Programming Question - A review

We conducted experiments as part of a test and comparison of three optimization methods: SVRG, SGD, and Mini-Batch SGD.
To ensure a comprehensive evaluation, we performed the experiments on two matrix sizes: 10x10 and 50x50.

It was apparent that the larger matrix performed less effectively than the smaller one in terms of orders of magnitude. However, the observed trend and the gaps between the various methods remained consistent, and I will now provide a detailed explanation for each method:

SVRG

Multiple experiments were conducted for the SVRG method. Initially, the matrix A was loaded with random values without attempting to control the conditioning number of A. Unfortunately, the performance was far from satisfactory, regardless of the chosen step size or the number of batches.
To improve the results, we adjusted the conditioning number of matrix A to a reasonably favorable value.
After employing various trial and error techniques, we achieved a conditioning number close to 1, which yielded significant enhancements.
In addition, to explore the theoretical concepts covered in lecture 8, we decided to determine the step size and number of batches based on the lemma discussed in class. Since A is positively definite, we calculated the alpha factor for strong convexity and the beta factor for smoothness, as demonstrated in lecture 8. Consequently, the SVRG method converged several times faster, and the function F values associated with the generated X values also exhibited substantial decay.

When utilizing a 50x50 matrix, the minimum recorded F value was significantly lower compared to the algorithms that will be discussed later. Moreover, for a 10x10 matrix, the F values approached 0 after just 10 epochs.

SGD

In the case of SGD, there was a decrease in F values, albeit not a significant one. Towards the later epochs, the F values even increased compared to the previous iterations.
Since the SGD method is already well-known and relatively basic, we did not find it necessary to employ different techniques to optimize the step size. Hence, we proceeded to review the next method: Mini-Batch SGD, referred to as MB SGD from this point forward.

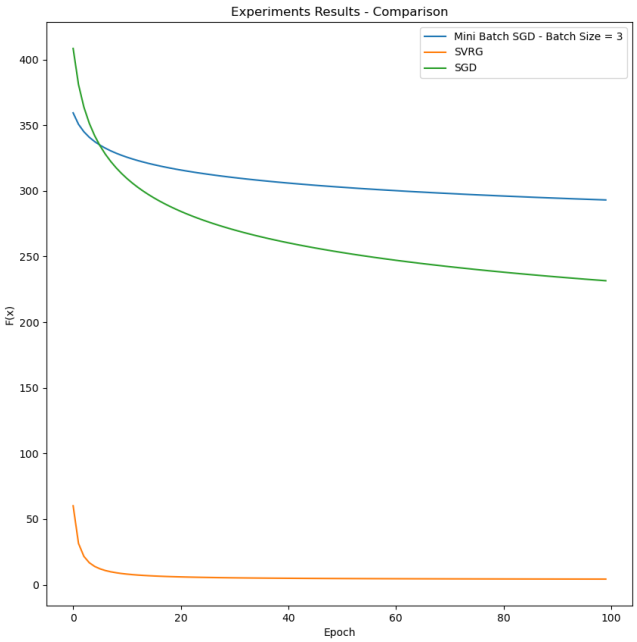
Mini-Batch SGD

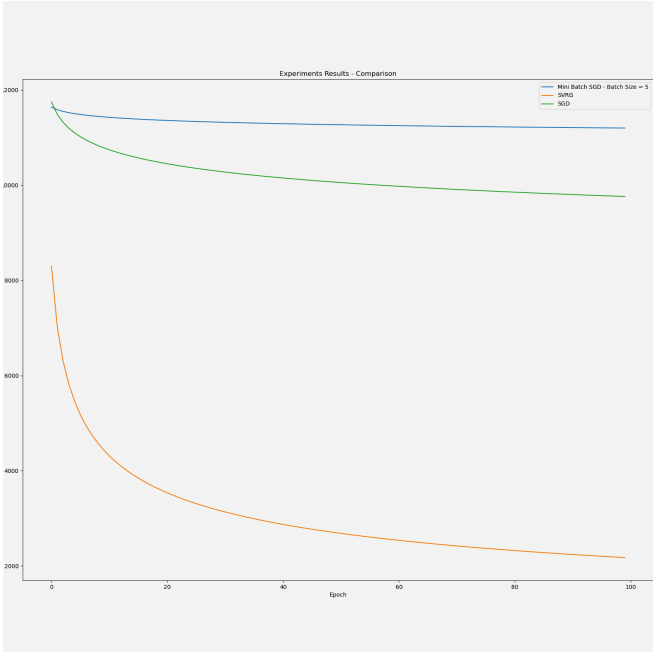
Regarding MB SGD, our observations were mixed.
For both the 50x50 and 10x10 matrices, MB SGD did not perform as well as SVRG but certainly showed more impressive results than the basic SGD.

