MASAGE: Model-Agnostic Sequential and Adaptive Game Estimation*

Yunian Pan[1 [0000−0002−7277−3657]], Guanze Peng[1 [0000−0003−4474−1129]], Juntao Chen[2 [0000−0001−7726−4926]], and Quanyan Zhu[1 [0000−0002−0008−2953]]

[1] New York University, Brooklyn, NY, USA; {yp1170,gp1363,qz494}@nyu.edu
[2] Fordham University, The Bronx, NY, USA; jchen504@forham.edu

Abstract. Zero-sum games have been used to model cybersecurity scenarios between an attacker and a defender. However, unknown and uncertain environments have made it difficult to rely on a prescribed zero-sum game to capture the interactions between the players. In this work, we aim to estimate and recover an unknown matrix game that encodes the uncertainties of nature and opponent based on the knowledge of historical games and the current observations of game outcomes. The proposed approach effectively transfers the past experiences that are encoded as expert games to estimate and inform future game plays. We formulate the game knowledge transfer and estimation problem as a sequential least-square problem. We characterize the structural properties of the problem and show that the non-convex problem has well-behaved gradient and Hessian under mild assumptions. We propose gradient-based methods to enable dynamic and adaptive estimation of the unknown game. A case study is used to corroborate the results and illustrate the behavior of the proposed algorithm.

Keywords: Zero-sum Games · Security Games · Neural Networks · Least-Square Estimation · Sensitivity Analysis · Gradient-based Methods

1 Introduction

In many adversarial scenarios, such as a battlefield and cyber threats, a defender plays against unknown opponents in uncertain environments. The prior knowledge or experience of the game may provide the defender a way to estimate the game by leveraging his past experience with the environment, or transfering other experiences of his own or from someone else. These experiences are encoded or represented by games that capture critical characteristics of an adversarial entity, including the incentives, the capabilities, and the information structures. The direct estimation of the game provides the defender a sufficient situational awareness of the unknown environment and enables dependable reasoning for making decisions.

Dealing with uncertainties in games has a long history. Harsanyi in 1967 [9] introduced Bayesian games and the notion of "type", encapsulating all uncertainties in

* This research is partially supported by awards ECCS-1847056, CNS-1544782, CNS-2027884, and SES-1541164 from National Science of Foundation (NSF), and grant W911NF-19-1-0041 from Army Research Office (ARO).

payoffs, actions, and psychological attributes of a player into the "type" space to overcome the technical difficulty created by the reasoning using infinite hierarchies of beliefs [15]. Built on Harsanyi's Bayesian game framework, many recent efforts have been on identifying and estimating structures of the game model, given the data of multiple equilibria [12,23] or the observed frequency of choices [10,19].

The estimation of games within Bayesian frameworks often requires the structural knowledge of baseline game models. However, in many security applications, this knowledge may not be directly available. It is difficult, if not impossible, to specify the set of uncertain parameters and the unknowns in security games, since mapping out the structural unknowns can be a challenging task, let alone the unknown unknowns. Hence, there is a need to shift the paradigm from a Bayesian-based approach to a completely data-driven and model-agnostic one. To this end, this work presents an estimation framework that is purely based on the past experiences and the real-time observations. We focus on the estimation of finite zero-sum static games, which are central to security applications, such as in network configurations [24], network provisioning [20], and jamming attacks [25].

We formulate MASAGE, a sequential least-square estimation problem over the game space, which is formed by the past transferable experiences. This approach dispenses with the knowledge of parametric uncertainties and the payoff structure of the game but takes the game as an object for estimation instead. In this work, we focus on a class of linear game estimators. Under mild assumptions, the static least-square game estimation problem is probably solvable by gradient-based algorithms. We extend the static framework to its sequential counterpart, in which the security game is estimated dynamically based on sequential observations. We characterize the structural properties of the estimation problem and show the convergence properties of the gradient-based data-driven adaptive algorithm.

2 Related Work

Game identification and estimation [18,19,2,10,12] have been investigated in economics literature. Hotz et al. in [10] have first considered a conditional choice probability estimator of the structural parameters in dynamic programming models. Following this work, [18,19] have proposed an identification and estimation framework based on time-series data using observed choices. They have considered a class of asymptotic least-square estimators defined by the equilibrium conditions. For discrete games and normal-form games, Bajari et al. in [1] and [2] have proposed simulation-based estimators for parametric games using algorithms that compute all the game equilibria. With a focus on the multiplicity of equilibira, Jovanovic in [12] has highlighted that the information of multiple solutions affects the statistical inference strategy. These works share a common structure that uses equilibria data from firms or companies to estimate the structural parameters of static or dynamic models. Our work studies this problem from a model-agnostic perspective by formulating the estimation directly on the game space. This work focuses on the class of zero-sum matrix games, which plays an important role in cybersecurity.

The analysis of the least-square game estimation problem relies on the perturbation theory of matrix games. Two closely related works are [8] and [6]. Gross in [8] has considered a general case of real matrices and computed the left and right value derivative with respect to arbitrary matrix entries. The author has observed that when the matrix has only one Nash equilibrium pair, the derivative exists, and the right and left derivatives are equal. Cohen et al. in [6] and [7] have studied the completely mixed matrix games and bi-matrix games, and have given the value derivatives with respect to the matrix entries. The authors have provided useful results of strategy derivative and higher-order derivatives of saddle-point values.

3 Problem Formulation

3.1 Preliminary

Game Description Consider a two-player zero-sum finite game G represented by a triplet $\langle \mathscr{N}, \{\mathscr{A}_1, \mathscr{A}_2\}, \{u_1, u_2\}\rangle$. Here, $\mathscr{N} = \{P_1, P_2\}$ is the player set containing a defender P_1 and an attacker P_2; $\mathscr{A}_1 = \{1, 2, \ldots, N_1\}$ and $\mathscr{A}_2 = \{1, 2, \ldots, N_1\}$ are action sets for P_1 and P_2, respectively, with $N_1 = |\mathscr{A}_1|$ and $N_2 = |\mathscr{A}_2|$; $u_1 : \mathscr{A}_1 \times \mathscr{A}_2 \to \mathbb{R}$ and $u_2 : \mathscr{A}_1 \times \mathscr{A}_2 \to \mathbb{R}$ are the utility functions of P_1 and P_2, respectively. Since the game is zero-sum, $u_1 + u_2 = 0$. The zero-sum game can be fully characterized by a single matrix of the size $N_1 \times N_2$. P_1 is the row player. P_2 is the column player. Each row and column is indexed by the corresponding actions of the player. Each entry of the matrix is associated with a payoff value that is viewed as cost to P_1 but utility to P_2.

We consider the scenario where the payoffs of the games are uncertain. To capture the uncertainties, we define a random matrix $\mathbf{M} : \Omega \to \mathbb{R}^{N_1 \times N_2}$ over an underlying probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Each entry of matrix $\mathbf{M}$ is a random variable defined on the probability space. The underlying distributions of the random variables are unknown to the players. Let $\mathrm{val}(\cdot)$ be the saddle-point value of a matrix game. Random matrix game $\mathbf{M}$ gives rise to its associated game value $\mathbf{z} = \mathrm{val}(\mathbf{M})$.

Expert Games and Game Estimation We consider the following scenario. The players do not know their game prior to the play. However, they are given a set of expert games that they have played before and know that their game will be similar and related to the set of expert games. The game is determined by nature, i.e., $\omega \in \Omega$ is realized when the game starts. Let $\bar{M} \in \mathbb{R}^{N_1 \times N_2}$ denote this game. The players cannot observe ω but can observe the outcome of the play of the game, i.e., the value of the sampled game $\bar{M}$, denoted by $\bar{z}$. $\bar{M}$ is also called the target game as the goal of the sequential play of the game is to estimate its value based on the prior information of the expert games and the sequential observations of $\bar{z}$. The formulation of this problem will be made clear later in Subsection 3.2.

