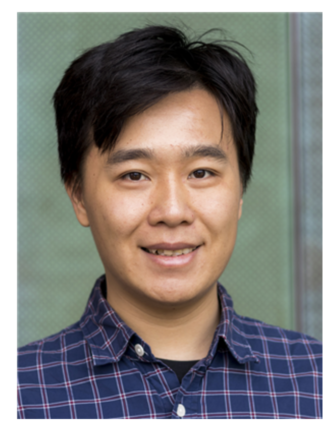
Batch Value-Function Approximation with Only Realizability



Tengyang Xie, Nan Jiang
University of Illinois at Urbana-Champaign

Value-based RL in Large State Spaces

Simple(?) Problem

- Given two Q-functions f_1, f_2, one of which is Q^*
- Can we identify Q^* from a "small" exploratory dataset of (s, a, r, s')?
  ("small" = no |S| or exponential-in-horizon dependence)

Value-based RL in Large State Spaces

Simple(?) Problem

- Given two Q-functions f_1, f_2, one of which is Q^*
- Can we identify Q^* from a "small" exploratory dataset of (s, a, r, s')? ("small" = no |S| or exponential-in-horizon dependence)

What was known

- Almost nothing, except hardness conjecture [CJ'19]

Value-based RL in Large State Spaces

Simple(?) Problem

- Given two Q-functions f_1, f_2, one of which is Q^*
- Can we identify Q^* from a "small" exploratory dataset of (s, a, r, s')? ("small" = no |S| or exponential-in-horizon dependence)

What was known

- Almost nothing, except hardness conjecture [CJ'19]

Importance

- Hyperparamter tuning for offline RL

Value-based RL in Large State Spaces

Simple(?) Problem

- Given two Q-functions f_1, f_2, one of which is Q^*
- Can we identify Q^* from a "small" exploratory dataset of (s, a, r, s')?
  ("small" = no |S| or exponential-in-horizon dependence)

What was known

- Almost nothing, except
  hardness conjecture [CJ'19]

Importance

- Hyperparamter tuning for offline RL
- Theoretical foundation for training
  - Is *realizability* alone sufficient for training?

Value-based RL in Large State Spaces

Simple(?) Problem

- Given two Q-functions f_1, f_2, one of which is Q^*
- Can we identify Q^* from a "small" exploratory dataset of (s, a, r, s')? ("small" = no |S| or exponential-in-horizon dependence)

What w

- Almo
  harc

Our contributions

- A surprising *positive* result: BVFT

Importance

- Hyperparamter tuning for offline RL
- Theoretical foundation for training
  - Is *realizability* alone sufficient for training?

Value-based RL in Large State Spaces

Simple(?) Problem

- Given two Q-functions f_1, f_2, one of which is Q^*
- Can we identify Q^* from a "small" exploratory dataset of (s, a, r, s')? ("small" = no |S| or exponential-in-horizon dependence)

What v

- Almo
  harc

Our contributions

- A surprising *positive* result: BVFT
- Handles *exponentially large* function space F with misspecification

Importance

- Hyperparamter tuning for offline RL
- Theoretical foundation for training
  - Is *realizability* alone sufficient for training?

Value-based RL in Large State Spaces

Simple(?) Problem

- Given two Q-functions f_1, f_2, one of which is Q^*
- Can we identify Q^* from a "small" exploratory dataset of (s, a, r, s')?
  ("small" = no |S| or exponential-in-horizon dependence)

What w

- Almo
  harc

Our contributions

- A surprising *positive* result: BVFT
- Handles *exponentially large* function space F with misspecification
- Potential application to hyperparameter tuning

Import

- Hyperparamter tuning for offline RL
- Theoretical foundation for training
  - Is *realizability* alone sufficient for training?

Why the problem seemed impossible

Simple(?) Problem

- Given two Q-functions f_1, f_2, one of which is Q^*
- Can we identify Q^* from a "small" exploratory dataset of (s, a, r, s')? ("small" = no |S| or exponential-in-horizon dependence)

Attempt 1: Off-policy Evaluation (OPE)

- Induce two greedy policies and evaluate them

Why the problem seemed impossible

Simple(?) Problem

- Given two Q-functions f_1, f_2, one of which is Q^*
- Can we identify Q^* from a "small" exploratory dataset of (s, a, r, s')? ("small" = no |S| or exponential-in-horizon dependence)

Attempt 1: Off-policy Evaluation (OPE)

- Induce two greedy policies and evaluate them
- Problem: OPE itself is a hard problem—importance sampling incurs *exponential-in-horizon* variance, and other methods (e.g., FQE/MIS) requires additional function approximation

Why the problem seemed impossible

Simple(?) Problem

- Given two Q-functions f_1, f_2, one of which is Q^*
- Can we identify Q^* from a "small" exploratory dataset of (s, a, r, s')?
  ("small" = no |S| or exponential-in-horizon dependence)

Attempt 2: Estimating Bellman Error

- $f = Q^\star \Leftrightarrow \|f - \mathcal{T} f\| = 0$, so try to estimate $\|f - \mathcal{T} f\|$?

