



ICML
International Conference
On Machine Learning

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

Gold standard human demonstrations

(Near) Optimal, minimum noise, known biases

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

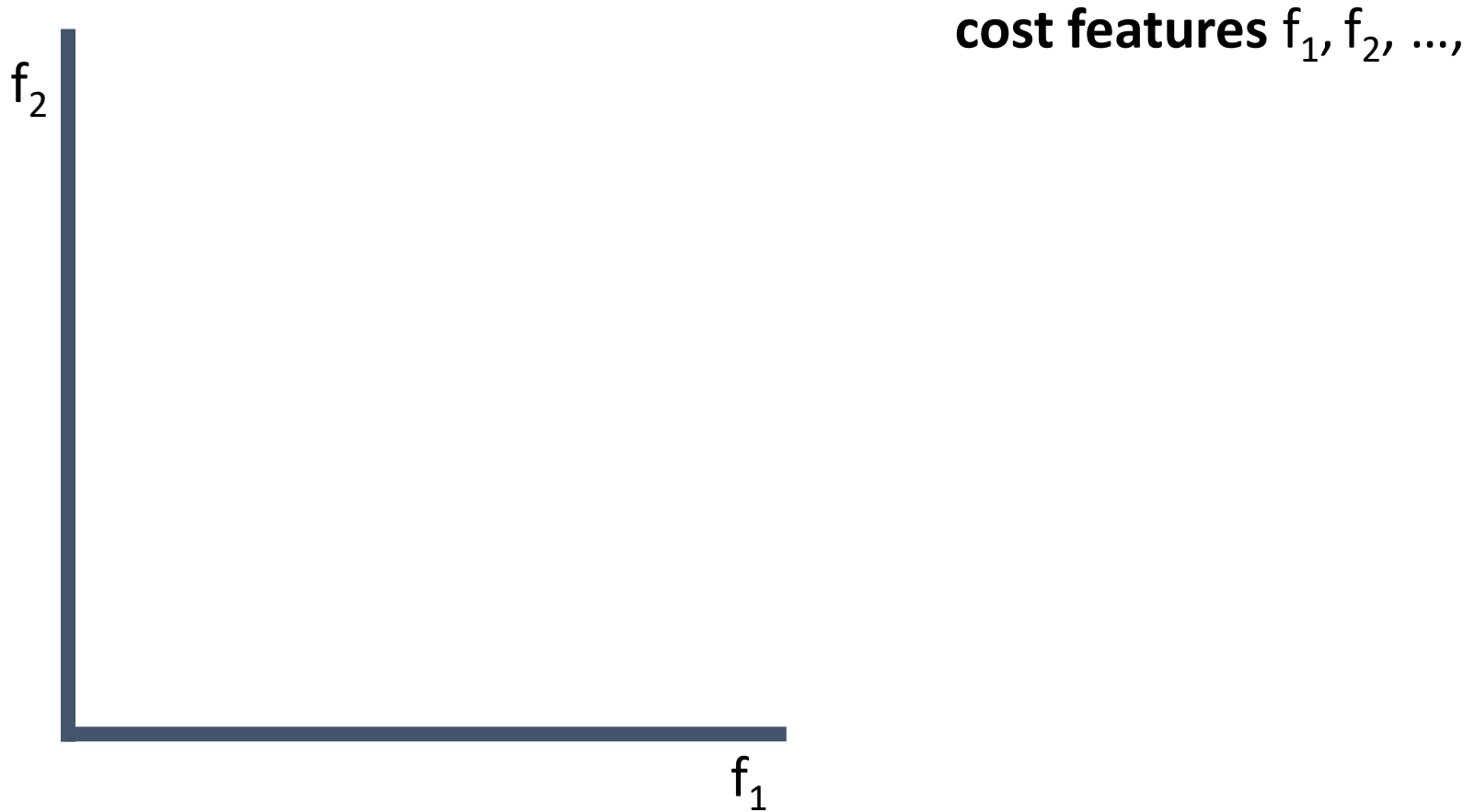
Oscar Wilde

Gold standard human demonstrations

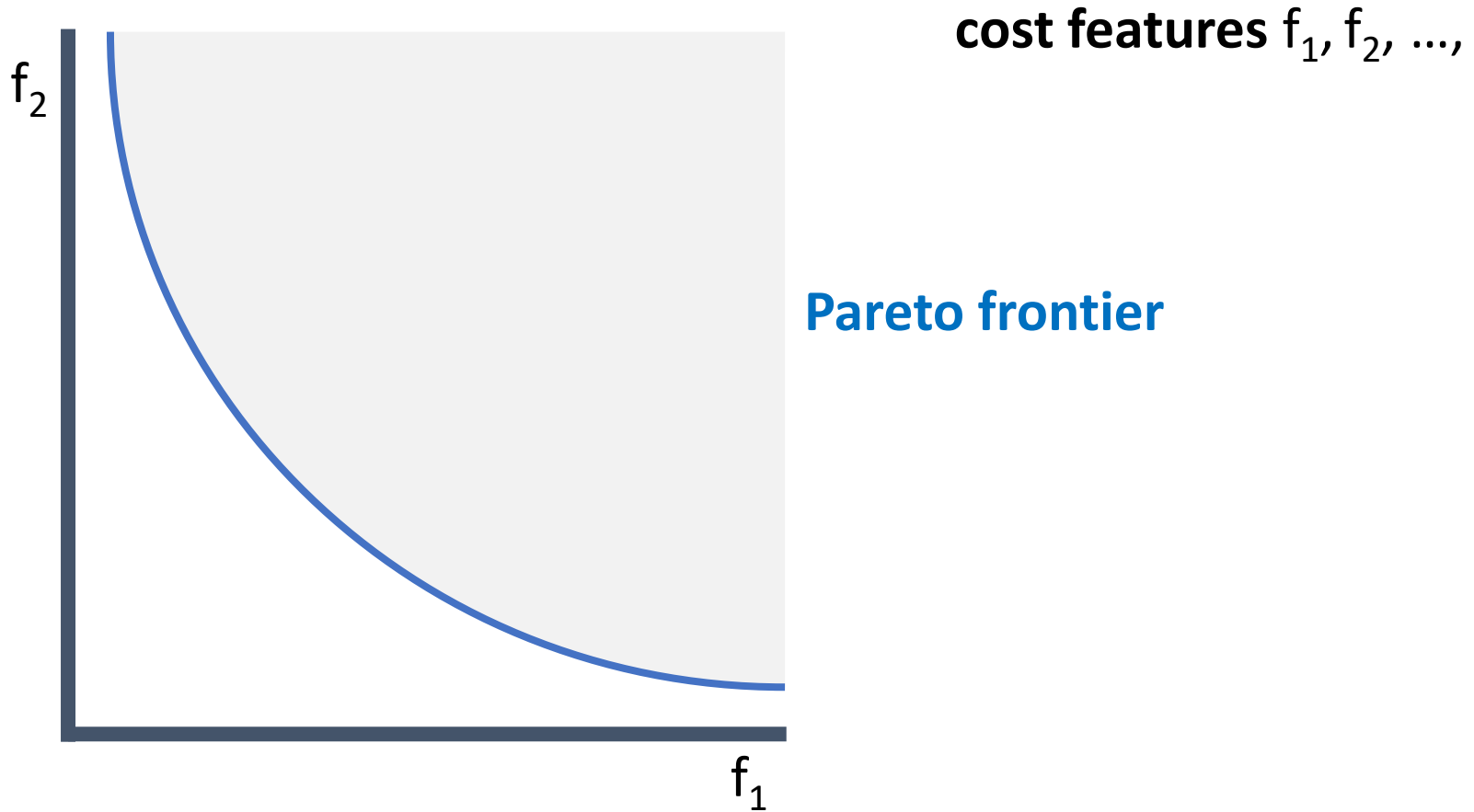
(Near) Optimal, minimum noise, known biases

