Reputation Design for Adaptive Networks with Selfish Agents

Chung-Kai Yu, Mihaela van der Schaar, and Ali H. Sayed

Department of Electrical Engineering
University of California, Los Angeles

***Abstract*—We consider a general information-sharing game over adaptive networks with selfish agents, in which a diffusion strategy is employed to estimate a common target parameter. The benefit and cost of sharing information are embedded into the individual utility functions. We formulate the interactions among selfish agents as successive one-shot games and show that the dominant strategy is for agents not to share information with each other. In order to encourage cooperation among selfish agents, we design a reputation scheme that enables agents to utilize the historic summary of other agents' past actions to predict future returns that would result from being cooperative i.e., from sharing information with other agents. Simulations illustrate the benefits of the combined diffusion and reputation strategies for learning over networks with selfish agents.**

I. Introduction

Adaptive networks consist of agents that are linked together by a topology and which are expected to cooperate through in-network processing to estimate some parameters of interest. Several distributed strategies exist to enable decentralized processing among cooperating agents, such as the consensus strategy (e.g., [1], [2]) and the diffusion strategy (e.g., [3]–[5]). In most prior works, agents have been modeled as cooperative players that participate willingly in the exchange and processing of information even though this incurs a cost for the agents. In this work, we study a broader class of networks where agents can behave in a selfish manner. In this case, agents would participate in the collaborative process and share information with their neighbors only if the cooperation is beneficial to them. For example, agents can weigh the cost of sharing information against the expected improvement in estimation accuracy and then decide on whether to cooperate or not with their neighbors based on the outcome of their evaluation. The interactions among selfish agents can be formulated as successive one-shot games.

Our arguments show that if left unattended, the dominant strategy for all selfish agents in each one-shot game is for them not to participate in the sharing of information, which leads to non-cooperating agents. This result follows from the fact that the agents are not provided with incentives to share information. One useful technique to avoid this inefficient scenario is to associate a reputation measure with each agent (e.g., [6]–[8]). In this method, reputation scores are used by the agents to assess the willingness of other agents to cooperate; the scores are also used to punish non-cooperative behavior. However, and different from conventional repeated games, the benefit of sharing information over adaptive networks generally decreases with time. This is because, as the estimation accuracy improves and, therefore, the benefit to continue cooperating for estimation purposes falls below the communication cost, the act of cooperating with other agents becomes unattractive and inefficient. Conventional reputation designs do not address this depreciation over time in the value of information, which will be examined more closely in our work. In addition, our formulation deals with a multi-user learning/decision game process where the decisions by the agents are coupled together in contrast to conventional Markov decision processes that deal with policies for single-user decision problems. For example, if multiple agents in the network decide not to share information, then other agents would end up learning this fact and may decide not to share data as well.

We therefore focus on studying an information-sharing game over adaptive networks, where agents are selfish and seek to minimize their own cost functions; these functions combine both the estimation accuracy and the communication cost. The purpose of the agents is to estimate a common target parameter. Following the randomly-paired protocol of [9], agents will be randomly matched into pairs at the beginning of each time interval. This situation could occur, for example, due to an exogenous matcher or the mobility of the agents. Based on some prior reference knowledge, each agent evaluates the expected cost of actions and decides whether to share estimates with the other agent. To motivate agents to cooperate with each other, we formulate a reputation scoring mechanism to help agents jointly assess the instantaneous benefit of depreciating information and the transmission cost of sharing information. A key contribution of the present work is that it studies dynamic adaptive scenarios with continuous learning and does not only focus on behavior at the equilibrium state.

Notation: We use lowercase letters to denote vectors and scalars, uppercase letters for matrices, plain letters for deterministic variables, and boldface letters for random variables. All vectors in our treatment are column vectors, with the exception of the regression vectors, $\boldsymbol{u}_{k,i}$.

This work was supported in part by NSF grants CCF-1011918 and CSR-1016081. Emails: {ckyuna,mihaela,sayed}@ee.ucla.edu

978-1-4673-5577-3/13/$31.00 ©2013 IEEE

TABLE I: The one-shot game for information sharing.

	$\boldsymbol{a}_k(i)=0$	$\boldsymbol{a}_k(i)=1$
$\boldsymbol{a}_\ell(i)=0$	$\text{MSD}_{\ell,i}(\boldsymbol{a}_k(i)=0\|\widetilde{\boldsymbol{w}}_{\ell,i-1})$ $\text{MSD}_{k,i}(\boldsymbol{a}_\ell(i)=0\|\widetilde{\boldsymbol{w}}_{k,i-1})$	$\text{MSD}_{\ell,i}(\boldsymbol{a}_k(i)=1\|\widetilde{\boldsymbol{w}}_{\ell,i-1})$ $\text{MSD}_{k,i}(\boldsymbol{a}_\ell(i)=0\|\widetilde{\boldsymbol{w}}_{k,i-1})+c$
$\boldsymbol{a}_\ell(i)=1$	$\text{MSD}_{\ell,i}(\boldsymbol{a}_k(i)=0\|\widetilde{\boldsymbol{w}}_{\ell,i-1})+c$ $\text{MSD}_{k,i}(\boldsymbol{a}_\ell(i)=1\|\widetilde{\boldsymbol{w}}_{k,i-1})$	$\text{MSD}_{\ell,i}(\boldsymbol{a}_k(i)=1\|\widetilde{\boldsymbol{w}}_{\ell,i-1})+c$ $\text{MSD}_{k,i}(\boldsymbol{a}_\ell(i)=1\|\widetilde{\boldsymbol{w}}_{k,i-1})+c$

