AMATH 515 Homework 3

Cade Ballew

March 9, 2022

1 Problem 1

1.1 Part a

Let $f(x) = \delta_{\mathbb{B}_\infty}(x)$. Then,

$$f^*(z) = \sup_x \left\{ z^T x - \delta_{\mathbb{B}_\infty}(x) \right\} = \sup_{x \in \mathbb{B}_\infty} z^T x.$$

This supremum is clearly obtained by taking

$$x_i = \begin{cases} 1, & z_i > 0 \\ -1, & z_i < 0. \end{cases}$$

Thus,

$$f^*(z) = \sum_{i=1}^n |z_i| = \|z\|_1.$$

1.2 Part b

Let $f(x) = \delta_{\mathbb{B}_2}(x)$. Then, by the Cauchy-Schwarz inequality,

$$f^*(z) = \sup_x \left\{ z^T x - \delta_{\mathbb{B}_2}(x) \right\} = \sup_{x \in \mathbb{B}_2} z^T x \leq \sup_{\|x\| < 1} \|z\| \|x\| \leq \|z\|.$$

However, this bound is attained by taking $x = z/\|z\|$ (which is on the boundary of the ball) which gives that $z^T x = \|z\|$. Thus, this is bound is in fact the supremum, so

$$f^*(z) = \|z\|.$$

1.3 Part c

Let $f(x) = \exp(x)$ where $x \in \mathbb{R}$. Then,

$$f^*(z) = \sup_x \left\{ z^T x - \exp(x) \right\} = \sup_x \left\{ zx - \exp(x) \right\},$$

because z and x are scalars. Setting the derivative with respect to x equal to zero to find the x that maximizes this quantity,

$$0 = z - \exp(x),$$

so $x = \log(z)$ which gives that

$$f^*(z) = z\log(z) - \exp(\log(z)) = z(\log(z) - 1).$$

1.4 Part d

Let $f(x) = \log(1 + \exp(x))$ be a scalar function. Then,

$$f^*(z) = \sup_x \left\{zx - \log(1 + \exp(x))\right\}.$$

Setting the derivative equal to 0,

$$0 = z - \frac{e^x}{1 + e^x}.$$

We can write this as

$$e^x = \frac{z}{1 - z},$$

so $x = \log \frac{z}{1-z}$ and

$$f^*(z) = z\log\frac{z}{1-z} - \log\left(1 + \frac{z}{1-z}\right) = z\log\frac{z}{1-z} - \log\frac{1}{1-z}.$$

1.5 Part e

Let $f(x) = x\log(x)$ be a scalar function. Then,

$$f^*(z) = \sup_x \left\{zx - x\log(x)\right\}.$$

Setting the derivative equal to 0,

$$0 = z - (\log x + 1),$$

so $x = \exp(z - 1)$ and

$$f^*(z) = z\exp(z-1) - \exp(z-1)(z-1) = \exp(z-1).$$

2 Problem 2

Let g be a given convex function.

2.1 Part a

If $f(x) = \lambda g(x)$, then

$$\begin{aligned} f^*(z) &= \sup_x \left\{ z^T x - \lambda g(x) \right\} = \sup_x \left\{ \lambda \left(\frac{1}{\lambda} z^T x - g(x) \right) \right\} \\ &= \lambda \sup_x \left\{ \left(\frac{z}{\lambda} \right)^T x - g(x) \right\} = \lambda g^* \left(\frac{z}{\lambda} \right). \end{aligned}$$

2.2 Part b

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

$$\begin{aligned} f^*(z) &= \sup_x \left\{ z^T x - g(x-a) - b^T x \right\} = \sup_x \left\{ (z-b)^T x - g(x-a) \right\} \\ &= \sup_x \left\{ (z-b)^T x - g(x-a) \right\} = \sup_x \left\{ (z-b)^T (x-a) - g(x-a) + (z-b)^T a \right\}. \end{aligned}$$

Now, we rename $x' = x - a$ to get

$$\begin{aligned} f^*(z) &= \sup_{x'} \left\{ (z-b)^T x' - g(x') + (z-b)^T a \right\} \\ &= \sup_{x'} \left\{ (z-b)^T x' - g(x') \right\} + a^T (z-b) = g^*(z-b) + a^T(z-b). \end{aligned}$$

2.3 Part c

If $f(x) = \inf_z \{ g(x,z) \}$, then

$$\begin{aligned} f^*(z) &= \sup_x \left\{ z^T x - \inf_y \{ g(x,y) \} \right\} = \sup_x \left\{ z^T x + \sup_y \{ -g(x,y) \} \right\} \\ &= \sup_{x,y} \left\{ z^T x - g(x,y) \right\}. \end{aligned}$$

