Dynamic Pricing

Jianyu Xu
PhD Student
Computer Science Department, UC Santa Barbara

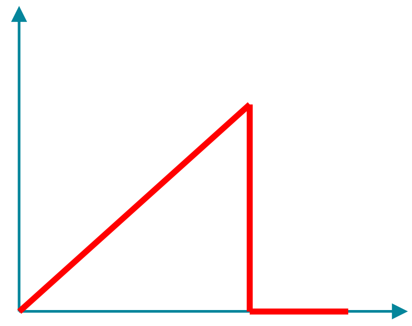
Generic Setting

For $t = 1,2,\ldots,T$:

- Customers generate a valuation y_t secretly;
- We propose a price p_t;
- If $p_t \leq y_t$, then we get a reward $r_t = p_t$;
- Else $p_t > y_t$, then we get a reward $r_t = 0$.

3 Properties:

- Boolean-censored valuation
- Bandit feedback
- Direction sensitive reward



Identical or Various Items

- Identical items: no feature
  - Aim at a best fixed price.
  - Here ``dynamics'' come from explorations.

- Various items: with features
  - For each time, a vector $x_t \in \mathbb{R}^d$ describes the feature of current item, and helps our pricing.
  - Aim at a best ``feature → price'' policy.
  - Here ``dynamics'' come from different items, and explorations.

Non-feature dynamic pricing

Non-feature setting

There are mainly three categories:

1. Fixed valuation

   y_t is fixed but secret

2. Stochastic valuation

   ① y_t is drawn i.i.d. from a fixed unknown distribution.

   ② y_t is drawn from adversarial unknown distribution $\mathbb{D}_t$'s.

3. Adversarial valuation

   y_t is obliviously arbitrarily chosen.

Non-feature setting -- Regret

Regret is defined as the difference between:

1. Max (expected) reward of a best **fixed** price, i.e.

$$\max_{p} \sum_{t=1}^{n} p \cdot 1(p \leq y_t)$$

2. (Expected) reward of our algorithm, i.e.

