ECE 602 - Introduction to Optimization

Solutions to Home Assignment 2

Exercise 1

(a) Find the conjugate functions for the following $f : \mathbb{R}^n \to \mathbb{R}$.

- $f(x) = I_C(x)$, where $I_C(x)$ is the indicator function of some non-empty, closed convex set C;
- $f(x) = \lambda \|\Sigma x\|_1$, where $\lambda > 0$ and $\Sigma \in \mathbb{S}^n_{++}$ is a diagonal matrix with its diagonal elements equal to $\sigma_i > 0$;
- $f(x) = \lambda \|\Sigma x\|_2$, where $\lambda > 0$ and $\Sigma \in \mathbb{S}^n_{++}$ are as above;
- $f(x) = \frac{1}{2} x^T A x + b^T x + c$, where $A \in \mathbb{S}^n_{++}$, $b \in \mathbb{R}^n$, and $c \in \mathbb{S}$.
- $f(x) = -\sum_{i=1}^n \log x_i$, with $\operatorname{dom} f = \mathbb{R}_{++}$.

(b) Provide closed-form expressions for the proximal operators of all the above functions. (Use Moreau decomposition where needed.)

Solution:

(a) (i) Substituting into the definition

$$\begin{aligned} f^*(y) &= \sup_{x \in \mathbf{dom}\, f} (y^\top x - I_C(x)) \\ &= \sup_{x \in C} (y^\top x) \end{aligned}$$

(ii) As shown in the lecture and in the textbook, for $g(x) = \|x\|$,

$$g^*(y) = \begin{cases} 0 & \text{if } \|y\|_* \le 1 \\ \infty & \text{otherwise} \end{cases}$$

We can represent the function $f(x)$ as $\lambda g(\Sigma x)$. Using the properties of conjugate functions we have

$$
\begin{aligned}
f^*(y) &= \lambda g^* \left(\frac{\Sigma^{-T} y}{\lambda} \right) \\
&= \begin{cases} 0 & \text{if } \left\| \frac{\Sigma^{-T} y}{\lambda} \right\|_* \leq 1 \\ \infty & \text{otherwise} \end{cases} \\
&= \begin{cases} 0 & \text{if } \left\| \Sigma^{-1} y \right\|_\infty \leq \lambda \\ \infty & \text{otherwise} \end{cases} \\
&= I_{\Lambda_\infty}
\end{aligned}
$$

where Λ_∞ represents the set of points y where $\left\| \Sigma^{-1} y \right\|_\infty \leq \lambda$.

(iii) Using the same procedure as for (ii), we have

$$
\begin{aligned}
f^*(y) &= \begin{cases} 0 & \text{if } \| \Sigma^{-1} y \|_2 \leq \lambda \\ \infty & \text{otherwise} \end{cases} \\
&= I_{\Lambda_2}
\end{aligned}
$$

where Λ_2 represents the set of points y where $\| \Sigma^{-1} y \|_2 \leq \lambda$.

(iv)

$$
\begin{aligned}
f^*(y) &= \sup_{x \in \mathbf{dom}\, f} (y^\top x - \frac{1}{2} x^\top A x - b^\top x - c) \\
0 &= \frac{\partial}{\partial x} \left(y^\top x^* - \frac{1}{2} x^{*\top} A x^* - b^\top x^* - c \right) \\
&= y - A x^* - b \\
x^* &= A^{-1}(y - b) \\
f^*(y) &= y^\top A^{-1}(y - b) - \frac{1}{2}(A^{-1}(y - b))^\top A A^{-1}(y - b) - b^\top A^{-1}(y - b) - c \\
&= y^\top A^{-1}(y - b) - \frac{1}{2}(y - b)^\top A^{-1}(y - b) - b^\top A^{-1}(y - b) - c \\
&= (y - b)^\top A^{-1}(y - b) - \frac{1}{2}(y - b)^\top A^{-1}(y - b) - c \\
&= \frac{1}{2}(y - b)^\top A^{-1}(y - b) - c
\end{aligned}
$$

