

---

# Distributed Learning with Strategic Users: A Repeated Game Approach

---

Anonymous Author(s)

Affiliation

Address

email

## Abstract

1 We consider a distributed learning setting where strategic users are incentivized, by  
2 a cost-sensitive fusion center, to train a learning model based on local data. The  
3 users are not obliged to provide their true gradient updates and the fusion center  
4 is not capable of validating the authenticity of reported updates. Thus motivated,  
5 we formulate the interactions between the fusion center and the users as repeated  
6 games, manifesting an under-explored interplay between machine learning and  
7 game theory. We then develop an incentive mechanism for the fusion center based  
8 on a joint gradient estimation and user action classification scheme, and study its  
9 impact on the convergence performance of distributed learning. Further, we devise  
10 an adaptive zero-determinant (ZD) strategy, thereby generalizing the celebrated ZD  
11 strategy to the repeated games with time-varying stochastic errors. Theoretical and  
12 empirical analysis show that the fusion center can incentivize the strategic users to  
13 cooperate and report informative gradient updates, thus ensuring the convergence.

## 14 1 Introduction

15 Distributed machine learning is becoming increasingly important in large-scale problems with data-  
16 intensive applications [18, 22, 26, 39]. Notably, federated learning has emerged as an attractive  
17 distributed computing paradigm that aims to learn an accurate model without collecting data from the  
18 owners and storing it in the cloud: The training data is kept locally on the computing devices which  
19 participate in the model training and report gradient updates (or its variants) based on local data [19].

20 In this work, we study a distributed learning scheme in which privacy-aware *users* train a global model  
21 with a *fusion center*. We consider the users to be rational, self-interested and risk-neutral. The users  
22 are not compelled to contribute their resources unconditionally, unless they are sufficiently rewarded,  
23 and the system may reach a noncooperative Nash equilibrium where the users do not participate in  
24 training. This departs from conventional distributed learning schemes where the agents directly follow  
25 the lead of the fusion center (FC)<sup>1</sup> and send their gradients. Since the users are strategic, a paramount  
26 objective for the FC is to *design an effective reward mechanism to incentivize self-interested users to*  
27 *provide informative gradient updates*. The repeated game enriches the distributed learning framework  
28 with the idea of many agents interacting within a common uncertain environment, and this framework  
29 provides a new perspective to specify how agents can strategically choose the learning updates how  
30 the resulting changes impact the performance of the learning efforts.

31 **Challenges and Contributions.** There are a number of challenges in distributed learning with  
32 strategic users. First, the users are not obliged to entirely dedicate their resources and they may not  
33 fulfill their roles in the training of the algorithm if it were not for their own interest. Secondly, the  
34 FC cannot directly validate data driven gradient updates due to their stochastic nature. The quality

---

<sup>1</sup>We refer to the fusion center as “she” and a user as “he”.

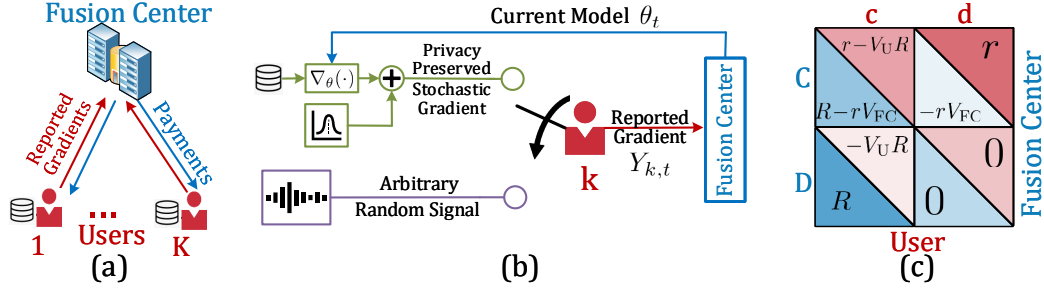


Figure 1: The fusion center (FC) trains the learning model with strategic users who are not obliged to report their gradients. (a) The objective of the FC is to incentivize users to cooperate by giving rewards so as to learn the model. (b) If the user is cooperative, he reports a privacy-preserved version of his gradient signal. Otherwise, the user is defective and sends an arbitrary uninformative signal. (c) The FC and the user each choose to cooperate or defect with respective payoffs as shown.

of the updates can vary over time and across the users since each user can control his own dataset. The interactions among users and the FC are repeated, and each user is capable of devising intricate strategies based on the past interactions. From a game-theoretic perspective, the fusion center's ability to reciprocate against non-cooperative user actions is significantly restricted since she cannot directly observe the user actions. Finally, the FC is not allowed to impose penalties on the users and positive rewards are the only options at her disposal to incentivize user participation. The work proposed here is, to the best of our knowledge, the first distributed learning framework to consider these challenges.

In this study, we model the interactions (in terms of gradient reporting and reward) between the FC and the users as repeated games, which intertwine with the updates in distributed learning. We propose a reward mechanism for the fusion center, based on an adaptive zero-determinant strategy, thereby generalizing the celebrated ZD strategy to the repeated games with time-varying stochastic errors. To tackle the challenge that the FC cannot directly verify the received reported gradients, we devise a gradient estimation and user action classification. Our findings demonstrate that, by employing adaptive ZD strategies, the FC can incentivize the strategic users to cooperate and report informative gradient updates, thus ensuring the convergence of distributed learning.

Detailed discussion on related work is relegated to Appendix A, due to space limitation.

## 2 Distributed Learning with Strategic Users as Repeated Games

We consider a distributed learning setting with  $K$  strategic users  $\mathcal{K} = \{1, \dots, K\}$  and a fusion center (FC), and the optimization problem is given as follows:

$$\min_{\theta \in \mathbb{R}^n} F(\theta) := \frac{1}{K} \sum_{k=1}^K \mathbb{E}_{Z_k \sim \mathcal{D}} [\mathcal{L}(\theta; Z_k)], \quad (1)$$

where  $\mathcal{L}(\cdot)$  is the loss function. In each iteration, each user gets a mini-batch of  $s$  i.i.d. samples from an unknown distribution  $\mathcal{D}$ , and computes the stochastic gradient signal as  $X_{k,t} := \frac{1}{s} \sum_{i=1}^s \nabla_{\theta} \mathcal{L}(\theta_t; z_{k,t}^i)$ , where  $z_{k,t}^i$  is the  $i^{\text{th}}$  sampled data of user  $k$  at time  $t$ .

**Stage Game Formulation: Actions and Payoffs.** The action and the reported signal of user  $k$  in iteration  $t$  are denoted with  $B_{k,t} \in \{c, d\}$  and  $Y_{k,t}$ , respectively. As depicted in Fig. 1, a user is cooperative ( $B_{k,t} = c$ ) if he is sending the privacy-preserved version of his gradient  $X_{k,t}$ . Otherwise, the user is defective and sends a noise signal  $Y_{k,t} \sim \mathcal{N}(0, \Xi_t)$  independent of  $X_{k,t}$ :

$$Y_{k,t} = \begin{cases} X_{k,t} + N_{k,t}, & \text{if } B_{k,t} = c \text{ (cooperative);} \\ \Upsilon_{k,t}, & \text{if } B_{k,t} = d \text{ (defective).} \end{cases} \quad (2)$$

**Remark 1.** Note that  $N_{k,t}$  is independent of  $X_{k,t}$  and  $N_{k,t} \sim \mathcal{N}(\vec{0}, \nu_t^2 \mathbf{I})$ . If  $\|\nabla_{\theta} \mathcal{L}(\theta; z)\|_2 \leq \ell$  for all  $\theta$  and  $z$ , then this privacy-preservation mechanism enjoys  $\epsilon_t$ -differential privacy, with  $\epsilon_t = \ell^2 / s^2 \nu_t^2$  for mini-batch size  $s$ . The details are provided in Appendix.

The payoff structure of a single interplay between the fusion center and a user is depicted in Fig 1b. In iteration  $t$ , when a user cooperates, he provides an information gain  $R$  to the FC at his privacy cost  $V_U R$  with  $0 < V_U \leq 1$ . When a user defects, he does not provide any information gain and does not incur any privacy cost. The FC may distribute rewards at the end of each iteration to incentivize the users. We denote the action of the FC toward user  $k$  as  $A_{k,t} \in \{C, D\}$ . The FC is cooperative ( $A_{k,t} = C$ ) if she makes a payment  $r$  to the user at her cost  $rV_{FC}$  with  $0 < V_{FC} \leq 1$ . The FC is defective ( $A_{k,t} = D$ ), if she does not make any payment to the user. The factor  $V_{FC}$  captures the difference in the valuation of the reward between the FC and the user; for instance, the reward can be a coupon which may be redeemed in the future. Denote the FC's payoff vector by  $\mathbf{S}_{FC} = [R - rV_{FC}, -rV_{FC}, R, 0]$  and that of the users by  $\mathbf{S}_U = [r - V_U R, r, -V_U R, 0]$ . In this paper, we only analyze the case where  $R > rV_{FC}$  and  $r > V_U R$ . Otherwise, the FC or users do not have any incentive to cooperate.

The FC cannot observe the actions of the users and her realized payoffs. We assume that users do not communicate or collude with each other. They cannot observe the actions of other users and the actions of the FC toward other users. Next, we will discuss how to devise effective strategies for the FC to incentivize cooperative user action for the repeated game in a cost-effective manner.

**Repeated Games between Users and Fusion Center.** A salient feature of  $2 \times 2$  repeated games is that players with longer memories of the history of the game have no advantage over those with shorter ones when each stage game is identically repeated infinite times [33]. Thus, without loss of generality, we assume the user strategies only depend on the outcomes of the last round. Let  $q_1, q_2, q_3$  and  $q_4$  denote the probabilities of cooperation for the user conditioned on the joint action pair of the previous iteration, that is  $(A_{k,t-1}, B_{k,t-1})$ , in the order of  $(C, c), (C, d), (D, c)$  and  $(D, d)$ . The user's strategy vector is defined as  $\mathbf{q} = [q_1, q_2, q_3, q_4]$ .

Analogous to the user strategies, let  $p_1, p_2, p_3$  and  $p_4$  denote the probabilities of cooperation for the FC conditioned on  $(A_{k,t-1}, B_{k,t})$ , in the order of  $(C, c), (C, d), (D, c)$  and  $(D, d)$ . The fusion center's strategy vector is defined as  $\mathbf{p} = [p_1, p_2, p_3, p_4]$ . The joint action pair of the user and the FC is considered as the state of the game in iteration  $t$ :  $(A_{k,t}, B_{k,t})$ . The strategy vectors  $\mathbf{p}$  and  $\mathbf{q}$  imply a Markov state transition matrix as follows:

$$\Omega = \begin{bmatrix} q_1 p_1 & (1 - q_1) p_2 & q_1 (1 - p_1) & (1 - q_1) (1 - p_2) \\ q_2 p_1 & (1 - q_2) p_2 & q_2 (1 - p_1) & (1 - q_2) (1 - p_2) \\ q_3 p_3 & (1 - q_3) p_4 & q_3 (1 - p_3) & (1 - q_3) (1 - p_4) \\ q_4 p_3 & (1 - q_4) p_4 & q_4 (1 - p_3) & (1 - q_4) (1 - p_4) \end{bmatrix}. \quad (3)$$

Let  $\Lambda^*$  be the stationary vector of the transition matrix  $\Omega$ , i.e.,  $\Lambda^* = \Lambda^* \Omega$ . We can find the expected payoffs of the FC and the user in the stationary state as  $s_{FC}^* = \Lambda^* \mathbf{S}_{FC}^\top$  and  $s_U^* = \Lambda^* \mathbf{S}_U^\top$ . The FC sets her strategy  $\mathbf{p}$  satisfying, for some real values  $\varphi_0, \varphi_1$  and  $\varphi_2$ , the equation

$$[p_1 - 1, p_2 - 1, p_3, p_4] = \varphi_0 \mathbf{S}_{FC} + \varphi_1 \mathbf{S}_U + \varphi_2 \mathbf{1}. \quad (4)$$

This class of strategies are called zero-determinant (ZD) strategies, which enforce a linear relation between the expected payoffs, given by  $\varphi_0 s_{FC}^* + \varphi_1 s_U^* + \varphi_2 = 0$ , regardless of the user strategy [33].

**Remark 2.** The ZD strategy is a powerful tool to incentivize the users cooperation for the FC because she can unilaterally set  $s_U^*$  or establish an extortionate linear relation between  $s_U^*$  and  $s_{FC}^*$ . Against such an FC strategy, the user's best response which maximizes his payoff is full cooperation,  $\mathbf{q}^* = [1 \ 1 \ 1 \ 1]$ . The details are provided in Appendix C.

Against the FC who is equipped with the ZD strategy, the user can increase his expected payoff only by cooperating more often, and consequently his best response is full cooperation. Assuming that there are sufficiently many participating users, the FC has the absolute leverage against any single user who tries to negotiate with her. Nevertheless, the FC cannot directly employ the ZD strategy since she cannot observe the true actions of the users. In the next section, we will study the use of ZD strategy can be extended in the scope of distributed learning.

### 3 Distributed Stochastic Gradient Descent with Strategic Users

For the ease of exposition, in this paper we focus on an interesting variant of the classical stochastic gradient descent algorithm using the gradient signals reported by strategic users (SGD-SU). In each iteration, the FC collects the reported gradients of the users and update the model as follows:

$$\theta_t = \theta_{t-1} - \eta_t \cdot \hat{m}_t(\mathbf{Y}_t), \quad (5)$$

**Algorithm 1:** Stochastic Gradient Descent with Strategic Users (SGD-SU)

```

1 for  $t = 1, 2, \dots, T - 1$  do
2   Fusion Center: broadcast the current iterate  $\theta_{t-1}$  to all the users
3   forall  $k \in \{1, 2, \dots, K\}$  do
4     User  $k$ : compute the gradient  $X_{k,t}$  and  $Y_{k,t} \leftarrow \begin{cases} X_{k,t} + N_{k,t} & \text{cooperative action,} \\ \Upsilon_{k,t} & \text{defective action,} \end{cases}$ 
5     Fusion Center: form the gradient estimate  $\hat{m}_t(\mathbf{Y}_t) \leftarrow \frac{1}{K(\Lambda_1 \Omega^{t-1}) \mathbf{q}^\top} \sum_{k=1}^K Y_{k,t}$ 
6     update model parameter  $\theta_t \leftarrow \theta_{t-1} - \eta_t \hat{m}_t(\mathbf{Y}_t)$ 
7     classify the users  $\hat{B}_{k,t}(\hat{m}_t, Y_{k,t}) \leftarrow \begin{cases} \hat{c} & \text{(cooperative) if } Y_{k,t}^\top \hat{m}_t > \|\frac{1}{2} \hat{m}_t\|_2^2 \\ \hat{d} & \text{(defective) else} \end{cases} \quad (7)$ 
8     compute the detection and false alarm probabilities using (8) and (11)
9     compute the adaptive strategies (9) and distribute the rewards accordingly

```

where  $\mathbf{Y}_t = [Y_{1,t} \dots Y_{K,t}]$ ,  $\eta_t$  is the step size and  $\hat{m}_t$  is the gradient estimator. The FC cannot directly observe user actions and verify the reported gradients. This gives rise to two coupled challenges:

- The gradient estimator  $\hat{m}_t$  should be resilient against the uninformative reports of defective users.
- Although the ZD strategies are powerful tools to incentivize user cooperation, the FC cannot directly employ a ZD strategy because she cannot observe the users' actions.

To tackle these difficulties, we will first introduce a gradient estimation and user classification scheme and discuss the impact of user action classification errors on the dynamics of repeated games. As outlined in Algorithm 1, we will develop adaptive FC strategies which generalize the classical ZD strategies to the repeated games with time-varying stochastic errors.

### 3.1 Joint Gradient Estimation and User Action Classification

The stochastic gradients can be decomposed as  $X_{k,t} = m_t + W_{k,t}$  where  $m_t := \nabla_{\theta} F(\theta_t)$  is the population gradient and  $W_{k,t}$  is the zero-mean noise term [31]. The unknown parameter  $m_t$  is the mean of the reported gradient  $Y_{k,t}$  when the user is cooperative ( $B_{k,t} = c$ ). The defective users send zero-mean random noise as their reported gradients. The FC needs to classify the reported gradients and obtain an estimate of  $m_t$  for the SGD-SU update in (5). These two problems are coupled with each other, and the joint scheme is, therefore, comprised of a gradient estimator  $\hat{m}_t$ , and a classification rule  $\hat{B}_{k,t}$ . To tackle this difficult problem, we first investigate gradient estimation. Let  $\Lambda_1$  be the initial state distribution of the games between the users and the FC. A modified empirical mean based gradient estimator can be employed as follows:

$$\hat{m}_t(\mathbf{Y}_t) := \frac{1}{K(\Lambda_1 \Omega^{t-1}) \mathbf{q}^\top} \sum_{k=1}^K Y_{k,t}. \quad (6)$$

It is easy to verify that  $\hat{m}_t(\cdot)$  is an unbiased estimator if the FC is able to employ her strategies  $\mathbf{p}$  without any errors and the state distribution of the repeated games are governed by the state transition matrix  $\Omega$  as in (3) without any perturbations.

Using the gradient estimator  $\hat{m}_t(\cdot)$ , the FC can form the user action classification rule as

$$\hat{B}_{k,t}(\hat{m}_t(\mathbf{Y}_t), Y_{k,t}) = \begin{cases} \hat{c} & \text{if } Y_{k,t}^\top \hat{m}_t > \frac{1}{2} \|\hat{m}_t\|_2^2, \\ \hat{d} & \text{else;} \end{cases} \quad (7)$$

where  $\hat{d}$  (or  $\hat{c}$ ) is the defective (or cooperative) label. The noise in the stochastic gradients,  $W_{k,t}$ , can be approximated as a zero mean Gaussian r.v. [17, 23, 27, 38]. Recall from (2) that cooperative users send the privacy-preserved versions of their gradient. This implies  $Y_{k,t} \sim \mathcal{N}(m_t, \Sigma_t)$ , given  $B_{k,t} = c$ , where  $\Sigma_t := \text{cov}[W_{k,t}] + \nu_t^2 \mathbf{I}$ . Thus, the detection and false alarm probabilities of the classifier, denoted by  $\Phi_t$  and  $\Psi_t$  respectively, can be found as

$$\Phi_t = 1 - \mathcal{Q} \left( \frac{m_t^\top \hat{m}_t - \frac{1}{2} \|\hat{m}_t\|_2^2}{\sqrt{\hat{m}_t^\top \Sigma_t \hat{m}_t}} \right) \quad \text{and} \quad \Psi_t = \mathcal{Q} \left( \frac{\frac{1}{2} \|\hat{m}_t\|_2^2}{\sqrt{\hat{m}_t^\top \Xi_t \hat{m}_t}} \right). \quad (8)$$

138 **Remark 3.** The linear classifier (7) is an effective tool under the homoscedasticity assumption. If  
 139 that is violated, the FC can employ different classifiers. The details are provided in Appendix for the  
 140 Classifier Design.

141 In the next subsection, we discuss how the FC can devise her strategies building on the joint gradient  
 142 estimation and user action classification scheme.

### 143 3.2 Adaptive Strategies for Fusion Center

144 Although the ZD strategies,  $\mathbf{p}$ , provide the FC an efficient and powerful mechanism to encourage  
 145 the user's cooperation; the FC cannot directly use  $\mathbf{p}$  since they are conditioned on the user's ac-  
 146 tion,  $B_{k,t}$ , which is not observable to her. Alternatively, the FC can use the classification results  
 147 after carefully *adapting* her strategies to mitigate the adverse effects of inevitable classification  
 148 errors. Let  $\pi_{t,1}, \pi_{t,2}, \pi_{t,3}$  and  $\pi_{t,4}$  denote the probabilities of cooperation for the FC conditioned  
 149 on  $(A_{k,t-1}, \hat{B}_{k,t})$ , in the order of  $(C, \hat{c}), (C, \hat{d}), (D, \hat{c})$  and  $(D, \hat{d})$ . These are referred to as *adaptive*  
 150 strategies and the FC sets these probabilities satisfying the following system of equations:

$$\begin{aligned} p_1 &= \pi_{t,1}\Phi_t + \pi_{t,2}(1 - \Phi_t), & p_2 &= \pi_{t,1}\Psi_t + \pi_{t,2}(1 - \Psi_t), \\ p_3 &= \pi_{t,3}\Phi_t + \pi_{t,4}(1 - \Phi_t), & p_4 &= \pi_{t,3}\Psi_t + \pi_{t,4}(1 - \Psi_t). \end{aligned}$$

151 Suppose  $\frac{\Phi_t}{\Psi_t} \geq \frac{p_1}{p_2}$  and  $\frac{\Phi_t}{\Psi_t} \geq \frac{p_3}{p_4}$ . Then the unique solution to the system above is given by

$$\pi_{t,1} = \frac{p_1(1 - \Psi_t) - p_2(1 - \Phi_t)}{\Phi_t - \Psi_t}, \quad \pi_{t,2} = \frac{p_2\Phi_t - p_1\Psi_t}{\Phi_t - \Psi_t}, \quad (9a)$$

$$\pi_{t,3} = \frac{p_3(1 - \Psi_t) - p_4(1 - \Phi_t)}{\Phi_t - \Psi_t}, \quad \pi_{t,4} = \frac{p_4\Phi_t - p_3\Psi_t}{\Phi_t - \Psi_t}. \quad (9b)$$

152 **Remark 4.** If the FC directly employed the ZD strategies without any adaptation, i.e., she cooperates  
 153 with probability  $p_i$  conditioned on classification output; the repeated games may not converge to  
 154 the stationary state  $\Lambda^*$  and a linear relation between the expected payoffs (4) may not be enforced  
 155 because the classification errors yield an additive disturbance on the state transition matrix as follows  
 156

$$\Omega - (p_1 - p_2) \{ \mathbf{q}^\top [1 - \Phi_t \ 0 \ 1 - \Phi_t \ 0] + (\mathbf{1} - \mathbf{q})^\top [0 \ \Psi_t \ 0 \ \Psi_t] \}. \quad (10)$$

157 Adaptive strategies (9) cancel out this adverse disturbance on the dynamics of the repeated games.

158 In the absence of classification errors ( $\Phi_t = 1$  and  $\Psi_t = 0$ ), the adaptive strategies reduce to the ZD  
 159 strategies, i.e.,  $\pi_t = \mathbf{p}$ . Classification errors force the FC to be more *retaliatory* than dictated by the  
 160 ZD strategy  $\mathbf{p}$ , i.e.,  $\pi_{t,1} > p_1$ ,  $\pi_{t,3} > p_3$ ,  $\pi_{t,2} < p_2$  and  $\pi_{t,4} < p_4$ . In general, detection and false alarm  
 161 probabilities,  $\Phi_t$  and  $\Psi_t$ , are time-varying; thus the adaptive strategies also change over time.