$$\sum_{t=1}^{n} p_t \cdot 1(p_t \leq y_t)$$

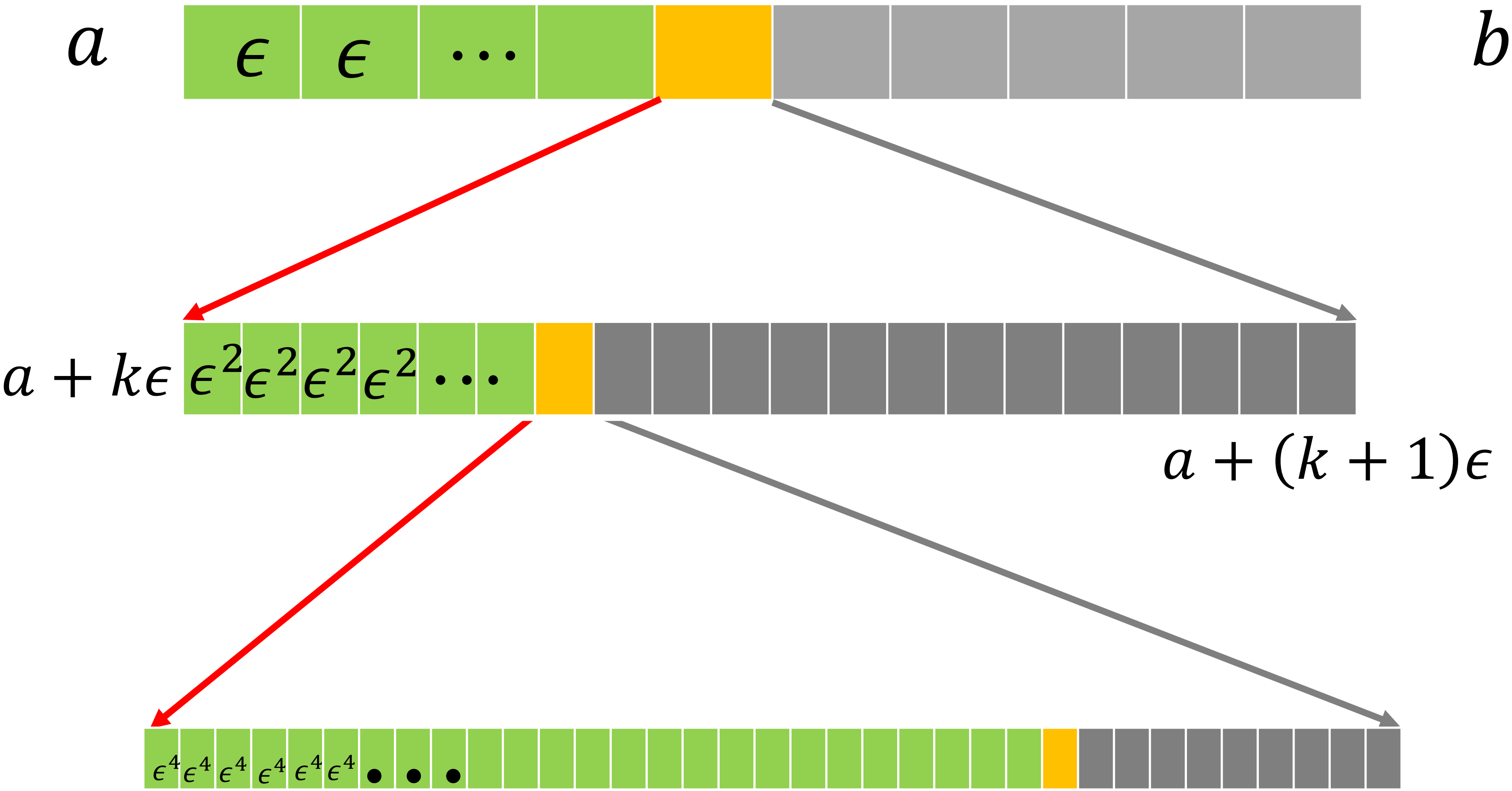
Non-feature setting -- Regrets

	Fixed valuation	Stochastic valuation	Arbitrary valuation
Upper Bound	$O(\log\log T)$ [KL03]	$O\left(T^{2/3}\right)$ in general $O\left(\sqrt{T}\right)$ for $\mathbb{C}^2$ demand curve [KL03]	$O\left(T^{2/3}\right)$ [KL03]
Lower Bound	$\Omega(\log\log T)$ [KL03]	$\Omega\left(T^{2/3}\right)$ in general, and for Lipschitz demand curve	$\Omega\left(T^{2/3}\right)$ [KL03]

Fixed valuation: searching algorithm

- Intuitively, we may apply binary search.
- But binary search can be as expensive as $O(\log T)$.
  - Indeed, a binary search will approach y_t in the fastest rate: at most $O(\log T)$ times of trials.
  - But it might fail in all $O(\log T)$ trials.
- To match the $\Omega(\log\log T)$ lower bound, we instead applies a *squaring search* method.

Squaring search [KL03]



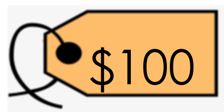
$\cdots\cdots$

Stochastic/Adversarial valuation: bandit algorithm

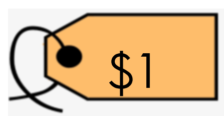
- Discretize the price space as: $\{T^{-\alpha}, 2T^{-\alpha}, \ldots, 1 - T^{-\alpha}, 1\}$, and play multi-armed bandits.

- Error caused by discretization intervals: $O(T^{1-\alpha})$;
- Error caused by multi-armed bandits: $O\left(\sqrt{TK}\right) = O\left(T^{\frac{1+\alpha}{2}}\right)$
- Minimax regret $= O\left(\min_{\alpha} \max\left\{T^{1-\alpha}, T^{\frac{1+\alpha}{2}}\right\}\right) = O\left(T^{\frac{2}{3}}\right)$.