Now, note that

$$g^*(z,w) = \sup_{x,y} \left\{ z^T x + w^T y - g(x,y) \right\},$$

so clearly,

$$f^*(z) = g^*(z, 0).$$

2.4 Part d

If $f(x) = \inf_z \left\{ \frac{1}{2} \|x - z\|^2 + g(z) \right\}$, define $h(x,y) = \frac{1}{2}\|x-y\|^2 + g(y)$ and compute

$$\begin{aligned} h^*(z,w) &= \sup_{x,y} \left\{ z^T x + w^T y - \frac{1}{2} \|x-y\|^2 - g(y) \right\} \\ &= \sup_{x,y} \left\{ z^T (x-y) + (w+z)^T y - \frac{1}{2} \|x-y\|^2 - g(y) \right\}. \end{aligned}$$

Now, define $x' = x - y$ to get

$$\begin{aligned} h^*(z,w) &= \sup_{x',y} \left\{ z^T x' + (w+z)^T y - \frac{1}{2}\|x'\|^2 - g(y) \right\} \\ &= \sup_{x'} \left\{ z^T x' - \frac{1}{2}\|x'\|^2 \right\} + \sup_{y} \left\{ (w+z)^T y - g(y) \right\} \\ &= p^*(z) + g^*(w+z) \end{aligned}$$

where $p(x) = \frac{1}{2}\|x\|^2$. From page 8 of lecture 12, we know that $p^*(z) = p(z)$, so

$$h^*(z,w) = \frac{1}{2}\|z\|^2 + g^*(w+z).$$

Now, we can apply part c of this problem to conclude that

$$f^*(z) = h^*(z,0) = \frac{1}{2}\|z\|^2 + g^*(z).$$

3 Problem 3

3.1 Part a

Let f be a closed proper convex function. To derive the Moreau identity, we first remark that for a closed proper convex function g, $u = \operatorname{prox}_g(v)$ iff $v - u \in \partial g(u)$. To see this, note that

$$\operatorname{prox}_g(v) \arg\min_x \left\{ \frac{1}{2}\|x - v\|^2 + g(x) \right\},$$

so the convexity of g and the differentiability of the 2-norm gives that $u = \operatorname{prox}_g(v)$ iff u minimizes the interior quantity iff $(u - v) + \partial g(u) \ni 0$ which occurs iff $v - u \in \partial g(u)$. Using this, let $x = \operatorname{prox}_f(z)$. Then, $z - x \in \partial f(x)$. Now, we apply the Fenchel flip to get that $x \in \partial f^*(z - x)$. This means that

$$\underbrace{z}_{v} - \underbrace{(z - x)}_{u} \in \partial f^*(\underbrace{z - x}_{u}),$$

so we again apply our property to get that this is true iff

$$z - x = \operatorname{prox}_{f^*}(z).$$

Thus,

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

3.2 Part b

From part a and problem 1 part a, we know that

$$z = \operatorname{prox}_{\mathbb{B}_\infty}(z) + \operatorname{prox}_{\mathbb{B}_\infty^*}(z) = \operatorname{prox}_{\mathbb{B}_\infty}(z) + \operatorname{prox}_{\|\cdot\|_1}(z).$$

Thus,

$$\text{prox}_{\|\cdot\|_1}(z) = z - \text{prox}_{\mathbb{B}_\infty}(z).$$

From homework 2, we know that the elements of $\text{prox}_{\mathbb{B}_\infty}(z)$ are given by

$$\text{prox}_{\mathbb{B}_\infty}(z)_i = \begin{cases} 1, & z_i > 1 \\ -1, & z_i < -1 \\ z_i, & |z_i| \leq 1. \end{cases}$$

Thus, the elements of $\text{prox}_{\|\cdot\|_1}(z)$ are given by

$$\text{prox}_{\|\cdot\|_1}(z)_i = \begin{cases} z_i - 1, & z_i > 1 \\ z_i + 1, & z_i < -1 \\ 0, & |z_i| \leq 1 \end{cases}$$

which matches what we derived on homework 2.
Similarly, problem 1 part b tells us that

$$z = \text{prox}_{\mathbb{B}_2}(z) + \text{prox}_{\mathbb{B}_2^*}(z) = \text{prox}_{\mathbb{B}_2}(z) + \text{prox}_{\|\cdot\|_2}(z),$$

so

$$\text{prox}_{\|\cdot\|_2}(z) = z - \text{prox}_{\mathbb{B}_2}(z).$$

From lecture 9 page 8, we know that

$$\text{prox}_{\mathbb{B}_2}(z) = \begin{cases} z, & \|z\| \leq 1 \\ \frac{z}{\|z\|}, & \|z\| > 1. \end{cases}$$

Thus,

$$\text{prox}_{\|\cdot\|_2}(z) = \begin{cases} 0, & \|z\| \leq 1 \\ z - \frac{z}{\|z\|}, & \|z\| > 1 \end{cases}$$

which matches what we derived on homework 2.