### 162 3.3 The Impact of Estimation Errors on Repeated Game Dynamics

163 The proposed adaptive strategies (9) requires the knowledge of detection probability,  $\Phi_t$ . However,  
 164 the FC cannot exactly compute  $\Phi_t$  using (8) since she does not have the knowledge of  $m_t$ . Instead,  
 165 she can form her estimate  $\hat{\Phi}_t$  using  $\hat{m}_t$ :

$$\hat{\Phi}_t = 1 - \mathcal{Q} \left( \frac{\frac{1}{2} \|\hat{m}_t\|^2}{\sqrt{\hat{m}_t^\top \Sigma_t \hat{m}_t}} \right) \quad (11)$$

166 Due to the inevitable gradient estimation errors, in general, we have  $\hat{\Phi}_t \neq \Phi_t$ . As a result, the FC  
 167 cannot exactly employ the adaptive FC strategies dictated by Eq. 9. With several steps of variable  
 168 substitutions, this yields an additive perturbation on the state transition matrix as follows:

$$\tilde{\Omega}_t = \Omega + V_t \Omega^\perp \text{ with } V_t := \frac{\hat{\Phi}_t - \Phi_t}{\Phi_t - \Psi_t} \text{ and } \Omega^\perp := (p_1 - p_2) \mathbf{q}^\top [-1 \ 0 \ 1 \ 0]. \quad (12)$$

169 Let  $\tilde{\Lambda}_t$  be the probability distribution over the state space of the games  $\{Cc, Cd, Dc, Dd\}$  at the start  
 170 of iteration  $t$ . According to (12), the state distributions follow the transition rule such that

$$\tilde{\Lambda}_{t+1} = \tilde{\Lambda}_t \tilde{\Omega}_t = \tilde{\Lambda}_t (\Omega + V_t \Omega^\perp).$$

171 Note that  $\Lambda_t$  can be considered as the state distribution of the repeated games in the absence of  
 172 perturbations on the state transition matrix. For the FC,  $\Lambda_t$  is the designed state distribution in which  
 173 the ZD strategy dominates against any user strategy.

174 Next, we study the time-varying perturbation terms. Using (8) and (11),  $V_t$  can be found as<sup>2</sup>:

$$V_t = \frac{\widehat{\Phi}_t - \Phi_t}{\widehat{\Phi}_t - \Psi_t} = \frac{\mathcal{Q}\left(\frac{\widehat{m}_t^\top (m_t - \frac{1}{2}\widehat{m}_t)}{\sqrt{\widehat{m}_t^\top \Sigma_t \widehat{m}_t}}\right) - \mathcal{Q}\left(\frac{\frac{1}{2}\|\widehat{m}_t\|^2}{\sqrt{\widehat{m}_t^\top \Sigma_t \widehat{m}_t}}\right)}{1 - \mathcal{Q}\left(\frac{\frac{1}{2}\|\widehat{m}_t\|^2}{\sqrt{\widehat{m}_t^\top \Sigma_t \widehat{m}_t}}\right) - \mathcal{Q}\left(\frac{\frac{1}{2}\|\widehat{m}_t\|^2}{\sqrt{\widehat{m}_t^\top \Xi_t \widehat{m}_t}}\right)} = \frac{\mathcal{Q}\left(\frac{\widehat{m}_t^\top (m_t - \widehat{m}_t) + \frac{1}{2}\|\widehat{m}_t\|}{\sqrt{\text{Ray}(\Sigma_t, \widehat{m}_t)}}\right) - \mathcal{Q}\left(\frac{\frac{1}{2}\|\widehat{m}_t\|}{\sqrt{\text{Ray}(\Sigma_t, \widehat{m}_t)}}\right)}{1 - \mathcal{Q}\left(\frac{\frac{1}{2}\|\widehat{m}_t\|}{\sqrt{\text{Ray}(\Sigma_t, \widehat{m}_t)}}\right) - \mathcal{Q}\left(\frac{\frac{1}{2}\|\widehat{m}_t\|}{\sqrt{\text{Ray}(\Xi_t, \widehat{m}_t)}}\right)}.$$

175 In the presence of these perturbations, to establish stability guarantees on the dynamics of the repeated  
 176 games, we impose the following assumption on the norm of the gradient estimator:

177 **Assumption 1.** Assume that  $\|\widehat{m}_t\| \geq \max\{2\sqrt{\text{Ray}(\widehat{m}_t, \Sigma_t)}, 2\sqrt{\text{Ray}(\widehat{m}_t, \Xi_t)}, \sqrt{|\widehat{m}_t^\top (m_t - \widehat{m}_t)|}\}$ .

178 Note that these conditions are primarily associated to the accuracy of the linear classifier (7) which  
 179 operates effectively when the mean vectors of the classes are sufficiently separated. The following  
 180 result indicates that, due to the perturbations on the state transition matrix, the real state distribution  
 181  $\tilde{\Lambda}_t$  is a noisy version of  $\Lambda_t$ .

182 **Lemma 1.** Let  $\Lambda_1$  denote the initial state distributions of the games between the FC and the users.  
 183 Under Assumption 1, we have that

$$\tilde{\Lambda}_t = \Lambda_t + \Lambda_1 \sum_{i=1}^{t-1} V_i \Omega^{i-1} \Omega^\perp \Omega^{t-1-i}. \quad (13)$$

184 This noise on the state distributions will manifest as a novel bias term in the gradient estimation. In  
 185 the next subsection, we will provide the convergence analysis of SGD-SU which will mainly focus  
 186 on the characterization of this bias term.

### 187 3.4 Convergence Results

188 In this section, we provide the convergence guarantee for SGD-SU (5). Let  $\mathcal{F}_t$  denote the  $\sigma$ -algebra,  
 189 generated by  $\{\theta_1, \mathbf{Y}_i, i < t\}$ . In particular,  $\mathcal{F}_t$  should be interpreted as the history of SGD-SU up to  
 190 iteration  $t$ , just before  $\mathbf{Y}_t$  is generated. Thus, conditioning on  $\mathcal{F}_t$  can be thought of as conditioning  
 191 on  $\{\theta_1, \tilde{\Lambda}_1, \mathbf{Y}_1, \dots, \theta_{t-1}, \tilde{\Lambda}_{t-1}, \mathbf{Y}_{t-1}, \theta_t, \tilde{\Lambda}_t\}$ . For convenience, denote  $\mathbb{E}_t[\cdot] := \mathbb{E}_t[\cdot|\mathcal{F}_t]$ . Observe  
 192 that, we can decompose the gradient estimator  $\widehat{m}_t$  as follows:

$$\widehat{m}_t(\cdot) = m_t(1 + \zeta_t) + \mathcal{E}_t, \quad (14)$$

193 where  $\zeta_t$  is the estimation bias term due to the perturbations on the state transition matrix, given by

$$\zeta_t = \frac{1}{m_t} (\mathbb{E}_t[\widehat{m}_t] - m_t) = \frac{\sum_{k=1}^K \mathbb{P}(B_{k,t} = c|\mathcal{F}_t)}{K(\Lambda_t \mathbf{q}^\top)} - 1$$

194 and  $\mathcal{E}_t$  is the estimation noise term, given by  $\mathcal{E}_t = \widehat{m}_t - \mathbb{E}_t[\widehat{m}_t]$ . Conditioned on  $\mathcal{F}_t$ , the probability of  
 195 a user taking the cooperative action, in iteration  $t$ , is given by  $\mathbb{P}(B_{k,t} = c|\mathcal{F}_t) = \tilde{\Lambda}_t \mathbf{q}^\top$ . The bias term,  
 196  $\zeta_t$ , can be found as follows:

$$\zeta_t = \frac{\tilde{\Lambda}_t \mathbf{q}^\top}{\Lambda_t \mathbf{q}^\top} - 1. \quad (15)$$

197 From Lemma 1 and (15), it is clear that the perturbations on the state transition matrix (12), directly  
 198 translates into a bias in the gradient estimation rule.

199 To establish convergence guarantees for the SGD-SU in (5),  $\Lambda_t \mathbf{q}^\top$  and  $\tilde{\Lambda}_t \mathbf{q}^\top$  must meet the following  
 200 criteria during the course of the algorithm:

201 **Assumption 2.** We assume that  $\Lambda_t \mathbf{q}^\top > \frac{1}{2}$  and  $\tilde{\Lambda}_t \mathbf{q}^\top > 0$ , for all  $t \in \{1, 2, \dots, T\}$ .

<sup>2</sup>The Rayleigh's quotient for a symmetric matrix  $M$  and nonzero vector  $x$  is defined as  $\text{Ray}(M, x) = \frac{x^\top M x}{x^\top x}$

202 The first condition  $\Lambda_t \mathbf{q}^\top \geq 0.5$  is very mild in the sense that it merely requires that the probability  
 203 of user cooperation dictated by the memory-1 strategies  $\mathbf{p}$  and  $\mathbf{q}$  ( $1 \times 4$  vectors), in the absence  
 204 of perturbations, is larger than 0.5. The second condition  $\tilde{\Lambda}_t \mathbf{q}^\top > 0$  states that, in the presence of  
 205 perturbations, the probability of user cooperation is always positive<sup>3</sup>.

206 By Assumption 2, there exists a positive constant  $H_T$  such that

$$0 < |\zeta_t| < H_T < 1, \forall t \in \{1, \dots, T\}. \quad (16)$$

207 Further, we have the following lemma characterizing the properties of estimation noise.

208 **Lemma 2.** *Conditioned on  $\mathcal{F}_t$ , the estimation noise in iteration  $t$ , denoted  $\mathcal{E}_t$ , is a zero-mean random*  
 209 *vector with the mean square error given by*

$$\mathbb{E}_t[\|\mathcal{E}_t\|^2] = \frac{1}{K(\Lambda_t \mathbf{q}^\top)} \left( (\zeta_t + 1) \text{tr}(\Sigma_t - \Xi_t) + \frac{1}{\Lambda_t \mathbf{q}^\top} \text{tr}(\Xi_t) \right). \quad (17)$$

210 By (16) and (17), we have that

$$\mathbb{E}_t[\|\mathcal{E}_t\|^2] \leq \frac{E_T}{K} \text{ with } E_T := \frac{1}{\Lambda_t \mathbf{q}^\top} \left[ (H_T + 1) \text{tr}(\Sigma_t - \Xi_t) + \frac{1}{\Lambda_t \mathbf{q}^\top} \text{tr}(\Xi_t) \right]. \quad (18)$$

211 We impose the following assumption on the objective function, which is standard for performance  
 212 analysis of stochastic gradient-based methods [3, 29].

213 **Assumption 3.** *The objective function  $F$  and the SGD-SU satisfy the following:*

214 (i)  *$F$  is  $L$ -smooth, that is,  $F$  is differentiable and its gradient is  $L$ -Lipschitz:*

$$\|\nabla F(\theta) - \nabla F(\theta')\| \leq L\|\theta - \theta'\|, \forall \theta, \theta' \in \mathbb{R}^n.$$

215 (ii) *The sequence of iterates  $\{\theta_t\}$  is contained in an open set over which  $F$  is bounded below by*  
 216 *a scalar  $F_{\inf}$ .*

217 Our next result describes the behavior of the sequence of gradients of  $F$  when fixed step sizes are  
 218 employed.

219 **Theorem 1.** *Under Assumptions 2 and 3, suppose that the SGD-SU (5) is run for  $T$  iterations with a*  
 220 *fixed stepsize  $\beta$  satisfying*

$$0 < \bar{\eta} \leq \frac{1}{L(1 + H_T)}. \quad (19)$$

221 *Then, the SGD algorithm with strategic users satisfies that*

$$\mathbb{E} \left[ \frac{1}{T} \sum_{t=1}^T \|\nabla F(\theta_t)\|^2 \right] \leq \frac{LE_T}{K(1 - H_T)} + \frac{2(F(\theta_1) - F_{\inf})}{\beta T(1 - H_T)}.$$

222 Theorem 1 illustrates the impact of the perturbations on the state transition matrix (12) on the  
 223 convergence rate of SGD-SU. When  $H_T$  is close to 0, SGD-SU performs similar to the basic  
 224 minibatch SGD. On the other hand, if  $H_T$  is close to 1, the optimality gap may be large. Our next  
 225 result will characterize the gradient estimation bias term  $\zeta_t$ . First, we have the following assumption  
 226 on the state transition matrix  $\Omega$ .

227 **Assumption 4.** *The state transition matrix  $\Omega$  can be diagonalized as  $\Omega = \Gamma \mathcal{U} \Gamma^{-1}$  with  $\mathcal{U}$  has the*  
 228 *eigenvalues of  $\Omega$  in descending order of magnitude:  $1 \geq |u_2| \geq |u_3| \geq |u_4| \geq 0$ .*

229 Denote the element of  $\Gamma^{-1}$  in the  $i^{\text{th}}$  row and  $j^{\text{th}}$  column as  $\Gamma_{ij}^{-1}$ . Denote the four rows of  $\Gamma^{-1}$  by  
 230  $\vec{\gamma}_1, \dots, \vec{\gamma}_4$ . Next, we define  $\delta$  as

$$\delta := \left( \max_{j \in \{2,3,4\}} |\Gamma_{3j} - \Gamma_{1j}| \right) \left( \max_{j \in \{2,3,4\}} |\vec{\gamma}_j \mathbf{q}^\top|^2 \right).$$

231 Further, the first order Taylor approximation of the scalar variable  $V_t$  can be found as follows:

$$V_t = \frac{m_t^\top (\hat{m}_t - m_t)}{\|m_t\|^2} h_t(m_t) \text{ with } h_t(m_t) := \frac{\frac{\|m_t\|}{\sqrt{2\pi \text{Ray}(\Sigma_t, m_t)}} \exp\left(-\frac{1}{8} \frac{\|m_t\|^2}{\text{Ray}(\Sigma_t, m_t)}\right)}{1 - \mathcal{Q}\left(\frac{\|m_t\|}{2\sqrt{\text{Ray}(\Sigma_t, m_t)}}\right) - \mathcal{Q}\left(\frac{\|m_t\|}{2\sqrt{\text{Ray}(\Xi_t, m_t)}}\right)}. \quad (20)$$

<sup>3</sup>A sufficient condition for this requirement is that user strategies are *forgiving* in nature, i.e.,  $q_1, q_2, q_3, q_4 > 0$ .

232 Define  $h_t^{\max} := \max_{i \in \{1, \dots, t\}} h_i(m_i)$ . Our next result indicates that, the estimation bias term  $\zeta_t$  can  
 233 be found in terms of the past gradient estimation errors.

234 **Theorem 2.** *Under Assumptions 1, 2 and 4, the gradient estimation bias term  $\zeta_t$ , can be found as*

$$\zeta_t = (p_1 - p_2) \sum_{i=1}^{t-1} \frac{\Lambda_i \mathbf{q}^\top}{\Lambda_t \mathbf{q}^\top} \frac{m_i^\top \mathcal{E}_i}{\|m_i\|^2} h_i(m_i) \Delta_{i,t} \quad (21a)$$

235 *with*

$$|\Delta_{i,t}| \leq \delta |u_2|^{t-1-i} + \delta^2 h_{t-1}^{\max} |u_2|^{t-2-i} (t-i-1). \quad (21b)$$

236 *Further, for some  $0 < \eta < 1$  we have*

$$\mathbb{P}(|\zeta_t| < \eta |\alpha_1, \dots, \alpha_{t-1}|) > 1 - \frac{\sum_{i=1}^{t-1} \alpha_i^2}{K \eta^2} \quad (22a)$$

237 *with*

$$\alpha_i^2 = \frac{2 \left| (\nu_i^2 - \xi_i^2) + \frac{m_i^\top \Sigma_i m_i}{\|m_i\|^2} \right| + \frac{\xi_i^2}{\Lambda_i \mathbf{q}^\top}}{\|m_i\|^2 (\Lambda_i \mathbf{q}^\top)} \left[ \frac{\Lambda_i \mathbf{q}^\top}{\Lambda_t \mathbf{q}^\top} \right]^2 h_i^2 \Delta_{i,t}^2. \quad (22b)$$

238 Note that Eq. (21) indicates that, the estimation bias term  $\zeta_t$  can be expanded in terms of past gradient  
 239 estimation errors. We prove that the absolute values of the coefficients,  $|\Delta_{i,t}|$ 's, are bounded as

$$|\Delta_{i,t}| \leq \delta |u_2|^{t-1-i} + \delta^2 h_{t-1}^{\max} |u_2|^{t-2-i} (t-i-1),$$

240 where  $u_2$  is the eigenvalue of  $\Omega$  with the second highest absolute value. Since  $\Omega$  is a row stochastic  
 241 matrix,  $|u_2| \leq 1$ . When  $|u_2|$  is strictly less than 1,  $\Delta_{i,t}$ 's decay fast as  $t-i$  grows. This can also be  
 242 interpreted as the impact of past gradient estimation errors fade away quickly. Using this result, in  
 243 Eq.(22), we derive a high probability upper bound on the estimation bias term  $\zeta_t$ .

## 244 4 Experiments

245 In this section, we evaluate the performance of SGD-SU (5) using real-life datasets. All the results in  
 246 the preceding section assert convergence for the SG method (5) under the assumption that the FC can  
 247 access  $\Sigma_t$  and  $\Xi_t$ . In a real-life machine learning setting with strategic users, this information may  
 248 not be available to the FC. For convenience, define  $\hat{\mathcal{K}}_t^c$  and  $\hat{\mathcal{K}}_t^d$  as the sets of users who are classified  
 249 as cooperative ( $\hat{c}$ ) and defective ( $\hat{d}$ ) at iteration  $t$ . Based on the user action classification, the FC can  
 250 form her estimates for the covariance matrices under the cooperative and defective actions as follows:

$$\hat{\Sigma}_t = \frac{1}{|\hat{\mathcal{K}}_t^c|} \sum_{k \in \hat{\mathcal{K}}_t^c} (Y_{k,t} - \bar{Y}_t^c) (Y_{k,t} - \bar{Y}_t^c)^\top \text{ and } \hat{\Xi}_t = \frac{1}{|\hat{\mathcal{K}}_t^d|} \sum_{k \in \hat{\mathcal{K}}_t^d} (Y_{k,t} - \bar{Y}_t^d) (Y_{k,t} - \bar{Y}_t^d)^\top, \quad (23)$$

251 where  $\bar{Y}_t^c = \frac{1}{|\hat{\mathcal{K}}_t^c|} \sum_{k \in \hat{\mathcal{K}}_t^c} Y_{k,t}$  and  $\bar{Y}_t^d = \frac{1}{|\hat{\mathcal{K}}_t^d|} \sum_{k \in \hat{\mathcal{K}}_t^d} Y_{k,t}$ .

252 In our first set of experiments, we consider a binary logistic classification problem and use the KDD-  
 253 Cup 04 dataset [6]. The goal of binary logistic classification experiments is to learn a classification  
 254 rule that differentiates between two types of particles generated in high energy collider experiments  
 255 based on 78 attributes [6]. In our second set of experiments, we consider a neural network trained on  
 256 the MNIST dataset. The number of users is chosen as  $K = 50$  and mini-batch size is  $s = 10$ . In the  
 257 experiments, we have tested the performance of two different ZD strategies, namely *equalizer* and  
 258 *extortion*[33].

