Towards Uniformly Superhuman Autonomy via Subdominance Minimization

Brian D. Ziebart, Sanjiban Choudhury,
Xinyan (Shane) Yan, and Paul Vernaza

How should we think about imitation learning?

“Imitation is the sincerest form of flattery that mediocrity can pay to greatness.”

Oscar Wilde

"Imitation is the sincerest form of flattery that mediocrity can pay to greatness."

Oscar Wilde

<u>**Gold standard human demonstrations**</u>

(Near) Optimal, minimum noise, known biases

"Imitation is the sincerest form of flattery that mediocrity can pay to greatness."

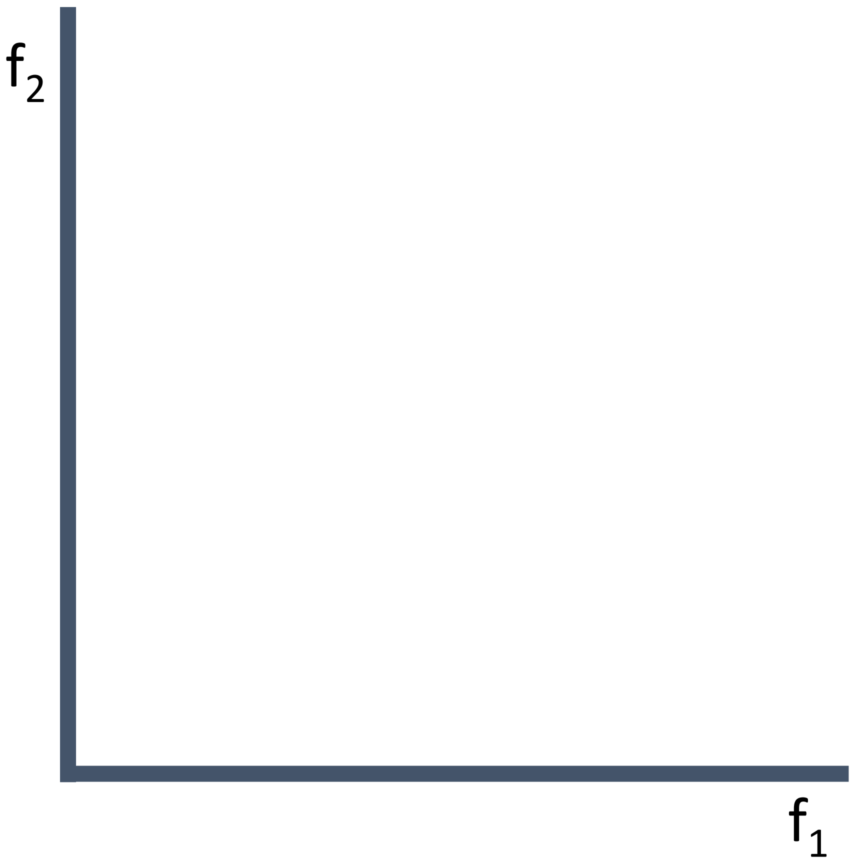
Oscar Wilde

<u>Gold standard human demonstrations</u>

(Near) Optimal, minimum noise, known biases

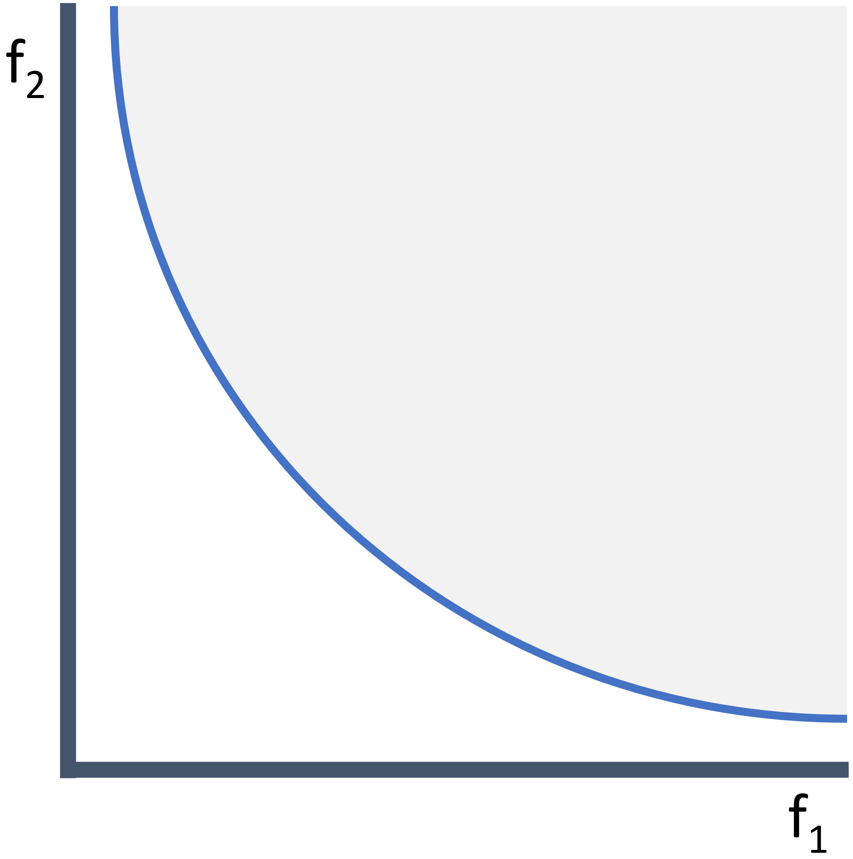
Formulation: Rationalize/match performance

Easy: Optimal demonstrations



cost features f_1, f_2, ...,

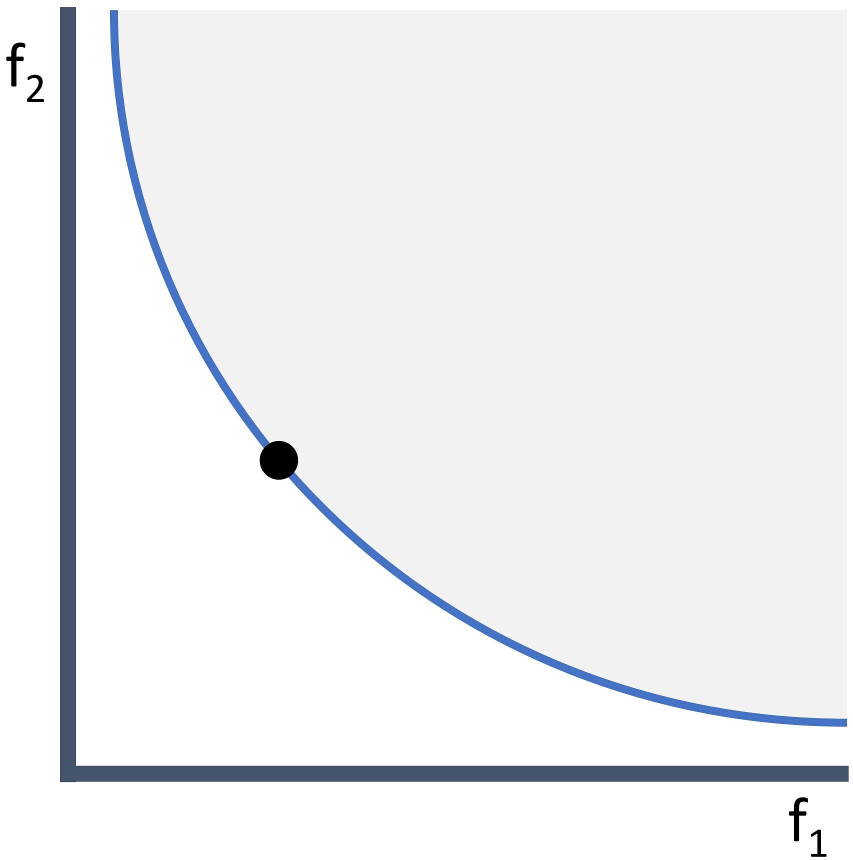
Easy: Optimal demonstrations



cost features $f_1, f_2, \ldots,$

Pareto frontier

Easy: Optimal demonstrations

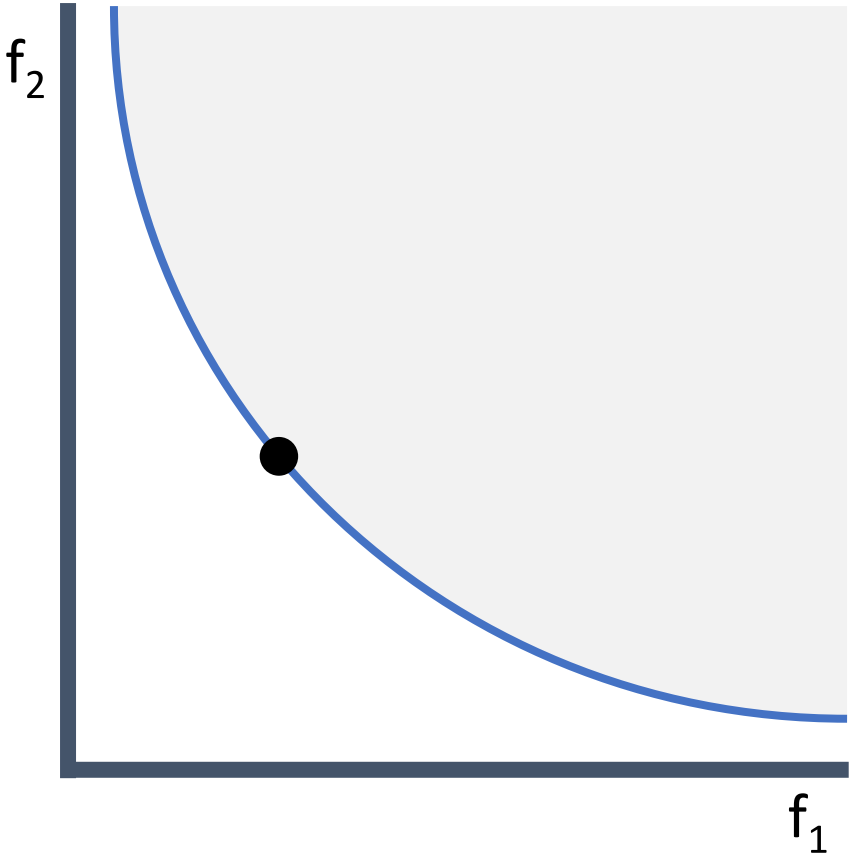


cost features f_1, f_2, ...,

human demonstration(s),

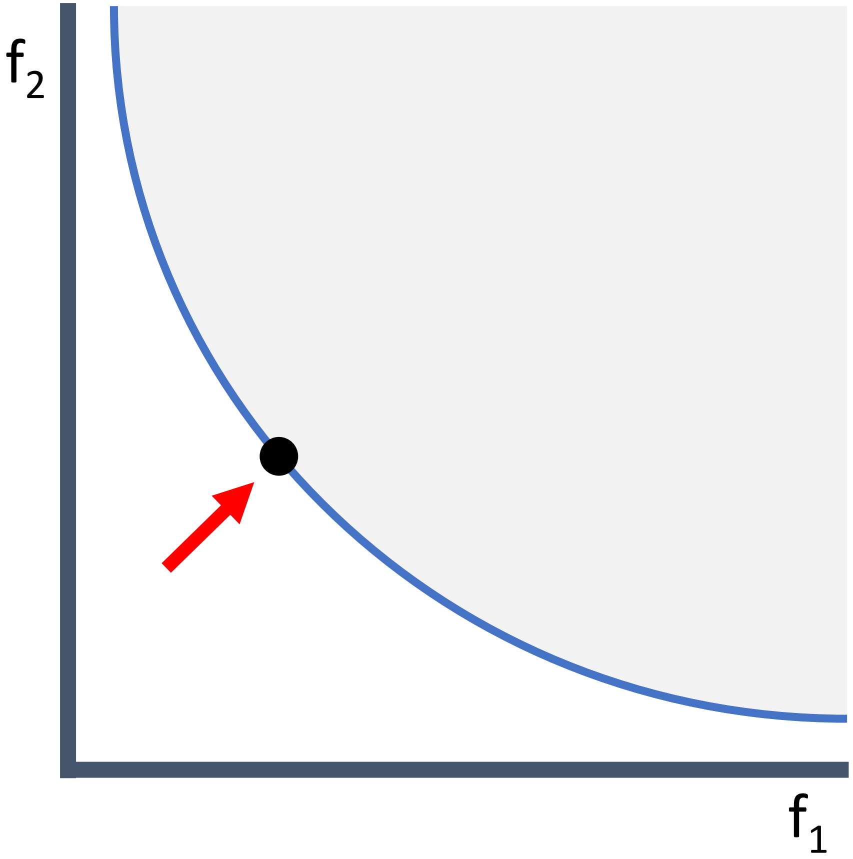
Pareto frontier

Easy: Optimal demonstrations



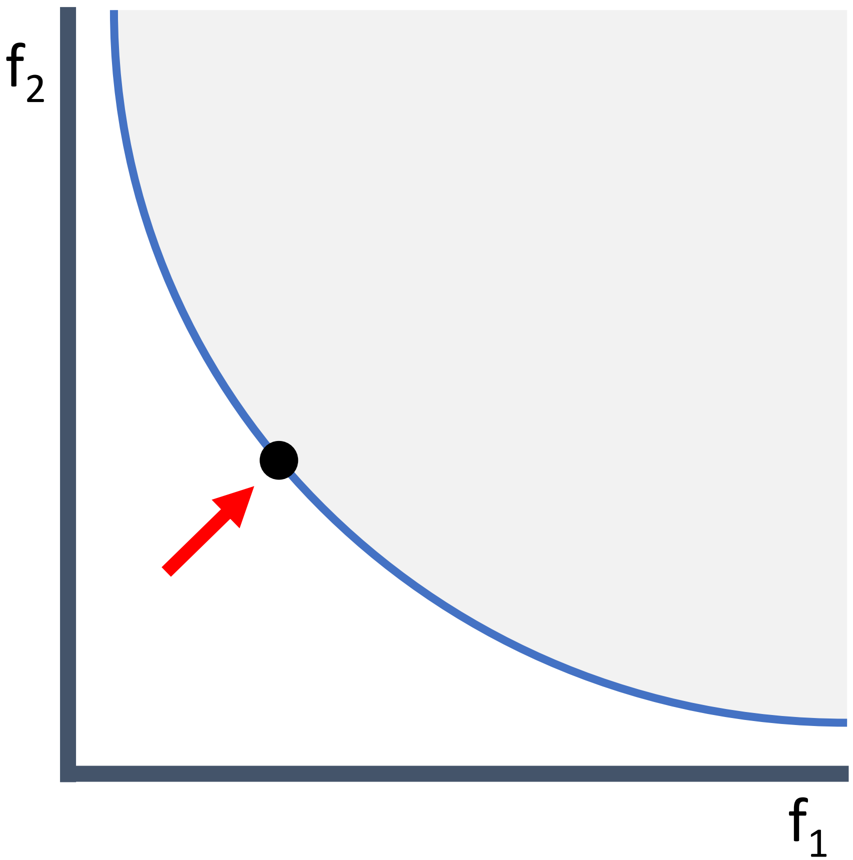
Given **cost features** f_1, f_2, …, learn weights **w** that make **human demonstration(s)**, which must reside on the **Pareto frontier**, **optimal.**