II. Network Model and Information Structure

A. Reference Knowledge and Transmission Cost

Consider a network with N selfish agents. At each discrete time i, pairs of agents randomly meet together to share information, say, agents k and ℓ. We assume the agents share some preliminary knowledge, denoted by $\mathbb{K}_k$ and $\mathbb{K}_\ell$, which could be exchanged at the time they are paired. Then, according to this knowledge, agent k decides whether to share additional information $\mathbb{I}_{k,i}$ with agent ℓ at time i, and vice-versa. Sharing the information $\mathbb{I}_{k,i}$ with agent ℓ bears some transmission cost for agent k, which is modeled as a positive coefficient $c>0$ and assumed to be known by each agent.

B. Diffusion Strategy for Distributed Estimation

At each time instant $i\geq 0$, each agent k in the network is assumed to have access to a scalar measurement $\boldsymbol{d}_k(i)\in\mathbb{C}$ and a $1\times M$ regression vector $\boldsymbol{u}_{k,i}\in\mathbb{C}^{1\times M}$ with covariance matrix $R_{u,k}\triangleq\mathbb{E}\boldsymbol{u}_{k,i}^*\boldsymbol{u}_{k,i}>0$. The data are assumed to be related via the linear regression model:

$$\boldsymbol{d}_k(i)=\boldsymbol{u}_{k,i}w^o+\boldsymbol{v}_k(i) \tag{1}$$

where $w^o\in\mathbb{C}^{M\times 1}$ is the common target vector to be estimated and $\boldsymbol{v}_k(i)\in\mathbb{C}$ is measurement noise with variance $\sigma_{v,k}^2$. Agents update their estimates of w^o based on their own data $\boldsymbol{d}_k(i)$ and $\boldsymbol{u}_{k,i}$, and on estimates from other neighboring agents if available. We employ the adapt-then-combine (ATC) diffusion strategy [3], [4] due to its enhanced performance in comparison to other strategies, including consensus strategies [10]. According to the diffusion implementation, agent k computes its successive estimates for w^o as follows:

$$\boldsymbol{\psi}_{k,i}=\boldsymbol{w}_{k,i-1}+\mu_k\boldsymbol{u}_{k,i}^*[\boldsymbol{d}_k(i)-\boldsymbol{u}_{k,i}\boldsymbol{w}_{k,i-1}] \tag{2}$$

$$\boldsymbol{w}_{k,i}=\alpha_k\boldsymbol{\psi}_{k,i}+(1-\alpha_k)\boldsymbol{\psi}_{\ell,i} \tag{3}$$

where the parameter μ_k is a positive step-size factor, which is assumed to be sufficiently small and common for all agents, i.e., $\mu_k=\mu\ll 1$. In the first step (2), an intermediate estimate $\boldsymbol{\psi}_{k,i}$ is determined by adapting the existing estimate $\boldsymbol{w}_{k,i-1}$ using local data. The second step (3) uses a non-negative coefficient $0\leq\alpha_k\leq 1$ to combine the intermediate estimates of agents k and ℓ. In the context of algorithm (2)-(3), the information $\mathbb{I}_{k,i}$ and $\mathbb{I}_{\ell,i}$ to be shared are, respectively, the intermediate estimates $\boldsymbol{\psi}_{k,i}$ and $\boldsymbol{\psi}_{\ell,i}$. We assume the reference knowledge $\mathbb{K}_k$ and $\mathbb{K}_\ell$ to be $\sigma_{v,k}^2$ and $\sigma_{v,\ell}^2$, respectively. We note that if agent ℓ decides not to share estimates, then α_k is set to 1.

C. Combined Cost Function

We denote the action of agent k at time i by $\boldsymbol{a}_k(i)$ where $\boldsymbol{a}_k(i)=1$ means "to share" and $\boldsymbol{a}_k(i)=0$ means "not to share". We also denote the instantaneous utility (or combined cost) functions of agents k and ℓ at time i by $J_{k,i}$ and $J_{\ell,i}$, respectively. These utilities are constructed as follows, where each agent k jointly considers the estimation performance and the transmission cost to define $J_{k,i}$ (similarly, for agent ℓ) in terms of the actions $(\boldsymbol{a}_k,\boldsymbol{a}_\ell)$ by both agents:

$$\begin{aligned}&J_{k,i}(\boldsymbol{a}_k(i),\boldsymbol{a}_\ell(i)|\widetilde{\boldsymbol{w}}_{k,i-1})\\&\quad\triangleq\begin{cases}\text{MSD}_{k,i}(\boldsymbol{a}_\ell(i)=0|\widetilde{\boldsymbol{w}}_{k,i-1}), & \text{if } (0,0)\\ \text{MSD}_{k,i}(\boldsymbol{a}_\ell(i)=1|\widetilde{\boldsymbol{w}}_{k,i-1}), & \text{if } (0,1)\\ \text{MSD}_{k,i}(\boldsymbol{a}_\ell(i)=0|\widetilde{\boldsymbol{w}}_{k,i-1})+c, & \text{if } (1,0)\\ \text{MSD}_{k,i}(\boldsymbol{a}_\ell(i)=1|\widetilde{\boldsymbol{w}}_{k,i-1})+c, & \text{if } (1,1)\end{cases}\\&\quad=\text{MSD}_{k,i}(\boldsymbol{a}_\ell(i)|\widetilde{\boldsymbol{w}}_{k,i-1})+\boldsymbol{a}_k(i)\cdot c\end{aligned} \tag{4}$$

