No-Regret Learning of Nash Equilibrium for Black-Box Games via Gaussian Processes

Minbiao Han*, Fengxue Zhang*, Yuxin Chen

University of Chicago

THE UNIVERSITY OF CHICAGO

Motivating Applications

Are you having trouble computing the Nash equilibrium of *black-box games*?

Is it because underlying agents' utility functions are *unknown* and can only be evaluated via *simulation*?

Is the evaluation on agents' utility information *noisy and too expensive*?

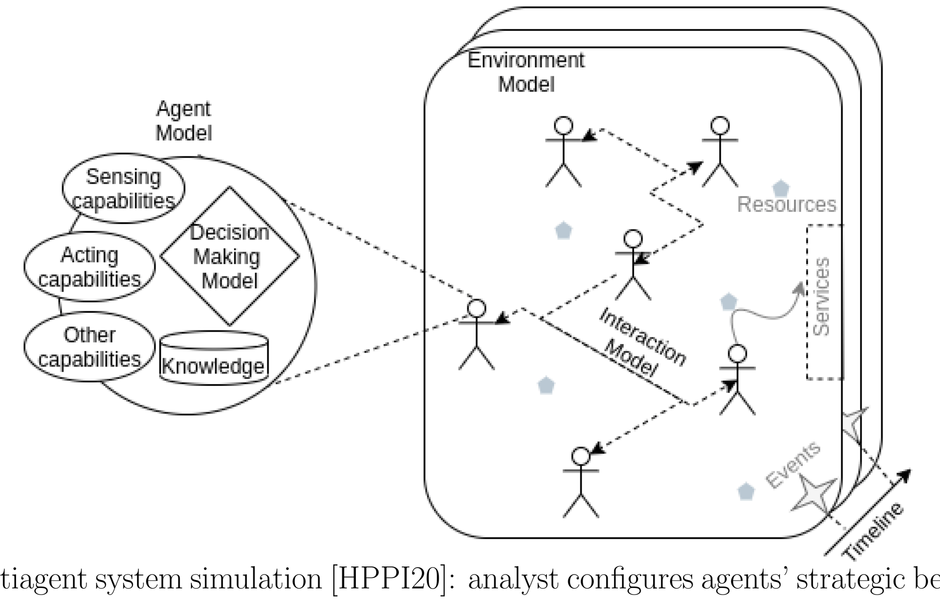

Figure 1: Multiagent system simulation [HPPI20]: analyst configures agents' strategic behavior models and simulates for an expected outcome.

Problem Formulation

- Nash Equilibrium (NE) strategy profile $\boldsymbol{x}^* = (x_1^*, \cdots, x_n^*)$ for n agents.

$$x_i^* \in \arg\max_{x_i} u_i(x_i, \boldsymbol{x}_{-i}^*), \quad \forall i \in [n].$$

No changes in an agent's strategy x_i will lead to any gains, given the strategy of other agents $\boldsymbol{x}_{-i}^*$ fixed.

- Given any strategy profile $\boldsymbol{x}$, loss function $f : \mathcal{X} \to \mathbb{R}_{\geq 0}$ is defined as:

$$f(\boldsymbol{x}) = \sum_{i \in [n]} \max_{x_i' \in \mathcal{X}_i} u_i(x_i', \boldsymbol{x}_{-i}) - u_i(\boldsymbol{x}) \tag{1}$$

The sum of all agents' gains from deviating from the given strategy; $f(\boldsymbol{x}^*) = 0$.

- Goal: to minimize the unknown objective function Equation (1).

Learning via Gaussian Process

Figure 2: Learning unknown utility function with Gaussian process

- For agent $i \in [n]$, the utility function u_i is modeled as a Gaussian Process (GP).
- The utility function $u_i : \mathcal{X} \to [0, 1]$ is represented as $u_i(\boldsymbol{x}) \sim \mathcal{GP}(\mu_{u_i}(\cdot), k_{u_i}(\cdot, \cdot))$.
- Given a history of observations $\mathcal{D}^{1:t}$, the posterior distribution under the $\mathcal{GP}(0, k(\boldsymbol{x}, \boldsymbol{x}'))$ prior remains Gaussian.
- The mean function is updated as: $\mu_{u_i,t}(\boldsymbol{x}) = \mathbf{k}_{u_i}^t(\boldsymbol{x})^\top(\mathbf{K}_{u_i}^t + \sigma^2\mathbf{I})^{-1}\boldsymbol{y}_i^{1:t}$
- The variance function is updated as:

$$\sigma_{u_i,t}(\boldsymbol{x})^2 = k_{u_i}(\boldsymbol{x}, \boldsymbol{x}) - \mathbf{k}_{u_i}^t(\boldsymbol{x})^\top(\mathbf{K}_{u_i}^t + \sigma^2\mathbf{I})^{-1}\mathbf{k}_{u_i}^t(\boldsymbol{x})$$

