EP3260 Machine Learning over Networks
Review of HW3 for Group 1

February 27, 2019

1 HW3(a)

The proof is correct and presented in a clear format.

2 HW3(b)

The proof is correct.

Good that the authors provided the equivalence between strongly convex and Lipschitz continuity of the gradient for f and $f^\star$ as a Lemma. However, the authors could have defined u and v as the variables related to the conjugate function definition at two different points $g_u(\mathbf{x}) = f(\mathbf{x}) - \mathbf{u}^\mathsf{T}\mathbf{x}$ and $g_v(\mathbf{x}) = f(\mathbf{x}) - \mathbf{v}^\mathsf{T}\mathbf{x}$. Then, it is easy to see that the minimizer $\mathbf{x}_v = \nabla f^\star(\mathbf{v})$ and $\mathbf{x}_u = \nabla f^\star(\mathbf{u})$. It would be more pedagogical and easier for the reader to understand the proof of Lemma 1.2.1.

3 HW3(c)

The solution seems to be correct. The references provided to show the convergence of primal and dual were very helpful.

4 CA 3

The problem is solved and clear figures are given to compare the convergence time and rate of different solvers with the two data sets.

It will be better if the report could contain more details about the process of generating the training code and the authors can analyze more based on the results from the figure.

If fixing the outer iterations, SVRG is expected to take the longest computation time as in the figure for the community dataset. The authors are suggested to explain the reason why it takes less time compared to BCD, GD, pGD for the power dataset.