where $\text{MSD}_{k,i}$ denotes the instantaneous mean-square-deviation measure at time i *conditioned* on $\widetilde{\boldsymbol{w}}_{k,i-1}$:

$$\text{MSD}_{k,i}\triangleq\mathbb{E}[\|\widetilde{\boldsymbol{w}}_{k,i}\|^2|\widetilde{\boldsymbol{w}}_{k,i-1}] \tag{5}$$

in terms of the error vector $\widetilde{\boldsymbol{w}}_{k,i}\triangleq w^o-\boldsymbol{w}_{k,i}$. Moreover, the notation $\text{MSD}_{k,i}(\boldsymbol{a}_\ell(i)|\widetilde{\boldsymbol{w}}_{k,i-1})$ is used to represent the $\text{MSD}_{k,i}$ that results from choosing actions $\boldsymbol{a}_k(i)$ and $\boldsymbol{a}_\ell(i)$ under estimation error $\widetilde{\boldsymbol{w}}_{k,i-1}$.

Assume that every agent plays the information-sharing game infinitely often. We consider that each agent k is foresighted and aims to minimize its long-term cost, defined as

$$\begin{aligned}&J_{k,i}^{\infty}[a_k(i)]\\&\triangleq\sum_{t=i}^{\infty}\delta^{t-i}\mathbb{E}[J_{k,t}(\boldsymbol{a}_k(t),\boldsymbol{a}_\ell(t)|\widetilde{\boldsymbol{w}}_{k,t-1})|\widetilde{\boldsymbol{w}}_{k,i-1},\boldsymbol{a}_k(i)=a_k(i)]\end{aligned}$$

where $\delta\in(0,1)$ is a discount factor due to the probability that agents could leave the network in the future, and $J_{k,t}(\boldsymbol{a}_k(t),\boldsymbol{a}_\ell(t)|\widetilde{\boldsymbol{w}}_{k,t-1})$ is defined in (4) for $t\geq i$. The expectation is taken over the random processes $\widetilde{\boldsymbol{w}}_{k,t-1}$, $\boldsymbol{a}_k(t)$ and $\boldsymbol{a}_\ell(t)$, conditioned on $\widetilde{\boldsymbol{w}}_{k,i-1}$ and $\boldsymbol{a}_k(i)$. Then, at each time i, agent k would like to choose $\boldsymbol{a}_k(i)=a_k(i)$ to minimize the long-term discounted cost:

$$\min_{a_k(i)}J_{k,i}^{\infty}[a_k(i)] \tag{6}$$

III. Pareto Inefficiency in One-Shot Information-Sharing Games

We now argue that, if left unattended, the dominant strategy is for the agents not to share estimates. The root cause for such non-cooperative behavior by the agents is due to their lack of belief in the other agents' actions.

A. Dominant Strategy

Minimizing the long-term discounted cost of agent k in (6) requires the prediction of agent ℓ's future actions, which is unavailable. Therefore, agent k can only choose its action to minimize the instantaneous combined cost (4) at every time instant, and thus this behavior is modeled as a one-shot

game. We remark that although the one-shot games are played over time, this situation does not correspond to a repeated game since the instantaneous combined cost function generally changes with time due to the randomness in $\widetilde{\boldsymbol{w}}_{k,t-1}$.

Table I summarizes the instant cost values for both agents under their respective actions. It is straightforward to conclude from the entries in the table that choosing action $\boldsymbol{a}_k(i) = 0$ is the dominant strategy for agent k regardless of the action chosen by agent ℓ because its utility will be the smallest it can be in that situation. Likewise, the dominant strategy for agent ℓ is $\boldsymbol{a}_\ell(i) = 0$ regardless of the action chosen by agent k. Therefore, the action profile $(\boldsymbol{a}_k, \boldsymbol{a}_\ell) = (0, 0)$ is the unique outcome as a dominant strategy equilibrium for the current one-shot game. However, this resulting action profile may be inefficient for the agents. For example, in the sequel we show that if the agents share their weight estimates then, under some conditions, the alternative action profile $(1, 1)$ where both agents cooperate leads to improved utility values for both agents in comparison to the dominant strategy solution. That is, we argue below that $(1, 1)$ can be superior to $(0, 0)$. In a later section, we explain how to introduce a reputation scheme to provide agents with incentives to share information in order to select the solution $(1, 1)$ whenever it is more efficient than the solution $(0, 0)$.

