

# AMATH 515 Homework 2

Cade Ballew

February 16, 2022

## 1 Problem 1

### 1.1 Part a

Let

$$f_t(y) = \min_x \frac{1}{2t} \|x - y\|^2 + f(x).$$

To see that  $f_t$  is convex, consider  $y, z \in \mathbb{R}^n$  and  $\lambda \in [0, 1]$ . Then,

$$\begin{aligned} f_t(\lambda y + (1 - \lambda)z) &= \min_x \frac{1}{2t} \|x - (\lambda y + (1 - \lambda)z)\|^2 + f(x) \\ &= \min_x \frac{1}{2t} \|\lambda(x - y) + (1 - \lambda)(x - z)\|^2 + \lambda f(x) + (1 - \lambda)f(x). \end{aligned}$$

Now, we use the result from homework 1 that norms (and therefore their squares since norms are nonnegative) are convex to get that

$$\begin{aligned} f_t(\lambda y + (1 - \lambda)z) &\leq \min_x \lambda \frac{1}{2t} \|x - y\| + (1 - \lambda) \frac{1}{2t} \|x - z\| + \lambda f(x) + (1 - \lambda)f(x) \\ &\leq \lambda \max_x \frac{1}{2t} \|x - y\| + f(x) + (1 - \lambda) \max_x \frac{1}{2t} \|x - z\| + f(x) \\ &= \lambda f_t(y) + (1 - \lambda)f_t(z). \end{aligned}$$

Therefore,  $f_t$  is convex.

### 1.2 Part b

To see that the prox operator is uniquely defined for any input  $y$  when  $f$  is convex, fix  $y$  and assume that there exist distinct  $x_1, x_2$  such that

$$\frac{1}{2t} \|x_1 - y\|^2 + f(x_1) = \frac{1}{2t} \|x_2 - y\|^2 + f(x_2).$$

Let  $\lambda \in (0, 1)$ . Then, by the convexity of  $f$  and the strict convexity of the 2-norm,

$$\begin{aligned}
& \frac{1}{2t} \|\lambda x_1 + (1 - \lambda)x_2 - y\|^2 + f(\lambda x_1 + (1 - \lambda)x_2) \\
& \leq \frac{1}{2t} \|\lambda(x_1 - y) + (1 - \lambda)(x_2 - y)\|^2 + \lambda f(x_1) + (1 - \lambda)f(x_2) \\
& < \lambda \frac{1}{2t} \|x_1 - y\|^2 + (1 - \lambda) \frac{1}{2t} \|x_2 - y\|^2 + \lambda f(x_1) + (1 - \lambda)f(x_2) \\
& = \lambda \left( \frac{1}{2t} \|x_1 - y\|^2 + f(x_1) \right) + (1 - \lambda) \left( \frac{1}{2t} \|x_2 - y\|^2 + f(x_2) \right) \\
& = \lambda \left( \frac{1}{2t} \|x_1 - y\|^2 + f(x_1) \right) + (1 - \lambda) \left( \frac{1}{2t} \|x_1 - y\|^2 + f(x_1) \right) \\
& = \frac{1}{2t} \|x_1 - y\|^2 + f(x_1).
\end{aligned}$$

Thus, the function

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

cannot be minimized at  $x = x_1, x_2$  since we have found that it is strictly smaller at  $x = \lambda x_1 + (1 - \lambda)x_2$ . Thus, any minimizer of this function must be unique, meaning that  $\text{prox}_{tf}(y)$  is uniquely defined for any input  $y$ .

### 1.3 Part c

Consider  $f(x) = \|x\|_1$ . Note that

$$\begin{aligned}
\text{prox}_{tf}(y) &= \arg \min_x \frac{1}{2t} \|x - y\|^2 + \|x\|_1 \\
&= \arg \min_x \sum_{i=1}^n \frac{1}{2t} (x_i - y_i)^2 + |x_i|
\end{aligned}$$

is separable, so we can minimize element-wise. Namely, we look for

$$\arg \min_{x_i} \frac{1}{2t} (x_i - y_i)^2 + |x_i|$$

by setting the gradient equal to zero. If  $x_i > 0$ , then

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

so  $x_i = -t + y_i$  and  $y_i = x_i + t > t$ . If  $x_i < 0$ , then

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

so  $x_i = t + y_i$  and  $y_i = x_i - t < -t$ . From this, we expect that  $x_i = 0$  when  $-t \leq y_i \leq t$ . To verify this, we look at the subgradient at 0

$$-\frac{1}{t} y_i + \partial |\cdot| (0) = -\frac{y_i}{t} + \{v \mid |y_i| \geq v y_i\}.$$

