The Pseudo-Dimension of Contracts

Tomek Ponitka
Tel Aviv University

Joint work with

Paul Dütting
Google Research

Michal Feldman
Tel Aviv University
Microsoft ILDC

Ermis Soumalias
University of Zurich
ETH AI Center
Meta

Contract Design

Incentivize an agent to act in your interest through a contract

Incentivize a salesman
to promote your product
through a fixed-percent commission.

Incentivize an insured person
to avoid risky behaviour
through co-pays and deductibles.

Challenge: Can we learn a good contract using past data?

The Principal-Agent Model

Outcomes and
principal's rewards

Actions and
agent's costs

	agent's cost	no sale $0	small sale $200	big sale $500
low effort	$0	50%	50%	
high effort	$100		50%	50%

The principal does not observe agent's action (effort).
The principal only observes the outcome (sale).

The Principal-Agent Model

contract:

	no sale $0	small sale $200	big sale $500
transfer from principal to agent	$0	$100	$400

agent's utility = expected transfer − agent's cost (determines the action)

	no sale	small sale	big sale	cost	agent's utility
low effort	50% × $0	50% × $100		$0	$50
high effort		50% × $100	50% × $400	$100	$150

principal's utility = expected reward − expected transfer (our objective)

	no sale	small sale	big sale	expected transfer	principal's utility
low effort					
high effort		50% × $200	50% × $500	$250	$100

Key Classes of Contracts

***Linear* contracts:**
pay α-fraction of reward
$\mathcal{C}_{\text{linear}} = [0, 1]$

example of a linear contract:
transfer 10% of reward

	no sale $0	small sale $200	big sale $500
transfer from principal to agent	$0	$20	$50

***Bounded* contracts:**
w.l.o.g. transfer at most 1
$\mathcal{C}_{\text{bounded}} = [0, 1]^{\#\text{outcomes}}$

***Unbounded* contracts:**
any transfer
$\mathcal{C}_{\text{unbounded}} = [0, \infty)^{\#\text{outcomes}}$

Our Model

Unknown agent type drawn from a probability distribution.
We only observe samples from that distribution.

agent (type 1)

	no sale $0	small sale $200	big sale $500	agent's cost
low effort	50%	50%		$0
high effort		50%	50%	$100

agent (type 2)

	no sale $0	small sale $200	big sale $500	agent's cost
low effort	80%	10%	10%	$0
high effort			100%	$300

Our Model

Unknown agent type drawn from a probability distribution.
We only observe samples from that distribution.

agent (type 1)

	no sale $0	small sale $200	big sale $500	agent's cost
low effort	50%	50%		$0
high effort		50%	50%	$100

agent's type space $\Theta = \left(\Delta^{\#\text{outcomes}}\right)^{\#\text{actions}} \times \mathbb{R}_{\geq 0}^{\#\text{actions}}$

outcome distributions — agent's costs

Our Model

Unknown agent type drawn from a probability distribution.
We only observe samples from that distribution.

salesman's type:
skillset

insured person's type:
health predisposition

Health predisposition (agent's type) affects the probabilities of requiring treatments (outcomes) if the agent acts recklessly (action).

Specify the insurance policy (contract) using a sample of health records.

Related Work

Learning contracts under different feedback models.

[Ho, Slivkins, Vaughan, 2014]
[Cohen, Koren, Deligkas, 2018]
[Zhu, Bates, Yang, Wang, Jiao, Jordan, 2023]
[Dütting, Guruganesh, Schneider, Wang, 2023]
[Chen, Chen, Deng, Huang, 2024]
[Bacchiocchi, Castiglioni, Marchesi, Gatti, 2024]

Optimizing for an agent drawn from a known distribution.

