AMATH 515 Problem Set 3

Tyler Chen

Problem 1

Compute the conjugates of the following functions.

(a) $f(x) = \delta_{\mathbb{B}_\infty}(x)$.
(b) $f(x) = \delta_{\mathbb{B}_2}(x)$.
(c) $f(x) = \exp(x)$.
(d) $f(x) = \log(1+\exp(x))$
(e) $f(x) = x\log(x)$

Solution

Recall that for a function $f : E \to \mathbb{R}$, the Fenchel conjugate $f^* : E \to \mathbb{R}$ is defined as,

$$f^*(y) = \sup_{x\in E}\{\langle y,x\rangle - f(x)\}$$

Note that if $E = \mathbb{R}$ then the x attaining the supremum,

$$\sup_{x\in\mathbb{R}}\{yx - f(x)\}$$

will also satisfy,

$$0 = \frac{\mathrm{d}}{\mathrm{d}x}\left(yx - f(x)\right) = y - f'(x)$$

(a) By definition,

$$f^*(y) = \sup_{x\in\mathbb{R}}\{\langle y,x\rangle - \delta_{\mathbb{B}_\infty}(x)\} = \sup_{x\in\mathbb{B}_\infty}\langle y,x\rangle$$

It is obvious that we should pick $x_i = \operatorname{sign}(y_i)$. Therefore,

$$f^*(y) = \|y\|_1$$

(b) By definition,

$$f^*(y) = \sup_{x\in\mathbb{R}}\{\langle y,x\rangle - \delta_{\mathbb{B}_2}(x)\} = \sup_{x\in\mathbb{B}_2}\langle y,x\rangle$$

By the geometric interpretation of the inner product and the Euclidian norm ball, it is obvious that we should pick $x = y/\|y\|_2$ so that,

$$f^*(y) = \|y\|_2$$

(c) We have,

$$f^*(y) = \sup_{x\in E}\{xy - e^x\} = y\log(y) - y$$

where we have solved,

$$0 = \nabla\left[xy - e^x\right] = y - e^x$$

to obtain,

$$x = \log(y)$$

(d) We have,

$$f^*(y) = \sup_{x \in E}\{xy - \log(1+\exp(x))\} = y\log\left(\frac{y}{1-y}\right) - \log\left(\frac{1}{1-y}\right)$$

where we have solved,

$$0 = \nabla\left[xy - \log(1+\exp(x))\right] = y - \frac{\exp(x)}{1+\exp(x)}$$

to obtain,

$$x = \log\left(\frac{y}{1-y}\right)$$

(e) We have,

$$f^*(y) = \sup_{x \in \mathbb{R}}\{xy - x\log x\} = y\exp(y-1) - \exp(y-1)(y-1) = \exp(y-1)$$

where we have solved,

$$0 = \nabla\left[xy - x\log(x)\right] = y - (1+\log(x))$$

to obtain,

$$x = \exp(y-1)$$

Problem 2

Let g be any convex function; f is formed using g. Compute f^* in terms of g^*.

(a) $f(x) = \lambda g(x)$.
(b) $f(x) = g(x-a) + \langle x, b\rangle$.
(c) $f(x) = \inf_z \{g(x,z)\}$.
(d) $f(x) = \inf_z \left\{\frac{1}{2}\|x-z\|^2 + g(z)\right\}$

Solution

Note that we do not explicitly write the domain over which the maximizations and minimizations occur. We note that occasionally this domain will be shifted based when variables are substituted or shifted. However, we leave this implicit in our solutions.

(a) Note that we must have $\lambda \geq 0$ so that λg is convex. Assume further that $\lambda > 0$ so that the problem is nontrivial. In this case,

$$\begin{aligned} f^*(y) &= \sup_x \{\langle y, x\rangle - \lambda g(x)\} \\ &= \sup_x \{\lambda(\langle y/\lambda, x\rangle - g(x))\} \\ &= \lambda \sup_x \{\langle y/\lambda, x\rangle - g(x)\} \\ &= \lambda g^*(y/\lambda) \end{aligned}$$

(b) By definition,

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

(c) Note that,

$$g^*(y,u) = \sup_{[x,z]} \{\langle [y,u], [x,z]\rangle - g(x,z)\} = \sup_{x,z} \{\langle y, x\rangle + \langle u, z\rangle - g(x,z)\}$$

Therefore,

$$\begin{aligned} f^*(y) &= \sup_x \left\{ \langle y, x \rangle - \inf_z \{ g(x,z) \} \right\} \\ &= \sup_x \left\{ \sup_z \{ \langle y, x \rangle - g(x,z) \} \right\} \\ &= \sup_{x,z} \{ \langle y, x \rangle - g(x,z) \} \\ &= g^*(y,0) \end{aligned}$$