Easy: Optimal demonstrations



Given **cost features** f_1, f_2, ..., learn weights **w** that make **human demonstration(s)**, which must reside on the Pareto frontier, **optimal.**

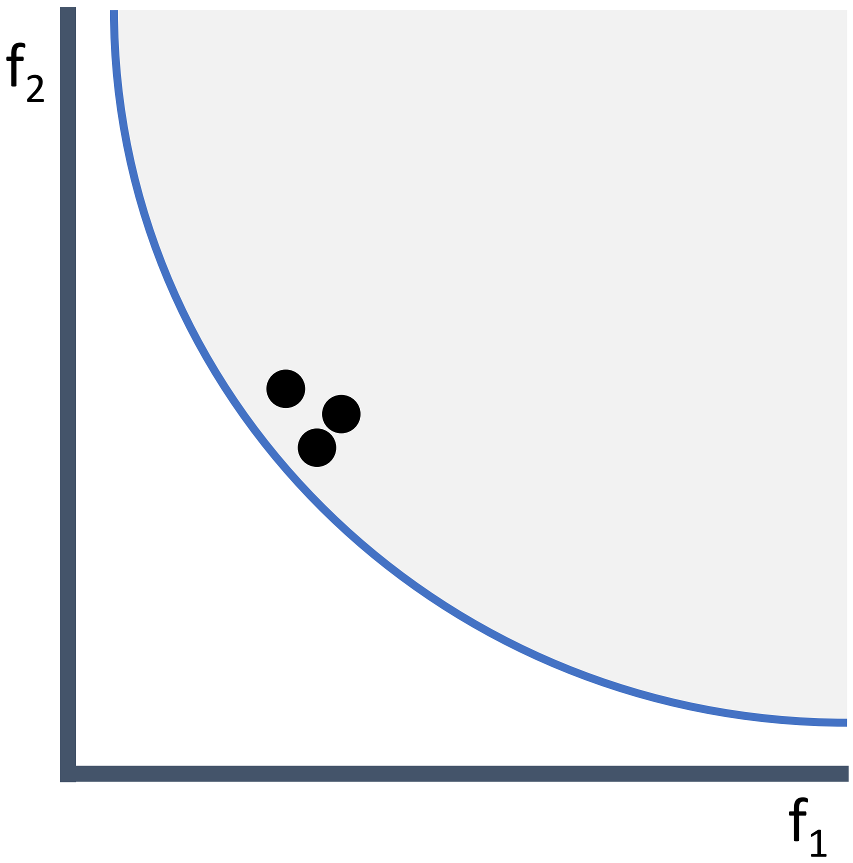
Easy: Optimal demonstrations



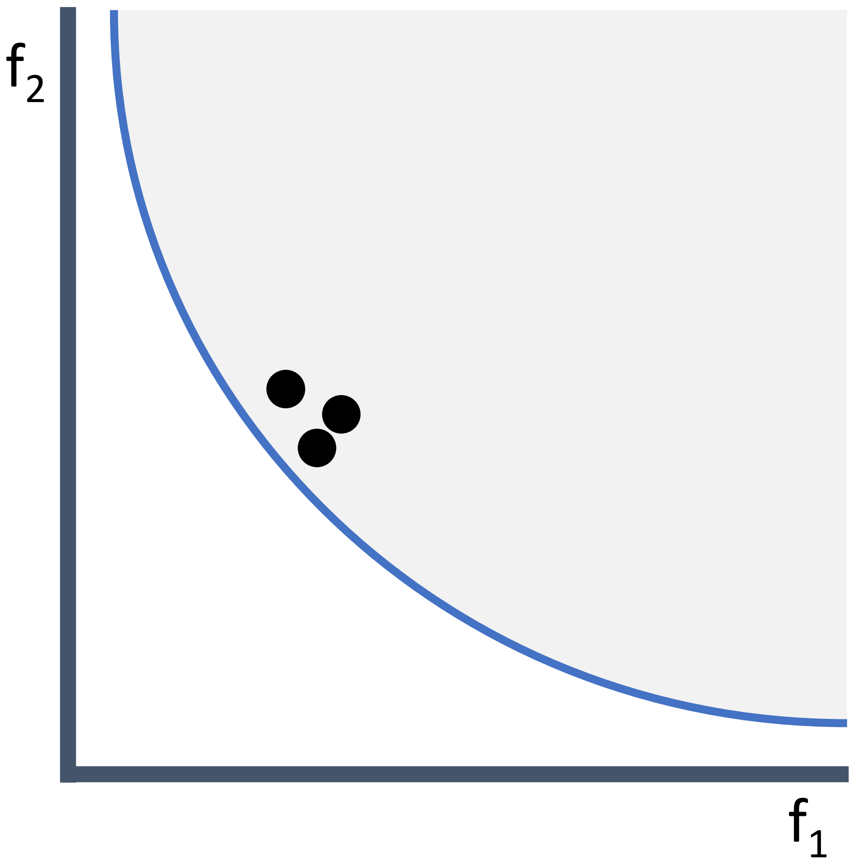
Given **cost features** f_1, f_2, ..., learn weights **w** that make **human demonstration(s)**, which must reside on the Pareto frontier, **optimal.**

Degenerate solutions (**w**=**0**) exist, but can be avoided (Ng & Russell 2000; Ratliff, Bagnell, Zinkevich 2006)

Harder: Suboptimal demonstrations

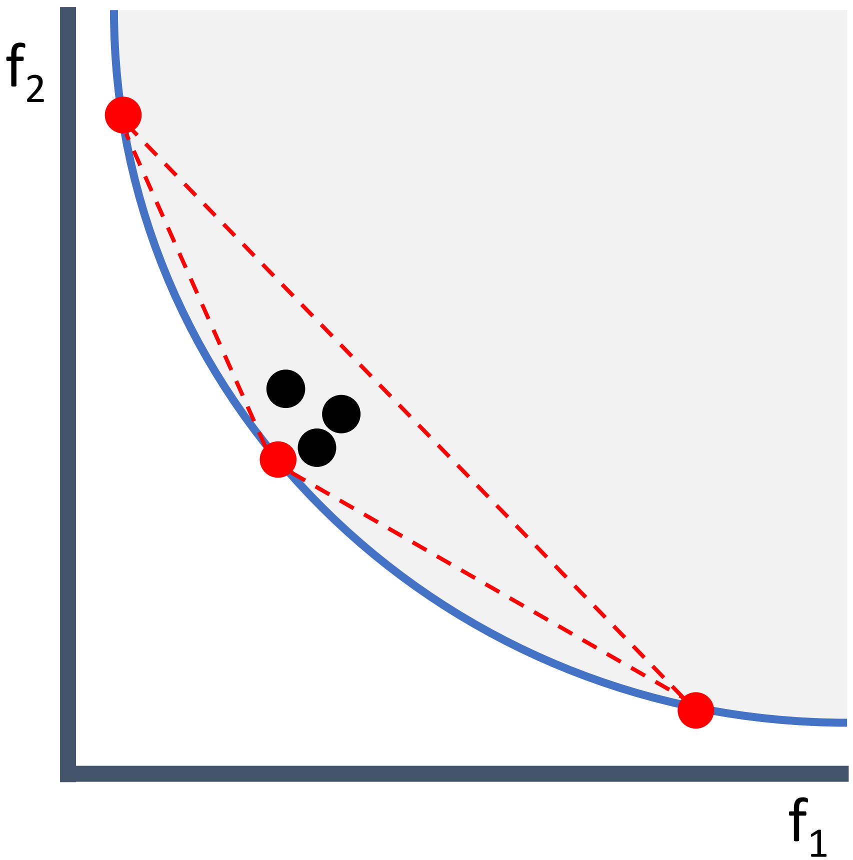


Harder: Suboptimal demonstrations



Feature Matching (Abbeel & Ng 2004): Matching expected features guarantees equal expected cost/reward (assuming linearity).

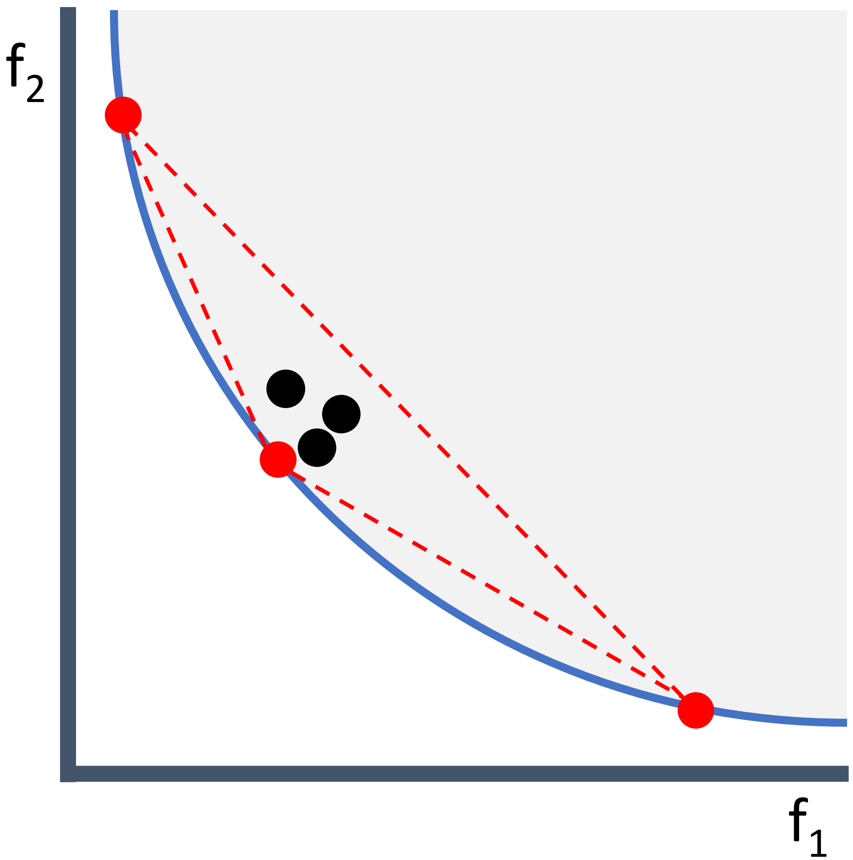
Harder: Suboptimal demonstrations



Feature Matching (Abbeel & Ng 2004): Matching expected features guarantees equal expected cost/reward (assuming linearity).

Approach: Mix optimal policies to match features.

Harder: Suboptimal demonstrations

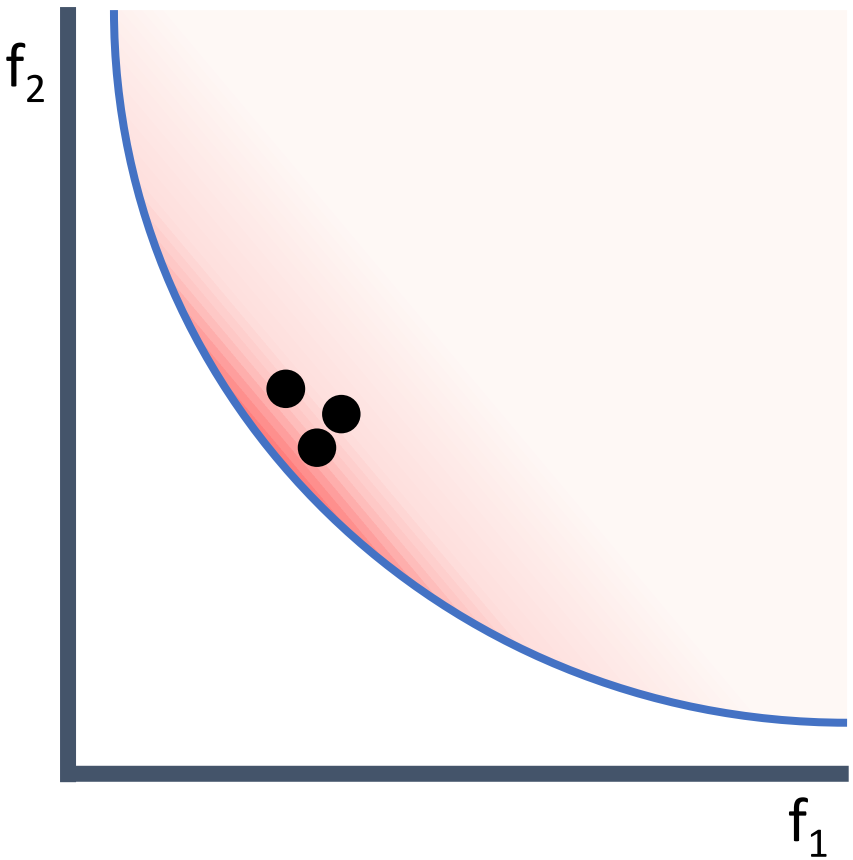


Feature Matching (Abbeel & Ng 2004): Matching expected features guarantees equal expected cost/reward (assuming linearity).