A Nonlinear Least-Square Estimator To provide a formal framework of the estimation problem, we first consider the following non-sequential estimation problem. At the start of the game, the defender has a set of S expert games $\mathscr{M} = \{M_1, \ldots, M_S\}$ that is non-random and observable, where $S \in \mathbb{N}$ is the number of expert games; Let $\mathscr{S} := \{1, \ldots, S\}$,

$M_i, i \in \mathscr{S}$ are informed to the player from past interactions or experiences that satisfy following properties:

(i) All expert games have nonzero saddle-point values, i.e., for all $i \in \mathscr{S}$,

$$\text{val}(M_i) \neq 0; \tag{1}$$

(ii) each pair of expert games are not strategically equivalent, i.e., for all $i, j \in \mathscr{S}$,

$$\forall c \in \mathbb{R}, \qquad M_i \neq cM_j; \tag{2}$$

(iii) entries of expert games are bounded, i.e., for all $i \in \mathscr{S},\ a \in \mathscr{A}_1,\ b \in \mathscr{A}_2$,

$$\exists B \in \mathbb{R}, \qquad (M_i)_{ab} \leq B. \tag{3}$$

The defender can observe the value of the game of the unknown game $\bar{M}$, $\bar{z}$ before the play of the game. The information that is available to the defender is $I = \{\mathscr{M}, \bar{z}\}$. The goal of the defender is to find an estimator $\mu : \mathscr{I} \to \mathbb{R}^{N_1 \times N_2}$ that maps the information set of the defender to find an estimate $\hat{M} = \mu(I)$. Here, $\mathscr{I}$ denotes the set of all possible information to the defender.

We consider a class of linear estimators $L(\mathscr{M}; \alpha)$ that are parameterized by a weight vector $\alpha \in \mathscr{X}$, where $\mathscr{X} \subseteq \mathbb{R}^S$ is the parameter space, $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_S]^{\mathrm{T}}$. The estimators take the following form:

$$\hat{M} = L(\mathscr{M}; \alpha) = \sum_{i=1}^{S} \alpha_i M_i \tag{4}$$

From (4), we can see that the linear estimator is taken as the linear combination of expert games. A natural criterion of an optimal estimator is the one that minimizes the error between the outcomes of the estimated game and the target game. The outcome of the estimated game is given by $\text{val}(L(\mathscr{M}; \alpha))$, while the outcome of the target game is assumed to be observable by the defender, which takes the value of $\bar{z}$. Hence, the residue error of the estimation is

$$\varepsilon = \bar{z} - \text{val}(L(\mathscr{M}; \alpha)) \tag{5}$$

An optimal linear estimator $\mu^* = L(\mathscr{M}; \alpha^*)$ with the optimal parameters α^* is the one that minimizes the residue error (5) using the following squared error criterion $J(\alpha)$:

$$J(\alpha) := |\text{val}(L(\mathscr{M}; \alpha)) - \bar{z}|^2. \tag{6}$$

To sum up, finding an optimal linear estimator is equivalent to solve the following finite-dimensional unconstrained problem (**SP**):

$$(\mathbf{SP}) \qquad \min_{\alpha} J(\alpha) \tag{7}$$

The solutions to optimization problem (SP) provide a foundation for sequential estimation of the game. One trivial solution to the problem is to let $\alpha^\star$ such that $J(\alpha^\star) = 0$. Consider ratio $\kappa_i := z/\text{val}(M_i),\ \ i \in \mathscr{S}$. A subset of optimal points $\alpha^\star$ would be $\{\kappa_i e_i\}_{i=1}^{S}$, where $\{e_i\}_{i=1}^{S}$ represents the standard basis of $\mathbb{R}^S$. These vectors are trivial solutions obtained by degenerating the set of multiple expert games into a singleton. The resulting estimation is a scaling of a chosen expert game. It is apparent that they are strategically equivalent games. However, these trivial solutions are arguably biased in terms of combining the information given by the experts and we need optimal points that take multiple expert games into consideration. In section 4, we study $J(\alpha)$ further to develop iterative algorithmic solutions.

3.2 Dynamic Linear Estimation Problem

Building on the estimation problem above, we formulate a dynamic linear estimation problem. Consider that the game is played sequentially. At the beginning of each time step t, the player has cumulated t expert-game sets $\{\mathscr{M}^{(t')}\}_{t'=1}^{t}$. At step t, an unknown game $\bar{M}^{(t)}$ is sampled from the underlying probability space $(\Omega, \mathscr{F}, \mathbb{P})$. The defender can observe the outcome of the play $\bar{z}^{(t)}$, which is the saddle-point value of the unknown game, i.e., $\bar{z}^{(t)} = \mathrm{val}(\bar{M}^{(t)})$. By the end of the play, the defender has accumulated information $I^{(t)} = \{ \{\mathscr{M}^{(t')}\}_{t'=1}^{t}, \{\bar{z}^{(t')}\}_{t'=1}^{t} \}$.

The goal of the defender is to find a sequential estimator $\mu_t(I^{(t)})$ to estimate a sequence of unknown games $\bar{M}^{(t)}$ based on his accumulated information.

At each step t, we consider a linear estimator $\mu_t(I^{(t)})$ taking the form of

$$\mu_t(I^{(t)}) := L(\mathscr{M}^{(t)}; \alpha) := \alpha_1 M_1^{(t)} + \alpha_2 M_2^{(t)} + \ldots + \alpha_S M_S^{(t)}.$$

Here, the linear mapping $L : \mathbb{R}^S \to \mathbb{R}^{N_1 \times N_2}$ is parameterized by a fixed vector α. At time t, the optimal parameters $\alpha^*(t)$ minimize the time-average accumulated residue error as follows:

$$J^{(t)}(\alpha) = \frac{1}{t} \sum_{t'=1}^{t} |\bar{z}^{(t')} - \mathrm{val}(L(\mathscr{M}^{(t')}; \alpha))|^2. \quad (8)$$

It is clear that $J^{(t)}$ depends on the samples of the game at each step t. We formulate the nonlinear regression problem at time t called **DP**-t.

$$(\mathbf{DP} - t) \qquad \min_{\alpha} J^{(t)}(\alpha) \quad (9)$$

Discussion on Asymptotic Behavior The formulated problem coincides with the standard form of nonlinear regression with a linearly parameterized function class, in which the following presumption holds:

$$\bar{z}^{t'} = \mathrm{val}(L(\mathscr{M}^{(t')}; \alpha_0)) + \varepsilon^{(t')} \qquad t' = 1, \ldots, t \quad (10)$$

where $\varepsilon^{(t')}$ are i.i.d. errors with zero mean and bounded variance, and α_0 is the true parameter. The least-square estimator $\alpha^*(t)$ is said to be strongly (weakly) consistent if $\alpha^*(t) \to \alpha_0$ a.s. (in prob.) as $t \to \infty$ [22].

The strong or weak consistency of $\alpha^*(t)$ depends on a series of conditions rigorously proved in [11,14,22]. Under the assumption of consistency, $\alpha^*(t)$ is asymptotically unbiased and induces minimum variance. In such case, while the estimation of game matrix is not necessarily unbiased, it still provides valuable information, since the value of estimated game enjoys asymptotic optimality.

4 Objective Function Analysis

In this section, we provide analytical results to give theoretical insights on the problem. We first characterize several properties of the objective functions including their continuity, differentiability, and convexity. In the second part of this section, we study parameter perturbations on the objective function.

4.1 Basic Properties

Let $v^{(t)}(\alpha)$ be the error between observations and value of output game at step t for a linear estimator with parameter α, given by

$$v^{(t)}(\alpha) := \mathrm{val}(L(\mathscr{M}^{(t)};\alpha)) - \bar{z}^{(t)} = \mathrm{val}(L(\mathscr{M}^{(t)};\alpha) - z^{(t)}E) \tag{11}$$

Let $f^{(t)}(\alpha)$, and $g^{(t)}(\alpha)$ be the saddle-point strategies of estimated game M for a given α. The error (11) can be rewritten as

$$v^{(t)}(\alpha) = f^{(t)\mathrm{T}}(\alpha)(L(\mathscr{M}^{(t)};\alpha) - z^{(t)}E)g^{(t)}(\alpha)$$

where $E \in \mathbb{R}^{N_1 \times N_2}$ is a matrix with all entries being 1. In dynamic estimation problems, the accumulated squared error up to time t is $J^{(t)}(\alpha) = \sum_{t'=1}^{t} \left(v^{(t')}(\alpha)\right)^2$.