Why the problem seemed impossible

Simple(?) Problem

- Given two Q-functions f_1, f_2, one of which is Q^*
- Can we identify Q^* from a "small" exploratory dataset of (s, a, r, s')? ("small" = no |S| or exponential-in-horizon dependence)

Attempt 2: Estimating Bellman Error

- $f = Q^\star \Leftrightarrow \|f - \mathcal{T}f\| = 0$, so try to estimate $\|f - \mathcal{T}f\|$?
- Problem: cannot be estimated in stochastic environments!
- The infamous *double-sampling difficulty:* the only natural estimator $\left(f(s,a) - (r + \gamma \max_{a'} f(s', a'))\right)^2$ is positively biased

New Algorithm: Batch Value-function Tournament (BVFT)

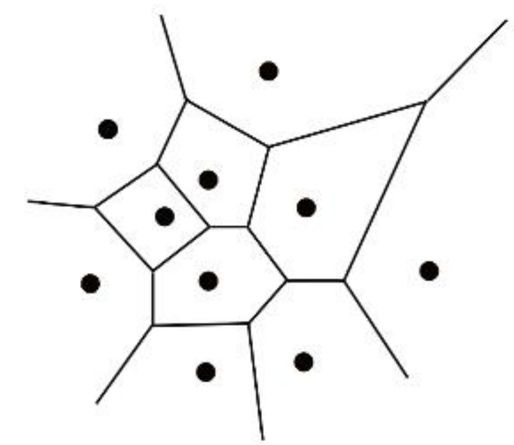
- $f = Q^\star \Leftrightarrow \|f - \mathcal{T}_\mathcal{G} f\| = 0$, where $\mathcal{T}_\mathcal{G} f$ is $\mathcal{T} f$ projected onto a function space G satisfying [Gordon'96]

New Algorithm: Batch Value-function Tournament (BVFT)

- $f = Q^\star \Leftrightarrow \|f - \mathcal{T}_\mathcal{G} f\| = 0$, where $\mathcal{T}_\mathcal{G} f$ is $\mathcal{T} f$ projected onto a function space G satisfying [Gordon'96]
  - $Q^\star \in \mathcal{G}$

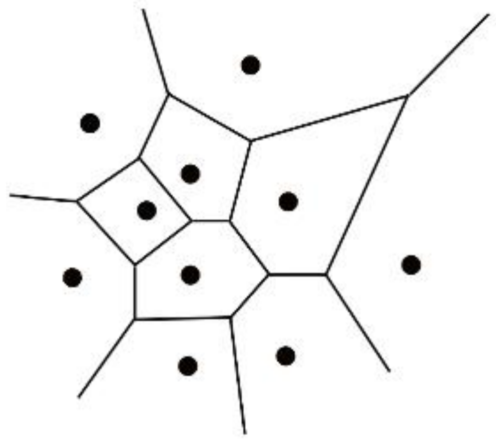
New Algorithm: Batch Value-function Tournament (BVFT)

- $f = Q^\star \Leftrightarrow \|f - \mathcal{T}_\mathcal{G} f\| = 0$, where $\mathcal{T}_\mathcal{G} f$ is $\mathcal{T} f$ projected onto a function space G satisfying [Gordon'96]
  - $Q^\star \in \mathcal{G}$
  - G is piecewise constant



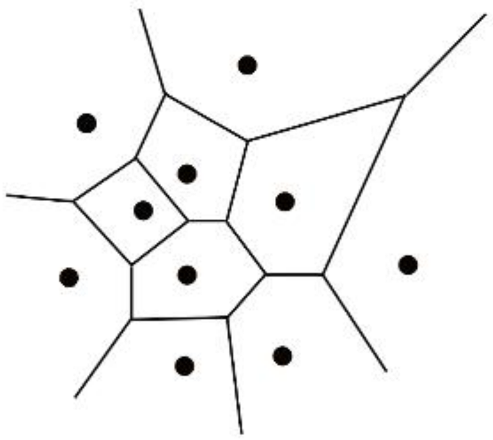
New Algorithm: Batch Value-function Tournament (BVFT)

- $f = Q^\star \Leftrightarrow \|f - \mathcal{T}_\mathcal{G} f\| = 0$, where $\mathcal{T}_\mathcal{G} f$ is $\mathcal{T} f$ projected onto a function space G satisfying [Gordon'96]
  - $Q^\star \in \mathcal{G}$
  - G is piecewise constant
- Where can we find such a *magical* G?



