - \left[\sum_{i=1}^{n} g(\langle a_i, x \rangle + u_i) - b^T Ax + R(x) \right] \right\} \\
&= \sup_u \left\{ \sum_{i=1}^{n} [v_i u_i - g(\langle a_i, x \rangle + u_i)] + \left[b^T Ax - R(x) \right] \right\} \\
&= b^T Ax - R(x) + \sup_u \left\{ \sum_{i=1}^{n} [v_i u_i - g(\langle a_i, x \rangle + u_i)] \right\} \\
&= b^T Ax - R(x) + \sum_{i=1}^{n} \sup_{u_i} \{ v_i u_i - g(\langle a_i, x \rangle + u_i) \} \\
&= b^T Ax - R(x) + \sum_{i=1}^{n} [g^{\star}(v_i) - v_i \langle a_i, x \rangle] \\
&= b^T Ax - R(x) - v^T Ax + \sum_{i=1}^{n} g^{\star}(v_i) .
\end{aligned}$$

Thus, we have that

$$\begin{aligned}
p^{\star}(u) &= \sup_x \left\{ b^T Ax - R(x) - u^T Ax + \sum_{i=1}^{n} g^{\star}(u_i) \right\} \\
&= \sum_{i=1}^{n} g^{\star}(u_i) + \sup_x \left\{ (b-u)^T Ax - R(x) \right\} \\
&= \sum_{i=1}^{n} g^{\star}(u_i) + R^{\star}\left(A^T (b-u) \right) .
\end{aligned}$$

□

(b) Specify your formula to Ridge-regularized logistic regression:

$$\min_x \sum_{i=1}^{n} \log(1 + \exp(\langle a_i, x \rangle)) - b^T Ax + \frac{\lambda}{2} \|x\|^2.$$

Proof. Using the results in from Problem (1)(d) and Problem (2)(a), we have that

$$g(x) = \log(1+\exp(x)) \Rightarrow g^\star(x) = \begin{cases} \infty, & |2x-1| > 1; \\ 0, & |2x-1| = 1; \\ x\log x + (1-x)\log(1-x), & \text{otherwise;} \end{cases}$$

$$R(x) = \frac{\lambda}{2}\|x\|^2 \Rightarrow R^\star(x) = \frac{1}{2\lambda}\|x\|^2$$

We can substitute to get

$$p(u) = \sum_{i=1}^{n} g^\star(u_i) + \frac{1}{2\lambda}\|A^T(b-u)\|^2.$$

□

(c) Specify your formula to 1-norm regularized Poisson regression:

$$\min_x \sum_{i=1}^{n} \exp(\langle a_i, x\rangle) - b^T A x + \lambda\|x\|_1.$$

Proof. Using the results in from Problem (1)(c), Problem (2)(a), and Problem (3)(b), we have that

$$g(x) = \exp(x) \Rightarrow g^\star(x) = \begin{cases} \infty, & x < 0; \\ 0, & x = 0; \\ x(\log x - 1), & \text{otherwise;} \end{cases}$$

$$R(x) = \lambda\|x\|_1 \Rightarrow R^\star(x) = \begin{cases} \infty, & \text{if any } |x_i| > \lambda; \\ 0, & \text{otherwise.} \end{cases}$$

We can substitute to get

$$p(u) = \sum_{i=1}^{n} g^\star(u_i) + R^\star\left(A^T(b-u)\right).$$

□

Coding Assignment

Please download `515Hw3_Coding.ipynb` and `proxes.py` to complete problem (5).

(5) In this problem you will write a routine to project onto the capped simplex.

The Capped Simplex Δ_k is defined as follows:

$$\Delta_k := \left\{x : \mathbf{1}^T x = k, \quad 0 \le x_i \le 1 \quad \forall i.\right\}$$

This is the intersection of the k-simplex with the unit box.

The projection problem is given by

$$\text{proj}_{\Delta_k}(z) = \arg\min_{x\in\Delta_k} \frac{1}{2}\|x-z\|^2.$$

(a) Derive the (1-dimensional) dual problem by focusing on the $\mathbf{1}^T x = k$ constraint.