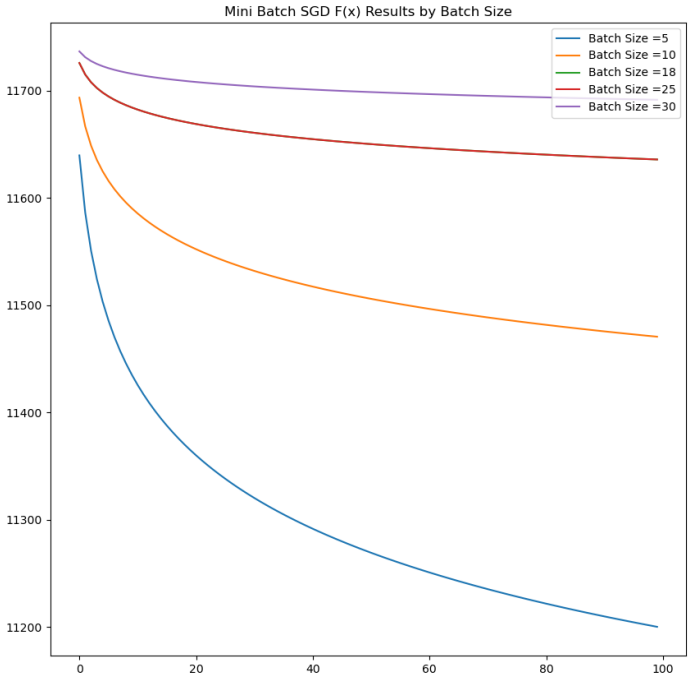
We experimented with different batch sizes and noticed that selecting smaller batch sizes led to better performance. Specifically, as the batch size decreased, the F values converged more rapidly during each epoch.
However, the convergence achieved by MB SGD was not on par with SVRG, irrespective of the matrix size, but it was definitely superior to basic SGD.

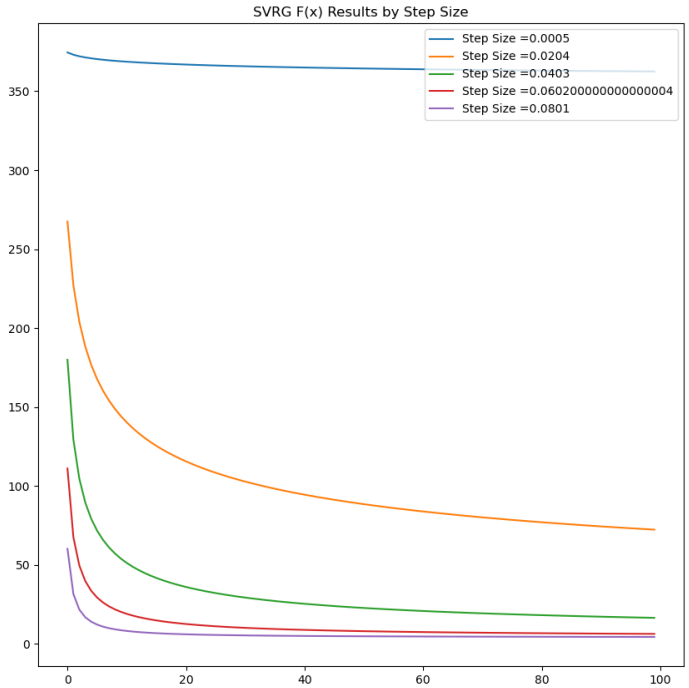
The results of the experiments are depicted in the following graphs, which illustrate the differences in convergence speed to the optimal solution among the different methods.