Lemma 1. *$v^{(t)}(\alpha)$ is continuous differentiable in domain $\mathbb{R}^S$, so is $J^{(t)}(\alpha)$.*

Proof. From [21], $|\mathrm{val}(A) - \mathrm{val}(B)| \leq d(A,B)$ for any real matrices $A, B \in \mathbb{R}^{N_1 \times N_2}$ with metric $d(A,B) = \max_{i\in\mathscr{A}_1, j\in\mathscr{A}_2} |A_{ij} - B_{ij}|$. For sufficiently small ε and all-one S dimension vector $\mathbf{1}_S$,

$$|v^{(t)}(L(\mathscr{M};\alpha+\varepsilon\mathbf{1}_S)) - v^{(t)}(L(\mathscr{M};\alpha))| \leq \varepsilon \max_{i\in\mathscr{A}_1\, j\in\mathscr{A}_2} |\sum_{s\in\mathscr{S}} (M_s)_{ij}|.$$

$v^{(t)}(\alpha)$ is continuous as the term $\max_{i\in\mathscr{A}_1\, j\in\mathscr{A}_2} |\sum_{s\in\mathscr{S}} (M_s)_{ij}|$ is bounded. Picking the $\|\cdot\|_2$ norm, we arrive at

$$\lim_{\varepsilon\to 0} \frac{|v^{(t)}(L(\mathscr{M};\alpha+\varepsilon\mathbf{1}_S)) - v^{(t)}(L(\mathscr{M};\alpha))|}{\|\alpha+\varepsilon\mathbf{1}_S - \alpha\|_2} \leq \frac{1}{\|\mathbf{1}_S\|_2} \max_{i\in\mathscr{A}_1\, j\in\mathscr{A}_2} |\sum_{s\in\mathscr{S}} (M_s)_{ij}|.$$

Thus, given bounded expert game matrices, $v^{(t)}(\alpha)$ is continuous differentiable in $\mathbb{R}^S$, and so is $J^{(t)}(\alpha)$ since it is a sum of squares of $v^{(t')}(\alpha)$. □

Lemma 2. *$J(\alpha)$ is non-convex in domain $\mathbb{R}^S$.*

Proof. We prove the result by contradiction. Suppose that $J(\alpha)$ is convex in the convex domain $\mathbb{R}^S$, then it must satisfy that $\forall \lambda \in [0,1]$ and $\forall \alpha_1, \alpha_2 \in \mathbb{R}^S$,

$$J(\lambda\alpha_1 + (1-\lambda)\alpha_2) \leq \lambda J(\alpha_1) + (1-\lambda)J(\alpha_2). \tag{12}$$

Pick arbitrary $\lambda \in (0,1)$ and two fundamental solutions: $\alpha_1 = \kappa_1 e_1$, $\alpha_2 = \kappa_2 e_2$ in (12) and yield

$$\begin{aligned} & |\mathrm{val}(L(\mathscr{M}^{(t)}; \lambda\alpha_1 + (1-\lambda)\alpha_2)) - \bar{z}|^2 \leq 0 \\ \Rightarrow \quad & \mathrm{val}\Big(\frac{M_2}{\mathrm{val}(M_2)} + \lambda\big(\frac{M_1}{\mathrm{val}(M_1)} - \frac{M_2}{\mathrm{val}(M_2)}\big)\Big) = 1 \end{aligned}$$

Thus, for bounded matrix M_1 and M_2 which has nonzero saddle-point values, it must hold that

$$M_1 = \frac{\mathrm{val}(M_1)}{\mathrm{val}(M_2)} M_2,$$

which contradicts to property (2). This contradiction indicates that $J(\alpha)$ is not convex. □

4.2 Perturbation Theory of Parameterized Matrix Game

In this subsection, we determine the first-order and second-order derivatives of the game value with respect to entries of the payoff matrix. We first introduce the concept of completely mixed games.

Definition 1. *A matrix game M is said to be completely mixed if, for every saddle-point solution (f,g), no element of f or g is zero. If M is completely mixed, then $N_1 = N_2$ and the saddle-point solution of M is unique.*

Let $\hat{M}^{(t)} := L(\mathscr{M}^{(t)}; \alpha)$ denote the estimation of the game at time t. We make the following assumptions on the parameter space and estimated game.

Assumption 1 *The parameter space $\mathscr{X}$ is a subset of Euclidean space $\mathbb{R}^S$ where for all $\alpha \in \mathscr{X}$, $c(\alpha) \leq \|\alpha\| \leq C(\alpha)$.*

Assumption 1 restricts the parameter to a compact space. It prevents the output estimation from approaching infinity or $\mathbf{0}$.

Assumption 2 *$\hat{M}^{(t)}$ is completely mixed for all t.*

Assumption 2 implies that the estimated game matrix is square and nonsingular. It enables the computation of first-order and second-order derivatives of the objective functions.

For games that are not completely mixed, their computations remain an open problem. Lloyd Shapley [6] has observed that the nonexistence of any derivatives as a function of a given matrix element correspond to degeneracies in the linear-programming solution of the game. Assumption 2 coincides with the facts in [5] that the set of $N_1 \times N_2$ matrices which have unique saddle-point points is open and everywhere dense in $N_1 \times N_2$-space; i.e., solutions are unique for most of the $N_1 \times N_2$ matrices. With assumption 2, we avoid equilibrium selection by degenerating saddle-point solution sets into singletons and ensure the uniqueness of $f^{(t)}(\alpha)$ and $g^{(t)}(\alpha)$. The explicit expression of saddle-point solutions are feasible under assumption 2, as shown in lemma 3 following [21].

Lemma 3. *Assume that $\mathbf{1}^{\mathrm{T}}[\hat{M}^{(t)}]^{-1}\mathbf{1}$ is nonzero. For every t and given α, under Assumption 2 and we have:*

(i) $v^{(t)}(\alpha) = 1/\mathbf{1}^{\mathrm{T}}[\hat{M}^{(t)}]^{-1}\mathbf{1} - z^{(t)}$.
(ii) $f^{(t)\mathrm{T}}(\alpha) = \mathbf{1}^{\mathrm{T}}[\hat{M}^{(t)}]^{-1}\mathrm{val}(\hat{M}^{(t)})$.
(iii) $g^{(t)}(\alpha) = [\hat{M}^{(t)}]^{-1}\mathbf{1}\mathrm{val}(\hat{M}^{(t)})$.

Here, vector $\mathbf{1}$ is a vector of appropriate dimension with all entries being 1. The assumption of $\mathbf{1}^{\mathrm{T}}[\hat{M}^{(t)}]^{-1}\mathbf{1}$ being nonzero is without loss of generality. Lemma 3 enables the following direct computation of the gradient of the error (11).

Theorem 1. *For every t, under Assumption 2, the gradient vector of the error (11) is given by*

$$\nabla v^{(t)}(\alpha) = (\delta_1^{(t)}(\alpha), \ldots, \delta_S^{(t)}(\alpha))^{\mathrm{T}}, \tag{13}$$

where $\delta_i(\alpha) = f^{(t)\mathrm{T}}(\alpha) M_i^{(t)} g^{(t)}(\alpha),\ i \in \mathscr{S}$. Furthermore, $\|\nabla v^{(t)}(\alpha)\|$ is bounded by some positive constant.

Proof. Given that $\hat{M}^{(t)}$ is completely mixed, the results in Lemma 3 hold. According to the product rule of derivatives, we have $\forall i \in \mathscr{S}$:

$$\begin{aligned}
\frac{\partial v^{(t)}(\alpha)}{\partial \alpha_i} &= f^{(t)\mathrm{T}}(\alpha) M_i^{(t)} g^{(t)}(\alpha) + \frac{\partial f^{(t)\mathrm{T}}(\alpha)}{\partial \alpha_i} \hat{M}^{(t)} g^{(t)}(\alpha) + f^{(t)\mathrm{T}}(\alpha) \hat{M}^{(t)} \frac{\partial g^{(t)}(\alpha)}{\partial \alpha} \\
&= f^{(t)\mathrm{T}}(\alpha) M_i^{(t)} g^{(t)}(\alpha) + \frac{\partial f^{(t)\mathrm{T}}(\alpha)}{\partial \alpha_i} \hat{M}^{(t)} [\hat{M}^{(t)}]^{-1} \mathbf{1} v^{(t)}(\alpha) \\
&\quad + v^{(t)}(\alpha) \mathbf{1}^{\mathrm{T}} [\hat{M}^{(t)}]^{-1} \hat{M}^{(t)} \frac{\partial g^{(t)}(\alpha)}{\partial \alpha_i} \\
&= f^{(t)\mathrm{T}}(\alpha) M_i^{(t)} g^{(t)}(\alpha) + v^{(t)}(\alpha) \left(\frac{\partial f^{(t)\mathrm{T}}(\alpha)\mathbf{1}}{\partial \alpha_i} + \frac{\partial \mathbf{1}^{\mathrm{T}} g^{(t)}(\alpha)}{\partial \alpha_i} \right) \\
&= f^{(t)\mathrm{T}}(\alpha) M_i^{(t)} g^{(t)}(\alpha).
\end{aligned}$$