Zero is in this subgradient precisely when

$$|y_i| \geq \frac{y_i}{t} y_i = \frac{y_i^2}{t},$$

i.e. when  $|y_i| \leq t$  which confirms our earlier result. Thus,

$$\text{prox}_{tf}(y)_i = \begin{cases} -t + y_i, & y_i > t \\ t + y_i, & y_i < -t \\ 0, & |y_i| \leq t. \end{cases}$$

If we let

$$f_t(y) = \sum_{i=1}^n f_t(y)_i,$$

then this enables us to plug these values in for  $x$  and compute

$$f_t(y)_i = \begin{cases} t/2 + |-t + y_i|, & y_i > t \\ t/2 + |t + y_i|, & y_i < -t \\ \frac{y_i^2}{2t}, & |y_i| \leq t. \end{cases}$$

#### 1.4 Part d

Now, let  $f(x) = \delta_{\mathbb{B}_\infty}(x)$ . Then, we know from lecture that

$$\text{prox}_{tf}(y) = \arg \min_x \frac{1}{2t} \|x - y\|^2 + \delta_{\mathbb{B}_\infty}(x) = \arg \min_{x \in \mathbb{B}_\infty} \frac{1}{2t} \|x - y\|^2 = \text{proj}_{\mathbb{B}_\infty}(y),$$

because  $f$  is an indicator function. Page 8 from lecture 9 also tells us that

$$\text{proj}_{\mathbb{B}_\infty}(y)_i = \max\{\min\{y_i, 1\}, -1\}.$$

Thus, the elements of the proximal operator are given by

$$\text{prox}_{tf}(y)_i = \max\{\min\{y_i, 1\}, -1\}$$

for  $i = 1, \dots, n$ . As before, if we let

$$f_t(y) = \sum_{i=1}^n f_t(y)_i,$$

then this enables us to plug these values in for  $x$  and compute

$$f_t(y)_i = \begin{cases} \frac{1}{2t}(1 - y_i)^2, & y_i > 1 \\ \frac{1}{2t}(-1 - y_i)^2, & y_i < -1 \\ 0, & |y_i| \leq 1. \end{cases}$$

## 2 Problem 2

### 2.1 Part a

Suppose  $f$  is convex and let  $g_s(x) = f(x) + \frac{1}{2s}\|x - x_0\|^2$ . Then, by the properties of inner products and completing the square,

$$\begin{aligned}
& \frac{1}{2t}\|x - y\|^2 + f(x) + \frac{1}{2s}\|x - x_0\|^2 \\
&= \frac{1}{2t}(\|x\|^2 - 2\langle x, y \rangle + \|y\|^2) + \frac{1}{2s}(\|x\|^2 - 2\langle x, x_0 \rangle + \|x_0\|^2) + f(x) \\
&= \frac{s+t}{2st}\|x\|^2 - \langle x, \frac{y}{t} + \frac{x_0}{s} \rangle + \frac{1}{2t}\|y\|^2 + \frac{1}{2s}\|x_0\|^2 + f(x) \\
&= \frac{s+t}{2st} \left( \|x\|^2 - \frac{2st}{s+t} \langle x, \frac{sy + tx_0}{st} \rangle \right) + \frac{1}{2t}\|y\|^2 + \frac{1}{2s}\|x_0\|^2 + f(x) \\
&= \frac{s+t}{2st} \left( \|x\|^2 - 2\langle x, \frac{sy + tx_0}{s+t} \rangle \right) + \frac{1}{2t}\|y\|^2 + \frac{1}{2s}\|x_0\|^2 + f(x) \\
&= \frac{s+t}{2st} \left( \|x - \frac{sy + tx_0}{s+t}\|^2 - \|\frac{sy + tx_0}{s+t}\|^2 \right) + \frac{1}{2t}\|y\|^2 + \frac{1}{2s}\|x_0\|^2 + f(x) \\
&= \left( \frac{s+t}{2st} \|x - \frac{sy + tx_0}{s+t}\|^2 + f(x) \right) - \frac{s+t}{2st} \|\frac{sy + tx_0}{s+t}\|^2 + \frac{1}{2t}\|y\|^2 + \frac{1}{2s}\|x_0\|^2.
\end{aligned}$$