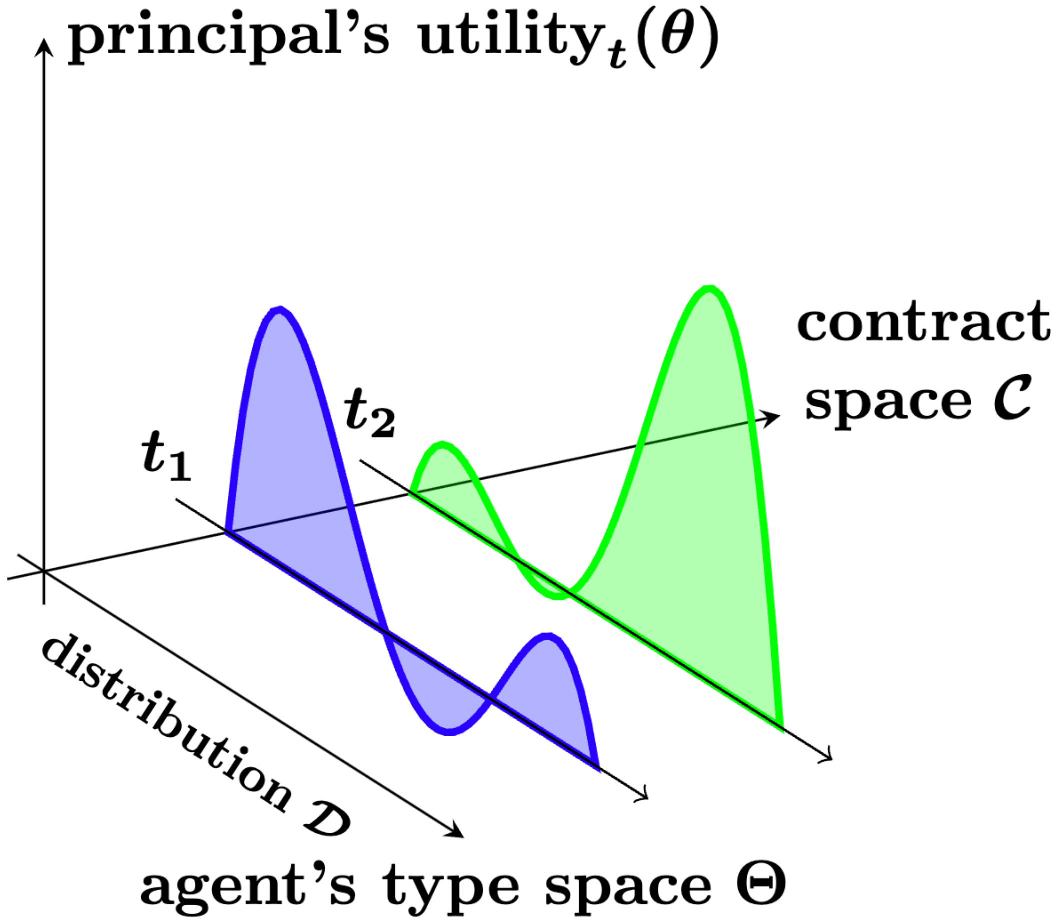
[Guruganesh, Schneider, Wang, 2020]
[Castiglioni, Marchesi, Gatti, 2021]
[Alon, Dütting, Talgam-Cohen, 2021]
[Castiglioni, Marchesi, Gatti, 2022]
[Guruganesh, Schneider, Wang, Zhao, 2023]
[Alon, Dütting, Li, Talgam-Cohen, 2023]

Similar techniques in learning auctions.

[Balcan, Blum, Hartline, Mansour, 2005]
[Cole, Roughgarden, 2015]
[Morgenstern, Roughgarden, 2015]
[Balcan, DeBlasio, Dick, Kingsford, Sandholm, Vitercik, 2021]
[Balcan, Sandholm, Vitercik, 2017]
[Beyeler, Brero, Lubin, Seuken, 2024]
[Soumalias, Heiss, Weissteiner, Seuken, 2024]
[Soumalias, Weissteiner, Heiss, Seuken, 2024]

The Learning Problem

Question 1: How many samples from the agent type distribution are needed to learn a near-optimal contract with high probability?



