Large-Scale Multi-Agent Deep FBSDEs

Tianrong Chen[1] Ziyi Wang[2] Ioannis Exarchos[3]
Evangelos A. Theodorou[4]

[1]Electical and Computer Engineering
Georgia Institute of Technology

[2]Center for Machine Learning
Georgia Institute of Technology

[3]Department of Computer Science
Stanford University

[4]School of Aerospace Engineering
Georgia Institute of Technology

Introduction

- Stochastic Differential Games (SDG) represent a framework for investigating scenarios where multiple players [1] make dicisions in a stochastic environment.

[1] Agent and player are used interchangeably in this paper

Introduction

- Stochastic Differential Games (SDG) represent a framework for investigating scenarios where multiple players [1] make dicisions in a stochastic environment.
- The shared environment is governed by a Stochastic Differential Equation (SDE), meanwhile the rationality of each player is characterirzed by Hamilton-Jacobi-Bellman (HJB) equation.

[1] Agent and player are used interchangeably in this paper

Mathematical Notations

Table: Mathematical notations.

CHARACTERS	DEFINITIONS
$\boldsymbol{X}$	quantities for all agents
X^i/X_i	quantities for ith agent
$\boldsymbol{X}_{-i}$	quantities from all agents except ith
$\boldsymbol{x}$	realization of $\boldsymbol{X}$

$\boldsymbol{X}$ represents for state and $\boldsymbol{U}$ represent for controls by default.

Problem Formulation

We consider a N-player non-cooperative SDG with dynamics:

$$\begin{aligned} \mathrm{d}\boldsymbol{X}_t &= \big(f(\boldsymbol{X}_t, t) + G(\boldsymbol{X}_t, t)\boldsymbol{U}(\boldsymbol{X}_t)\big)\mathrm{d}t + \Sigma(\boldsymbol{X}_t, t)\mathrm{d}\boldsymbol{W}_t \\ \boldsymbol{X}_{t_0} &= \boldsymbol{x}_{t_0}. \end{aligned} \tag{1}$$

The stochastic optimal control problem for agent i is defined as minimizing the expectation of the cumulative cost functional J_t^i:

$$\begin{aligned} & J_t^i(\boldsymbol{X}, U_{i,m}; \boldsymbol{U}_{-i,m-1}) = \\ & \mathbb{E}\left[g(\boldsymbol{X}_T) + \int_t^T C^i(\boldsymbol{X}_\tau, U_{i,m}(\boldsymbol{X}_\tau); \boldsymbol{U}_{-i,m-1})\mathrm{d}\tau\right], \end{aligned} \tag{2}$$

Individual Rationality – HJB

Based on the cumulative cost functional (2), one can write the HJB function for individual player as:

$$V_t^i + \inf_{u^i \in \mathcal{U}_i} \left\{ V_x^{i\mathrm{T}}(f + G\boldsymbol{U}) + C^i \right\} + \frac{1}{2}\mathrm{tr}(V_{xx}^i \Sigma\Sigma^{\mathrm{T}}) = 0 \tag{3}$$

If the optimal control is accessible, one can rewrite the HJB equation (3) by plugging in the optimal control:

$$V_t^i + h + V_x^{i\mathrm{T}}(f + G\boldsymbol{U}_{0,*}) + \frac{1}{2}\mathrm{tr}(V_{xx}^i \Sigma\Sigma^{\mathrm{T}}) = 0, \tag{4}$$

where $h^i = C^{i*} + G\boldsymbol{U}_{*,0}$.The $*$ denotes the optimal control,and the 'zero' represents for taking zero control $\boldsymbol{U}_{-i} = 0$. For instance,

$$\boldsymbol{U}_{0,*} = (U_1^*, \cdots, U_{i-1}^*, 0, U_{i+1}^*, \cdots, U_N^*).$$