B. Pareto Optimal Action Profile

Agent k cannot evaluate its cost (4) directly since agent k does not know $\boldsymbol{\psi}_{\ell,i}$ before receiving it from agent ℓ. Therefore, $\boldsymbol{\psi}_{\ell,i}$ can be modeled by agent k as an additive perturbation of w^o, say,

$$\boldsymbol{\psi}_{\ell,i} = w^o + \boldsymbol{n}_{\ell,i} \tag{7}$$

Then, steps (2) and (3) lead to the recursion

$$\begin{aligned}\widetilde{\boldsymbol{w}}_{k,i} = \alpha_k(I - \mu\boldsymbol{u}_{k,i}^*\boldsymbol{u}_{k,i})\widetilde{\boldsymbol{w}}_{k,i-1} - \alpha_k\mu\boldsymbol{u}_{k,i}^*\boldsymbol{v}_k(i) \\ - (1-\alpha_k)\boldsymbol{n}_{\ell,i}\end{aligned} \tag{8}$$

Assuming the processes $\{\boldsymbol{u}_{k,i}, \boldsymbol{v}_k(i), \boldsymbol{n}_{\ell,i}\}$ are zero mean and mutually independent, we arrive at

$$\begin{aligned}\text{MSD}_{k,i} &\approx \alpha_k^2\widetilde{\boldsymbol{w}}_{k,i-1}^*(I - 2\mu R_{u,k})\widetilde{\boldsymbol{w}}_{k,i-1} \\ &\quad + (1-\alpha_k)^2\mathbb{E}\|\boldsymbol{n}_{\ell,i}\|^2 \\ &\triangleq \alpha_k^2\boldsymbol{s}_{kk}(i) + (1-\alpha_k)^2 s_\ell\end{aligned} \tag{9}$$

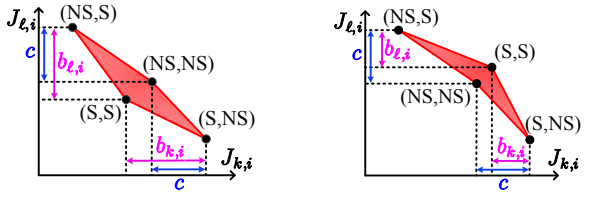
where we ignored terms that depend on μ^2 due to $\mu \ll 1$, and

$$\boldsymbol{s}_{kk}(i) \triangleq \widetilde{\boldsymbol{w}}_{k,i-1}^*(I - 2\mu R_{u,k})\widetilde{\boldsymbol{w}}_{k,i-1} \geq 0 \tag{10}$$

$$s_\ell \triangleq \mathbb{E}\|\boldsymbol{n}_{\ell,i}\|^2 > 0 \tag{11}$$

Expression (9) is valid for both cases in which agent ℓ cooperates or not; this information is controlled through the choice of α_k. To compute (9), agent k still needs to estimate the variance of $\boldsymbol{n}_{\ell,i}$. Using the reference knowledge of $\sigma_{v,\ell}^2$, one useful approximation is to note that if agent ℓ operates independently of the other agents in the network and runs an LMS filter on its own, then the variance of $\boldsymbol{n}_{\ell,i}$ would approach the following value in steady-state [11]:

$$s_\ell \approx \frac{\mu M}{2}\sigma_{v,\ell}^2 \tag{12}$$



(a) $\gamma_{k,i} > 1$ and $\gamma_{\ell,i} > 1$ (b) $\gamma_{k,i} < 1$ and $\gamma_{\ell,i} < 1$

Fig. 1: Illustration of the behavior of the combined costs in terms of the sizes of the benefit-cost ratio parameters ("S" and "NS" refer to the actions "share" and "do not share", respectively).

If agent ℓ decides not to share $\boldsymbol{\psi}_{\ell,i}$, then $\alpha_k = 1$ and expression (9) leads to

$$\text{MSD}_{k,i}(\boldsymbol{a}_\ell(i) = 0|\widetilde{\boldsymbol{w}}_{k,i-1}) = \boldsymbol{s}_{kk}(i) \tag{13}$$

On the other hand, if agent ℓ is willing to share $\boldsymbol{\psi}_{\ell,i}$ with agent k, i.e., $\boldsymbol{a}_\ell(i) = 1$, agent k can predict the resulting $\text{MSD}_{k,i}$ based on the value of α_k. For example, if one assumes that uniform combination weights are used so that $\alpha_k = 1/2$, then

$$\text{MSD}_{k,i}(\boldsymbol{a}_\ell(i) = 1|\widetilde{\boldsymbol{w}}_{k,i-1}) = \frac{\boldsymbol{s}_{kk}(i) + s_\ell}{4} \tag{14}$$

We define the estimation benefit that agent k obtains from a cooperating agent ℓ as the improvement in MSD value and denote it by

$$\begin{aligned}\boldsymbol{b}_{k,i} \triangleq \text{MSD}_{k,i}(\boldsymbol{a}_\ell(i) = 0|\widetilde{\boldsymbol{w}}_{k,i-1}) \\ - \text{MSD}_{k,i}(\boldsymbol{a}_\ell(i) = 1|\widetilde{\boldsymbol{w}}_{k,i-1})\end{aligned} \tag{15}$$

We further introduce the benefit-cost ratio as the ratio of the estimation benefit to the cost of communication:

$$\boldsymbol{\gamma}_{k,i} \triangleq \frac{\boldsymbol{b}_{k,i}}{c} \tag{16}$$

Then, cost values corresponding to the action profile $(1, 1)$ in Table I will be smaller than the cost values corresponding to the action profile $(0, 0)$ if

$$\boldsymbol{\gamma}_{k,i} > 1, \quad \boldsymbol{\gamma}_{\ell,i} > 1 \tag{17}$$