(d) By (c) we know that,

$$f^*(y) = G^*(y,0), \qquad G(x,z) = \frac{1}{2}\|x-z\|^2 + g(z)$$

We now compute,

$$\begin{aligned} G^*(y,u) &= \sup_{[x,z]} \left\{ \langle [y,u], [x,z] \rangle - \frac{1}{2}\|x-z\|^2 - g(z) \right\} \\ &= \sup_{x,z} \left\{ \langle y, x \rangle + \langle u, z \rangle - \frac{1}{2}\|x-z\|^2 - g(z) \right\} \\ &= \sup_{x,z} \left\{ \langle y, x-z \rangle - \frac{1}{2}\|x-z\|^2 + \langle u+y, z \rangle - g(z) \right\} \end{aligned}$$

We now define $w = x - z$ and note that w is still free from z so that,

$$\begin{aligned} G^*(y,u) &= \sup_{w,z} \left\{ \langle y, w \rangle - \frac{1}{2}\|w\|^2 + \langle u+y, z \rangle - g(z) \right\} \\ &= \sup_{w} \left\{ \langle y, w \rangle - \frac{1}{2}\|w\|^2 \right\} + \sup_z \left\{ \langle u+y, z \rangle - g(z) \right\} \\ &= \frac{1}{2}\|y\|^2 + g^*(u+y) \end{aligned}$$

Therefore,

$$G^*(y,0) = \frac{1}{2}\|y\|^2 + g^*(y)$$

Problem 3

Moreau Identities.

(a) Derive the Moreau Identity:

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

You may find the 'Fenchel flip' useful.

(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.

Solution

Recall that when f is closed, proper, and convex,

$$z \in \partial f(x) \Longleftrightarrow x \in \partial f^*(z)$$

(a) Fix z and assume f is closed, proper, and convex. By definition,

$$\operatorname{prox}_f(z) = \arg\min_{x_1} \left(\frac{1}{2}\|x_1 - z\|^2 + f(x_1) \right)$$
$$\operatorname{prox}_{f^*}(z) = \arg\min_{x_2} \left(\frac{1}{2}\|x_2 - z\|^2 + f^*(x_2) \right)$$

Since the proximal operator is a well defined function, $x_1 = \operatorname{prox}_f(z)$ is the *unique* point so that,

$$0 \in \partial \left(\frac{1}{2}\|x_1 - z\|^2 + f(x_1) \right) = x_1 - z + \partial f(x_1)$$

Similarly, $x_2 = \operatorname{prox}_{f^*}(z)$ is the *unique* point so that,

$$0 \in \partial \left(\frac{1}{2}\|x_2 - z\|^2 + f^*(x_2) \right) = x_2 - z + \partial f^*(x_2)$$

Equivalently, $x_1 = \operatorname{prox}_f(z)$ and $x_2 = \operatorname{prox}_{f^*}(z)$ are the unique points so that,

$$z - x_1 \in \partial f(x_1), \qquad\qquad z - x_2 \in \partial f^*(x_2)$$

Now, using the Frenchel flip we find that $x_1 = \operatorname{prox}_f(z)$ is the unique points so that,

$$x_1 \in \partial f^*(z - x_1)$$

Writing $x_3 = z - x_1$ we have that x_3 is the unique points so that,

$$z - x_3 \in \partial f^*(x_3)$$

Therefore $x_3 = x_2 = z - x_2$ so that,

$$z = x_1 + x_2 = \operatorname{prox}_f(z) + \operatorname{prox}_{f^*}(z)$$

(b) Suppose $f = \|\cdot\|_1$. Then $f^* = \delta_{\mathbb{B}_\infty}$. We have previously derived,

$$\operatorname{prox}_f(z) = \begin{cases} z_i + 1, & z_i < -1 \\ 0, & z_i \in [-1,1] \\ z_i - 1, & z_i > 1 \end{cases}$$

Note that when computing $\operatorname{prox}_{f^*}$ we must keep $x \in \mathbb{B}_\infty$ so $x_i \in [-1,1]$. Moreover, we want x_i as near to z_i as possible. Therefore,

$$\operatorname{prox}_{f^*}(z) = \arg\min_x \left(\frac{1}{2} \|x - z\|^2 + \delta_{\mathbb{B}_\infty}(x) \right) = \begin{cases} -1, & z_i < 1 \\ z_i, & z_i \in [-1,1] \\ 1, & z_i > 1 \end{cases}$$

The identity is clearly satisfied. □

Now, suppose $f = \|\cdot\|_2$. Then $f^* = \delta_{\mathbb{B}_2}$. We have previously derived,

$$\operatorname{prox}_f(z) = \begin{cases} \left(1 - \frac{1}{\|z\|}\right) z, & \|z\| \geq 1 \\ 0 & \|z\| < 1 \end{cases}$$