Non-linear Feynman-Kac Lemma

A non-linear PDE (eq.4) can be related to a set of Forward SDE (FSDE) and Backward SDE (BSDE) via Non-linear Feynman-Kac Lemma (Karatzas & Shreve 1991):

$$\begin{aligned} \mathrm{d}\boldsymbol{X}_t &= (f + G\boldsymbol{U}_{0,*})\mathrm{d}t + \Sigma \mathrm{d}\boldsymbol{W}_t,\ \boldsymbol{X}_{t_0} = \boldsymbol{x}_{t_0} \quad &\text{(FSDE)},\\ \mathrm{d}Y_t^i &= -h_t^i \mathrm{d}t + Z_t^i \mathrm{d}W_t,\ Y_T^i = g(\boldsymbol{X}_T), \quad &\text{(BSDE)}, \end{aligned} \tag{5}$$

where the backward process Y_t corresponds to the value function $V(\boldsymbol{x}, t)$.

Remark 1

The problem of solving HJB PDE (4) will be transformed to solve a FBSDE system. (eq.5). The *i*th player will provide zero control in the drift term in the forward process, and the rest of agents will execute the optimal policy according to the Value function.

Deep Fictitious Play FBSDEs Framework

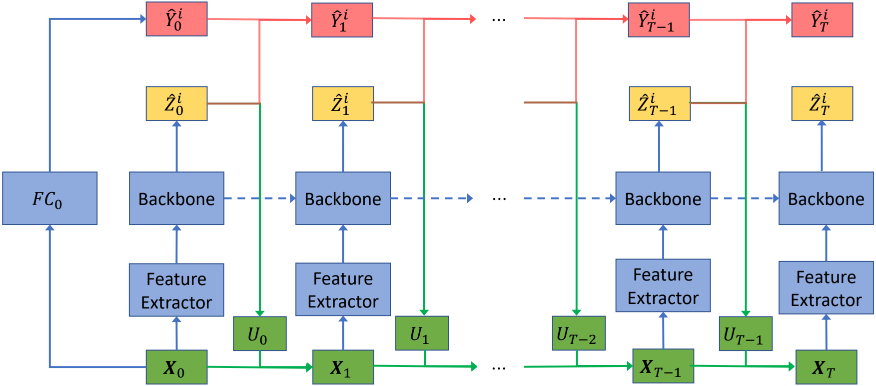

Figure: FBSDE Network for a single agent. The dashed arrow indicates hidden states propagation if LSTM is chosen as backbone. The dash arrow would disappear when FC is chosen.

Training Procedures

The training procedures of this framework follow the typical approaches in deep learning community.

ITEMS	DEFINITIONS
Training Data	random generated $\boldsymbol{x}_0$
Label	terminal cost $Y^* = \boldsymbol{g}(\boldsymbol{x}_T)$
Training Loss Function	$\|\|\hat{\boldsymbol{Y}}^T - \boldsymbol{Y}^*\|\|_2^2$
Trainable Parameters	parameters of backbones and FCs

Importance Sampling (IS) in FBSDEs

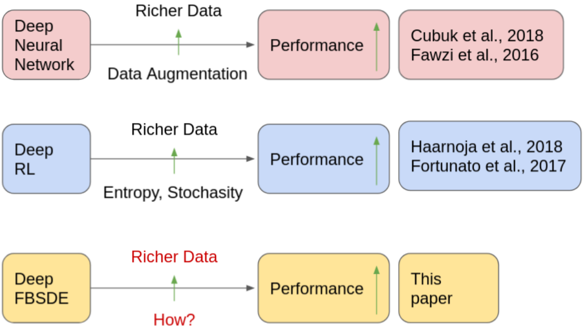

Figure: Training Deep FBSDE can be similar to regular deep learning model. However a proper way to obtain richer data (Y components) while still solving same HJB-PDE is not clear.

Importance Sampling (IS) in FBSDEs

Assumption 1

There exists a measurable function $\phi : [0, T] \times \mathcal{X} \rightarrow \mathcal{X}$ and $\Gamma : [0, T] \times \mathcal{U} \rightarrow \mathcal{X}$ so that $\Sigma(t, \boldsymbol{X})\phi(t, \boldsymbol{X}) = f(t, \boldsymbol{X})$, and $\Sigma(t, \boldsymbol{X})\Gamma(t, \boldsymbol{X}) = G(t, \boldsymbol{X})$.