(v) Let $f_i : \mathbb{R} \to \mathbb{R}$. As seen in the lecture,

$$
\begin{aligned}
f_i^*(y_i) &= \sup_{x_i \in \mathbf{dom}\, f_i} (y_i x_i + \log x_i) \\
0 &= \frac{\partial}{\partial x_i} (y_i x_i^* + \log x_i^*) \\
&= y_i - \frac{1}{x_i^*} \\
x_i^* &= -\frac{1}{y_i} \\
f_i^*(y_i) &= y_i \left(-\frac{1}{y_i} \right) + \log \left(-\frac{1}{y_i} \right) \\
&= \begin{cases} -1 - \log(-y_i) & \text{if } y_i < 0 \\ \infty & \text{if } y_i \geq 0 \end{cases}
\end{aligned}
$$

Based on the property that for $f(x) = f_1(x_1) + f_2(x_2)$, the corresponding conjugate function is $f^*(x) = f_1^*(x_1) + f_2^*(x_2)$,

$$
f^*(y) = \begin{cases} -n - \sum_{i=1}^{n} \log(-y_i) & \text{if } \forall i, y_i < 0 \\ \infty & \text{otherwise} \end{cases}
$$

(b) (i) Substituting into the definition

$$\begin{aligned}\underset{f}{\mathbf{prox}}(y) &= \underset{x}{\operatorname{argmin}}\left\{\frac{1}{2}\|x-y\|_2^2 + I_C(x)\right\}\\ &= \underset{x\in C}{\operatorname{argmin}}\left\{\frac{1}{2}\|x-y\|_2^2\right\}\\ &= \mathcal{P}_C(y)\end{aligned}$$

which is the orthogonal projection of y onto C.

(ii) First, we note that the problem is separable

$$\begin{aligned}\underset{\lambda f}{\mathbf{prox}}(y) &= \underset{x}{\operatorname{argmin}}\left\{\frac{1}{2}\|x-y\|_2^2 + \lambda\|x\|_1\right\}\\ &= \underset{x}{\operatorname{argmin}}\sum_{i=1}^{n}\left(\frac{1}{2}(x_i-y_i)^2 + \lambda|\sigma_i x_i|\right)\\ &= (\underset{\lambda f_i}{\mathbf{prox}}(y_i))_i^n\\ \underset{\lambda f_i}{\mathbf{prox}}(y_i) &= \underset{x_i}{\operatorname{argmin}}\left\{\frac{1}{2}(x_i-y_i)^2 + \lambda|\sigma_i x_i|\right\}\end{aligned}$$

Next, we solve for a single component of the proximal mapping.

$$\begin{aligned}\partial f_i(y_i) &= (x_i - y_i) + \lambda\partial(|\sigma_i x_i|)\\ 0 &\in (x_i^* - y_i) + \lambda\partial(|\sigma_i x_i|)\\ y_i &\in x_i^* + \lambda\partial(|\sigma_i x_i|)\end{aligned}$$

where, since $\sigma_i > 0$,

$$\partial(|\sigma_i x_i|) = \begin{cases}\sigma_i & \text{if } x_i > 0\\ [-\sigma_i, \sigma_i] & \text{if } x_i = 0\\ -\sigma_i & \text{if } x_i < 0\end{cases}$$

Taking the inverse, we have

$$\begin{aligned}x_i^* &= \begin{cases}y_i - \lambda\sigma_i & \text{if } y_i > \lambda\sigma_i\\ 0 & \text{if } -\lambda\sigma_i \le y_i \le \lambda\sigma_i\\ y_i + \lambda\sigma_i & \text{if } y_i < \lambda\sigma_i\end{cases}\\ &= S_{\lambda\sigma_i}(y_i)\end{aligned}$$

Therefore, the proximal mapping of f is

$$\underset{\lambda f}{\mathbf{prox}}(y) = \mathcal{S}_{\lambda\Sigma}(y)$$

where Σ is applied to the soft-thresholding function $\mathcal{S}$ element-wise.