Approach: Mix optimal policies to match features.

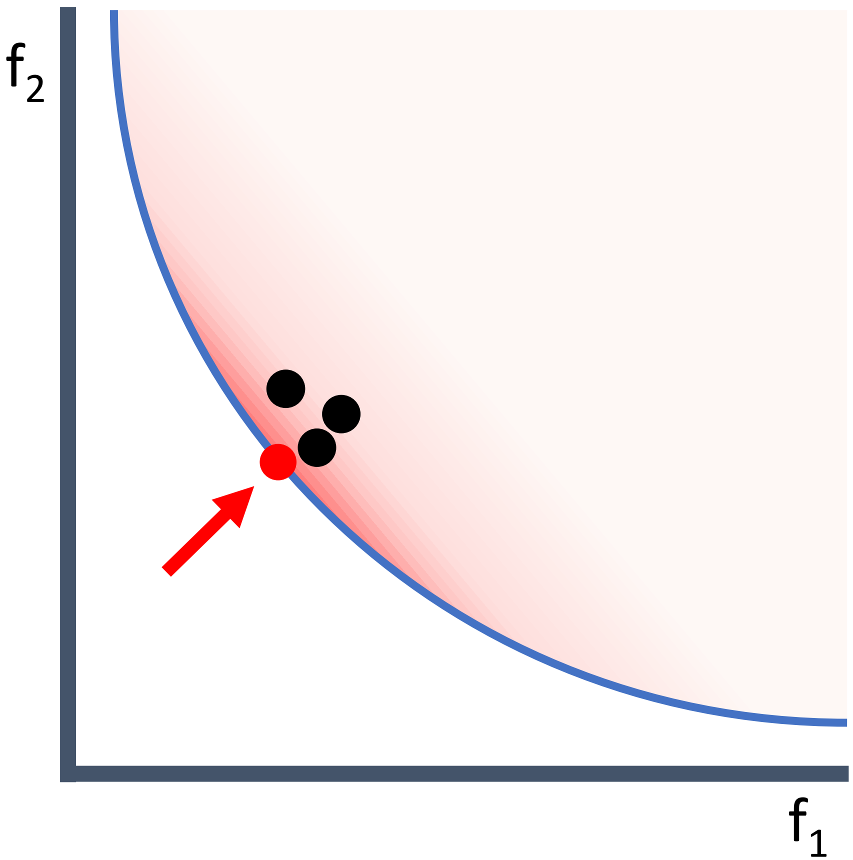
Limitations: Many solutions exist; which policy to deploy?

Harder: Suboptimal demonstrations



MaxEnt IRL (Ziebart et al. 2008): Demonstrations are noisy with probability $\propto e^{-\mathbf{w} \cdot \mathbf{f}}$

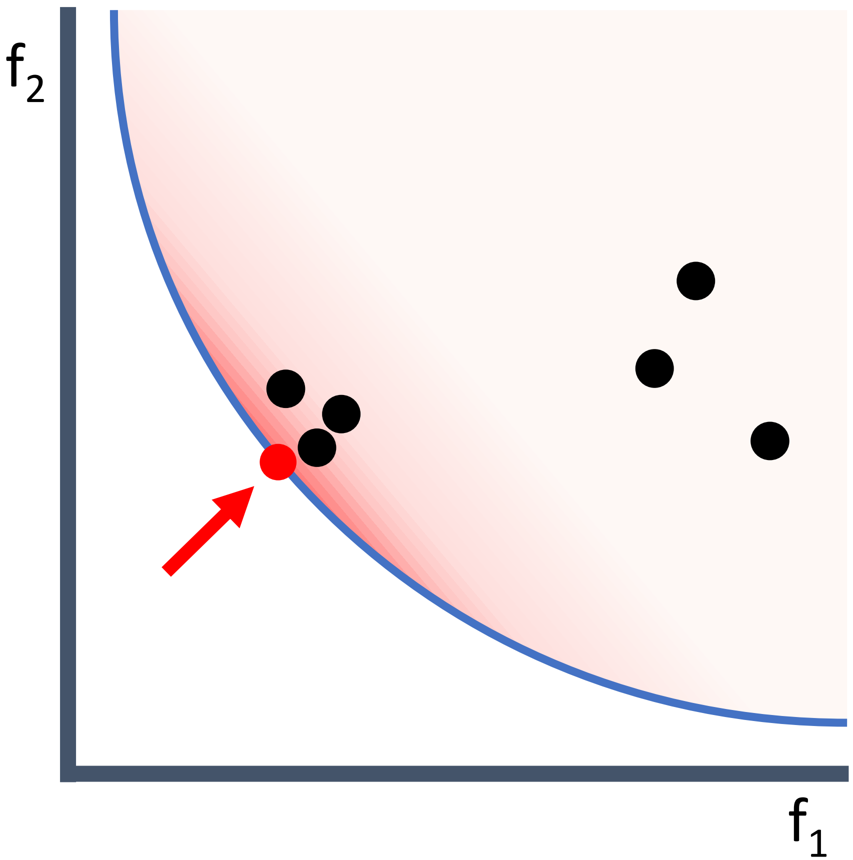
Harder: Suboptimal demonstrations



MaxEnt IRL (Ziebart et al. 2008): Demonstrations are noisy with probability $\propto e^{-\mathbf{w}\cdot\mathbf{f}}$

Apprenticeship learning: estimate **w**, employ distribution mode (i.e., minimize **w**·**f**).

Harder: Suboptimal demonstrations

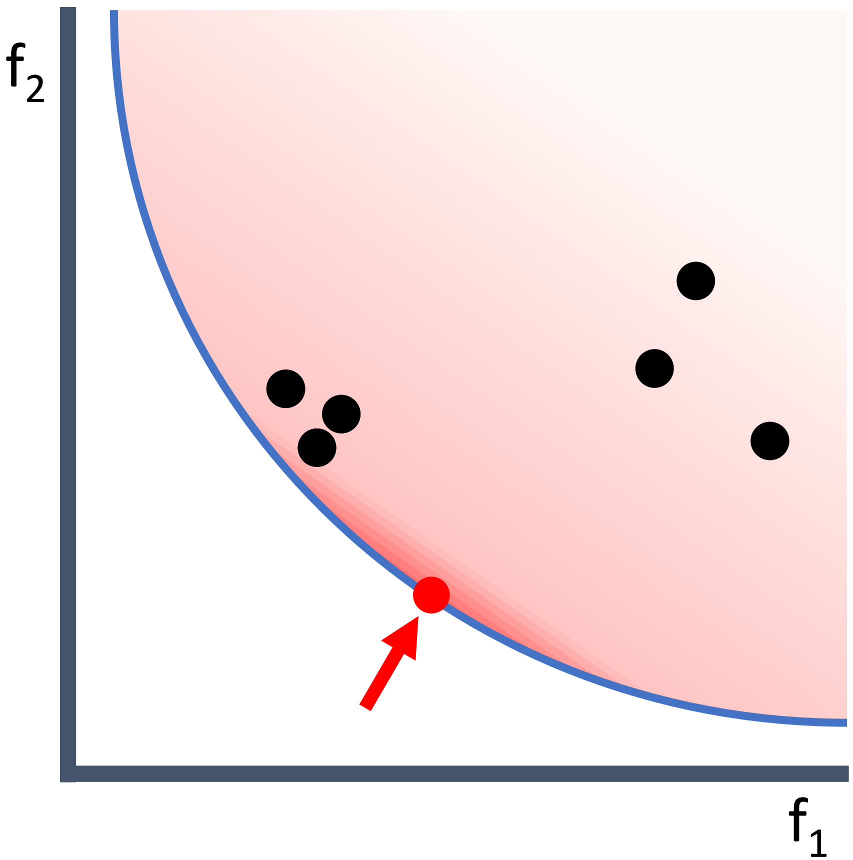


MaxEnt IRL (Ziebart et al. 2008): Demonstrations are noisy with probability $\propto e^{-\mathbf{w}\cdot\mathbf{f}}$

Apprenticeship learning: estimate **w**, employ distribution mode (i.e., minimize **w**·**f**).

Outliers violating noise model can significantly shift the mode.

Harder: Suboptimal demonstrations

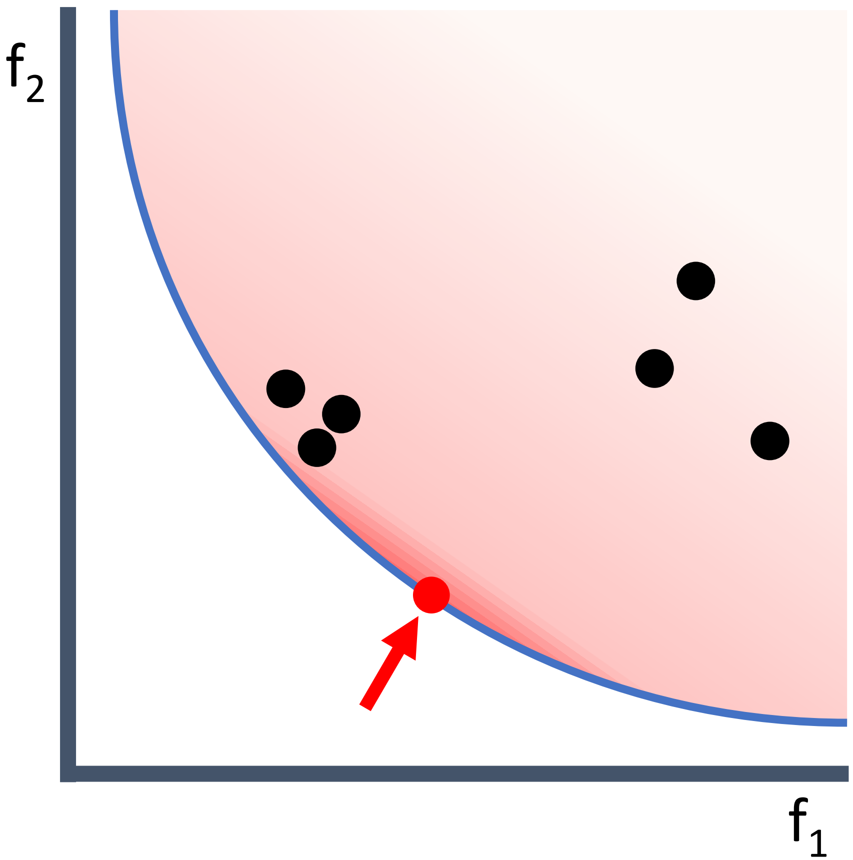


MaxEnt IRL (Ziebart et al. 2008): Demonstrations are noisy with probability $\propto e^{-\mathbf{w}\cdot\mathbf{f}}$

Apprenticeship learning: estimate **w**, employ distribution mode (i.e., minimize **w**·**f**).

Outliers violating noise model can significantly shift the mode.

Harder: Suboptimal demonstrations



MaxEnt IRL (Ziebart et al. 2008): Demonstrations are noisy with probability $\propto e^{-\mathbf{w}\cdot\mathbf{f}}$

Apprenticeship learning: estimate **w**, employ distribution mode (i.e., minimize $\mathbf{w}\cdot\mathbf{f}$).

Outliers violating noise model can significantly shift the mode.

Substantial amounts of related work seek to “ignore” outliers.

Rankings/Confidences (Ibarz et al., 2018; Brown et al., 2019; Brown et al., 2020, Novoseller et al., 2020; Zhang et al., 2021; Myers et al., 2021; Tangkaratt et al., 2020; 2021; Wang et al., 2021a; Wang et al. 2021b; Bıyık et al., 2022)
Noise models (Evans et al., 2016; Majumdar et al., 2017; Reddy et al., 2018; Kwon et al., 2020; Zhi-Xuan et al., 2020)


“I visualize a time when we will be to robots what dogs are to humans, and I'm rooting for the machines.”
Claude Shannon

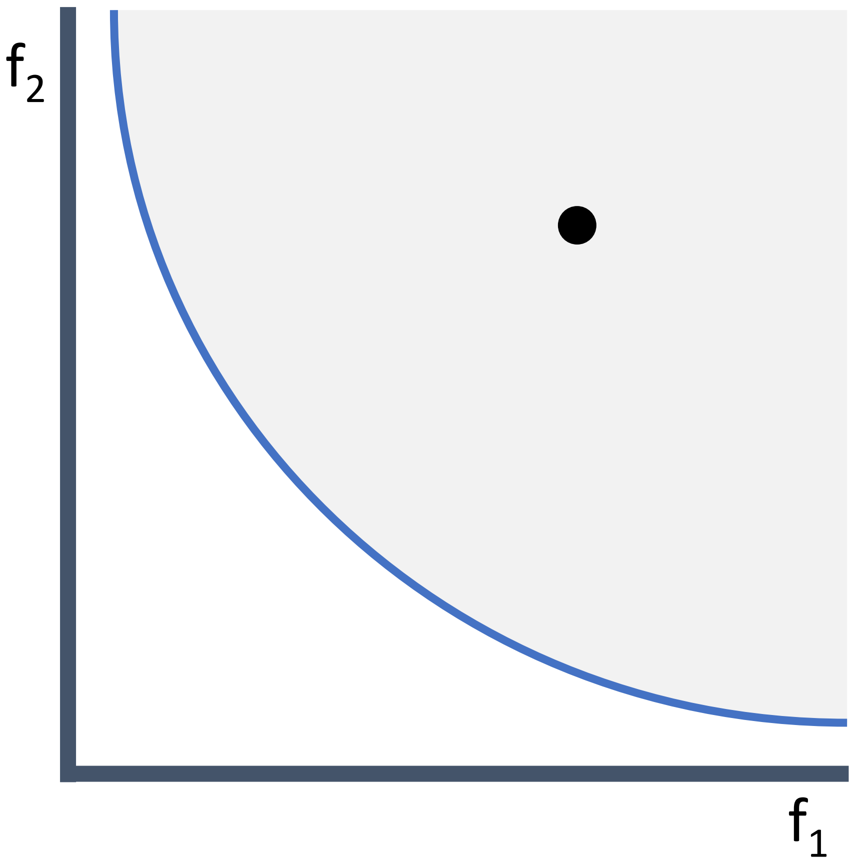
"I visualize a time when we will be to robots what dogs are to humans, and I'm rooting for the machines."