Assumption 2

For a general FBSDE system,

$$\begin{aligned} \boldsymbol{X}_T^{t,\boldsymbol{x}} &= \boldsymbol{x} + \int_t^T \mu_s \mathrm{d}s + \int_t^T \Sigma_s \mathrm{d}\boldsymbol{W}_s && \text{(FSDE)}, \\ Y_t^{T,\boldsymbol{x}} &= g(\boldsymbol{X}_T^{t,\boldsymbol{x}}) - \int_t^T H_s \mathrm{d}s + \int_t^T Z_s \mathrm{d}\boldsymbol{W}_s && \text{(BSDE)}, \end{aligned} \tag{6}$$

Functions μ_s, H_s,Σ,$g(\cdot)$, and $U(\cdot)$ satisfy Lipschitz continuous properties with Lipschitz constants $\mu_x, \mu_u, \Sigma_x, H_x, H_z, g_x, u_x$ respectively. The detailed and formal description can be found in Appendix.

Importance Sampling (IS) in FBSDEs

We first analyze the FBSDE without IS.

Definition

Denote $(\boldsymbol{X}_s^{t,\boldsymbol{x}}, Y_s^{t,\boldsymbol{x}}, Z_s^{t,\boldsymbol{x}})_{t\leq s\leq T}$ as the solution for the FBSDE system (6) satisfying assumptions 1 and 2. Denote the difference of components at two different states $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ as:

$$\begin{aligned} &\delta \boldsymbol{X}_t = \boldsymbol{X}_t^{t_0,\boldsymbol{x}_1} - \boldsymbol{X}_t^{t_0,\boldsymbol{x}_2}, \delta Y_t = Y_t^{t_0,\boldsymbol{x}_1} - Y_t^{t_0,\boldsymbol{x}_2}. \\ &\delta Z_t = Z_t^{t_0,\boldsymbol{x}_1} - Z_t^{t_0,\boldsymbol{x}_2} \end{aligned} \tag{7}$$

Importance sampling in FBSDEs

Lemma 1

$$|\delta Y_T|^2 \leq L_1|\boldsymbol{x}_1 - \boldsymbol{x}_2|^2, |\delta Y_{t_0}|^2 \leq L_2|\boldsymbol{x}_1 - \boldsymbol{x}_2|^2, \tag{8}$$

Where L_1 and L_2 are defined as:

$$\begin{aligned} L_1 &= g_x e^{\xi T} \\ L_2 &= e^{H_z(T-t_0)}\left[g_x e^{\xi(T-t_0)} + H_x \frac{e^{\xi(T-t_0)} - 1}{\Sigma_x^{-1} H_z^{-1}}\right], \\ \xi &= I + \mu_x + \mu_u u_x + \Sigma_x, \end{aligned} \tag{9}$$

Following arguments in (Ma et al.,2002), one further has,

$$||Z_t||_S^2 \leq ||\Sigma||_S^2 ||\nabla_x Y_t||_S^2 \leq M_\Sigma L_2 \tag{10}$$

Where $\mu_x, \mu_u, u_x, \Sigma_x, H_x, H_z, g_x$ are Lipschitz constants defined in Assumptions. M_Σ is the upper bound of Σ.

Importance sampling in FBSDEs

- Lemma 1 bridges the connection between the states and their corresponding value functions.
- In the next slide, we will state the definition of Imporatance Samping.