259 For the logistic classification problem, Fig. 4a and 4b, depict the optimality gap under four different  
 260 user strategies:  $\mathbf{q} = [0.9 \ 0.15 \ 0.9 \ 0.15]$  (stubborn),  $\mathbf{q} = [0.9 \ 0.9 \ 0.15 \ 0.15]$  (tit-for-tat),  $\mathbf{q} =$   
 261  $[0.9 \ 0.15 \ 0.15 \ 0.9]$  (win-stay-lose-switch) and  $\mathbf{q} = [0.9 \ 0.9 \ 0.9 \ 0.9]$  (full cooperation). For the full  
 262 cooperation, coin toss, tit-for-tat and stubborn user strategies, SGSU converges quickly. For Pavlov  
 263 user strategies, SGSU can eventually approach, albeit more slowly than other cases. Fig 4c and 4d  
 264 illustrate the probability of user cooperation,  $\hat{\Lambda}_t \mathbf{q}^\top$ , across different user strategies. The experimental  
 265 results validate Lemma 1 and the empirical user cooperation probabilities match the theoretical except  
 266 when the users are Pavlov. Unsurprisingly, when the users follow full cooperation (or coin toss)  
 267 strategy, they cooperate with probability 0.9 (or 0.5) regardless of the actual states of the repeated



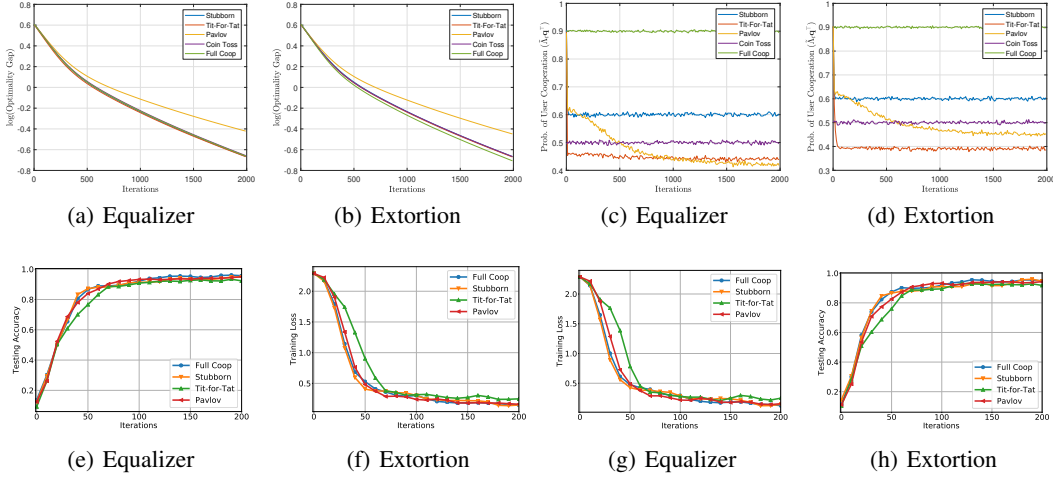


Figure 2: Stochastic Descent Algorithm with Strategic Users

games. For the cases with stubborn and tit-for-tat users, the games quickly converge to the steady state distribution. Interestingly, for the cases with Pavlov users, the probability of user cooperation decreases over time. This is associated to the performance of the linear classifier. For the image classification problem, Fig 4e-h depict the training loss and testing accuracy across iterations for different FC and user strategies. In all experiments, SGSU converges in the presence of strategic users. Further details regarding the Experimental results are relegated to Appendix.

## 5 Future Directions

In this work, we study a distributed learning framework where strategic users train a learning model with a fusion center. The main objective of the FC is to encourage users to be cooperative by distributing rewards. Based on this, we devise a reward mechanism for the FC based on the ZD-strategies. Further, we examine the performance of SGD algorithm in the presence of strategic users. Our findings reveal that the algorithm has provable convergence and our empirical results verify our theoretical analysis.

We are also working on the development of robust estimation tools in distributed learning with strategic users. The geometric median is a reliable estimation technique when the collected data contain outliers of large magnitude [10, 14, 25, 28]:

$$\text{Med}(\mathbf{Y}_t) := \arg \min_{y \in \mathbb{R}^n} \sum_{k=1}^K \|y - Y_{k,t}\|_2. \quad (24)$$

The FC can use Med as a robust gradient estimator, especially when the variance of the uninformative signals,  $\xi_t^2$ , reported by the defective users, is very high. The geometric median (24) can be computed by the Weiszfeld's algorithm [36, 37], which is a special case of iteratively reweighted least squares. In contrast, with the knowledge of  $\mathbf{q}$ , the modified sample mean estimator (6) allows the FC to trade robustness for overall tractability of the algorithm with reduced computational complexity.

The linear classifier is vulnerable to vanishing gradients as the stochastic gradient descent algorithm with strategic users (SGD-SU) converges to  $\theta^*$ . This can be addressed by modifying the classifier to incorporate the information contained in the norm of the reported gradients. Furthermore, we discuss how to extend the convergence guarantee for SGSU to allow heterogeneous user strategies. The details are presented in Appendix.

## References

- [1] ALISTARH, D., ALLEN-ZHU, Z., AND LI, J. Byzantine stochastic gradient descent. In *Advances in Neural Inform. Proc. Systems 31* (2018), NIPS'18, pp. 4613–4623.

- [2] BLANCHARD, P., EL MHAMDI, E. M., GUERRAOU, R., AND STAINER, J. Machine learning with adversaries: Byzantine tolerant gradient descent. In *Advances in Neural Inform. Proc. Systems 30* (2017), NIPS'17, pp. 119–129.
- [3] BOTTOU, L., CURTIS, F. E., AND NOCEDAL, J. Optimization methods for large-scale machine learning. *SIAM Review* 60, 2 (2018), 223–311.
- [4] CAI, Y., DASKALAKIS, C., AND PAPADIMITRIOU, C. Optimum statistical estimation with strategic data sources. *Journal of Machine Learning Research* 40, 2015 (2015), 1–17.
- [5] CARAGIANNIS, I., PROCACCIA, A. D., AND SHAH, N. Truthful univariate estimators. *33rd International Conference on Machine Learning, ICML 2016 1* (2016), 200–210.
- [6] CARUANA, R., JOACHIMS, T., AND BACKSTROM, L. Kdd-cup 2004: Results and analysis. *SIGKDD Explor. Newsl.* 6, 2 (Dec. 2004), 95–108.
- [7] CHEN, Y., IMMORLICA, N., LUCIER, B., SYRGKANIS, V., AND ZIANI, J. Optimal data acquisition for statistical estimation. In *Proceedings of the 2018 ACM Conference on Economics and Computation* (New York, NY, USA, 2018), EC '18, Association for Computing Machinery, p. 27–44.
- [8] CHEN, Y., PODIMATA, C., PROCACCIA, A. D., AND SHAH, N. Strategyproof Linear Regression in High Dimensions. In *Proceedings of the 2018 ACM Conference on Economics and Computation* (New York, NY, USA, jun 2018), vol. 76, ACM, pp. 9–26.
- [9] CHEN, Y., SU, L., AND XU, J. Distributed statistical machine learning in adversarial settings: Byzantine gradient descent. *Proc. ACM Meas. Anal. Comput. Syst.* 1, 2 (Dec. 2017).
- [10] COHEN, M. B., LEE, Y. T., MILLER, G., PACHOCKI, J., AND SIDFORD, A. Geometric median in nearly linear time. In *Proc. ACM Symp. on Theory of Comp.* (New York, NY, USA, 2016), STOC '16, ACM, p. 9–21.
- [11] CUMMINGS, R., IOANNIDIS, S., AND LIGETT, K. Truthful linear regression. *Journal of Machine Learning Research* 40, 2015 (2015), 1–36.
- [12] DEKEL, O., FISCHER, F., AND PROCACCIA, A. D. Incentive compatible regression learning. *Journal of Computer and System Sciences* 76, 8 (2010), 759–777.
- [13] DWORK, C. Differential privacy. In *Proc. Int. Conf. Automata, Languages and Programming - Volume Part II* (Berlin, Heidelberg, 2006), ICALP'06, Springer-Verlag, pp. 1–12.
- [14] FLETCHER, P. T., VENKATASUBRAMANIAN, S., AND JOSHI, S. Robust statistics on riemannian manifolds via the geometric median. In *2008 IEEE Conference on Computer Vision and Pattern Recognition* (2008), pp. 1–8.
- [15] HAO, D., RONG, Z., AND ZHOU, T. Extortion under uncertainty: Zero-determinant strategies in noisy games. *Phys. Rev. E* 91 (May 2015), 052803.
- [16] HORN, R., HORN, R., AND JOHNSON, C. *Matrix Analysis*. Cambridge University Press, 1990.
- [17] JASTRZKEBSKI, S., KENTON, Z., ARPIT, D., BALLAS, N., FISCHER, A., BENGIO, Y., AND STORKEY, A. J. Three factors influencing minima in SGD. *arXiv:1711.04623v3 [cs.LG]* (Nov. 2017).
- [18] JORDAN, M. I., LEE, J. D., AND YANG, Y. Communication-efficient distributed statistical inference. *Journal of the American Statistical Association* 114, 526 (2019), 668–681.
- [19] KONEČNÝ, J., MCMAHAN, H. B., RAMAGE, D., AND RICHTARIK, P. Federated optimization: Distributed machine learning for on-device intelligence. *arXiv:1610.02527 [cs.LG]* (Oct. 2016).
- [20] KONG, Y., SCHOENEBECK, G., TAO, B., AND YU, F.-Y. Information elicitation mechanisms for statistical estimation. *Proceedings of the AAAI Conference on Artificial Intelligence* 34, 02 (Apr. 2020), 2095–2102.
- [21] KRAINES, D. P., AND KRAINES, V. Y. Natural selection of memory-one strategies for the iterated prisoner's dilemma. *Journal of Theoretical Biology* 203, 4 (2000), 335 – 355.
- [22] LI, M., ANDERSEN, D. G., PARK, J. W., SMOLA, A. J., AHMED, A., JOSIFOVSKI, V., LONG, J., SHEKITA, E. J., AND SU, B.-Y. Scaling distributed machine learning with the parameter server. In *Proc. of the 11th USENIX Conf. on Operating Systems Design and Implementation* (USA, 2014), OSDI'14, USENIX Association, p. 583–598.

- [23] LIN, T., STICH, S. U., PATEL, K. K., AND JAGGI, M. Don't use large mini-batches, use local sgd. In *Int. Conf. Learning Representations* (2020), ICLR'20.
- [24] LIU, Y., AND WEI, J. Incentives for Federated Learning: A Hypothesis Elicitation Approach. *arXiv:2007.10596v1 [cs.LG]* (July 2020).
- [25] LOPUHAA, H. P., AND ROUSSEEUW, P. J. Breakdown points of affine equivariant estimators of multivariate location and covariance matrices. *Ann. Statist.* 19, 1 (03 1991), 229–248.
- [26] LOW, Y., BICKSON, D., GONZALEZ, J., GUESTRIN, C., KYROLA, A., AND HELLERSTEIN, J. M. Distributed graphlab: A framework for machine learning and data mining in the cloud. *Proc. VLDB Endow.* 5, 8 (Apr. 2012), 716–727.
- [27] MANDT, S., HOFFMAN, M. D., AND BLEI, D. M. A variational analysis of stochastic gradient algorithms. In *Proc. Int. Conf. Machine Learning* (2016), vol. 48 of *ICML'16*, JMLR.org, p. 354–363.
- [28] MINSKER, S. Geometric median and robust estimation in banach spaces. *Bernoulli* 21, 4 (11 2015), 2308–2335.
- [29] NEMIROVSKI, A., JUDITSKY, A., LAN, G., AND SHAPIRO, A. Robust stochastic approximation approach to stochastic programming. *SIAM J. on Optimization* 19, 4 (2009), 1574–1609.
- [30] NG, K. L., CHEN, Z., LIU, Z., YU, H., LIU, Y., AND YANG, Q. A multi-player game for studying federated learning incentive schemes. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20* (7 2020), C. Bessiere, Ed., International Joint Conferences on Artificial Intelligence Organization, pp. 5279–5281.
- [31] POLYAK, B. T., AND JUDITSKY, A. B. Acceleration of stochastic approximation by averaging. *SIAM Journal on Control and Optimization* 30, 4 (1992), 838–855.
- [32] POOR, H. V. *An Introduction to Signal Detection and Estimation*, 2nd ed.. ed. Springer Texts in Electrical Engineering. Springer-Verlag, New York, NY, 1994.
- [33] PRESS, W. H., AND DYSON, F. J. Iterated prisoner's dilemma contains strategies that dominate any evolutionary opponent. *Proc. Natl. Acad. Sci* 109, 26 (2012), 10409–10413.
- [34] RICHARDSON, A., FILOS-RATSIKAS, A., AND FALTINGS, B. *Budget-Bounded Incentives for Federated Learning*. Springer International Publishing, Cham, 2020, pp. 176–188.
- [35] SU, L., AND XU, J. Securing distributed gradient descent in high dimensional statistical learning. *SIGMETRICS Perform. Eval. Rev.* 47, 1 (Dec. 2019), 83–84.
- [36] VARDI, Y., AND ZHANG, C.-H. The multivariate  $L_1$ -median and associated data depth. *Proc. Natl. Acad. Sci.* 97, 4 (2000), 1423–1426.
- [37] WEISZFELD, E. Sur un problème de minimum dans l'espace. *Tohoku Math. J.* 42 (1936), 274–280.
- [38] XING, C., ARPIT, D., TSIRIGOTIS, C., AND BENGIO, Y. A Walk with SGD. *arXiv:1802.08770 [stat.ML]* (Feb. 2018).
- [39] XING, E. P., HO, Q., XIE, P., AND WEI, D. Strategies and principles of distributed machine learning on big data. *Engineering* 2, 2 (2016), 179 – 195.

## Checklist

### 1. For all authors...

- (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
- (b) Did you describe the limitations of your work? [Yes]
- (c) Did you discuss any potential negative societal impacts of your work? [N/A]
- (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]

### 2. If you are including theoretical results...

- (a) Did you state the full set of assumptions of all theoretical results? [Yes]
- (b) Did you include complete proofs of all theoretical results? [Yes] Proofs are included in the Appendix.

### 3. If you ran experiments...

- (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] The code is included in supplementary material complying to NeurIPS instructions with details on how to run the code.
- (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes]
- (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes]
- (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] The hardware used is described in the Appendix

### 4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...

- (a) If your work uses existing assets, did you cite the creators? [Yes]
- (b) Did you mention the license of the assets? [N/A]
- (c) Did you include any new assets either in the supplemental material or as a URL? [Yes] Our code is included in the supplementary material.
- (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
- (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]

### 5. If you used crowdsourcing or conducted research with human subjects...

- (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
- (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
- (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

## 426 A Related Work

### 427 A.1 Game-Theoretical Approaches in Machine Learning

428 There are several papers that study incentives in the context of statistical estimation and inference  
 429 from strategic data sources [4, 5, 7, 8, 11, 12, 20, 24]. These works consider single-stage games and  
 430 their fundamental goals differ from ours. Since we consider a federated learning setting where the  
 431 FC only collects stochastic gradients rather than raw data or estimated models. Further, the FC and  
 432 the users interact repeatedly and it intertwines with the stochastic gradient descent updates. Thus,  
 433 it is crucial for us to evaluate the impact of the repeated games on the convergence performance of  
 434 distributed learning with strategic users.

435 In [34], a federated learning setting with independent and self-interested participants is considered.  
 436 One key difference is that, their work focuses on the economics of a federated learning system at  
 437 a single iteration rather than the impact of untruthful reporting on the overall performance of the  
 438 learning scheme throughout the entire training process. We formulate the interactions between the  
 439 FC and the users as repeated games, and introduce zero-determinant strategies for the FC. This is  
 440 especially significant to analyze the impact of defective user actions on the overall convergence  
 441 performance of the system. In [30], a multi-player game is proposed to study the reactions of strategic  
 442 participants, in various federated learning ecosystems, for various incentive mechanisms. However,  
 443 the scope of this study is limited to the development of an interactive user interface to collect data for  
 444 future experimental studies.

### 445 A.2 Repeated Games

446 In a 2x2 repeated game, it is possible for a player to unilaterally impose a linear relationship between  
 447 their and the opponent’s payoff employing “zero-determinant” (ZD) strategies [33]. In Press and  
 448 Dyson’s work, both players can observe the action of their opponent in a perfect environment without  
 449 any noise. In [15], the ZD strategies in noisy games is examined under the assumption that the players  
 450 know the time-invariant error distribution. In our paper, however, the FC cannot directly receive any  
 451 (noisy or noiseless) observation of the user action. In order to address this key difficulty, using the  
 452 collected reported gradients of the users, she forms a user action classifier and assigns cooperative or  
 453 defective labels to the users. Due to the nature of the data driven gradient updates, the user action  
 454 classification incurs time-varying stochastic errors, which adds another non-trivial complexity.

### 455 A.3 Byzantine-Resilient Machine Learning

456 In the presence of malicious devices, the robustness issues in distributed learning has received much  
 457 attention. In these studies, it is assumed that *good* devices dominate the entire set of devices and it is  
 458 proposed that fault-tolerant algorithms can trim the outliers from the candidates [1, 2, 9, 35]. The  
 459 basic goal of these studies differs from ours, since we consider a game-theoretic setting in which the  
 460 users are utility-driven who have the ability to formulate strategies to choose their actions, *cooperative*  
 461 or *defective*, which can depend on the outcome of previous interactions with the FC.

## 462 B Remark 1 on Differential Privacy

463 When a user is cooperative, he sends a privacy-protected version of his stochastic gradient  $X_{k,t}$  with  
 464 noise injection to preserve *differential privacy*. The privacy parameter  $\epsilon$  quantifies privacy loss; and  
 465 the lower  $\epsilon$  the stronger privacy protection.

466 **Definition 1.** [13] A randomized function  $\mathcal{A}$  is  $\epsilon$ -differentially private if for all data sets  $\mathbf{Z}$  and  $\mathbf{Z}'$   
 467 that differ in the value of a single sample, and all  $S \subseteq \text{Range}(\mathcal{A})$ ,

$$\mathbb{P}(\mathcal{A}(\mathbf{Z}) \in S) \leq e^\epsilon \mathbb{P}(\mathcal{A}(\mathbf{Z}') \in S),$$

468 where the probability is over the coin flips of  $\mathcal{A}$ .

469 Formally, when a user is cooperative, he forms his report through an oracle  $\mathcal{G}$ . Given the current iterate  
 470  $\theta_t$ , the user draws  $s$  independent samples  $\mathbf{Z}_{k,t} := \{z_{k,t}^1, \dots, z_{k,t}^s\}$  from the unknown underlying data

471 distribution  $\mathcal{D}$ , and outputs a perturbed version of local stochastic gradient:

$$\mathcal{G}(\theta_t) = N_{k,t} + \frac{1}{s} \sum_{i=1}^s \nabla_{\theta} \mathcal{L}(\theta; z_{k,t}^i) \Big|_{\theta=\theta_t},$$

472 where  $N_{k,t} \sim \mathcal{N}(\vec{0}, \nu_t^2 \mathbf{I})$ . For any  $t$ , and any  $\mathbf{Z}_{k,t}, \mathbf{Z}'_{k,t}$  that differ in the value of a single sample, it  
473 follows that

$$\begin{aligned} \frac{f(\mathcal{G}(\theta_t) = y | \mathbf{Z}_t)}{f(\mathcal{G}(\theta_t) = y | \mathbf{Z}'_t)} &= \frac{f\left(N_{k,t} = y - \frac{1}{s} \sum_{z \in \mathbf{Z}_{k,t}} \nabla \mathcal{L}(\theta; z)\right)}{f\left(N_{k,t} = y - \frac{1}{s} \sum_{z \in \mathbf{Z}'_{k,t}} \nabla \mathcal{L}(\theta; z)\right)} \\ &= \frac{\exp\left\{-\frac{1}{2\nu_t^2} \left\|y - \frac{1}{s} \sum_{z \in \mathbf{Z}_{k,t}} \nabla \mathcal{L}(\theta; z)\right\|_2^2\right\}}{\exp\left\{-\frac{1}{2\nu_t^2} \left\|y - \frac{1}{s} \sum_{z \in \mathbf{Z}'_{k,t}} \nabla \mathcal{L}(\theta; z)\right\|_2^2\right\}} \\ &\leq \exp\left\{\frac{1}{2\nu_t^2} \left\|\frac{1}{s} [\nabla \mathcal{L}(\theta_t; z) - \nabla \mathcal{L}(\theta_t; z')]\right\|_2^2\right\} \leq \exp\left\{\frac{\ell^2}{\nu_t^2 s^2}\right\}. \end{aligned}$$

## 474 C Remark 2 on Zero-Determinant Strategies

475 If the FC sets her strategy  $\mathbf{p}$  satisfying, for some real values  $\varphi_0, \varphi_1$  and  $\varphi_2$ , the equation

$$[p_1 - 1, p_2 - 1, p_3, p_4] = \varphi_0 \mathbf{S}_{\text{FC}} + \varphi_1 \mathbf{S}_{\text{U}} + \varphi_2 \vec{1},$$

476 then a linear relation between the expected payoffs, given by

$$\varphi_0 s_{\text{FC}}^* + \varphi_1 s_{\text{U}}^* + \varphi_2 = 0 \quad (25)$$

477 is enforced, regardless of  $\mathbf{q}$  [33]. This is called the zero-determinant (ZD) strategy. In this study, we  
478 focus on two important specializations of ZD strategies: *equalizers* and *extortioners*. An important  
479 subclass of ZD strategies, *extortionate* strategies, enable the ZD player to guarantee that an increase  
480 in the player's own payoff exceeds the increase in the opponent's payoff. *Equalizer* strategies, another  
481 subset of ZD strategies, allow the ZD player to set the opponent player's expected long term payoff.

482 Equalizers are those ZD strategies for which  $\varphi_0 = 0 \neq \varphi_1$  and  $\vec{\mathbf{p}} = \varphi_1 \mathbf{S}_{\text{U}} + \varphi_2 \vec{1}$ . With the adoption  
483 of equalizers, in the stationary state, the FC can assign the expected payoff of the user to a fixed value  
484 between 0 (mutual defection) and  $r - V_{\text{U}}R$  (mutual cooperation):

$$s_{\text{U}}^* = -\frac{\varphi_2}{\varphi_1} = (r - V_{\text{U}}R) \frac{p_4}{1 - p_1 + p_4}.$$

485 Extortioners are those ZD strategies for which  $\varphi_2 = 0$ , with  $\chi := -\varphi_1/\varphi_0 > \frac{r - rV_{\text{FC}}}{r - RV_{\text{U}}}$ . Then, the  
486 FC can enforce

$$s_{\text{FC}}^* = \chi s_{\text{U}}^*.$$

487 In this case,  $s_{\text{FC}}^*$  and  $s_{\text{U}}^*$  are maximized when  $\mathbf{q} = [1 \ 1 \ 1 \ 1]$ , ergo full cooperation is the best response  
488 strategy for the user. If the user does not accept full cooperation, by employing equalizer strategies,  
489 the FC may claim a unilateral control on the user's expected payoff.