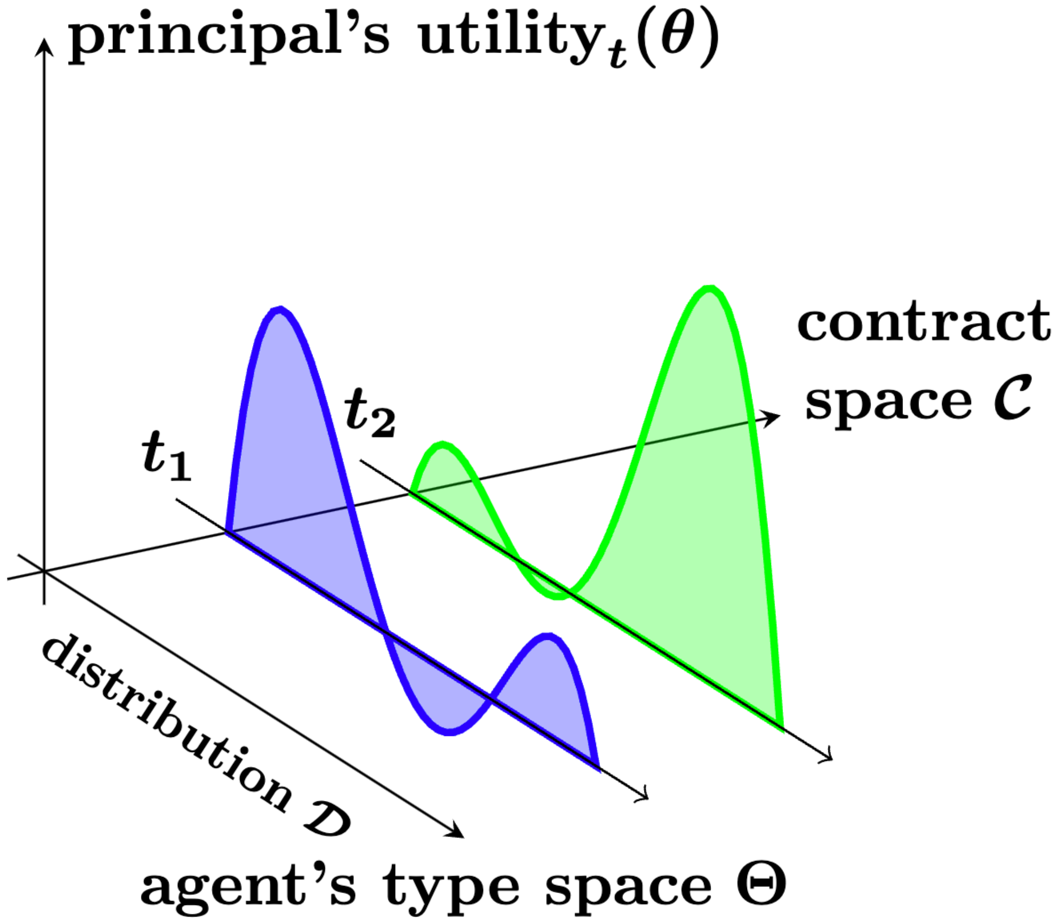
Find a contract t maximizing

$$\mathbb{E}_{\theta \sim \mathcal{D}}[\text{principal's utility}_t(\theta)]$$

up to an additive error of ϵ, with probability at least $1 - \delta$.

The Pseudo-Dimension of Contracts

The pseudo-dimension is a combinatorial measure of complexity of a class of real-valued functions. [Pollard, 1984]



It can be applied to contract classes, viewed as classes of functions from agent's type to principal's utility.

It can be used to bound sample complexity (next slide).

It offers a new perspective on the simplicity vs optimality tradeoff.

The Pseudo-Dimension of Contracts

Classic Theorem:

For any class $\mathcal{C} \subseteq \mathcal{C}_{\text{bounded}}$, it suffices to have

$N = O\left((1/\epsilon)^2 \cdot \left(\text{Pdim}(\mathcal{C}) + \log(1/\delta)\right)\right)$ samples,

to learn a contract in $\mathcal{C}$ that is optimal up to an additive error of ϵ, with probability at least $1 - \delta$.

Research Direction

Question 1: How many samples from the agent type distribution are needed to learn a near-optimal contract with high probability?

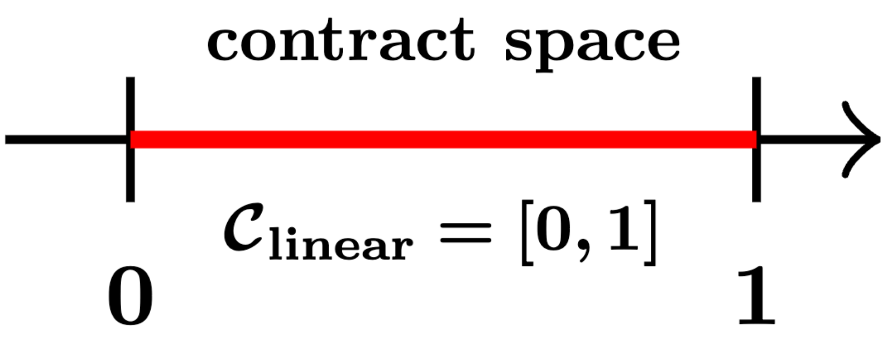
Question 2: What is the pseudo-dimension of key contract classes: linear, bounded, and unbounded?

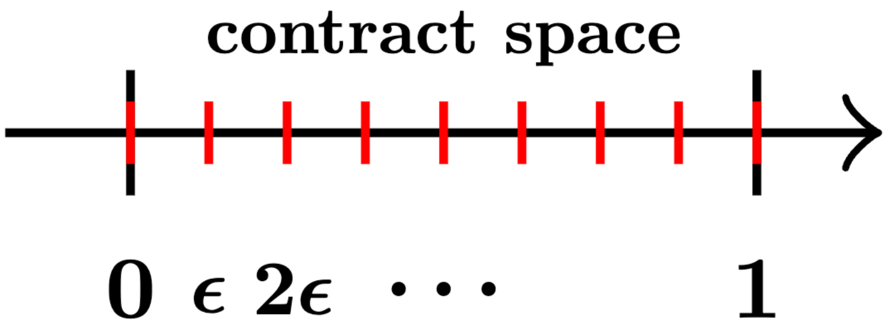
Question 3: Are there contract classes with low pseudo-dimension that closely approximate key contract classes?

Approximation quality is measured by the representation error: the additive loss in principal's utility compared to original class.

Linear Contracts

Linear contracts:
pay α-fraction of reward
$\mathcal{C}_{\text{linear}} = [0, 1]$





***Theorem** (All Linear Contracts):*
Pdim($\mathcal{C}_{\text{linear}}$) = $\Theta(\log(\#\text{actions}))$

Issue: #actions can be infinite, e.g., when effort levels are [0,1] rather than {low, high}.