Importance sampling in FBSDEs

Theorem 1 (Bender & Moseler, 2010)

Let $(X_s^{i,t,x}, Y_s^{i,t,x}, Z_s^{i,s,x})$ be the solution of the FBSDE system (5) for ith agent, and let $K_s : [0,T] \times \Omega \to \mathbb{R}^{n_x}$ be any bounded and square integrable process for ith agent. Consider the forward process whose drift term is modified by K_s

$$\mathrm{d}\tilde{\boldsymbol{X}}_s = [\mu_s + \Sigma K_s]\mathrm{d}s + \Sigma \mathrm{d}\boldsymbol{W}_s,\ \tilde{\boldsymbol{X}}_t = \boldsymbol{x}_t, \tag{11}$$

along with the corresponding BSDE

$$\mathrm{d}\tilde{Y}_s^i = [-h_s^i + \tilde{Z}_s K_s]\mathrm{d}s + \tilde{Z}_s^i \mathrm{d}W_s,\ Y_T^i = g(\boldsymbol{X}_T). \tag{12}$$

Here we denote $(\tilde{X}_s^{i,t,x}, \tilde{Y}_s^{i,t,x}, \tilde{Z}_s^{i,s,x})$ as the solution for modified FBSDE system (11,12). For all $s \in [t,T]$,
$(\tilde{X}_s^{i,t,x}, \tilde{Y}_s^{i,t,x}, \tilde{Z}_s^{i,s,x}) = (X_s^{i,t,x}, Y_s^{i,t,x}, Z_s^{i,s,x})$ a.s. If $(\tilde{Y}_s^{i,t,x}, \tilde{Z}_s^{i,s,x})$ are defined as $(\tilde{V}^i, \Sigma^{\mathrm{T}} \tilde{V}_x^i)$ with $\tilde{V}^i$ being the solution to 4, then $V^i \equiv \tilde{V}^i$ a.e.

Importance sampling in FBSDEs

Theorem 2

Denote $(\tilde{\boldsymbol{X}}_s^{t,\boldsymbol{x}}, \tilde{Y}_s^{t,\boldsymbol{x}}, \tilde{Z}_s^{t,\boldsymbol{x}})_{t\leq s\leq T}$ is the solution for the FBSDE system with IS (11,12), and $(\boldsymbol{X}_s^{t,\boldsymbol{x}}, Y_s^{t,\boldsymbol{x}}, Z_s^{t,\boldsymbol{x}})_{t\leq s\leq T}$ is the solution for the FBSDE system (6). and they satisfy the assumption 1 and 2. Then given the identical training data $\boldsymbol{X}_0$ for FBSDE w/ and w/o IS, one can have,

$$\begin{aligned} \max|\delta Y_T|^2 &\leq \max|\delta \tilde{Y}_T|^2, \\ \max|\delta Y_0|^2 &\leq \max|\delta \tilde{Y}_0|^2 \end{aligned} \tag{13}$$

Theorem 1 and Theorem 2 show that, Importance sampling can provide richer region of training target while still solve the same HJB-PDE.

Importance sampling in FBSDEs

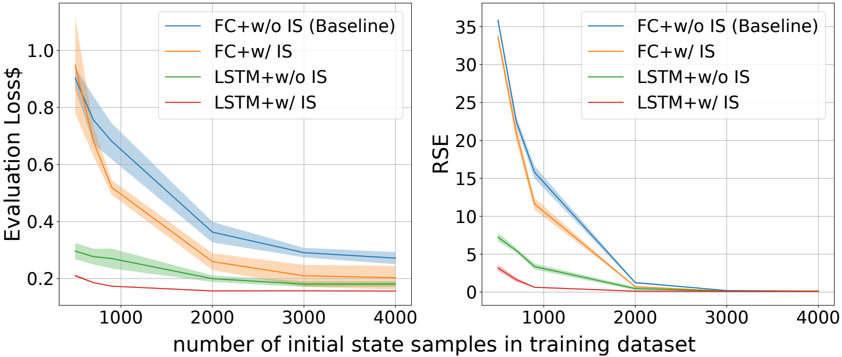

Figure: Performance difference of DFP-FBSDE w/ and w/o importance samping over limited training dataset. The simulation is executed on 100 agents inter-bank game.