Observe that when computing $\operatorname{prox}_{f^*}$ we must keep $x \in \mathbb{B}_\infty$. It is obviously best to pick x in the direction of z with magnitude to cancel as much of z as possible. Therefore,

$$\operatorname{prox}_{f^*}(z) = \arg\min_x \left(\frac{1}{2} \|x - z\|^2 + \delta_{\mathbb{B}_2}(x) \right) = \begin{cases} \frac{z}{\|z\|}, & \|z\| \geq 1 \\ z & \|z\| < 1 \end{cases}$$

The identity is again clearly satisfied. □

Problem 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

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

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

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

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

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

Solution

(a) For convenience define,

$$\varphi(x,u) = h(\tilde{b} - \tilde{A}x + u) + \langle \tilde{c}, x\rangle + k(x)$$

where,

$$h(z) = \sum_{i=1}^{n} g(z_i), \qquad \tilde{A} = -A, \qquad \tilde{b} = 0, \qquad \tilde{c} = -A^T b, \qquad k(x) = R(x)$$

Then,

$$\begin{aligned}
\varphi^*(z,v) &= k^*(z + \tilde{A}^T v - \tilde{c}) + h^*(v) - \left\langle v, \tilde{b}\right\rangle \\
&= k^*(z - A^T v + A^T b) + h^*(v) - \langle v, 0\rangle \\
&= k^*(z - A^T(v - b)) + h^*(v)
\end{aligned}$$

By the definition of convex conjugate we have,

$$h^*(v) = \sup_z \{\langle v, z\rangle - h(z)\} = \sup_{z_i} \left\{ \sum_{i=1}^{n} v_i z_i - g(z_i) \right\} = \sum_{i=1}^{n} g^*(v_i)$$

Moreover, the dual problem is,

$$\sup_v \{-p^*(v)\} = \sup_v \{-\varphi(0,v)\} = \sup_v \left\{ -\sum_{i=1}^{n} g^*(v_i) - R^*(A^T(b - v)) \right\}$$

(b) Here $g(z) = \log(1+\exp(z))$ and $R(x) = \frac{\lambda}{2}\|x\|^2$. Therefore,

$$h^*(v) = \sup_x \left\{ \langle v, x\rangle - \sum_{i=1}^{n} \log(1+\exp(x_i)) \right\}$$

Taking the gradient to be zero we find,

$$0 = \nabla\left[\langle v, x\rangle - \sum_{i=1}^{n} \log(1+\exp(x_i)) \right] = v - \frac{\exp(x)}{1+\exp(x)}$$

Now, solving for x, we have,

$$x = \log\left(\frac{v}{1-v}\right)$$

Thus,

$$h^*(v) = \left\langle v, \log\left(\frac{v}{1-v}\right)\right\rangle - \sum_{i=1}^{n} \log\left(\frac{1}{1-v_i}\right)$$

Similarly,

$$k^*(w) = \sup_x \left\{ \langle w, x\rangle - \frac{\lambda}{2}\|x\|^2 \right\} = \frac{1}{\lambda}\frac{\|w\|^2}{2}$$

Therefore, the dual problem is,

$$\sup_v \left\{ -\left\langle v, \log\left(\frac{v}{1-v}\right)\right\rangle + \sum_{i=1}^{n} \log\left(\frac{1}{1-v_i}\right) - \frac{1}{\lambda}\frac{\|A^T(b-v)\|^2}{2} \right\}$$

(c) Here $g(z) = \exp(z)$ and $R(x) = \lambda\|x\|_1$. Therefore,

$$h^*(v) = \sup_x \left\{ \langle v, x\rangle - \sum_{i=1}^{n} \exp(\langle a_i, x\rangle) \right\}$$

Taking the gradient to be zero we find,

$$0 = \nabla\left[\langle v, x\rangle - \sum_{i=1}^{n} \exp(x_i) \right] = v - \exp(x)$$

Now, solving for x we have,

$$x = \log(v)$$

Thus,

$$h^*(v) = \langle v, \log(v)\rangle - \sum_{i=1}^{n} v_i$$

Similarly,

$$k^*(w) = \sup_x \{\langle w, x\rangle - \lambda\|x\|_1\} = \delta_{\lambda\mathbb{B}_\infty}(w)$$

Therefore, the dual problem is,

$$\sup_v \left\{ - \langle v, \log(v)\rangle + \sum_{i=1}^{n} v_i - \delta_{\lambda\mathbb{B}_\infty}(A^T(v-b)) \right\}$$

Problem 5

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

The Capped Simplex Δ_k is defined as follows:

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

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

The projection problem is given by

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

(a) Derive the (1-dimensional) dual problem by focusing on the $1^T x = k$ constraint.
(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.
(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 id working correctly.