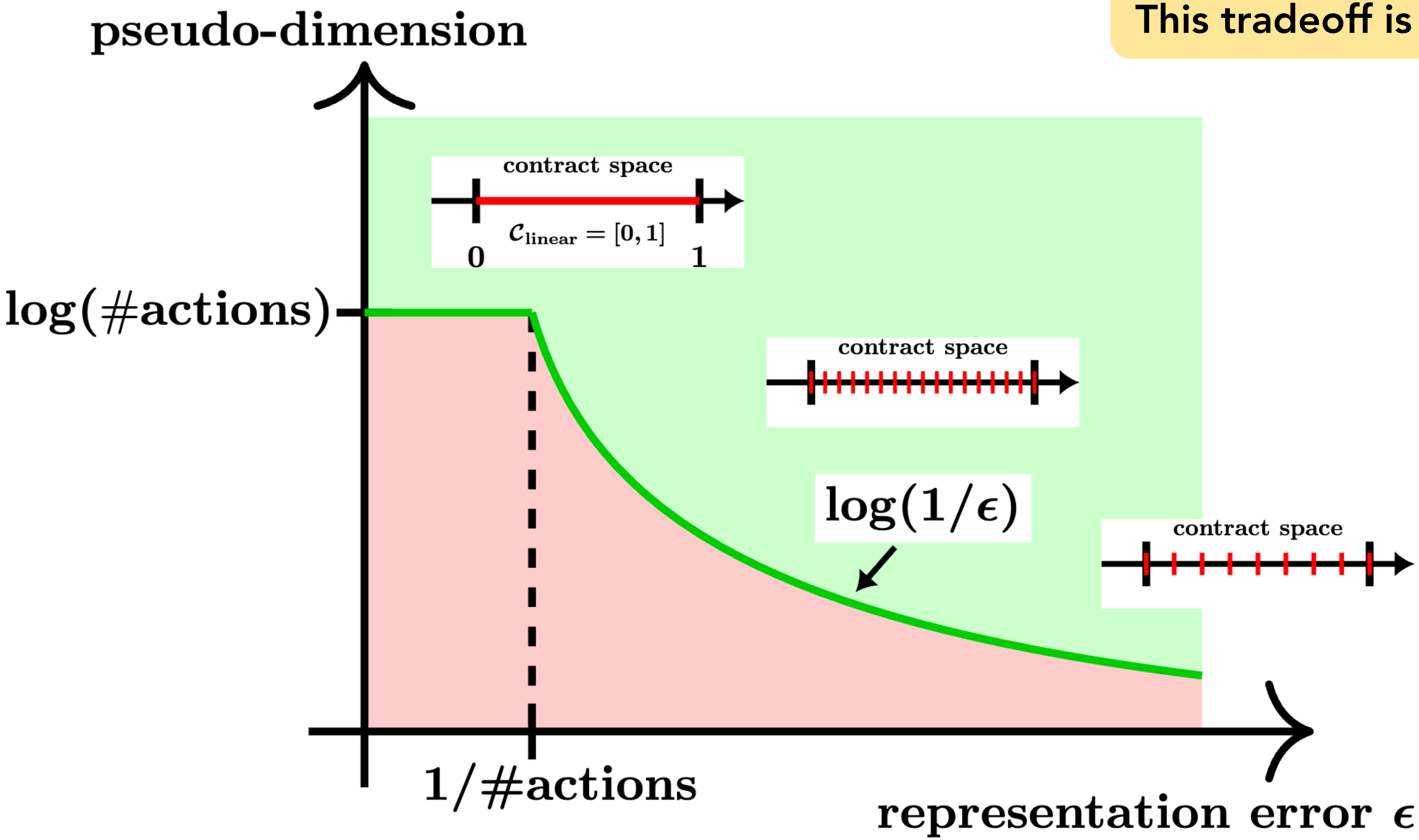
***Theorem** (Discretized Linear Contracts):*
Pdim = $\Theta(\log(1/\epsilon))$
representation error = ϵ

(i.e., principal's utility ≥ OPT−LINEAR − ϵ)

Works even for a continuous action space!

