1. Given a vector $y$, define the function $f(x)=\|x-y\|^{2}$. Notice $f(x)$ is separable, and therefore we have

$$
[y]_{+}^{i}=\underset{a^{i} \leq x^{i} \leq b^{i}}{\arg \min }\left(x^{i}-y^{i}\right)^{2}= \begin{cases}y^{i}, & \text { if } a_{i} \leq y_{i} \leq b_{i} \\ a^{i}, & y_{i}<a_{i} \\ b^{i} & y_{i}>b_{i}\end{cases}
$$

2. 

(a) At every iteration $k$, ISTA first calculates $h_{k}=x_{k}-\alpha \nabla f\left(x_{k}\right)$ and then apply the shrinkage operation for each entry of $h_{k}$ as follows

$$
x_{k+1}^{j}=\left\{\begin{array}{cl}
h_{k}^{j}-\mu \alpha & \text { if } h_{k}^{j} \geq \mu \alpha \\
0 & \text { if }-\mu \alpha<h_{k}^{j}<\mu \alpha \\
h_{k}^{j}+\mu \alpha & \text { if } h_{k}^{j} \leq-\mu \alpha
\end{array}\right.
$$

where $h_{k}^{j}$ is the $j$-th entry of $h_{k}$.
(b) The implementation steps are as follows:

- Generate the data set using random matrices, choose $\alpha, \mu$, as the step size and the parameter indicating the level of sparsity.
- Find $h_{k}=x_{k}-\alpha \nabla f\left(x_{k}\right)$ and apply the shrinkage operation for each entry of $h_{k}$.
- Iterate until convergence

It was observed that by increasing $\mu$, we will have larger error but more sparse solutions as expected. Moreover, smaller step sizes provides better stability at the expense of more required iterations. Starting from a random point and using $\alpha=0.0001, \mu=1$, we have the sample plots for $(p, m)=(500,160),(p, m)=(500,120),(p, m)=(1500,500)$, and $(p, m)=(1500,350)$ as shown in Figures 1, 2, 3, and 4.


Figure 1. Log error using ISTA


Figure 2. Log error using ISTA


Figure 3. Log error using ISTA


Figure 4. Log error using ISTA