Proof. We can rewrite the the problem as

$$\begin{aligned}
\text{proj}_{\Delta_k}(z) &= \arg\min_{x\in\Delta_k} \frac{1}{2}\|x-z\|^2 \\
&= \arg\min_{x\in\mathbb{R}^n} \frac{1}{2}\|x-z\|^2 + \delta_{\{0\}}\left(\mathbf{1}^T x - k\right) + \delta_{[0,1]^n}(x) \\
&= \arg\min_{x\in\mathbb{R}^n} \frac{1}{2}\|x-z\|^2 + \sup_{\lambda\in\mathbb{R}}\left\{\lambda\left(\mathbf{1}^T x - k\right)\right\} + \delta_{[0,1]^n}(x) \\
&= \arg\min_{x\in\mathbb{R}^n}\sup_{\lambda\in\mathbb{R}} \left\{\frac{1}{2}\|x-z\|^2 + \lambda\left(\mathbf{1}^T x - k\right) + \delta_{[0,1]^n}(x)\right\}.
\end{aligned}$$

If $k \in (0, n]$, we can apply Slater's condition since there is guaranteed to be an x such that $x \in [0,1]^n$ and $\mathbf{1}^T x = k$. Thus, we have strong duality, and

$$\min_{x\in\mathbb{R}^n}\sup_{\lambda\in\mathbb{R}} \left\{\frac{1}{2}\|x-z\|^2 + \lambda\left(\mathbf{1}^T x - k\right) + \delta_{[0,1]^n}(x)\right\} =$$

$$\boxed{\sup_{\lambda\in\mathbb{R}}\min_{x\in\mathbb{R}^n} \frac{1}{2}\|x-z\|^2 + \lambda\left(\mathbf{1}^T x - k\right) + \delta_{[0,1]^n}(x).}$$

□

(b) Implement a routine to solve this dual. It's a scalar root finding problem, so you can use the root-finding algorithm provided in the code.

Proof. Suppose we fix λ. The inner minimization problem is solved by $\text{prox}_{g+h}(z)$, where $g(x) = \lambda\mathbf{1}^T x - \lambda k$ and $h(x) = \delta_{[0,1]^n}(x)$. Since g is affine, we can apply

Theorem 3 of Yu's *On Decomposing the Proximal Map* to obtain

$$\begin{aligned}\operatorname{prox}_{g+h}(z) &= \operatorname{prox}_h\left(\operatorname{prox}_g(z)\right)\\ &= \operatorname{prox}_h\left(x - \lambda\mathbf{1}\right)\\ &= \min\left((x-\lambda\mathbf{1})_+, \mathbf{1}\right),\end{aligned}$$

where $(\cdot)_+ = \max(\cdot, 0)$, and the min and $(\cdot)_+$ operation are taken to be element-wise.

Denote $x_\lambda = \min\left((x-\lambda\mathbf{1})_+, \mathbf{1}\right)$. With this notation, we have that

$$\sup_{\lambda\in\mathbb{R}} \min_{x\in\mathbb{R}^n} \frac{1}{2}\|x-z\|^2 + \lambda\left(\mathbf{1}^T x - k\right) + \delta_{[0,1]^n}(x) = \sup_{\lambda\in\mathbb{R}} \frac{1}{2}\|x_\lambda - z\|^2 + \lambda\left(\mathbf{1}^T x_\lambda - k\right),$$

where the last term falls away since x_λ is always in the unit box. By duality, the solution must occur when $\mathbf{1}^T x_\lambda - k = 0$, so we need to solve for λ such that

$$\mathbf{1}^T \min\left((z-\lambda\mathbf{1})_+, \mathbf{1}\right) - k = 0.$$

The first term is monotonically decreasing as a function of λ, so we can use bisection to solve for λ. See `prox_csimplex` in Listing 1 for a Python implementation.

□

(c) Using the dual solution, write down a closed form formula for the projection. Use this formula, along with your dual solver, to implement the projection. You can use the unit test provided to check if your code is working correctly.