New Algorithm: Batch Value-function Tournament (BVFT)

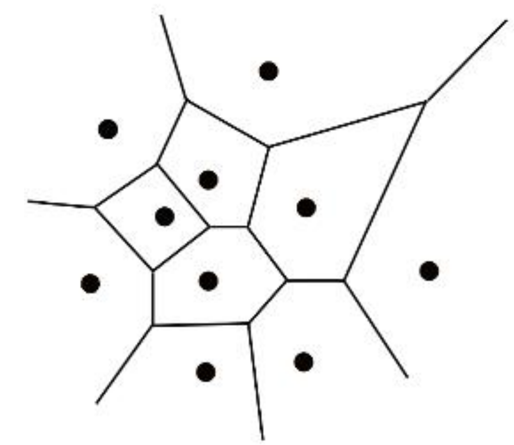
- $f = Q^\star \Leftrightarrow \|f - \mathcal{T}_\mathcal{G} f\| = 0$, where $\mathcal{T}_\mathcal{G} f$ is $\mathcal{T} f$ projected onto a function space G satisfying [Gordon'96]
  - $Q^\star \in \mathcal{G}$
  - G is piecewise constant



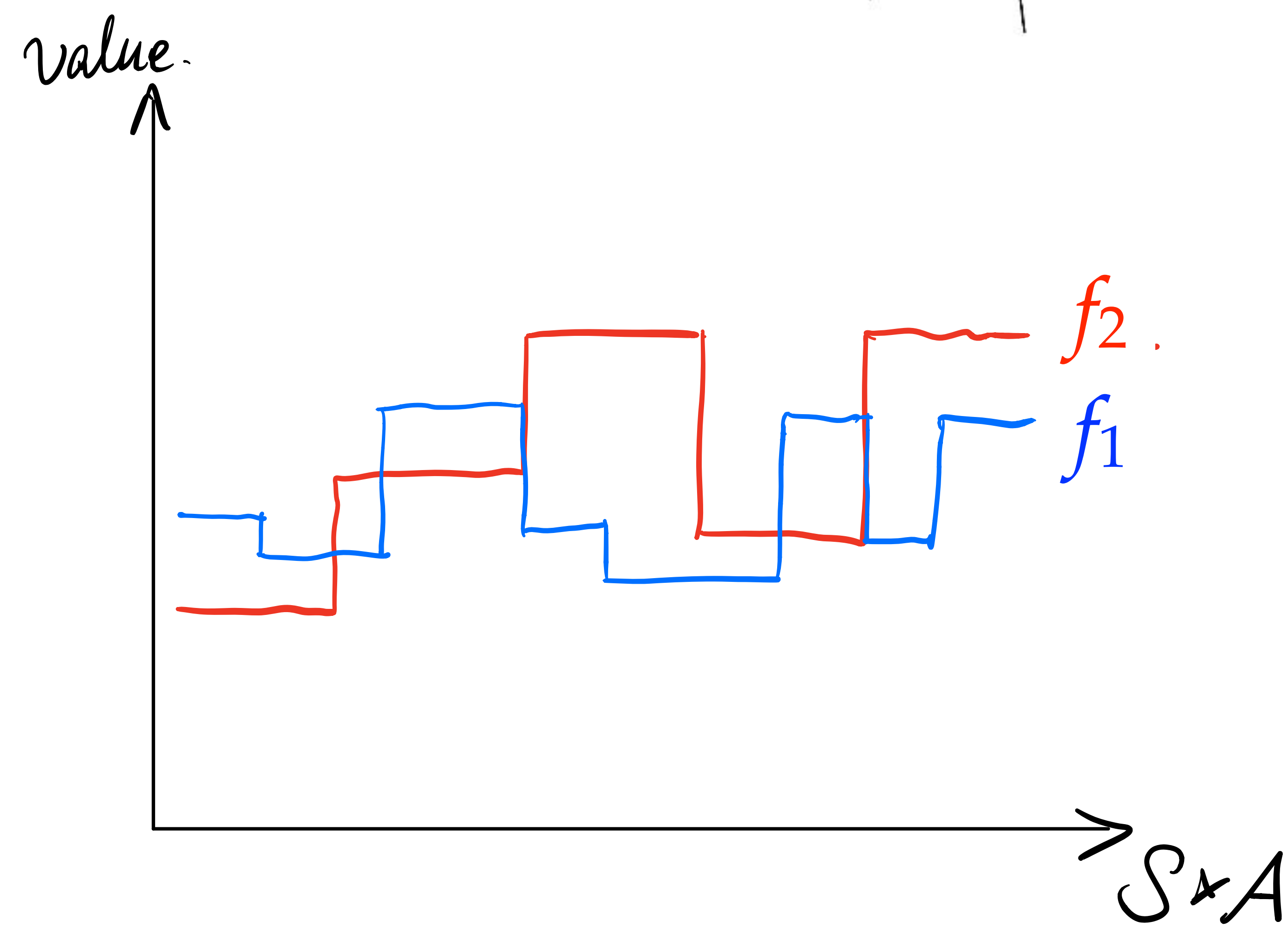
- Where can we find such a *magical* G?
  - out of f_1, f_2 themselves!