490 Next, we show that  $p_1 > p_2$  and  $p_3 > p_4$  with  $p_1 - p_2 = p_3 - p_4$  when the FC adopts equalizer or  
491 extortion strategies. This property is useful in Section 3.3 when we discuss the impact of gradient  
492 estimation errors on the dynamics of the repeated games. First, consider the case where the FC  
493 employs equalizer strategies. We have that

$$p_2 = \frac{p_1 r - (1 + p_4) V_{\text{U}}R}{r - V_{\text{U}}R} \quad \text{and} \quad p_3 = \frac{(1 - p_1) V_{\text{U}}R + p_4 r}{r - V_{\text{U}}R}.$$

494 It follows that

$$p_1 - p_2 = p_3 - p_4 = \frac{V_{\text{U}}R}{r - V_{\text{U}}R} (1 + p_4 - p_1) > 0.$$

Furthermore, for sufficiently small  $\gamma$  and where  $\chi \geq 1$  is the extortion factor, the extortionate strategies of the FC satisfy that

$$p_1 = 1 - \gamma \left[ \frac{r}{1 + \chi V_U} - \frac{R}{V_{FC} + \chi} \right], \quad p_2 = 1 - \gamma \frac{r}{1 + \chi V_U}, \quad p_3 = \gamma \frac{R}{V_{FC} + \chi}, \quad p_4 = 0,$$

implying that

$$p_1 - p_2 = p_3 - p_4 = \gamma \frac{R}{V_{FC} + \chi} > 0.$$

## D Details on Time Varying State Transition Matrix in Section 3

In Section 3.2, we propose the adaptive strategies (9) which use the detection probability  $\Phi_t$  and false alarm rate  $\Psi_t$  of the linear classifier. The FC does not have the knowledge of true detection probability, which is a function of  $m_t$ . Instead, she forms her estimate  $\hat{\Phi}_t$  using  $\hat{m}_t$ . Under the assumption that  $\hat{\Phi}_t/\hat{\Psi}_t \geq \max\{p_1/p_2, p_3/p_4\}$ , we can write the probability that the FC takes the cooperative action conditioned on the joint action pair of  $(A_{k,t-1}, B_{k,t})$ , as follows:

$$\begin{aligned} P(A_{k,t}=C|A_{k,t-1}=C, B_{k,t}=c) &= \Phi_t \pi_{t,1} + (1 - \Phi_t) \pi_{t,2} \\ &= \Phi_t \frac{p_1(1 - \Psi_t) - p_2(1 - \hat{\Phi}_t)}{\hat{\Phi}_t - \Psi_t} + (1 - \Phi_t) \frac{p_2\hat{\Phi}_t - p_1\Psi_t}{\hat{\Phi}_t - \Psi_t}, \end{aligned}$$

$$\begin{aligned} P(A_{k,t}=C|A_{k,t-1}=C, B_{k,t}=d) &= \Psi_t \pi_{t,2} + (1 - \Psi_t) \pi_{t,2} \\ &= \Psi_t \frac{p_1(1 - \Psi_t) - p_2(1 - \hat{\Phi}_t)}{\hat{\Phi}_t - \Psi_t} + (1 - \Psi_t) \frac{p_2\hat{\Phi}_t - p_1\Psi_t}{\hat{\Phi}_t - \Psi_t}, \end{aligned}$$

504

$$\begin{aligned} P(A_{k,t}=C|A_{k,t-1}=D, B_{k,t}=c) &= \Phi_t \pi_{t,3} + (1 - \Phi_t) \pi_{t,4} \\ &= \Phi_t \frac{p_3(1 - \Psi_t) - p_4(1 - \hat{\Phi}_t)}{\hat{\Phi}_t - \Psi_t} + (1 - \Phi_t) \frac{p_4\hat{\Phi}_t - p_3\Psi_t}{\hat{\Phi}_t - \Psi_t}, \end{aligned}$$

$$\begin{aligned} P(A_{k,t}=C|A_{k,t-1}=D, B_{k,t}=d) &= \Psi_t \pi_{t,3} + (1 - \Psi_t) \pi_{t,4} \\ &= \Psi_t \frac{p_3(1 - \Psi_t) - p_4(1 - \hat{\Phi}_t)}{\hat{\Phi}_t - \Psi_t} + (1 - \Psi_t) \frac{p_4\hat{\Phi}_t - p_3\Psi_t}{\hat{\Phi}_t - \Psi_t}. \end{aligned}$$

After some algebra, it follows that

$$P(A_{k,t}=C|A_{k,t-1}=C, B_{k,t}=c) = p_1 - (p_1 - p_2) \frac{\hat{\Phi}_t - \Phi_t}{\hat{\Phi}_t - \Psi_t},$$

$$P(A_{k,t}=C|A_{k,t-1}=C, B_{k,t}=d) = p_2 - (p_1 - p_2) \frac{\hat{\Psi}_t - \Psi_t}{\hat{\Phi}_t - \Psi_t},$$

506

$$P(A_{k,t}=C|A_{k,t-1}=D, B_{k,t}=c) = p_3 - (p_3 - p_4) \frac{\hat{\Phi}_t - \Phi_t}{\hat{\Phi}_t - \Psi_t},$$

$$P(A_{k,t}=C|A_{k,t-1}=D, B_{k,t}=d) = p_4 - (p_3 - p_4) \frac{\hat{\Psi}_t - \Psi_t}{\hat{\Phi}_t - \Psi_t}.$$

This is to say, memory-one user strategies  $\mathbf{q}$  and adaptive FC strategies  $\pi_t$  imply a time-varying Markov state transition matrix given by

$$\tilde{\Omega}_t = \Omega + V_t \Omega^\perp \quad \text{with } V_t := \frac{\hat{\Phi}_t - \Phi_t}{\hat{\Phi}_t - \Psi_t} \text{ and } \Omega^\perp := \begin{bmatrix} -q_1(p_1 - p_2) & 0 & q_1(p_1 - p_2) & 0 \\ -q_2(p_1 - p_2) & 0 & q_2(p_1 - p_2) & 0 \\ -q_3(p_3 - p_4) & 0 & q_3(p_3 - p_4) & 0 \\ -q_4(p_3 - p_4) & 0 & q_4(p_3 - p_4) & 0 \end{bmatrix}. \quad (26)$$

509 Recall from Remark 2, when the memory-one FC strategies  $\mathbf{p}$  are in the form of extortioners or  
 510 equalizers, we have that

$$p_1 - p_2 = p_3 - p_4.$$

511 Thus,  $\Omega^\perp$  is a rank-one matrix and can be decomposed as

$$\Omega^\perp = (p_1 - p_2)\mathbf{q}^\top [-1 \ 0 \ 1 \ 0]1. \quad (27)$$

512 Using (8) and (11), we can express  $V_t$  as follows:

$$V_t = \frac{\hat{\Phi}_t - \Phi_t}{\hat{\Phi}_t - \Psi_t} = \frac{\mathcal{Q}\left(\frac{\hat{m}_t^\top (m_t - \frac{1}{2}\hat{m}_t)}{\sqrt{\hat{m}_t^\top \Sigma_t \hat{m}_t}}\right) - \mathcal{Q}\left(\frac{\frac{1}{2}\|\hat{m}_t\|^2}{\sqrt{\hat{m}_t^\top \Sigma_t \hat{m}_t}}\right)}{1 - \mathcal{Q}\left(\frac{\frac{1}{2}\|\hat{m}_t\|^2}{\sqrt{\hat{m}_t^\top \Sigma_t \hat{m}_t}}\right) - \mathcal{Q}\left(\frac{\frac{1}{2}\|\hat{m}_t\|^2}{\sqrt{\hat{m}_t^\top \Xi_t \hat{m}_t}}\right)}.$$

513 For  $\|\hat{m}_t\|_2 > 0$ ,  $V_t$  is a differentiable function of  $\hat{m}_t$  and the first order Taylor approximation of  $V_t$  at  
 514  $\hat{m}_t = m_t$  is given by

$$V_t \approx V_t^\perp|_{\hat{m}_t=m_t} + (\hat{m}_t - m_t)^\top \nabla_{\hat{m}_t} V_t|_{\hat{m}_t=m_t}.$$

515 Observe that

$$V_t|_{\hat{m}_t=m_t} = 0, \\ \nabla_{\hat{m}_t} V_t^\perp|_{\hat{m}_t=m_t} = m_t \frac{\frac{1}{\sqrt{2\pi m_t^\top \Sigma_t m_t}} \exp\left(-\frac{\|m_t\|^4}{8m_t^\top \Sigma_t m_t}\right)}{1 - \mathcal{Q}\left(\frac{\frac{1}{2}\|m_t\|^2}{\sqrt{m_t^\top \Sigma_t m_t}}\right) - \mathcal{Q}\left(\frac{\frac{1}{2}\|m_t\|^2}{\sqrt{m_t^\top \Xi_t m_t}}\right)}.$$

516 It follows that

$$V_t = \frac{m_t^\top (\hat{m}_t - m_t)}{\|m_t\|^2} h_t(m_t) \text{ with } h_t(m_t) := \frac{\frac{\|m_t\|^2}{\sqrt{2\pi m_t^\top \Sigma_t m_t}} \exp\left(-\frac{\|m_t\|^4}{8m_t^\top \Sigma_t m_t}\right)}{1 - \mathcal{Q}\left(\frac{\frac{1}{2}\|m_t\|^2}{\sqrt{m_t^\top \Sigma_t m_t}}\right) - \mathcal{Q}\left(\frac{\frac{1}{2}\|m_t\|^2}{\sqrt{m_t^\top \Xi_t m_t}}\right)}.$$

517 In Linear Algebra, the Rayleigh's quotient for a given symmetric matrix  $M$  and nonzero column  
 518 vector  $x$  is defined as:

$$\text{Ray}(M, x) = \frac{x^\top M x}{x^\top x}. \quad (28)$$

519 It can be shown that  $\lambda_{\min}(M) \leq \text{Ray}(M, x) \leq \lambda_{\max}(M)$ , where  $\lambda_{\min}(M)$  and  $\lambda_{\max}(M)$  are  
 520 respectively the smallest and the largest eigenvalues of  $M$  [16]. Further,  $\text{Ray}(M, x)$  reaches its  
 521 maximum (or minimum) when  $x$  is the eigenvector corresponding to  $\lambda_{\max}(M)$  (or  $\lambda_{\min}(M)$ ). After  
 522 some algebra, we can rewrite  $h_t(m_t)$  as follows:

$$V_t = \frac{m_t^\top (\hat{m}_t - m_t)}{\|m_t\|^2} h_t(m_t) \text{ with } h_t(m_t) := \frac{\frac{\|m_t\|}{\sqrt{2\pi \text{Ray}(\Sigma_t, m_t)}} \exp\left(-\frac{1}{8} \frac{\|m_t\|^2}{\text{Ray}(\Sigma_t, m_t)}\right)}{1 - \mathcal{Q}\left(\frac{\|m_t\|}{2\sqrt{\text{Ray}(\Sigma_t, m_t)}}\right) - \mathcal{Q}\left(\frac{\|m_t\|}{2\sqrt{\text{Ray}(\Xi_t, m_t)}}\right)}.$$

## 523 E Proof of Lemma 1

524 **Lemma 1.** Let  $\Lambda_1$  denote the initial state distributions of the games between the FC and the users.  
 525 Under Assumption 1, we have that

$$\tilde{\Lambda}_t = \Lambda_t + \Lambda_1 \sum_{i=1}^{t-1} V_i \Omega^{i-1} \Omega^\perp \Omega^{t-1-i}.$$



526 *Proof.* Recall that  $V_t$  can be found as

$$V_t = \frac{\mathcal{Q}\left(\frac{\hat{m}_t(m_t - \hat{m}_t) + 0.5\|\hat{m}_t\|^2}{\|m_t\|\sqrt{\text{Ray}(\Sigma_t, \hat{m}_t)}}\right) - \mathcal{Q}\left(\frac{0.5\|\hat{m}_t\|}{\sqrt{\text{Ray}(\Sigma_t, \hat{m}_t)}}\right)}{1 - \mathcal{Q}\left(\frac{0.5\|\hat{m}_t\|}{\sqrt{\text{Ray}(\Sigma_t, \hat{m}_t)}}\right) - \mathcal{Q}\left(\frac{0.5\|\hat{m}_t\|}{\sqrt{\text{Ray}(\Xi_t, \hat{m}_t)}}\right)}.$$

527 Under Assumption 1, it follows that

$$\left| \mathcal{Q}\left(\frac{\hat{m}_t(m_t - \hat{m}_t) + 0.5\|\hat{m}_t\|^2}{\|m_t\|\sqrt{\text{Ray}(\Sigma_t, \hat{m}_t)}}\right) - \mathcal{Q}\left(\frac{0.5\|\hat{m}_t\|}{\sqrt{\text{Ray}(\Sigma_t, \hat{m}_t)}}\right) \right| \leq 0.1499,$$

$$1 - \mathcal{Q}\left(\frac{0.5\|\hat{m}_t\|}{\sqrt{\text{Ray}(\Sigma_t, \hat{m}_t)}}\right) - \mathcal{Q}\left(\frac{0.5\|\hat{m}_t\|}{\sqrt{\text{Ray}(\Xi_t, \hat{m}_t)}}\right) \geq 0.6826.$$

528 Consequently,  $|V_t| \leq 0.2196$ . Recall that, given the initial state distributions of the repeated games  
529 as  $\tilde{\Lambda}_1 = \Lambda_1$ , we define the deterministic process:  $\Lambda_t = \Lambda_1 \Omega^{t-1}$ . For  $t = 2$ , we have

$$\tilde{\Lambda}_2 = \tilde{\Lambda}_1 (\Omega + V_1 \Omega^\perp) = \tilde{\Lambda}_1 \Omega + V_1 \tilde{\Lambda}_1 \Omega^\perp = \Lambda_2 + V_1 \Lambda_2^\perp, \quad \text{with } \Lambda_2^\perp := \Lambda_1 \Omega^\perp.$$

530 For  $t = 3$ ,

$$\begin{aligned} \tilde{\Lambda}_3 &= \tilde{\Lambda}_2 (\Omega + V_2 \Omega^\perp) = (\Lambda_2 + V_1 \Lambda_2^\perp) (\Omega + V_2 \Omega^\perp), \\ &= \Lambda_3 + V_1 \Lambda_3^\perp + V_2 \Lambda_2 \Omega^\perp + V_1 V_2 \Lambda_2^\perp \Omega^\perp, \quad \text{with } \Lambda_3^\perp := \Lambda_2 \Omega^\perp, \\ &\approx \Lambda_3 + V_1 \Lambda_3^\perp + V_2 \Lambda_2 \Omega^\perp, \end{aligned}$$

531 where the approximation follows since  $|V_1 V_2| \ll |V_1|, |V_2|$ . Further, suppose (13) is true. Then,

$$\begin{aligned} \tilde{\Lambda}_{t+1} &= \tilde{\Lambda}_t (\Omega + V_t \Omega^\perp) = \left[ \Lambda_t + \Lambda_1 \sum_{i=1}^{t-1} V_i \Omega^{i-1} \Omega^\perp \Omega^{t-1-i} \right] (\Omega + V_t \Omega^\perp), \\ &= \Lambda_{t+1} + V_t \Lambda_t \Omega^\perp + \Lambda_t \sum_{i=1}^{t-1} V_i \Omega^{i-1} \Omega^\perp \Omega^{t-i} + \Lambda_1 \sum_{i=1}^{t-1} V_i V_t \Omega^{i-1} \Omega^\perp \Omega^{t-1-i} \Omega^\perp, \\ &\approx \Lambda_{t+1} + V_t \Lambda_t \Omega^\perp + \Lambda_t \sum_{i=1}^{t-1} V_i \Omega^{i-1} \Omega^\perp \Omega^{t-i} = \Lambda_{t+1} + \Lambda_1 \sum_{i=1}^t V_i \Omega^{i-1} \Omega^\perp \Omega^{t-i}, \end{aligned}$$

532 where the approximation follows since the higher order terms of  $V_t$  are negligible under Assumption 1.  
533  $\square$

## 534 F Proof of Lemma 2

535 **Lemma 2.** Conditioned on  $\mathcal{F}_t$ , the estimation noise in iteration  $t$ , denoted  $\mathcal{E}_t$ , is a zero-mean random  
536 vector with the mean square error given by

$$\mathbb{E}_t[\|\mathcal{E}_t\|^2] = \frac{1}{K(\Lambda_t \mathbf{q}^\top)} \left( (\zeta_t + 1) \text{tr}(\Sigma_t - \Xi_t) + \frac{1}{\Lambda_t \mathbf{q}^\top} \text{tr}(\Xi_t) \right).$$

537 *Proof.* Conditioned on  $\mathcal{F}_t$ , the user reports  $Y_{1,t}, \dots, Y_{K,t}$  which are independent random vectors  
538 following a 2-component multivariate Gaussian mixture distribution<sup>4</sup>

$$f_{Y_{k,t}}(y) = \tilde{\Lambda}_t \mathbf{q}^\top \phi(y, m_t, \Sigma_t) + [1 - \tilde{\Lambda}_t \mathbf{q}^\top] \phi(y, \vec{0}, \Xi_t \mathbf{I}).$$

539 According to the definition of gradient estimator,  $\hat{m}_t$  (6), we can find the distribution of  $\hat{m}_t$  as follows:

$$\hat{m}_t(\mathbf{Y}_t) \sim \sum_{\ell=0}^K \binom{K}{\ell} (\tilde{\Lambda}_t \mathbf{q}^\top)^\ell (1 - \tilde{\Lambda}_t \mathbf{q}^\top)^{K-\ell} \mathcal{N}\left(\ell \frac{m_t}{K \Lambda_t \mathbf{q}^\top}, \frac{\ell \Sigma_t + (K - \ell) \Xi_t \mathbf{I}}{K^2 (\Lambda_t \mathbf{q}^\top)^2}\right).$$

<sup>4</sup>We denote the multivariate Gaussian distribution of an  $N$ -dimensional random vector with mean vector  $\mu$  and covariance matrix  $\Sigma$  as follows:

$$\phi(\mathbf{x}, \mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^N \det(\Sigma)}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^\top \Sigma^{-1}(\mathbf{x} - \mu)\right).$$

540 Recall that in (14), we decompose the gradient estimator as  $\hat{m}_t = m_t(1 + \zeta_t) + \mathcal{E}_t = m_t \frac{\tilde{\Lambda}_t \mathbf{q}^\top}{\Lambda_t \mathbf{q}^\top}$ . After  
 541 minor rearranging, it is easy to show that

$$f_{\mathcal{E}_t}(\epsilon) = \sum_{\ell=0}^K \binom{K}{\ell} (\tilde{\Lambda}_t \mathbf{q}^\top)^\ell (1 - \tilde{\Lambda}_t \mathbf{q}^\top)^{K-\ell} \phi\left(\epsilon, \frac{m_t}{\Lambda_t \mathbf{q}^\top} \left(\frac{\ell}{K} - \tilde{\Lambda}_t \mathbf{q}^\top\right), \frac{\ell \Sigma_t + (K-\ell) \Xi_t}{K^2 (\Lambda_t \mathbf{q}^\top)^2}\right).$$

542 Then, we have

$$\begin{aligned} \mathbb{E}_t[\mathcal{E}_t] &= \sum_{\ell=0}^K \binom{K}{\ell} (\tilde{\Lambda}_t \mathbf{q}^\top)^\ell (1 - \tilde{\Lambda}_t \mathbf{q}^\top)^{K-\ell} \frac{m_t}{\Lambda_t \mathbf{q}^\top} \left(\frac{\ell}{K} - \tilde{\Lambda}_t \mathbf{q}^\top\right), \\ &= \frac{m_t}{\Lambda_t \mathbf{q}^\top} \left(\frac{K \tilde{\Lambda}_t \mathbf{q}^\top}{K} - \tilde{\Lambda}_t \mathbf{q}^\top\right) = \vec{0} \end{aligned}$$

543 and<sup>5</sup>

$$\begin{aligned} \mathbb{E}_t[\|\mathcal{E}_t\|_2^2] &= \text{tr} \left( \sum_{\ell=0}^K \binom{K}{\ell} (\tilde{\Lambda}_t \mathbf{q}^\top)^\ell (1 - \tilde{\Lambda}_t \mathbf{q}^\top)^{K-\ell} \frac{\ell \Sigma_t + (K-\ell) \Xi_t \mathbf{I}}{K^2 (\Lambda_t \mathbf{q}^\top)^2} \right), \\ &= \text{tr} \left( \frac{1}{K (\Lambda_t \mathbf{q}^\top)} \left( \frac{\tilde{\Lambda}_t \mathbf{q}^\top}{\Lambda_t \mathbf{q}^\top} \Sigma_t + \frac{1 - \tilde{\Lambda}_t \mathbf{q}^\top}{\Lambda_t \mathbf{q}^\top} \Xi_t \right) \right), \\ &= \frac{1}{K (\Lambda_t \mathbf{q}^\top)} \left( \frac{\tilde{\Lambda}_t \mathbf{q}^\top}{\Lambda_t \mathbf{q}^\top} \text{tr}(\Sigma_t - \Xi_t) + \frac{\text{tr}(\Xi_t)}{\Lambda_t \mathbf{q}^\top} \right), \\ &= \frac{1}{K (\Lambda_t \mathbf{q}^\top)} \left( (\zeta_t + 1) \text{tr}(\Sigma_t - \Xi_t) + \frac{1}{\Lambda_t \mathbf{q}^\top} \text{tr}(\Xi_t) \right). \end{aligned}$$

544 □

## 545 G Proof of Theorem 1

546 **Theorem 1.** Under Assumptions 2 and 3, suppose that the SGSU (5) is run for  $T$  iterations with a  
 547 fixed stepsize  $\bar{\beta}$  satisfying

$$0 < \bar{\beta} \leq \frac{1}{L(1 + H_T)}.$$

548 Then, the SGD algorithm with strategic users satisfies that

$$\mathbb{E} \left[ \frac{1}{T} \sum_{t=1}^T \|\nabla F(\theta_t)\|^2 \right] \leq \frac{LE_T}{K(1 - H_T)} + \frac{2(F(\theta_1) - F_{\inf})}{\bar{\beta}T(1 - H_T)}.$$

