

ADAPTIVE LEARNING FOR STOCHASTIC GENERALIZED NASH EQUILIBRIUM PROBLEMS

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

Department of Electrical Engineering, University of California, Los Angeles

ABSTRACT

This work examines a stochastic formulation of the generalized Nash equilibrium problem (GNEP) where agents are subject to randomness in the environment of unknown statistical distribution. Three stochastic gradient strategies are developed by relying on a penalty-based approach where the constrained GNEP formulation is replaced by a penalized unconstrained formulation. It is shown that this penalty solution is able to approach the Nash equilibrium in a stable manner within $O(\mu)$, for small step-size values μ . The operation of the algorithms is illustrated by considering the Cournot competition problem.

Index Terms— Adaptive learning, generalized Nash equilibrium, penalized approximation, diffusion learning.

1. INTRODUCTION AND RELATED WORK

The generalized Nash equilibrium problem (GNEP) refers to a setting where each agent in a collection of agents seeks to minimize its own cost function subject to certain constraints *and* where both the cost function and the constraints are generally dependent on the actions selected by the other agents [1–4]. In these types of problems, the Nash equilibrium is a desired and stable solution since at the Nash equilibrium no agent can benefit by unilaterally deviating from the solution. In general, GNEP formulations do not admit closed-form solutions. It is therefore common to appeal to penalty-based formulations where the original cost function is modified by including a penalty term. The purpose of the penalty term is to assign large penalties to deviations from the constraints [5–8].

In most prior works, the individual cost functions are assumed to be *deterministic* and agents are able to compute exactly their gradient vectors [4, 6, 9]. However, when the agents are subject to randomness in the environment, it is customary to define the cost functions in terms of *expectations* of certain loss functions. The expectation operations are in relation to the distribution of the random data. Since this distribution is rarely known beforehand, it becomes impossible for the agents to compute the exact gradients of their cost functions to permit gradient-descent implementations. Instead, the agents need to resort to *stochastic* gradient implementations where the actual gradient vectors are replaced by approximations. One stochastic implementation along these lines is considered in [10] albeit with a *vanishing* step-size parameter and without constraints. The use of step-sizes that decay to zero is problematic in scenarios that require continuous adaptation and learning. For example, in nonstationary environments, the Nash equilibrium may drift with time due to changes in the statistical distribution of the data. When the step-size approaches zero, adaptation stops and the resulting stochastic gradient algorithm will lose its ability to track the drift.

For this reason, we focus in this work on the use of *constant* step-sizes to enable adaptation and learning. When this is done, gradient noise seeps into the operation of the algorithm. By gradient noise we mean the difference between the true gradient vector and its approximation. In decaying step-size implementations, this gradient noise component is annihilated over time by the diminishing step-size parameter at the expense of a deteriorating tracking performance. In contrast, in the constant step-size implementation, the gradient noise process is persistently present in the operation of the algorithm. One main challenge in the analysis, is to establish that the stochastic-gradient implementation is able to keep the influence of gradient noise under check and to deliver an accurate estimation of the minimizer. Arriving at these conclusions for networked agents is the key contribution of this work. In the simulations section, we illustrate the theoretical results and apply the algorithms to the Cournot competition problem, which is widely used in applications such as economic trading with geographical considerations, power management over smart grids, and resource allocation [11–13].

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.

2. PROBLEM FORMULATION

Consider a connected network of N agents indexed by the set $\mathcal{N} = \{1, \dots, N\}$. The neighborhood of each agent k , denoted by \mathcal{N}_k , includes agent k and the neighboring agents connected to k . We denote the action of each agent k by a vector $w_k \in \mathbb{R}^{M_k}$ and associate with k an individual convex cost function denoted by $J_k(\cdot)$. The argument of $J_k(\cdot)$ does not depend solely on w_k but also on the action vectors of the neighboring agents. Let us collect the actions of all agents in \mathcal{N}_k into the block vector $w^k = \text{col}\{w_\ell; \ell \in \mathcal{N}_k\} \in \mathbb{R}^{M^k}$, and the actions of all agents in \mathcal{N} into $w = \text{col}\{w_1, \dots, w_N\} \in \mathbb{R}^M$ where

$$M^k \triangleq \sum_{\ell \in \mathcal{N}_k} M_\ell, \quad M \triangleq \sum_{\ell=1}^N M_\ell \quad (1)$$