(iii) Let $f(x) = \lambda\|\Sigma x\|_2 = \lambda g(x)$, where $g(x) = \|\Sigma x\|_2$. Using Moreau decomposition, we have

$$\underset{\lambda g}{\mathbf{prox}}(y) = y - \lambda \underset{\lambda^{-1} g^*}{\mathbf{prox}}\left(\frac{y}{\lambda}\right)$$

$$g^*(x) = \begin{cases} 0 & \text{if } \left\|\Sigma^{-1} x\right\|_2 \leq 1 \\ \infty & \text{otherwise} \end{cases} = \mathcal{I}_{\mathcal{N}_2}$$

where $\mathcal{N}_2$ represents the set of points x where $\|\Sigma^{-1}x\|_2 \leq 1$.

$$\underset{\lambda f}{\mathbf{prox}}(y) = y - \lambda \mathcal{P}_{\mathcal{N}_2}\left(\frac{y}{\lambda}\right)$$

$$\left(\mathcal{P}_{\mathcal{N}_2}\left(\frac{y}{\lambda}\right)\right)_i = \begin{cases} \|\Sigma\|_2 \frac{\frac{y}{\lambda}}{\|\frac{y}{\lambda}\|_2} & \text{if } \|\Sigma^{-1}(\frac{y}{\lambda})\|_2 > 1 \\ \frac{y}{\lambda} & \text{if } \|\Sigma^{-1}(\frac{y}{\lambda}))\|_2 \leq 1 \end{cases}$$

$$\underset{\lambda g}{\mathbf{prox}}(y) = \begin{cases} \left(1 - \lambda \frac{1}{\|\Sigma^{-1} y\|_2}\right) y & \text{if } \|\Sigma^{-1} y\|_2 > \lambda \\ 0 & \text{if } \|\Sigma^{-1} y\|_2 \leq \lambda \end{cases}$$

(iv)

$$\begin{aligned}
\underset{f}{\mathbf{prox}}(y) &= \underset{x}{\operatorname{argmin}} \left\{\frac{1}{2}\|x-y\|_2^2 + \frac{1}{2}x^\top A x + b^\top x + c\right\} \\
0 &= \frac{\partial}{\partial x}\left(\frac{1}{2}\|x-y\|_2^2 + \frac{1}{2}x^{*\top} A x^* + b^\top x^* + c\right) \\
&= x^* - y + x^* A + b \\
(A+I)x^* &= y - b \\
\underset{f}{\mathbf{prox}}(y) &= (A+I)^{-1}(y-b)
\end{aligned}$$

(v)

$$\begin{aligned}
\underset{f_i}{\mathbf{prox}}(y_i) &= \underset{x}{\operatorname{argmin}} \left\{\frac{1}{2}\|x_i - y_i\|_2^2 - \log x_i\right\} \\
0 &= \frac{\partial}{\partial x_i}\left\{\frac{1}{2}\|x_i - y_i\|_2^2 - \log x_i\right\} \\
&= x_i^* - y_i - \frac{1}{x_i^*} \\
0 &= x_i^{*2} - y_i x_i^* - 1 \\
x_i^* &= \frac{y_i + \sqrt{y_i^2 + 4}}{2} \\
\underset{f}{\mathbf{prox}}(y_i) &= \frac{y_i + \sqrt{y_i^2 + 4}}{2}
\end{aligned}$$

Based on the property that if f is a fully separable function, i.e. $f(x) = \sum_{i=1}^n f_i(x_i)$, then $(\mathbf{prox}_f(y))_i = (\mathbf{prox}_{f_i}(y_i))_{i=1}^n$ Therefore,

$$\underset{f}{\mathbf{prox}}(y) = \left(\frac{y_i + \sqrt{y_i^2 + 4}}{2}\right)_{i=1}^n$$

Exercise 2

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function of $x \in \mathbb{R}^n$, and let its associated *proximal operator* be $\text{prox}_f(x)$ (which is a mapping from $\mathbb{R}^n$ to $\mathbb{R}^n$).

(a) Assuming that $f(x)$ is separable, *viz.*

$$f(x) = \sum_{i=1}^{n} f_i(x_i),$$

derive an expression for $\text{prox}_f(x)$ in terms of $\text{prox}_{f_i}(x_i)$.

(b) Let $W \in \mathbb{R}^{n\times n}$ be an orthogonal matrix (i.e., $W^TW = WW^T = I$). Derive an expression for the proximal operator of $f(Wx)$ in terms of $\text{prox}_f(x)$. How does the expression changes when f is separable?