Importance sampling in FBSDEs

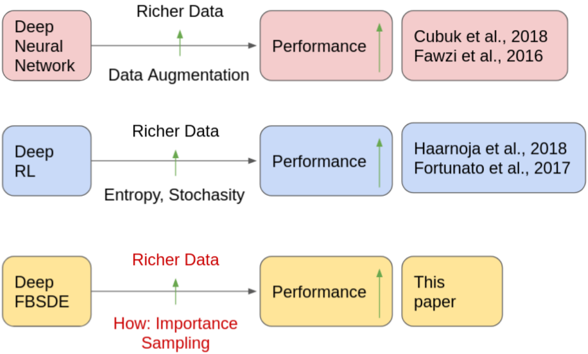

Figure: According to the theoratical and experimental result, Importance Sampling is included in our framework.

Invariant Layers

- We incoperate the Invariant Layers (Zaheer et al.,2017) to extract the permutation invariant features.

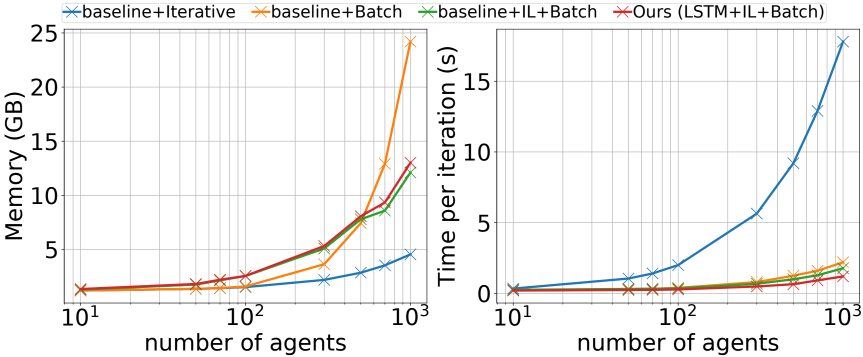

Figure: Time and memory complexity comparison between batch, iterate and IL+batch implementations.

Invariant Layers

- Additionally, it helps to mitigate the curse of many agents when the number of agent is increasing.

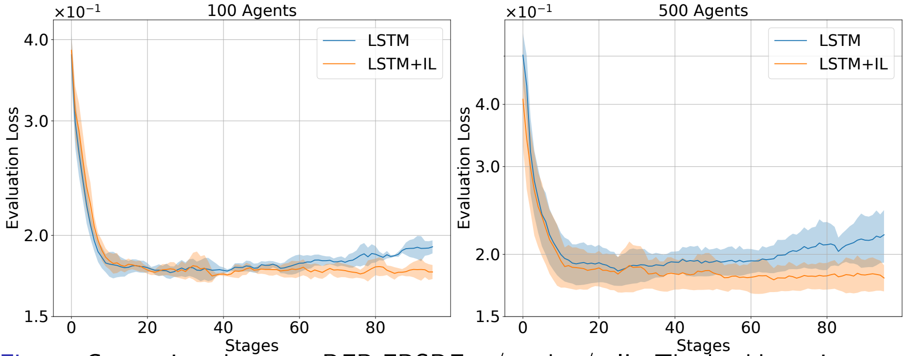

Figure: Comparison between DFP-FBSDE w/ and w/o IL. The backbone is chosen as LSTM. The simulation is inter-bank game.

Experiments: Inter-Bank Game

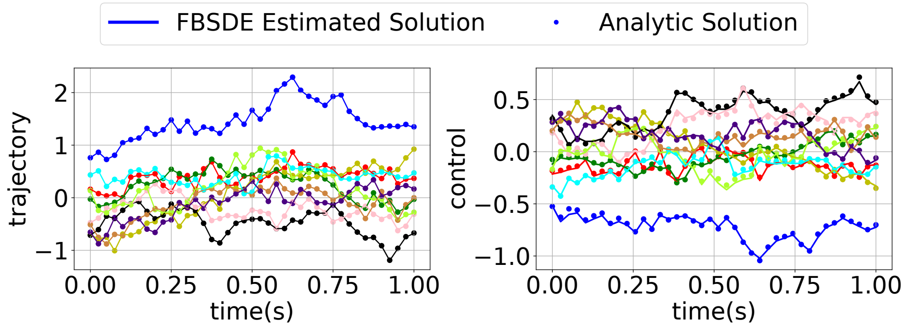

Figure: Comparison of SDFP and analytical solution for the inter-bank problem. Both the state (*left*) and control (*right*) trajectories are aligned with the analytical solution (represented by dots).