We consider that the action of each agent k should satisfy a set of local constraints $\{g_{k,q}(w^k) \leq 0, q = 1, \dots, L_k\}$. The local constraint functions $\{g_{k,q}(w^k)\}$ are known to agent k and assumed to be twice-differentiable and convex in w^k . We further assume that the constraints are shared by the neighbors, i.e., if the argument of $g_{k,q}(w^k)$ contains the action of some neighbor ℓ , then agent ℓ is subject to the same constraint $g_{k,q}(w^k) \leq 0$. Figure 1 illustrates an example showing the dependence of the cost functions and the shared constraint within agents 3 and 5 over a network topology. We note that while there is no direct link between agents 2 and 4, the action vectors w_2 and w_4 are coupled through the intermediate agent 3. Therefore, in general, the actions of agents are affected explicitly

This work was supported in part by NSF grants ECCS-1407712 and CIF-1524250. Emails: ckyuna@ucla.edu, {mihaela,sayed}@ee.ucla.edu

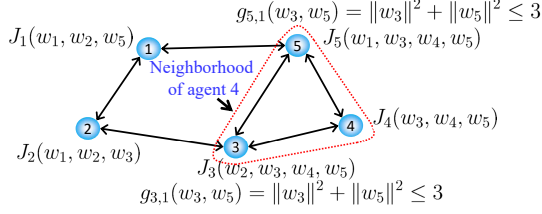


Fig. 1. Illustration of the cost functions and the shared constraints over a network topology.

by the neighbors and also implicitly by other agents in the network. This scenario is common in applications [5, 6, 9, 14].

Each agent k then seeks an optimal action vector that solves the following constrained optimization problem:

$$\begin{cases} \min_{w^k} & J_k(w^k) \\ \text{subject to} & g_{k,q}(w^k) \leq 0, \quad q = 1, \dots, L_k \end{cases} \quad (2)$$

which is known as the generalized Nash equilibrium problem (GNEP).

For convenience, we shall collect all distinct individual constraints into a global set denoted by $\{g_q(w) \leq 0, q = 1, \dots, L\}$ by removing the repeated shared constraints. We assume that the GNEP in (2) is feasible for each agent, meaning that the set $\{w | g_q(w) \leq 0, q = 1, \dots, L\}$ is non-empty. In this work, we focus on proposing distributed learning strategies by which agents can adaptively learn to solve (2) using local observations of the actions of neighboring agents. Although the treatment can be extended to more general cases, we illustrate the main construction and results by focusing in this manuscript on the important case of quadratic cost functions, namely, on functions of the following form:

$$\begin{aligned} J_k(w^k) &= \mathbb{E}[w^{\top} B_k w + b_k^{\top} w + \epsilon_k] \\ &= \mathbb{E} \left[\sum_{s \in \mathcal{N}_k} \sum_{\ell \in \mathcal{N}_k} w_s^{\top} B_{s\ell}^k w_{\ell} + \sum_{\ell \in \mathcal{N}_k} b_{k\ell}^{\top} w_{\ell} + \epsilon_k \right] \\ &= \sum_{s \in \mathcal{N}_k} \sum_{\ell \in \mathcal{N}_k} w_s^{\top} B_{s\ell}^k w_{\ell} + \sum_{\ell \in \mathcal{N}_k} b_{k\ell}^{\top} w_{\ell} + \epsilon_k \end{aligned} \quad (3)$$