4 Problem 4

4.1 Part a

Consider

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

where g is convex and R is any regularizer. From lecture 13 page 11, we know that the problem

$$\min_x f(Qx - d) + h(x) + c^T x$$

has dual

$$\sup_z -z^T d - f^*(z) - h^*(-Q^T z - c).$$

To write our problem in this form, take $Q = A$, $d = 0$, $h(x) = R(x)$, $c = -A^T b$ ($c^T = -b^T A$), and $f(x) = \sum_{i=1}^n g(x_i)$. Now, compute

$$\begin{aligned} f^*(z) &= \sup_x \left\{ z^T x - \sum_{i=1}^n g(x_i) \right\} = \sup_{x_1,\dots,x_n} \left\{ \sum_{i=1}^n (z_i x_i - g(x_i)) \right\} \\ &= \sum_{i=1}^n \sup_{x_i} \{ z_i x_i - g(x_i) \} = \sum_{i=1}^n g^*(z_i). \end{aligned}$$

Thus, our problem has dual

$$\sup_z \left\{ -\sum_{i=1}^n g^*(z_i) - R^*(-A^T z + A^T b) \right\} = \sup_z \left\{ -\sum_{i=1}^n g^*(z_i) - R^*(A^T(b-z)) \right\}.$$

4.2 Part b

Now, we compute the dual of

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

by taking $g(x) = \log(1 + \exp(x))$ and $R(x) = \frac{\lambda}{2}\|x\|^2$. From problem 1 part d, we know that

$$g^*(z) = z \log \frac{z}{1-z} - \log \frac{1}{1-z}.$$

From problem 2 part a and the conjugate of $\frac{1}{2}\|x\|^2$ that we already established, we find that

$$R^*(z) = \frac{\lambda}{2}\|z/\lambda\|^2 = \frac{1}{2\lambda}\|z\|^2.$$

Thus, the dual of our problem is given by

$$\sup_z \left\{ -\sum_{i=1}^n \left(z_i \log \frac{z_i}{1-z_i} - \log \frac{1}{1-z_i} \right) - \frac{1}{2\lambda}\|A^T(b-z)\|^2 \right\}.$$

4.3 Part c

Now, we compute the dual of

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

by taking $g(x) = \exp(x)$ and $R(x) = \lambda\|x\|_1$. From problem 1 part c, we have that

$$g^*(z) = z(\log(z) - 1).$$

We also know from problem 1 part a that $\delta^*_{\mathbb{B}_\infty}(z) = \|z\|_1$, so using the fact that the 1-norm is closed convex, $\|z\|_1^* = \delta_{\mathbb{B}_\infty}(z)$ ($f^{**} = f$ under these conditions from lecture 12 page 13). Applying problem 2 part a,

$$R^*(z) = \lambda \delta_{\mathbb{B}_\infty}(z/\lambda) = \delta_{\mathbb{B}_\infty}(z/\lambda) = \delta_{\lambda\mathbb{B}_\infty}(z).$$

Thus, the dual of our problem is given by

$$\sup_z \left\{ -\sum_{i=1}^n z_i(\log(z_i) - 1) - \delta_{\lambda\mathbb{B}_\infty}(A^T(b-z)) \right\}.$$

5 Problem 5

5.1 Part a

Consider the Capped Simplex Δ_k

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

and the projection problem is given by

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

From lecture 13 page 13, we know that the dual constraint for the condition $1^T x = k$ is given by

$$\sup_z z(1^T x - k),$$

so we can write our dual problem as

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

Taking the gradient the interior and setting it equal to zero to minimize with respect to x.

$$0 = (x - y) + z1$$

which gives $x^* = y - z1$. However, we need to ensure that $x^* \in [0,1]^n$ which we do by projecting onto that space as

$$x_i^* = \begin{cases} y_i - z, & 0 \le y_i - z \le 1 \\ 1, & y_i - z > 1 \\ 0, & y_i - z < 0. \end{cases}$$

Plugging this in for x, our dual problem becomes

$$\sup_z \sum_{i=1}^n f(z)_i$$

where

$$f(z)_i = \begin{cases} -\frac{z^2}{2} + zy_i, & 0 \le y_i - z \le 1 \\ \frac{(y_i-1)^2}{2} + z, & y_i - z > 1 \\ \frac{y_i^2}{2}, & y_i - z < 0. \end{cases}$$

5.2 Part b

To solve this dual, we note that our function is separable, so we set its gradient equal to 0 by computing

$$f(z)'_i = \begin{cases} -z + y_i, & 0 \leq y_i - z \leq 1 \\ 1, & y_i - z > 1 \\ 0, & y_i - z < 0. \end{cases}$$

and writing

$$0 = \sum_{i=1}^{n} f'(z)_i.$$

We find the solution to numerically which yields an optimal z^*.

5.3 Part c

Once we find an optimal z^*, we can then find our optimal x^* as described above, meaning that elementwise,

$$\mathrm{proj}_{\Delta_k}(y)_i = x_i^* = \begin{cases} y_i - z^*, & 0 \leq y_i - z \leq 1 \\ 1, & y_i - z > 1 \\ 0, & y_i - z < 0. \end{cases}$$