Mean field games: Large number of small, anonymous agents with negligible individual impact

### Stationary MF equilibrium (SMFE)

A stationary mean-field equilibrium (SMFE) is a pair of policy \( \pi \in \Pi \) and mean-field \( z \in \Delta(\mathcal{X}) \) which satisfies the following two properties:

1. **Sequential rationality**: For any other policy \( \pi' \), \( V_{\pi'}(x) \geq V_{\pi}(x) \), \( \forall x \in \mathcal{X} \).
2. **Consistency**: The mean-field \( z \) is stationary under policy \( \pi \), i.e., \( z = \Phi(z, \pi) \).

### Local SMFE (LSMF)

A local stationary mean-field equilibrium (LSME) is a pair of policy \( \pi_0 \in \Pi \) and mean-field \( z \in \Delta(\mathcal{X}) \) which satisfies the following two properties:

1. **Local sequential rationality**: \( \partial_z V_{\pi_0}/\partial \theta = 0 \).
2. **Consistency**: \( z = \Phi(z, \pi_0) \).

**RL algorithm for learning LSMFE**

Suppose \( G_{02} \) is an unbiased estimator of \( \partial_{z_{\pi_0}}/\partial \theta \). Then, we start with an initial guess \( \theta_0 \in \Theta \) and \( z_0 \in \Delta(\mathcal{X}) \) and at each step of the iteration, update the guess \( \{z_t, \theta_t\} \) using two-timescale stochastic gradient ascent:

\[
\begin{align*}
    z_{t+1} &= z_t + \theta_t [\Phi(z_t, \pi_0) - z_t] ; \\
    \theta_{t+1} &= \{ \theta_t + \alpha_t G_{\theta_{t+1}} \theta \}
\end{align*}
\]

where \( z_0 = \Phi(z_0, \pi_0) \), \( \alpha_t \), and \( \beta_t \) are chosen s.t.

\[
\begin{align*}
    \sum_{t=0}^{\infty} \alpha_t &= \infty, \quad \sum_{t=0}^{\infty} \beta_t < \infty, \quad \lim \alpha_t = 0, \quad \lim \beta_t = 0, \quad \lim \alpha_t / \beta_t = 0.
\end{align*}
\]

**Local MF-SO (LSMF-SO)**

A policy \( \pi_0 \in \Pi \) is local stationary mean-field social-welfare optimal (LSMF-SO) if it satisfies the following property:

- **Locality**: For any other policy \( \pi' \in \Pi \), \( V_{\pi_0}(x) \geq V_{\pi'}(x) \), \( \forall x \in \mathcal{X} \), where \( z \) and \( z' \) are the stationary mean-field distributions: \( z = \Phi(z, \pi_0) \) and \( z' = \Phi(z', \pi_0) \).

**RL algorithm for learning LSMF-SO**

Suppose \( T_0 \) is an unbiased estimator of \( \partial_{z_{\pi_0}}/\partial \theta \), where \( z_0 \) is the fixed point of \( z = \Phi(z, \pi_0) \). Then, we start with an initial guess \( \theta_0 \in \Theta \) and at each step of the iteration, update the guess using stochastic gradient ascent:

\[
\begin{align*}
    \theta_{t+1} &= \{ \theta_t + \alpha_t G_{\theta_{t+1}} \theta \}
\end{align*}
\]

**Solution concept**

- Mean-field equilibrium—competitive agents.
- Mean-field social-welfare optimal policy—cooperative agents.
- Extension to stationary mean-field games:
  - Stationary mean-field equilibrium (SMFE)
  - Stationary social-welfare optimal policy (SMF-SO)

**Our contribution**

- Generalization of these solution concepts to their local variants using bounded rationality based arguments.
- Development of policy gradient based reinforcement learning algorithms to predict these solution concepts.
- Proof of convergence of these algorithms to the right solution concept under mild technical conditions.

---

**Numerical examples**

Fig. 3: Malware spread

- LSMFE [52%]:
  - LSMF-SO Threshold
  - LSMF-SO

Fig. 4: Product investments

- LSMFE [42%]:
  - LSMF-SO Threshold
  - LSMF-SO [42%]