where B_k is a random symmetric matrix of size $M^k \times M^k$, b_k is a random vector of size $1 \times M^k$, ϵ_k is a random scalar variable with mean ϵ_k . In (3), it holds that the J_k depends solely on w^k instead of the entire vector w if B_k and b_k are constructed in the following manner. We partitioned B_k and b_k , respectively, into block matrices $\{B_{s\ell}^k \in \mathbb{R}^{M_s \times M_{\ell}}\}$ and block vectors $\{b_{k\ell} \in \mathbb{R}^{M_{\ell} \times 1}\}$ in conformity with the block structure of w^k . We then set $b_{k\ell} = \mathbf{0}_{M_{\ell} \times 1}$ and $B_{s\ell}^k = \mathbf{0}_{M_s \times M_{\ell}}$ if $s \notin \mathcal{N}_k$ or $\ell \notin \mathcal{N}_k$. Furthermore, we denote the means of these quantities by $B_{s\ell}^k = \mathbb{E} B_{s\ell}^k$ and $b_{k\ell} = \mathbb{E} b_{k\ell}$. Note that since B_k is symmetric, we have $B_{s\ell}^k = B_{\ell s}^{k\top}$ and $B_{s\ell}^k = B_{\ell s}^{k\top}$ for any s and ℓ . The expectation in (3) is over the distribution of the random data $\{B_k, b_k, \epsilon_k\}$, which are assumed to be independent.

Now note that the gradient vector of $J_k(w^k)$ with respect to w_k^{\top} is the $M_k \times 1$ vector given by

$$\nabla_{w_k^{\top}} J_k(w^k) = \sum_{\ell \in \mathcal{N}_k} 2B_{k\ell}^k w_{\ell} + b_{kk} \quad (4)$$

If we collect these individual gradient vectors into

$$F(w) \triangleq \text{col}\{\nabla_{w_1^{\top}} J_1(w^1), \dots, \nabla_{w_N^{\top}} J_N(w^N)\} \quad (5)$$

then this block column vector satisfies

$$F(w) = Bw + b \quad (6)$$

where $b = \text{col}\{b_{11}, \dots, b_{NN}\}$ and

$$B \triangleq \begin{bmatrix} 2B_{11}^1 & \dots & 2B_{1N}^1 \\ \vdots & \ddots & \vdots \\ 2B_{N1}^N & \dots & 2B_{NN}^N \end{bmatrix} \in \mathbb{R}^{M \times M} \quad (7)$$

To continue, we introduce the following condition.

Condition 1. *There exists a positive constant ν such that $B \geq \nu I$, i.e., for any $M \times 1$ vector a we have*

$$a^{\top} (B - \nu I) a \geq 0 \quad \Leftrightarrow \quad a^{\top} B a \geq \nu \|a\|^2 \quad (8)$$

□

Note that since B is not necessarily symmetric, we know from [15, pp. 259] that Condition 1 holds if, and only if, the symmetric part of B satisfies:

$$\frac{1}{2}(B + B^{\top}) \geq \nu I \quad (9)$$

It follows from this condition that the largest singular value of B , denoted by σ_{\max} , is greater than or equal to ν since

$$\sigma_{\max} = \|B\| \geq \left\| \frac{1}{2}(B + B^{\top}) \right\| \geq \nu \quad (10)$$

Under Condition 1, it is easy to verify that for any two action profiles $w = w^{\circ}$ and $w = w^{\bullet}$ we have

$$(w^{\circ} - w^{\bullet})^{\top} [F(w^{\circ}) - F(w^{\bullet})] \geq \nu \|w^{\circ} - w^{\bullet}\|^2 \quad (11)$$

which means that the mapping $F(w) : \mathbb{R}^M \rightarrow \mathbb{R}^M$ is strongly monotone [16].

3. STOCHASTIC PENALTY-BASED LEARNING

3.1. Penalty Approximation for Coupled Constraints

Solving the constrained optimization problem (2) is generally demanding and may not admit a closed-form solution. Alternatively, in this work, we resort to a penalty-based approach to replace the original problem by an unconstrained optimization problem and then show that the solution to the approximate problem tends asymptotically with the penalty parameter to the desired solution to (2). More importantly, we will show that the penalty-based approach enables the agents to employ *adaptive* learning strategies, which in turn endow the agents with the ability to track variations in the location of the Nash equilibrium due to changes that may occur in the constraint conditions or cost measures.