(c) Let $y \in \mathbb{R}^n$ be a measured signal contaminated by additive noises. For a properly chosen (square and orthogonal) W, the problem of *de-noising* of y can be formulated as an optimization problem of the form

$$x^* = \arg\min_x \left\{ \frac{1}{2}\|y - x\|_2^2 + \lambda \|Wx\|_1 \right\},$$

where $\lambda > 0$ is a user-defined regularization parameter. Derive a closed-form solution to the above problem using Moreau decomposition.

Solution:

(a)

$$\begin{aligned}
prox_f(x) &= \text{argmin}_y \left\{ \frac{1}{2}\|y - x\|_2^2 + f(y) \right\} \\
&= \text{argmin}_y \left\{ \frac{1}{2}\|y - x\|_2^2 + \sum_{i=1}^{n} f_i(y_i) \right\} \\
&= \text{argmin}_y \left\{ \sum_{i=1}^{n} \frac{1}{2}\|y_i - x_i\|^2 + f_i(y_i) \right\} \\
&= \left[prox_{f_1}(x_1), \dots, prox_{f_n}(x_n) \right].
\end{aligned}$$

(b) Lert $g(x) = f(Wx)$, then

$$\begin{aligned} prox_g(x) &= \text{argmin}_y \left\{ \frac{1}{2} \|y - x\|_2^2 + g(y) \right\} \\ &= \text{argmin}_y \left\{ \frac{1}{2} \|y - x\|_2^2 + f(Wy) \right\}. \end{aligned}$$

Let $z = Wy$, then by the orthogonal property of W we have $y = W^T z$ and

$$\begin{aligned} prox_g(x) &= \text{argmin}_y \left\{ \frac{1}{2} \|y - x\|_2^2 + f(Wy) \right\} \\ &= \text{argmin}_z \left\{ \frac{1}{2} \left\|W^T z - x\right\|_2^2 + f(z) \right\} \\ &= W^T \text{argmin}_z \left\{ \frac{1}{2} \|z - Wx\|_2^2 + f(z) \right\} \quad \text{since } WW^T = I \\ &= W^T prox_f(Wx). \end{aligned}$$

For separable functions, we have

$$prox_g(x) = W^T \left[prox_{f_1}(W_(:, 1)^T x), \ldots, prox_{f_n}(W_(:, n)^T x) \right].$$

(c) First define $f(x) = \lambda \|Wx\|_1$, then by the definition of the proximal mapping we have

$$x^* = \text{argmin}_x \left\{ \frac{1}{2} \|y - x\|_2^2 + \lambda \|Wx\|_1 \right\} = prox_f(y).$$

Then by the property of the Moreau decomposition, we have,

$$\begin{aligned} y &= prox_f(y) + prox_{f^*}(y) \\ &= x^* + prox_{f^*}(y). \end{aligned} \tag{5}$$

Now we need to find the conjugate of function $f(x) = \lambda \|Wx\|_1$ as follows:

$$f^*(y) = \sup_x \left\{ x^T y - \lambda \|Wx\|_1 \right\}.$$

Note that f is a scaled norm, therefore, f^* is the indicator function for the unit ball of the dual norm. Recall that the dual of ℓ_1-norm is ℓ_∞-norm. Thus, we have

$$f^*(y) = \begin{cases} 0 & \text{if } \|Wy\|_\infty \leq 1/\lambda, \\ \infty & \text{otherwise.} \end{cases}$$

Now we need the proximal mapping for the conjugate function.

$$\begin{aligned} prox_{f^*}(y) &= \text{argmin}_x \left\{ \frac{1}{2} \|y - x\|_2^2 + f^*(x) \right\} \\ &= \text{argmin}_{\|Wx\|_\infty \leq 1/\lambda} \left\{ \frac{1}{2} \|y - x\|_2^2 \right\} \end{aligned}$$

Note that $prox_{f^*}(y)$ is the projection of y on to the set $\|Wx\|_\infty \leq 1/\lambda$. Then for each $i \in [n]$, the i the element of $prox_{f^*}(y)$ is as follows:

$$\text{element } i \text{ of } prox_{f^*}(y) = \begin{cases} -1 & \text{if } W(:,i)^T y < -1/\lambda \\ y_i & \text{if } -1/\lambda \leq W(:,i)^T y \leq 1/\lambda \\ 1 & \text{if } W(:,i)^T y > 1/\lambda \end{cases}$$

Observe that if $\|Wy\|_\infty \leq 1/\lambda$ then $prox_{f^*}(y) = y$. Finally we substitute $prox_{f^*}(y)$ into Equation (5).