549 *Proof.* By Lemma 4.2 in [3], under Assumption 3(i), the iterates generated by SGSU satisfy

$$\begin{aligned} F(\theta_{t+1}) - F(\theta_t) &\leq m_t^\top (\theta_{t+1} - \theta_t) + \frac{1}{2} L \|\theta_{t+1} - \theta_t\|_2^2 \\ &\leq -\bar{\beta} m_t^\top \hat{m}_t(\mathbf{Y}_t) + \frac{1}{2} \bar{\beta}^2 L \|\hat{m}_t(\mathbf{Y}_t)\|_2^2. \end{aligned}$$

550 Taking the expectation of this inequality conditioned on  $\mathcal{F}_t$ , and noting that  $\theta_t$  is  $\mathcal{F}_t$ -measurable, we  
 551 obtain

$$\mathbb{E}_t[F(\theta_{t+1})] - F(\theta_t) \leq -\bar{\beta} m_t^\top \mathbb{E}_t[\hat{m}_t(\mathbf{Y}_t)] + \frac{1}{2} \bar{\beta}^2 L \mathbb{E}_t[\|\hat{m}_t(\mathbf{Y}_t)\|_2^2]. \quad (29)$$

552 Recall that in (14), we decompose the gradient estimator as  $\hat{m}_t = m_t(1 + \zeta_t) + \mathcal{E}_t$ . Note that  $m_t$   
 553 and  $\zeta_t$  are  $\mathcal{F}_t$ -measurable. Further, by Lemma 2, we have  $\mathbb{E}_t[\mathcal{E}_t] = \vec{0}$ . Consequently, it follows that

$$m_t^\top \mathbb{E}_t[\hat{m}_t(\mathbf{Y}_t)] = m_t^\top \mathbb{E}_t[m_t(1 + \zeta_t)] = \|m_t\|_2^2 (1 + \zeta_t) \quad (30)$$

<sup>5</sup>The trace of an  $n \times n$  square matrix  $\mathbf{A}$ , where  $a_{ii}$  denotes the entry on the  $i^{\text{th}}$  row and  $i^{\text{th}}$  column of  $\mathbf{A}$ , is defined as

$$\text{tr}(\mathbf{A}) = \sum_{i=1}^n a_{11} + a_{22} + \cdots + a_{nn}.$$

554 and

$$\mathbb{E}_t [\|\hat{m}_t(\mathbf{Y}_t)\|_2^2] = \|m_t\|_2^2(1 + \zeta_t)^2 + \mathbb{E}_t [\|\mathcal{E}_t\|_2^2]. \quad (31)$$

555 Using (30) and (31) in (29), we have

$$\mathbb{E}_t[F(\theta_{t+1})] - F(\theta_t) \leq \frac{1}{2}\bar{\beta}^2 L \mathbb{E}_t [\|\mathcal{E}_t\|_2^2] - \bar{\beta} \|m_t\|_2^2(1 + \zeta_t) \left(1 - \frac{1}{2}\bar{\beta} L(1 + \zeta_t)\right).$$

556 From (16) and (19), it follows that

$$(1 + \zeta_t) \left[1 - \frac{1}{2}\bar{\beta} L(1 + \zeta_t)\right] \geq (1 + \zeta_t) \left[1 - \frac{1}{2}\bar{\beta} L(1 + H_T)\right] \geq \frac{1}{2}(1 + \zeta_t)$$

557 and

$$\begin{aligned} \mathbb{E}_t[F(\theta_{t+1})] - F(\theta_t) &\leq \frac{1}{2}\bar{\beta}^2 L \mathbb{E}_t [\|\mathcal{E}_t\|_2^2] - \frac{1}{2}\bar{\beta} \|m_t\|_2^2(1 + \zeta_t), \\ &\leq \frac{1}{2}\bar{\beta}^2 L \mathbb{E}_t [\|\mathcal{E}_t\|_2^2] - \frac{1}{2}\bar{\beta} \|m_t\|_2^2(1 - H_T). \end{aligned}$$

558 Using (18) in the above inequality, we obtain

$$\mathbb{E}_t[F(\theta_{t+1})] - F(\theta_t) \leq -\frac{1}{2}\bar{\beta} \|m_t\|^2(1 - H_T) + \frac{1}{2}\bar{\beta}^2 L \frac{E_T}{K}.$$

559 Taking the total expectation of this inequality yields that

$$\mathbb{E}[F(\theta_{t+1})] - \mathbb{E}[F(\theta_t)] \leq -\frac{1}{2}\bar{\beta} \mathbb{E} [\|m_t\|^2] (1 - H_T) + \frac{1}{2}\bar{\beta}^2 L \frac{E_T}{K}.$$

560 Summing both sides of this inequality for  $t \in \{1, \dots, T\}$  and by Assumption 3(ii), it follows that

$$F_{\inf} - F(\theta_1) \leq \mathbb{E}[F(\theta_{T+1})] - F(\theta_1) \leq -\frac{1}{2}\bar{\beta}(1 - H_T) \sum_{t=1}^T \mathbb{E}_t [\|m_t\|_2^2] + \frac{1}{2}T\bar{\beta}^2 L \frac{E_T}{K}.$$

561 Rearranging the terms in the above inequality, we conclude that

$$\mathbb{E} \left[ \sum_{t=1}^T \|m_t\|_2^2 \right] \leq T \frac{\bar{\beta} L E_T}{K(1 - H_T)} + \frac{2(F_{\inf} - F(\theta_1))}{\bar{\beta}(1 - H_T)}.$$

562

□

## 563 H Proof of Theorem 2

564 **Theorem 2.** Under Assumption 2 and 4, the gradient estimation bias term  $\zeta_t$ , has the following form:

$$\zeta_t = (p_1 - p_2) \sum_{i=1}^{t-1} \frac{\Lambda_i \mathbf{q}^\top}{\Lambda_t \mathbf{q}^\top} \frac{m_i^\top \mathcal{E}_i}{\|m_i\|^2} h_i(m_i) \Delta_{i,t}$$

565 with

$$|\Delta_{i,t}| \leq \delta |u_2|^{t-1-i} + \delta^2 h_{t-1}^{\max} |u_2|^{t-2-i} (t - i - 1).$$

566 Further, for some  $0 < \eta < 1$  we have

$$\mathbb{P}(|\zeta_t| < \eta | \alpha_1, \dots, \alpha_{t-1}) > 1 - \frac{\sum_{i=1}^{t-1} \alpha_i^2}{K\eta^2}$$

567 with

$$\alpha_i^2 = \frac{2 \left| (\nu_i^2 - \xi_i^2) + \frac{m_i^\top \Sigma_i m_i}{\|m_i\|^2} \right| + \frac{\xi_i^2}{\Lambda_i \mathbf{q}^\top}}{\|m_i\|^2 (\Lambda_i \mathbf{q}^\top)} \left[ \frac{\Lambda_i \mathbf{q}^\top}{\Lambda_t \mathbf{q}^\top} \right]^2 h_i^2 \Delta_{i,t}^2.$$

568 *Proof.* Recall that, in (14), the gradient estimator can be decomposed as

$$\hat{m}_t = m_t(1 + \zeta_t) + \mathcal{E}_t, \quad \text{with} \quad \zeta_t = \frac{\tilde{\Lambda}_t \mathbf{q}^\top}{\Lambda_t \mathbf{q}^\top} - 1. \quad (32)$$

569 For convenience, define  $\rho_t$  as  $\rho_t := \tilde{\Lambda}_t - \Lambda_t$  and note that

$$\zeta_t = \frac{\rho_t \mathbf{q}^\top}{\Lambda_t \mathbf{q}^\top}. \quad (33)$$

570 From Lemma 2.1, it follows that

$$\begin{aligned} \rho_t &= \tilde{\Lambda}_t - \Lambda_t = \Lambda_1 \sum_{i=1}^{t-1} V_i \Omega^{i-1} \Omega^\perp \Omega^{t-1-i}, \\ &= \Lambda_1 \left( \sum_{i=1}^{t-2} V_i \Omega^{i-1} \Omega^\perp \Omega^{t-2-i} \right) \Omega + V_{t-1} \Lambda_1 \Omega^{t-2} \Omega^\perp, \\ &= \rho_{t-1} \Omega + V_{t-1} \Lambda_{t-1} \Omega^\perp. \end{aligned} \quad (34)$$

571 Use (32) and (33) in the linear approximation of  $V_t$  in (20). It follows that

$$\begin{aligned} V_t &= \frac{m_t^\top (\hat{m}_t - m_t)}{\|m_t\|_2^2} h_t(m_t) = \left[ \frac{m_t^\top \mathcal{E}_t}{\|m_t\|_2^2} + \zeta_t \right] h_t(m_t), \\ &= h_t(m_t) \frac{m_t^\top \mathcal{E}_t}{\|m_t\|_2^2} + h_t(m_t) \frac{\rho_t \mathbf{q}^\top}{\Lambda_t \mathbf{q}^\top}. \end{aligned} \quad (35)$$

572 Recall that in (12), we define  $\Omega^\perp$  and  $\omega^\perp$  as

$$\Omega^\perp = \mathbf{q}^\top \omega^\perp, \quad \text{with } \omega^\perp := (p_1 - p_2)[-1 \ 0 \ 1 \ 0]. \quad (36)$$

573 For the purpose of brevity, we may simply write  $h_t$  instead of  $h_t(m_t)$ . Use (35) and (36) in (34), and  
574 it follows that

$$\rho_t = \rho_{t-1} \Omega + V_{t-1} \Lambda_{t-1} \Omega^\perp = \rho_{t-1} \Omega + \left( h_{t-1} \frac{m_{t-1}^\top \mathcal{E}_{t-1}}{\|m_{t-1}\|_2^2} + h_{t-1} \frac{\rho_{t-1} \mathbf{q}^\top}{\Lambda_{t-1} \mathbf{q}^\top} \right) \Lambda_{t-1} \mathbf{q}^\top \omega^\perp.$$

575 Recall that  $\mathbf{q}$  and  $\Lambda_t$ 's are  $1 \times 4$  row-vectors, whereas  $m_t$  and  $\mathcal{E}_t$  are  $n \times 1$  column-vectors. After  
576 minor rearrangements, we have

$$\rho_t = \Lambda_{t-1} \mathbf{q}^\top \frac{m_{t-1}^\top \mathcal{E}_{t-1}}{\|m_{t-1}\|_2^2} h_{t-1} \omega^\perp + \rho_{t-1} (\Omega + h_{t-1} \mathbf{q}^\top \omega^\perp). \quad (37)$$

577 Repeat (37) and we obtain

$$\begin{aligned} \rho_t &= \Lambda_{t-2} \mathbf{q}^\top \frac{m_{t-2}^\top \mathcal{E}_{t-2}}{\|m_{t-2}\|_2^2} h_{t-2} \omega^\perp (\Omega + h_{t-1} \mathbf{q}^\top \omega^\perp) + \Lambda_{t-1} \mathbf{q}^\top \frac{m_{t-1}^\top \mathcal{E}_{t-1}}{\|m_{t-1}\|_2^2} h_{t-1} \omega^\perp \\ &\quad + \rho_{t-2} (\Omega + h_{t-2} \mathbf{q}^\top \omega^\perp) (\Omega + h_{t-1} \mathbf{q}^\top \omega^\perp). \end{aligned}$$

578 For convenience, we define  $\mathcal{W}_{t,t'}$  as follows:

$$\mathcal{W}_{t,t'} := \begin{cases} (\Omega + h_t \Omega^\perp) \times \cdots \times (\Omega + h_{t'} \Omega^\perp) & \text{if } t \leq t', \\ \mathbf{I} & \text{if } t > t'. \end{cases}$$

579 Applying (37) repeatedly through iteration  $t \in \mathbb{N}$ , we can obtain that

$$\rho_t = \rho_1 \mathcal{W}_{1,t-1} + \sum_{i=1}^{t-1} \Lambda_i \mathbf{q}^\top \frac{m_i^\top \mathcal{E}_i}{\|m_i\|_2^2} h_i \omega^\perp \mathcal{W}_{i+1,t-1} \quad (38a)$$

580 and

$$\zeta_t = \frac{\rho_t \mathbf{q}^\top}{\Lambda_t \mathbf{q}^\top} = \frac{\rho_1 \mathcal{W}_{1,t-1} \mathbf{q}^\top}{\Lambda_t \mathbf{q}^\top} + \sum_{i=1}^{t-1} \frac{\Lambda_i \mathbf{q}^\top}{\Lambda_t \mathbf{q}^\top} \frac{m_i^\top \mathcal{E}_i}{\|m_i\|_2^2} h_i \omega^\perp \mathcal{W}_{i+1,t-1} \mathbf{q}^\top. \quad (38b)$$

581 Since  $h_t$ 's are small positive values, along the same line as in the proof of Lemma 1, we have conclude  
582 that for  $i < t-2$ ,

$$\mathcal{W}_{i+1,t-1} \approx \Omega^{t-i-1} + \sum_{j=0}^{t-i-2} h_{i+j+1} \Omega^j \Omega^\perp \Omega^{t-2-i-j}. \quad (39)$$

By Approximation 2.3,  $\Omega$  can be diagonalized as  $\Omega = \Gamma \mathcal{U} \Gamma^{-1}$  with  $\mathcal{U}$  has the eigenvalues of  $\Omega$  in descending order of magnitude. Writing  $\Gamma$  (and  $\Gamma^{-1}$ ) as a block matrix of its column (and row) vectors  $\vec{b}_\ell$  (and  $\vec{\gamma}_\ell$ ),  $\omega^\perp \Omega^j \mathbf{q}^\top$  can be expressed as

$$\omega^\perp \Omega^j \mathbf{q}^\top = \sum_{\ell=1}^4 u_\ell^j (\omega^\perp \vec{b}_\ell) (\vec{\gamma}_\ell \mathbf{q}^\top).$$

Since  $\Omega$  is row stochastic,  $u_1 = 1$  and  $1 \geq |u_2| \geq |u_3| \geq |u_4| \geq 0$ . Further, the column eigenvector associated to the eigenvalue 1 is  $\vec{b}_1 = [1 \ 1 \ 1 \ 1]^\top$ . Thus,  $\omega^\perp \vec{b}_1 = 0$ . It follows that

$$\begin{aligned} \omega^\perp \Omega^{j_1} \Omega^\perp \Omega^{j_2} \mathbf{q}^\top &= \omega^\perp \Gamma \mathcal{U}^{j_1} \Gamma^{-1} \mathbf{q}^\top \omega^\perp \Gamma \mathcal{U}^{j_2} \Gamma^{-1} \mathbf{q}^\top, \\ &= \sum_{\ell_1, \ell_2=2}^4 u_{\ell_1}^{j_1} u_{\ell_2}^{j_2} \omega^\perp \vec{b}_{\ell_1} \omega^\perp \vec{b}_{\ell_2} \vec{\gamma}_{\ell_1} \mathbf{q}^\top \vec{\gamma}_{\ell_2} \mathbf{q}^\top. \end{aligned}$$

Thus,

$$|\omega^\perp \Omega^{j_1} \Omega^\perp \Omega^{j_2} \mathbf{q}^\top| \leq \delta^2 |u_2|^{j_1+j_2}, \quad \delta := \left( \max_{\ell \in \{2,3,4\}} |\omega^\perp \vec{b}_\ell| \right) \left( \max_{\ell \in \{2,3,4\}} |\vec{\gamma}_\ell \mathbf{q}^\top|^2 \right). \quad (40)$$

Recall that we define  $h_t^{\max}$  as  $h_t^{\max} := \max_{i \in \{1, \dots, t\}} h_i(m_i)$ . By using (40) in (39), we obtain that

$$\begin{aligned} |\omega^\perp \mathcal{W}_{i+1, t-1} \mathbf{q}^\top| &\leq |\omega^\perp \Omega^{t-i-1} \mathbf{q}^\top| + \sum_{j=0}^{t-i-2} h_{i+j+1} |\omega^\perp \Omega^j \Omega^\perp \Omega^{t-2-i-j} \mathbf{q}^\top|, \\ &\leq \delta |u_2|^{t-1-i} + \delta^2 |u_2|^{t-2-i} \sum_{j=i+1}^{t-1} h_j, \\ &\leq \delta |u_2|^{t-1-i} + \delta^2 |u_2|^{t-2-i} (t-i-1) h_{t-1}^{\max}. \end{aligned} \quad (41)$$

Use (41) in (38), and we obtain

$$\zeta_t = \frac{\rho_1 \mathcal{W}_{1, t-1} \mathbf{q}^\top}{\Lambda_t \mathbf{q}^\top} + \sum_{i=1}^{t-1} \frac{\Lambda_i \mathbf{q}^\top}{\Lambda_t \mathbf{q}^\top} \frac{m_i^\top \mathcal{E}_i}{\|m_i\|_2^2} h_i \Delta_{i, t}, \quad (42)$$

with  $|\Delta_{i, t}| \leq \delta |u_2|^{t-1-i} + \delta^2 |u_2|^{t-2-i} (t-i-1) h_{t-1}^{\max}$ . We assume the FC has the knowledge of the initial state distribution of the repeated games between, i.e.,  $\Lambda_1 = \tilde{\Lambda}_1$ . Thus,  $\rho_1 = [0 \ 0 \ 0 \ 0]$  and we obtain

$$\zeta_t = \sum_{i=1}^{t-1} \frac{\Lambda_i \mathbf{q}^\top}{\Lambda_t \mathbf{q}^\top} \frac{m_i^\top \mathcal{E}_i}{\|m_i\|_2^2} h_i \Delta_{i, t}, \quad \text{with } |\Delta_{i, t}| \leq \delta |u_2|^{t-1-i} + \delta^2 |u_2|^{t-2-i} (t-i-1) h_{t-1}^{\max}. \quad (43)$$

Now, we are going to show that, for some  $0 < \eta < 1$ , we have

$$\mathbb{P}(|\zeta_t| < \eta | \alpha_1, \dots, \alpha_{t-1}) > 1 - \frac{\sum_{i=1}^{t-1} \alpha_i^2}{K \eta^2},$$

with

$$\alpha_i^2 = \frac{2 \left| (\nu_i^2 - \xi_i^2) + \frac{m_i^\top \Sigma_i m_i}{\|m_i\|_2^2} \right| + \frac{\xi_i^2}{\Lambda_i \mathbf{q}^\top}}{\|m_i\|_2^2 (\Lambda_i \mathbf{q}^\top)} \left( \frac{\Lambda_i \mathbf{q}^\top}{\Lambda_t \mathbf{q}^\top} \right)^2 h_i^2(m_i) \Delta_{i, t}^2.$$

In the proof of Lemma 3.2, noting that  $m_t$  is  $\mathcal{F}$ -measurable, we derive the conditional distribution of  $\mathcal{E}_t$  given  $\mathcal{F}_t$  as

$$f_{\mathcal{E}_t}(\epsilon) = \sum_{\ell=0}^K \binom{K}{\ell} (\tilde{\Lambda}_t \mathbf{q}^\top)^\ell (1 - \tilde{\Lambda}_t \mathbf{q}^\top)^{K-\ell} \phi \left( \epsilon, \frac{m_t}{\Lambda_t \mathbf{q}^\top} \left( \frac{\ell}{K} - \tilde{\Lambda}_t \mathbf{q}^\top \right), \frac{\ell (\Sigma_t + \nu_t^2 \mathbf{I}) + (K-\ell) \xi_t^2 \mathbf{I}}{K^2 (\Lambda_t \mathbf{q}^\top)^2} \right).$$

Thus, for  $i \in \{1, 2, \dots, t\}$ , conditioned on  $\mathcal{F}_i$ , we obtain

$$m_i^\top \mathcal{E}_i \sim \sum_{\ell=0}^K \binom{K}{\ell} (\tilde{\Lambda}_i \mathbf{q}^\top)^\ell (1 - \tilde{\Lambda}_i \mathbf{q}^\top)^{K-\ell} \mathcal{N} \left( \frac{\|m_i\|_2^2 \ell}{K \Lambda_i \mathbf{q}^\top}, \frac{\ell (m_i^\top \Sigma_i m_i + \nu_i^2 \|m_i\|_2^2) + (K-\ell) \xi_i^2 \|m_i\|_2^2}{K^2 (\Lambda_i \mathbf{q}^\top)^2} \right).$$

599 After some algebra

$$\mathbb{E} [m_i^\top \mathcal{E}_i | \mathcal{F}_t] = 0$$

600 and

$$\text{Var} [m_i^\top \mathcal{E}_i | \mathcal{F}_t] = \frac{(\zeta_i + 1) (\|m_i\|_2^2 (\nu_i^2 - \xi_i^2) + m_i^\top \Sigma_i m_i) + \frac{\|m_i\|_2^2 \xi_i^2}{\Lambda_i \mathbf{q}^\top}}{K (\Lambda_i \mathbf{q}^\top)}.$$

601 Under Assumption 1, for  $t \in \{1, 2, \dots, T\}$ , we have that  $|\zeta_t| \leq 1$ . Thus,

$$\text{Var} [m_i^\top \mathcal{E}_i | \mathcal{F}_t] \leq \frac{\tilde{\alpha}_i^2}{K} \text{ with } \tilde{\alpha}_i^2 := \frac{2 (\|m_i\|_2^2 (\nu_i^2 - \xi_i^2) + m_i^\top \Sigma_i m_i) + \frac{\|m_i\|_2^2 \xi_i^2}{\Lambda_i \mathbf{q}^\top}}{\Lambda_i \mathbf{q}^\top}.$$

602 For convenience, we define  $\varphi_{i,t}$  as

$$\varphi_{i,t} := \frac{\Lambda_i \mathbf{q}^\top h_i \Delta_{i,t}}{\Lambda_t \mathbf{q}^\top \|m_i\|_2^2}.$$

603 From (43), it is clear that

$$\zeta_t = \sum_{i=1}^{t-1} \varphi_{i,t} m_i^\top \mathcal{E}_i.$$

604 Further,

$$\begin{aligned} \mathbb{E} [\zeta_t | \tilde{\alpha}_1^2 \varphi_{1,t}^2, \dots, \tilde{\alpha}_{t-1}^2 \varphi_{t-1,t}^2] &= 0, \\ \text{Var} [\zeta_t | \tilde{\alpha}_1^2 \varphi_{1,t}^2, \dots, \tilde{\alpha}_{t-1}^2 \varphi_{t-1,t}^2] &\leq \sum_{i=1}^{t-1} \varphi_{i,t}^2 \frac{\tilde{\alpha}_i^2}{K}. \end{aligned}$$

605 Noting that  $\alpha_i^2 = \tilde{\alpha}_i^2 \varphi_{i,t}^2$ , via Chebyshev's inequality, we obtain the desired result

$$\begin{aligned} \mathbb{P} (|\zeta_t| < \eta | \alpha_1, \dots, \alpha_{t-1}) &> 1 - \frac{\sum_{i=1}^{t-1} \varphi_{i,t}^2 \alpha_i^2}{K \eta^2}, \\ &> 1 - \frac{\sum_{i=1}^{t-1} \alpha_i^2}{K \eta^2}. \end{aligned}$$

606 □

## 607 I Experiments

608 In this section, we provide additional details and results of our empirical studies. In our first set of  
 609 experiments, we consider a binary logistic classification problem and use the KDD-Cup 04 dataset [6].  
 610 The goal of binary logistic classification experiments is to learn a classification rule that differentiates  
 611 between two types of particles generated in high energy collider experiments based on 78 attributes  
 612 [6]. The dataset is of 50000 samples and it is publicly available. The learning rate is chosen as  
 613  $\bar{\eta} = 0.01$ . When the users take the defective action, they send zero-mean Gaussian noise,  $Y_{k,t} = \Upsilon_{k,t}$   
 614 and  $\Upsilon_{k,t} \sim \mathcal{N}(0, \Xi_t)$ , where the covariance matrix is  $\Xi_t = \xi_t^2 \mathbf{I}$  and  $\xi_t = 0.5$ . When the users  
 615 take the cooperative action, they send a privacy-preserved version of their stochastic gradients,  
 616  $Y_{k,t} = X_{k,t} + N_{k,t}$  and  $N_{k,t} \sim \mathcal{N}(0, \nu_t^2 \mathbf{I})$  with  $\nu_t = 0.1$ . The number of iterations is 2000. The  
 617 number of users is chosen as  $K = 50$  and mini-batch size is  $s = 10$ . Each experiment is repeated  
 618 250 times and their average is taken.