For example, this situation would arise if the estimation errors $\widetilde{\boldsymbol{w}}_{k,i-1}$ and $\widetilde{\boldsymbol{w}}_{\ell,i-1}$ are large enough. In Fig. 1(a), we illustrate schematically how the values of the utility functions would compare to each other when (17) holds for the four possibilities of action profiles. It is seen from this figure that the action profile $(1, 1)$ is Pareto optimal and that the dominant strategy choice $(0, 0)$ is inefficient and leads to worse performance (which is a manifestation of the prisoner's dilemma problem in game theory [12]). On the other hand, if the estimation errors $\widetilde{\boldsymbol{w}}_{k,i-1}$ and $\widetilde{\boldsymbol{w}}_{\ell,i-1}$ are small enough to ensure $\boldsymbol{\gamma}_{k,i} < 1$ and $\boldsymbol{\gamma}_{\ell,i} < 1$, then we are led to Figure 1(b), where the action profile $(0, 0)$ becomes Pareto optimal and Pareto superior to $(1, 1)$. It follows from these observations that it would be advantageous if each selfish agent k could adaptively adjust its actions to choose $\boldsymbol{a}_k(i) = 1$ whenever $\boldsymbol{\gamma}_{k,i}$ is large and $\boldsymbol{a}_k(i) = 0$ whenever $\boldsymbol{\gamma}_{k,i}$ is small.

IV. Adaptive Reputation Design

One way to enable agents to assess the belief of other agents is to associate a reputation score with each agent. Agents that cooperate when it is beneficial are rewarded with higher scores; agents that do not cooperate are penalized with lower scores. It therefore becomes important to enable agents to look

ahead and to assess how their long-term benefit and reputation will depend on their actions.

Here, we assume the agents are bounded rational so that each agent is regarded as an automaton that follows a pre-determined strategy to select actions. Such games are called "machine games" since agents behave like machines [12]. Let us define the reputation parameter $\boldsymbol{\theta}_{k,i}^{\ell} \in (0,1)$ as a scalar score summarizing the history of agent ℓ's actions as viewed by agent k. Similarly, we define $\boldsymbol{\theta}_{\ell,i}^{k} \in (0,1)$. Being an automaton, agent k uses the following recursive strategy to take actions and to update reputations:

$$\boldsymbol{a}_k(i) = f(\widetilde{\boldsymbol{w}}_{k,i-1}, \boldsymbol{\theta}_{k,i}^{\ell}), \quad \boldsymbol{\theta}_{k,i+1}^{\ell} = \tau(\boldsymbol{\theta}_{k,i}^{\ell}, \boldsymbol{a}_{\ell}(i)) \tag{18}$$

where f is the action-choosing policy and τ is the reputation update rule based on the action of agent ℓ. To proceed, we will restrict our attention to threshold-based policies $f(\cdot)$ since they require simple computations. We now motivate one particular choice for $\tau(\cdot)$ and $f(\cdot)$. We begin by designing the reputation update rule and provide an intuitive interpretation in terms of adaptive learning. Then, based on this rule, we analyze the corresponding optimal action-choosing policy.

A. Reputation Update Rule

The reputation scores should be able to reflect the past actions of agents and give more recent actions higher weights. Therefore, we adopt the following form for the reputation update rule:

$$\boxed{\boldsymbol{\theta}_{k,i+1}^{\ell} = r\boldsymbol{\theta}_{k,i}^{\ell} + (1-r)\boldsymbol{a}_{\ell}(i), \quad 0 < r < 1} \tag{19}$$

where r controls the dynamics of the reputation updates. The value of $\boldsymbol{\theta}_{k,i+1}^{\ell}$ can be interpreted by agent k as a measure of its belief in the willingness of agent ℓ to cooperate [8]. Update (19) admits an intuitive interpretation if we assume the past actions are the result of realizations by an independent and stationary Bernoulli random process with probability p_{ℓ}. The probability p_{ℓ} can be estimated from a time window of observations $\boldsymbol{a}_{\ell}(i)$ from $i = 1$ to L as

$$\hat{\boldsymbol{p}}_{\ell,L} = \frac{1}{L}\sum_{i=1}^{L} \boldsymbol{a}_{\ell}(i) \tag{20}$$

or in a recursive form:

$$\hat{\boldsymbol{p}}_{\ell,L+1} = \tfrac{L}{L+1}\hat{\boldsymbol{p}}_{\ell,L} + \Big(1 - \tfrac{L}{L+1}\Big)\boldsymbol{a}_{\ell}(L+1) \tag{21}$$

If we use a sliding window of fixed size $L+1$ to adaptively estimate p_{ℓ}, we can rewrite (21) for arbitrary time instants as:

$$\hat{\boldsymbol{p}}_{\ell,i+1} = r\hat{\boldsymbol{p}}_{\ell,i} + (1-r)\boldsymbol{a}_{\ell}(i) \tag{22}$$

where $r \triangleq L/(L+1)$. Observe that (22) has a form similar to (19).

