Name: Hongda Li
HW: AMATH 515 HW2 Theory

Problem 1

(1) Let $x, y \in \mathbb{R}^n$, and consider a function $f : \mathbb{R}^n \to \mathbb{R}$. We make the following definitions:

$$\text{prox}_{tf}(y) := \arg\min_x \frac{1}{2t}\|x-y\|^2 + f(x)$$
$$f_t(y) := \min_x \frac{1}{2t}\|x-y\|^2 + f(x)$$

Notice that $\text{prox}_{tf}(y)$ is the minimizer of an optimization problem; in particular it is a vector in $\mathbb{R}^n$, On the other hand $f_t(y)$ is a function from $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, just as f. Suppose f is convex.

(a)

Objective: Show that f_t is convex.
Claim 1a1: Sum of 2 convex functions is convex. This is proved in HW1.
Claim 1a2: $||\cdot||$ is convex. This is proved in HW1.
Therefore, $\frac{1}{2t}\|x-y\|^2$ is a convex function and by hypothesis from the statement, $f(x)$ is also convex. By 1a1, the sum of the 2 function is convex.
Define a bi-variable convex function:

$$F(x,y) := \frac{1}{2t}\|x-y\|^2 + f(x)$$

Then it can be said that:

$$f_t(y) = \min_x F(x,y)$$

Notice that, the minimizing along one of the dimension preserves convexity of the function.
Proof: The function $\min_x F(x,y)$ is convex wrt to variable y. And this means that:

$$\min_u(F(u, x+\lambda(y-x))) \le \min_u(F(u,x) + \lambda F(u, y-x)) \le \min_u(F(u,x)) + \lambda \min_u(F(u,y-x))$$

The first $\le$ is justified by the fact that $F(x,y)$ is convex wrt both x, y, and the second $\le$ is justified by the properties of the minimizing operator.
The convexity is preserved, therefore $f_t(y)$ is a convex function.

(b)

Objective: Show that $\text{prox}_{t,f}(y)$ produces a unique result.
Claim 1b1: A strictly convex function that is level bounded and proper has a unique minimizer.
Proof: If a function is level bounded and proper then there exists some minimizers, choose any of them, name it x^+, then using strict convexity we have:

$$f(y) > f(x^+) + \nabla f(x^+)(y - x^+)$$

Which is true for all $y \ne x^+$, therefore x^+ is a uniquer global minimizer.
Claim 1b2: $||\cdot||^2$ is strictly convex. This is not hard to convince. Using the differential characteristic of convexity and $y \ne x \ne 0$, which makes the triangular inequality strict, and then it will show that 2-norm squared is a strictly convex function. Or, we can use the fact that the function is smooth and the second derivative is non-zero.
Claim 1b3: The following function is level bounded:

$$g(x) := \frac{1}{2t}\|x-y\|^2 + f(x)$$

Proof: By convexity of the function $f(x)$:

$$f(y) \geq f(x) + \partial f(x)(y-x) \implies f(x) \leq f(y) - \partial f(x)^T(y-x)$$

Then we can say that:

$$g(x) \leq \frac{1}{2t}\|x-y\|^2 + f(y) - \partial f(x)^T(y-x)$$

Notice that:

$$g(x) \leq \underbrace{\frac{1}{2t}\|x-y\|^2 + f(y) - \partial f(x)^T(y-x)}_{(1)} \leq \alpha$$

The middle expression (1) is level bounded because it's a quadratic function that is convex, with all leading coefficients on quadratic term to be positive, therefore:

$$\underset{\alpha}{\text{lev}}(g(x)) \subseteq \left\{x : \frac{1}{2t}\|x-y\|^2 + f(y) - \partial f(x)^T(y-x) \leq \alpha\right\}$$

The subset of a bounded set is bounded. Therefore $g(x)$ is level bounded.
Claim 1b4: $g(x)$ defined in previous claim is strictly convex. This is true because buy Claim 1b2 2-norm squared is strictly convex, in addition $f(x)$ is convex. The sum of a strict convex function and a convex function is strictly convex (Showed in HW1).
By Claim 1b3, Claim 1b4 the function $g(x)$ is level bounded and strictly convex, therefore, a unique minimizer exists for $g(x)$, and $g(x)$ is the definition of the proximity operation on function f with t. Therefore, the proximity operator a unique result for all y.

(c)

Objective: Compute $\underset{t,f}{\text{prox}}(y)$ and f_t when $f(x) = \|x\|_1$.