Claude Shannon

Visuomotor imprecision (Wolpert et al. 1995)

Bounded rationality (Simon, 1997)

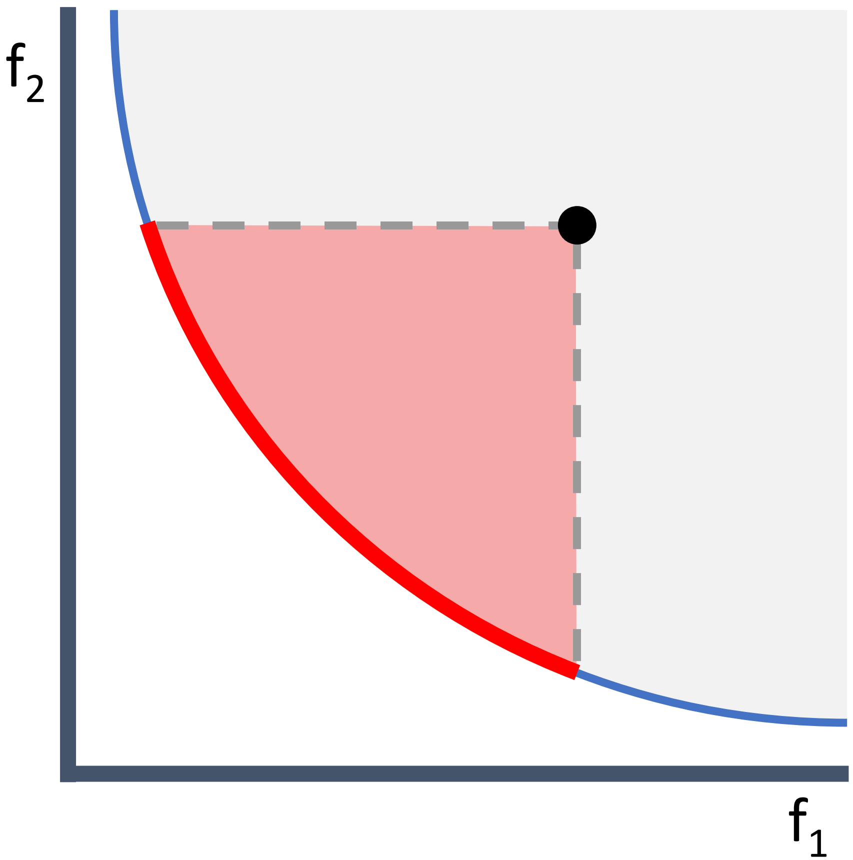
Defining Superhuman Behavior



cost features f_1, f_2, ...
human demonstrations

Pareto frontier

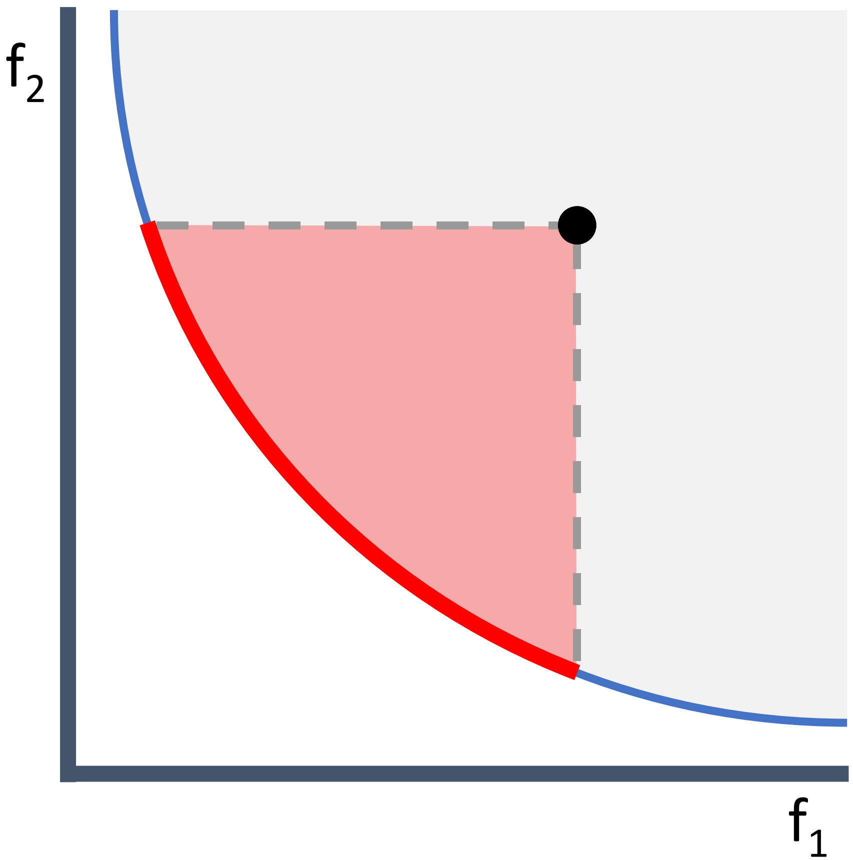
Defining Superhuman Behavior



A **policy** is **superhuman** if it has smaller **cost features** f_1, f_2, … for all **human demonstrations**

Pareto frontier

Defining Superhuman Behavior

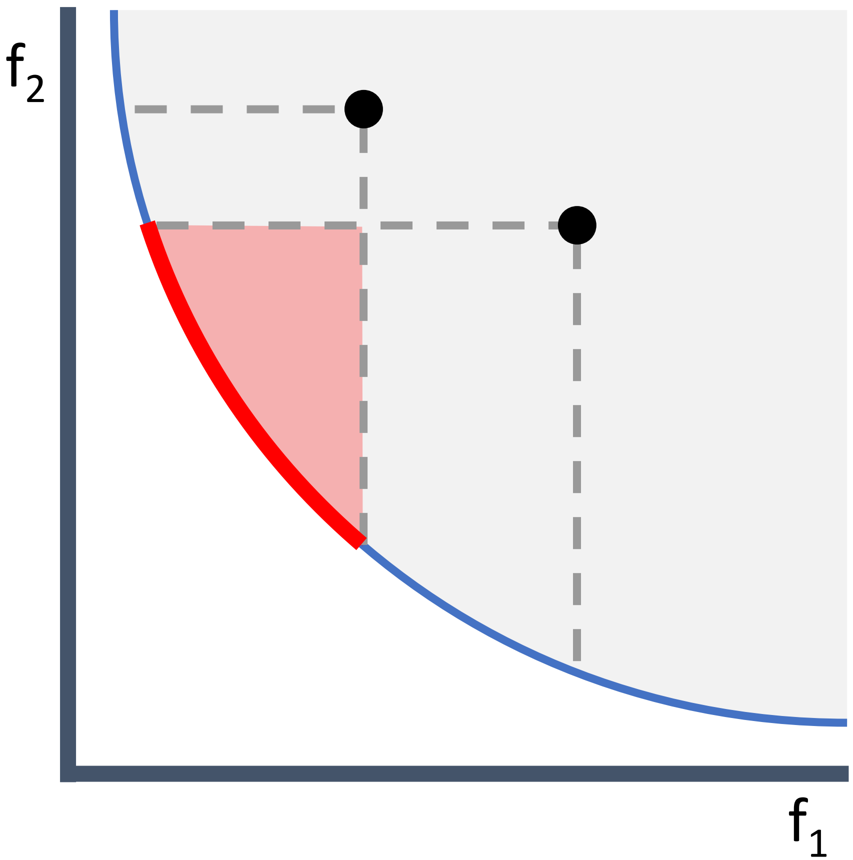


A **policy** is **superhuman** if it has smaller **cost features** f_1, f_2, … for all **human demonstrations**

Guarantees lower cost than demonstration costs for family of additive cost functions

Pareto frontier

Defining Superhuman Behavior

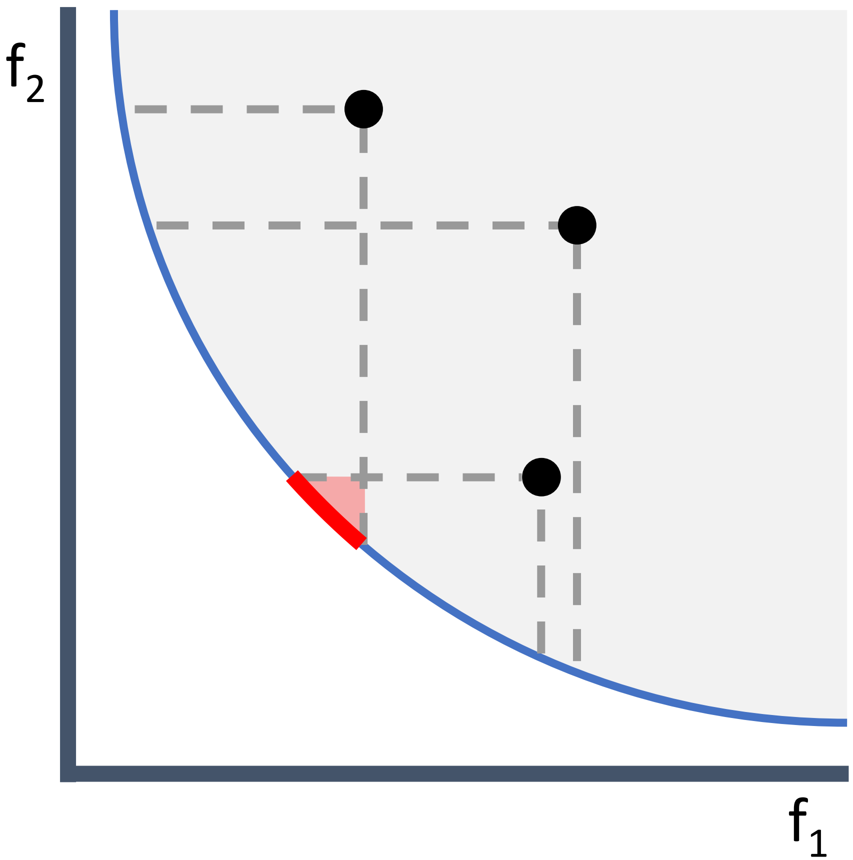


A **policy** is **superhuman** if it has smaller **cost features** f_1, f_2, … for all **human demonstrations**

Guarantees lower cost than demonstration costs for family of additive cost functions

Set of **superhuman policies** on the **Pareto frontier** shrinks as demonstrations grow

Defining Superhuman Behavior

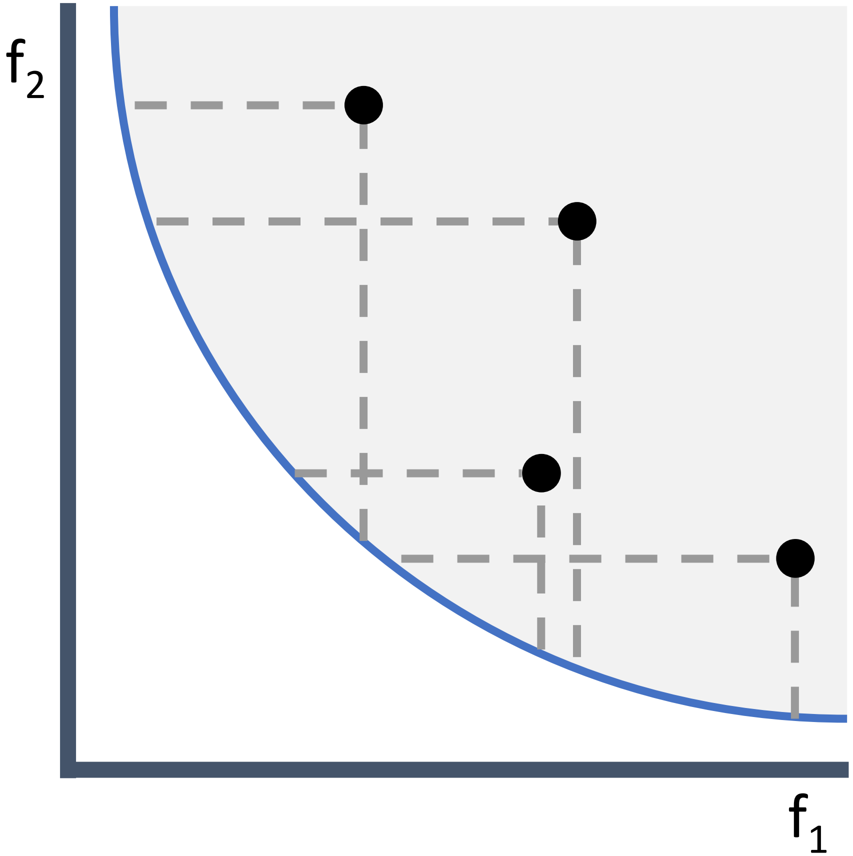


A **policy** is **superhuman** if it has smaller **cost features** f_1, f_2, … for all **human demonstrations**