When agents use the reputation scores of their neighbors to evaluate their willingness to share estimates, we observe that the higher the values of these scores, the more likely it is that the agents will share information. It is therefore justified to assume that the probability of $\boldsymbol{a}_{\ell}(t) = 1$ is proportional to $\boldsymbol{\theta}_{k,t}^{\ell}$ and $\boldsymbol{\theta}_{\ell,t}^{k}$ and we assume that the following approximation is used for agent k to predict agent ℓ's future behavior:

$$\mathbb{P}(\boldsymbol{a}_{\ell}(t) = 1) \approx \boldsymbol{\theta}_{k,t}^{\ell} \cdot \boldsymbol{\theta}_{\ell,t}^{k}, \quad t \geq i \tag{23}$$

It is clear that the reputation scheme encourages agent k to keep its reputation score high to obtain rewarding cooperation from other agents in the future.

B. Action-Choosing Policy

Under the reputation update rule (19), we next need to analyze the optimal action-choosing policy $f(\cdot)$. To begin with, it is clear that the solution of (6) satisfies

$$\boldsymbol{a}_k(i) = \begin{cases} 1, & \text{if } J_{k,i}^{\infty}[a_k(i)=1] < J_{k,i}^{\infty}[a_k(i)=0] \\ 0, & \text{otherwise} \end{cases} \tag{24}$$

Solving the minimization problem (6) therefore requires that we find the condition for which $J_{k,i}^{\infty}[a_k(i)=1] < J_{k,i}^{\infty}[a_k(i)=0]$, which means conditions for the following inequality to hold:

$$J_{k,i}^{\infty}[a_k(i)=1] - J_{k,i}^{\infty}[a_k(i)=0] = \sum_{t=i}^{\infty} \delta^{t-i} \triangle J_{k,t} < 0 \tag{25}$$

where

$$\begin{aligned} \triangle J_{k,t} \triangleq & \mathbb{E}[J_{k,t}(\boldsymbol{a}_k(t), \boldsymbol{a}_{\ell}(t) | \widetilde{\boldsymbol{w}}_{k,t-1}) | \widetilde{\boldsymbol{w}}_{k,i-1}, \boldsymbol{a}_k(i)=1] - \\ & \mathbb{E}[J_{k,t}(\boldsymbol{a}_k(t), \boldsymbol{a}_{\ell}(t) | \widetilde{\boldsymbol{w}}_{k,t-1}) | \widetilde{\boldsymbol{w}}_{k,i-1}, \boldsymbol{a}_k(i)=0] \end{aligned}$$

From (4), we know that $\triangle J_{k,i} = c$. Now, let us consider any time $t > i$. Then,

$$\begin{aligned} \triangle J_{k,t} = & \mathbb{E}[\mathrm{MSD}_{k,t}(\boldsymbol{a}_{\ell}(t) | \widetilde{\boldsymbol{w}}_{k,t-1}) | \widetilde{\boldsymbol{w}}_{k,i-1}, \boldsymbol{a}_k(i)=1] - \\ & \mathbb{E}[\mathrm{MSD}_{k,t}(\boldsymbol{a}_{\ell}(t) | \widetilde{\boldsymbol{w}}_{k,t-1}) | \widetilde{\boldsymbol{w}}_{k,i-1}, \boldsymbol{a}_k(i)=0] + \\ & c \cdot \Big(\mathbb{E}[\boldsymbol{a}_k(t) | \widetilde{\boldsymbol{w}}_{k,i-1}, \boldsymbol{a}_k(i)=1] \\ & - \mathbb{E}[\boldsymbol{a}_k(t) | \widetilde{\boldsymbol{w}}_{k,i-1}, \boldsymbol{a}_k(i)=0] \Big) \end{aligned} \tag{26}$$

Using (15) and (13), we have

$$\begin{aligned} & \mathbb{E}[\mathrm{MSD}_{k,t}(\boldsymbol{a}_{\ell}(t) | \widetilde{\boldsymbol{w}}_{k,t-1}) | \widetilde{\boldsymbol{w}}_{k,i-1}, \boldsymbol{a}_k(i)=j] \\ & \quad = \mathbb{E}[\boldsymbol{s}_{kk}(t) - \boldsymbol{b}_{k,t}\boldsymbol{a}_{\ell}(t) | \widetilde{\boldsymbol{w}}_{k,i-1}, \boldsymbol{a}_k(i)=j] \end{aligned} \tag{27}$$

where $j = 0$ or 1. The future estimation errors $\widetilde{\boldsymbol{w}}_{k,t-1}$ are highly dynamic due to the random matching of agents. One useful approximation[1] is to use the current values:

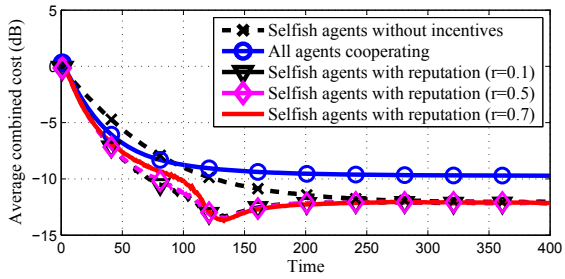
$$\widetilde{\boldsymbol{w}}_{k,t-1} \approx \widetilde{\boldsymbol{w}}_{k,i-1}, \quad \boldsymbol{\theta}_{k,t}^{\ell} \approx \boldsymbol{\theta}_{k,i}^{\ell} \tag{28}$$