Stacking all the partial derivatives of i's gives the gradient. For any $\alpha \in \mathscr{X}$ that satisfies Assumption 2, we have

$$\begin{aligned}
\|\nabla v^{(t)}(\alpha)\| &\leq \| \big(\max_{i \in \mathscr{A}_1,\ j \in \mathscr{A}_2} |(M_1^{(t)})_{ij}|, \ldots, \max_{i \in \mathscr{A}_1,\ j \in \mathscr{A}_2} |(M_S^{(t)})_{ij}| \big)^{\mathrm{T}} \| \\
\|\nabla v^{(t)}(\alpha)\| &\leq \| \big(\min_{i \in \mathscr{A}_1,\ j \in \mathscr{A}_2} |(M_1^{(t)})_{ij}|, \ldots, \min_{i \in \mathscr{A}_1,\ j \in \mathscr{A}_2} |(M_S^{(t)})_{ij}| \big)^{\mathrm{T}} \|.
\end{aligned}$$

Thus, for bounded expert matrices, $\|\nabla v^{(t)}(\alpha)\|$ is bounded too, which can be viewed as a corollary of Lemma 1. □

Corollary 1. *Under Assumption 2, the gradient of $J^{(t)}(\alpha)$ is given by*

$$\nabla J^{(t)}(\alpha) = \frac{2}{t} \sum_{t'=1}^{t'} \big((\delta_1^{(t)}(\alpha), \ldots, \delta_S^{(t')}(\alpha) \big)^{\mathrm{T}} v^{(t')}(\alpha). \tag{14}$$

Remark 1. The entry $\delta_i^{(t)}(\alpha)$ indicates the sensitivity or the change in the accumulated square error with respect to a perturbation of α_i. It can be interpreted as the partial contribution by expert i to the reduction of the error. Note that $f^{(t)\mathrm{T}}(\alpha)M_i^{(t)}g^{(t)}(\alpha)$ is the expected outcome of the expert game i, $M_i^{(t)}$, achieved with the saddle-point strategies of $\hat{M}^{(t)}$.

We are also interested in the sensitivity of $\nabla J^{(t)}(\alpha)$ with respect to the changes in variable α.

Theorem 2. *For every t, under Assumption 1 and 2, $v^{(t)}(\alpha)$ is twice continuously differentiable, and so is $J^{(t)}(\alpha)$. The Hessian of $v^{(t)}(\alpha) := \left[\frac{\partial^2 v^{(t)}(\alpha)}{\partial\alpha_i\partial\alpha_j}\right]_{i,j\in\mathscr{S}}$ is given by*

$$\frac{\partial^2 v^{(t)}(\alpha)}{\partial\alpha_i\partial\alpha_j} = \phi_{ij}^{(t)} M_i^{(t)} g^{(t)}(\alpha) + f^{(t)\mathrm{T}}(\alpha) M_i^{(t)} \varphi_{ij}^{(t)},\ \ i,j\in\mathscr{S}, \tag{15}$$

where

$$\phi_{ij}^{(t)} = \left(\mathbf{1}^{\mathrm{T}} f^{(t)\mathrm{T}}(\alpha) M_j^{(t)} g^{(t)}(\alpha) - f^{(t)\mathrm{T}}(\alpha) M_i^{(t)}\right)[\hat{M}^{(t)}]^{-1}$$
$$\varphi_{ij}^{(t)} = [\hat{M}^{(t)}]^{-1}\left(f^{(t)\mathrm{T}}(\alpha) M_j^{(t)} g^{(t)}(\alpha)\mathbf{1} - M_i^{(t)} g^{(t)}(\alpha)\right).$$

Furthermore, the Hessian $\nabla^2 J^{(t)}(\alpha)$ is bounded; i.e., there exists a positive constant, such that $\|\nabla^2 J^{(t)}(\alpha)\| \le \frac{1}{2}\beta$, where $\|\nabla^2 J^{(t)}(\alpha)\|$ is the maximum (real) eigenvalue.

Proof. Under Assumption 2, the derivative of (13) exists, for $i,j\in\mathscr{S}$:

$$\frac{\partial^2 v^{(t)}(\alpha)}{\partial\alpha_i\partial\alpha_j} = \frac{\partial f^{(t)}(\alpha)}{\partial\alpha_j} M_i^{(t)} g^{(t)}(\alpha) + f^{(t)\mathrm{T}}(\alpha) M_i^{(t)} \frac{\partial g^{(t)}(\alpha)}{\partial\alpha_j}$$

From Lemma 3, we have

$$f^{(t)\mathrm{T}}(\alpha)\hat{M}^{(t)} = \mathbf{1}^{\mathrm{T}}\mathrm{val}(\hat{M}^{(t)})$$
$$\hat{M}^{(t)} g^{(t)}(\alpha) = \mathrm{val}(\hat{M}^{(t)})\mathbf{1}.$$

Take derivative w.r.t α_j on both sides and we arrive at the derivative of the saddle-point strategies:

$$\phi_{ij}^{(t)} = \frac{\partial f^{(t)}(\alpha)}{\partial\alpha_j} = \left(\mathbf{1}^{\mathrm{T}} f^{(t)\mathrm{T}}(\alpha) M_j^{(t)} g^{(t)}(\alpha) - f^{(t)\mathrm{T}}(\alpha) M_i^{(t)}\right)[\hat{M}^{(t)}]^{-1}$$
$$\varphi_{ij}^{(t)} = \frac{\partial g^{(t)}(\alpha)}{\partial\alpha_j} = [\hat{M}^{(t)}]^{-1}\left(f^{(t)\mathrm{T}}(\alpha) M_j^{(t)} g^{(t)}(\alpha)\mathbf{1} - M_i^{(t)} g^{(t)}(\alpha)\right)$$

The Hessian $\nabla^2 J^{(t)}(\alpha)$ can be constructed using the first and second-order derivatives of $v^{(t)}(\alpha)$. Its entry takes the following form:

$$[\nabla^2 J^{(t)}(\alpha)]_{ij} = \sum_{t'=1}^{t} \frac{\partial v^{(t')}(\alpha)}{\partial\alpha_i}\frac{\partial v^{(t')}(\alpha)}{\partial\alpha_j} + \frac{\partial^2 v^{(t')}(\alpha)}{\partial\alpha_i\partial\alpha_j} v^{(t')}(\alpha).$$

Using triangular inequality, we obtain

$$\|[\nabla^2 J^{(t)}(\alpha)]_{ij}\| \leq \sum_{t'=1}^{t} \|\frac{\partial v^{(t')}(\alpha)}{\partial \alpha_i}\frac{\partial v^{(t')}(\alpha)}{\partial \alpha_j} + \frac{\partial^2 v^{(t')}(\alpha)}{\partial \alpha_i \partial \alpha_j} v^{(t')}(\alpha)\|$$
$$\leq \sum_{t'=1}^{t} \underbrace{\|\frac{\partial v^{(t')}(\alpha)}{\partial \alpha_i}\frac{\partial v^{(t')}(\alpha)}{\partial \alpha_j}\|}_{\text{first term}} + \underbrace{\|\frac{\partial^2 v^{(t')}(\alpha)}{\partial \alpha_i \partial \alpha_j} v^{(t')}(\alpha)\|}_{\text{second term}}.$$

The boundedness of Hessian entry is determined by the first term and the second term. We have for any $t' \in \{1,\dots,t\}$, the first term is bounded according to Theorem 1:

$$\|\frac{\partial v^{(t')}(\alpha)}{\partial \alpha_i}\frac{\partial v^{(t')}(\alpha)}{\partial \alpha_j}\| \leq \|\max_{a,b}[M_i^{(t')}]_{ab}\| \cdot \|\max_{a,b}[M_j^{(t')}]_{ab}\|.$$

For the second term,

$$\|\frac{\partial^2 v^{(t')}(\alpha)}{\partial \alpha_i \partial \alpha_j} v^{(t')}(\alpha)\| \leq \|v^{(t')}(\alpha)\phi_{ij}^{(t')} M_i^{(t')} g^{(t')}(\alpha)\| + \|f^{(t')\mathrm{T}}(\alpha) M_i^{(t')} \varphi_{ij}^{(t')} v^{(t')}(\alpha)\|$$
$$\leq (Q+P)\|\mathrm{val}(\hat{M}) - \bar{z}^{(t')}\|\|[\hat{M}^{(t')}]^{-1}\|$$
$$\leq (Q+P)(\|\mathrm{val}(\hat{M})\|\|[\hat{M}^{(t')}]^{-1}\| + \|\bar{z}^{(t')}\|\|[\hat{M}^{(t')}]^{-1}\|),$$

where Q and P are positive constants such that

$$\|\mathbf{1}^{\mathrm{T}} f^{(t)\mathrm{T}}(\alpha) M_j^{(t)} g^{(t)}(\alpha) - f^{(t)\mathrm{T}}(\alpha) M_i^{(t)}\| \cdot \|M_i^{(t)} g^{(t)}(\alpha)\| \leq Q$$
$$\|f^{(t)\mathrm{T}}(\alpha) M_j^{(t)} g^{(t)}(\alpha)\mathbf{1} - M_i^{(t)} g^{(t)}(\alpha)\| \cdot \|f^{(t)\mathrm{T}}(\alpha) M_i^{(t)}\| \leq P.$$