Note that only the things inside the parentheses depend on  $x$ , so

$$\begin{aligned}
\text{prox}_{t_g}(y) &= \arg \min_x \frac{1}{2t}\|x - y\|^2 + f(x) + \frac{1}{2s}\|x - x_0\|^2 \\
&= \arg \min_x \left( \frac{s+t}{2st} \|x - \frac{sy + tx_0}{s+t}\|^2 + f(x) \right) - \frac{s+t}{2st} \|\frac{sy + tx_0}{s+t}\|^2 + \frac{1}{2t}\|y\|^2 + \frac{1}{2s}\|x_0\|^2 \\
&= \arg \min_x \left( \frac{s+t}{2st} \|x - \frac{sy + tx_0}{s+t}\|^2 + f(x) \right) = \arg \min_x \left( \frac{1}{2} \frac{1}{\frac{st}{s+t}} \|x - \frac{sy + tx_0}{s+t}\|^2 + f(x) \right) \\
&= \text{prox}_{af}(z)
\end{aligned}$$

where

$$a = \frac{st}{s+t}$$

and

$$z = \frac{sy + tx_0}{s+t}.$$

### 2.2 Part b

Now, let  $f(x) = \|x\|_2$  and let

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

Then,

$$\nabla g(x) = \frac{1}{t}(x - y) + \frac{x}{\|x\|}$$

when  $\|x\| \neq 0$ , i.e. when  $\|x\| > 0$ . Setting the gradient equal to zero to minimize  $g$ ,

$$y = x \left( 1 + \frac{t}{\|x\|} \right).$$

In this case,

$$\|y\| = \left\| x \left( 1 + \frac{t}{\|x\|} \right) \right\| = \left( 1 + \frac{t}{\|x\|} \right) \|x\| = \|x\| + t$$

if we assume  $t \geq 0$ . Thus,  $\|x\| = \|y\| - t$  which we now plug into our expression for the gradient.

$$0 = \frac{1}{t}(x - y) + \frac{x}{\|y\| - t}.$$

Solving this for  $x$ ,

$$x = \frac{y}{1 + \frac{t}{\|y\| - t}} = \frac{y(\|y\| - t)}{\|y\|} = \left( 1 - \frac{t}{\|y\|} \right) y.$$

Now, note that this is valid when  $0 < \|x\| = \|y\| - t$ , i.e. when  $\|y\| > t$ . Otherwise  $\|x\| = 0$ , so we expect  $g(x)$  to be minimized at  $x = 0$  when  $\|y\| \leq t$ . To verify this, we investigate the subgradient at  $x = 0$ .

$$\partial g(0) = \frac{1}{t}(x - y) \Big|_{x=0} + \partial \|\cdot\|(0) = -\frac{y}{t} + \{v \mid \|y\| \geq v^T y\}.$$

Therefore, zero is in this subgradient precisely when

$$\|y\| \geq \frac{1}{t} y^T y = \frac{1}{t} \|y\|^2,$$

i.e. when  $\|y\| \leq t$ . Thus, we can conclude that

$$\text{prox}_{t f}(y) = \arg \min_x g(x) = \begin{cases} \left( 1 - \frac{t}{\|y\|} \right) y, & \|y\| > t \\ 0, & \|y\| \leq t. \end{cases}$$

### 2.3 Part c

Now, let  $f(x) = \frac{1}{2} \|x\|^2$ . If we let  $g(x)$  be given by

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

then

$$\nabla g(x) = \frac{1}{t}(x - y) + x$$

for all  $x$ . Setting this equal to 0 and solving for  $x$ ,

$$x = \frac{y}{t+1}.$$

Because the gradient is defined everywhere, this is precisely the minimizer of  $g$  in all cases. Thus,

$$\text{prox}_{tf}(y) = \arg \min_x g(x) = \frac{y}{t+1}.$$

## 2.4 Part d

Now, let  $f(x) = \frac{1}{2}\|Cx\|^2$  and let  $g(x)$  be given by

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

Then,

$$\nabla g(x) = \frac{1}{t}(x - y) + C^T Cx$$

for all  $x$ . Setting this equal to zero and solving for  $x$  (and implicitly assuming that  $I + tC^T C$  is invertible),

$$x = (I + tC^T C)^{-1}y.$$

Because the gradient is defined everywhere, this is precisely the minimizer of  $g$  in all cases. Thus,

$$\text{prox}_{tf}(y) = \arg \min_x g(x) = (I + tC^T C)^{-1}y.$$

## 3 Problem 4

### 3.1 Part b

We do appear to be able to recover the signal when the proximal gradient algorithm is applied to the sparse regression problem.

### 3.2 Part c

Applying the accelerated proximal gradient to this same problem appears to have faster convergence.

## 4 Problem 5

### 4.1 Part b

With the logistic regression on MINST data, Newton's method appears to converge the fastest while gradient descent converges the slowest (note that proximal gradient also reaches the maximum number of iterations).

## 4.2 Part c

We achieve 100% accuracy when classifying the test data.