Pareto Frontier for Linear Contracts

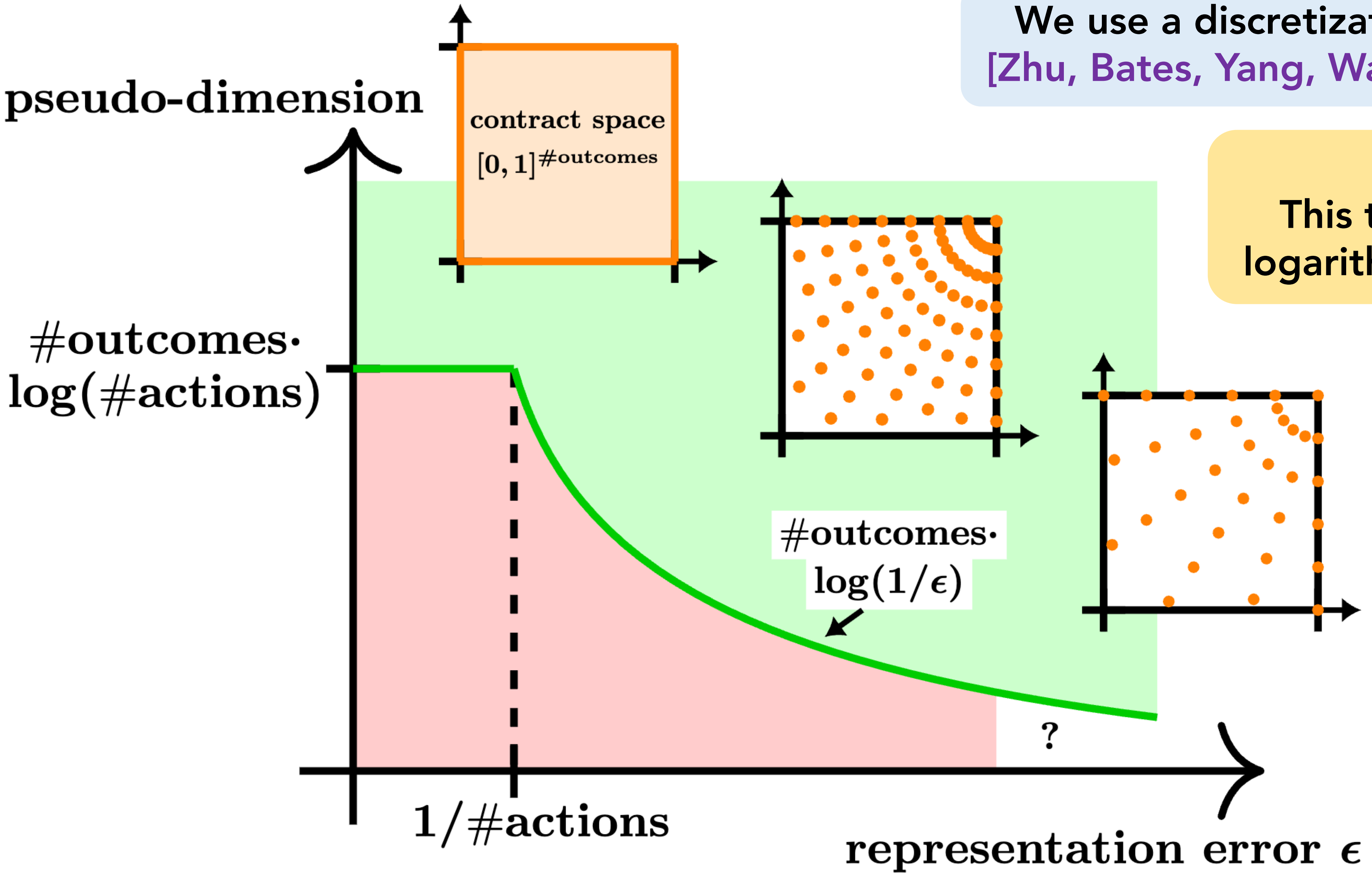
Theorem:
This tradeoff is tight up to a constant!



Pareto Frontier for Bounded Contracts

We use a discretization of $\mathcal{C}_{\text{bounded}}$ due to [Zhu, Bates, Yang, Wang, Jiao, Jordan, 2023]

Theorem:
This tradeoff is tight up to logarithmic factors for small ϵ!



Sample Complexity

Our pseudo-dimension analysis leads to essentially tight bounds on sample complexity for linear and bounded contracts.

Theorem (Positive): We can learn linear contracts with sample complexity of $\widetilde{\Theta}\left((1/\epsilon)^2 \cdot \log(1/\delta)\right)$.

Theorem (Positive): We can learn bounded contracts with sample complexity of $\widetilde{\Theta}\left((1/\epsilon)^2 \cdot (\#\text{outcomes} + \log(1/\delta))\right)$.

In contrast, for unbounded contracts, we establish impossiblity.

Theorem (Negative): There is <u>no</u> algorithm with finite sample complexity for learning unbounded contracts.

Main Insights

Main Results:
Near-tight bounds on pseudo-dimension and sample complexity.