Formulation: Rationalize/match performance

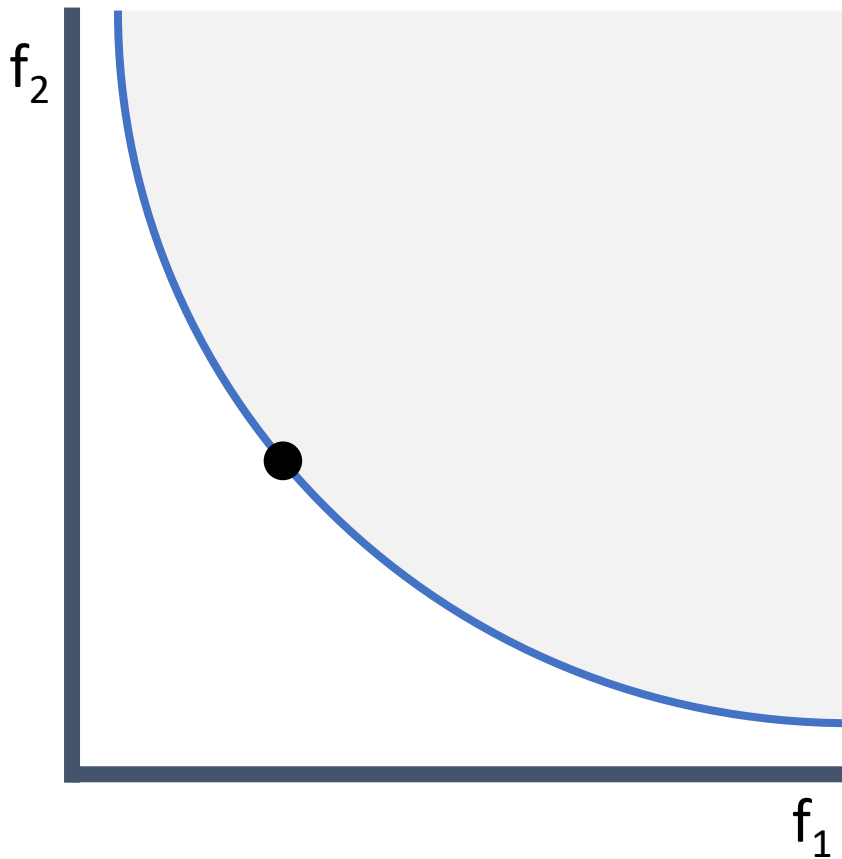
Easy: Optimal demonstrations



Easy: Optimal demonstrations



Easy: Optimal demonstrations

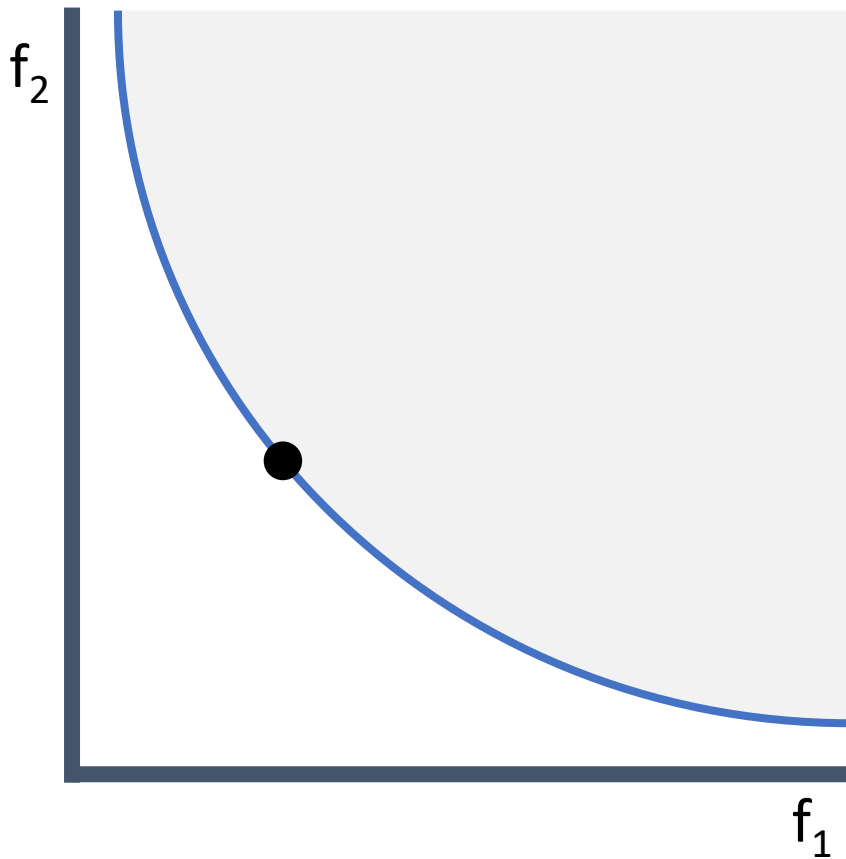


cost features f_1, f_2, \dots ,

human demonstration(s),

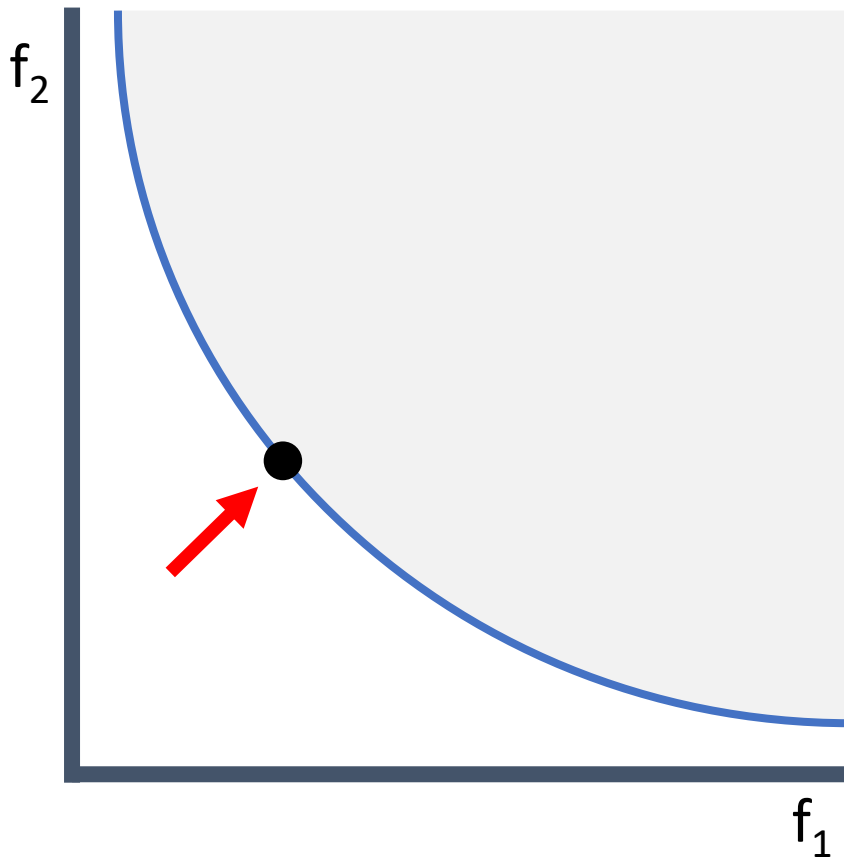
Pareto frontier

Easy: Optimal demonstrations



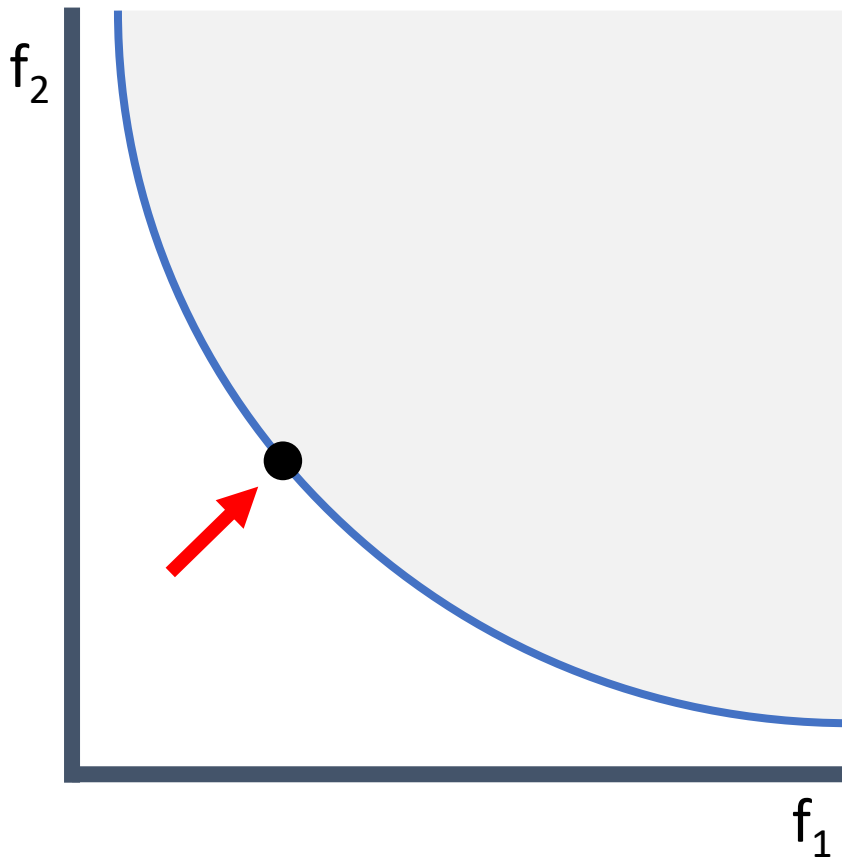
Given **cost features** f_1, f_2, \dots , learn weights **w** that make **human demonstration(s)**, which must reside on the **Pareto frontier**, optimal.

Easy: Optimal demonstrations



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

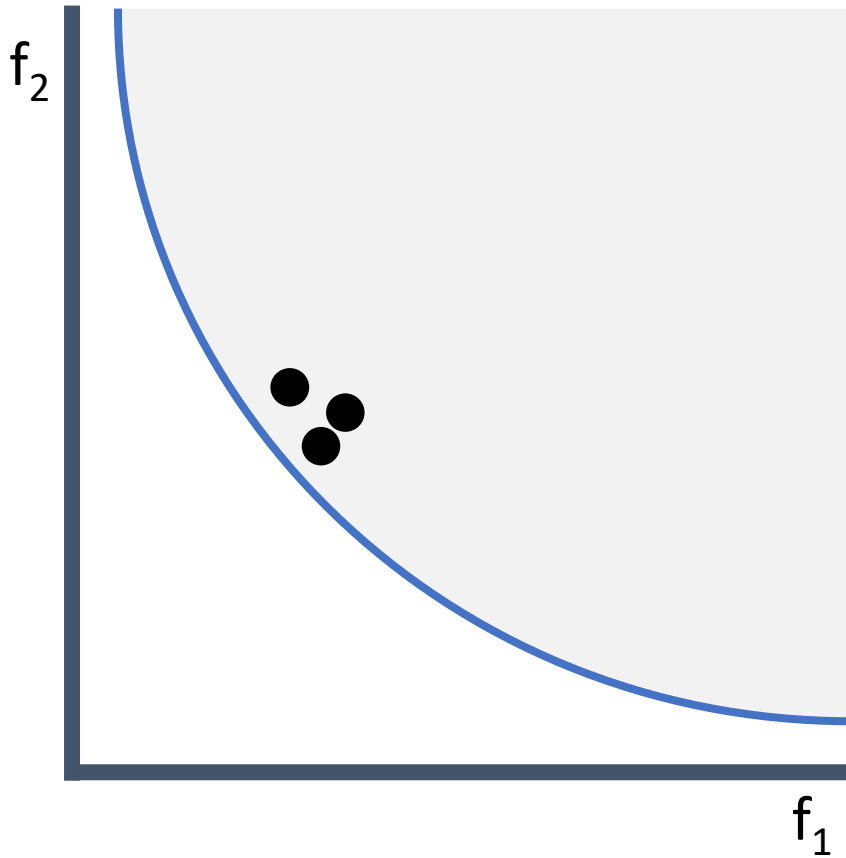
Easy: Optimal demonstrations



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

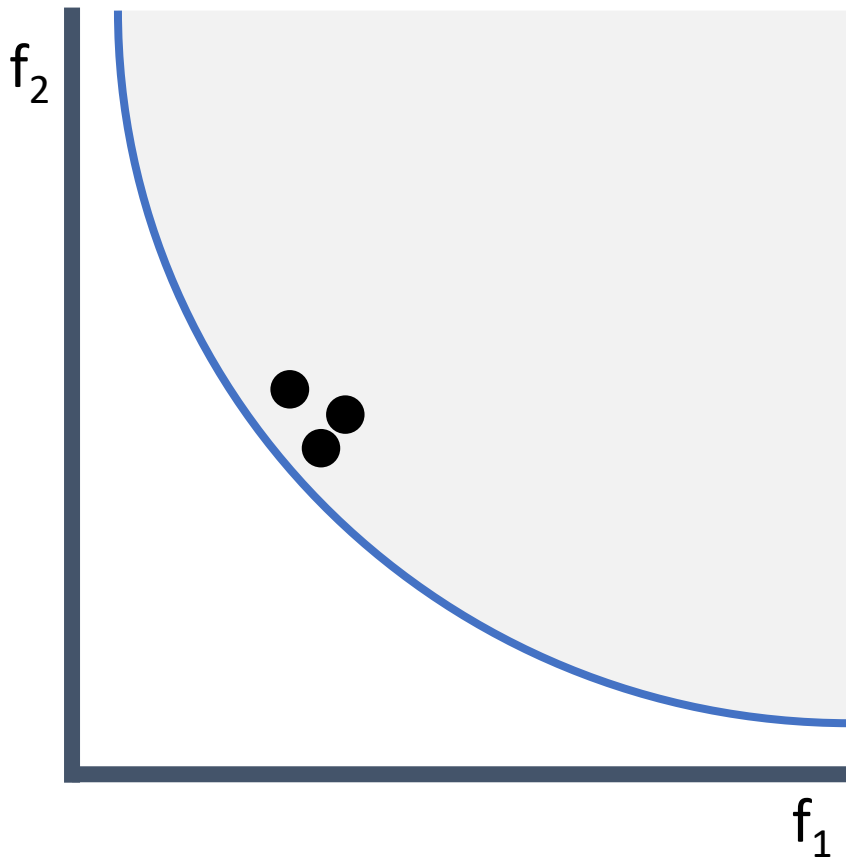
Degenerate solutions (**w=0**) exist, but can be avoided (Ng & Russell 2000; Ratliff, Bagnell, Zinkevich 2006)