Solution

(a) Note that we can write,

$$\begin{aligned}
\operatorname*{proj}_{\Delta_k}(z) &= \arg\min_{x \in \Delta_k} \left(\frac{1}{2}\|x - z\|^2 \right) \\
&= \left\{ x \in \Delta_k : \frac{1}{2}\|x - z\|^2 = \min_{x \in \Delta_k} \frac{1}{2}\|x - z\|^2 \right\} \\
&= \left\{ x < \infty : \max_{\lambda} \left(\frac{1}{2}\|x - z\|^2 + \lambda(1^T x - k) \right) \right. \\
&\qquad\qquad \left. = \max_{\lambda} \min_{x \in [0,1]^n} \left(\frac{1}{2}\|x - z\|^2 + \lambda(1^T x - k) \right) \right\}
\end{aligned}$$

As such, we focus on the dual problem,

$$\max_{\lambda} \min_{x \in [0,1]^n} \left(\frac{1}{2}\|x - z\|^2 + \lambda(1^T x - k) \right)$$

Define,

$$[f(x)](\lambda) = \frac{1}{2}\|x - z\|^2 + \lambda(1^T x - k)$$

We note that the minimum of f over $\mathbb{R}^n$ (for a fixed λ) occurs at the solution to,

$$0 = \nabla \left(\frac{1}{2}\|x - z\|^2 + \lambda(1^T x - k) \right) = (x - z) + \lambda \cdot 1$$

That is, at $x = z - \lambda \cdot 1$.

However, since we must constrain x to be in the unit box, this will not be the minimizer of the constrained problem. Since $\frac{1}{2}\|x-z\|^2 + \lambda(1^T x - k)$ is a quadratic with all the coefficients of the quadratic terms are equal (so that the level curves are circles), we can find the constraint minimizer by projecting to the unit box. That is,

$$x_{\text{opt}} = \max(\min(z - \lambda \cdot 1, 1), 0) = \begin{cases} 1, & z_i > \lambda + 1 \\ z_i - \lambda, & z_i \in [\lambda, \lambda+1] \\ 0, & z_i < \lambda \end{cases}$$

Plugging this in to f we find,

$$\begin{aligned} [f(x_{\text{opt}})](\lambda) &= -\lambda k + \sum_i ((x_{\text{opt}})_i - z_i)^2 + \lambda x_i \\ &= -\lambda k + \sum_i \begin{cases} \frac{1}{2}(1-z_i)^2 + \lambda, & z_i > \lambda + 1 \\ \frac{1}{2}\lambda^2 + \lambda(z_i - \lambda), & z_i \in [\lambda, \lambda+1] \\ \frac{1}{2}(z_i)^2, & z_i < \lambda \end{cases} \end{aligned}$$

For convenience define,

$$f_i(\lambda) = \begin{cases} \frac{1}{2}(1-z_i)^2 + \lambda, & \lambda < z_i - 1 \\ \frac{1}{2}\lambda^2 + \lambda(z_i - \lambda), & \lambda \geq z_i - 1, \lambda \leq z_i \\ \frac{1}{2}z_i^2, & \lambda > z_i \end{cases}$$

Then clearly,

$$f_i'(\lambda) = \begin{cases} 1, & \lambda < z_i - 1 \\ z_i - \lambda, & \lambda \geq z_i - 1, \lambda \leq z_i \\ 0, & \lambda > z_i \end{cases}$$

This is obviously continuous in λ so that $f_i(\lambda)$ is smooth. Therefore, $[f(x_{\text{opt}})](\lambda)$ is a smooth function of λ.

(b) To solve the dual problem $\min_\lambda [f(x_{\text{opt}})](\lambda)$ we set the derivative to zero and solve for λ. That is solve,

$$0 = [f(x_{\text{opt}})]'(\lambda) = -k + \sum_i \begin{cases} 1, & \lambda < z_i - 1 \\ z_i - \lambda, & \lambda \geq z_i - 1, \lambda \leq z_i \\ 0, & \lambda > z_i \end{cases}$$

We will use bisection to do this. This means we need to find λ_1 such that $f'(\lambda_1) \geq 0$ and λ_2 such that $f'(\lambda_2) \leq 0$. In particular, since $0 \leq k \leq n$ we can pick $\lambda_1 < \min_i(z_i) - 1$ and $\lambda_2 > \max_i(z_i)$.

(c) To find the solution x we return to the expression,

$$x_{\text{opt}} = \max(\min(z - \lambda \cdot 1, 1), 0) = \begin{cases} 1, & z_i > \lambda + 1 \\ z_i - \lambda, & z_i \in [\lambda, \lambda + 1] \\ 0, & z_i < \lambda \end{cases}$$

Note that $f_i'(\lambda) = x_{\text{opt}}(\lambda) = \max(\min(z - \lambda, 1), 0)$ so that we can conveniently implement these functions using `np.clip`.