The parameterized $\|[\hat{M}^{(t')}]\|^{-1}$ is bounded since α is lower bounded by positive constant according to assumption 1. Since the eigenvalue of a square matrix is bounded by its maximum entry multiplied by its order, $\|[\hat{M}^{(t)}]^{-1}\|\|\mathrm{val}(\hat{M}^{(t)})\|$ is also bounded, according to Lemma 3:

$$\|[\hat{M}^{(t)}]^{-1}\|\|\mathrm{val}(\hat{M}^{(t)})\| = \|[\hat{M}^{(t)}]^{-1}\| / \|\mathbf{1}^{\mathrm{T}}[\hat{M}^{(t)}]^{-1}\mathbf{1}\|$$
$$\leq \frac{N_1 \max_{i,j}\left([\hat{M}^{(t')}]^{-1}\right)_{ij}}{\sum_{i,j}\left([\hat{M}^{(t')}]^{-1}\right)_{ij}}$$

Similarly, boundedness of Hessian entries implies that its eigenvalues are bounded by some constant, and thus we arrive at a bound β. □

In the following, we provide a lemma that establishes the relation between bounded Hessian and Lipschitz continuity, and then give the main theorem that ensures the convergence of gradient-based algorithms.

Lemma 4. *Let $f : \mathbb{R}^S \to \mathbb{R}$ be a twice continuously differentiable function. If there exists a positive constant β such that $\|\nabla^2 f\| \leq \beta$, where $\|\nabla^2 f\|$ is the matrix norm, then*

$$\forall \alpha, \tilde{\alpha} \in \mathbb{R}^S : \quad \|\nabla f(\alpha) - \nabla f(\tilde{\alpha})\| \leq \beta \|\alpha - \tilde{\alpha}\|.$$

Proof. The result can be proved by using a second-order Taylor expansion around α and $\tilde{\alpha}$, i.e.,

$$\begin{aligned} f(\alpha) - f(\tilde{\alpha}) &= \nabla f(\tilde{\alpha})^{\mathrm{T}}(\alpha - \tilde{\alpha}) + \frac{1}{2}(\tilde{\alpha} - \alpha)^{\mathrm{T}} \nabla^2 f(\xi_1)(\tilde{\alpha} - \alpha) \\ &= -\nabla f(\alpha)^{\mathrm{T}}(\tilde{\alpha} - \alpha) - \frac{1}{2}(\alpha - \tilde{\alpha})^{\mathrm{T}} \nabla^2 f(\xi_2)(\alpha - \tilde{\alpha}), \end{aligned}$$

where $\xi_1 = \alpha + t_1(\tilde{\alpha} - \alpha)$ and $\xi_2 = \tilde{\alpha} + t_2(\alpha - \tilde{\alpha})$ and $t_1, t_2 \in (0,1)$. We combine the two relations and obtain

$$\begin{aligned} \|\nabla f(\alpha) - \nabla f(\tilde{\alpha})\| &\leq \frac{1}{2}\|\nabla^2 f(\xi_2)\|\|\tilde{\alpha} - \alpha\| + \frac{1}{2}\|\nabla^2 f(\xi_2)\|\|\alpha - \tilde{\alpha}\| \\ &\leq \beta\|\tilde{\alpha} - \alpha\|. \end{aligned}$$

□

Theorem 3. *For every t, under Assumption 2, the vector functions $|v^{(t)}(\alpha)|^2$ are Lipschitz continuous; i.e., there exists a Lipschitz constant $\beta > 0$, such that for all $\alpha, \tilde{\alpha} \in \mathscr{X}$ that satisfies*

$$\|\nabla v^{(t')}(\alpha) v^{(t')}(\alpha) - \nabla v^{(t')}(\tilde{\alpha}) v^{(t')}(\tilde{\alpha})\| \leq \beta\|\alpha - \tilde{\alpha}\|; \tag{16}$$

and

$$\|\nabla J^{(t)}(\alpha) - \nabla J^{(t)}(\tilde{\alpha})\| \leq 2\beta\|\alpha - \tilde{\alpha}\|. \tag{17}$$

Furthermore, the following holds:

$$J^{(t)}(\alpha) - J^{(t)}(\tilde{\alpha}) \leq (\nabla J^{(t)}(\tilde{\alpha}))^{\mathrm{T}}(\alpha - \tilde{\alpha}) + \beta\|\alpha - \tilde{\alpha}\|^2. \tag{18}$$

Proof. Inequality (16) immediately follows Lemma 4 and the analysis in Theorem 2. To obtain (17), we add up (16) for all t' and use the triangular inequality.

$$\begin{aligned} \|\nabla J^{(t)}(\alpha) - \nabla J^{(t)}(\tilde{\alpha})\| &\leq 2\sum_{t'=1}^{t} \|\nabla v^{(t')}(\alpha) v^{(t')}(\alpha) - \nabla v^{(t')}(\tilde{\alpha}) v^{(t')}(\tilde{\alpha})\| \\ &\leq 2\beta\|\alpha - \tilde{\alpha}\|. \end{aligned}$$

Inequality (18) is a basic result following (17):

$$\begin{aligned} J^{(t)}(\alpha) - J^{(t)}(\tilde{\alpha}) &= \int_0^1 (\alpha - \tilde{\alpha})^{\mathrm{T}} \nabla J^{(t)}(\tilde{\alpha} + \xi(\alpha - \tilde{\alpha})) d\xi \\ &\leq \int_0^1 (\alpha - \tilde{\alpha})^{\mathrm{T}} \nabla J^{(t)}(\tilde{\alpha}) d\xi \\ &\quad + \int_0^1 \|\alpha - \tilde{\alpha}\|\|\nabla J^{(t)}(\tilde{\alpha} + \xi(\alpha - \tilde{\alpha})) - \nabla J^{(t)}(\tilde{\alpha})\| d\xi \\ &\leq (\nabla J^{(t)}(\tilde{\alpha}))^{\mathrm{T}}(\alpha - \tilde{\alpha}) + \beta\|\alpha - \tilde{\alpha}\|^2. \end{aligned}$$

□

The gradient and the Hessian of the errors, together with the property of Lipschitz continuity, provides a theoretical foundation for developing gradient-based algorithms, which will be discussed in Section 5.

5 Algorithmic Analysis

In this section, we develop gradient-based algorithms to find the linear optimal estimator, and study their convergence properties. We first formally present the optimality conditions that characterize the solutions to the dynamic problem 8.

Proposition 1 *(Stationary Points) For every t, due to non-convexity, we are satisfied at finding a solution $\alpha^*(t)$ for $J^{(t)}(\alpha)$ in **DP**-t that satisfies the first-order conditions,*

$$\nabla J^{(t)}(\alpha^*) = 0$$

which we refer to as stationary points.

A descent algorithm starts from initial point α^0, proceeding iteratively as follows:

$$\alpha^{k+1} = \alpha^k + \gamma^k s^k, \qquad k = 0, 1, 2, \ldots,$$

where $\gamma^k \in \mathbb{R}_+$ is the stepsize and $s_k \in \mathbb{R}^S$ represents the descent direction. Many choices are plausible for the descent direction, resulting in different algorithmic implementations; e.g., *steepest gradient* (i.e., $s^k = -\nabla J^{(t)}(\alpha^k)$), *Newton's method* (i.e., $s^k = -(\nabla^2 J^{(t)}(\alpha^k))^{-1} \nabla J^{(t)}(\alpha^k)$), and other variants (e.g., quasi-Newton methods). Algorithm 1 gives a steepest gradient descent algorithm, which is well known to achieve a linear convergence rate. The tolerance ε denotes the stopping criteria.