Then, we apply (23) to (27) and obtain

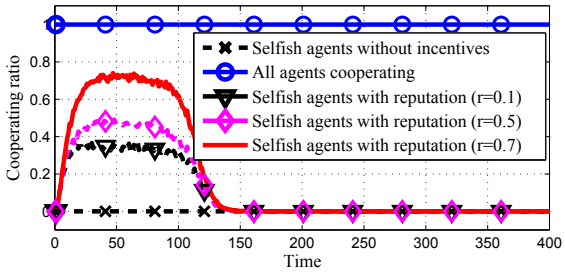
$$\begin{aligned} & \mathbb{E}[\mathrm{MSD}_{k,t}(\boldsymbol{a}_{\ell}(t) | \widetilde{\boldsymbol{w}}_{k,t}) | \widetilde{\boldsymbol{w}}_{k,i}, \boldsymbol{a}_k(i)=j] \\ & \quad \approx \boldsymbol{s}_{kk}(i) - \boldsymbol{b}_{k,i}\boldsymbol{\theta}_{k,i}^{\ell} \cdot \mathbb{E}[\boldsymbol{\theta}_{\ell,t}^{k} | \widetilde{\boldsymbol{w}}_{k,i}, \boldsymbol{a}_k(i)=j] \end{aligned} \tag{29}$$

Therefore, the only parameter varying with time t is the reputation $\boldsymbol{\theta}_{\ell,t}^{k}$. An important observation from the physical interpretation of $\boldsymbol{\theta}_{\ell,t}^{k}$ is that given (28), agent k expects that agent ℓ holding higher $\boldsymbol{\theta}_{\ell,t}^{k}$ will have more probability to share estimates, and then in return, agent k will have more willingness to share estimates back, and vice-versa. Thus, under assumption (28), agent k choosing $\boldsymbol{a}_k(i) = 1$ introduces a positive outcome for the future actions $\boldsymbol{a}_k(t) = 1$ since $\boldsymbol{\theta}_{\ell,t}^{k}$ increases, i.e., agent k expects that choosing $\boldsymbol{a}_k(i) = 1$ will give $\boldsymbol{a}_k(t) = 1$. On the other hand, selecting $\boldsymbol{a}_k(i) = 0$ introduces a negative outcome for agent k to select $\boldsymbol{a}_k(t) = 0$ since $\boldsymbol{\theta}_{\ell,t}^{k}$ decreases, i.e., agent k expects that choosing $\boldsymbol{a}_k(i) = 0$ will give $\boldsymbol{a}_k(t) = 0$. Based on this argument, it is reasonable to

[1] Such stationary approximations are common in the literature of learning game theory [12].



(a) Average combined cost



(b) Cooperating ratio

Fig. 2: Learning curves for various reputation factors r.

predict the future actions $\boldsymbol{a}_k(t) \approx \boldsymbol{a}_k(i)$. Let $\boldsymbol{\theta}_{\ell,t}^{k(1)}$ and $\boldsymbol{\theta}_{\ell,t}^{k(0)}$ denote the future reputations given $\boldsymbol{a}_k(i) = 1$ and $\boldsymbol{a}_k(i) = 0$, respectively. Then,

$$\mathbb{E}[\mathrm{MSD}_{k,t}(\boldsymbol{a}_\ell(t)|\widetilde{\boldsymbol{w}}_{k,t-1})|\widetilde{\boldsymbol{w}}_{k,i}, \boldsymbol{a}_k(i) = j] \approx \boldsymbol{s}_{kk}(i) - \boldsymbol{b}_{k,i}\boldsymbol{\theta}_{k,i}^{\ell}\boldsymbol{\theta}_{\ell,t}^{k(j)} \quad (30)$$

Therefore, (26) can be rewritten as

$$\triangle J_{k,t} \approx -\boldsymbol{b}_{k,i}\boldsymbol{\theta}_{k,i}^{\ell}(\boldsymbol{\theta}_{\ell,t}^{k(1)} - \boldsymbol{\theta}_{\ell,t}^{k(0)}) + c \quad (31)$$

Since $\boldsymbol{a}_k(t) \approx \boldsymbol{a}_k(i)$, we use (19) to write:

$$\boldsymbol{\theta}_{\ell,t}^{k(1)} - \boldsymbol{\theta}_{\ell,t}^{k(0)} = \boldsymbol{\theta}_{\ell,i}^{k} r^{t-i} + (1-r)\sum_{q=0}^{t-i-1} r^q - \boldsymbol{\theta}_{\ell,i}^{k} r^{t-i} = 1 - r^{t-i} \quad (32)$$

Then,

$$\triangle J_{k,t} \approx c - \boldsymbol{b}_{k,i}\boldsymbol{\theta}_{k,i}^{\ell}(1 - r^{t-i}) \quad (33)$$

Using (33) and the fact $\triangle J_{k,i} = c$, agent k chooses action $\boldsymbol{a}_k(i) = 1$ if the benefit-cost ratio $\boldsymbol{\gamma}_{k,i}$ is larger than a threshold:

$$J_{k,i}^{\infty}(a_k(i) = 1) < J_{k,i}^{\infty}(a_k(i) = 0) \Leftrightarrow \boldsymbol{\gamma}_{k,i} > \frac{1-r\delta}{\delta(1-r)\boldsymbol{\theta}_{k,i}^{\ell}} \quad (34)$$