619 In our second set of experiments, we consider a three layer neural network trained on the MNIST  
 620 dataset. We conduct image classification experiments on MNIST image classification dataset, which  
 621 is composed of 50K images for training and 10K images for testing. All experiments are conducted  
 622 in PyTorch on a server. We use convolutional neural network (CNN) with 2 convolutional layers  
 623 followed by 1 fully connected layer, with a total of 5142 parameters. We repeat each experiment  
 624 10 times and take the average. Training loss is cross-entropy loss and testing accuracy is true

Equalizer Fusion Center Strategy ( $\mathbf{p} = [0.8 \ 0.5 \ 0.4 \ 0.1]$ )		
User Strategy	Steady State Distribution	Prob. of Coop.
Full Coop.: $\mathbf{q} = [0.90 \ 0.90 \ 0.90 \ 0.90]$	$\Lambda^* = [0.58 \ 0.03 \ 0.32 \ 0.07]$	$\Lambda^* \mathbf{q}^\top = 0.90$
Stubborn: $\mathbf{q} = [0.90 \ 0.15 \ 0.90 \ 0.15]$	$\Lambda^* = [0.38 \ 0.08 \ 0.22 \ 0.32]$	$\Lambda^* \mathbf{q}^\top = 0.60$
Tit-for-tat: $\mathbf{q} = [0.90 \ 0.90 \ 0.15 \ 0.15]$	$\Lambda^* = [0.32 \ 0.07 \ 0.12 \ 0.49]$	$\Lambda^* \mathbf{q}^\top = 0.44$
Pavlov: $\mathbf{q} = [0.90 \ 0.15 \ 0.15 \ 0.90]$	$\Lambda^* = [0.42 \ 0.08 \ 0.24 \ 0.26]$	$\Lambda^* \mathbf{q}^\top = 0.67$
Coin Toss: $\mathbf{q} = [0.50 \ 0.50 \ 0.50 \ 0.50]$	$\Lambda^* = [0.28 \ 0.13 \ 0.22 \ 0.37]$	$\Lambda^* \mathbf{q}^\top = 0.50$
Extortion Fusion Center Strategy ( $\mathbf{p} = [0.95 \ 0.75 \ 0.2 \ 0]$ )		
User Strategy	Steady State Distribution	Prob. of Coop.
Full Coop.: $\mathbf{q} = [0.90 \ 0.90 \ 0.90 \ 0.09]$	$\Lambda^* = [0.67 \ 0.05 \ 0.23 \ 0.05]$	$\Lambda^* \mathbf{q}^\top = 0.90$
Stubborn: $\mathbf{q} = [0.90 \ 0.15 \ 0.90 \ 0.15]$	$\Lambda^* = [0.40 \ 0.08 \ 0.20 \ 0.32]$	$\Lambda^* \mathbf{q}^\top = 0.60$
Tit-for-tat: $\mathbf{q} = [0.90 \ 0.90 \ 0.15 \ 0.15]$	$\Lambda^* = [0.28 \ 0.02 \ 0.10 \ 0.60]$	$\Lambda^* \mathbf{q}^\top = 0.38$
Pavlov: $\mathbf{q} = [0.90 \ 0.15 \ 0.15 \ 0.90]$	$\Lambda^* = [0.44 \ 0.09 \ 0.22 \ 0.25]$	$\Lambda^* \mathbf{q}^\top = 0.67$
Coin Toss: $\mathbf{q} = [0.50 \ 0.50 \ 0.50 \ 0.50]$	$\Lambda^* = [0.25 \ 0.15 \ 0.25 \ 0.35]$	$\Lambda^* \mathbf{q}^\top = 0.50$

Table 1: The steady state distribution  $\Lambda^*$  without perturbations for different FC and user strategies.

classification rate. The learning rate is chosen as,  $\bar{\eta} = 0.1$ . For the image classification problem,  $\Xi_t = \xi_t^2 \mathbf{I}$  with  $\xi_t = 0.071$  and  $\nu_t = 0.032$ .

In all experiments, the payoff parameters are set as  $R = r = 1$  and  $V_{\text{FC}} = V_{\text{U}} = 0.5$ . Recall that, the FC can set the expected payoff of the user to a fixed value between 0 (mutual defection payoff) and  $r - V_{\text{U}} R$  (mutual cooperation payoff) with the adoption of equalizers. In the experiments, we have tested the performance of an equalizer strategy  $\mathbf{p} = [0.8 \ 0.5 \ 0.4 \ 0.1]$  ( $\varphi_1 = -0.6$  and  $\varphi_2 = 0.1$ ). With these choices, the expected payoff of the users can be found as:

$$s_{\text{U}}^* = -\frac{\varphi_2}{\varphi_1} = (r - V_{\text{U}} R) \frac{p_4}{1 - p_1 + p_4} = \frac{1}{6} \approx 0.167. \quad (44)$$

Alternatively, the FC can enforce an extortionate share of payoffs in the steady state of the game with the adoption of extortioners. In the experiments, we have also tested the performance of an extortioner strategy  $\mathbf{p} = [0.95 \ 0.75 \ 0.2 \ 0]$  ( $\chi = 2$  and  $\gamma = 0.5$ ). With these choices, the ratio of the expected payoffs can be found as:

$$\frac{s_{\text{FC}}^*}{s_{\text{U}}^*} = \chi = 2. \quad (45)$$

We consider how ZD strategies fare against the stochastic versions of some of the most important memory-one strategies:

1. Full cooperation:  $\mathbf{q}^\top = [0.9 \ 0.9 \ 0.9 \ 0.9]$ ,
2. Stubborn:  $\mathbf{q}^\top = [0.9 \ 0.15 \ 0.9 \ 0.15]$ ,
3. Tit-for-tat:  $\mathbf{q}^\top = [0.9 \ 0.9 \ 0.15 \ 0.15]$ ,
4. Pavlov:  $\mathbf{q}^\top = [0.9 \ 0.15 \ 0.15 \ 0.9]$ ,
5. Coin Toss:  $\mathbf{q}^\top = [0.5 \ 0.5 \ 0.5 \ 0.5]$ .

In Table 1, the steady state distribution  $\Lambda^*$  of the Markov transition matrix  $\Omega$  and the probability of user cooperation at the steady state are listed. As Lemma 1 indicates, due to the perturbations on the state transition matrix, the real state distribution at the steady state will be a noisy version of  $\Lambda^*$ . Thus, in the experiments, we expect the probability of user cooperation  $\bar{\Lambda}_t \mathbf{q}^\top$  converge to  $\Lambda^* \mathbf{q}^\top$ .

Fig. 3a and Fig. 3b depict the optimality gap of the optimization problem across iterations. For the full cooperation, coin toss, tit-for-tat and stubborn user strategies, SGSU converges quickly. For

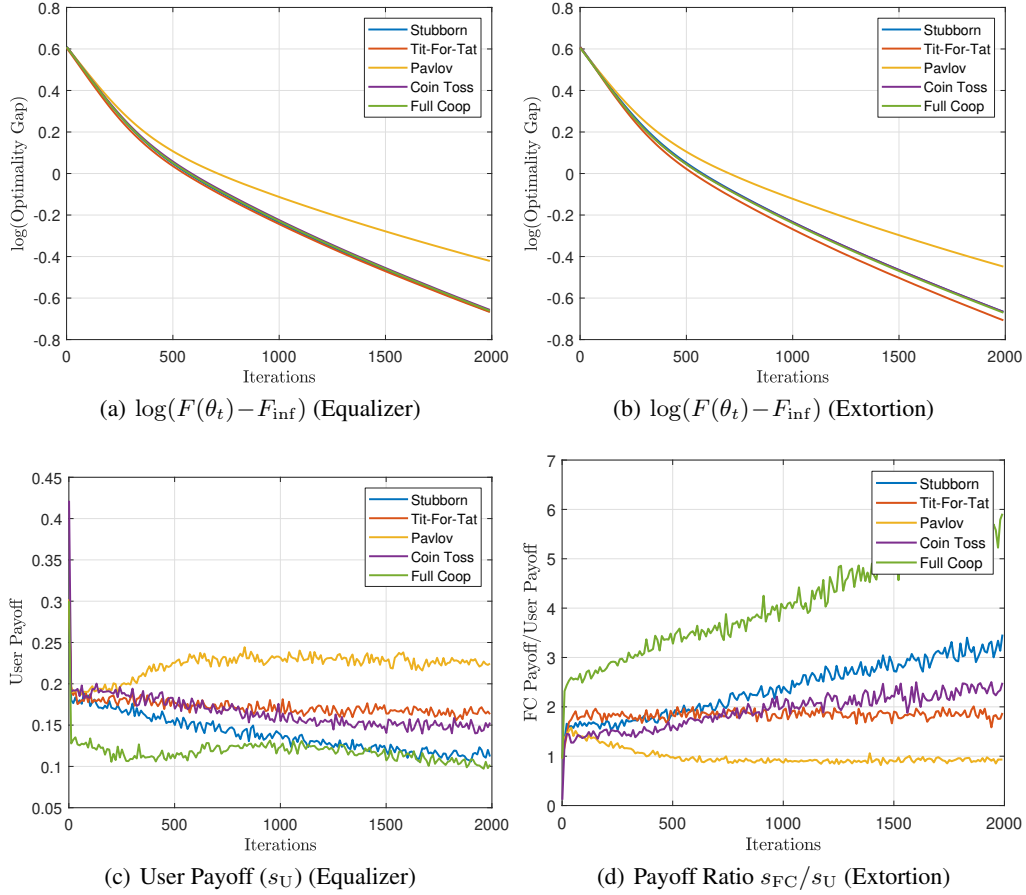


Figure 3: SGD with strategic users (logistic classification problem)

649 Pavlov user strategies, SGSU can eventually approach, albeit more slowly than other cases. Fig. 3c  
 650 shows average user payoff across iterations when the FC employs the equalizer strategy. For coin  
 651 toss and tit-for-tat user strategies, the average user payoffs are very close to 0.167 as expected from  
 652 (44). For full cooperation and stubborn strategies, the average user payoffs are observed to be slightly  
 653 lower than 0.167. One unanticipated finding is that the average user payoffs are significantly higher  
 654 than 0.167 when the users employ Pavlov strategies. Fig. 3d shows the ratio of average FC and user  
 655 payoffs across iterations when the FC employs the extortion strategy. For coin toss and tit-for-tat  
 656 user strategies, the ratio is very close to 2. This is consistent with (45). Contrary to expectations, the  
 657 payoff ratio for the Pavlov strategies is almost 1.

658 This discrepancy could be attributed to that Pavlov users are very sensitive to both type I ( $\hat{B}_{k,t} = \hat{d}$   
 659 given that  $B_{k,t} = c$ ) and type II ( $\hat{B}_{k,t} = \hat{c}$  given that  $B_{k,t} = d$ ) classification errors. When the users  
 660 employ the Pavlov strategy, they are *trusting avengers who is exploiting yet repentant* [21], i.e., they  
 661 cooperate if rewarded for cooperating or punished for defecting. In contrast, stubborn users are not  
 662 sensitive to the classification errors and they repeat their last action with a high probability. Tit-for-tat  
 663 users are retaliatory players and they cooperate if rewarded and they are not sensitive to Type II  
 664 errors.

665 Fig. 4a and 4b illustrate the probability of user cooperation,  $\tilde{\Lambda}_t \mathbf{q}^\top$ , across different user strategies.  
 666 The experimental results validate Lemma 1 and the empirical user cooperation probabilities match  
 667 the values in Table 1, except when the users are Pavlov. Unsurprisingly, when the users follow full  
 668 cooperation (or coin toss) strategy, they cooperate with probability 0.9 (or 0.5) regardless of the actual  
 669 states of the repeated games. For the cases with stubborn and tit-for-tat users, the games quickly  
 670 converge to the steady state distribution. Interestingly, for the cases with Pavlov users, the probability  
 671 of user cooperation decreases over time. This is associated to the performance of the linear classifier.



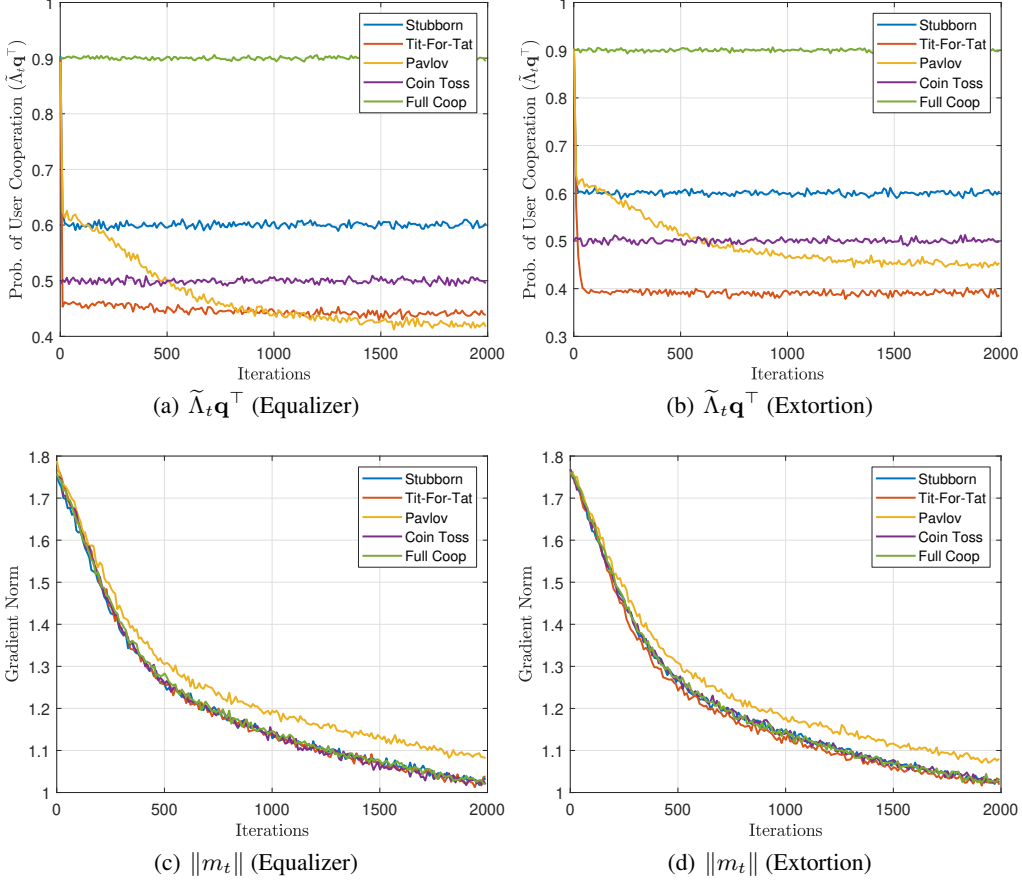


Figure 4: Dynamics of the repeated games (logistic classification problem)

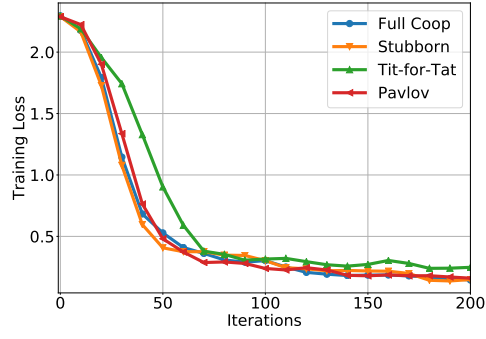
Fig. 4c and 4d illustrate that the gradients vanish as the algorithm converges to  $\theta^*$ . Thus, the linear classifier cannot differentiate between the cooperative and defective hypothesis.

Fig. 5(a) and Fig. 5(b) plot the training loss and the testing accuracy for the image classification experiments. As expected, the best performance is achieved when the users employ the full cooperation strategy. For the cases with Pavlov or stubborn users, the convergence rate is similar to the case with full cooperation. When the users employ tit-for-tat strategy, the convergence rate is slower.

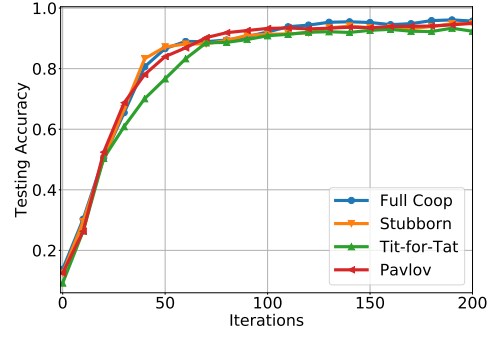
Fig. 6 plots further experiments for different scenarios. Fig. 6(a) depicts the optimality gap (TFT users) for different mini-batch sizes and number of users for the logistic class. problem. Fig 6(b) illustrates an intriguing case where the strategy vector of user  $k$  is  $\mathbf{q}_k = [0.9 + \Delta_k, 0.9 + \Delta_k, 0.1 - \Delta_k, 0.1 - \Delta_k]$  s.t.  $\Delta_k \sim \text{Unif}[-0.1, 0.1]$  and  $\xi_k \sim \mathcal{N}(0.5, 0.1)$ . The FC knows the means of  $\mathbf{q}_k$  and  $\xi_k$ ; however, she does not have the knowledge of exact realizations. In these experiments, the imperfect knowledge of  $\mathbf{q}_k$  and  $\xi_k$  does not make a significant impact on the convergence rate. As we discuss in Section 5, we are trying to extend our work using the robust statistics. Fig 7(a) shows that the performance of the proposed sample mean based gradient estimator diminishes as the noise level of the defective user signals increase. The FC can overcome this problem by employing a geometric median based gradient estimator.

## J Classifier Design for the User Action Classification Problem

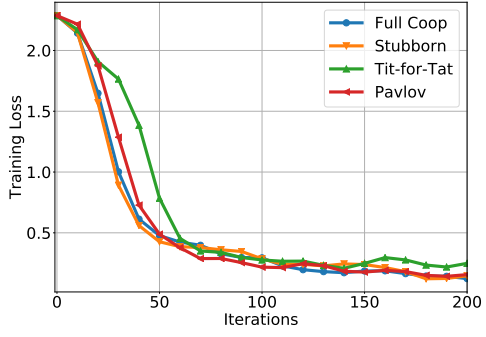
In Section 3, we study the joint gradient estimation and user action classification problem. Recall that a user is cooperative ( $B_{k,t} = c$ ) if he is sending the privacy-preserved version of his gradient  $X_{k,t}$ .