Algorithm 1: Optimal Linear Estimation Using Steepest Gradient

```
Data: {𝓜^(t')}_{t'=1}^t, {z^(t')}_{t'=1}^t;
Input: α^0, {γ^k}, ε;
for k ← 1,2,... do
    foreach i ← 1 to t do
        (f^(i), g^(i)) ← saddle-point(L(𝓜^(i), α) − z^(i) E)
    end
    ∇J^(t)(α^k) ← (1/t) Σ_{i=1}^t (∇v^(i)(α)) v^(i)(α);
    if ||∇J^(t)(α^k)|| ≤ ε ;
     then
        Break
    end
    α^{k+1} ← α^k − γ^k ∇J^(t)(α^k);
end
Result: α*
```

Pseudo-Gradient Approximation As saddle-point strategies are computationally costly to obtain, determining a steepest direction is relatively inefficient. In fact, the descent direction can be approximated once the approximation error is sufficiently small. We hereby provide a pseudo gradient method that uses a surrogate descent direction $\bar{s}^k$, where for all $i \in \mathscr{S}$

$$\bar{s}_i^k = \sum_{t'=1}^{t} \frac{1}{N_1 N_2} \sum_{i,j} (M_i^{(t')})_{ij} \left(\frac{1}{N_1 N_2} \sum_{i,j} (\hat{M}^{(t')})_{ij} - \bar{z}^{(t')} \right). \tag{19}$$

In short, the pseudo-gradient approximates the gradient by replacing $\delta_i^{(t')}$ with the mean value of $M_i^{(t')}$ and replacing $\mathrm{val}(\hat{M}^{(t')})$ with average entry value of $\hat{M}^{(t')}$. By doing so, we eliminate the problem for computing the saddle-point strategies and game values, significantly reducing the computational complexity.

5.1 Sequential Observation and Adaptation

When t becomes large, steepest gradient methods are inefficient as it needs to sweep through the entire dataset. It is more attractive to use an incremental method that can sequentially update the gradient. The incremental gradient method is described as follows:

$$\alpha^{k+1} = \alpha^k - \gamma^k \left(\sum_{i=1}^{t} \nabla v^{(i)}(\psi^{i-1}) v^{(i)}(\psi^{i-1}) \right), \tag{20}$$

where at iteration k:

$$\psi^i = \psi^{i-1} - \gamma^k \nabla v^{(i)}(\psi^{i-1}) v^{(i)}(\psi^{i-1}) \qquad i = 1, \ldots, t.$$

The stepsize selection is essential to ensure the convergence of the iterations. Usually when γ^k does not diminish to 0, there will be an oscillation within ψ^i.

Assumption 3 *The following conditions are satisfied:*

(a) The product of every error (11) and its gradient is bounded for all $\alpha \in \mathscr{X}$ and every t', t; i.e.,

$$\|\nabla v^{(t')}(\alpha) v^{(t')}(\alpha)\| \leq c_1 + c_2 \|\nabla J^{(t)}(\alpha)\| \tag{21}$$

for positive constants c_1 and c_2;

(b) Diminishing stepsize, i.e., $\sum_{k=0}^{\infty} \gamma^k = \infty$ and $\sum_{k=0}^{\infty} (\gamma^k)^2 < \infty$.

Corollary 2. *Under Assumption 3, for all $\alpha \in \mathscr{X}$, we have*

$$(1 - 2c_2) \|\nabla J^{(t)}(\alpha)\| \leq 2c_1. \tag{22}$$

Particularly, when $0 < c_2 < \frac{1}{2}$, $\|\nabla J^{(t)}(\alpha)\|$ is bounded by $\dfrac{2c_1}{1-2c_2}$.

This bound can be obtained through triangular inequality:

$$\begin{aligned} \|\nabla J^{(t)}(\alpha)\| &= \frac{2}{t} \| \sum_{t'=1}^{t} \nabla v^{(t')}(\alpha) v^{(t')}(\alpha) \| \\ &\leq \frac{2}{t} \sum_{t'=1}^{t} \|\nabla v^{(t')}(\alpha) v^{(t')}(\alpha)\| \\ &\leq 2c_1 + 2c_2 \|\nabla J^{(t)}(\alpha)\| \end{aligned}$$

Proposition 2 *Under Assumption 3, the incremental gradient method 20 applied to 8 generates a sequence $\{\alpha^k\}$. $J^{(t)}(\alpha^k)$ converges to a finite value and $\lim_{k\to\infty} \nabla J^{(t)}(\alpha^k) = 0$. Every limit point of α^k is a stationary point of problem 8.*

Proof. We provide a sketch of the proof here. At iteration k, we have

$$\begin{aligned}
\psi^1 &= \alpha^k - \gamma^k \nabla v^{(1)}(\alpha^k) v^{(1)}(\alpha^k) \\
\psi^2 &= \alpha^k - \gamma^k \nabla v^{(2)}(\psi^1) v^{(2)}(\psi^1) \\
\vdots &\quad\quad \vdots \\
\psi^t &= \alpha^k - \gamma^k \nabla v^{(t)}(\psi^{t-1}) v^{(t)}(\psi^{t-1})
\end{aligned}$$

Adding them up, we obtain

$$\begin{aligned}
\alpha^{k+1} &= \alpha^k - \gamma^k \big(\nabla J^{(t)}(\alpha^k) - \sum_{t'=2}^{t} (\nabla v^{(t')}(\alpha^k) v^{(t')}(\alpha^k) - \nabla v^{(t')}(\psi^{t'-1}) v^{(t')}(\psi^{t'-1}))\big) \\
&= \alpha^k - \gamma^k \big(\nabla J^{(t)}(\alpha^k) - w^k\big)
\end{aligned}$$

Using Theorem 3, we see that the error term $w^k = \sum_{t'=2}^{t} \nabla v^{(t')} v^{(t')}(\alpha^k) - \nabla v^{(t')} v^{(t')}(\psi^{t'-1}) = \sum_{t'=2}^{t} w_{t'}^k$ is bounded, for every t':

$$\begin{aligned}
w_{t'}^k &\leq \sum_{i=2}^{t-1} \|\nabla v^{(t')} v^{(t')}(\psi^i) - \nabla v^{(t')} v^{(t')}(\psi^{i-1})\| \\
&\quad + \|\nabla v^{(t')} v^{(t')}(\alpha^k) - \nabla v^{(t')} v^{(t')}(\psi^1)\| \\
&\leq \beta \big(\|\alpha^k - \psi^1\| + \sum_{i=2}^{t-1} \|\psi^i - \psi^{i-1}\|\big) \\
&= \beta\gamma^k (\|\nabla v^{(t')} v^{(t')}(\alpha^k)\| + \sum_{i=1}^{t-2} \|\nabla v^{(t')} v^{(t')}(\psi^i)\|).
\end{aligned}$$

According to Assumption 3 (21),

$$w_{t'}^k \leq \beta\gamma^k ((t-1)(c_1 + c_2 \|\nabla J^{(t')}(\alpha^k))\| + \sum_{i=1}^{t-2} \|\nabla J^{(t')}(\alpha^k) - \nabla J^{(t')}(\psi^i)\|)$$

Leveraging Corollary 2, we recursively eliminate $\nabla J^{(t')}(\psi^i)$ and see that the error term w_t is bounded; i.e., there exist positive constants C_1 and C_2 such that

$$w^k \leq \gamma^k (C_1 + C_2 \|\nabla J^{(t)}(\alpha^k)\|) \tag{23}$$

Here, we omit the algebraic calculation of constants C_1 and C_2. Note that the elimination procedures are similar. Using (18), we obtain

$$\begin{aligned}
J^{(t)}(\alpha^{k+1}) - J^{(t)}(\alpha^k) &\leq \gamma^k (-\|\nabla J^{(t)}(\alpha^k)\|^2 + \|\nabla J^{(t)}(\alpha^k)\| \|w^k\|) \\
&\quad + \gamma^2 \beta \|\nabla J^{(t)}(\alpha) + w^k\|^2 \\
&\leq \gamma^k (-1 + \gamma^k (C_2 + 2\beta) + 2(\gamma^k)^3 C_2^2 \beta) \|\nabla J^{(t)}(\alpha)\|^2 \\
&\quad + (\gamma^k)^2 (C_1 + 4\gamma^2 C_1 C_2 \beta) \|\nabla J^{(t)}(\alpha)\| + 2(\gamma^k)^4 C_1^2 \beta
\end{aligned}$$

As Assumption 3 states that $(\gamma^k)^2$ diminishes to 0, the terms multiplying γ^k with order 2 or higher will go to 0. For k sufficiently large, $\gamma^k \to 0$, for some positive constants c_1' and c_2',

$$J^{(t)}(\alpha^{k+1}) - J^{(t)}(\alpha^k) \le -\gamma^k c_1' \|\nabla J^{(t)}(\alpha)\|^2 + (\gamma^k)^2 c_2' \|\nabla J^{(t)}(\alpha)\| + 2(\gamma^k)^4 C_1^2 \beta.$$

