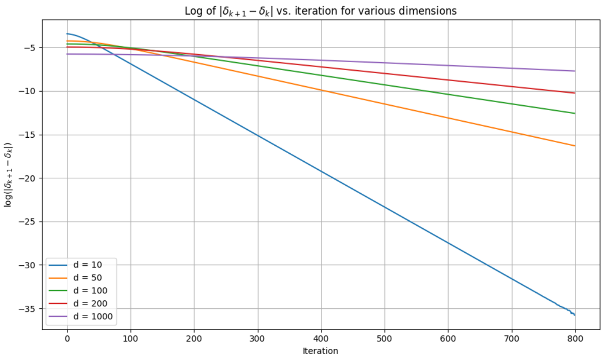
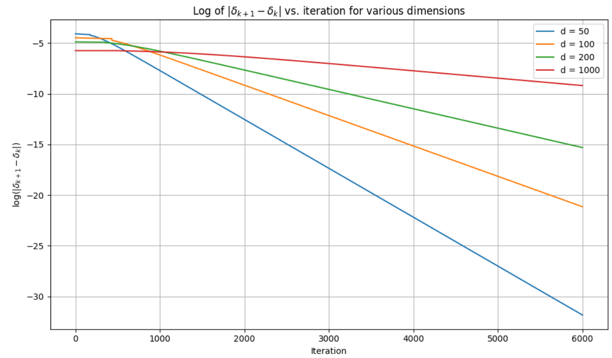
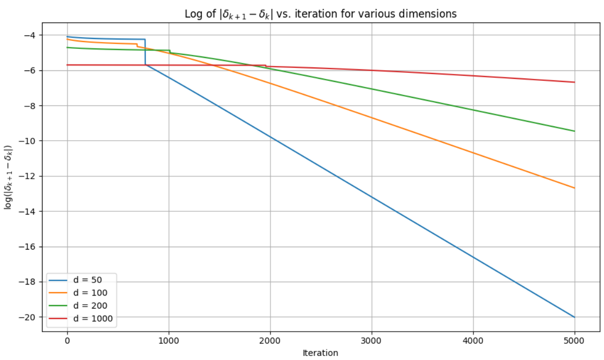


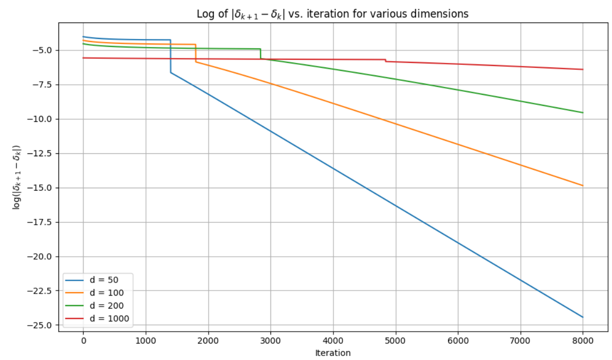
(a) $\epsilon = 1$



(b) $\epsilon = 5$



(c) $\epsilon = 10$



(d) $\epsilon = 20$

Figure S1. The convergence of $\boldsymbol{\delta}_k$ produced by PGD for solving optimization Log-Sum-Exp (1) with different dimensions d. We use a logarithmic coordinate system where the horizontal axis represents the number of iterations k and the vertical axis represents $\log_{10}(\|\boldsymbol{\delta}_{k+1} - \boldsymbol{\delta}_k\|)$.

Our experiments include two additional convex and smooth objective functions that are commonly used in machine learning:

Log-Sum-Exp (used for constructing probability distributions):

$$\Phi(\boldsymbol{\delta}) = \log\left(\sum_{i=1}^{d} \exp(\delta_i)\right). \tag{1}$$

Logistic Regression Loss (used for binary classification):

