I implemented LLA + ISTA to optimize the Lasso objective with SCAD penalty and compared it to a standard lasso implemented in the SK-learn library optimized via coordinate descent. I performed a variety of experiments and optimized λ_0 via 5-fold cross validation. Noise is generated from a $\mathcal{N}(0, 1.5)$ distribution. I found the optimal λ_0 to be approximately **0.85**. For my implementation, I utilized the numpy numerical computation library. I have uploaded my implementation and code to run the experiments here: `https://github.com/choltz95/LASSO-SCAD`

The data in the following tables are aggregated over 200 simulations with $\lambda_0 = 0.85$.

1 P1

Cell entries are of the form l2-error/false pos/false neg

n\d	256	512	1024
100	3.29/2.015/0.335	3.34/3.955/0.355	3.54/7.41/0.57
200	2.88/0.04/0.095	2.87/0.04/0.055	2.96/0.12/0.07

Table 1: LLA + ISTA, $\sigma_{i,i} = 1$

n\d	256	512	1024
100	3.41/1.695/0.425	3.46/3.26/0.5	3.67/6.955/0.68
200	2.95/0.045/0.065	2.90/0.045/0.04	2.98/0.18/0.05

Table 2: Lasso, $\sigma_{i,i} = 1$

In an effort to find parameters which result in a more signficant improvement over lasso, I found that by artificially increasing the variance of the covariates improves results of ISTA + LLA over standard CD:

n\d	256	512	1024
100	0.849/0.345/0.0	0.884/0.77/0.0	1.021/2.56/0.0
200	0.691/0.0/0.0	0.723/0.01/0.0	0.736/0.0/0.0

Table 3: LLA + ISTA, $\sigma_{i,i} = 1.5$

n\d	256	512	1024
100	1.63/4.02/0.005	1.68/6.86/0.0	1.83/12.37/0.005
200	1.40/0.12/0.0	1.41/0.215/0.0	1.39/0.515/0.0

Table 4: Lasso, $\sigma_{i,i} = 1.5$

2 P2

n\d	256	512	1024
100	4.31/1.27/2.71	4.29/2.225/2.795	4.27/3.66/2.755
200	3.98/0.015/2.21	3.99/0.025/2.245	3.90/0.075/2.21

Table 5: LLA + ISTA, $\sigma_{i,j} = 0.5^{|i-j|}$

From the above tables, we see that while neither algorithm performs well, the standard lasso-based formulation is more robust to correlated covariates. We can conclude that neither formulation is applicable to the case where covariates are correlated and we can hypothesize that decorrelation of the variables as a preprocessing step may have a marked effect on their effectiveness.

n\d	256	512	1024
100	3.33/1.15/1.4	3.34/2.395/1.1495	3.34/3.685/1.5
200	3.088/0.025/0.99	3.15/0.02/0.975	3.11/0.09/0.975

Table 6: Lasso, $\sigma_{i,j} = 0.5^{|i-j|}$