New Algorithm: Batch Value-function Tournament (BVFT)

- $f = Q^\star \Leftrightarrow \|f - \mathcal{T}_\mathcal{G} f\| = 0$, where $\mathcal{T}_\mathcal{G} f$ is $\mathcal{T} f$ projected onto a function space G satisfying [Gordon'96]
  - $Q^\star \in \mathcal{G}$
  - G is piecewise constant

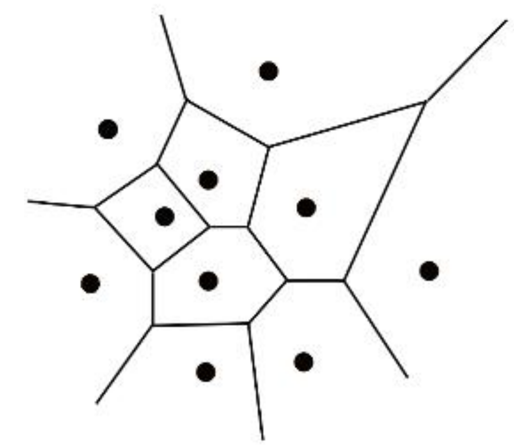


- Where can we find such a *magical* G?
  - out of f_1, f_2 themselves!

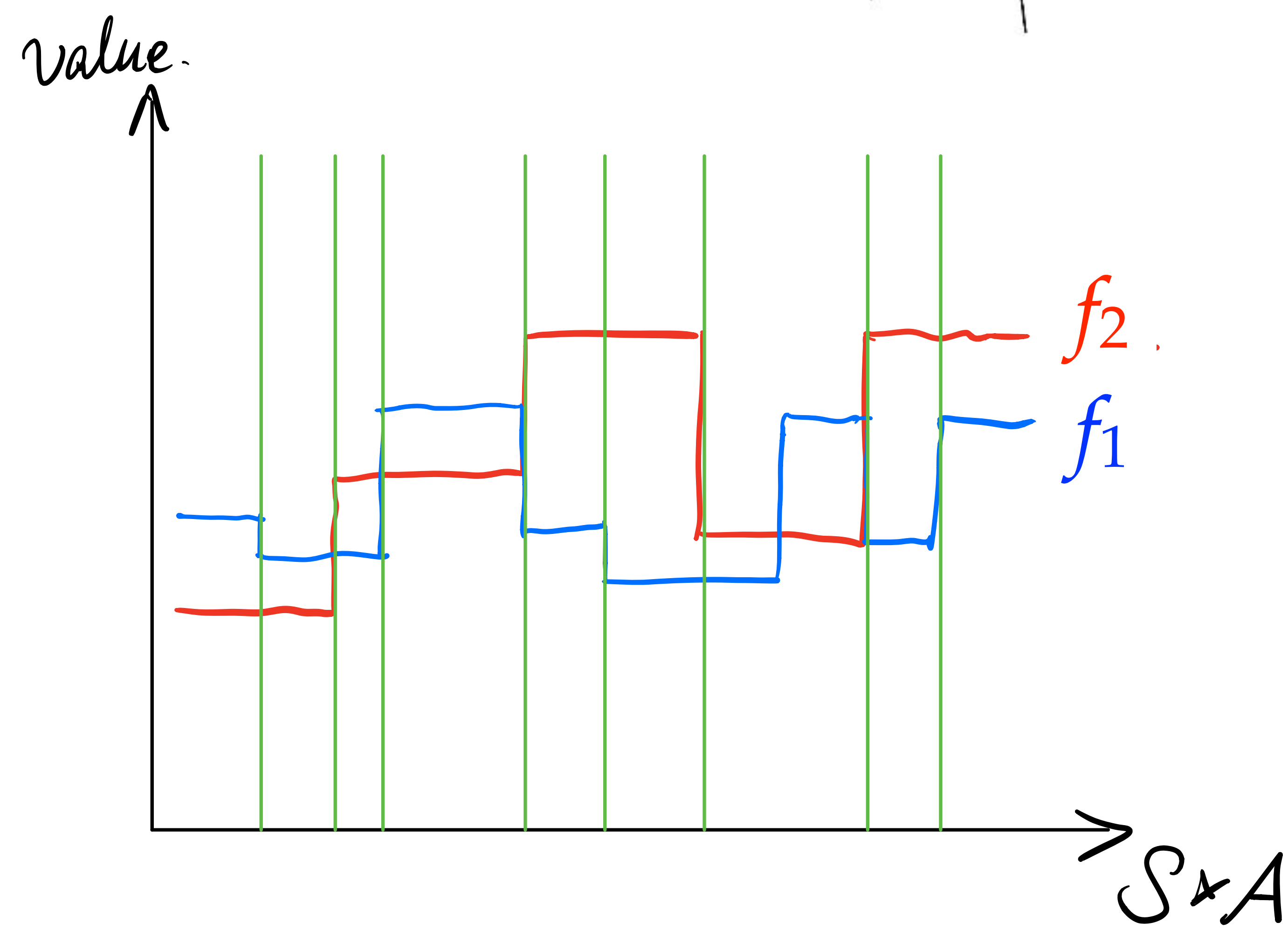


New Algorithm: Batch Value-function Tournament (BVFT)

- $f = Q^\star \Leftrightarrow \|f - \mathcal{T}_\mathcal{G} f\| = 0$, where $\mathcal{T}_\mathcal{G} f$ is $\mathcal{T} f$ projected onto a function space G satisfying [Gordon'96]
  - $Q^\star \in \mathcal{G}$
  - G is piecewise constant

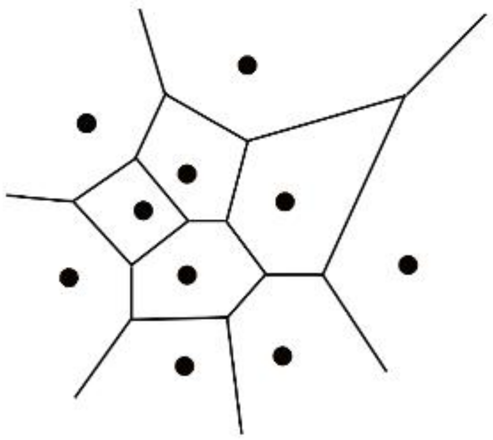


- Where can we find such a *magical* G?
  - **out of** f_1**,** f_2 **themselves!**