The main motivation for penalty methods is to assign a large penalty weight whenever constraints are violated and a smaller or zero weight when the constraints are satisfied [7, 17, 18]. More specifically, problem (2) is replaced by the following unconstrained formulation:

$$\min_{w_k} J_k(w^k) + \rho p_k(w^k) \quad (12)$$

where $\rho \geq 0$ is a penalty parameter, $p_k(w^k)$ denotes the penalty function for agent k and is assumed to be of the following aggregate form, with one penalty factor applied to each constraint:

$$p_k(w^k) = \sum_{q=1}^{L_k} \theta(g_{k,q}(w^k)) \quad (13)$$

where $\theta(x)$ is a convex function. The penalty factor $\theta(\cdot)$ returns zero value if the constraint is satisfied, i.e., when $g_{k,q}(w^k) \leq 0$, and introduces a large positive penalty if the constraint is violated, i.e., when $g_{k,q}(w^k) > 0$. While there exist numerous penalty-type functions in the literature, e.g., γ -norm [19], exponential and shifted logarithmic functions [20], in this work we assume that $\theta(x)$ is chosen to be continuous, convex, nondecreasing, and twice-differentiable — see [7, 21] and (43) for an example.

We denote the penalized cost by

$$J_k^p(w^k) \triangleq J_k(w^k) + \rho p_k(w^k) \quad (14)$$

An action profile w^* that minimizes simultaneously all penalized costs $\{J_k^p(w^k)\}$ is called a Nash equilibrium for the penalized formulation (12). By the results in [7, pp. 3930], we get that the distance between the solutions to problem (2) and problem (12) can be made arbitrarily small by choosing ρ appropriately. Consequently, the resulting Nash equilibrium for the unconstrained problem in (12) can be arbitrarily close to a Nash equilibrium for the GNEP in (2). Since the $\{J_k^p(w^k)\}$ are convex functions, it follows from results in [9, pp. 522] that a Nash equilibrium exists. The following theorem ensures its uniqueness.

Theorem 1. (Uniqueness): *Under Condition 1 and for any convex choice of $\theta(x)$, problem (12) has a unique Nash equilibrium w^* , and it satisfies*

$$F^p(w^*) \triangleq Bw^* + b + \rho \nabla_{w^T} p(w^*) = \mathbf{0} \quad (15)$$

where

$$F^p(w) \triangleq \text{col}\{\nabla_{w^T} J_1^p(w^1), \dots, \nabla_{w^T} J_N^p(w^N)\} \quad (16)$$

$$p(w) \triangleq \sum_{q=1}^L \theta(g_q(w)) \quad (17)$$

□

We note that it holds that

$$\nabla_{w^T} p(w) = \text{col}\{\nabla_{w^T} p_1(w^1), \dots, \nabla_{w^T} p_N(w^N)\} \quad (18)$$

3.2. Stochastic Learning Dynamics

Again, a closed form solution to problem (12) is not generally possible. If this were possible, then agents could learn w^* given knowledge of the other agents' actions; this solution method would lead to the best response dynamics [22]. Since this approach is rarely applicable, agents can instead appeal to learning strategies where they gradually approach the desired w^* through successive estimation from streaming data. For example, one well-known solution is to use the gradient information $\nabla_{w^T} J_k^p(w^k)$ to update the agents' actions at discrete-time instants i [23–25]. However, knowledge of the mean quantities $\{B_{k\ell}^k, b_{kk}\}$ is required to compute $\nabla_{w^T} J_k(w^k)$.

When the statistics of $B_{k\ell}^k$ and b_{kk} are unavailable, we need to resort to instantaneous realizations of these random variables, which we denote by $B_{k\ell,i}^k$ and $b_{kk,i}$ at iteration i . We assume these realizations are independent over both k, ℓ , and i . Using the realizations $\{B_{k\ell,i}^k, b_{kk,i}\}$, agents can update their actions at each time instant i by employing the following localized stochastic-gradient rule:

$$\begin{aligned} w_{k,i} &= w_{k,i-1} - \mu \left(\sum_{\ell \in \mathcal{N}_k} 2B_{k\ell,i}^k w_{\ell,i-1} + b_{kk,i} \right) \\ &\quad - \mu \rho \nabla_{w^T} p_k(w_{k,i-1}^k) \end{aligned} \quad (19)$$

where μ is the step-size. Alternatively, motivated by the arguments from [7] and [26], one can implement (19) incrementally by using a two-step learning strategy to improve the individual and penalty