Observe that if $\|\nabla J^{(t)}(\alpha)\| \ge 1$, then $\|\nabla J^{(t)}(\alpha)\| < \|\nabla J^{(t)}(\alpha)\|^2$, or else $\|\nabla J^{(t)}(\alpha)\|^2 \le \|\nabla J^{(t)}(\alpha)\| \le 1$, and thus $\|\nabla J^{(t)}(\alpha)\| \le 1 + \|\nabla J^{(t)}(\alpha)\|^2$. Then,

$$J^{(t)}(\alpha^{k+1}) - J^{(t)}(\alpha^k) \le -\gamma^k (c_1' - \gamma^k c_2') \|\nabla J^{(t)}(\alpha)\|^2 + o((\gamma^k)^2). \tag{24}$$

For k sufficiently large, $c_1' - \gamma^k c_2' \le 0$, so that $J^{(t)}(\alpha^{k+1}) \le J^{(t)}(\alpha^k)$ and $J^{(t)}(\alpha^{k+1}) \ge 0$. (24) satisfies the deterministic form of supermartingale theorem. Hence $J(\alpha)$ converges to some finite value and it must have $\sum_{k=0}^{\infty} \gamma^k \|\nabla J^{(t)}(\alpha^k)\|^2 \le \infty$. Since we assume $\sum_{k=0}^{\infty} \gamma^k = \infty$, it also has to satisfy $\liminf_{k\to\infty} \|\nabla J^{(t)}(\alpha^k)\| = 0$. Due to Lipschitz continuity, $\limsup_{k\to\infty} \nabla J^{(t)}(\alpha^k)$ is also 0 (the proof is omitted here), and hence the limit points are stationary points. □

Stochastic Gradient Descent (SGD) The surrogate estimated gradient is:

$$\alpha^{k+1} = \alpha^k - \gamma^k \nabla \widehat{J}^{(t)}(\alpha^k) \tag{25}$$

$$= \alpha^k - \gamma^k \frac{1}{|B|} \sum_{b \in B} \nabla v^{(b)}(\alpha^k) v^{(b)}(\alpha^k), \tag{26}$$

where the indices b is chosen from batch set B. SGD is a stochastic version of incremental method, exhibiting a lower computational cost in one single iteration with less gradient memory storage. SGD guarantees weak convergence in non-convex systems under Lipschitz-smoothness, pseudo-gradient property, and bounded variance of the descent direction [4]. In our problem where there may exist multiple minimum, SGD potentially admits global optimum.

5.2 Extended Kalman Filter

We consider a commonly used iterative method for nonlinear least-square estimation, *Gauss-Newton* method, which is given as follows:

$$\alpha^{k+1} = \alpha^k - \gamma^k (\mathbf{J}_v \mathbf{J}_v^{\mathrm{T}} + \lambda I)^{-1} \mathbf{J}_v \mathbf{v}(\alpha^k), \tag{27}$$

where $\mathbf{J}_v = \big(\nabla v^{(1)}(\alpha^k), \ldots, \nabla v^{(t)}(\alpha^k)\big)$ is the Jacobian of the vector $\mathbf{v}(\alpha^k) = \big(v^{(1)}(\alpha^k), \ldots, v^{(t)}(\alpha^k)\big)^{\mathrm{T}}$ and λI stands for a possitive multiple of the identity matrix as proposed in *Levenberg-Marquardt* method [17] to ensure nonsingularity caused by the rank deficiency of $\mathbf{J}_v$.

Gauss-Newton iteration (27) is obtained by approximating Hessian with $(\mathbf{J}_v \mathbf{J}_v^{\mathrm{T}} + \Delta_t)$ as result of solving quadratic subproblems iteratively using linearized objective function around every α^k. This approximation avoids computing the individual residue Hessian $\nabla^2 v^{(t')}(\alpha), t' = 1, \ldots, t$, in Theorem 2.

Extended Kalman Filter (EKF) [4,3,16] is an incremental version of the Gauss-Newton method. Starting with some point α^0, a single cycle of the method updates the α via iterations that aims to minimize the partial sums $\sum_{t'=1}^{j} |v^{(t')}(\alpha)|^2$ $j = 1,\ldots,t$ successively. Thus, it sequentially generates the vectors:

$$\psi^{t'} = \arg\min_{\alpha} \sum_{i=1}^{t'} \left| v^{(i)}(\psi^{i-1}) + \left(\nabla v^{(i)}(\psi^{i-1})\right)^{\mathrm{T}}(\alpha - \psi^{i-1}) \right|^2 \qquad t' = 1,\ldots,t$$

We consider the algorithm where $\psi^{t'}$ are obtained through increments:

$$\psi^i = \psi^{i-1} - (H^i)^{-1} \nabla v^{(i)}(\psi^{i-1}) v^{(i)}(\psi^{i-1}), \qquad i = 1,\ldots,t, \tag{28}$$

with $\psi^0 = \alpha^k$ at step k, where matrices H^i are generated by:

$$H^i = \lambda H^{i-1} + \nabla v^{(i)}\left(\psi^{i-1}\right) \nabla v^{(i)}\left(\psi^{i-1}\right)^{\mathrm{T}}, \qquad i = 1,\ldots,t, \tag{29}$$

with λ being a positive constant and $H^0 = \lambda I$ at iteration $k = 0$. The algorithm uses ψ^t at the end of an iteration to update α^k:

$$\alpha^{k+1} = \alpha^k - (H^{t(k+1)})^{-1} \left(\sum_{i=1}^{t} \nabla v^{(i)}(\psi^{kt+i-1}) v^{(i)}(\psi^{kt+i-1}) \right), \tag{30}$$

where

$$H^{t(k+1)} = \lambda I + \sum_{j=0}^{k} \sum_{i=1}^{t} \nabla v^{(i)} \left(\psi^{kt+i-1}\right) \nabla v^{(i)} \left(\psi^{kt+i-1}\right)^{\mathrm{T}}. \tag{31}$$

Proposition 3 *(Extended Kalman Filter (EKF) [3]) Assuming that there is a constant* $c > 0$ *such that scalar* λ_k *used in the EKF algorithm at iteration k satisfies:*

$$0 \leq 1 - \lambda_k^t \leq \frac{c}{k}, \qquad k = 1,2,\ldots.$$

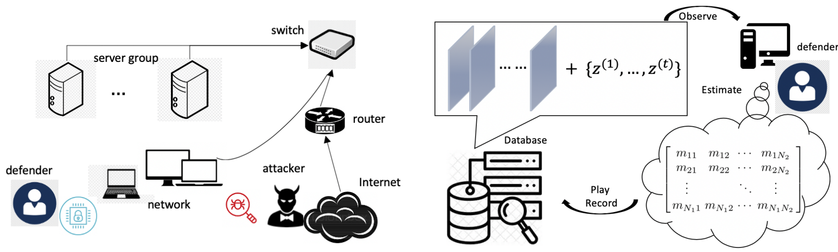
Then, the EKF algorithm generates a bounded sequence of vectors ψ^i*. Each of the limit points of* $\{\alpha^k\}$ *is a stationary point of the least-square problem 8.*

Proof. One can follow the argument in Proposition 2 of [3] to show the convergence of EKF, when a series of conditions are satisfied, among which the Lipschitz condition has been verified.

Remark 2. λ represents the discount factor that discounts the effects of old information. An interpretation of this algorithm is that, as the defender proceeds to estimate, the previous experience tends to be gradually out-of-date, while newly encountered ones should be highly valued in the estimation.

6 Case Study

In this section, we study a network configuration game to corroborate the results and investigate the numerical properties of the algorithms. Consider a game with an attacker



(a) Network configuration (b) Game estimation

Fig. 1. Illustration of adversarial interaction and estimation process.

and a defender in a network of server group. The defender chooses a subset of servers to monitor and protect, while the attacker selects a subset of them to attack. The interactions induce some value for both players.

Assuming that each player has four strategies and the defender does not know the game, we can use a $N_1 \times N_2$ matrix game with random entries to capture this scenario. The defender sequentially estimates the game based on past experiences (i.e., expert games) and value observation. This situation is illustrated in Fig. 1.

6.1 Experimental Setting and Results

Here, we conduct the experiment by fixing configuration parameters shown in Table 1. We generate the matrices $M^{(t')}$ and values of $\bar{z}^{(t')}$ from i.i.d. distributions $\mathcal{N}(\mu \mathbf{1}_4, \sigma^2 I_{4\times4})$ and $\mathcal{N}(\mu_z, \sigma_z^2)$, with a fixed random seed. As a result, the differences between values of expert games and target games scale well. We compare the performances of different methods for both **SP** and **DP**-t, and show their convergences in Fig. 2.