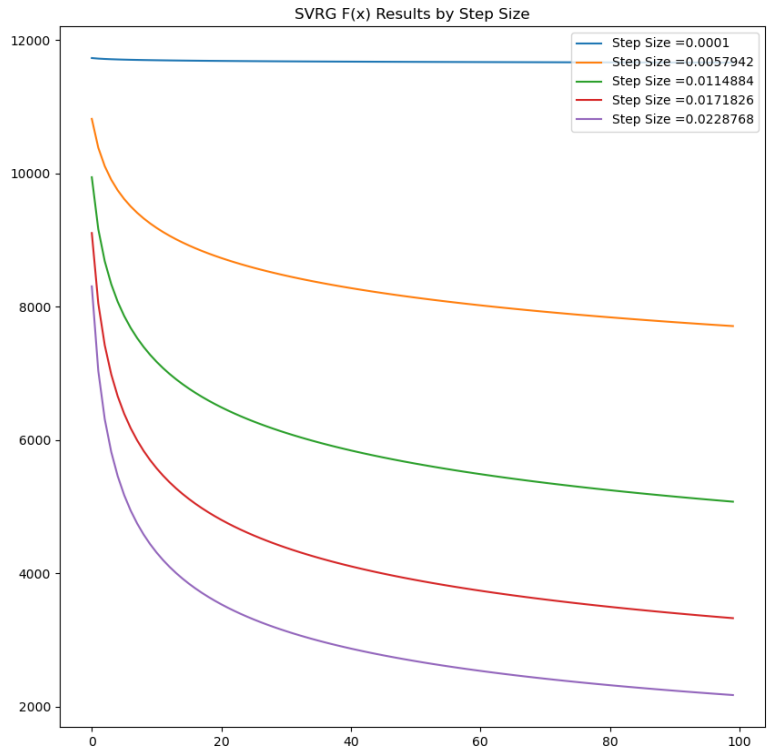
In conclusion, regardless of the matrix size, it is evident that SVRG was the most effective method, exhibiting satisfactory performance regardless of the other methods. It was followed by MB SGD, and finally the SGD method.

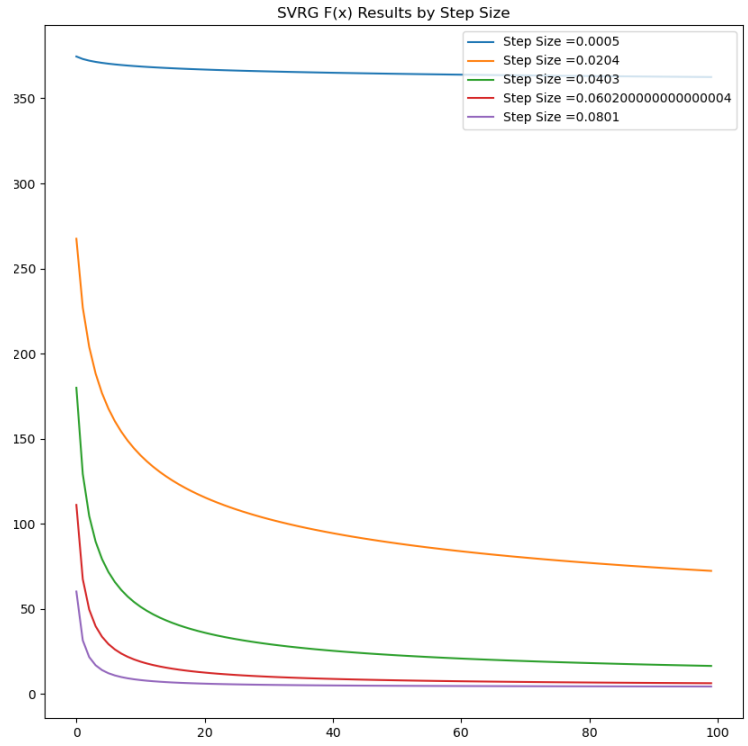












SVRG Without setting A's condition number