costs separately. For example, agent k can use an Adapt-then-Penalize (ATP) diffusion learning strategy to update first the iterate along the negative gradient direction of $J_k(\cdot)$ and then correct along the gradient of the penalty term:

$$\text{(ATP)} \begin{cases} \psi_{k,i} = w_{k,i-1} - \mu \left(\sum_{\ell \in \mathcal{N}_k} 2B_{k\ell,i}^k w_{\ell,i-1} + b_{kk,i} \right) \\ w_{k,i} = \psi_{k,i} - \mu \rho \nabla_{w^T} p_k(\psi_{k,i}^k) \end{cases} \quad (20)$$

$$(21)$$

Agents can also switch the order of these two steps and use a Penalize-then-Adapt (PTA) diffusion learning strategy:

$$\text{(PTA)} \begin{cases} \psi_{k,i} = w_{k,i-1} - \mu \rho \nabla_{w^T} p_k(w_{k,i-1}^k) \\ w_{k,i} = \psi_{k,i} - \mu \left(\sum_{\ell \in \mathcal{N}_k} 2B_{k\ell,i}^k \psi_{\ell,i} + b_{kk,i} \right) \end{cases} \quad (22)$$

$$(23)$$

Observe that we are denoting the weight iterates in boldface since they are random quantities due to the randomness in the realizations $\{B_{k\ell,i}^k, b_{kk,i}\}$.

4. PERFORMANCE ANALYSIS

We now examine the convergence and stability properties of these stochastic algorithms. In particular, we examine how close their limiting point gets to the equilibrium point, w^* . To continue, we introduce the following condition on the penalty function. This condition is not restrictive since the choice of the penalty function is under the designer's control.

Condition 2. *Consider two arbitrary block vectors w° and w^\bullet collecting all agents' actions. For each k , and the corresponding w_{\circ}^k and w_{\bullet}^k , the gradient vector $\nabla_{w^T} p_k(\cdot)$ is assumed to satisfy the Lipschitz condition:*

$$\left\| \nabla_{w^T} p_k(w_{\circ}^k) - \nabla_{w^T} p_k(w_{\bullet}^k) \right\| \leq \gamma_k \|w_{\circ}^k - w_{\bullet}^k\| \quad (24)$$

where γ_k is a positive constant. □

4.1. Stochastic Gradient Dynamics

We can describe the evolution of the dynamics of the first algorithm (19) in the following manner:

$$w_i = w_{i-1} - \mu(B_i w_{i-1} + b_i) - \mu \rho \nabla_{w^T} p(w_{i-1}) \quad (25)$$

in terms of the aggregate quantities $w_i \triangleq \text{col}\{w_{1,i}, \dots, w_{N,i}\}$, $b_i \triangleq \text{col}\{b_{1,i}, \dots, b_{NN,i}\}$, and

$$B_i \triangleq \begin{bmatrix} 2B_{11,i}^1 & \cdots & 2B_{1N,i}^1 \\ \vdots & \ddots & \vdots \\ 2B_{N1,i}^N & \cdots & 2B_{NN,i}^N \end{bmatrix} \quad (26)$$

Subtracting w^* from both sides of (25), introducing the error vector $\tilde{w}_i = w^* - w_i$ and using (15) we find that

$$\tilde{w}_i = \tilde{w}_{i-1} - \mu[F^p(w^*) - F^p(w_{i-1})] + \mu s_i(w_{i-1}) \quad (27)$$

where $s_i(w) = -\tilde{B}_i w - \tilde{b}_i$ is the gradient noise, $\tilde{B}_i = B - B_i$ and $\tilde{b}_i = b - b_i$. From the assumed independence of B_i , b_i , and w_{i-1} , it can be verified that

$$\mathbb{E} s_i(w_{i-1}) = \mathbf{0} \quad (28)$$

$$\mathbb{E} \|s_i(w_{i-1})\|^2 \leq \alpha \mathbb{E} \|\tilde{w}_{i-1}\|^2 + \beta \quad (29)$$

for some constants α and β . The proof of the following result is omitted for brevity.