Harder: Suboptimal demonstrations

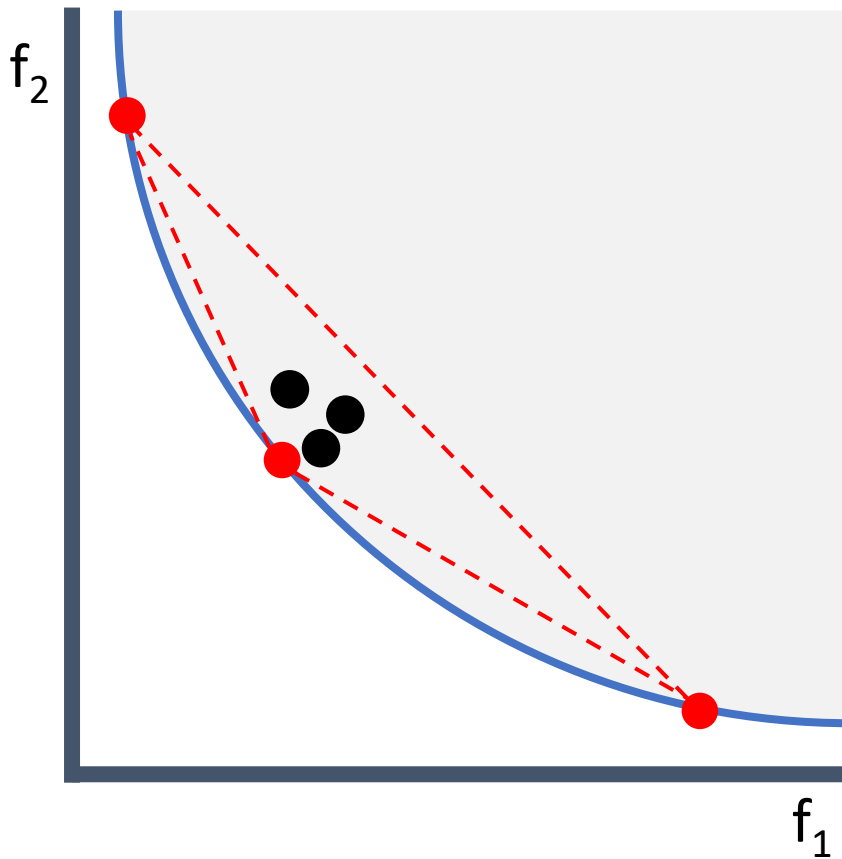


Harder: Suboptimal demonstrations

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



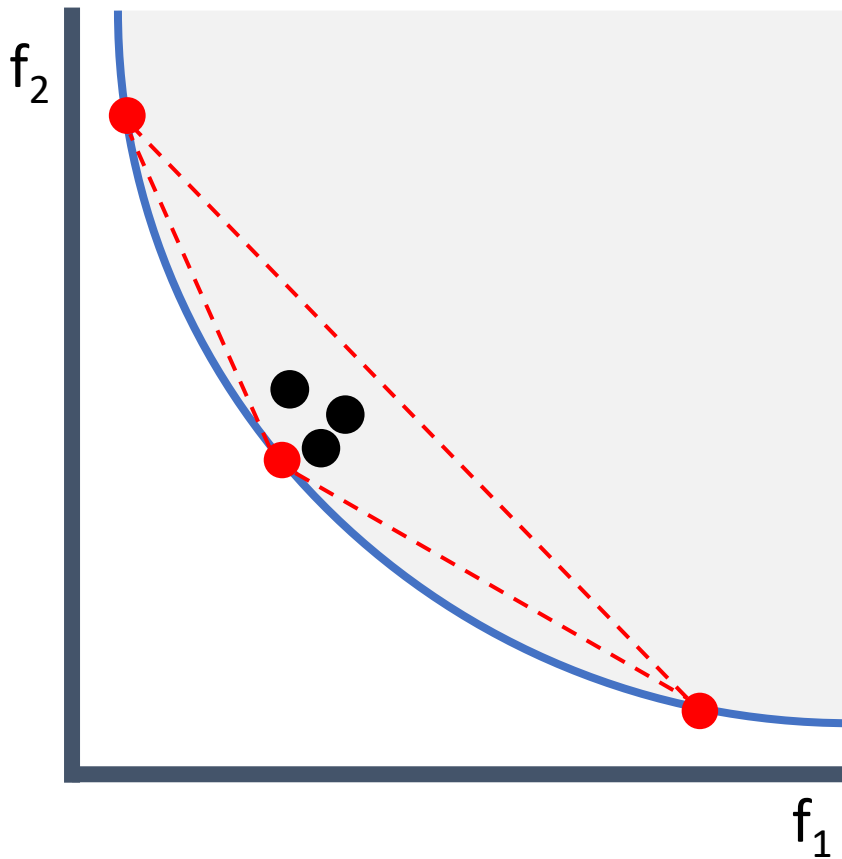
Harder: Suboptimal demonstrations



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



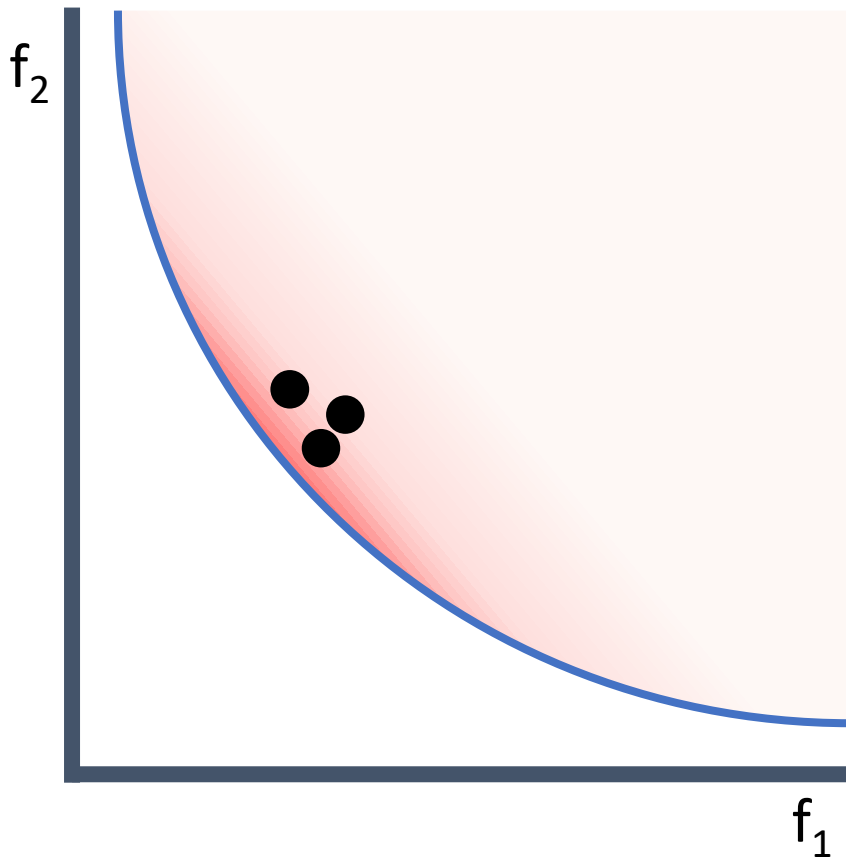
Feature Matching (Abbeel & Ng 2004): Matching expected features guarantees equal expected cost/reward (assuming linearity).

Approach: Mix optimal policies to match features.

Limitations: Many solutions exist; which policy to deploy?

Harder: Suboptimal demonstrations

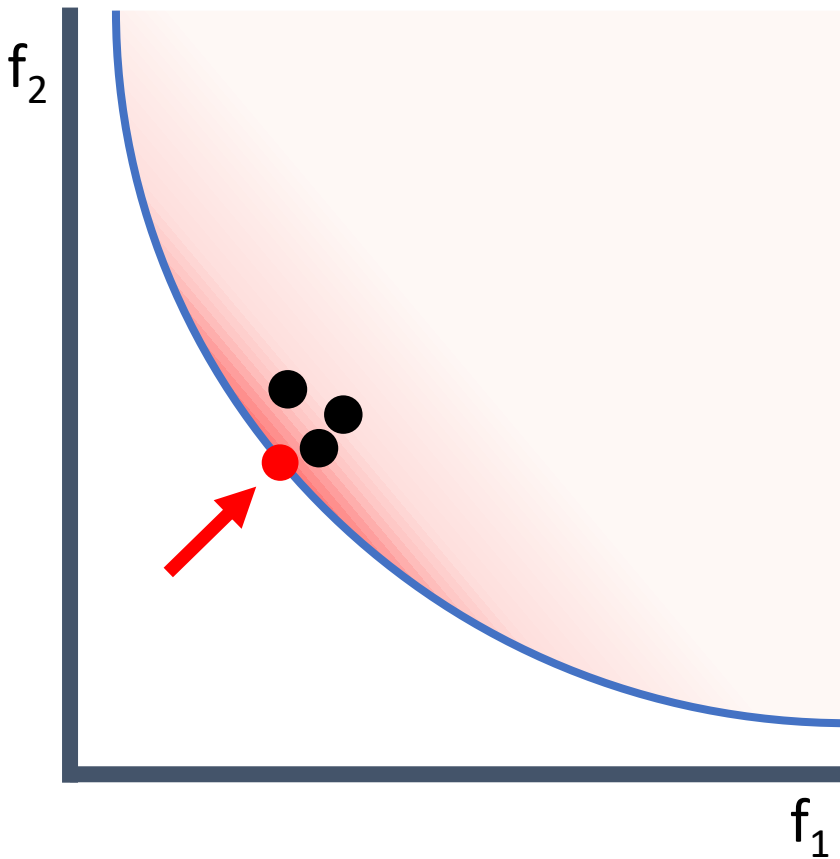
MaxEnt IRL (Ziebart et al. 2008):
Demonstrations are noisy with
probability $\propto e^{-w \cdot f}$



Harder: Suboptimal demonstrations

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

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

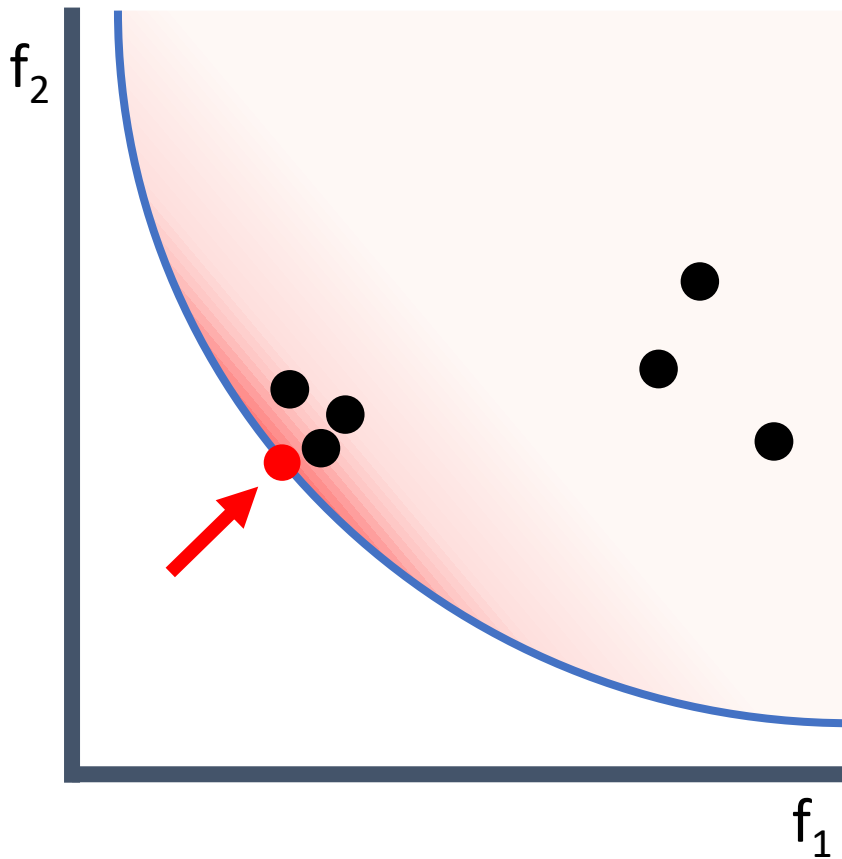


Harder: Suboptimal demonstrations

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

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

Outliers violating noise model
can significantly shift the mode.

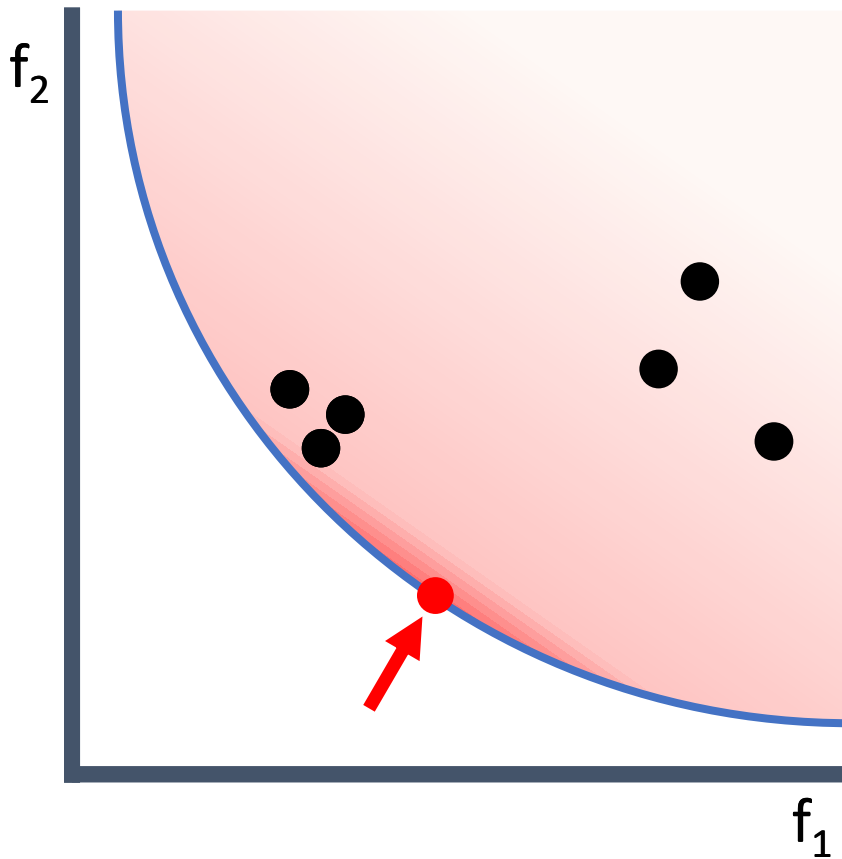


Harder: Suboptimal demonstrations

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

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

Outliers violating noise model
can significantly shift the mode.



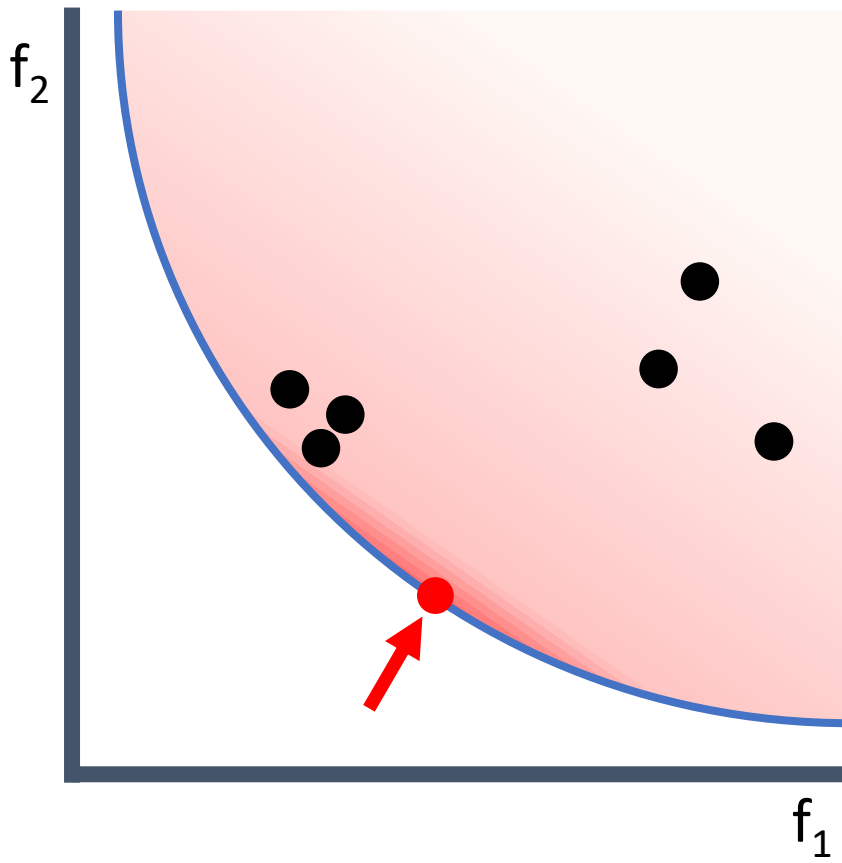
Harder: Suboptimal demonstrations

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

Apprenticeship learning:
estimate \mathbf{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; Büyük 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



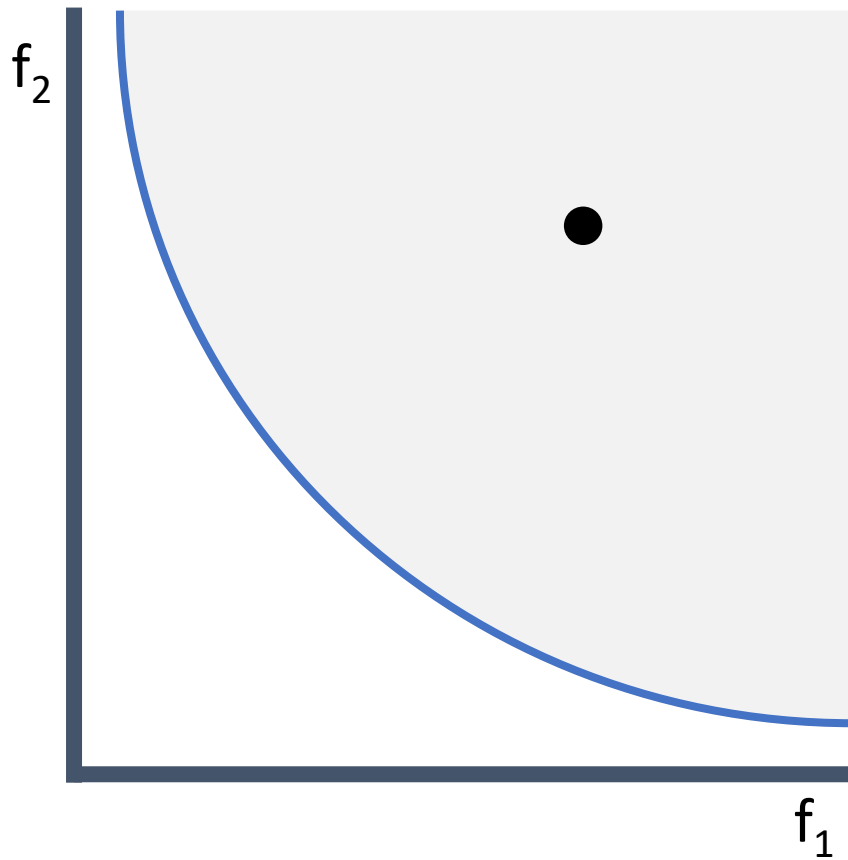
“I visualize a time when we will be to robots what dogs are to humans, and I'm rooting for the machines.”