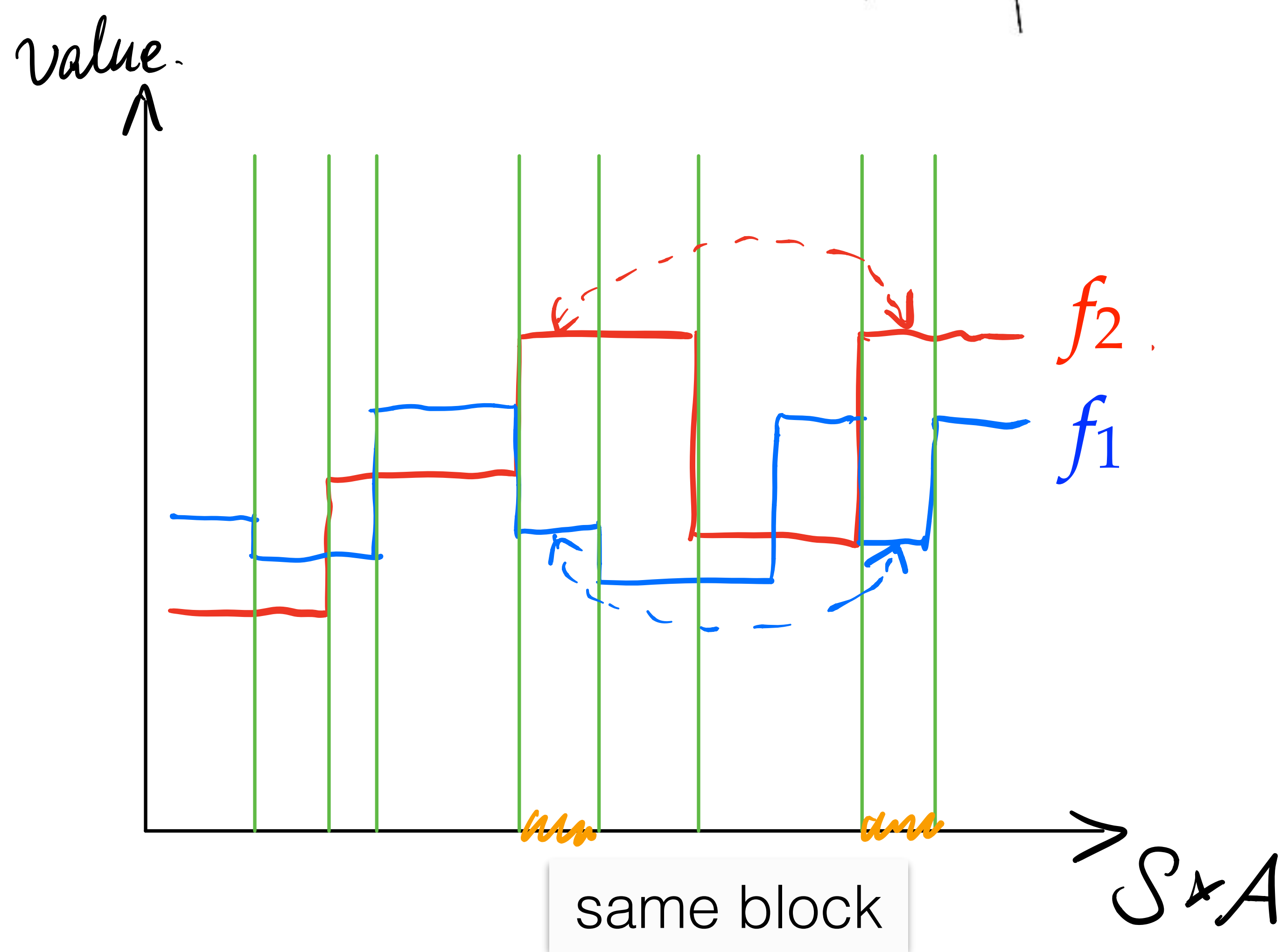


New Algorithm: Batch Value-function Tournament (BVFT)

- $f = Q^\star \Leftrightarrow \|f - \mathcal{T}_\mathcal{G} f\| = 0$, where $\mathcal{T}_\mathcal{G} f$ is $\mathcal{T} f$ projected onto a function space G satisfying [Gordon'96]
  - $Q^\star \in \mathcal{G}$
  - G is piecewise constant