Guarantees <u>lower cost</u> than demonstration costs for family of additive cost functions

Set of **superhuman policies** on the **Pareto frontier** shrinks as demonstrations grow

Defining Superhuman Behavior



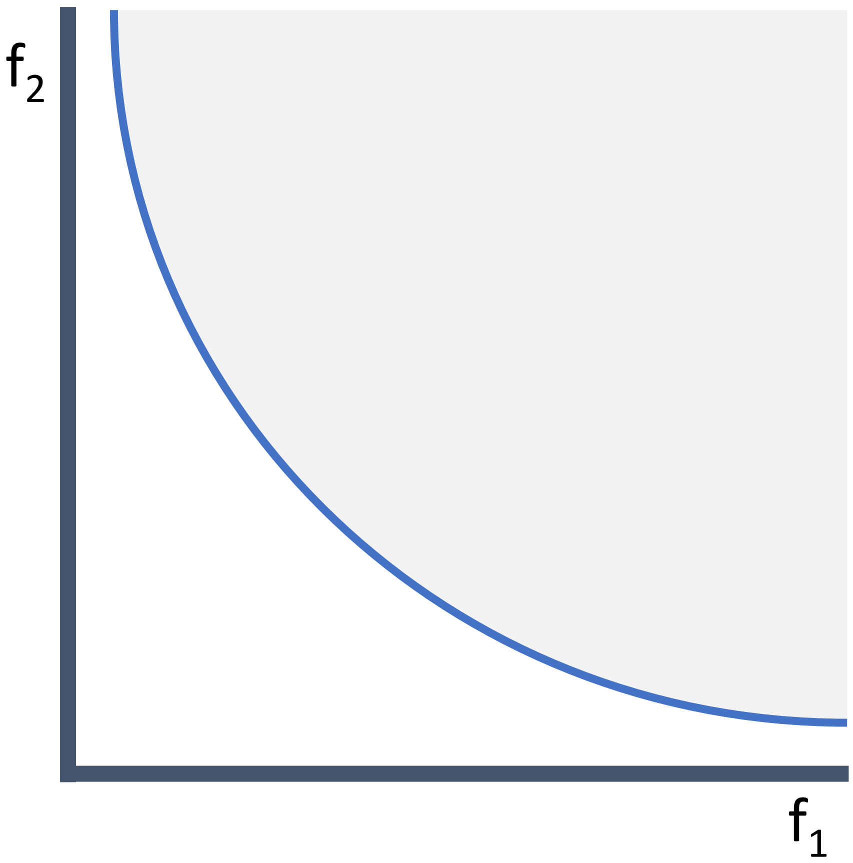
A **policy** is **superhuman** if it has smaller **cost features** f_1, f_2, … for all **human demonstrations**

Guarantees <u>lower cost</u> than demonstration costs for family of additive cost functions

Set of **superhuman policies** on the **Pareto frontier** shrinks as demonstrations grow

Can become empty!

Superhuman Percentile & Subdominance



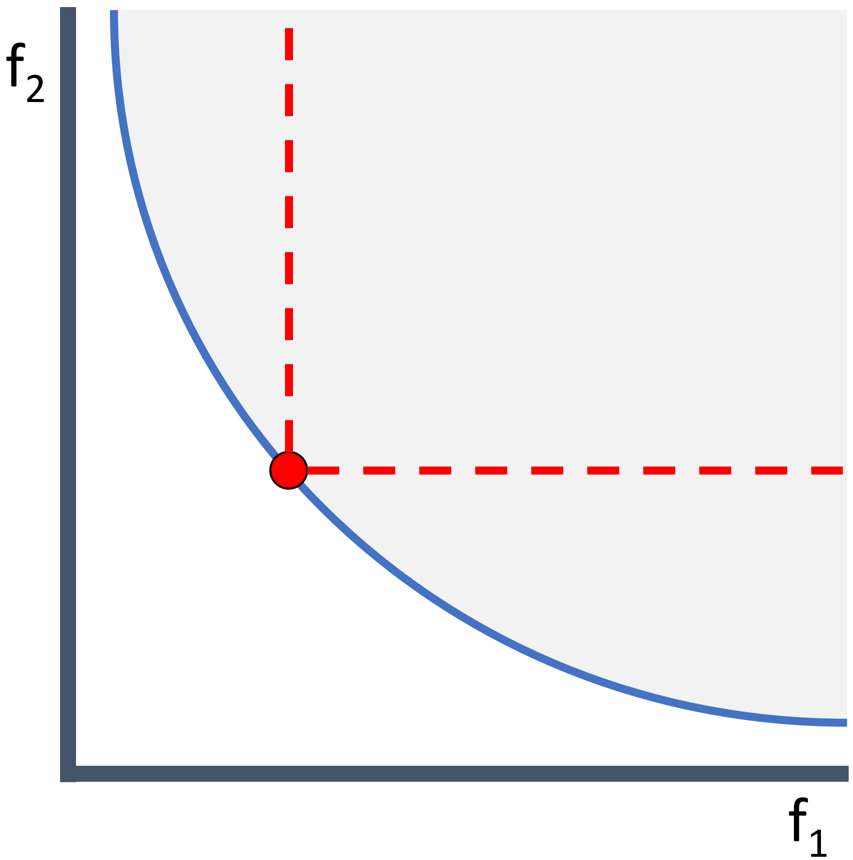
cost features f_1, f_2, ...

Pareto frontier

Superhuman Percentile & Subdominance

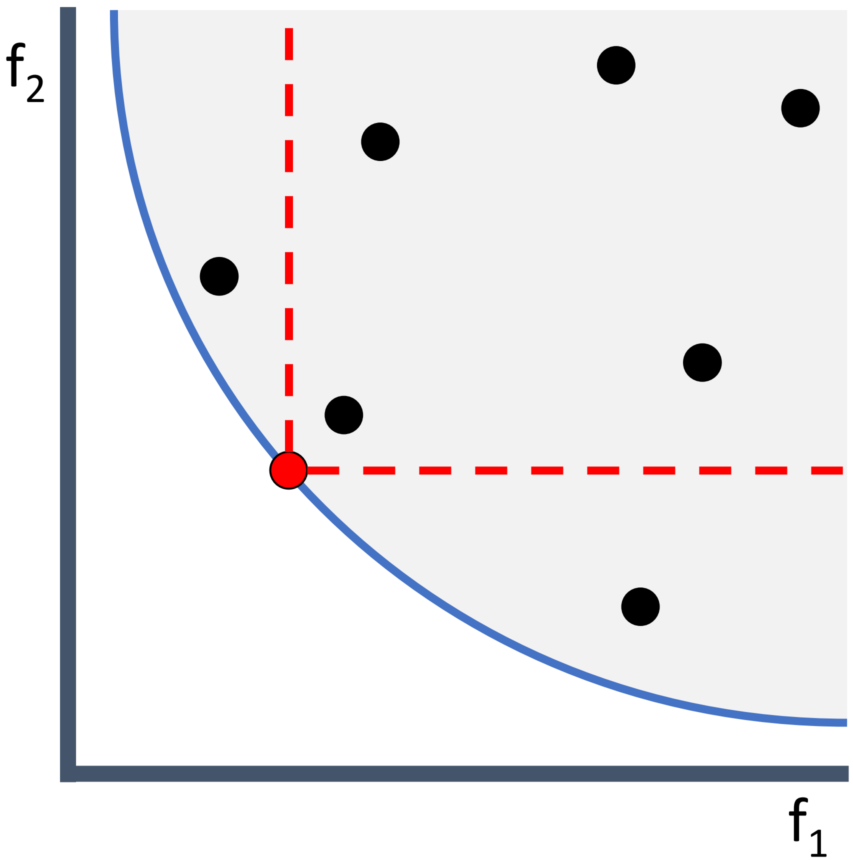
policy

cost features f_1, f_2, ...



Pareto frontier

Superhuman Percentile & Subdominance



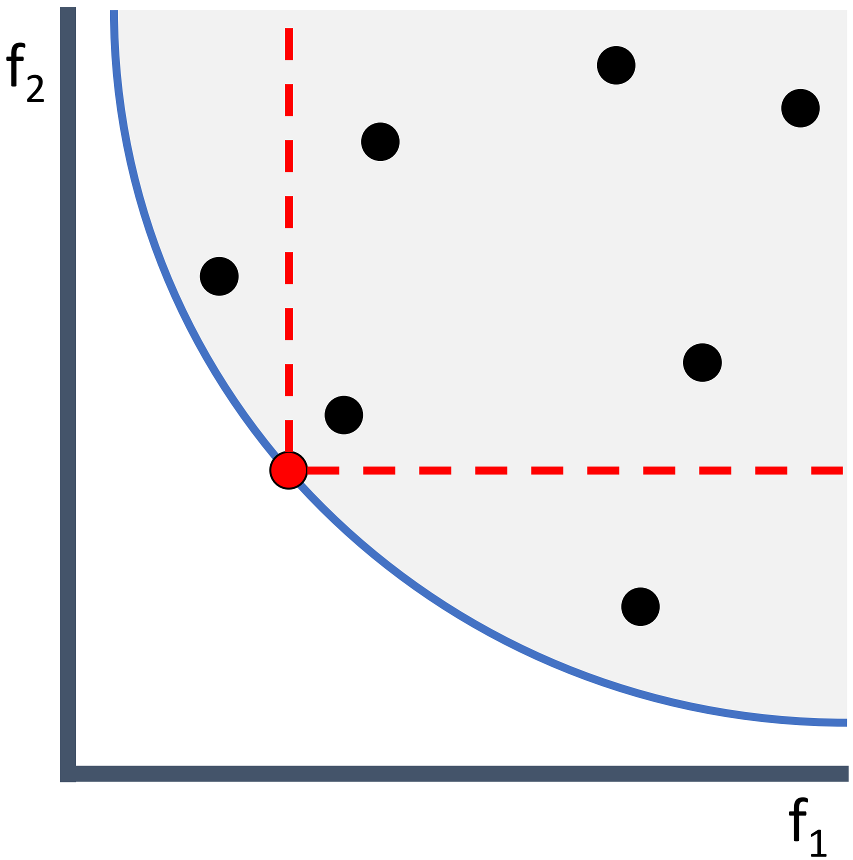
policy

cost features f_1, f_2, ...

human demonstrations

Pareto frontier

Superhuman Percentile & Subdominance

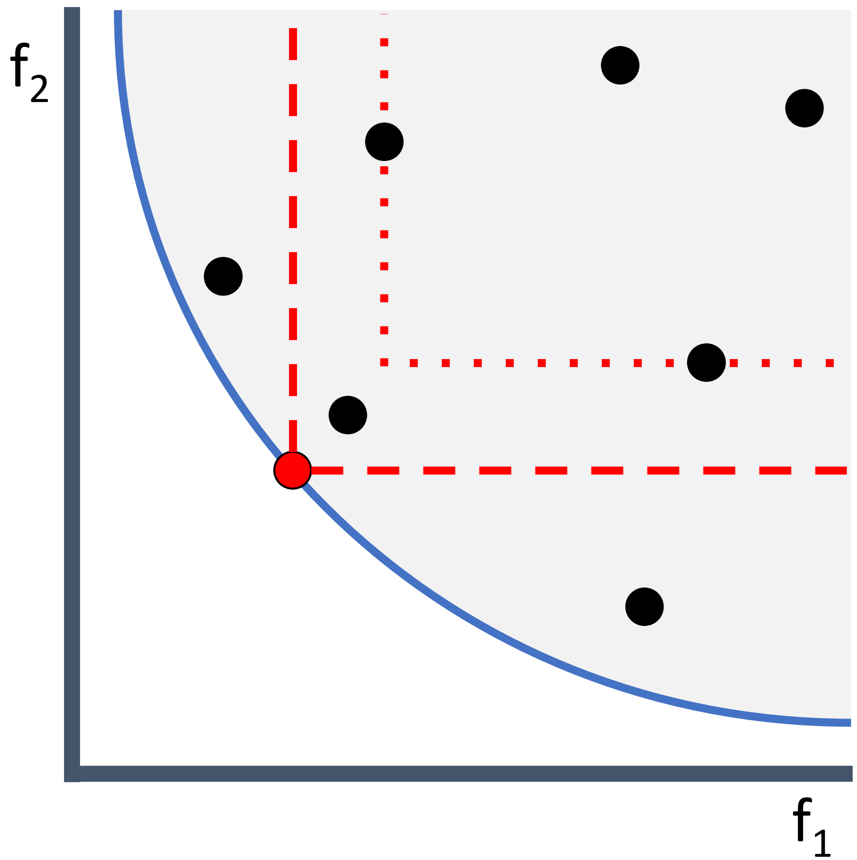


A **policy** is **γ-superhuman** if it has smaller **cost features** f_1, f_2, … than γ% of **human demonstrations**