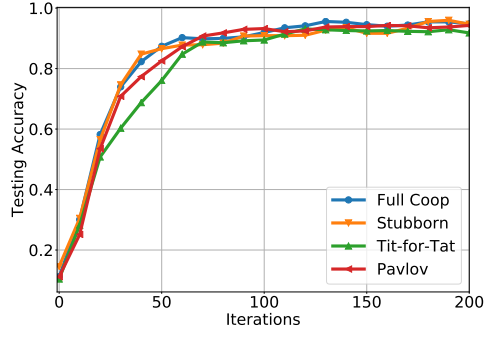
(a) Equalizer



(b) Equalizer

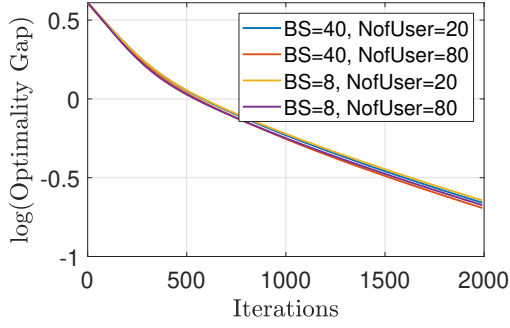


(c) Extortion

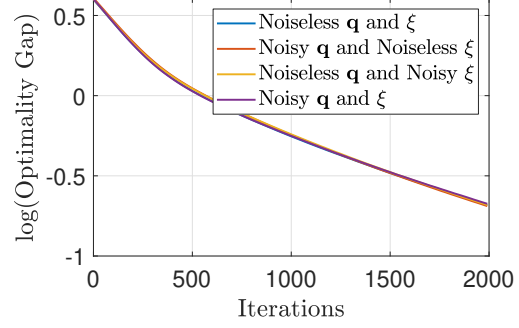


(d) Extortion

Figure 5: SGD with strategic users (image classification experiments on MNIST)

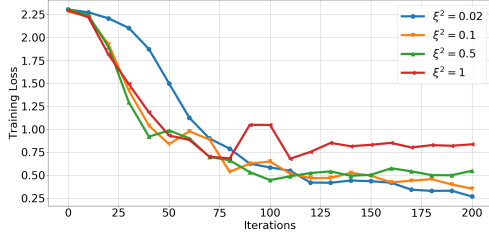


(a) Batch size and num. of users

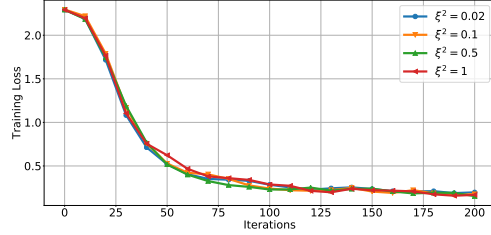


(b) Imperfect knowledge of  $\mathbf{q}$  and  $\xi$

Figure 6: Dynamics of the repeated games (logistic classification problem)



(a) Sample Mean Based Estimation



(b) Geometric Median Based Estimation

Figure 7: SGD with strategic users (image classification experiments on MNIST - Tit-For-Tat Users)

691 Otherwise, the user is defective and sends a noise signal  $\Upsilon_{k,t} \sim \mathcal{N}(0, \Xi_t)$  independent of  $X_{k,t}$ :

$$Y_{k,t} = \begin{cases} X_{k,t} + N_{k,t}, & \text{if } B_{k,t} = c \text{ (cooperative);} \\ \Upsilon_{k,t}, & \text{if } B_{k,t} = d \text{ (defective).} \end{cases}$$

692 Since the stochastic gradient in any iteration (at each user) is computed using many i.i.d. samples,  
693 the noise in the stochastic gradient can be approximated by a zero-mean Gaussian random vector  
694 [27, 17, 38, 23]:

$$X_{k,t} = m_t + W_{k,t} \quad (46)$$

695 where  $m_t$  is the population gradient,  $m_t = \nabla_{\theta} F(\theta_t)$ , and  $W_{k,t}$  is the noise in the stochastic gradients.  
696 This implies  $Y_{k,t} \sim \mathcal{N}(m_t, \Sigma_t)$ , given  $B_{k,t} = c$ , where  $\Sigma_t := \text{cov}[W_{k,t}] + \nu_t^2 \mathbf{I}$ . Thus, the user  
697 action classification problem may be cast as follows:

$$\mathcal{H}_c : Y_{k,t} \sim \mathcal{N}(m_t, \Sigma_t) \quad \text{vs.} \quad \mathcal{H}_d : Y_{k,t} \sim \mathcal{N}(0, \Xi_t), \quad (47)$$

698 where the hypothesis  $\mathcal{H}_c$  (or  $\mathcal{H}_d$ ) is that the user is cooperative (or defective). The maximum a  
699 posteriori decision rule, given  $Y_{k,t}$ , for the binary hypothesis testing problem (47) is of the form [32]:

$$\delta_t(Y_{k,t}) = \begin{cases} \mathcal{H}_c & \text{if } L_t(Y_{k,t}) > \ln \frac{P(\mathcal{H}_d)}{P(\mathcal{H}_c)} \\ \mathcal{H}_d & \text{else} \end{cases}$$

700 where  $L_t(y)$  is the log-likelihood ratio between  $\mathcal{H}_c$  and  $\mathcal{H}_d$ :

$$L_t(Y_{k,t}) = \frac{1}{2} Y_{k,t}^\top (\Xi_t^{-1} - \Sigma_t^{-1}) Y_{k,t} + m_t^\top \Sigma_t^{-1} \left( Y_{k,t} - \frac{1}{2} m_t \right) + \frac{1}{2} \ln \frac{|\Xi_t|}{|\Sigma_t|}.$$

701 Note that  $L_t(Y_{k,t})$  here consists of a quadratic term, a linear term and a constant. The resulting  
702 classifier structure is known as a linear-quadratic (LQ) classifier. In general, we cannot evaluate  
703 the detection probability  $\Phi_t$  and false alarm rate  $\Psi_t$  associated to the LQ classifier. Recall that  
704 the adaptive strategies for fusion center (9) use the knowledge of  $\Phi_t$  and  $\Psi_t$ . Therefore, direct  
705 employment of LQ detector introduces formidable technical difficulties and is not practical. To get  
706 a more concrete sense, in what follows we study two special cases where the LQ detector can be  
707 further simplified.

### 708 J.1 Linear Classifier vs. Quadratic Classifier

709 If the two covariance matrices have roughly the same energy level under the cooperative and defective  
710 hypotheses; i.e.,  $\Sigma_t \approx \Xi_t$ , then the quadratic term in the log-likelihood ratio disappears

$$L_t(Y_{k,t}) = m_t^\top \Sigma_t^{-1} \left( Y_{k,t} - \frac{1}{2} m_t \right) + \frac{1}{2} \ln \frac{|\Xi_t|}{|\Sigma_t|}.$$

711 and we have a linear test statistic:

$$\delta_t^L(Y_{k,t}) = \begin{cases} \mathcal{H}_c & \text{if } m_t^\top \Sigma_t^{-1} Y_{k,t} > \frac{1}{2} m_t^\top \Sigma_t^{-1} m_t + \ln \frac{P(\mathcal{H}_d)}{P(\mathcal{H}_c)} \\ \mathcal{H}_d & \text{else.} \end{cases}$$

712 This yields a linear classifier which has the structure of a matched filter or a correlation detector. The  
713 associated detection and false alarm probabilities can be found in terms of Gaussian  $\mathcal{Q}$ -functions.

714 **Remark 5.** In practice, the FC does not have the knowledge of  $m_t$  and  $\Sigma_t$ ; therefore, she cannot  
715 directly employ  $\delta_t^L$ . Under the homoscedasticity assumption ( $\Sigma_t = \sigma_t^2 \mathbf{I}$ ),  $\delta_t^L$  reduces to the proposed  
716 user action classification rule in (7). The proposed classifier uses the gradient estimator  $\hat{m}_t$  (6)  
717 instead of  $m_t$ . Following the same rationale, it may be argued that  $\Sigma_t$  can also be estimated in  
718 a similar manner and employed in the user action classification. In practice, this poses the risk  
719 of introducing matrix estimation errors into the user action classification scheme. In contrast, the  
720 proposed classifier is more robust and tractable with reduced computational complexity.

721 In general, the linear classifier allows us to trade accuracy for overall tractability of the algorithm by  
722 “neglecting” the information contained in the average energy of the reported gradients compared to  
723 the LQ classifier. This may be reasonable in the early stages of the SGD algorithm when the reported  
724 gradients still significantly differ under the cooperative and defective hypotheses in terms of their

means. Nevertheless, when the iterates of SGD are confined to a small enough region around a local optimum of the loss, the gradients vanish and the linear classifier starts to fail differentiating  $\mathcal{H}_c$  and  $\mathcal{H}_d$ . To tackle this challenge, in the late stages of the training, the classifier can be altered to utilize the information contained in the statistical energies of the reported gradients.

For the ease of exposition, we will first consider the case where  $\Xi_t = \xi_t^2 \mathbf{I}$ . If the (statistical) average amplitude and energy of the population gradient is very small, i.e.,  $m_t \approx 0$  and  $\text{Cov}[W_{k,t}] \approx 0$ , then the structure of log-likelihood ratio is quadratic:

$$L_t(Y_{k,t}) = \frac{1}{2} \left( \frac{1}{\xi_t^2} - \frac{1}{\nu_t^2} \right) \|Y_{k,t}\|^2 + \frac{n}{2} \ln \frac{\xi_t^2}{\nu_t^2}$$

where  $n$  is the dimension of the optimization problem (1). The quadratic classifier can be found as:

$$\delta_t(Y_{k,t}) = \begin{cases} \mathcal{H}_c & \text{if } \|Y_{k,t}\|^2 < \mathcal{T}_t := \left( n \ln \frac{\xi_t^2}{\nu_t^2} - 2 \ln \frac{\text{P}(\mathcal{H}_d)}{\text{P}(\mathcal{H}_c)} \right) \frac{\xi_t^2 \nu_t^2}{\xi_t^2 - \nu_t^2} \\ \mathcal{H}_d & \text{else.} \end{cases} \quad (48)$$

**Remark 6.** If  $\text{Cov}[W_{k,t}] = \sigma_t^2 \mathbf{I}$  and  $\Xi_t = \xi_t^2 \mathbf{I}$ , then the false alarm rate,  $\Psi_t^Q$ , and the detection probability,  $\Phi_t^Q$ , associated to the quadratic classifier (48) can be found as <sup>6</sup>:

$$\Psi_t^Q = \text{P}(\|Y_{k,t}\|^2 < \mathcal{T}_t | B_{k,t} = d) = \frac{1}{\Gamma(n/2)} \gamma \left( \frac{n}{2}, \frac{\mathcal{T}_t}{2\xi_t^2} \right), \quad (49a)$$

$$\Phi_t^Q = \text{P}(\|Y_{k,t}\|^2 < \mathcal{T}_t | B_{k,t} = c) = 1 - \mathcal{Q}_{\frac{n}{2}} \left( \frac{\|m_t\|_2}{\sqrt{\sigma_t^2 + \nu_t^2}}, \sqrt{\frac{\mathcal{T}_t}{\sigma_t^2 + \nu_t^2}} \right). \quad (49b)$$

*Proof.* Under the defective action hypothesis,  $\|Y_{k,t}\|^2$  is the sum of independent and zero-mean Gaussian random variables with variance  $\xi_t^2$ . Thus, it follows that

$$\Psi_t^Q = \text{P}(\|Y_{k,t}\|^2 < \mathcal{T}_t | B_{k,t} = d) = \frac{1}{\Gamma(n/2)} \gamma \left( \frac{n}{2}, \frac{\mathcal{T}_t}{2\xi_t^2} \right). \quad (50)$$

Under the cooperative action hypothesis, we have  $Y_{k,t} \sim \mathcal{N}(m_t, \Sigma_t)$ . Assuming  $\Sigma_t$  is invertible,  $\Sigma_t^{-1}$  is a positive semi-definite matrix and it can be decomposed into  $\Sigma_t^{-1} = \Sigma_t^{-1/2} \Sigma_t^{-1/2}$  where  $\Sigma_t^{-1/2}$  is also positive semi-definite. Use the spectral theorem and write  $\Sigma_t = \mathbf{S}_t^\top \mathbf{E}_t \mathbf{S}_t$  where  $\mathbf{S}_t$  is an orthogonal matrix (i.e.,  $\mathbf{S}_t^\top \mathbf{S}_t = \mathbf{S}_t \mathbf{S}_t^\top = \mathbf{I}$ ) and  $\mathbf{E}_t$  is diagonal with positive elements. It follows that

$$\begin{aligned} Y_{k,t}^\top Y_{k,t} &= Y_{k,t}^\top \Sigma_t^{-1/2} \Sigma_t \Sigma_t^{-1/2} Y_{k,t} = \left( \Sigma_t^{-1/2} Y_{k,t} \right)^\top \Sigma_t \left( \Sigma_t^{-1/2} Y_{k,t} \right), \\ &= \left( \Sigma_t^{-1/2} Y_{k,t} \right)^\top \mathbf{S}_t^\top \mathbf{E}_t \mathbf{S}_t \left( \Sigma_t^{-1/2} Y_{k,t} \right) = \left( \mathbf{S}_t \Sigma_t^{-1/2} Y_{k,t} \right)^\top \mathbf{E}_t \left( \mathbf{S}_t \Sigma_t^{-1/2} Y_{k,t} \right), \\ &= (U_{k,t} + \mu_t)^\top \mathbf{E}_t (U_{k,t} + \mu_t) \end{aligned}$$

<sup>6</sup>For complex numbers, the gamma function is defined as

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx, \quad \Re(z) > 0.$$

For any positive integer  $z$ ,  $\Gamma(z) = (z-1)!$ . The lower incomplete gamma function is defined as

$$\gamma(s, x) = \int_0^x t^{s-1} e^{-t} dt, \quad \text{with } s > 0.$$

The Marcum-Q-function  $Q_M$  is defined as

$$Q_M(a, b) = \int_b^\infty x \left( \frac{x}{a} \right)^{M-1} \exp \left( -\frac{x^2 + a^2}{2} \right) I_{M-1}(ax) dx,$$

with modified Bessel function  $I_{M-1}$  of order  $M-1$ .

742 where  $U_{k,t} := \mathbf{S}_t \Sigma_t^{-1/2} (Y_{k,t} - m_t)$  and  $\mu_t = \mathbf{S}_t \Sigma_t^{-1/2} m_t$ .

743 Note that  $U_{k,t}$  is multivariate normal with identity covariance matrix and expectation zero. In general,  
 744 there is no known analytical expression for the PDF and CDF of  $Y_{k,t}^\top Y_{k,t}$ . We can find the closed  
 745 form expression of  $\Phi_t^Q$  for the special case of  $\text{cov}[W_{k,t}] = \sigma_t^2 \mathbf{I}$ . In this case,  $\Sigma_t = (\sigma_t^2 + \nu_t^2) \mathbf{I}$  and  
 746  $\|Y_{k,t}\|^2 / (\sigma_t^2 + \nu_t^2)$  is distributed according to the noncentral chi-square distribution with  $n$  degrees  
 747 of freedom and the noncentrality parameter  $\frac{1}{\sigma_t^2 + \nu_t^2} \|m_t\|_2^2$ . Consequently,

$$\begin{aligned} \Phi_t^Q &= \text{P}(\|Y_{k,t}\|^2 < \mathcal{T}_t | B_{k,t} = c) = \text{P}\left(\|U_{k,t} + \mu_t\|_2^2 < \frac{\mathcal{T}_t}{\sigma_t^2 + \nu_t^2}\right), \\ &= 1 - \mathcal{Q}_{\frac{n}{2}}\left(\frac{\|m_t\|}{\sqrt{\sigma_t^2 + \nu_t^2}}, \sqrt{\frac{\mathcal{T}_t}{\sigma_t^2 + \nu_t^2}}\right). \end{aligned}$$

748

□

749 In the next subsection, we discuss how the FC can employ the quadratic classifier in practice.  
 750 Furthermore, we compare the performances of linear and quadratic classifiers using real-life datasets.

## 751 J.2 Numerical Results

752 The proposed adaptive strategies (9) require the knowledge of the detection probability  $\Phi_t$  and  
 753 the false alarm rate  $\Psi_t$  of the classifier employed by the fusion center. For the quadratic detector,  
 754 in general,  $\Phi_t^Q$  does not have a closed form expression. Instead, the fusion center can form her  
 755 approximated estimate  $\hat{\Phi}_t^Q$  under the assumptions that  $m_t \approx \vec{0}$  and  $\Sigma_t \approx \nu_t^2 \mathbf{I}$ :

$$\hat{\Phi}_t^Q = \frac{1}{\Gamma(n/2)} \gamma\left(\frac{n}{2}, \frac{\mathcal{T}_t}{2\nu_t^2}\right).$$

756 We repeat the experiments for the binary logistic classification problem with the same parameters  
 757 (recall that the noise parameters are set as  $\nu_t = 0.1$  and  $\xi_t = 0.5$ ). Fig 8a and Fig 8b depict the  
 758 optimality gap of the optimization problem across iterations. We observe that SGSU converges  
 759 quickly for all cases. Fig 8c shows average user payoff across iterations when the FC employs the  
 760 equalizer strategy. For all case, the average user payoffs are very close to the theoretic result 0.167 in  
 761 (44). Fig 8d shows the ratio of average FC and user payoffs across iterations when the FC employs the  
 762 extortion strategy, and it is very close to 2. This is consistent with (45). Further, Fig. 8e and Fig. 8f  
 763 illustrate the probability of user cooperation  $\hat{\Lambda}_t \mathbf{q}_t$  for different user strategies. The experimental  
 764 results validate Lemma 1 and the empirical user cooperation probabilities match the values in Table  
 765 1. The improvement in the experimental results is due to the performance of the user classification  
 766 scheme.

767 Next, we consider another set of experiments where the noise parameters are set as  $\xi_t = 0.4$  and  
 768  $\nu_t = 0.2$ . In these experiments, we test the performance of an equalizer strategy  $\mathbf{p} = [0.8 \ 0.5 \ 0.4 \ 0.1]$   
 769 against probabilistic Pavlov strategies  $\mathbf{q} = [0.9 \ 0.15 \ 0.15 \ 0.9]$ . Fig. 9a depicts the detection probabil-  
 770 ity of the linear and quadratic classifiers. We observe that, in the early stages of the optimization,  
 771 the linear classifier outperforms the quadratic classifier because the average energy of the reported  
 772 gradients under cooperative and defective hypotheses are close to each other. As the gradients vanish,  
 773 the performance of the quadratic classifier improves. Fig. 9b illustrates the optimality gap of the  
 774 optimization problem across iterations. We also include a third option where the fusion center starts  
 775 with the linear classifier and switches to the quadratic classifier during the run of the optimization.  
 776 The results indicate that this third option outperforms the cases where the FC only employs the  
 777 quadratic classifier or the linear classifier.

778 We also repeat the experiments for the image classification problem using the quadratic classifier  
 779 with threshold  $\mathcal{T}_t = 200$ , and  $\xi_t^2 = 0.71$  and  $\nu_t^2 = 0.032$ . Note that the noise level of the defective  
 780 user signals is significantly increased in contrast to the experiments with the linear classifier. Fig 10  
 781 plots the training loss and the testing accuracy for different scenarios.

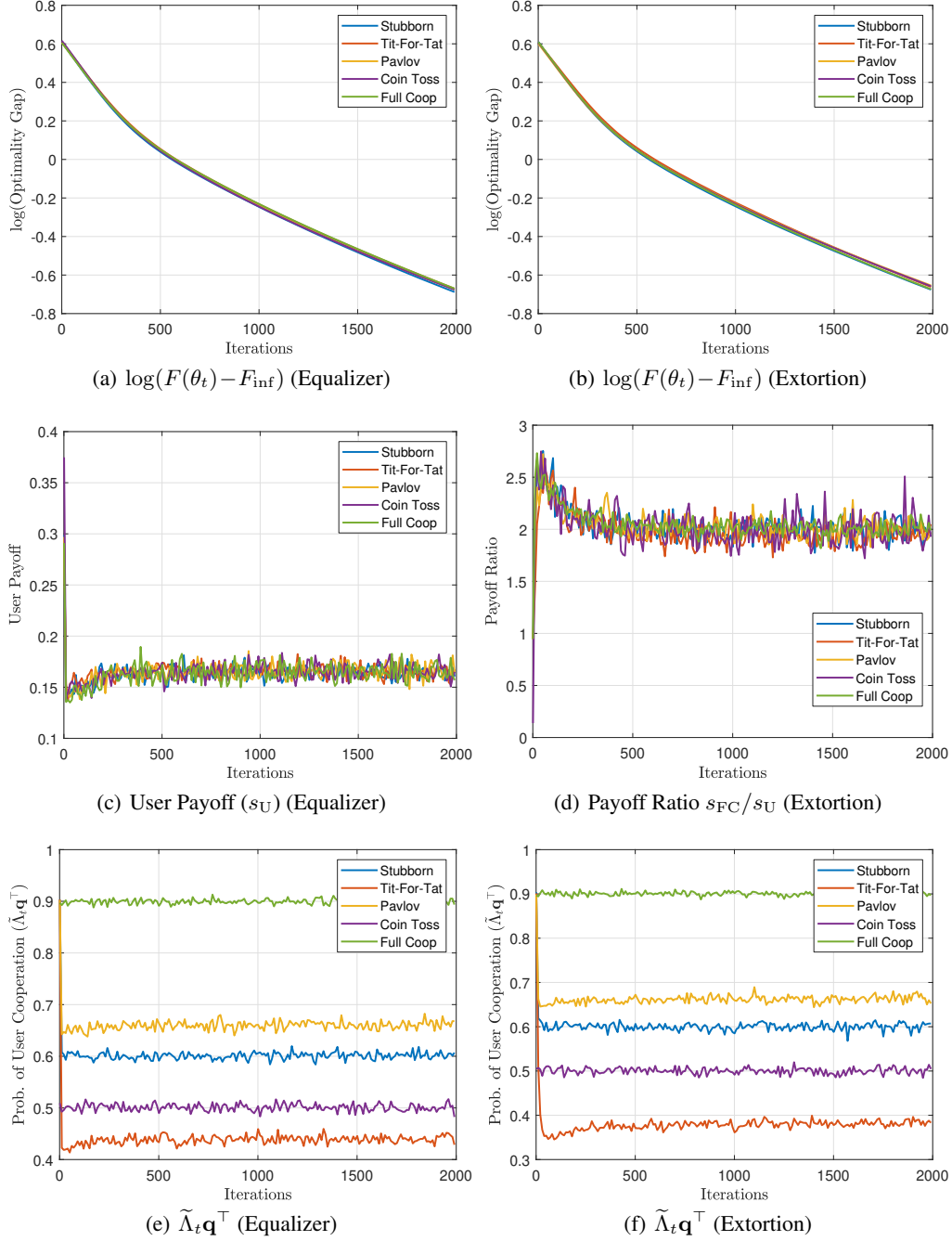


Figure 8: SGSU with quadratic classifier (logistic classification problem)