Experiments: Inter-Bank Game

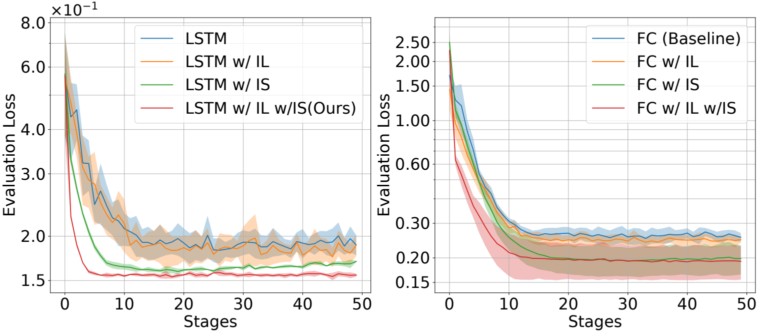

Figure: Ablation experiments on LSTM and FC backbone.

Experiments:Inter-Bank game

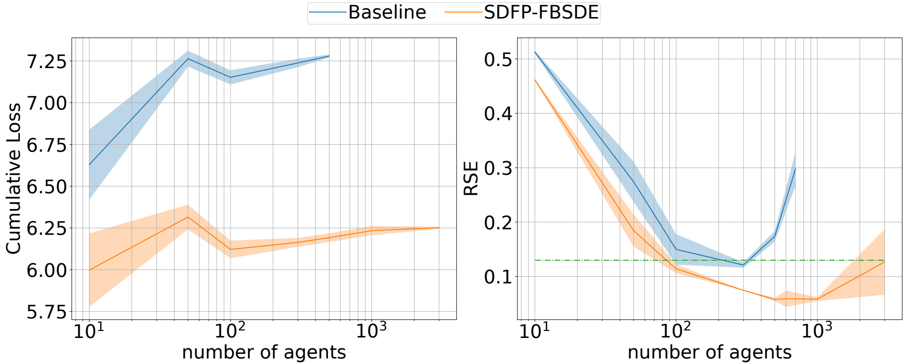

Figure: Comparison of SDFP-FBSDE and Baseline for inter-bank problem with different number of agents evaluated on cumulative loss and RSE.

Experiments:Partial Observed Racing-Car

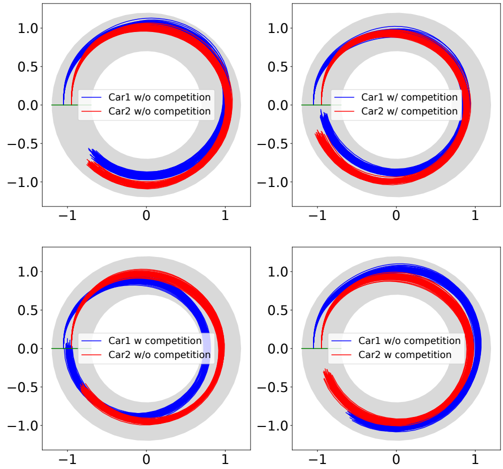

Figure: The plots contains 64 trials of racing. The performance varies with respect to the competition loss.

Conclusion

- In this paper, we first extend the theoretical analysis from (Han & Long 2020) and introduce importance sampling to improve the sample efficiency and convergence rate.
- To further push our work to handle larger number of agents with appreciable time and memory complexity, batch query scheme and invariant layer implementation are proposed.
- Our framework achieves better performance in different metrics and scales to significantly higher dimensions.
- The general applicability of our framework is showcased on a belief space racing problem in the partially observed scenario.

Thank you!