Theorem 2. (Mean-square-error stability) For the stochastic gradient implementation (19), if the step-size μ satisfies

$$0 < \mu < \frac{2\nu}{(\sigma_{\max} + \rho\delta_p)^2 + \alpha} \quad (30)$$

where $\alpha > 0$ is a constant parameter, $\delta_p \triangleq \sqrt{\sum_{k=1}^N \gamma_k^2}$, and σ_{\max} is the largest singular value of B , then it holds that

$$\lim_{i \rightarrow \infty} \sup \mathbb{E} \|\tilde{w}_i\|^2 = O(\mu) \quad (31)$$

4.2. Stochastic ATP and PTA Strategies

We can aggregate the ATP and PTA strategies in (20)–(21) and (22)–(23), respectively, across all agents into the following unified description:

$$w_i = \psi_{i-1} - c_1 \mu \rho \nabla_{w_i} p(\psi_{i-1}) \quad (32)$$

$$\phi_i = w_i - \mu(B_i w_i + b_i) \quad (33)$$

$$\psi_i = \phi_i - c_2 \mu \rho \nabla_{w_i} p(\phi_i) \quad (34)$$

for some constants (c_1, c_2) . By setting $(c_1, c_2) = (0, 1)$ we get the ATP recursions (20)–(21) while for $(c_1, c_2) = (1, 0)$ we obtain the PTA recursions (22)–(23). The following result establishes that without gradient noise, the deterministic version of recursions (32)–(34) converges to a unique fixed point.

Theorem 3. (Unique fixed point) Without gradient noise, the mapping from ψ_{i-1} to ψ_i in recursions (32)–(34) converges to a unique fixed point, denoted by ψ^∞ , for small step-sizes that satisfy:

$$0 < \mu < \frac{2\nu}{\sigma_{\max}^2 + \rho^2 \delta_p^2} \quad (35)$$

Let us denote by w^∞ the fixed point corresponding to the deterministic variable w_i , and introduce the fixed-point error $\tilde{w}_i^\infty = w^\infty - w_i$ for the stochastic variable w_i with gradient noise. The following theorem shows that the variance of this error is bounded.

Theorem 4. (Bounded MSE) For the stochastic recursion (32)–(34), if the step-size μ satisfies

$$0 < \mu < \frac{2\nu}{\sigma_{\max}^2 + \rho^2 \delta_p^2 + \alpha} \quad (36)$$

then it holds that

$$\lim_{i \rightarrow \infty} \sup \mathbb{E} \|\tilde{w}_i^\infty\|^2 = O(\mu) \quad (37)$$

We note that condition (36) for the stochastic recursion implies condition (35) for the deterministic case. Therefore, any step-size μ satisfying (36) ensures the existence of the fixed point w^∞ . Furthermore, comparing (36) with (30) we observe that the stochastic ATP and PTA learning strategies allow to use a larger step-size, which can assist with faster convergence performance. However, the fixed point w^∞ would be different from the Nash equilibrium, w^* . In the following theorem, we examine the bias $\tilde{w} = w^* - w^\infty$. We show that for small μ , the norm of the bias is asymptotically upper bounded by $O(\mu)$.

Theorem 5. (Small bias) For sufficiently small step-size μ satisfying the following condition:

$$0 < \mu < \frac{2\nu}{\sigma_{\max}^2 + 2\rho^2 \delta_p^2} \quad (38)$$

it holds that

$$\lim_{i \rightarrow \infty} \sup \mathbb{E} \|w^* - w_i\|^2 \leq O(\mu) + O(\mu^2 \rho^2) \quad (39)$$

5. SIMULATION RESULTS

In this section, we consider the Cournot competition problem [11, 12] with stochastic parameters. Thus, consider a network with $N = 10$ factories and $L = 3$ markets connected to the factories as shown in Fig. 2. Each factory k needs to determine a continuous quantity of products to be produced and delivered to each connected market, which is defined as the action of factory k denoted by $w_k = [w_k(1), \dots, w_k(M_k)]^T$ where we assumed M_k markets are connected to factory k . The individual cost function of each factory k is defined as

$$J_k(w^k) = \mathbb{E} \left(C_{k,i}(w_k) - \sum_{u \sqsubset \ell} w_k(u) \cdot P_{\ell,i}(r_\ell) \right) \quad (40)$$