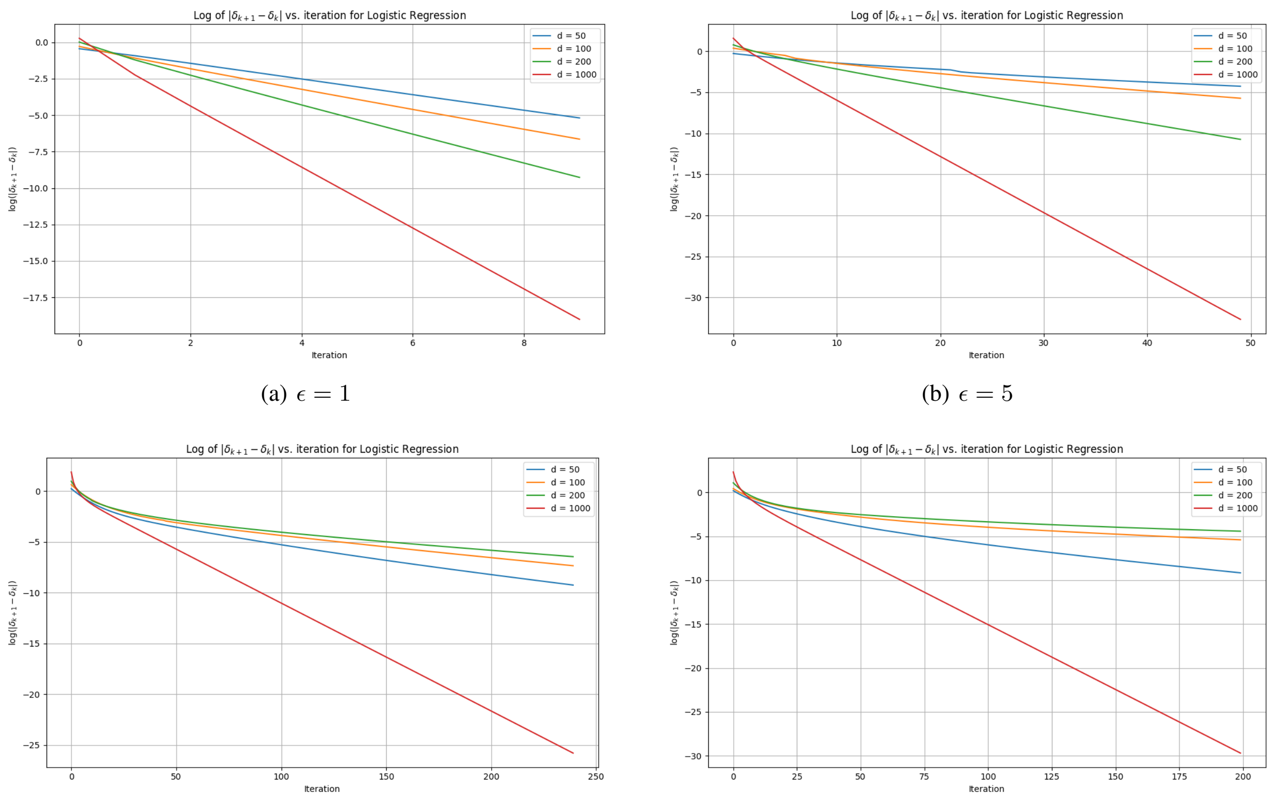
$$\Phi(\boldsymbol{\delta}) = \frac{1}{n}\sum_{i=1}^{n} \log\left(1 + \exp(-y_i \boldsymbol{\delta}^\top \boldsymbol{x}_i)\right), \tag{2}$$

where $\boldsymbol{x}_i$ and y_i denote the input features and binary labels generated for simulation. In both cases, PGD is applied to solve the constrained optimization problem:

$$\min_{\boldsymbol{\delta} \in \mathbb{B}(\mathbf{0}, \epsilon)} \Phi(\boldsymbol{\delta}),$$

where $\Phi(\boldsymbol{\delta})$ represents either the Log-Sum-Exp or Logistic Regression Loss, both of which are convex but not strongly convex. No prior knowledge of the optimal solution $\boldsymbol{\delta}^\triangle$ is assumed, and the initialization is randomly sampled from the constraint ball $\mathbb{B}(\mathbf{0}, \epsilon)$.

We consider various dimensions $d \in \{50, 100, 200, 1000\}$ and constraint radii $\epsilon \in \{1, 5, 10, 20\}$. The experimental results for Log-Sum-Exp and Logistic Regression Loss are shown in Figures S1 and S2, respectively. As illustrated, PGD demonstrates both global convergence and local linear convergence, which is consistent with our theoretical analysis.



(a) $\epsilon = 1$

(b) $\epsilon = 5$

(c) $\epsilon = 10$

(d) $\epsilon = 20$

Figure S2. The convergence of $\boldsymbol{\delta}_k$ produced by PGD for solving optimization Logistic Regression Loss (2) with different dimensions d. We use a logarithmic coordinate system where the horizontal axis represents the number of iterations k and the vertical axis represents $\log_{10}(\|\boldsymbol{\delta}_{k+1} - \boldsymbol{\delta}_k\|)$.

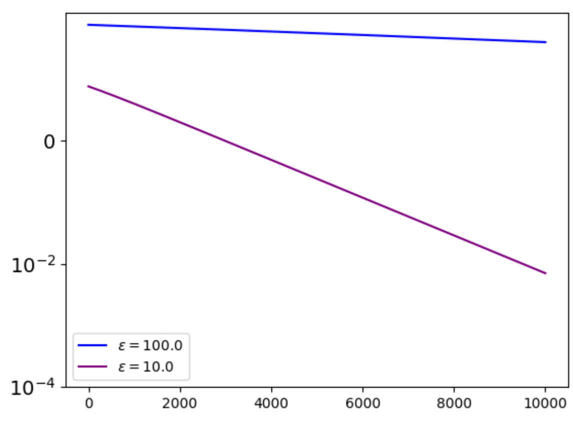


Figure S3. The convergence of PGD under large ϵ for solving the OP in Appendix D. We use a logarithmic coordinate system where the horizontal axis represents the number of iterations k and the vertical axis represents $\log_{10}(\|\boldsymbol{\delta}_k - \boldsymbol{\delta}^{\triangle}\|)$.