Approximation of the Partial Maximum

We approximate the unknown $v_i(\boldsymbol{x}_{-i}) \triangleq \max_{x_i'} u_i(x_i', \boldsymbol{x}_{-i})$ with its corresponding upper confidence bound (UCB) and lower confidence bound (LCB) derived from the marginalized $\mathcal{GP}_{u_i|\boldsymbol{x}_{-i}}$.

- **Confidence Bounds**:
  - The upper confidence bound(UCB) and lower confidence bound(LCB) are defined as:

$$\mathrm{UCB}_{v_i,t}(\boldsymbol{x}_{-i}, \mathcal{S}) \triangleq \max_{x_i':(x_i',\boldsymbol{x}_{-i})\in\mathcal{S}} \mu_{u_i,t-1}(x_i', \boldsymbol{x}_{-i}) + \beta^{1/2}\sigma_{u_i,t-1}(x_i', \boldsymbol{x}_{-i}), \tag{2}$$

$$\mathrm{LCB}_{v_i,t}(\boldsymbol{x}_{-i}, \mathcal{S}) \triangleq \max_{x_i':(x_i',\boldsymbol{x}_{-i})\in\mathcal{S}} \mu_{u_i,t-1}(x_i', \boldsymbol{x}_{-i}) - \beta^{1/2}\sigma_{u_i,t-1}(x_i', \boldsymbol{x}_{-i}), \tag{3}$$

  - β controls confidence level. A higher β value increases the confidence level, making the bounds wider.
  - The UCB and LCB provide a high confidence bound for v_i. The width of this bound becomes more accurate after a certain number of iterations.
- **Domain of the Marginal Maximum**:
  - $\mathcal{S}$ denotes the domain over which the marginal maximum is taken.
  - This ensures that the bounds are computed considering the relevant subset of the input space.

Adaptive Level-set Estimation for Global Optimization

Example 1. *A two-player game from [ADHO18; Par+08] as a running example, where the utility functions of the two players are defined as* $u_1(x_1, x_2) = (x_2 - x_2^*)^2 - (x_1 - x_1^*)^2$ *and* $u_2(x_1, x_2) = (x_1 - x_1^*)^2 - (x_2 - x_2^*)^2$. $\boldsymbol{x}^* = (x_1^*, x_2^*) = (0.5, 0.5)$ *denotes the NE.*

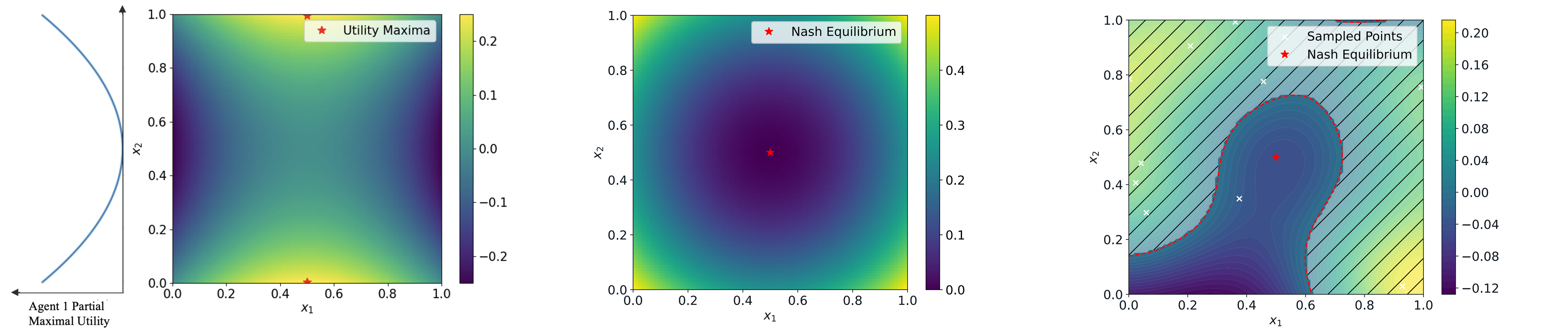

Figure 3: Function visualizations of Example 1, where x-axis represents agent 1's action and y-axis represents agent 2's action.