Proof:
By definition of the proximity operator, we have:

$$\underset{t,f}{\text{prox}}(y) = \underset{x}{\text{argmin}}\left(\frac{1}{2t}\|x-y\|^2 + \|x\|_1\right) = \underset{x_1,x_2\cdots x_n}{\text{argmin}}\left(\sum_{i=1}^{n}\left(\frac{1}{2t}(x_i-y_i)^2 + |x_i|\right)\right) \tag{1c1}$$

Observe that x_i is independent to each other and that means:

$$\forall 1 \leq i \leq n \ \underset{x_i}{\text{argmin}}\left(\frac{1}{2t}(x_i-y_i)^2 + |x_i|\right) \tag{1c2}$$

Let's consider the scalar optimization problem by cases:

1. $x > 0$:

   $$\frac{d}{dx}\left(\frac{1}{2t}(x_i-y_i)^2 + |x_i|\right) = \frac{1}{t}(x_i-y_i) + 1$$

   Setting the derivative to zero we have:

   $$\frac{1}{t}(x_i-y_i) + 1 = 0 \implies x_i - y_i + t = 0 \implies x_i = y_i - t$$

   Notice that, it assert constraint on y_i for which: $y_i - t > 0 \implies y > t$.
   Subtitutes it back we can get the objective value to as:

   $$\begin{aligned}&\frac{1}{2t}(y_i - t - y_i)^2 + y_i - t\\ &= \frac{t^2}{2t} + y_i - t\\ &= \frac{t}{2} + y_i - t\\ &= y_i - \frac{t}{2}\end{aligned}$$

2. $x = 0$:

$$\frac{d}{dx}\left(\frac{1}{2t}(x_i - y_i)^2 + |x_i|\right) = \frac{1}{t}(x_i - y_i) + \partial|x_i|$$

Using sub-differential, and setting $x_i = 0$, we know that:

$$0 \in \frac{-y_i}{t} + [-1, 1] \implies 0 \in -y_i + [-t, t] \implies -y_i - t \le 0 \le -y_i + t \implies y_i \in [-t, t]$$

Substituting $x_i = 0$ beck and we have: $\frac{y_i^2}{2t}$.

3. $x < 0$:

$$\frac{d}{dx}\left(\frac{1}{2t}(x_i - y_i)^2 + |x_i|\right) = \frac{1}{t}(x_i - y_i) - 1$$

Setting the derivative to zero:

$$x_i - y_i - t = 0 \implies 0 \ge x_i = y_i + t$$

And it means that: $y_i \le -t$
Substituting $x_i = y_i + t$:

$$\begin{aligned}&\frac{1}{2t}(y_i + t - y_i)^2 - (y_i + t)\\ &= \frac{t}{2} - y_i - t\\ &= -y_i - \frac{t}{2}\end{aligned}$$

Notice that, by assuming cases for x and solve for the optimal x^* under each cases, we got the optimal solution for x^* given different y values, and it can be summarized as:

$$\left(\underset{t,\|\cdot\|_1}{\text{prox}}(y)\right)_i = \underset{x_i}{\text{argmin}}\left(\frac{1}{2t}(x_i - y_i)^2 + |x_i|\right) = \begin{cases}0 & y \in [-t, t]\\ y_i + t & y_i < -t\\ y_i - t & y_i > t\end{cases} \tag{1c3}$$

In addition, we also have a way of computing the envelope for the function for each value x_i. Which means that:

$$\begin{cases}\frac{y_i^2}{2t} & y_i \in [-t, t]\\ y_i - \frac{t}{2} & y > t\\ -y_i - \frac{t}{2} & y < -t\end{cases} \tag{1c4}$$

And, using some element wise operations on vector y, we can get the expression for the envelope of the function like:

$$f_t(y) = \frac{y^2}{2t}\text{sign}(\max(0, t - \text{abs(y)}))) + \text{sign}(\max(0, \text{abs}(y) - t)))(\text{sign}(y - t)y - \frac{t}{2}) \tag{1c4.1}$$

(d)

Objective: Compute the envelope and proximity when non-smooth convex function is a infinity norm ball.
Claim 1d1: The proximity operator on the ∞ norm ball can be reduced to projection and easily evaluated by the following expression (Mentioned as an example from the lecture):

$$\underset{t,f}{\text{prox}}(y) = \underset{x}{\text{argmin}}\left(\frac{1}{2t}\|x - y\|^2 + \delta_{\mathbb{B}_\infty}(x)\right) = \underset{\mathbb{B}_\infty}{\text{proj}}(y) = \min(1, \max(-1, y))$$

Next, consider the following quantity: $\|\min(1, \max(-1, y)) - y\|^2$ is essentially:

$$\||y| - \mathbb{J}\|^2$$

Where $\mathbb{J}$ is a vector full of 1s, and it has the same length as vector y. Combining it with the definition for $f_t(y)$ we have:

$$f_t(y) = \frac{1}{2t}\||y| - \mathbb{J}\|^2 + \|x\|_\infty$$

And this is how I would Compute the f_t and Prox for infinity norm.