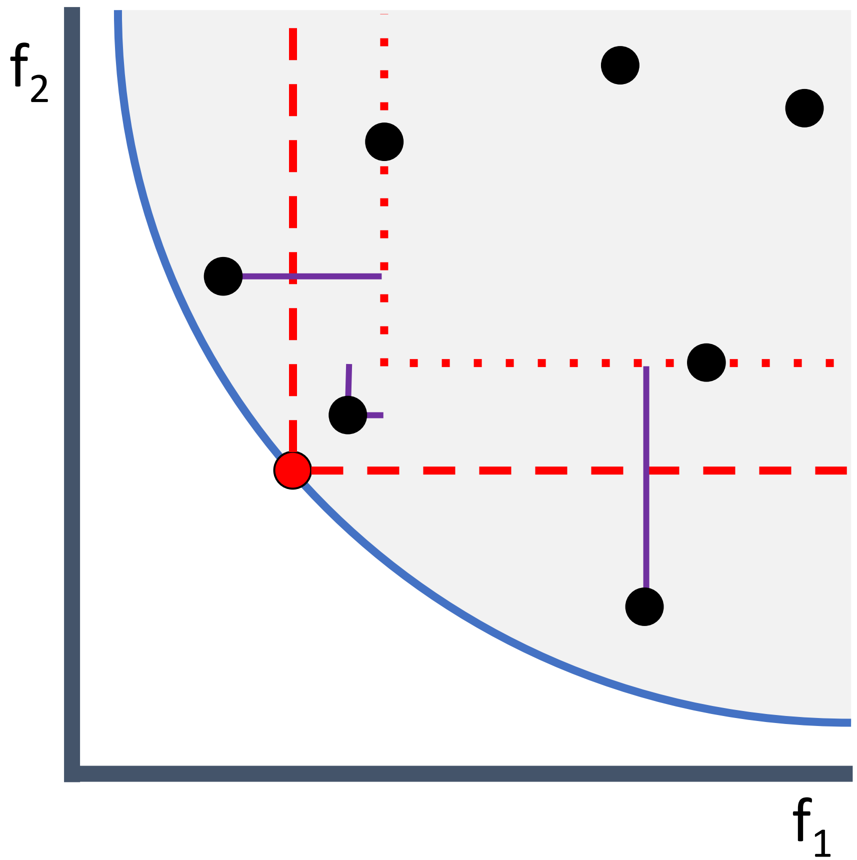
Pareto frontier

Superhuman Percentile & Subdominance

A **policy** is **γ-superhuman** if it has smaller **cost features** f_1, f_2, … than γ% of **human demonstrations**



Superhuman Percentile & Subdominance

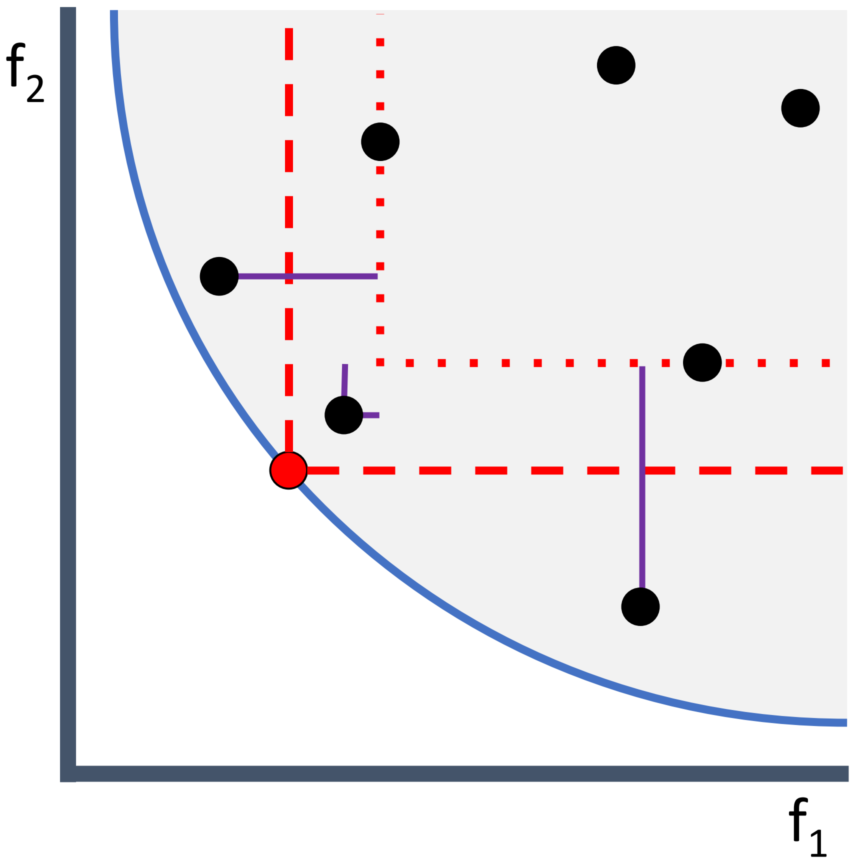


A **policy** is **γ-superhuman** if it has smaller **cost features** $f_1, f_2, \ldots$ than γ% of **human demonstrations**

Subdominance measures how far a policy is from superhuman by some **margins**

Pareto frontier

Superhuman Percentile & Subdominance

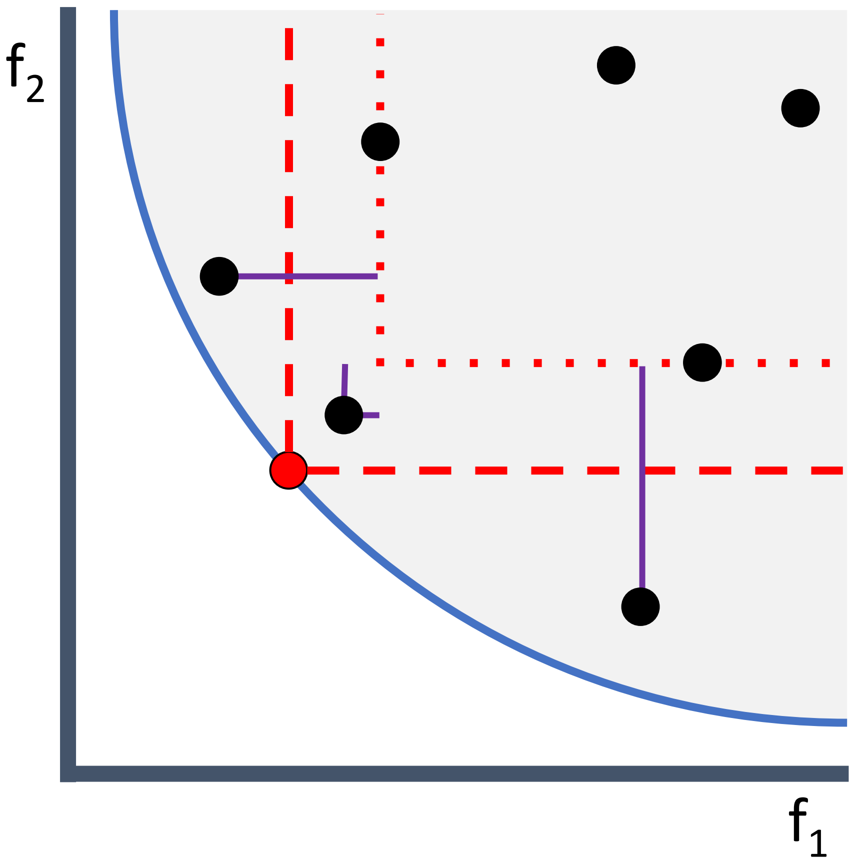


A **policy** is **γ-superhuman** if it has smaller **cost features** $f_1, f_2, \ldots$ than γ% of **human demonstrations**

Subdominance measures how far a policy is from superhuman by some **margins**

Minimum Subdominance Inverse Optimal Control seeks policies on the **Pareto frontier** minimizing this

Superhuman Percentile & Subdominance



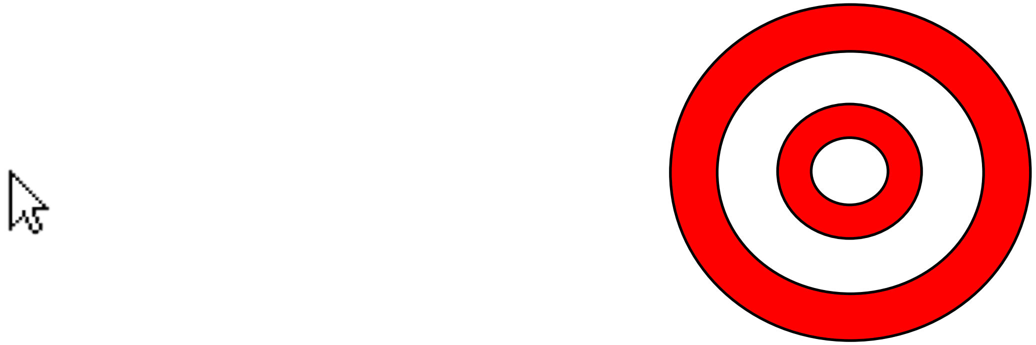
A **policy** is **γ-superhuman** if it has smaller **cost features** $f_1, f_2, \ldots$ than γ% of **human demonstrations**

Subdominance measures how far a policy is from superhuman by some **margins**

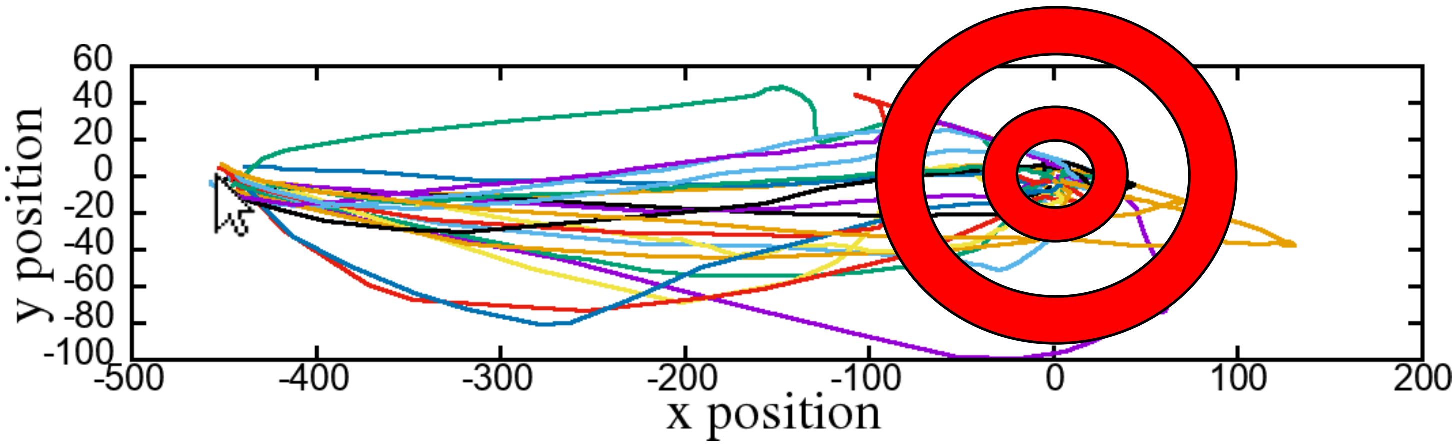
Minimum Subdominance Inverse Optimal Control seeks policies on the **Pareto frontier** minimizing this