The UCB and LCB for f are $\mathrm{UCB}_{f,t}(\boldsymbol{x}, \mathcal{S}) \triangleq \sum_{i\in[n]} \mathrm{UCB}_{v_i,t}(\boldsymbol{x}_{-i}, \mathcal{S}) - \mathrm{LCB}_{u_i,t}(\boldsymbol{x})$, $\mathrm{LCB}_{f,t}(\boldsymbol{x}, \mathcal{S}) \triangleq \sum_{i\in[n]} \mathrm{LCB}_{v_i,t}(\boldsymbol{x}_{-i}, \mathcal{S}) - \mathrm{UCB}_{u_i,t}(\boldsymbol{x})$. We define the filtering threshold as $\mathrm{UCB}_{f,t,\min} \triangleq \min_{\boldsymbol{x}\in\mathcal{X}} \mathrm{UCB}_{f,t}(\boldsymbol{x})$. The sublevel-set

$$\hat{\mathcal{X}}^t \triangleq \{\boldsymbol{x} \in \mathcal{X} \mid \mathrm{LCB}_{f,t}(\boldsymbol{x}) \leq \min(\mathrm{UCB}_{f,t,\min}, 0)\}$$

serve as the region of interest for global optimization.

Bayesian Optimization with Adaptive Level-Set Estimation

Algorithm 1 Adaptive Region of Interest Search for Nash Equilibrium (ARISE)

1: **Input**:Search space $\mathcal{X}$, initial observation $\mathcal{D}^0$, horizon T;
2: **for** $t = 1$ *to* T **do**
3: Fit the Gaussian processes $\mathcal{GP}_{u_i,t}$: $\theta_{u_i,t} \leftarrow \arg\min_{\theta_{u_i}} -\log \mathbb{P}\left[y_i^{1:t-1} \mid \boldsymbol{x}^{1:t-1}, \theta_{u_i}\right]$
4: Identify ROIs via sublevel-set estimation $\hat{\mathcal{X}}^t \leftarrow \{\boldsymbol{x} \in \mathcal{X} \mid \mathrm{LCB}_{f,t}(\boldsymbol{x}) \leq 0\}$
5: Optimize the sublevel-set acquisition function: $\boldsymbol{x}^t \leftarrow \arg\max_{\boldsymbol{x}\in\hat{\mathcal{X}}^t} \alpha_{f,t}(\boldsymbol{x}, \hat{\mathcal{X}}^t)$
6: $\mathcal{D}^{1:t} \leftarrow \mathcal{D}^{1:t-1} \cup \{(\boldsymbol{x}^t, \boldsymbol{y}^t)\}$
7: **end for**
8: **Output**: $\arg\min_{\boldsymbol{x}\in\hat{\mathcal{X}}^T} \mathrm{LCB}_{f,T}(\boldsymbol{x})$

Key Theoretical Results

Here, we justify that the proposed acquisition function efficiently reduces the width of the confidence interval of the global optimum.

Theorem 1. *The width of the resulting confidence interval of the global optimum $f^* = f(\boldsymbol{x}^*)$ has an upper bound. That is, under the assumptions above, with a constant $\beta = 2\log(n|\tilde{S}|T/\delta)$, and $\boldsymbol{x}^t = \arg\max_{\boldsymbol{x}\in\mathcal{X}} \alpha_{f,t}(\boldsymbol{x}, \mathcal{X})$, after at most $T \geq \frac{\beta\widehat{\gamma}_T\hat{C}_1}{\epsilon^2}$ iterations, we have*

$$\mathbb{P}\left[|CI_{f^*,T}| \leq \epsilon, f^* \in CI_{f^*,T}\right] \geq 1 - \delta$$

Assuming that the Nash-Equilibrium exists, and the points of ROI are sufficiently close to $\boldsymbol{x}^*$, we have with probability at least $1 - \delta$ that ARISE achieves ϵ-Nash Equilibrium.

