Name:

AMATH 515 Homework Set 2

(1) Recall that

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

Suppose f is convex.

(a) Prove that f_t is convex.

(b) Prove that prox_{tf} is a single-valued mapping.

(c) Compute prox_{tf} and f_t, where $f(x) = \|x\|_1$.

(d) Compute prox_{tf} and f_t for $f = \delta_{\mathbb{B}_\infty}(x)$, where $\mathbb{B}_\infty = [-1,1]^n$.

(2) More prox identities.

(a) Suppose f is convex and let $g(x) = f(x) + \frac{1}{2}\|x - x_0\|^2$. Find formulas for prox_{tg} and g_t in terms of prox_{tf} and f_t.

(b) The elastic net penalty is used to detect groups of correlated predictors:

$$g(x) = \beta\|x\|_1 + (1-\beta)\frac{1}{2}\|x\|^2, \quad \beta \in (0,1).$$

Write down the formula for prox_{tg} and g_t.

(c) Let $f(x) = \frac{1}{2}\|Cx\|^2$. Write $\mathrm{prox}_{tf}(y)$ in closed form.

(d) Let $f(x) = \|x\|_2$. Write $\mathrm{prox}_{tf}(y)$ in closed form.

Coding Assignment

Please download `515Hw2_Coding.ipynb` `solvers.py` and `mnist01.npy` to complete the coding problem (3), (4) and (5).

(3) Complete three generic solvers we learned from the class in `solvers.py`, including,
- proximal gradient descent,
- accelerated gradient descent.
- accelerated proximal gradient descent.

(4) Compressive sensing, consider the sparse regression problem,

$$\min_x \ \frac{1}{2}\|Ax - b\|^2 + \lambda\|x\|_1$$

where $A \in \mathbb{R}^{m \times n}$ and $m < n$ and this is an under determine system. Fortunately, we have the prior knowledge of x being sparse, by adding the ℓ_1 regularizer, we could recover the original signal.

Remark: we choose $\lambda = \|A^\top b\|_\infty / 10$, the reason of it will be more clear when come to duality.

(a) By treating $f(x) = \frac{1}{2}\|Ax - b\|^2$ and $g(x) = \lambda\|x\|_1$, complete the function w.r.t. to f and g.

(b) Apply the proximal gradient algorithm, can you recover the signal?

(c) Apply the accelerated proximal gradient algorithm, is it faster compare to (b)?

(5) Logistic regression on MNIST data, recall the logistic regression problem,

$$\min_x \ \sum_{i=1}^{m} \{\ln(1 + \exp(\langle a_i, x\rangle)) - b_i\langle a_i, x\rangle\} + \frac{\lambda}{2}\|x\|^2.$$

We will try to use logistic regression to classify the "0" and "1" images from MNIST.

In this specific example, a_i is our image (vectorized), and b_i is the corresponding label. By solving the above optimization problem, we want to obtain an classifier, so that for a new image a_{new}, we could say

$$\begin{cases} a_{\text{new}} \text{ is a } 0, & \text{if } \langle a_{\text{new}}, x\rangle \le 0 \\ a_{\text{new}} \text{ is a } 1, & \text{if } \langle a_{\text{new}}, x\rangle > 0 \end{cases}.$$

(a) Complete the function, gradient and Hessian for the logistic regression.

(b) Apply gradient, accelerate gradient and Newton's method to solve the problem. Which one is the fastest and which one is the slowest?

(c) What is your accuracy of the classification for the test data.