Subdominance bounds the **superhuman percentile**

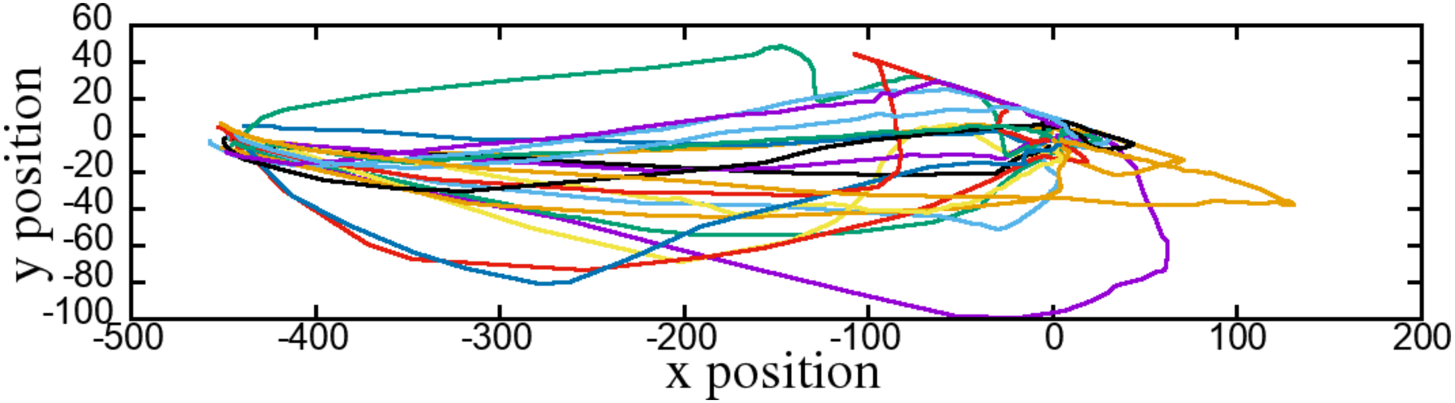
Cursor Pointing Task



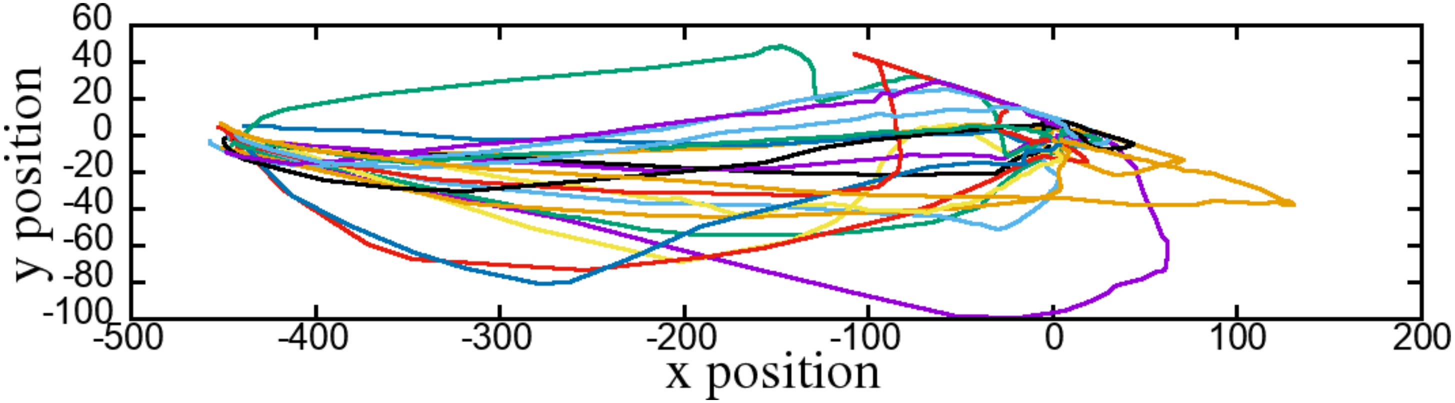
Cursor Pointing Task

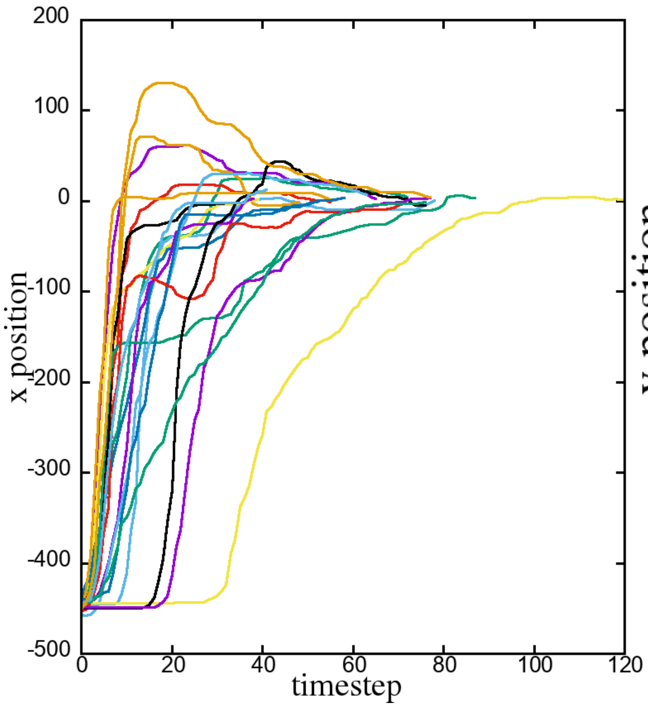


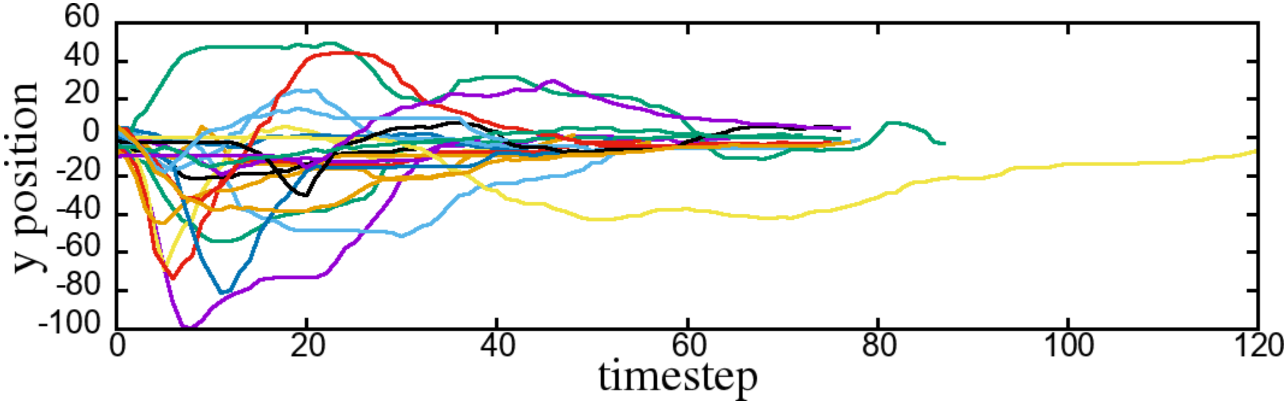
Cursor Pointing Task



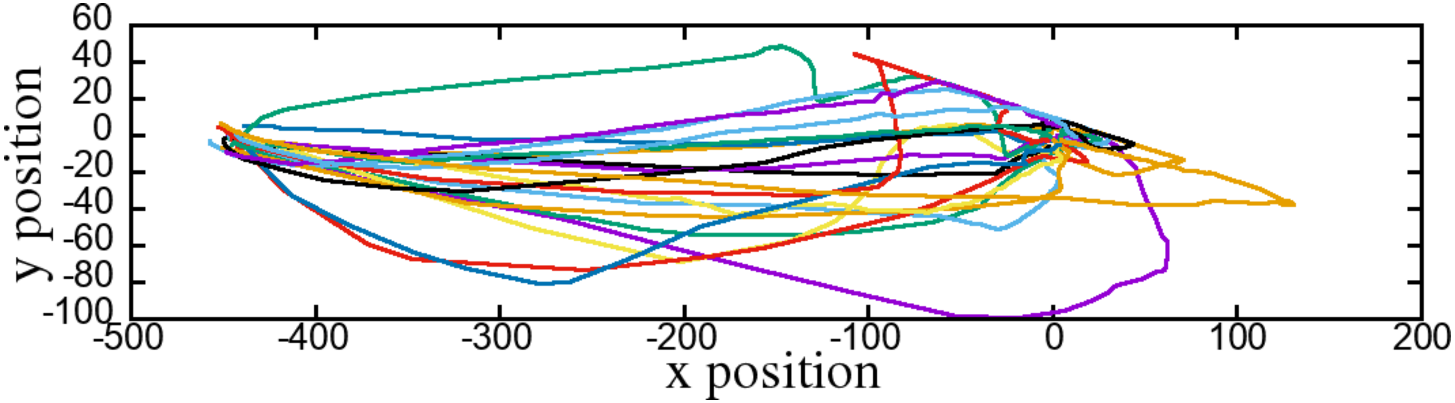
Cursor Pointing Task

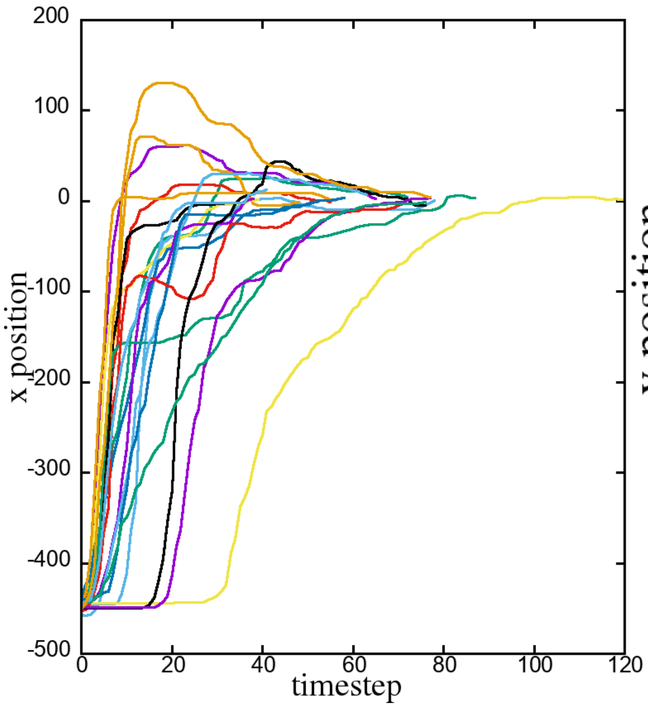


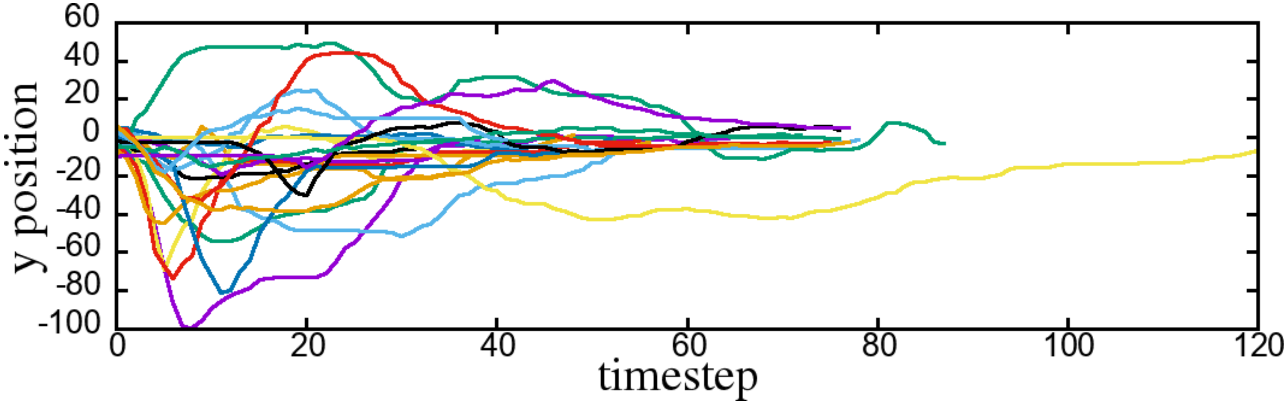




Cursor Pointing Task

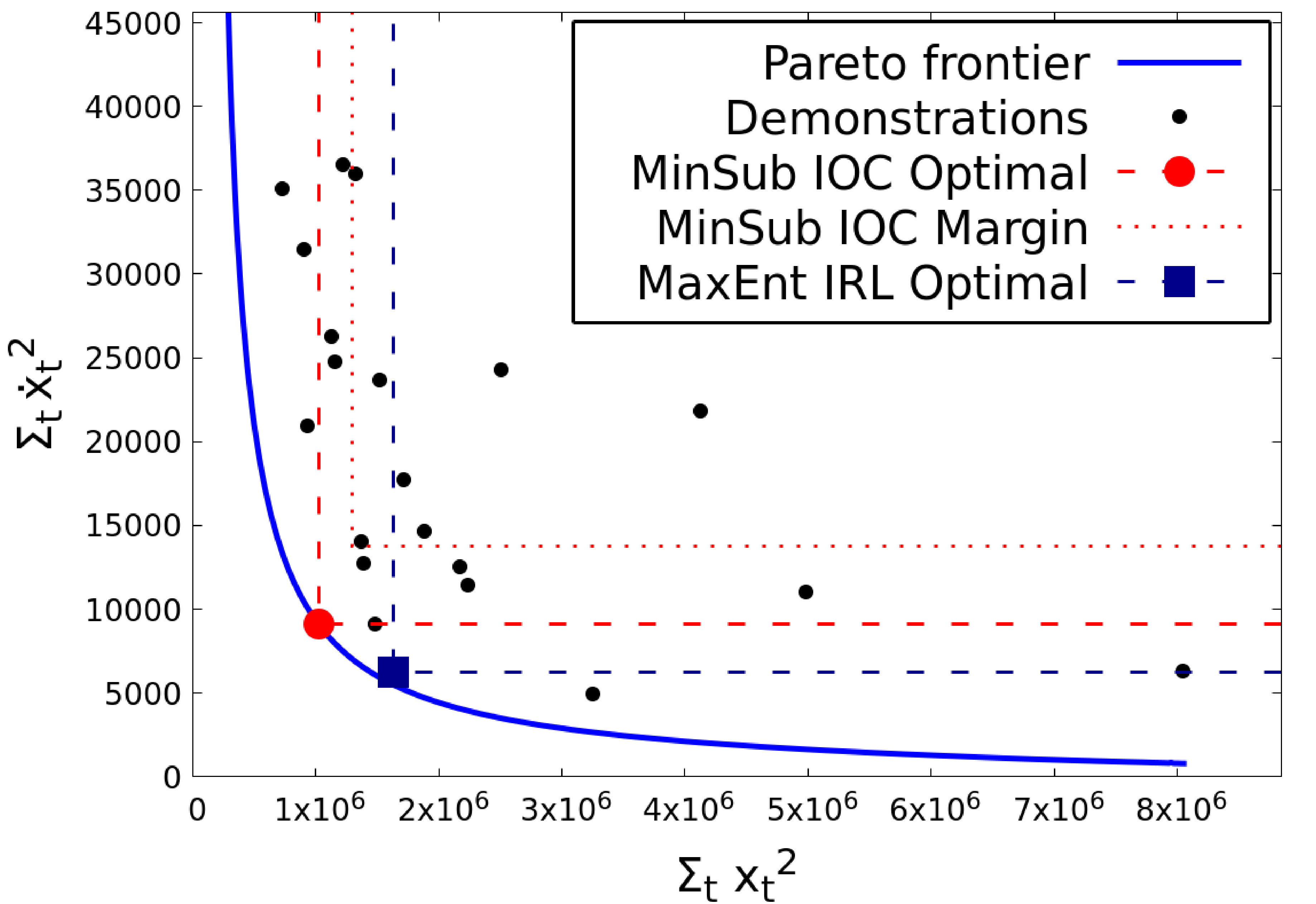




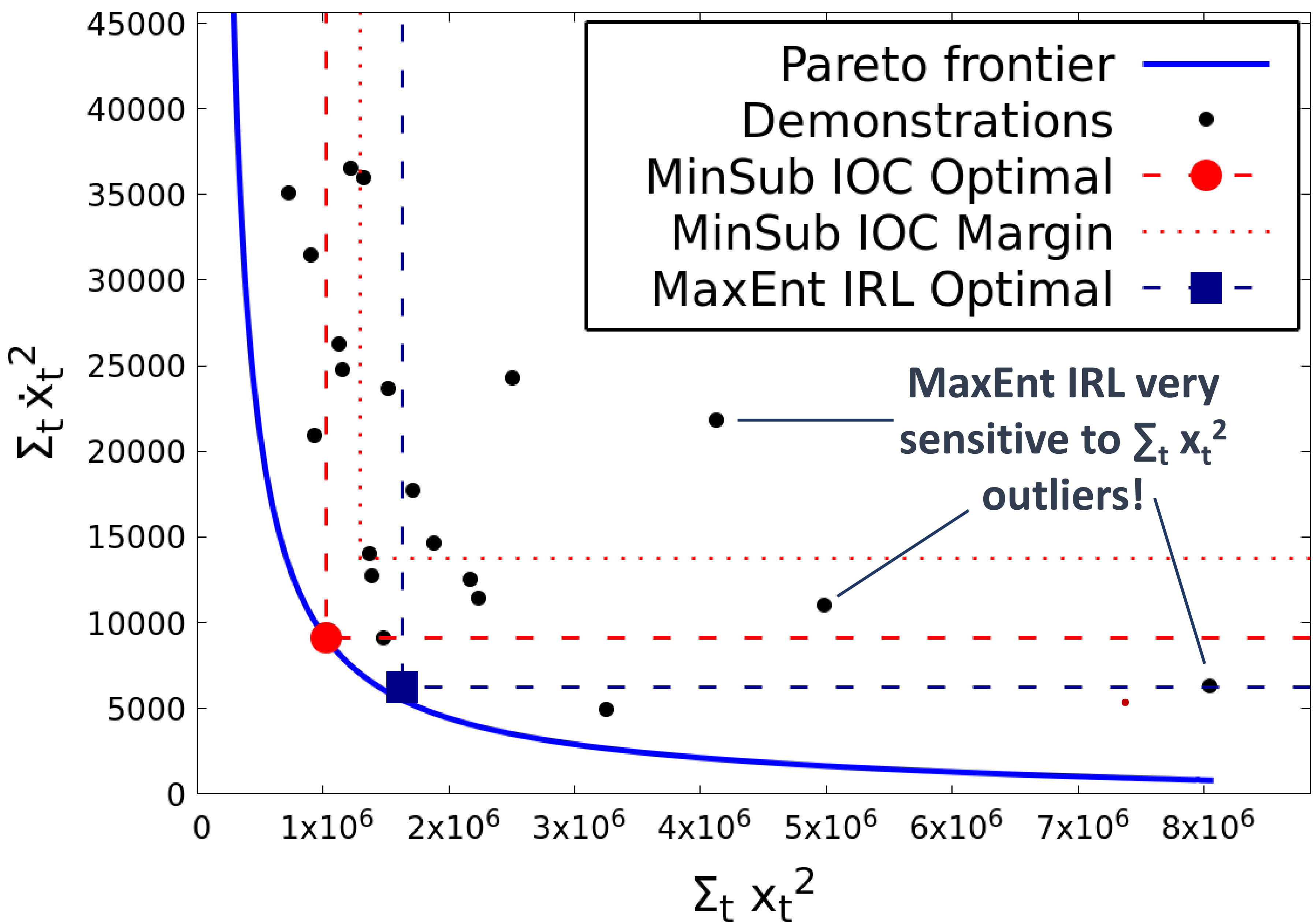


Linear-quadratic regulation formulation:

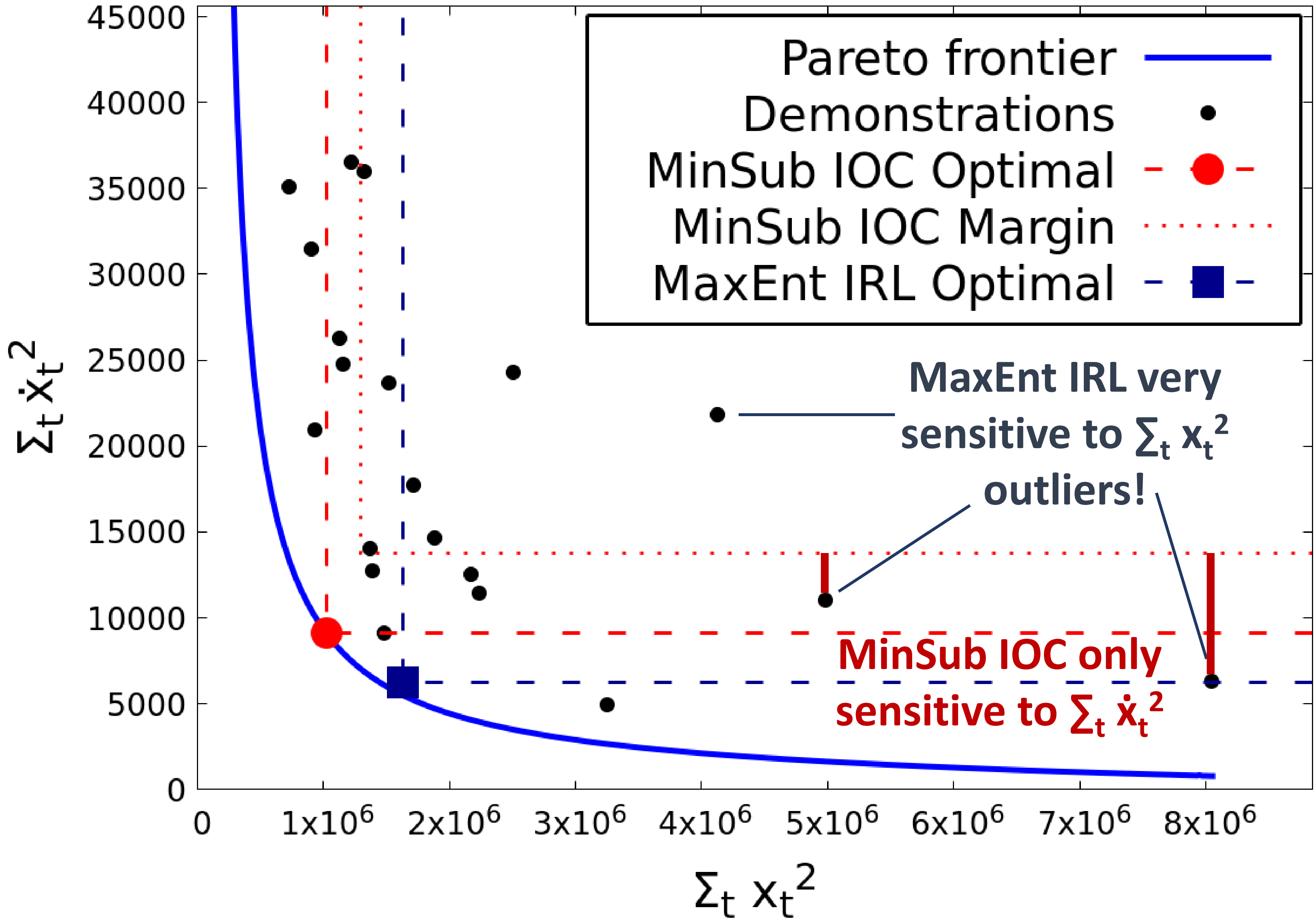
$$\mathrm{Cost}(s_t) = \alpha_{x,x}\, x_t^2 + \alpha_{\dot{x},\dot{x}}\, \dot{x}_t^2 + \alpha_{\ddot{x},\ddot{x}}\, \ddot{x}_t^2 + \ldots$$



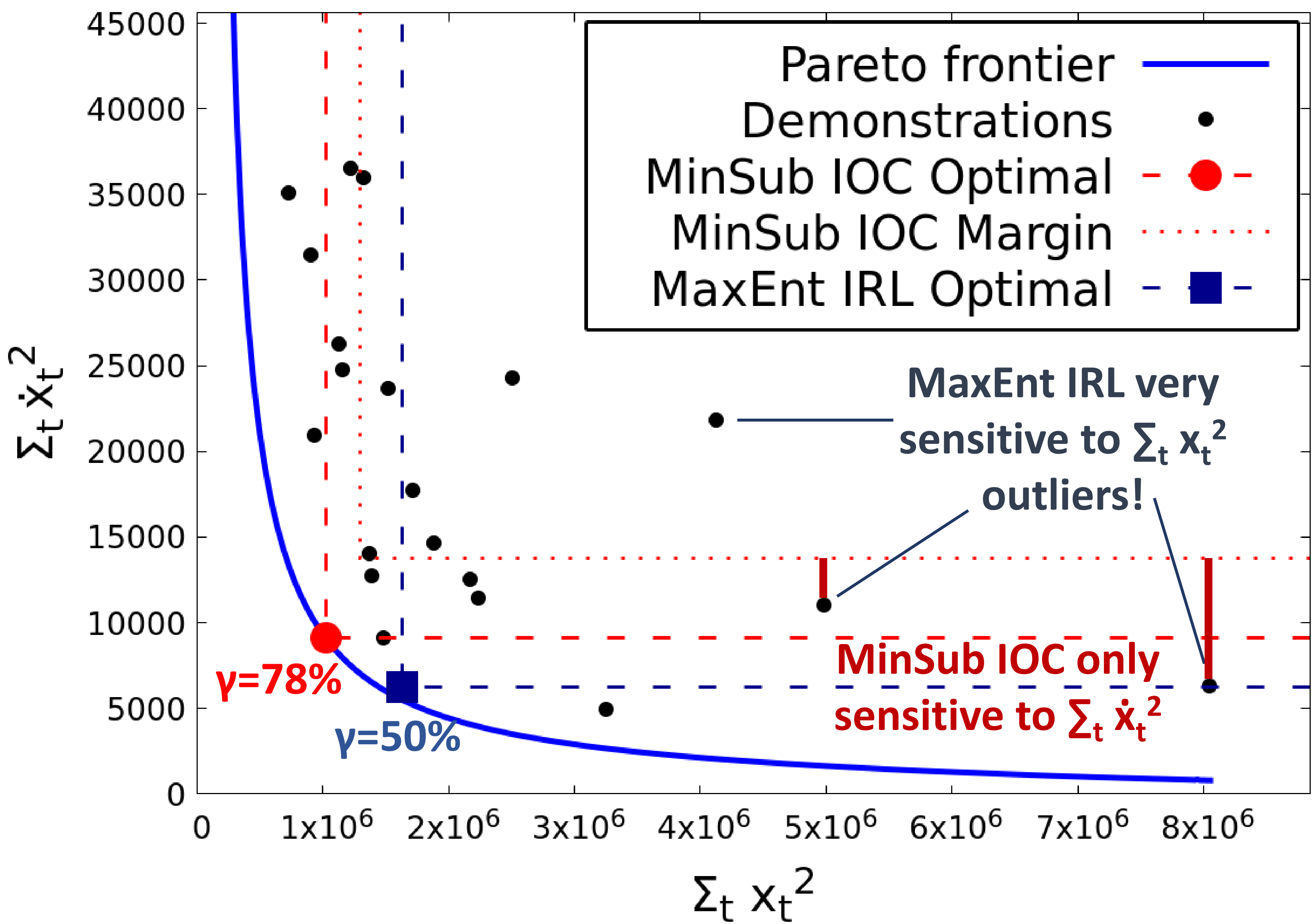
Pareto frontier
Demonstrations
MinSub IOC Optimal
MinSub IOC Margin
MaxEnt IRL Optimal
45000
40000
35000
30000
25000
20000
15000
10000
5000
0
0
1x10^6
2x10^6
3x10^6
4x10^6
5x10^6
6x10^6
7x10^6
8x10^6
$\Sigma_t \dot{x}_t^2$
$\Sigma_t x_t^2$



Pareto frontier
Demonstrations
MinSub IOC Optimal
