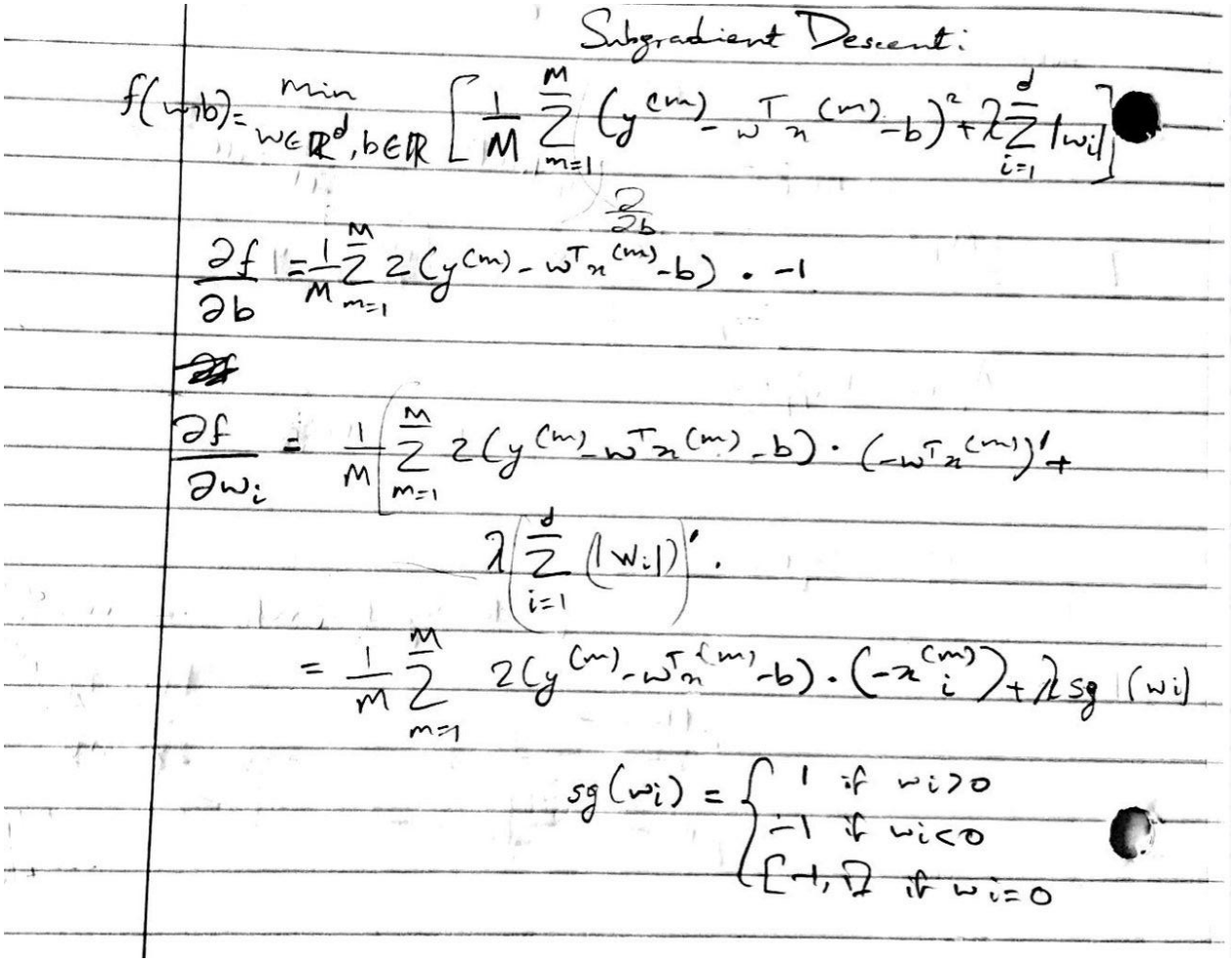
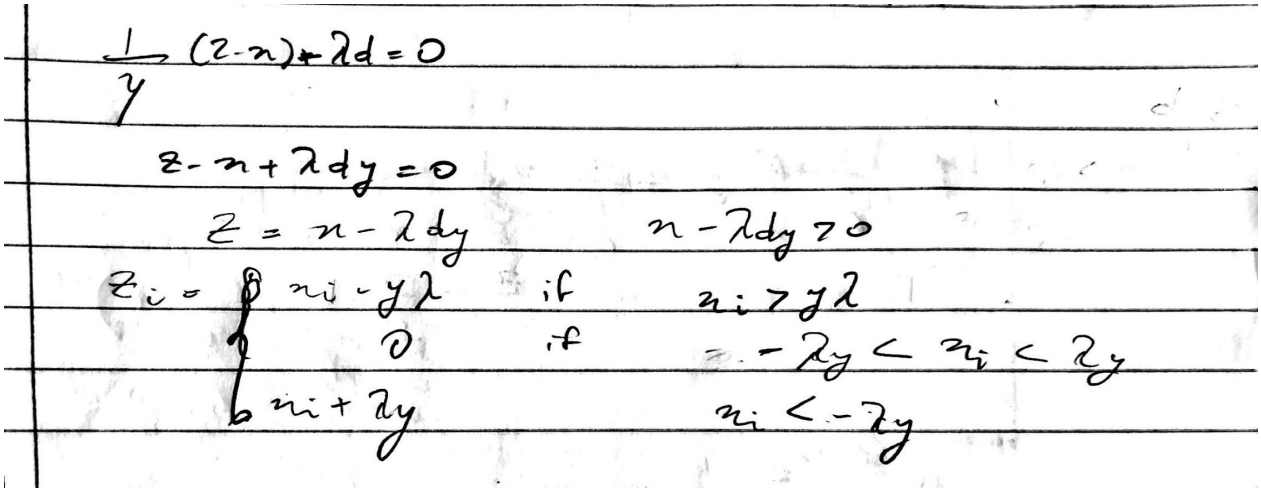
[Special Topics] Homework 5
Sudarshana Jagadeeshi
CS 4301.001

1.
Subgradient is:



https://colab.research.google.com/drive/1UMqkdz6VEnTBaFyPb26AaJw7Lviq-XMm?usp=sharing

2.
Prox is:



https://colab.research.google.com/drive/1I5MaDIFTWjXe0nkbTAWDTFq6APjtta5n?usp=sharing

3.
It seems that proximal gradient descent works best, especially as lambda grows. It also converges far faster. I'm not sure why the error is different between the methods however.

lambda= 0.5:
SGD Best Value: 26.762535289480475 in 78 iterations
Proximal Best Value: 17.417805312443992 in 62 iterations

lambda= 2:
SGD Best Value: 94.53712638013762 in 78 iterations
Proximal Best Value: 31.75211773369509 in 37 iterations

4.
The new function h(w) is differentiable, so we can find a closed form solution:

$$\arg\min_w \frac{1}{2\gamma}\|w-x\|_2^2 + \frac{1}{2} w^T A w + b^T w$$

The gradient is:

$$\frac{1}{\gamma}(w-x) + (Aw+b) = 0$$

$$(w-x) + \gamma A w + \gamma b = 0$$

$$w(1+\gamma A) - x + \gamma b = 0$$

$$w = \frac{x-\gamma b}{(I+\gamma A)^{-1}}$$

The update:

$$w^{(t+1)} = \frac{w^{(t)} - \gamma \nabla f(w^{(t)}) - \gamma b}{(I+\gamma A)^{-1}}$$

Of course for b, the update remains the same regardless of choice of h(w).