Problem (2)

(a)

Objective: Figure out:

$$\begin{aligned} \underset{t,g_s}{\text{prox}}(y) &:= \underset{x}{\text{argmin}}\left(\frac{1}{2t}\|x-y\|^2 + g_s(x)\right) \\ g_t(y) &:= \min_x\left(\frac{1}{2t}\|x-y\|^2 + g_s(x)\right) \\ g_s(x) &:= f(x) + \frac{1}{2s}\|x - x_0\|^2 \end{aligned} \tag{2a0}$$

In terms of prox, envelope wrt to function f.
Strategies: There are 2 ways to do it, completing the square, or by the optimality conditions on the prox and use template matching.
Expanding using definition of $g_s(x)$ we have:

$$\underset{t,g_s}{\text{prox}}(y) := \underset{x}{\text{argmin}}\left(\frac{1}{2t}\|x-y\|^2 + \frac{1}{2s}\|x-x_0\|^2 + f(x)\right) \tag{2a1}$$

Using the optimality condition on the prox operator using sub-differential, we have:

$$\begin{aligned} 0 &\in \frac{1}{t}(x^+ - y) + \frac{1}{s}(x^+ - x_0) + \partial f(x^+) \\ 0 &\in \left(\frac{1}{t} + \frac{1}{s}\right)x^+ - \left(\frac{y}{t} + \frac{x_0}{s}\right) + \partial f(x^+) \\ 0 &\in \left(\frac{1}{t} + \frac{1}{s}\right)\left(x^+ - \left(\frac{y}{t} + \frac{x_0}{x}\right)\left(\frac{1}{t} + \frac{1}{s}\right)^{-1}\right) + \partial f(x^+) \\ 0 &\in \left(\frac{1}{t} + \frac{1}{s}\right)\left(x^+ - (ys + tx_0)(t+s)^{-1}\right) + \partial f(x^+) \\ 0 &\in \left(\frac{1}{t} + \frac{1}{s}\right)\left(x^+ - \frac{ys + tx_0}{t+s}\right) + \partial f(x^+) \end{aligned} \tag{2a2}$$

And notice that, this is the optimality conditions of:

$$\underset{x}{\text{argmin}}\left(\frac{1}{2}\left(\frac{1}{t} + \frac{1}{s}\right)\left\|x - \left(\frac{ys + tx_0}{t+s}\right)\right\|^2 + f(x)\right) \tag{2a3}$$

And observe that, the corresponding proximal operator will be:

$$\underset{t,g_s}{\text{prox}}(y) = \underset{f,h}{\text{prox}}\left(\frac{ys + tx_0}{t+s}\right) = \underset{x}{\text{argmin}}\left(\frac{1}{2h}\left\|x - \left(\frac{ys + tx_0}{t+s}\right)\right\|^2 + f(x)\right) \quad h = \frac{st}{s+t} \tag{2a4}$$

And this is the answer we want. Once the prox is defined, then we can get the envelop by just switching the argmin into min, giving us that:

$$g_t(y) = \min_x \left(\frac{1}{2h}\left\|x - \left(\frac{ys + tx_0}{t+s}\right)\right\|^2 + f(x)\right) = f_h\left(\frac{ys+tx_0}{t+s}\right) \quad h = \frac{st}{s+t} \tag{2a5}$$

(b)

Objective: Find $\text{prox}_{t,f}$ when $f(x) = \|x\|_2$. Notice that, we can take the derivative on the norm Here is the quick justification:

$$\nabla\|x\|_2 = \nabla\sqrt{\|x\|^2} = \frac{1}{2\sqrt{\|x\|^2}}2x = \frac{x}{\|x\|} \tag{2b1}$$

And it leaves with the edge case when $x = 0$ to be handled by the sub-differential. And then we will have:

$$\partial(\|x\|) = \begin{cases} \frac{x}{\|x\|} & x \neq \mathbf{0} \\ \{x : \|x\| \le 1\} & x = \mathbf{0} \end{cases} \tag{2b1.1}$$