Theorem 2. *We assume the aforementioned assumptions hold. We apply the same β and the acquisition function as illustrated in figure 4. In addition, we assume after $T \geq \frac{\beta\widehat{\gamma}_T\hat{C}_1}{\epsilon^2}$ iterations, when $\forall \boldsymbol{x} \in \tilde{S}_{\hat{\mathcal{X}}^t}$, it holds that $UCB_{u_i,t}(\boldsymbol{x}_{-i}, \tilde{S}_{\hat{\mathcal{X}}^t}) = UCB_{u_i,t}(\boldsymbol{x}_{-i}, \tilde{S})$ and $LCB_{u_i,t}(\boldsymbol{x}_{-i}, \tilde{S}_{\hat{\mathcal{X}}^t}) = LCB_{u_i,t}(\boldsymbol{x}_{-i}, \tilde{S})$, we have*

$$\mathbb{P}\left[f(\boldsymbol{x}^T) \leq \sqrt{\frac{\beta\widehat{\gamma}_T\hat{C}_1}{T}} \leq \epsilon\right] \geq 1 - \delta$$

Here $\hat{C}_1$. Previous work by [Sri+09] bounds the maximum information gain γ to be sublinear to T.

Experimental Results

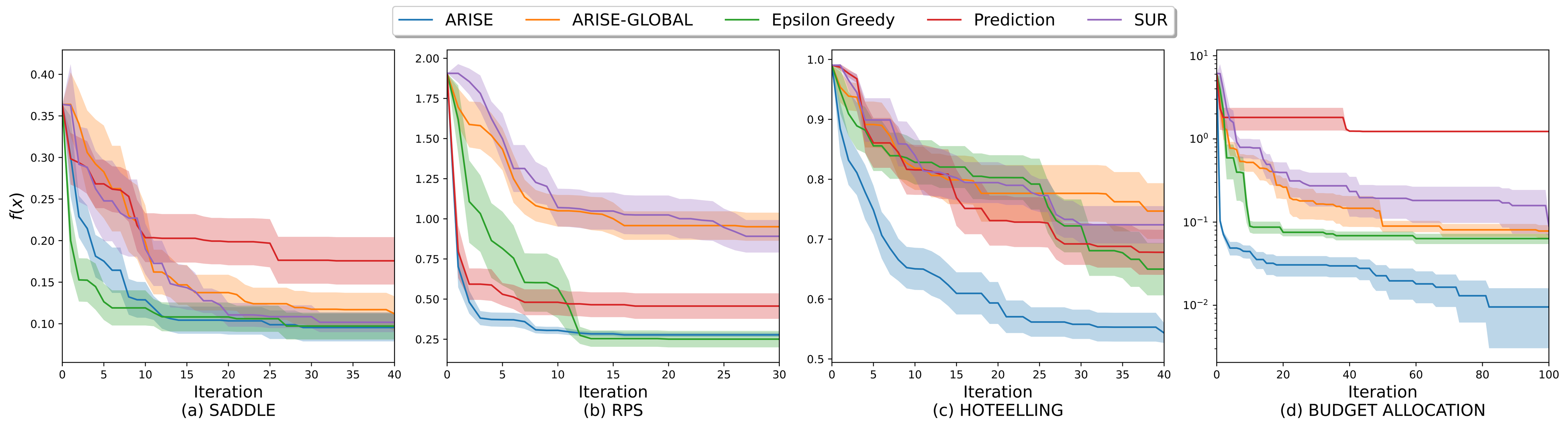

Figure 4: In each plot, the x-axis denotes the number of function evaluations. The curves show the $f(\boldsymbol{x}^t)$ values averaged over at least ten independent trials. Shaded area denotes the standard error.

- **Saddle:** The running example presented in Example 1 [ADHO18; PBH19].
- **RPS:** The Rock-Paper-Scissors game, with Nash Equilibrium at equal probabilities for all options.
- **Hotelling:** Firms choose locations on a 2D grid to attract customers, balancing proximity and competition [Bre05].
- **Budget Allocation:** Advertisers allocate budgets to media channels to maximize customer reach, using a bipartite graph model with activation probabilities [MYK15].

Additional Resources

Scan For Paper

Scan For Code