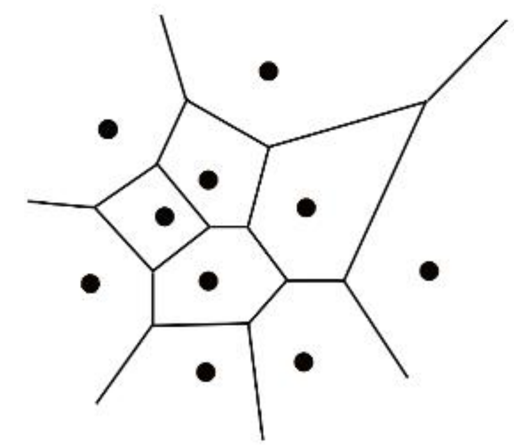


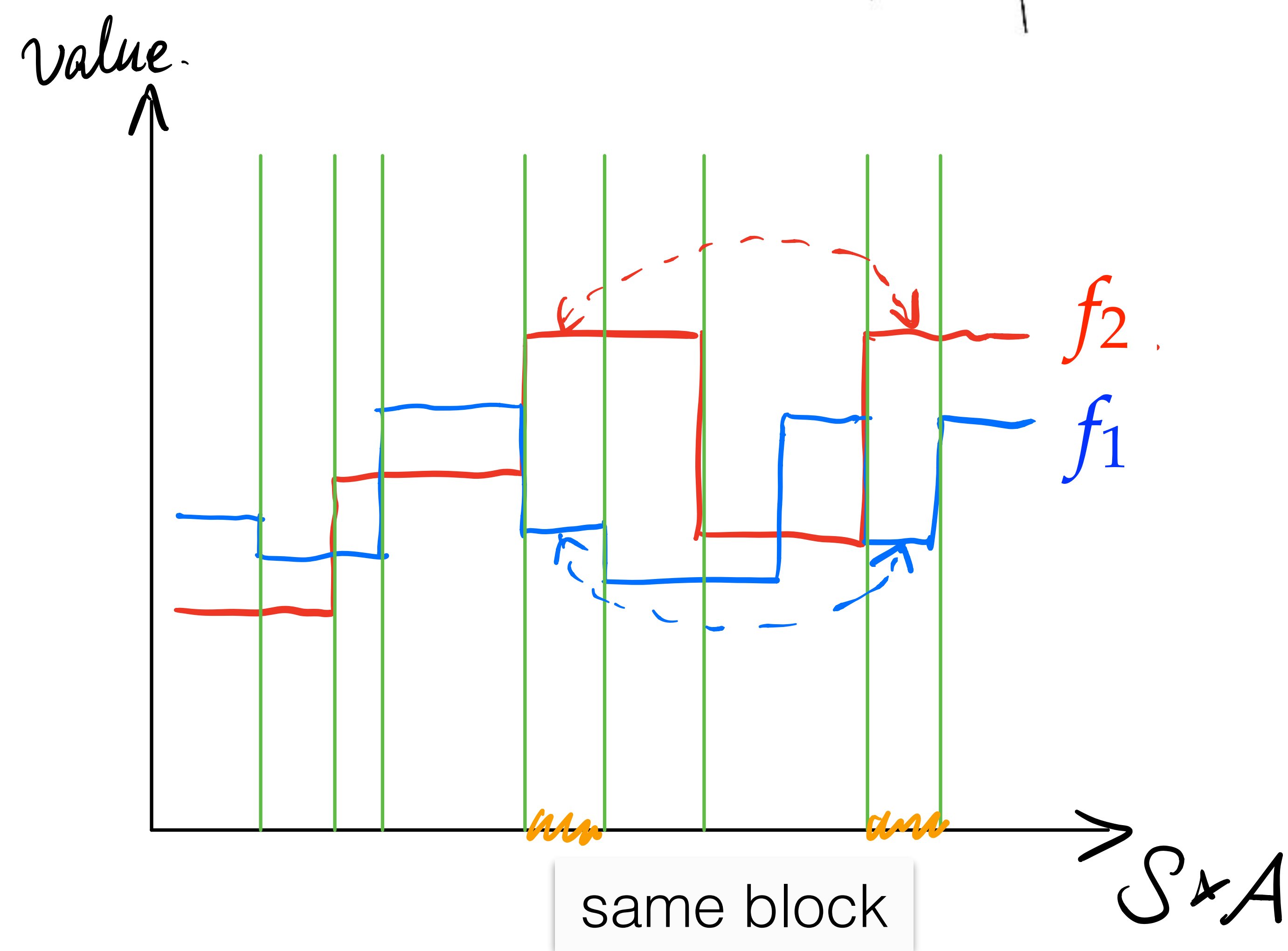
- Where can we find such a *magical* G?
  - **out of f_1, f_2 themselves!**



New Algorithm: Batch Value-function Tournament (BVFT)

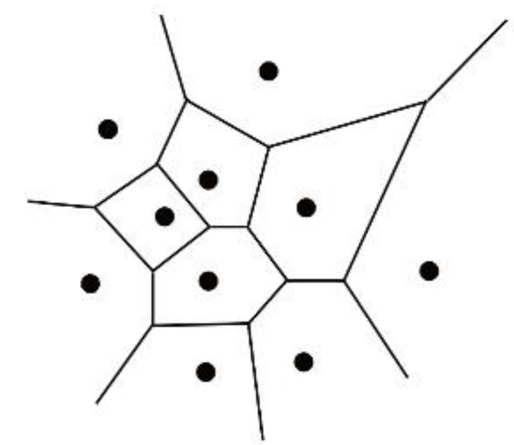
- $f = Q^\star \Leftrightarrow \|f - \mathcal{T}_\mathcal{G} f\| = 0$, where $\mathcal{T}_\mathcal{G} f$ is $\mathcal{T} f$ projected onto a function space G satisfying [Gordon'96]
  - $Q^\star \in \mathcal{G}$ ← $Q^\star \in \{f_1, f_2\} \subseteq \mathcal{G}$
  - G is piecewise constant
- Where can we find such a *magical* G?
  - **out of f_1, f_2 themselves!**



