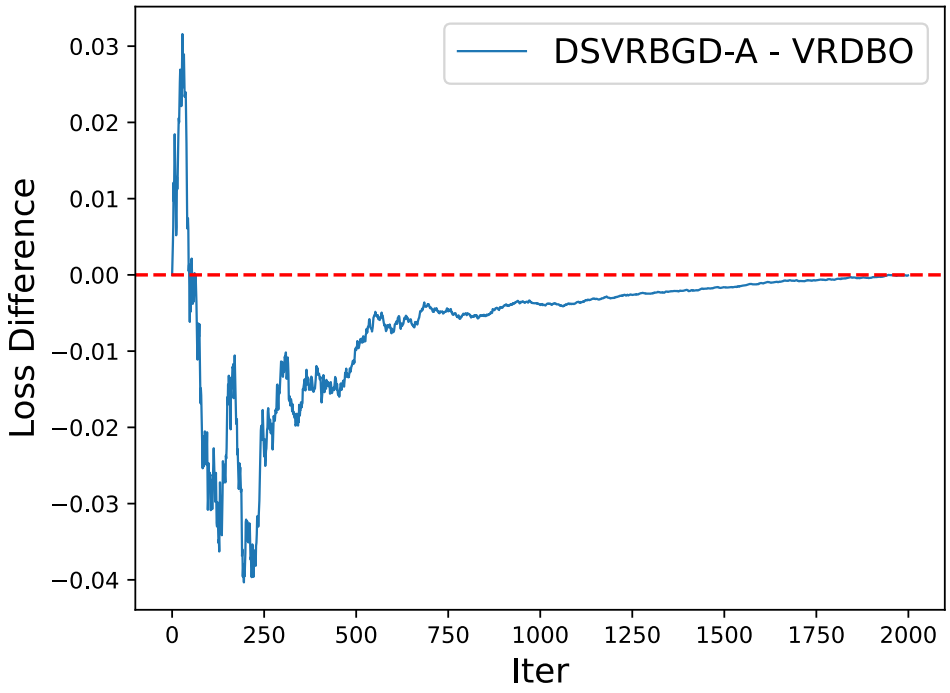
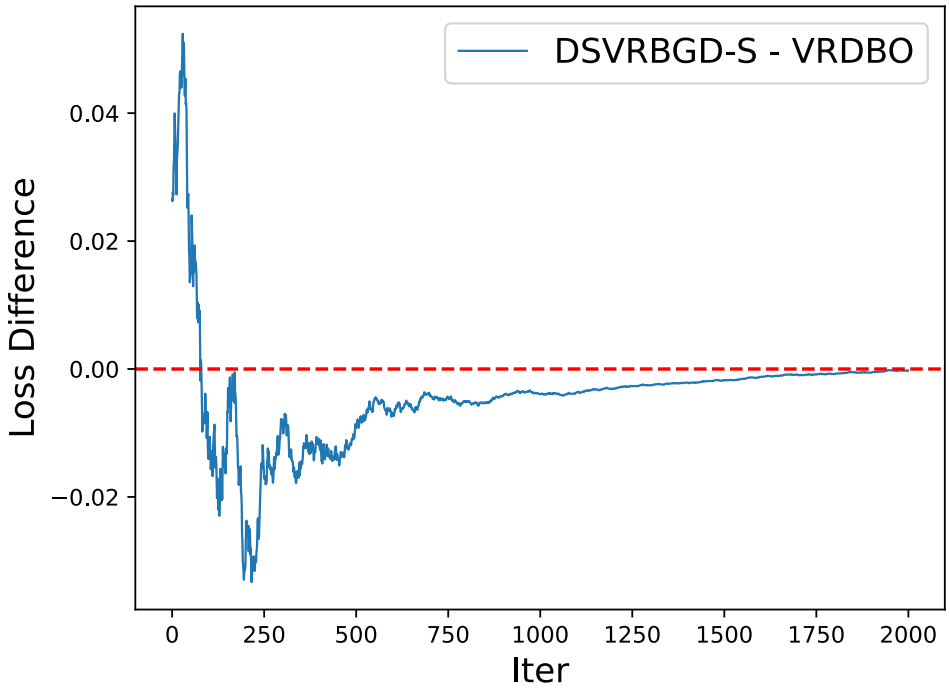


(a) $\text{LOSS}_{\text{DSVRBGD-A}} - \text{LOSS}_{\text{VRDBO}}$



(b) $\text{LOSS}_{\text{DSVRBGD-S}} - \text{LOSS}_{\text{VRDBO}}$

Figure 1: The loss difference between our two methods and VRDBO. **A negative difference indicates that our methods converge faster to the stationary point than VRDBO** (The potential reason for the positive difference in the first several steps is that the one-step gradient descent for estimating Hessian-inverse-vector product is not as good as the Neuman series expansion method in the initial stage.). There are 8 workers in this experiment. The training sample's feature x on the k-th worker is generated from a Guassian distribution $\mathcal{N}(\mu_k, \sigma_k)$, where μ_k is generated from a Uniform distribution $\mathcal{U}(-3, 3)$ and σ_k is generated from a Uniform distribution $\mathcal{U}(1, 25)$.