And here is the justification for the sub-differential at $x = \mathbf{0}$:

$$\begin{aligned} \|y\| &\ge x + v^T(y - x) \\ \|y\| &\ge v^T y \quad \text{set: } x = \mathbf{0} \\ c &\ge \underbrace{cv^T\widehat{y}}_{y = c\widehat{y}} \end{aligned} \tag{2b1.2}$$

And it's not hard to see that the only v that makes 2b1.2 true for all y is when $\|v\| \le 1$
Therefore, the sub-differential for $\|x\|$ at $x = \mathbf{0}$ is $\{x : \|x\| \le 1\}$
By the optimality of the proximal operator, we should be investigating that:

$$\mathbf{0} \in \frac{1}{t}(x - y) + \partial(\|x\|)$$

Which is harder, but we can make it easier by noticing the fact that $\widehat{x}$ is $x/\|x\|$, and this means that, we can write everything in terms of unit vector $\widehat{x}$, hence assume that $x = c\widehat{x}$, where $c > 0$:

1. $x \neq \mathbf{0}$:

$$\mathbf{0} = \frac{1}{t}(cx - y) + \widehat{x} \tag{2b2}$$

$$\begin{aligned} \mathbf{0} &= (c\widehat{x} - y) + t\widehat{x} \\ \mathbf{0} &= (c + t)\widehat{x} - y \end{aligned}$$

$$y = (c+t)\widehat{x} \tag{2b2.1}$$

$$\begin{aligned} c + t &= \|y\| \\ c &= \|y\| - t \end{aligned}$$

   Notice that 2b2.1 implies that x must point to the same direction as y. Which means that the optimal x^+ is:

$$x^+ = cx = (\|y\| - t)\widehat{y}$$

   But does this means for the threshold of y? Notice that assumption that $c > 0$, which will mean that:

$$\|y\| - t > 0 \implies \|y\| > t$$

2. $x = \mathbf{0}$ Substituting we have:

$$\begin{aligned} \mathbf{0} &\in \frac{1}{t}(-y) + \partial(\|x\|)|_{x=\mathbf{0}} \\ \mathbf{0} &\in -y + t\,\partial(\|x\|)|_{x=\mathbf{0}} \\ y &\in t\,\partial(\|x\|)|_{x=\mathbf{0}} \\ \|y\| &\leq t \end{aligned} \tag{2b3}$$

And, the optimal $x^+ = \mathbf{0}$, because that is the only value that triggers the sub-differential.

Summarizing it we have the following cases for the proximal operator:

$$\underset{t,\|\cdot\|}{\text{prox}}(y) = \begin{cases} (\|y\| - t)\widehat{y} & \|y\| > t \\ \mathbf{0} & \|y\| \leq t \end{cases}$$

(c)

Objective: Figure out:

$$\underset{t,\frac{1}{2\|x\|^2}}{\text{prox}}(y) = \underset{x}{\text{argmin}}\left(\frac{1}{2t}\|x-y\|^2 + \frac{1}{2}\|x\|^2\right)$$

Notice that the function are all smooth, so we don't need the sub-differential anymore, by just taking the derivative and setting it to zero, we will get our optimal solution, and this is given by:

$$\begin{aligned} \mathbf{0} &= \frac{1}{t}(x-y) + x \\ \mathbf{0} &= \frac{1+t}{t}x - \frac{y}{t} \\ \frac{y}{t} &= \frac{1+t}{t}x \\ \underbrace{x}_{x^+} &= \frac{y}{1+t} \end{aligned} \tag{2c1}$$

(d)

Objective: Figure out:

$$\underset{t,\frac{1}{2\|C\cdot\|^2}}{\text{prox}}(y) = \underset{x}{\text{argmin}}\left(\frac{1}{2t}\|x-y\|^2 + \frac{1}{2}\|Cx\|^2\right)$$

Because this is smooth, therefore we can just take the gradient of it can set it equal to zero to obtain the optimal x^+ and it will be the output for the proximal operator.

$$\begin{aligned} \mathbf{0} &= \frac{1}{t}(x-y) + \frac{1}{2}(C^T 2(Cx)) \\ \mathbf{0} &= \frac{1}{t}(x-y) + C^T Cx \\ \mathbf{0} &= x - y + tCC^T \\ y &= (I + tC^T C)x \\ \underbrace{x}_{x^+} &= (I + tC^T C)^{-1} y \end{aligned} \tag{2d1}$$

And that is the optimal, and using 2a4 we have:

$$\underset{t,g_s}{\text{prox}}(y) = \underset{f,\frac{st}{s+t}}{\text{prox}}\left(\frac{sy + tx_0}{s+t}\right)$$