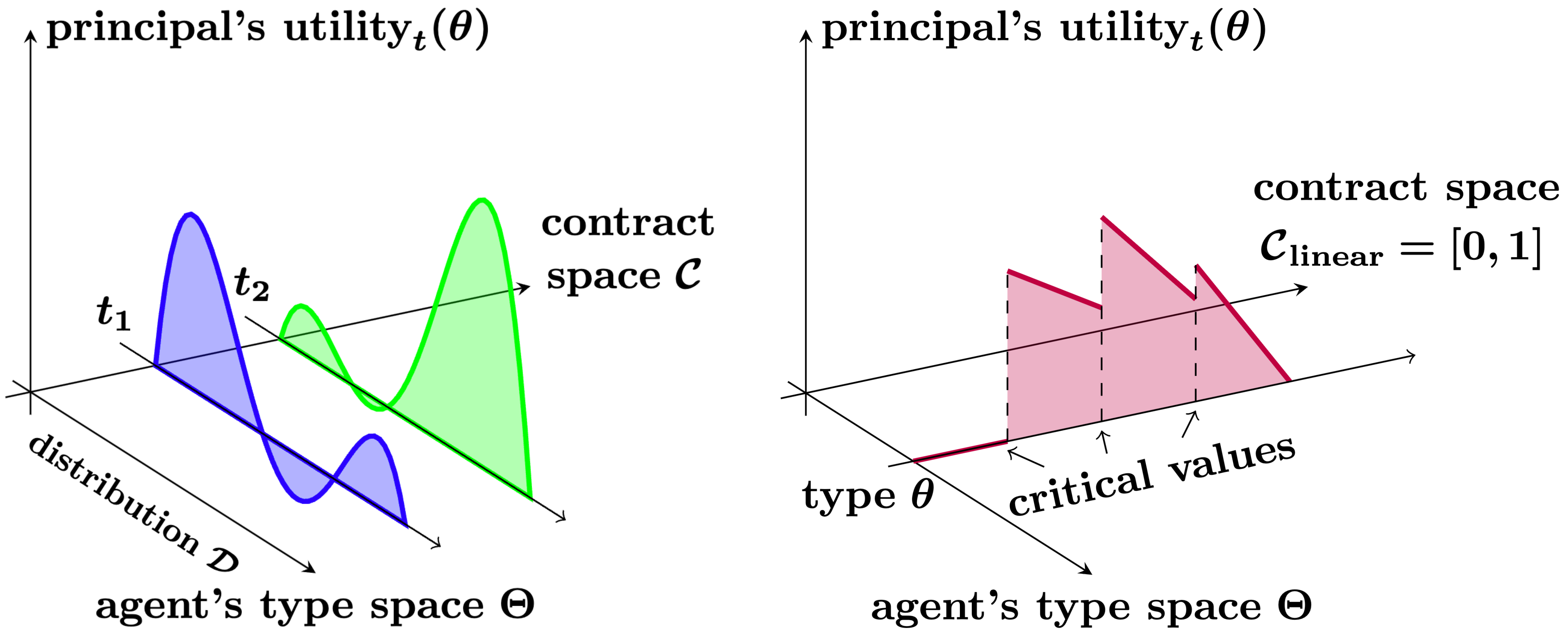
We also extend our analysis to piecewise linear contracts and menus of contracts (see the paper for details).

Structural Insight #1: Sample complexity of learning linear contracts depends on the number of critical values.

Structural Insight #2: We establish a strong separation between expert advice and bandit feedback in our setting.

Structural Insight #1: Critical Values

Lemma [Dütting, Ezra, Feldman, Kesselheim, 2021]: For linear contracts, for any fixed agent's type θ, the principal's utility is piecewise linear.



Structural Insight #1: Critical Values

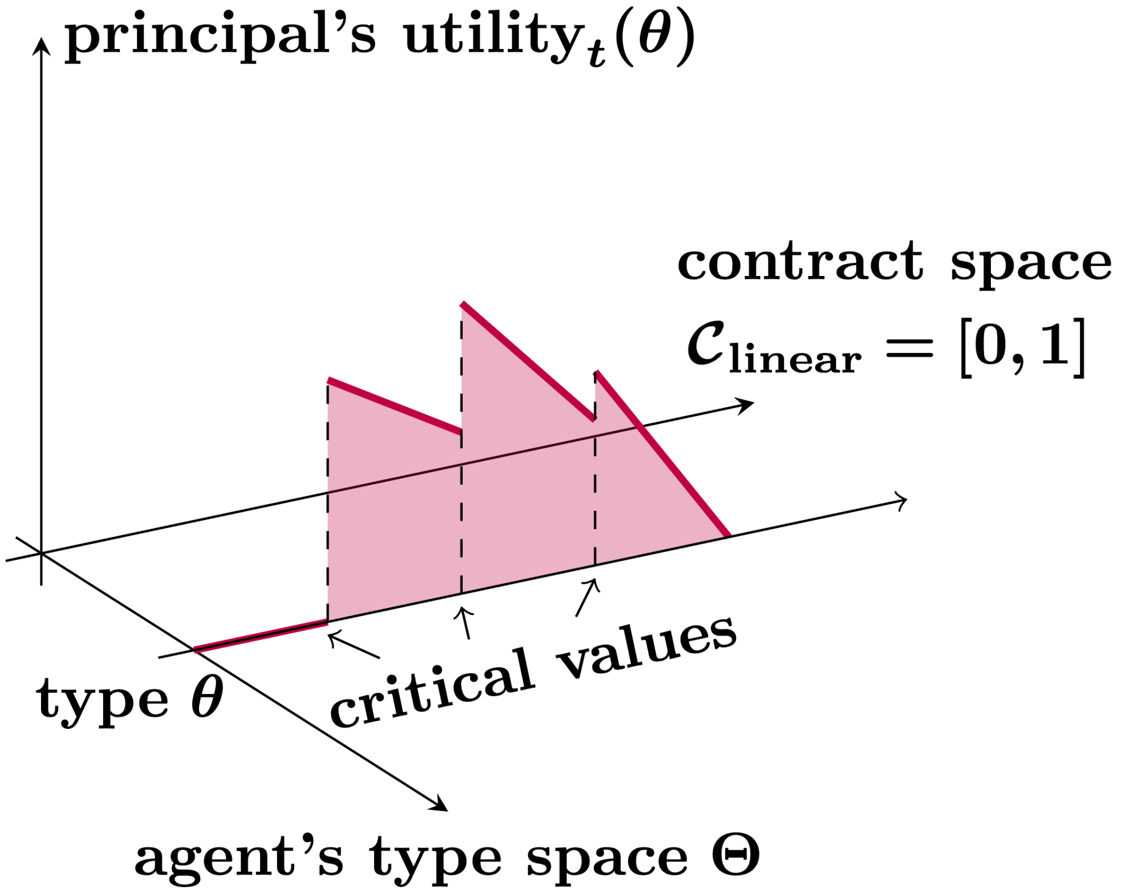
Proof: $\mathrm{Pdim}(\mathcal{C}_{\mathrm{linear}}) \leq \log(\#\text{critical values}) \leq \log(\#\text{actions})$