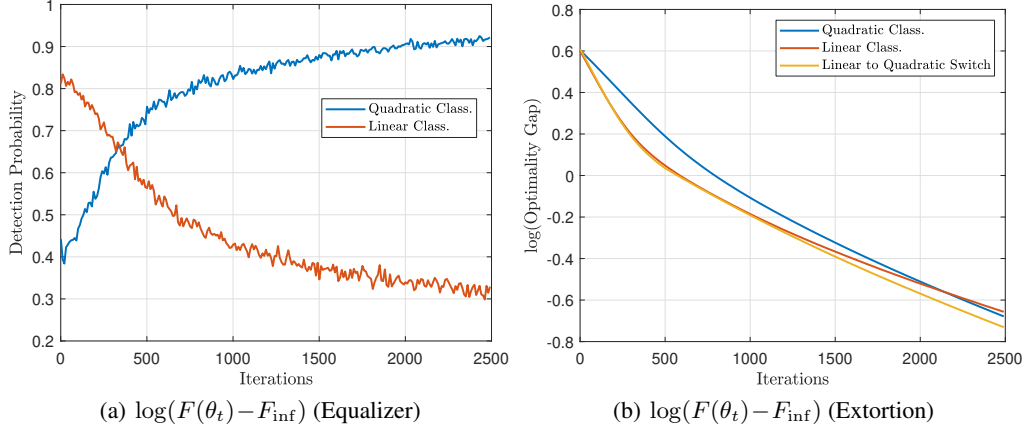


Figure 9: SGSU with linear and quadratic classifier (logistic classification problem)

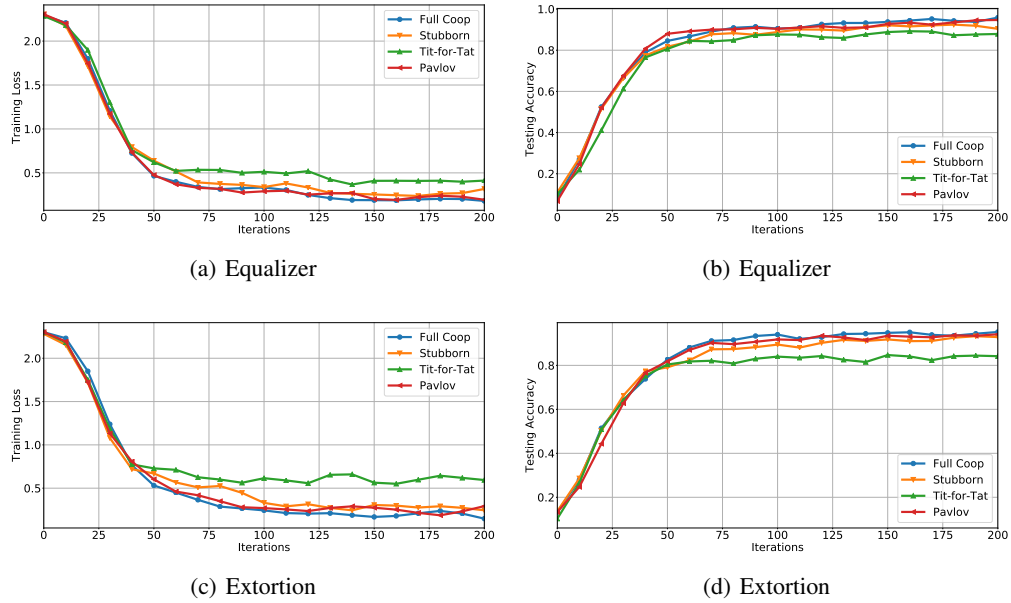


Figure 10: SGSU with quadratic classifier (image classification problem)

## 782 K Heterogeneous User Strategies

783 In this section, we generalize the adaptive FC strategies for the case where the users employ  
 784 different strategies. Let  $q_{k,1}, q_{k,2}, q_{k,3}$  and  $q_{k,4}$  denote the probabilities of cooperation for the  
 785 user conditioned on the joint action pair of the previous iteration, that is  $(A_{k,t-1}, B_{k,t-1})$ , in the  
 786 order of  $(C, c), (C, d), (D, c)$  and  $(D, d)$ . Furthermore, the user's strategy vector is defined as  
 787  $\mathbf{q}_k = [q_{k,1} \ q_{k,2} \ q_{k,3} \ q_{k,4}]$ . The FC takes her action after collecting the user reports and her strategies  
 788 are conditioned on the user action,  $B_{k,t}$ . In the repeated game against user  $k$ , analogous to the user  
 789 strategies, let  $p_{k,1}, p_{k,2}, p_{k,3}$  and  $p_{k,4}$  denote the probabilities of cooperation for the FC conditioned  
 790 on  $(A_{k,t-1}, B_{k,t})$ , in the order of  $(C, c), (C, d), (D, c)$  and  $(D, d)$ . The fusion center's strategy vector  
 791 is defined as  $\mathbf{p}_k = [p_{k,1} \ p_{k,2} \ p_{k,3} \ p_{k,4}]$ .

792 In any iteration round, we treat the two actions of the user and the FC as the state of the game in  
 793 iteration  $t$ :  $(A_{k,t}, B_{k,t})$ . The strategy vectors  $\mathbf{p}$  and  $\mathbf{q}$  imply a Markov state transition matrix as

$$\Omega_k = \begin{bmatrix} q_{k,1}p_{k,1} & (1-q_{k,1})p_{k,2} & q_{k,1}(1-p_{k,1}) & (1-q_{k,1})(1-p_{k,2}) \\ q_{k,2}p_{k,1} & (1-q_{k,2})p_{k,2} & q_{k,2}(1-p_{k,1}) & (1-q_{k,2})(1-p_{k,2}) \\ q_{k,3}p_{k,3} & (1-q_{k,3})p_{k,4} & q_{k,3}(1-p_{k,3}) & (1-q_{k,3})(1-p_{k,4}) \\ q_{k,4}p_{k,3} & (1-q_{k,4})p_{k,4} & q_{k,4}(1-p_{k,3}) & (1-q_{k,4})(1-p_{k,4}) \end{bmatrix}. \quad (51)$$

794 Let  $\Lambda_{k,1}$  be the initial state distribution of the game between the FC and user  $k$ . Analogous to the  
 795 case with homogeneous user strategies, we define the deterministic process  $\Lambda_{k,t}$  as follows:

$$\Lambda_{k,t} := \Lambda_{k,1} \Omega_k^{t-1}. \quad (52)$$

796 Given  $\mathbf{Y}_t = \mathbf{y}_t$ , we use the following modified sample mean estimator:

$$\hat{m}_t(\mathbf{y}_t) = \frac{1}{\sum_{k=1}^K \Lambda_{k,t} \mathbf{q}_k^\top} \sum_{k=1}^K y_{k,t} = \frac{1}{\sum_{k=1}^K \Lambda_{k,1} \Omega_k^{t-1} \mathbf{q}_k^\top} \sum_{k=1}^K y_{k,t}. \quad (53)$$

797 The FC can form her prediction about the user's action using the following linear classifier:

$$\hat{B}_{k,t}(\hat{m}_t(\mathbf{y}_t), y_{k,t}) = \begin{cases} \hat{c} & \text{if } y_{k,t} \hat{m}_t > \frac{1}{2} \hat{m}_t^\top \hat{m}_t + \tau_{k,t}, \\ \hat{d} & \text{if } y_{k,t} \hat{m}_t \leq \frac{1}{2} \hat{m}_t^\top \hat{m}_t + \tau_{k,t}. \end{cases} \quad (54)$$

798 The detection probabilities and the false alarm rates associated to the linear classifiers can be found  
 799 as follows:

$$\Phi_{k,t} := \mathbb{P}(\hat{B}_{k,t} = \hat{c} | B_{k,t} = c) = 1 - \mathcal{Q}\left(\frac{\hat{m}_t^\top (m_t - \frac{1}{2} \hat{m}_t) - \tau_{k,t}}{\sqrt{\hat{m}_t^\top \Sigma_t \hat{m}_t}}\right), \quad (55a)$$

$$\Psi_{k,t} := \mathbb{P}(\hat{B}_{k,t} = \hat{c} | B_{k,t} = d) = \mathcal{Q}\left(\frac{\frac{1}{2} \hat{m}_t^\top \hat{m}_t + \tau_{k,t}}{\sqrt{\xi_t^2 \hat{m}_t^\top \hat{m}_t}}\right). \quad (55b)$$

800 The FC forms her estimate  $\hat{\Phi}_{k,t}$  using  $\hat{m}_t$ :

$$\hat{\Phi}_{k,t} = 1 - \mathcal{Q}\left(\frac{\frac{1}{2} \hat{m}_t^\top \hat{m}_t - \tau_{k,t}}{\sqrt{\hat{m}_t^\top (\Sigma_t + \nu_t^2 \mathbf{I}) \hat{m}_t}}\right). \quad (56)$$

801 Due to the errors in gradient estimations, the state transition matrices are time-varying, given by:

$$\tilde{\Omega}_{k,t} = \Omega_k + V_{k,t} \Omega_k^\perp \quad \text{with } V_{k,t} := \frac{\hat{\Phi}_{k,t} - \Phi_{k,t}}{\hat{\Phi}_{k,t} - \Psi_{k,t}} \quad (57a)$$

802 and

$$\Omega_k^\perp := \begin{bmatrix} -q_{k,1}(p_{k,1} - p_{k,2}) & 0 & q_{k,1}(p_{k,1} - p_{k,2}) & 0 \\ -q_{k,2}(p_{k,1} - p_{k,2}) & 0 & q_{k,2}(p_{k,1} - p_{k,2}) & 0 \\ -q_{k,3}(p_{k,3} - p_{k,4}) & 0 & q_{k,3}(p_{k,3} - p_{k,4}) & 0 \\ -q_{k,4}(p_{k,3} - p_{k,4}) & 0 & q_{k,4}(p_{k,3} - p_{k,4}) & 0 \end{bmatrix}. \quad (57b)$$

803 Due to the perturbations on the state transition matrix, the real state distribution  $\tilde{\Lambda}_{k,t}$  is a noisy version  
 804 of  $\Lambda_{k,t}$ :

$$\tilde{\Lambda}_{k,t} = \Lambda_{k,t} + \Lambda_{k,1} \sum_{i=1}^{t-1} V_i \Omega_k^{i-1} \Omega_k^\perp \Omega_k^{t-1-i}. \quad (58)$$

805 For convenience, we define  $\tilde{\Lambda}_t := \{\tilde{\Lambda}_{1,t}, \tilde{\Lambda}_{2,t}, \dots, \tilde{\Lambda}_{K,t}\}$ . Let  $\mathcal{F}_t$  denote the  $\sigma$ -algebra, generated by  
 806  $\{\theta_1, \tilde{\Lambda}_1, \mathbf{Y}_1, \dots, \theta_{t-1}, \tilde{\Lambda}_{t-1}, \mathbf{Y}_{t-1}, \theta_t, \tilde{\Lambda}_t\}$ . Observe that, we can decompose the gradient estimator  
 807  $\hat{m}_t$  as follows:

$$\hat{m}_t(\cdot) = m_t(1 + \zeta_t) + \mathcal{E}_t, \quad (59)$$

808 where  $\zeta_t$  is the estimation bias term due to the perturbations on the state transition matrices, given by

$$\zeta_t = \frac{1}{m_t} (\mathbb{E}_t[\hat{m}_t] - m_t) = \frac{\frac{1}{K} \sum_{k=1}^K \mathbb{P}(B_{k,t} = c | \mathcal{F}_t)}{\frac{1}{K} \sum_{k=1}^K \Lambda_{k,t} \mathbf{q}_k^\top} - 1 = \frac{\frac{1}{K} \sum_{k=1}^K \tilde{\Lambda}_{k,t} \mathbf{q}_k^\top}{\frac{1}{K} \sum_{k=1}^K \Lambda_{k,t} \mathbf{q}_k^\top} - 1$$



809 and  $\mathcal{E}_t$  is the estimation noise, given by

$$\mathcal{E}_t = \hat{m}_t - \mathbb{E}_t[\hat{m}_t].$$

810 To establish convergence guarantees for the SGSU under heterogeneous user strategies in (53),  
 811  $\sum_{k=1}^K \tilde{\Lambda}_{k,t} \mathbf{q}_k^\top$  and  $\sum_{k=1}^K \Lambda_{k,t} \mathbf{q}_k^\top$  must meet the following criteria during the course of the algorithm:

812 **Assumption 5.** We assume that  $\frac{1}{K} \sum_{k=1}^K \Lambda_{k,t} \mathbf{q}_k^\top > \frac{1}{2}$  and  $\sum_{k=1}^K \tilde{\Lambda}_{k,t} \mathbf{q}_k^\top > 0$ , for all  $t \in$   
 813  $\{1, 2, \dots, T\}$ .

814 Assumption 5 is a counterpart of Assumption 2 for the case with heterogeneous user strategies. The  
 815 first condition requires that the average probability of user cooperation dictated by the memory-1  
 816 strategies, in the absence of perturbations, is larger than 0.5. The second condition states that, in  
 817 the presence of perturbations, the average probability of user cooperation always stays positive. By  
 818 Assumption 5, there exists a positive constant  $H_T$  such that

$$0 < |\zeta_t| < H_T < 1, \forall t \in \{1, \dots, T\}. \quad (60)$$

819 We have the following lemma characterizing the properties of estimation noise.

820 **Lemma 3.** Conditioned on  $\mathcal{F}_t$ , the estimation noise in iteration  $t$ , denoted

$$\mathbb{E}_t[\|\mathcal{E}_t\|^2] = \frac{1}{\sum_{k=1}^K \Lambda_{k,t} \mathbf{q}_k^\top} \left( (\zeta_t + 1)[\text{tr}(\Sigma_t) + n(\nu_t^2 - \xi_t^2)] + \frac{n\xi_t^2}{\frac{1}{K} \sum_{k=1}^K \Lambda_{k,t} \mathbf{q}_k^\top} \right). \quad (61)$$

821 *Proof.* Conditioned on  $\mathcal{F}_t$ , the user reports  $Y_{1,t}, \dots, Y_{K,t}$  which are independent random vectors  
 822 following a 2-component multivariate Gaussian mixture distribution

$$f_{Y_{k,t}}(y) = \tilde{\Lambda}_{k,t} \mathbf{q}_k^\top \phi(y, m_t, \Sigma_t + \nu_t^2 \mathbf{I}) + [1 - \tilde{\Lambda}_{k,t} \mathbf{q}_k^\top] \phi(y, \vec{0}, \xi_t^2 \mathbf{I}).$$

823 According to the definition of gradient estimator,  $\hat{m}_t$  (53), we can find the distribution of  $\hat{m}_t$  as the  
 824 following:

$$\begin{aligned} \hat{m}_t(\mathbf{Y}_t) &\sim \sum_{\ell_1=0}^1 \cdots \sum_{\ell_K=0}^1 [\tilde{\Lambda}_{1,t} \mathbf{q}_1^\top]^{\ell_1} [1 - \tilde{\Lambda}_{1,t} \mathbf{q}_1^\top]^{1-\ell_1} \cdots [\tilde{\Lambda}_{K,t} \mathbf{q}_K^\top]^{\ell_K} [1 - \tilde{\Lambda}_{K,t} \mathbf{q}_K^\top]^{1-\ell_K} \\ &\quad \times \mathcal{N} \left( \frac{m_t \sum_{k=1}^K \ell_k}{\sum_{k=1}^K \Lambda_{k,t} \Omega^{t-1} \mathbf{q}_k^\top}, \frac{K \xi_t^2 \mathbf{I} + (\Sigma_t + (\nu_t^2 - \xi_t^2) \mathbf{I}) \sum_{k=1}^K \ell_k}{\left( \sum_{k=1}^K \Lambda_{k,t} \Omega^{t-1} \mathbf{q}_k^\top \right)^2} \right). \end{aligned}$$

825 Thus,

$$\begin{aligned} f_{\mathcal{E}_t}(\epsilon) &= \sum_{\ell_1=0}^1 \cdots \sum_{\ell_K=0}^1 [\tilde{\Lambda}_{1,t} \mathbf{q}_1^\top]^{\ell_1} [1 - \tilde{\Lambda}_{1,t} \mathbf{q}_1^\top]^{1-\ell_1} \cdots [\tilde{\Lambda}_{K,t} \mathbf{q}_K^\top]^{\ell_K} [1 - \tilde{\Lambda}_{K,t} \mathbf{q}_K^\top]^{1-\ell_K} \\ &\quad \times \phi \left( \epsilon, m_t \frac{\sum_{k=1}^K \ell_k - \tilde{\Lambda}_{k,t} \mathbf{q}_k^\top}{\sum_{k=1}^K \Lambda_{k,t} \mathbf{q}_k^\top}, \frac{K \xi_t^2 \mathbf{I} + (\Sigma_t + (\nu_t^2 - \xi_t^2) \mathbf{I}) \sum_{k=1}^K \ell_k}{\left( \sum_{k=1}^K \Lambda_{k,t} \Omega^{t-1} \mathbf{q}_k^\top \right)^2} \right). \end{aligned}$$

826 It follows that

$$\begin{aligned} \mathbb{E}_t[\mathcal{E}_t] &= \sum_{\ell_1=0}^1 \cdots \sum_{\ell_K=0}^1 [\tilde{\Lambda}_{1,t} \mathbf{q}_1^\top]^{\ell_1} [1 - \tilde{\Lambda}_{1,t} \mathbf{q}_1^\top]^{1-\ell_1} \cdots [\tilde{\Lambda}_{K,t} \mathbf{q}_K^\top]^{\ell_K} [1 - \tilde{\Lambda}_{K,t} \mathbf{q}_K^\top]^{1-\ell_K} \\ &\quad \times m_t \frac{\sum_{k=1}^K \ell_k - \tilde{\Lambda}_{k,t} \mathbf{q}_k^\top}{\sum_{k=1}^K \Lambda_{k,t} \mathbf{q}_k^\top} = \frac{m_t}{\sum_{k=1}^K \Lambda_{k,t} \mathbf{q}_k^\top} \left( \sum_{k=1}^K \tilde{\Lambda}_{k,t} \mathbf{q}_k^\top - \sum_{k=1}^K \tilde{\Lambda}_{k,t} \mathbf{q}_k^\top \right) = \vec{0} \end{aligned}$$

827 and

$$\begin{aligned}
\mathbb{E}_t [\|\mathcal{E}_t\|_2^2] &= \text{tr} \left( \sum_{\ell_1=0}^1 \cdots \sum_{\ell_K=0}^1 \left[ \tilde{\Lambda}_{1,t} \mathbf{q}_1^\top \right]^{\ell_1} \left[ 1 - \tilde{\Lambda}_{1,t} \mathbf{q}_1^\top \right]^{1-\ell_1} \cdots \left[ \tilde{\Lambda}_{K,t} \mathbf{q}_K^\top \right]^{\ell_K} \left[ 1 - \tilde{\Lambda}_{K,t} \mathbf{q}_K^\top \right]^{1-\ell_K} \right. \\
&\quad \times \left. \frac{K \xi_t^2 \mathbf{I} + (\Sigma_t + (\nu_t^2 - \xi_t^2) \mathbf{I}) \sum_{k=1}^K \ell_k}{\left( \sum_{k=1}^K \Lambda_{k,1} \Omega^{t-1} \mathbf{q}_k^\top \right)^2} \right) = \frac{\text{tr} \left( K \xi_t^2 \mathbf{I} + (\Sigma_t + (\nu_t^2 - \xi_t^2) \mathbf{I}) \sum_{k=1}^K \tilde{\Lambda}_{k,t} \mathbf{q}_k^\top \right)}{\left( \sum_{k=1}^K \Lambda_{k,t} \mathbf{q}_k^\top \right)^2}, \\
&= \frac{nK \xi_t^2 + [\text{tr}(\Sigma_t) + n(\nu_t^2 - \xi_t^2)] \sum_{k=1}^K \tilde{\Lambda}_{k,t} \mathbf{q}_k^\top}{\left( \sum_{k=1}^K \Lambda_{k,t} \mathbf{q}_k^\top \right)^2}, \\
&= \frac{1}{\sum_{k=1}^K \Lambda_{k,t} \mathbf{q}_k^\top} \left( \frac{\sum_{k=1}^K \tilde{\Lambda}_{k,t} \mathbf{q}_k^\top}{\sum_{k=1}^K \Lambda_{k,t} \mathbf{q}_k^\top} [\text{tr}(\Sigma_t) + n(\nu_t^2 - \xi_t^2)] + \frac{nK \xi_t^2}{\sum_{k=1}^K \Lambda_{k,t} \mathbf{q}_k^\top} \right), \\
&= \frac{1}{\sum_{k=1}^K \Lambda_{k,t} \mathbf{q}_k^\top} \left( (\zeta_t + 1) [\text{tr}(\Sigma_t) + n(\nu_t^2 - \xi_t^2)] + \frac{n \xi_t^2}{\frac{1}{K} \sum_{k=1}^K \Lambda_{k,t} \mathbf{q}_k^\top} \right).
\end{aligned}$$

828

□

829 By (60) and (61), we have that

$$\mathbb{E}_t [\|\mathcal{E}_t\|^2] \leq \frac{E_T}{K} \tag{62}$$

830 with

$$E_T := \frac{1}{\sum_{k=1}^K \Lambda_{k,t} \mathbf{q}_k^\top} \left( (H_T + 1) [\text{tr}(\Sigma_t) + n(\nu_t^2 - \xi_t^2)] + \frac{n \xi_t^2}{\frac{1}{K} \sum_{k=1}^K \Lambda_{k,t} \mathbf{q}_k^\top} \right).$$