Claude Shannon

Visuomotor imprecision
(Wolpert et al. 1995)

Bounded rationality
(Simon, 1997)

Defining Superhuman Behavior

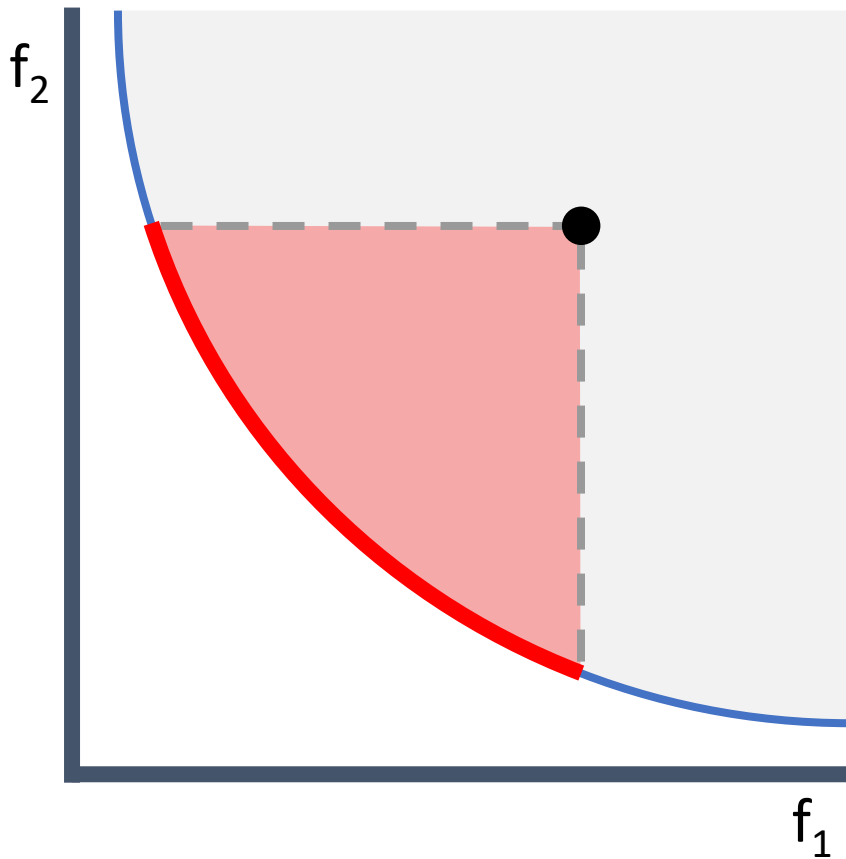


cost features f_1, f_2, \dots
human demonstrations

Pareto frontier

Defining Superhuman Behavior

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



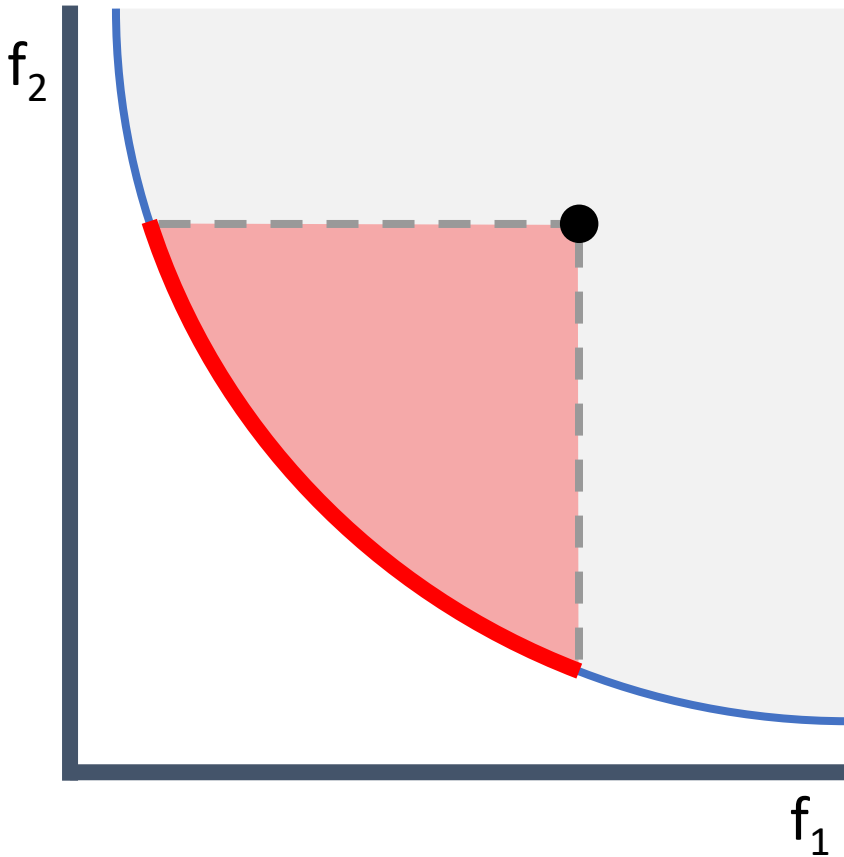
Pareto frontier

Defining Superhuman Behavior

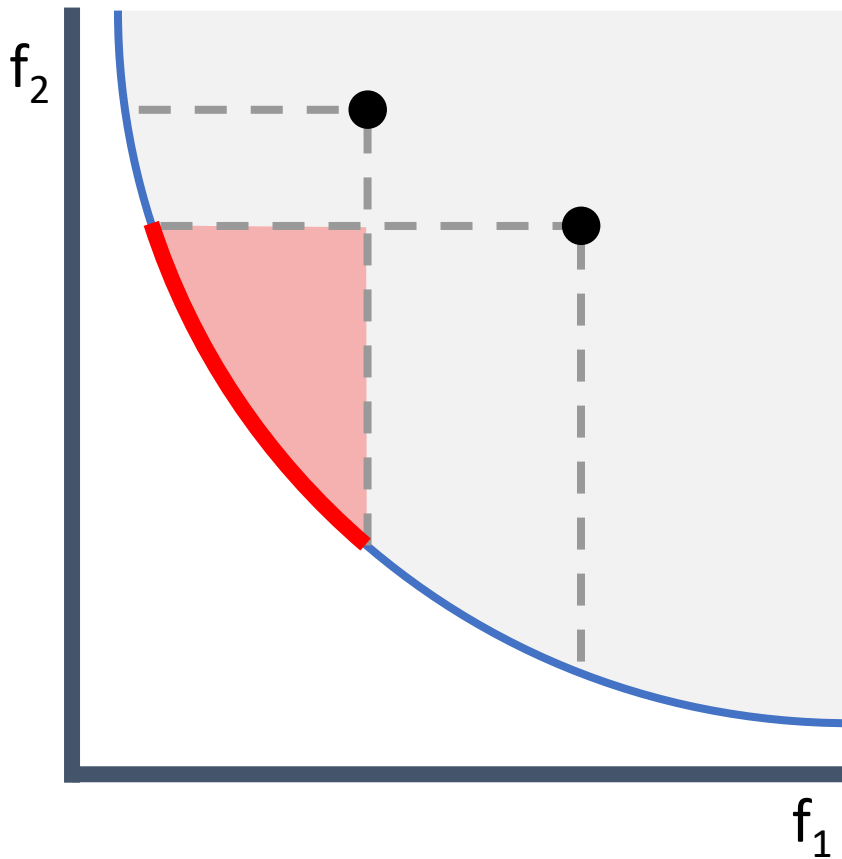
A **policy** is **superhuman** if it has smaller **cost features** f_1, f_2, \dots for all **human demonstrations**

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

Pareto frontier



Defining Superhuman Behavior

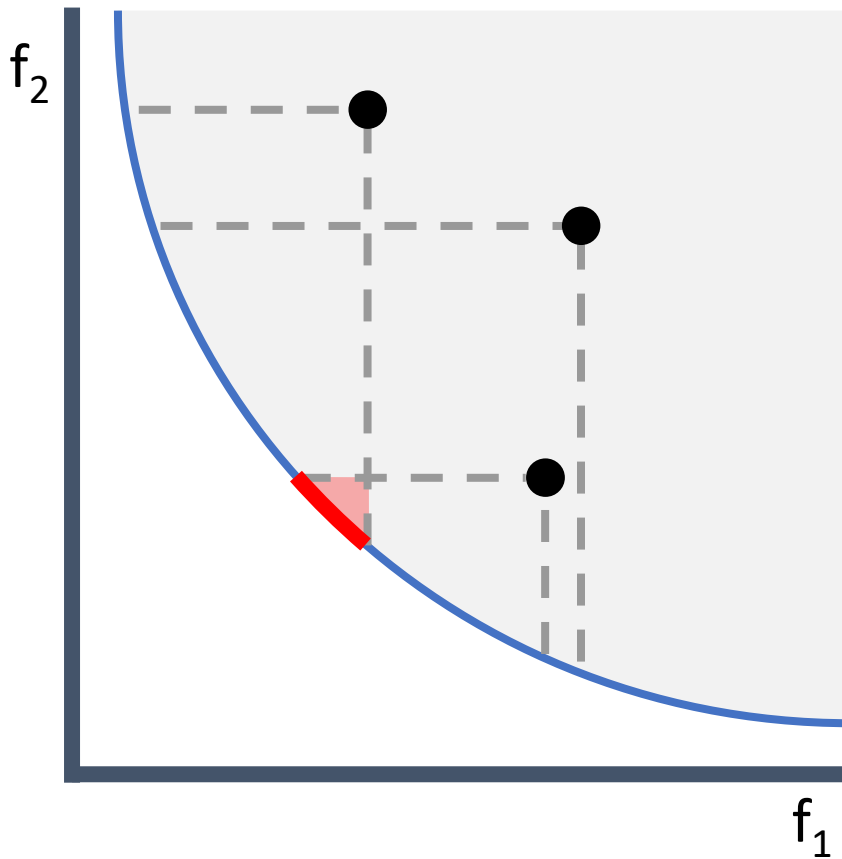


A **policy** is **superhuman** if it has smaller **cost features** f_1, f_2, \dots for all **human demonstrations**

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

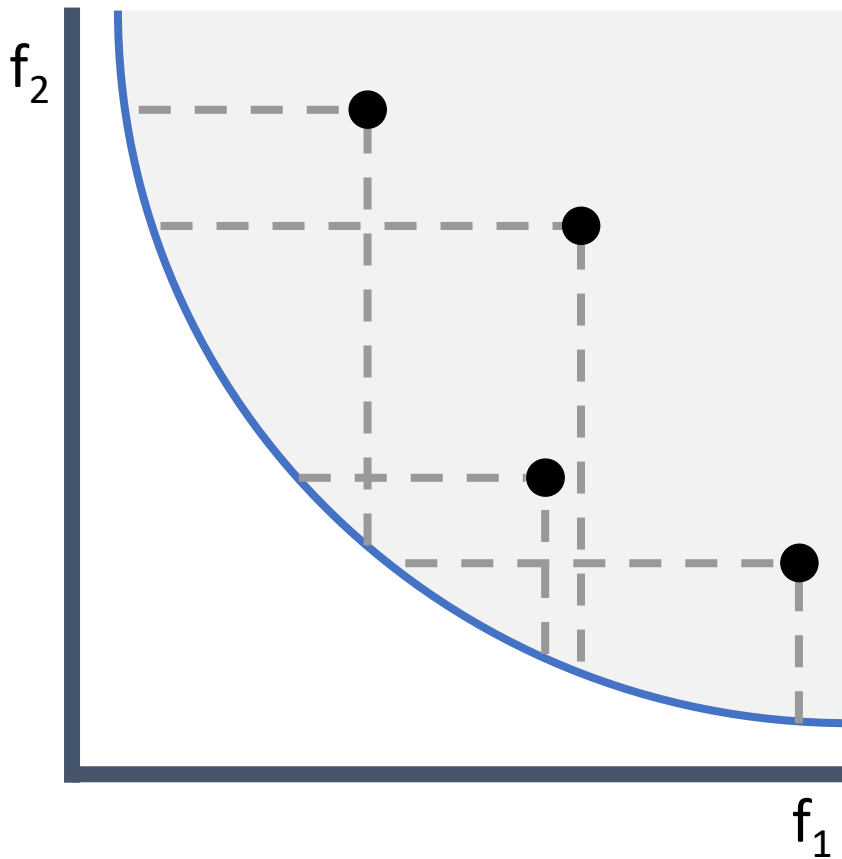


A **policy** is **superhuman** if it has smaller **cost features** f_1, f_2, \dots for all **human demonstrations**

