

As the reviewer suggested, we conduct experiments to show the performance of our methods with different hyperparameters, where the dataset is a9a and the task is hyperparameter optimization in our Eq. (11).

In Figure 9, we show the effect of β_i (where $i \in \{1, 2, 3\}$) on the convergence performance of our DSVRBGD-S and DSVRBGD-A. Specifically, we set $\beta_i = \{0.1, 0.3, 0.5, 0.7, 0.9\}$ while fixing $\eta = 0.05$. Since the actual learning rate is $\beta_i \eta$ from the global view as $\bar{X}_{t+1} = \bar{X}_t - \beta_1 \eta \bar{P}_t$, a smaller β_i results in a smaller learning rate. Therefore, a smaller β_i leads to a slower convergence rate as shown in Figure 9.

In Figure 10, we show the effect of α_i (where $i \in \{1, 2, 3\}$) on the convergence performance of our DSVRBGD-S and DSVRBGD-A. In this experiment, we fix $\eta = 0.05$ and then vary α_i such that $\alpha_i \eta^2 = \{0.1, 0.3, 0.5, 0.7, 0.9\}$. It can be observed that it does not affect the convergence rate significantly.

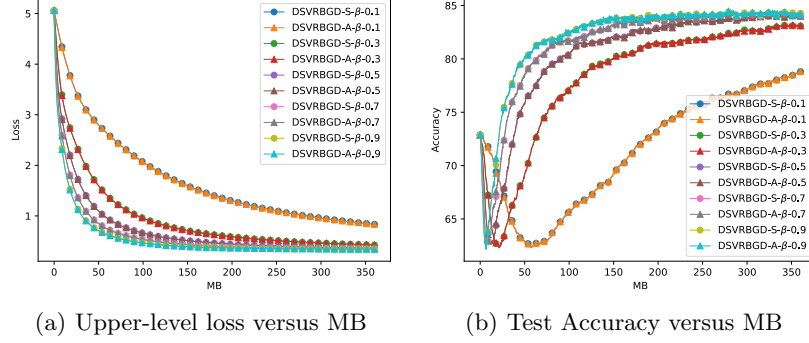


Figure 9: The convergence performance of our DSVRBGD-S and DSVRBGD-A for different values of β_i .

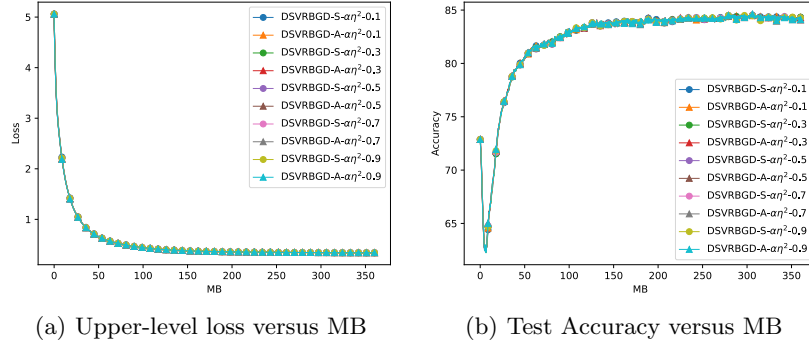


Figure 10: The convergence performance of our DSVRBGD-S and DSVRBGD-A for different values of α_i .