where

$$C_k(w_k) = (\bar{\alpha}_k + v_k^\alpha) \left(\sum_{n=1}^{M_k} w_k(n) \right)^2 \quad (41)$$

is a random quadratic production cost function to generate $\sum_{n=1}^{M_k} w_k(n)$ amount of products for $\bar{\alpha}_k > 0$ and with some random disturbance $v_{k,i}^\alpha$, and

$$P_{\ell,i}(r_\ell) = \bar{q}_\ell - (\bar{s}_\ell + v_{\ell,i}^s) r_\ell \quad (42)$$

is the price of products in market ℓ and r_ℓ is the total amount of products sold in market ℓ , e.g., $r_2 = w_3(2) + w_4(1) + w_5(1) + w_6(2)$ and so on. In (40), we use the notation $u \sqsubset \ell$ to represent that $w_k(u)$ is the quantity of factory k sold in market ℓ . We set $\bar{\alpha}_k = 1$, $\bar{q}_\ell = 3$, and $\bar{s}_\ell = 1$ for all k and ℓ . The random variables $v_{k,i}^\alpha$ and $v_{\ell,i}^s$ for all k, ℓ and i are assumed to be independent and follow normal distributions with zero mean and standard deviation 10. Each market ℓ has an upper bound of 0.5 for the capacity of products, i.e., $r(\ell) \leq 0.5$. The penalty function $\theta(x)$ is chosen as [7]:

$$\theta(x) \triangleq \max \left\{ 0, \frac{x^3}{\sqrt{x^2 + \eta^2}} \right\} \quad (43)$$

with $\eta = 0.015$. We set the penalty parameter $\rho = 1000$ and the step-size $\mu = 0.001$. In Fig. 3, we simulate the learning curves for the stochastic gradient, ATP, and PTA strategies. To be specific, we compare the mean-squared error $\mathbb{E} \|w^* - w_i\|^2$ in these three cases. We observe from the figure that while the three strategies exhibit approximately the same steady-state performance, the stochastic ATP and PTA strategies have better convergence performance in the transient phase.

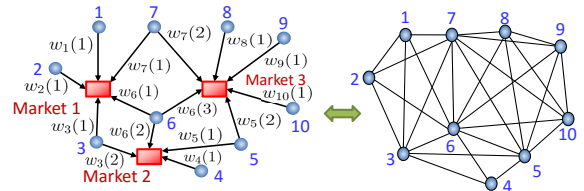


Fig. 2. Network topology used for the simulations of networked Cournot competition.

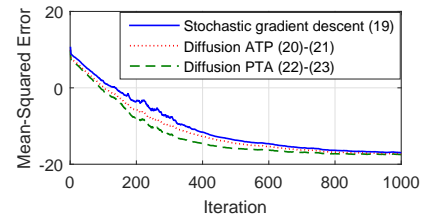


Fig. 3. Learning curves for the stochastic gradient dynamic, diffusion ATP, and diffusion PTA.