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



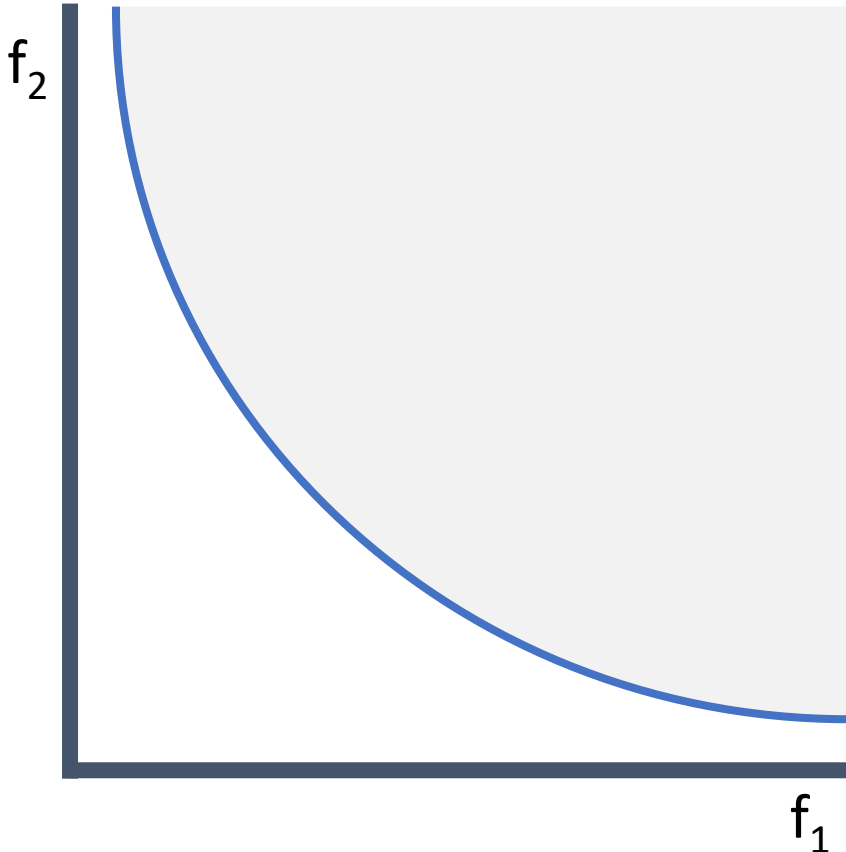
A **policy** is **superhuman** if it has smaller **cost features** f_1, f_2, \dots for all **human demonstrations**

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

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

Can become empty!

Superhuman Percentile & Subdominance



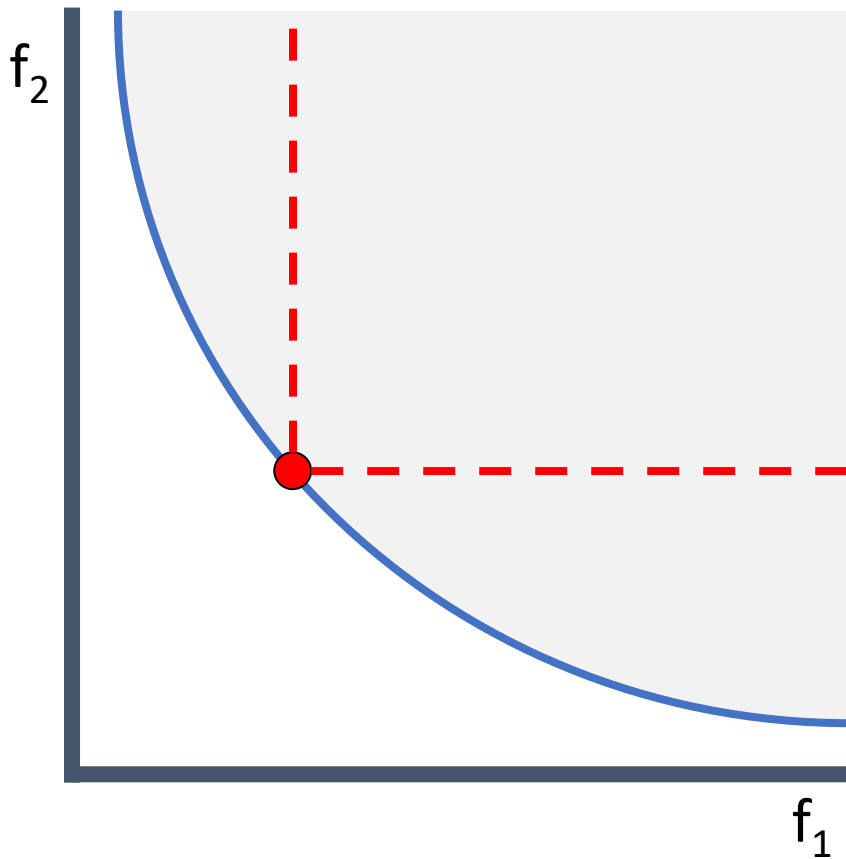
cost features f_1, f_2, \dots

Pareto frontier

Superhuman Percentile & Subdominance

policy

cost features f_1, f_2, \dots

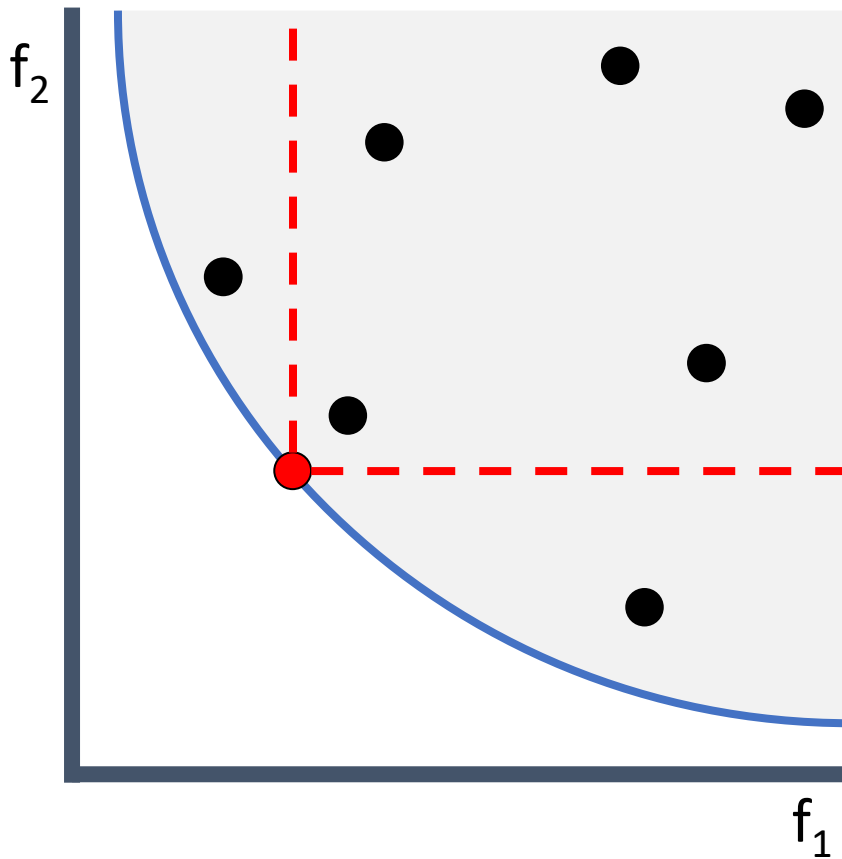


Pareto frontier

Superhuman Percentile & Subdominance

policy

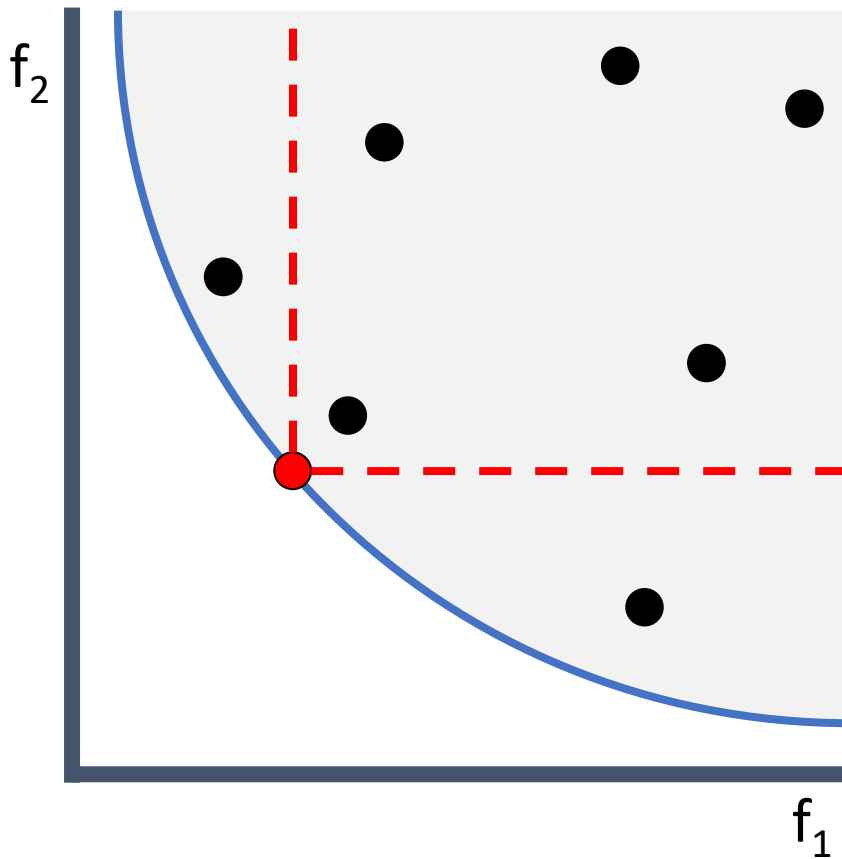
**cost features f_1, f_2, \dots
human demonstrations**



Pareto frontier

Superhuman Percentile & Subdominance

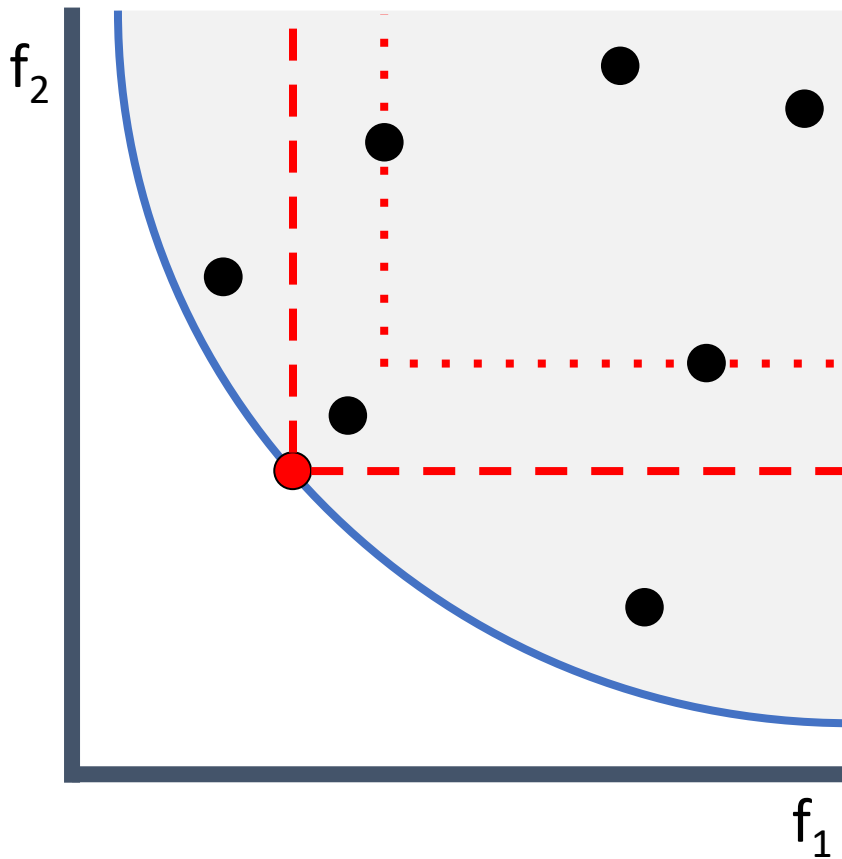
A **policy** is γ -superhuman if it has smaller **cost features** f_1, f_2, \dots than $\gamma\%$ of **human demonstrations**