Therefore, the action-choosing policy $f(\cdot)$ becomes

$$\boldsymbol{a}_k(i) = \begin{cases} 1, & \text{if } \boldsymbol{\gamma}_{k,i} > \frac{1-r\delta}{\delta(1-r)\boldsymbol{\theta}_{k,i}^{\ell}} \\ 0, & \text{otherwise} \end{cases} \quad (35)$$

C. Instantaneous Approximations

In (35), the benefit-cost ratio $\boldsymbol{b}_{k,i}$ requires knowledge of $R_{u,k}$ and $\widetilde{\boldsymbol{w}}_{k,i-1}$ for real-time implementations. One common way to instantaneously approximate $R_{u,k}$ is to use $R_{u,k} \approx u_{k,i}^* u_{k,i}^*$ [11]. For $\widetilde{\boldsymbol{w}}_{k,i-1}$, we use the following equation to recursively approximate w^o in a moving average sense:

$$\hat{\boldsymbol{w}}_{k,i}^o = (1-\nu_k)\hat{\boldsymbol{w}}_{k,i-1}^o + \nu_k \boldsymbol{\psi}_{k,i} \quad (36)$$

where $\nu_k < 1$ is a positive forgetting factor, and then $\widetilde{\boldsymbol{w}}_{k,i-1}$ is approximated by $\widetilde{\boldsymbol{w}}_{k,i-1} \approx \hat{\boldsymbol{w}}_{k,i}^o - \boldsymbol{w}_{k,i-1}$.

V. Simulation Results

The network has 10 agents which are randomly paired at each time instant. The length of w^o is $M = 30$ and we randomly choose its entries and normalize them to satisfy $\|w^o\| = 1$. The regressor $\{\boldsymbol{u}_{k,i}\}$ is zero-mean and $R_{u,k}$ is diagonal with entries uniformly generated between $[0, 10]$. The background noise $\boldsymbol{v}_k(i)$ is temporally white and spatially independent Gaussian distributed with zero-mean and $\sigma_{v,k}^2$ uniformly selected between $[-10, 0]$ (dB). We set the step-size to $\mu = 0.005$, the discounted parameter to $\delta = 0.99$, and the transmission cost to $c = 0.1$. All reputation scores are set to 1 at time $i = 0$. Each agent k uses the combination coefficient $\alpha_k = 1/2$ when the shared estimates are available.

In Fig. 2, the average combined costs and cooperating ratios over all agents are simulated. We compare the behavior of selfish agents using our reputation scheme to the behavior of selfish agents without incentives to cooperate and to the behavior of agents that are fully cooperative all the time. In the latter case, the cost of communication adds to the overall cost and continuous sharing of information can therefore degrade performance from this perspective. In Fig. 2(a), the benefit of sharing information is observed in terms of the convergence rate. Our reputation design not only encourages the selfish agents to cooperate, but also enables them to adaptively choose to not cooperate when the benefit of sharing information depreciates (which starts at around time $i = 100$) as shown in Fig. 2(b). We also observe that a larger r facilitates the willingness to cooperate.

References

[1] S. Kar and J. M. F. Moura, "Convergence rate analysis of distributed gossip (linear parameter) estimation: Fundamental limits and tradeoffs," *IEEE J. Sel. Topics in Signal Process.*, vol. 5, no. 5, pp. 674–690, Aug. 2011.

[2] A. Nedic and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Trans. Autom. Control*, vol. 54, no. 1, pp. 48–61, Jan. 2009.

[3] C. G. Lopes and A. H. Sayed, "Diffusion least-mean squares over adaptive networks: Formulation and performance analysis," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3122–3136, Jul. 2008.

[4] F. S. Cattivelli and A. H. Sayed, "Diffusion LMS strategies for distributed estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1035–1048, Mar. 2010.

[5] J. Chen and A. H. Sayed, "Diffusion adaptation strategies for distributed optimization and learning over networks," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 4289–4305, Aug. 2012.

[6] J. Xu and M. van der Schaar, "Social norm design for information exchange systems with limited observations," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 11, pp. 2126–2135, Dec. 2012.

[7] Y. Zhang and M. van der Schaar, "Reputation-based incentive protocols in crowdsourcing applications," in *Proc. IEEE INFOCOM*, Orlando, Florida, USA, Mar. 2012, pp. 2140–2148.

[8] J. Carter, E. Bitting, and A. A. Ghorbani, "Reputation formalization for an information-sharing multi-agent system," in *Computational Intelligence*, vol. 18, no. 4, 2002, pp. 515–534.

[9] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2508–2530, Jun. 2006.

[10] S.-Y. Tu and A. H. Sayed, "Diffusion strategies outperform consensus strategies for distributed estimation over adaptive networks," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6217–6234, Dec. 2012.

[11] A. H. Sayed, *Adaptive Filters*. Wiley, NJ, 2008.

[12] Y. Shoham and K. Leyton-Brown, *Multiagent Systems: Algorithmic, Game Theoretic and Logical Foundations*. Cambridge University Press, 2008.