6. REFERENCES

- [1] P. T. Harker, "Generalized Nash games and quasi-variational inequalities," *European Journal of Operational Research*, vol. 54, no. 1, pp. 81–94, Sep. 1991.
- [2] J. Contreras, M. Klusch, and J. B. Krawczyk, "Numerical solutions to Nash-Cournot equilibria in coupled constraint electricity markets," *IEEE Trans. Power Syst.*, vol. 19, no. 1, pp. 195–206, Feb. 2004.
- [3] J.-S. Pang and M. Fukushima, "Quasivariational inequalities, generalized Nash equilibria, and multi-leader-follower games," *Comput. Manag. Sci.*, vol. 2, no. 1, pp. 21–56, Jan. 2005.
- [4] F. Facchinei and J.-S. Pang, "Exact penalty functions for generalized Nash problems," *Large-Scale Nonlin. Optimiz.*, vol. 83, pp. 115–126, 2006.
- [5] M. Fukushima, "Restricted generalized nash equilibria and controlled penalty algorithm," *Computat. Manage. Sci.*, vol. 8, no. 3, pp. 201–218, Aug. 2011.
- [6] F. Facchinei and C. Kanzow, "Generalized Nash equilibrium problems," *4OR, Quart. J. Oper. Res.*, vol. 5, no. 3, pp. 173–210, Sep. 2007.
- [7] Z. J. Towfic and A. H. Sayed, "Adaptive penalty-based distributed stochastic convex optimization," *IEEE Trans. Signal Process.*, vol. 62, no. 15, pp. 3924–3938, Aug. 2014.
- [8] M. C. Pinar and S. Zenios, "On smoothing exact penalty functions for convex constrained optimization," *SIAM J. Optim.*, vol. 4, no. 3, pp. 486–511, 1994.
- [9] J. B. Rosen, "Existence and uniqueness of equilibrium points for concave n-person games," *Econometrica*, vol. 33, no. 3, pp. 520–534, 1965.
- [10] J. Koshal, A. Nedic, and U. V. Shanbhag, "Regularized iterative stochastic approximation methods for stochastic variational inequality problems," *IEEE Trans. Autom. Control*, vol. 58, no. 3, pp. 594–609, Mar. 2013.
- [11] B. F. Hobbs and J. S. Pang, "Nash-cournot equilibria in electric power markets with piecewise linear demand functions and joint constraints," *Oper. Res.*, vol. 55, no. 1, pp. 113–127, 2007.
- [12] K. Bimpikis, S. Ehsani, and R. Ilklic, "Cournot competition in networked markets," in *Proc. of the 15th ACM Conf. on Economics and Computation*, Stanford, CA, USA, Jun. 2014, pp. 733–733.
- [13] M. Parzy and H. Bogucka, "QoS support in radio resource sharing with Cournot competition," in *2nd International Workshop on Cognitive Information Processing (CIP)*, Elba, Italy, Jun. 2010, pp. 93–98.
- [14] G. Scutari, D. Palomar, F. Facchinei, and J.-S. Pang, "Convex optimization, game theory, and variational inequality theory," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 35–49, May 2010.
- [15] C. R. Johnson, "Positive definite matrices," *Amer. Math. Monthly*, vol. 77, no. 3, pp. 259–264, Mar. 1970.
- [16] F. Facchinei and J.-S. Pang, *Finite Dimensional Variational Inequalities and Complementarity Problems*. Springer-Verlag, NY, Inc., 2003.
- [17] B. T. Polyak, *Introduction to Optimization*. New York, NY, USA: Optimization Software, 1987.
- [18] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming: Theory and Algorithms*. Wiley, NY, 1993.
- [19] F. Facchinei and C. Kanzow, "Penalty methods for the solution of generalized nash equilibrium problems," *SIAM J. Optim.*, vol. 20, no. 5, pp. 2228–2253, 2010.
- [20] A. Ben-Tal and M. Zibulevsky, "Penalty/barrier multiplier methods for convex programming problems," *SIAM J. Optim.*, vol. 7, no. 2, pp. 347–366, 1997.
- [21] A. H. Sayed, "Adaptive networks," *Proceedings of the IEEE*, vol. 102, no. 4, pp. 460–497, Apr. 2014.
- [22] A. Matsui, "Best response dynamics and socially stable strategies," *Journal of Economic Theory*, vol. 57, no. 2, pp. 343–362, Aug. 1992.
- [23] S. D. Flam, "Equilibrium, evolutionary stability and gradient dynamics," *Int. Game Theory Rev.*, vol. 4, no. 4, pp. 357–370, 2002.
- [24] J. Shamma and G. Arslan, "Dynamic fictitious play, dynamic gradient play, distributed convergence to Nash equilibria," *IEEE Trans. Autom. Control*, vol. 50, no. 3, pp. 312–327, Mar. 2005.
- [25] N. Li and J. Marden, "Designing games for distributed optimization," *IEEE J. Sel. Topics in Signal Process.*, vol. 7, no. 2, pp. 230–242, Apr. 2013.
- [26] A. H. Sayed, *Adaptation, Learning, and Optimization over Nnetworks*. Foundations and Trends in Machine Learning, vol. 7, no. 4-5, pp. 311–801, NOW Publishers, Jul. 2014.