Pareto frontier

Superhuman Percentile & Subdominance

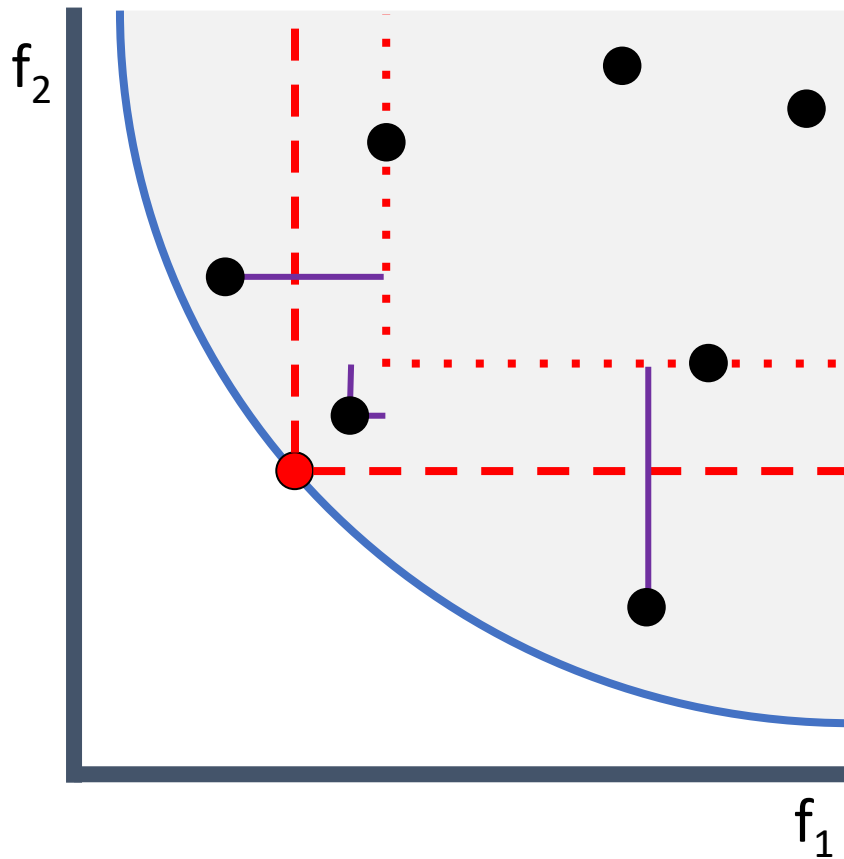
A **policy** is γ -superhuman if it has smaller **cost features** f_1, f_2, \dots than $\gamma\%$ of **human demonstrations**



margins

Pareto frontier

Superhuman Percentile & Subdominance

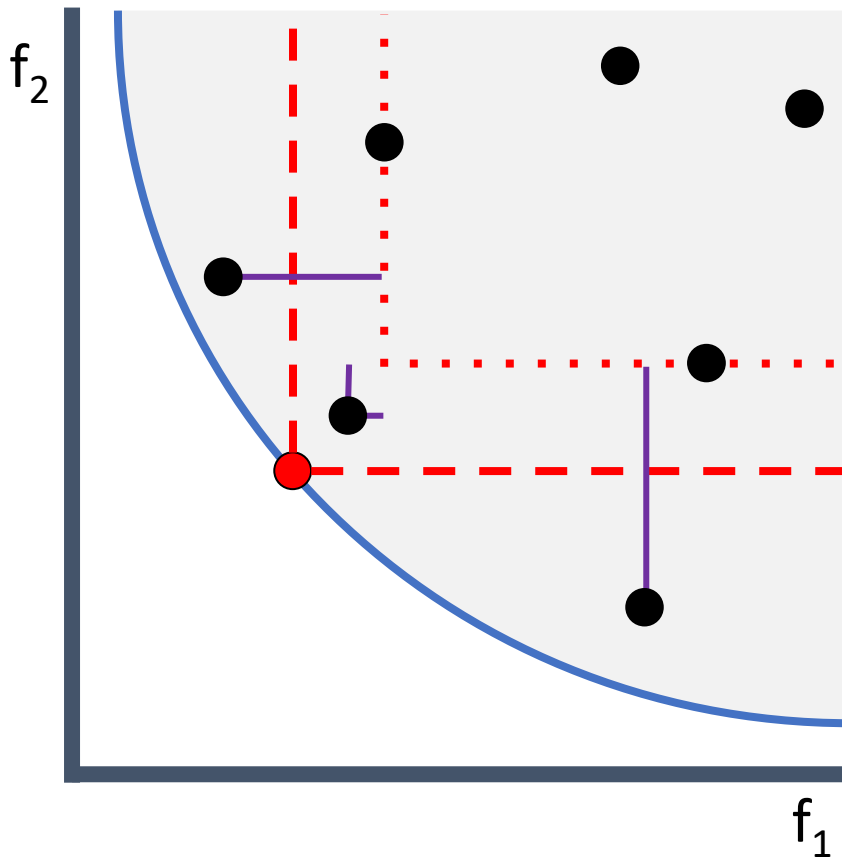


A **policy** is γ -superhuman if it has smaller **cost features** f_1, f_2, \dots than $\gamma\%$ of **human demonstrations**

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

Pareto frontier

Superhuman Percentile & Subdominance

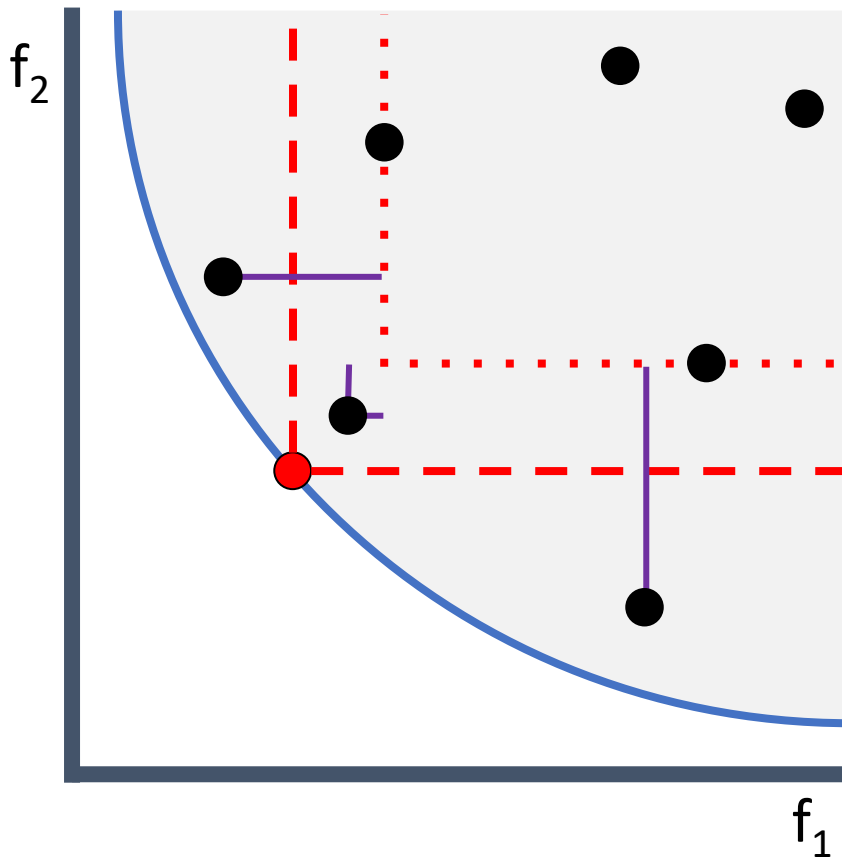


A **policy** is γ -superhuman if it has smaller **cost features** f_1, f_2, \dots than $\gamma\%$ of **human demonstrations**

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



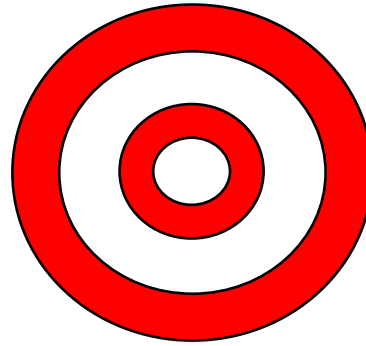
A **policy** is γ -superhuman if it has smaller **cost features** f_1, f_2, \dots than $\gamma\%$ of **human demonstrations**

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

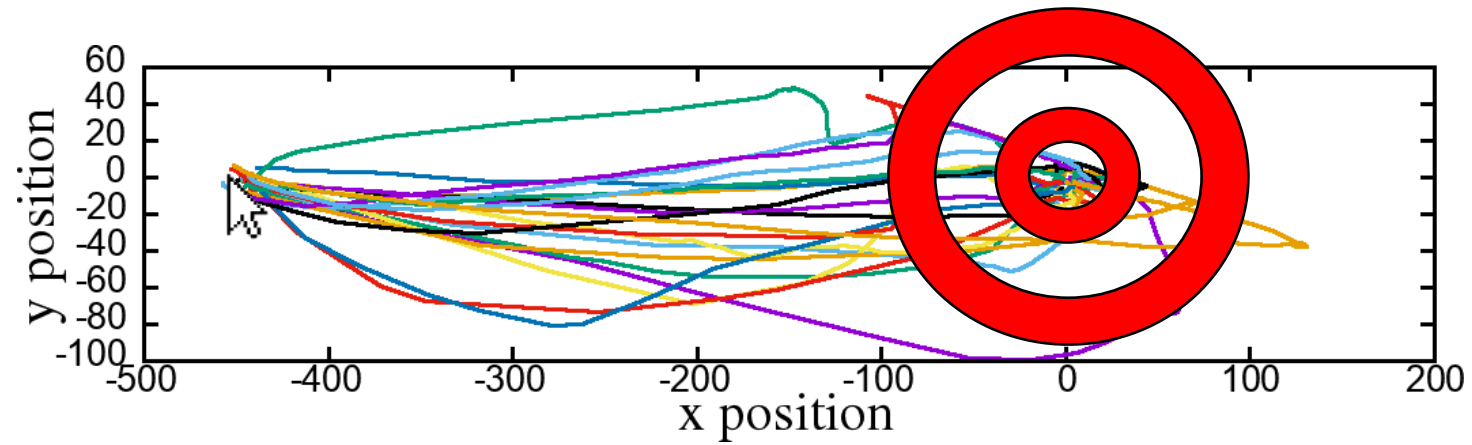
Minimum Subdominance Inverse Optimal Control seeks policies on the **Pareto frontier** minimizing this