MinSub IOC Margin
MaxEnt IRL Optimal
MaxEnt IRL very
sensitive to $\sum_t x_t^2$
outliers!
45000
40000
35000
30000
25000
20000
15000
10000
5000
0
$\sum_t \dot{x}_t^2$
0
1×10^6
2×10^6
3×10^6
4×10^6
5×10^6
6×10^6
7×10^6
8×10^6
$\sum_t x_t^2$

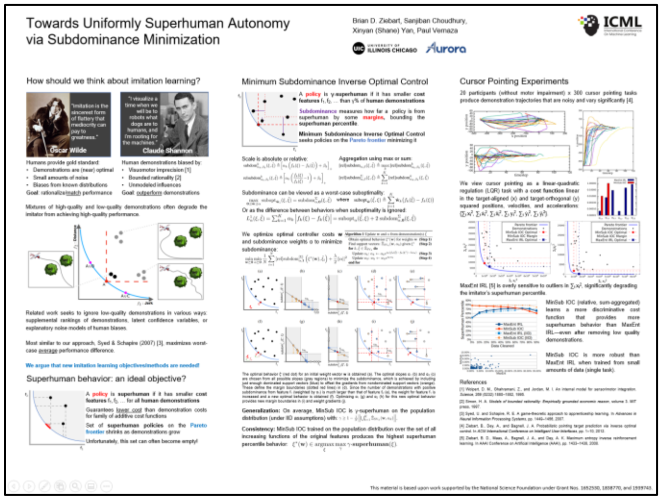


Pareto frontier
Demonstrations
MinSub IOC Optimal
MinSub IOC Margin
MaxEnt IRL Optimal
MaxEnt IRL very sensitive to $\sum_t x_t^2$ outliers!
MinSub IOC only sensitive to $\sum_t \dot{x}_t^2$
$\sum_t \dot{x}_t^2$
$\sum_t x_t^2$
0
5000
10000
15000
20000
25000
30000
35000
40000
45000
0
1x10^6
2x10^6
3x10^6
4x10^6
5x10^6
6x10^6
7x10^6
8x10^6



Pareto frontier
Demonstrations
MinSub IOC Optimal
MinSub IOC Margin
MaxEnt IRL Optimal
MaxEnt IRL very sensitive to $\sum_t x_t^2$ outliers!
MinSub IOC only sensitive to $\sum_t \dot{x}_t^2$
γ=78%
γ=50%
$\sum_t \dot{x}_t^2$
$\sum_t x_t^2$
45000
40000
35000
30000
25000
20000
15000
10000
5000
0
1x10^6
2x10^6
3x10^6
4x10^6
5x10^6
6x10^6
7x10^6
8x10^6

And much more…



- Relationships to suboptimality
- SVM analogies
- Consistency/generalization
- Cleaning/noise experiments

Poster: Hall E #827