First step is based on delineability.
[Balcan, Sandholm, Vitercik, 2023]

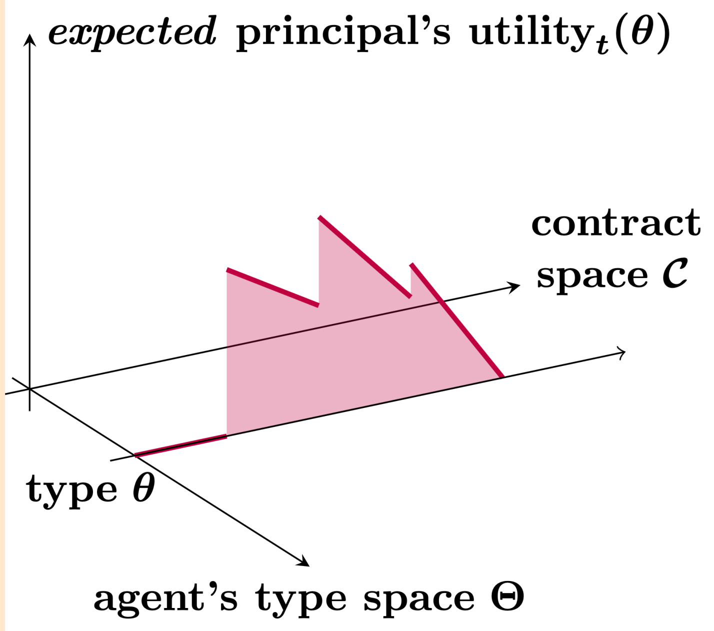
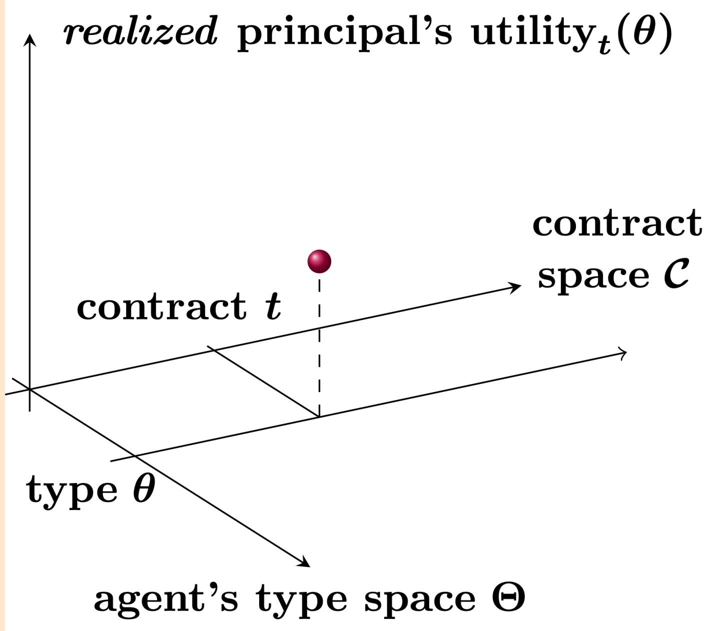
New connection: sample complexity depends on #critical values.

In related problems, time complexity depends on #critical values.

Better bounds on #critical values are known for many special cases.



Structural Insight #2: Experts vs Bandits

	Our Model: Expert Advice	*Prior Work: Bandit Feedback*
Samples	full agent's type	realized outcome
Sample complexity (bounded contracts)	Polynomial: $\tilde{\Theta}\left((1/\epsilon)^2 \cdot \#\text{outcomes}\right)$	Exponential (even for fixed agent): $(1/\epsilon)^{\Theta(\#\text{outcomes})}$
Given a sample, we observe:	 expected principal's utility for all contracts	 realized principal's utility for one contract

Structural Insight #2: Experts vs Bandits

	Our Model: Expert Advice	*Prior Work: Bandit Feedback*
Samples	full agent's type	realized outcome
Sample complexity (bounded contracts)	Polynomial: $\tilde{\Theta}\left((1/\epsilon)^2 \cdot \#\text{outcomes}\right)$	Exponential (even for fixed agent): $(1/\epsilon)^{\Theta(\#\text{outcomes})}$
	We have to learn the agent's type distribution.	We have to learn both the agent's type distribution and the outcome distributions. The hardness comes from learning the outcome distributions.