Subdominance bounds the **superhuman percentile**

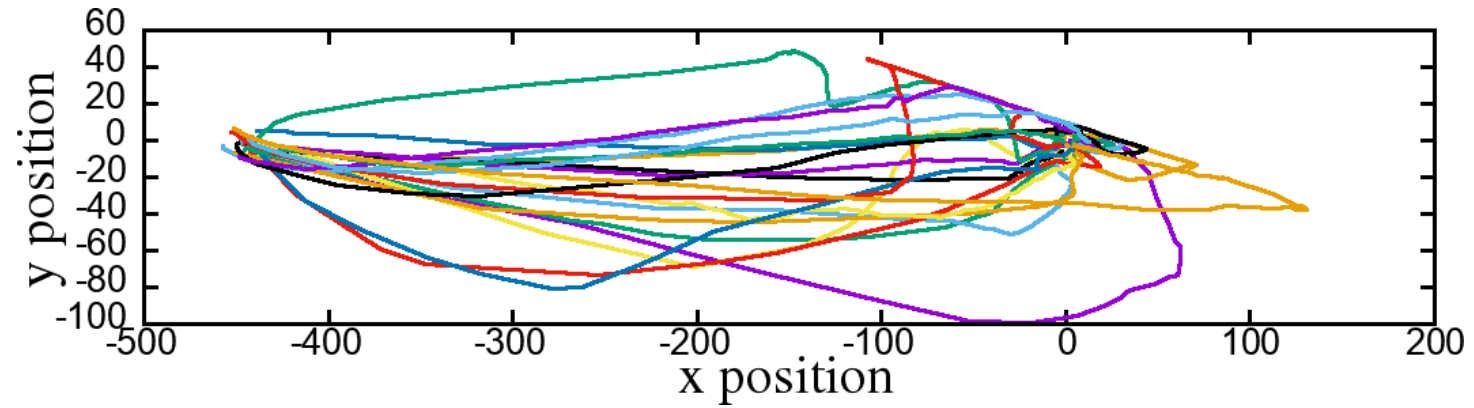
Cursor Pointing Task



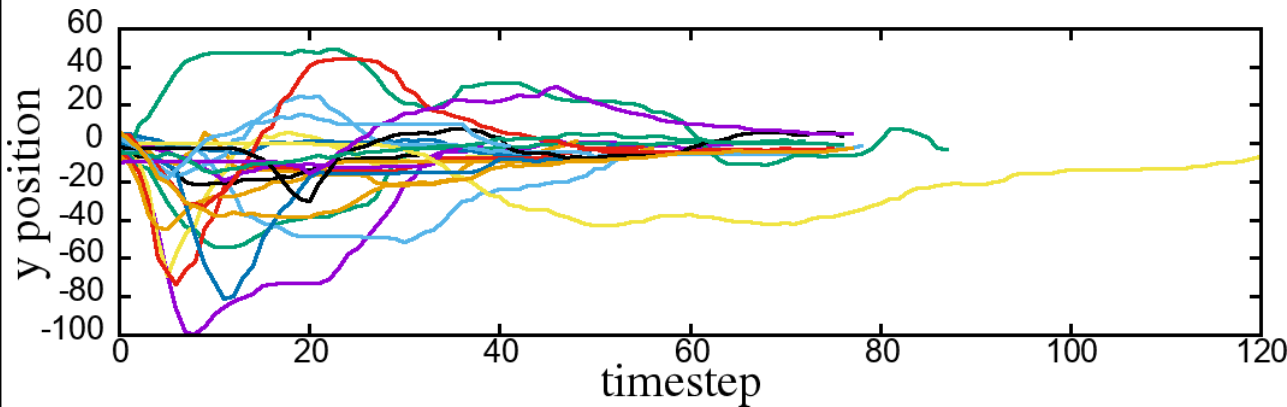
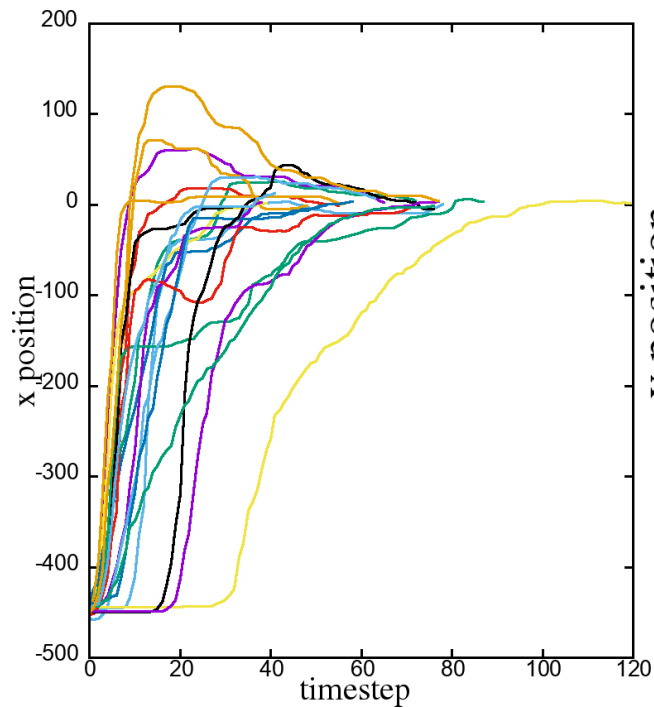
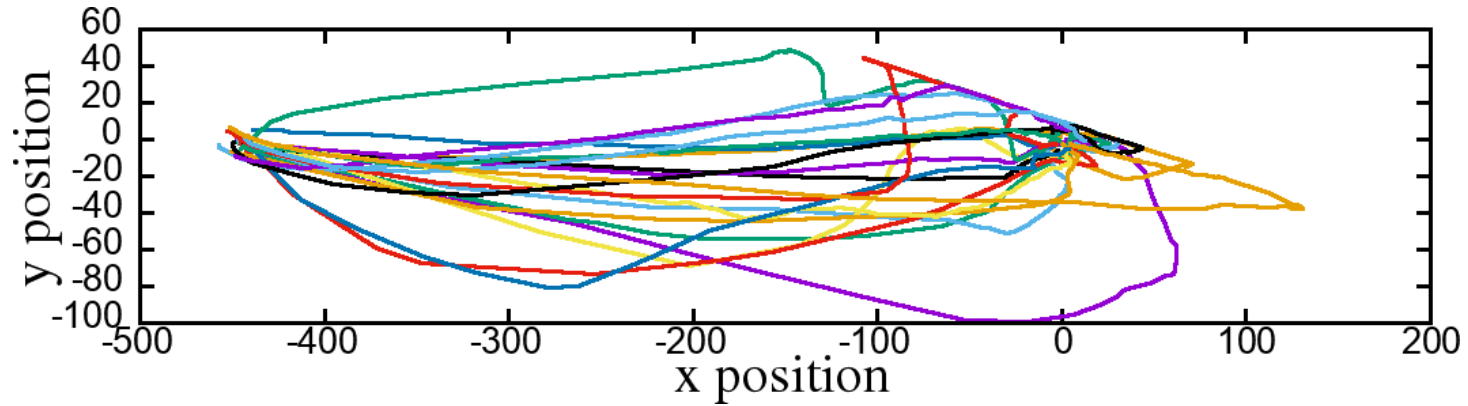
Cursor Pointing Task



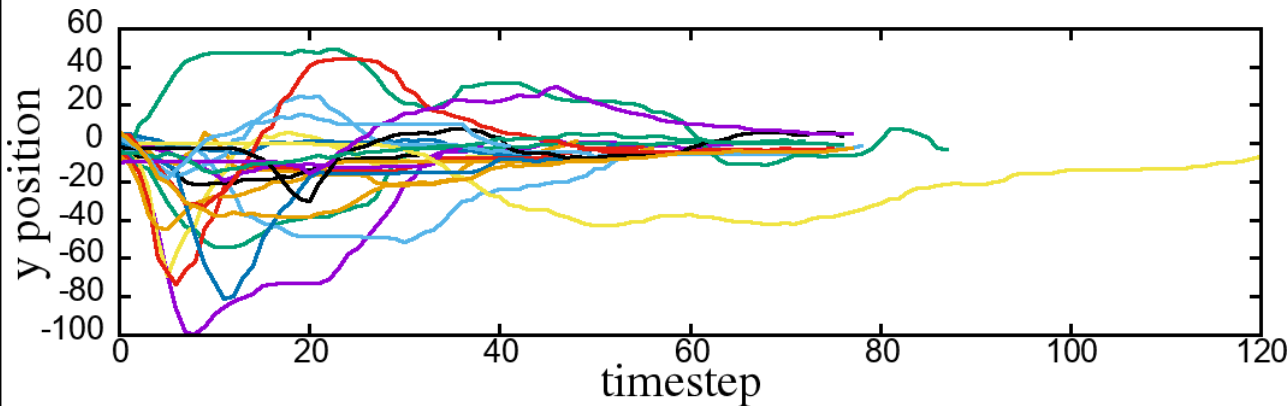
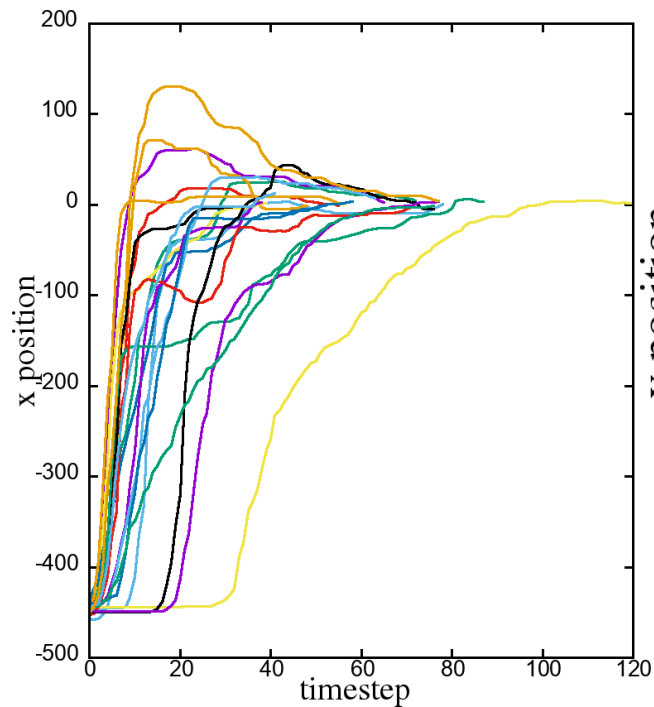
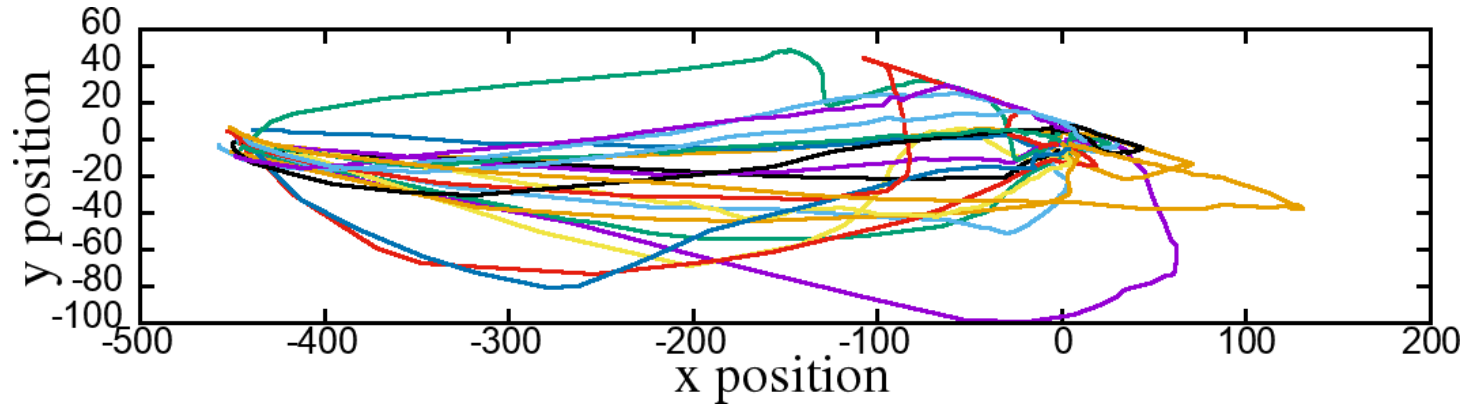
Cursor Pointing Task