Proof.

$$\operatorname{proj}_{\Delta_k}(z) = \min\left((z-\lambda\mathbf{1})_+, \mathbf{1}\right), \tag{1}$$

$$\text{where } \lambda \text{ satisfies } 0 = \mathbf{1}^T \min\left((z-\lambda\mathbf{1})_+, \mathbf{1}\right) - k,$$

where the min and $(\cdot)_+$ operation are taken to be element-wise.

See Figure 1 for an example of this algorithm. □

```
"""Implementation of `prox_csimplex`."""
import numpy as np
from scipy.optimize import bisect

UW_ID = '1772371'
FIRST_NAME = 'Philip'
LAST_NAME = 'Pham'

# Prox of capped simplex
# -----------------------------------------------------------------------------
def prox_csimplex(z, k):
  """Prox of capped simplex argmin_x 1/2||x - z||^2 s.t. x in k-capped-simplex.

  Args:
    z: arraylike, reference point
    k: float, positive number between 0 and z.size, denote simplex cap

  Returns:
    arraylike, projection of z onto the k-capped simplex
  """
  # safe guard for k
  if k < 0 or k > z.size:
    raise ValueError(
      'k: k must be between 0 and dimension of the input. k = {}'.format(k))
  # 1. Construct the scalar dual object.
  def f(y):
    return np.sum(np.clip(z - y, 0, 1)) - k
  # 2. Use `bisect` to solve it.
  lower_bound, upper_bound = -1., 1.
  while f(lower_bound) < 0:
    lower_bound *= 2
  while f(upper_bound) >= 0:
    upper_bound *= 2
  y = bisect(f, lower_bound, upper_bound)
  # 3. Obtain primal variable from optimal dual solution and return it.
  return np.clip(z - y, 0, 1)
```

Listing 1: An implementation of Equation 1 in the `proxes.py` file.

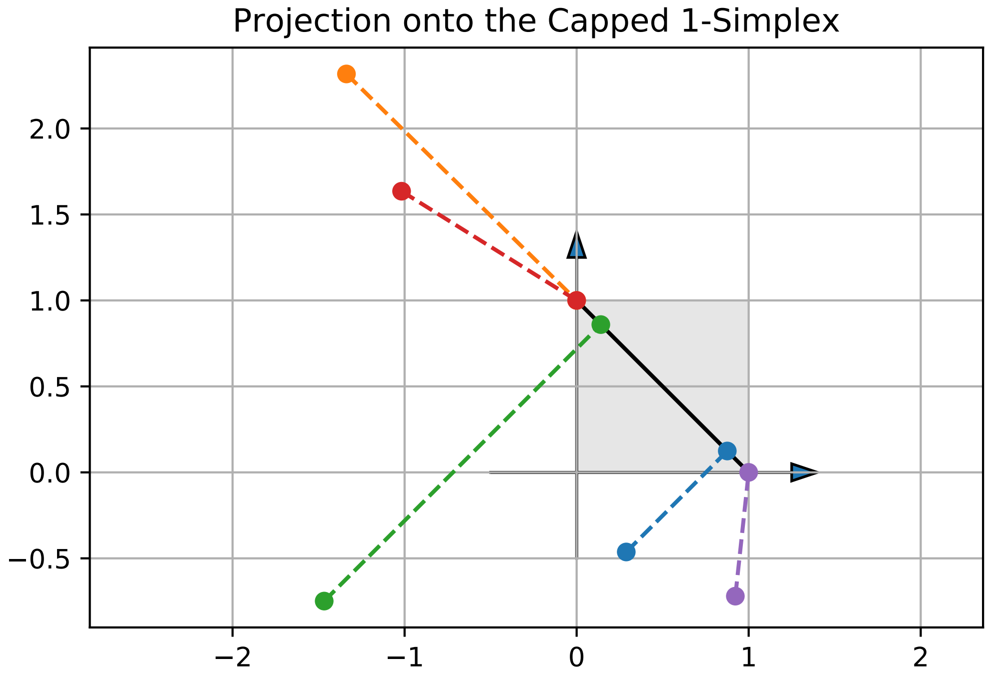

Figure 1: Projection of 5 randomly-generated points using Equation 1.