Exercise 3

Let I_0 be a grey-scale image represented by a real-valued matrix of size $n \times n$. Suppose that only half of the values of I_0 are known, while the other half has to be recovered through an optimization procedure. In what follows, you will be asked to compare the results produced by two alternative approaches.

Let Ω denotes the set of pixel coordinates over which the values of I_0 are known. Then, one can estimate the remaining values of the image by solving the following optimization problem

$$\begin{aligned} &\underset{I}{\text{minimize}} \ f_0(I) \\ &\text{subject to } I(i,j) = I_0(i,j), \ (i,j) \in \Omega. \end{aligned}$$

Note that, in this formulation, f_0 could be viewed as a "roughness measure" representing our a priori belief as to how variable the final result should be.

In this exercise, we consider two possible choices of f_0 as given below.

1. *Tikhonov regularizer (TR)*

$$f_0(I) = \sum_{i=2}^{n} \sum_{j=2}^{n} |I(i,j) - I(i-1,j)|^2 + |I(i,j) - I(i,j-1)|^2.$$

2. *Anisotropic total variation (aTV)*

$$f_0(I) = \sum_{i=2}^{n}\sum_{j=2}^{n} |I(i,j) - I(i-1,j)| + |I(i,j) - I(i,j-1)|.$$

Compute the above two solutions using the `cvx` package (`http://cvxr.com/cvx/`). Which of the two approaches provides a more natural result?

As an input, use a 64×64 image `I0` defined as

```
f = imread('cameraman.tif');
f = double(f(33:96,81:144));
```

Also, define Ω by means of a binary mask `M` given by

```
rng(2000);
M = false(64);
ind = randperm(64*64);
M(ind(1:64*64/2)) = true;
```

In this case, the observed values of I_0 are simply given by `I0(M)`, which is a column vector of length 2,048.

Finally, to visualize images, use

```
imagesc(I0)
axis square
colormap gray
```

Solution:

The code for the interpolation is attached as follows.

```
%% demo for TV image interpolation
clc; clear; close all;

I0 = imread('cameraman.tif'); % Original image
```

```
I0 = double(I0(33:96,81:144));

rng(2000);
M = false(64);
ind = randperm(64*64);
M(ind(1:64*64/2)) = true;

I1 = zeros(64,64);
I1(M) = I0(M); % Obscured image

cvx_begin
    variable Il2(64, 64);
    Il2(M) == I0(M); % Fix known pixel values.
    Ix = Il2(2:end,2:end) - Il2(2:end,1:end-1); % x (horiz) differences
    Iy = Il2(2:end,2:end) - Il2(1:end-1,2:end); % y (vert) differences
    minimize(norm([Ix(:); Iy(:)], 2)); % l2 roughness measure
cvx_end


cvx_begin
    variable Itv(64, 64);
    Itv(M) == I0(M); % Fix known pixel values.
    Ix = Itv(2:end,2:end) - Itv(2:end,1:end-1); % x (horiz) differences
    Iy = Itv(2:end,2:end) - Itv(1:end-1,2:end); % y (vert) differences
    minimize(norm([Ix(:); Iy(:)], 1)); % tv roughness measure
cvx_end

figure,
subplot(2,2,1), imshow(uint8(I0)); title('Original image');
subplot(2,2,2), imshow(uint8(I1)); title('Obscured image');
subplot(2,2,3), imshow(uint8(Il2)); title('L_2 reconstructed image');
subplot(2,2,4), imshow(uint8(Itv)); title ('TV reconstructed image');
```

We get the following images, from which we observe the TV reconstructed image is of better perceptual quality due to sharper details and less distortions.

Figure 1: Image Interpolation. (a) Original image. (b) Obscured image. (c) ℓ_2 reconstructed image. (d) TV reconstructed image.