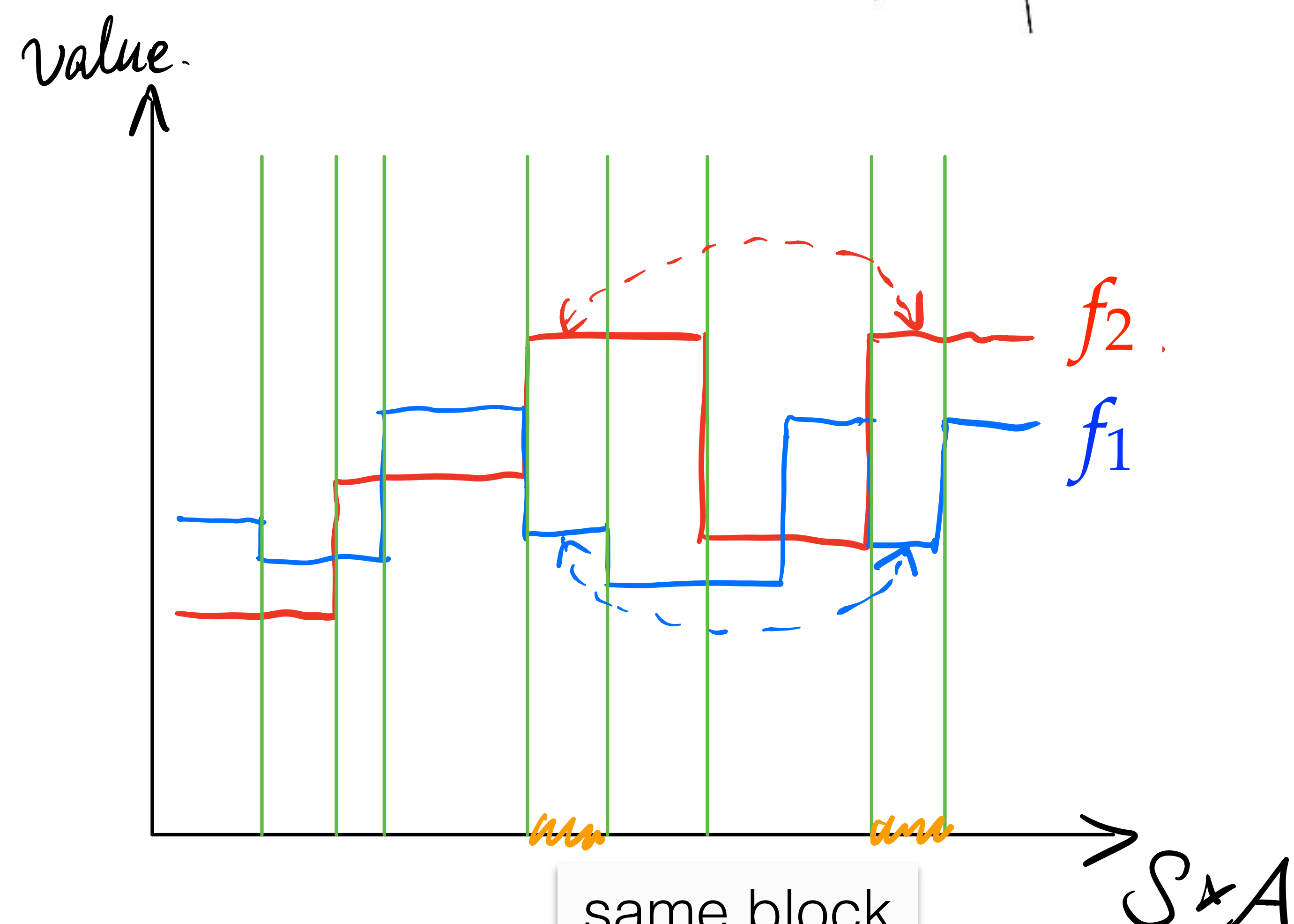


New Algorithm: Batch Value-function Tournament (BVFT)

- $f = Q^\star \Leftrightarrow \|f - \mathcal{T}_\mathcal{G} f\| = 0$, where $\mathcal{T}_\mathcal{G} f$ is $\mathcal{T} f$ projected onto a function space G satisfying [Gordon'96]
  - $Q^\star \in \mathcal{G}$ ⟵ $Q^\star \in \{f_1, f_2\} \subseteq \mathcal{G}$
  - G is piecewise constant

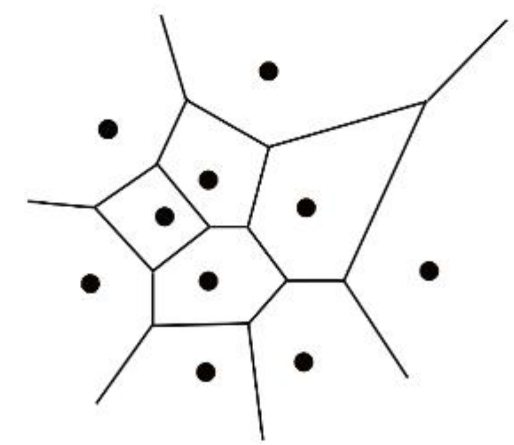


- Where can we find such a *magical* G?
  - **out of f_1, f_2 themselves!**
  - $O(1/\varepsilon^2)$ complexity



New Algorithm: Batch Value-function Tournament (BVFT)

- $f = Q^\star \Leftrightarrow \|f - \mathcal{T}_\mathcal{G} f\| = 0$, where $\mathcal{T}_\mathcal{G} f$ is $\mathcal{T} f$ projected onto a function space G satisfying [Gordon'96]
  - $Q^\star \in \mathcal{G}$ ⟵ $Q^\star \in \{f_1, f_2\} \subsetneq \mathcal{G}$
  - G is piecewise constant



- Where can we find such a *magical* G?
  - **out of f_1, f_2 themselves!**
  - $O(1/\varepsilon^2)$ complexity
- Extend to exponentially many candidates by *pairwise comparison ("tournament")*