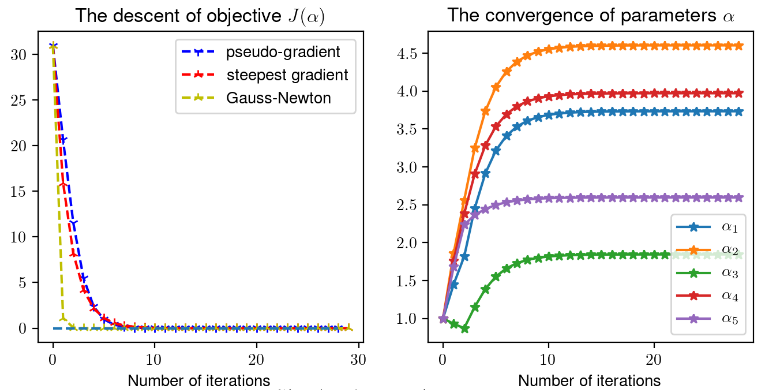
Table 1. Configurations

Variables	Values	Variables	Values
Data horizon t	30	$\mathscr{M}^{(t')}$ entry distribution (μ, σ)	$(1,1)$
Vector α Size S	5	$\bar{z}^{(t')}$ value distribution (μ_z, σ_z)	$(1,1)$
Stepsize γ^k	$0.98^k \times 0.01$	Parameter α initialization	$\mathbf{1}_S$
Tolerance ε	1e-5	Batch size $\lvert B\rvert$	1
Game size	4×4	Fading factor λ for EKF	0.9

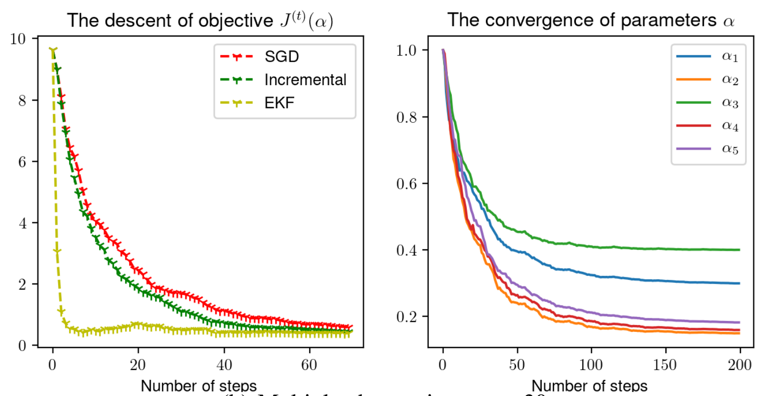
The well-known Lemke-Howson algorithm [13] is implemented to find the saddle-point strategies and values of matrix games.

6.2 Discussions

From Fig. 2, one shall see Gauss-Newton method as well as EKF exhibit convergence faster than others as they naturally tune the stepsize. Meanwhile, the pseudo-gradient



(a) Single observation at $t = 1$



(b) Multiple observation at $t = 30$

Fig. 2. Estimation curve for both static (a) and dynamic (b) problems

method displays promising convergence behavior. It can be seen in (a) that the partial contribution by expert 2 dominates the learning process, indicating greater similarity between expert game $M_2^{(1)}$ and $\hat{M}$.

We notice that the output square matrix $\hat{M}$ usually does not satisfy Assumption 2 as the estimated saddle-point mixed strategies have 0 elements in the iterative process. However, despite this, the algorithms still converge, indicating that assumption 2 is a conservative assumption for practice.

7 Conclusions and Future Research

This work has formulated and analyzed static and dynamic least-square game estimation problems for a class of finite zero-sum security games. The formulation captures the scenario where the players do not know the adversarial environments they interact with. We have studied the basic properties of least-square errors and developed iterative algorithms to solve the game estimation problem. The proposed approach effectively transfers the past experiences that are encoded as expert games to estimate the unknown game and inform future game plays. We have seen that the algorithms work over randomly generated datasets despite certain assumptions are not strictly satisfied.

There are many open research problems that could be addressed as future work. First, it has been observed that the assumption for completely mixed game is conservative. The future work would investigate the properties of the error functions when the assumption does not hold. Second, it would be possible to extend this framework for stochastic games. We would capture the dynamic adversarial environment using a stochastic game representation, and estimate the environment using multi-time scale observations.

References

1. Bajari, P., Hong, H., Ryan, S.: Identification and estimation of a normal form game. manuscript, Duke University (2004)
2. Bajari, P., Hong, H., Ryan, S.P.: Identification and estimation of a discrete game of complete information. Econometrica **78**(5), 1529–1568 (2010)
3. Bertsekas, D.P.: Incremental least squares methods and the extended kalman filter. SIAM Journal on Optimization **6**(3), 807 (1996)
4. Bertsekas, D.P., Tsitsiklis, J.N.: Neuro-dynamic programming. Athena Scientific (1996)
5. Bohnenblust, H., Karlin, S., Shapley, L.: Solutions of discrete, two-person games. Contributions to the Theory of Games **1**, 51–72 (1950)
6. Cohen, J.E.: Perturbation theory of completely mixed matrix games. Linear Algebra and its Applications **79**, 153 – 162 (1986). https://doi.org/https://doi.org/10.1016/0024-3795(86)90297-1, `http://www.sciencedirect.com/science/article/pii/0024379586902971`
7. Cohen, J.E., Marchi, E., Oviedo, J.A.: Perturbation theory of completely mixed bimatrix games. Linear Algebra and its Applications **114**, 169–180 (1989)
8. Gross, O.A.: The derivatives of the value of a game. (1954)
9. Harsanyi, J.C.: Games with incomplete information played by “bayesian” players, i–iii part i. the basic model. Management science **14**(3), 159–182 (1967)
10. Hotz, V.J., Miller, R.A.: Conditional choice probabilities and the estimation of dynamic models. The Review of Economic Studies **60**(3), 497–529 (1993)
11. Jennrich, R.I.: Asymptotic properties of non-linear least squares estimators. The Annals of Mathematical Statistics **40**(2), 633–643 (1969)
12. Jovanovic, B.: Observable implications of models with multiple equilibria. Econometrica: Journal of the Econometric Society pp. 1431–1437 (1989)
13. Lemke, C.E., Howson, Jr, J.T.: Equilibrium points of bimatrix games. Journal of the Society for industrial and Applied Mathematics **12**(2), 413–423 (1964)
14. Malinvaud, E.: The consistency of nonlinear regressions. The Annals of Mathematical Statistics **41**(3), 956–969 (1970)
15. Mertens, J.F., Zamir, S.: Formulation of bayesian analysis for games with incomplete information. International Journal of Game Theory **14**(1), 1–29 (1985)
16. Moriyama, H., Yamashita, N., Fukushima, M.: The incremental gauss-newton algorithm with adaptive stepsize rule. Computational Optimization and Applications **26**(2), 107–141 (2003)
17. Nocedal, J., Wright, S.: Numerical optimization. Springer Science & Business Media (2006)
18. Pesendorfer, M., Schmidt-Dengler, P.: Identification and estimation of dynamic games. Tech. rep., National Bureau of Economic Research (2003)

19. Pesendorfer, M., Schmidt-Dengler, P.: Asymptotic least squares estimators for dynamic games. The Review of Economic Studies **75**(3), 901–928 (2008)
20. Rass, S.: On game-theoretic network security provisioning. Journal of network and systems management **21**(1), 47–64 (2013)
21. Shapley, L.S., Snow, R.: Basic solutions of discrete games. Contributions to the Theory of Games **1**, 27–35 (1952)
22. Wu, C.F.: Asymptotic theory of nonlinear least squares estimation. The Annals of Statistics pp. 501–513 (1981)
23. Xiao, R.: Identification and estimation of incomplete information games with multiple equilibria. Journal of Econometrics **203**(2), 328–343 (2018)
24. Zhu, Q., Başar, T.: Dynamic policy-based ids configuration. In: Proceedings of the 48h IEEE Conference on Decision and Control (CDC) held jointly with 2009 28th Chinese Control Conference. pp. 8600–8605. IEEE (2009)
25. Zhu, Q., Li, H., Han, Z., Başar, T.: A stochastic game model for jamming in multi-channel cognitive radio systems. In: 2010 IEEE International Conference on Communications. pp. 1–6. IEEE (2010)