Cursor Pointing Task

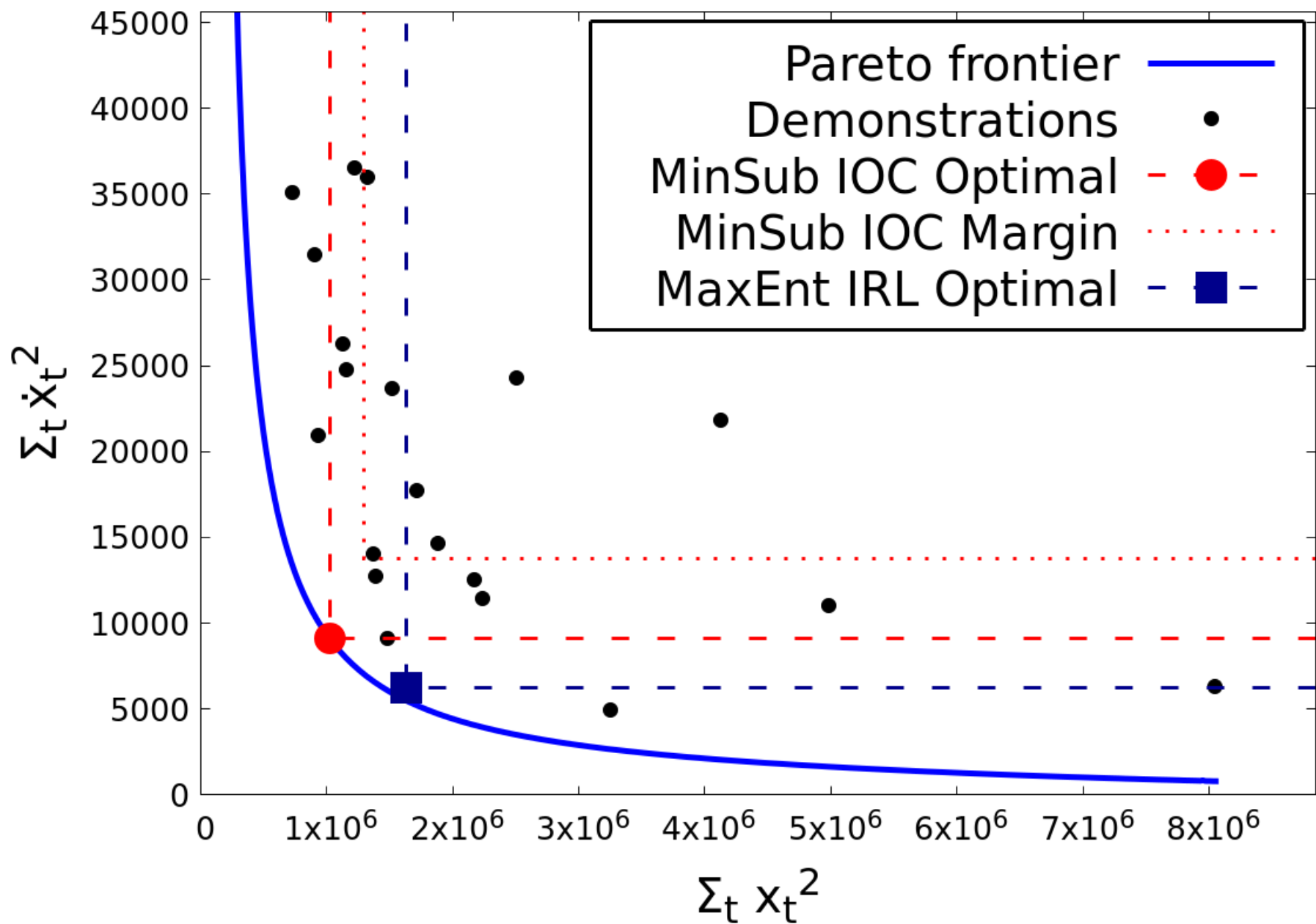


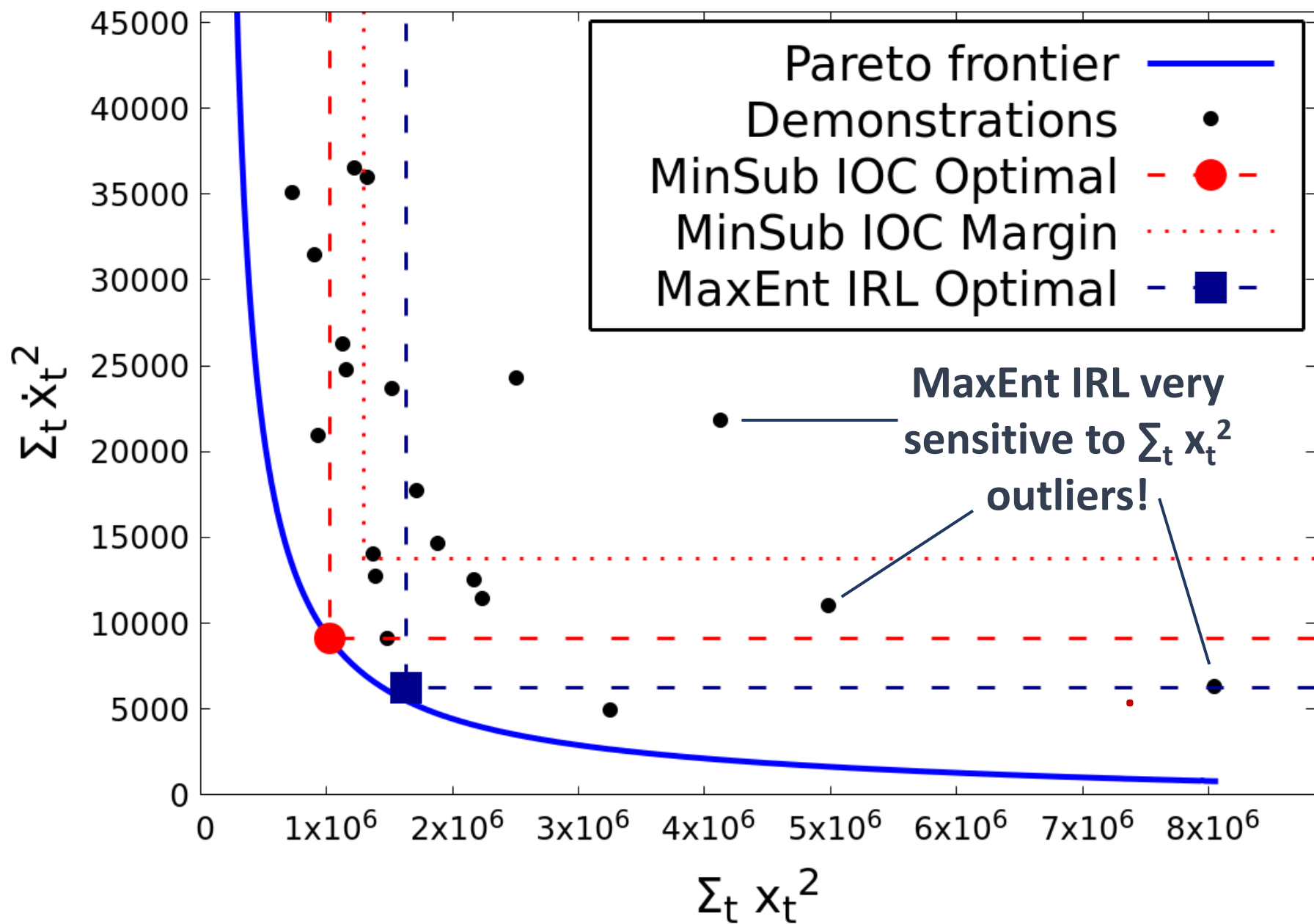
Cursor Pointing Task

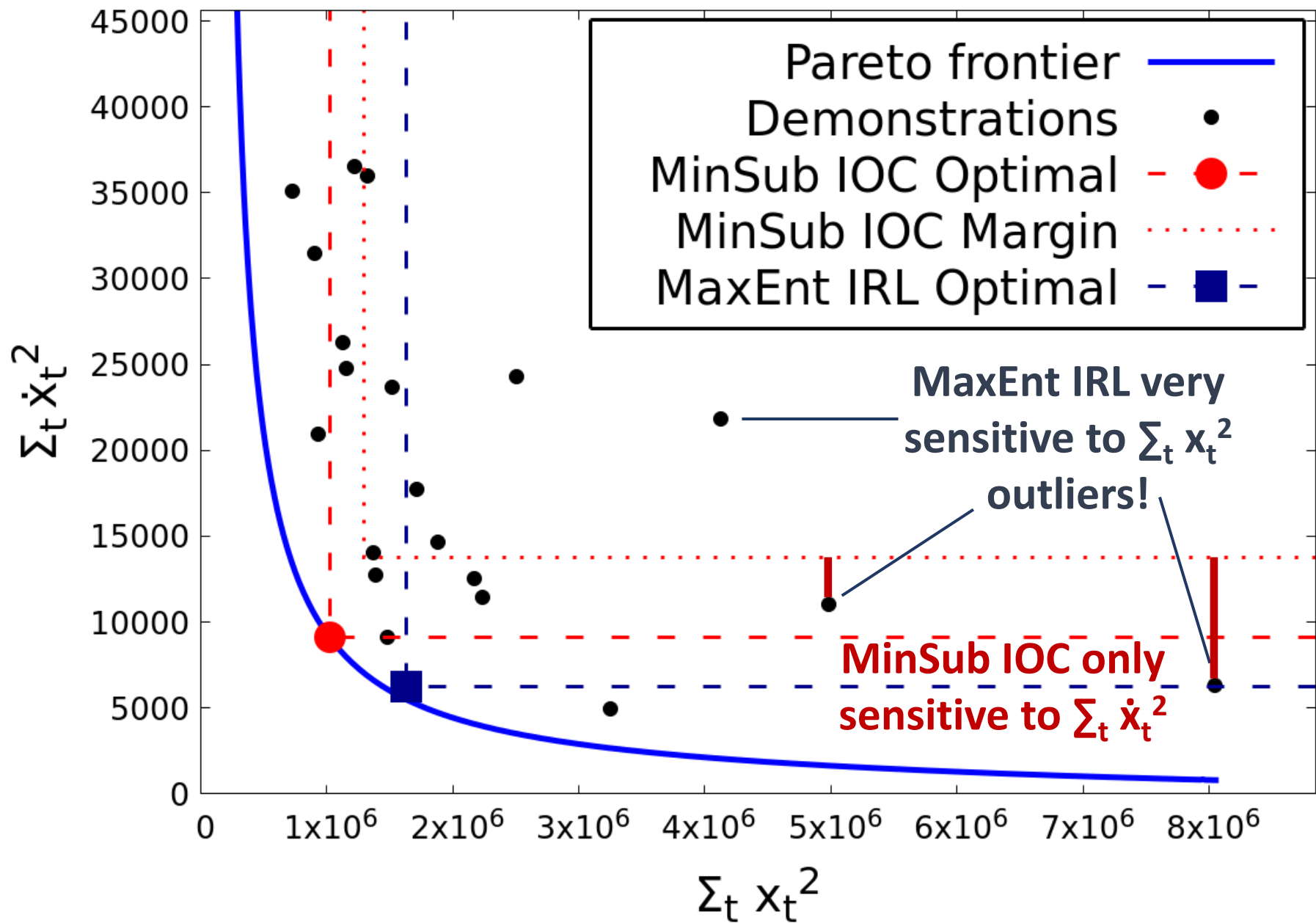


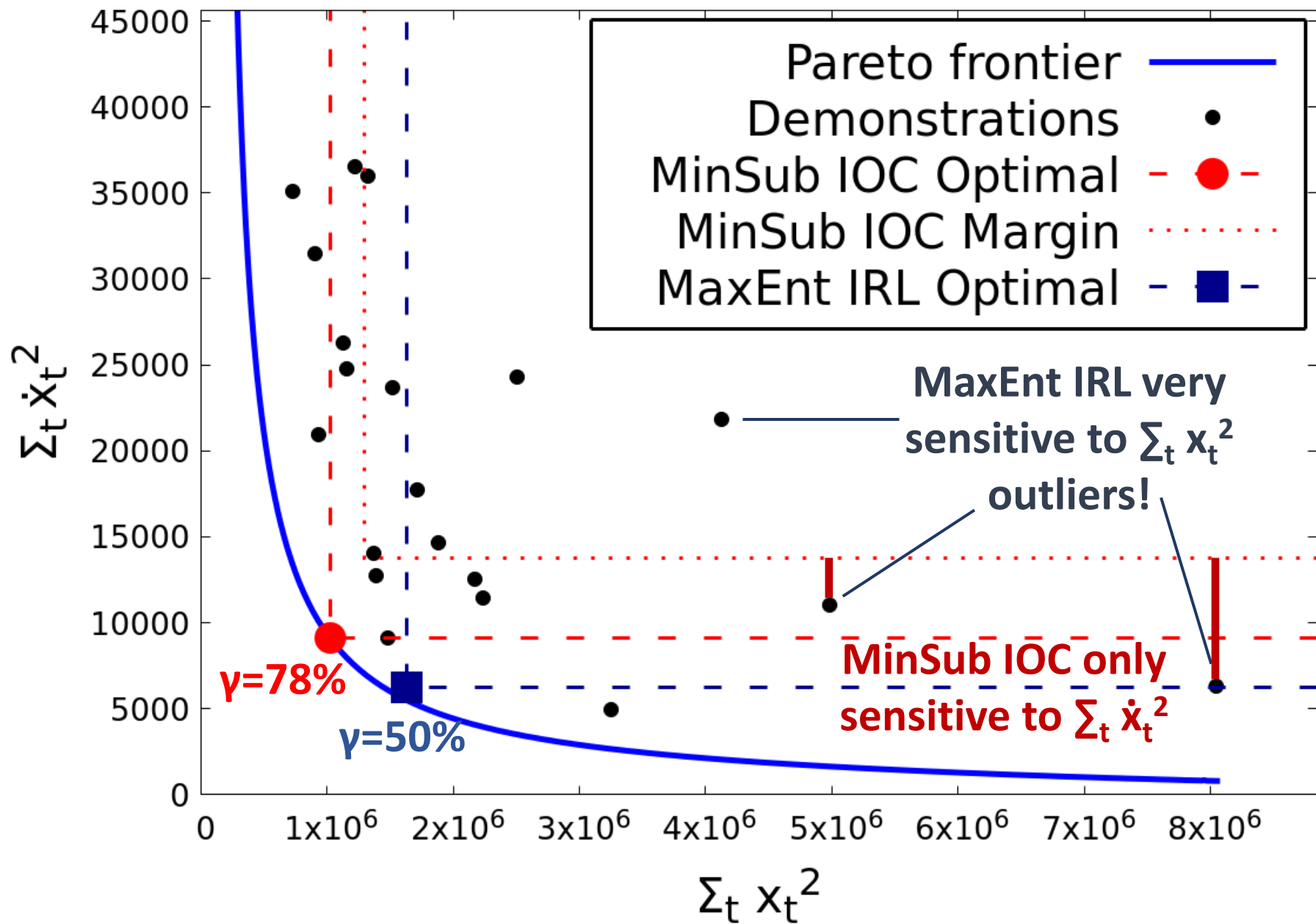
Linear-quadratic regulation formulation:

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









And much more...

Towards Uniformly Superhuman Autonomy via Subdominance Minimization

Brian D. Ziebart, Sreyash Choudhury, Xinyue (Shane) Yao, Paul Vercos

UNIVERSITY OF ILLINOIS CHICAGO **Aurora** ICMIL

How should we think about imitation learning?

Human-provided and observed demonstrations are often imperfect. Small amounts of noise. States from known distributions. Goal information (if any). Metrics of high-quality and low-quality demonstrations often depend on the intent from achieving high-quality performance.

Related work seeks to learn low-quality demonstrations in various ways: supervised learning of demonstrations, linear confidence variables, or supervisory noise models of human intent.

Most similar to our approach, Szepesvári & Schapire (2007) [5] maximize consistent imitation performance difference.

We argue that one imitation learning objective/methods are useful:

Superhuman behavior: an ideal objective?

A policy is superhuman if it has smaller cost than $\gamma\%$ of human demonstrations. Consistent lower cost for demonstration costs from arbitrary cost functions. Set of superhuman policies on the Pareto frontier. Metrics on demonstration goals. Unintentionally, this set can often become empty!

Minimum Subdominance Inverse Optimal Control

A policy is superhuman if it has smaller cost than $\gamma\%$ of human demonstrations. Subdominance measures how far a policy is from superhuman by some metric, bounding the superhuman guarantee.

Minimum Subdominance Inverse Optimal Control seeks policies on the Pareto frontier achieving γ .

Scale to absolute or relative. Aggregation using max or sum.

Subdominance can be viewed as worst case optimality.

Or as the difference between behavior when suboptimally is ignored.

The optimal control controller costs, as subdominance weights is to minimize subdominance.

Consistent: On average, Meta-IC is superhuman on the population distribution (over γ demonstrations, $\gamma \in \{0.1, 0.2, \dots, 0.9\}$).

Consistency: Meta-IC based on the population distribution over the set of all increasing functions of the original features produces the highest superhuman percentage behavior. (ICW) is superior to γ -superhuman(0).

Cursor Pointing Experiments

20 participants (colored notes represent) + 300 cursor pointing trials produce demonstration trajectories that are noisy and vary significantly [1].

We view cursor pointing as a linear-quadratic regulator (LQR) task with a cost function linear in the target aligned (x) and target orthogonal (y) spread, position, velocity, and acceleration.

Meta-IC (0.1) is nearly needed to obtain γ significantly regarding the imitator's superhuman percentage.

Meta-IC (0.1) is more robust than Meta-IC (0.5), when tested from small amounts of data (single test).

Meta-IC (0.1) is more aggregated than a more distributed cost function. Meta-IC (0.1) provides more superhuman behavior than Meta-IC (0.5)-noise after learning low quality demonstrations.

Meta-IC (0.1) is more robust than Meta-IC (0.5), when tested from small amounts of data (single test).

References:

- [1] Sreyash S. Choudhury, Brian D. Ziebart, and Paul Vercos. An actor-critic for consistent imitation learning. In AAAI Conference on Artificial Intelligence, 2016.
- [2] Sreyash S. Choudhury, Brian D. Ziebart, and Paul Vercos. A generative approach to supervised learning in imitation learning. In Proceedings of the AAAI Conference on Artificial Intelligence, 2017.
- [3] Sreyash S. Choudhury, Brian D. Ziebart, and Paul Vercos. A generative approach to supervised learning in imitation learning. In Proceedings of the AAAI Conference on Artificial Intelligence, 2017.
- [4] Sreyash S. Choudhury, Brian D. Ziebart, and Paul Vercos. A generative approach to supervised learning in imitation learning. In Proceedings of the AAAI Conference on Artificial Intelligence, 2017.
- [5] Csaba Szepesvári and Robert E. Schapire. Consistent imitation learning with noisy observations. In Proceedings of the AAAI Conference on Artificial Intelligence, 2007.

This material is based upon work supported by the National Science Foundation under Grant Nos. IRI-03201, IRI-03170, and IRI-03070.

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

Poster: Hall E #827