Summary

We study sample complexity of contract design.

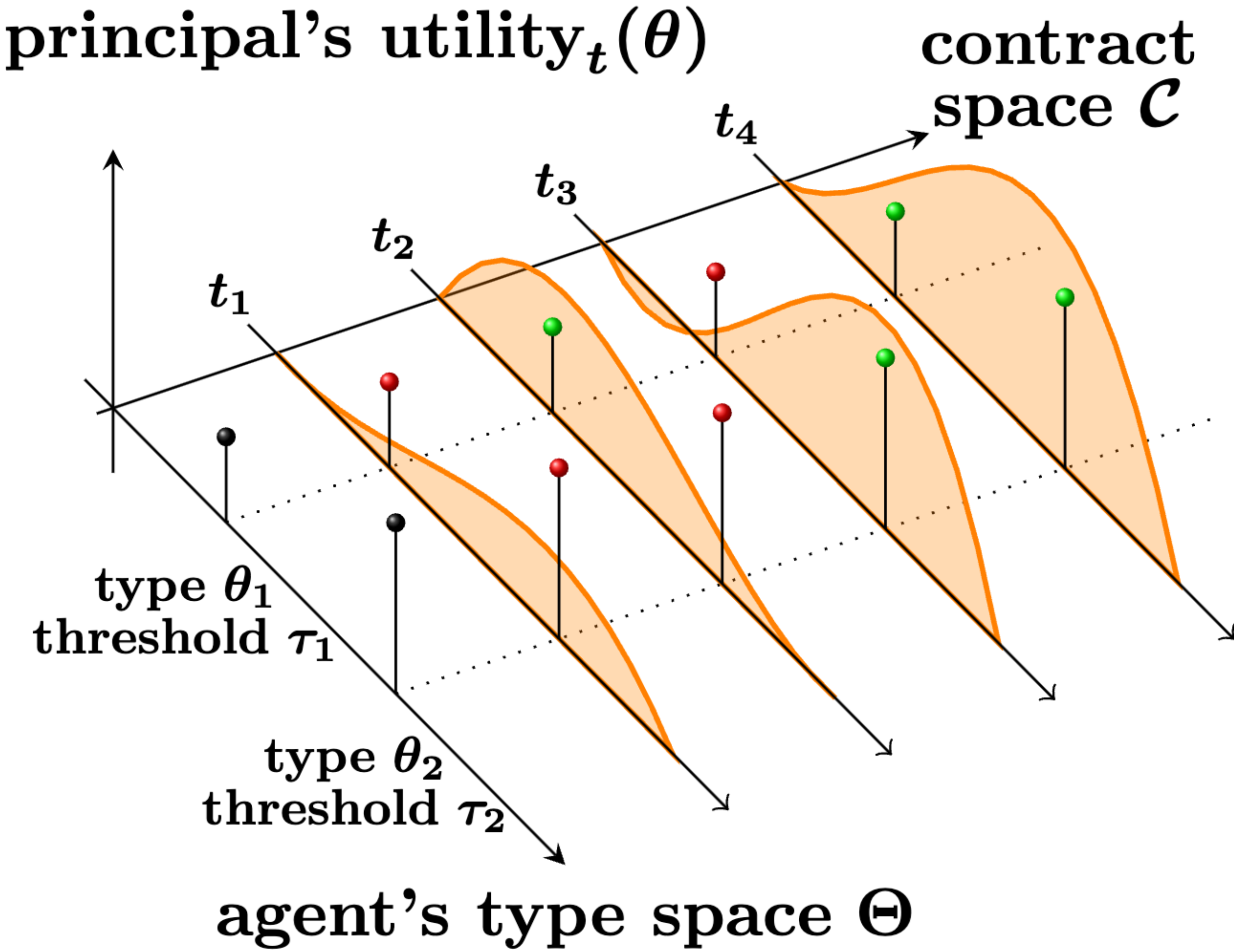
Key Takeaway:
Pseudo-dimension leads to near-tight bounds on sample complexity.

Structural Insight #1: Sample complexity of learning linear contracts depends on the number of critical values.

Structural Insight #2: We establish a strong separation between expert advice and bandit feedback in our setting.

Thank you!

Pseudo-Dimension of Contracts



Definition (safe to skip):
pseudo-dimension of $\mathcal{C}$ =
size of maximal shattering of types

Example: shattering of types $\{\theta_1, \theta_2\}$ with thresholds $\{\tau_1, \tau_2\}$ implies that pseudo-dimension is at least 2.

Pseudo-dimension is defined with respect to the agent's type space Θ. It doesn't depend on distribution $\mathcal{D}$.