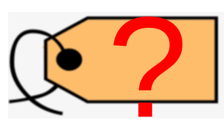
Feature-based dynamic pricing

 + 

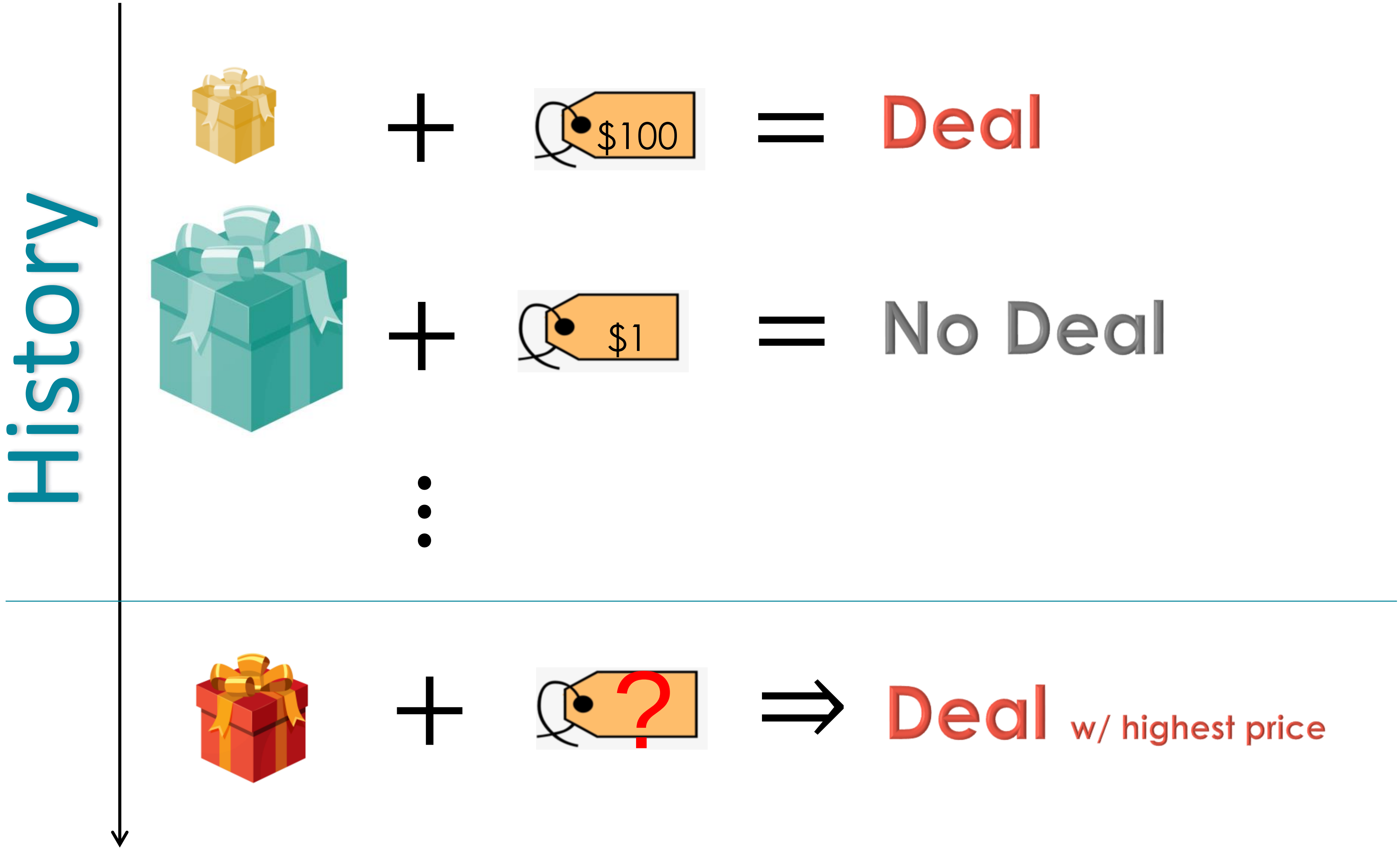
= Deal

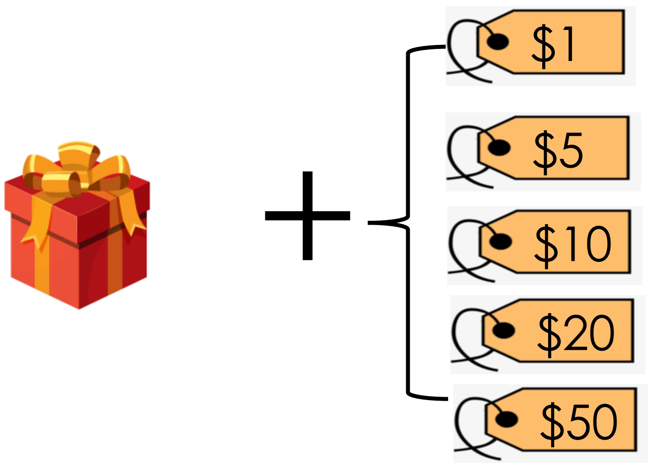
 + 

= No Deal

 + 

= Deal with a high price


History
$100
Deal
$1
No Deal
?
Deal w/ highest price


+
$1
$5
$10
$20
$50

Feature-based setting

We assume *a linear model* for the feature-valuation relationships: $y_t = x_t^\top \theta^* + N_t$, where θ^* is fixed but unknown over time, and N_t can be:

1. $N_t \equiv 0, \forall t = 1,2,\ldots,T$.

2. $N_t \sim_{i.i.d.} \mathbb{D}$, where $\mathbb{D}$ is fixed and could be:
   1. Totally known (e.g. Standard Gaussian)
   2. Parametric (e.g. Gaussian, or Laplacian)
   3. Totally unknown beside bounds

Feature-based setting -- Regret

In this setting, a regret is defined as the difference between:

1. Max expected reward of an **omniscient** seller, i.e.

$$\sum_{t=1}^{n} \max_{p_t^*} \mathbb{E}_{N_t \sim \mathbb{D}}[p_t^* \cdot 1(p_t^* \leq x_t^\top \theta^* + N_t) | x_t, \theta^*, \mathbb{D}]$$

2. Expected reward of our algorithm, i.e.

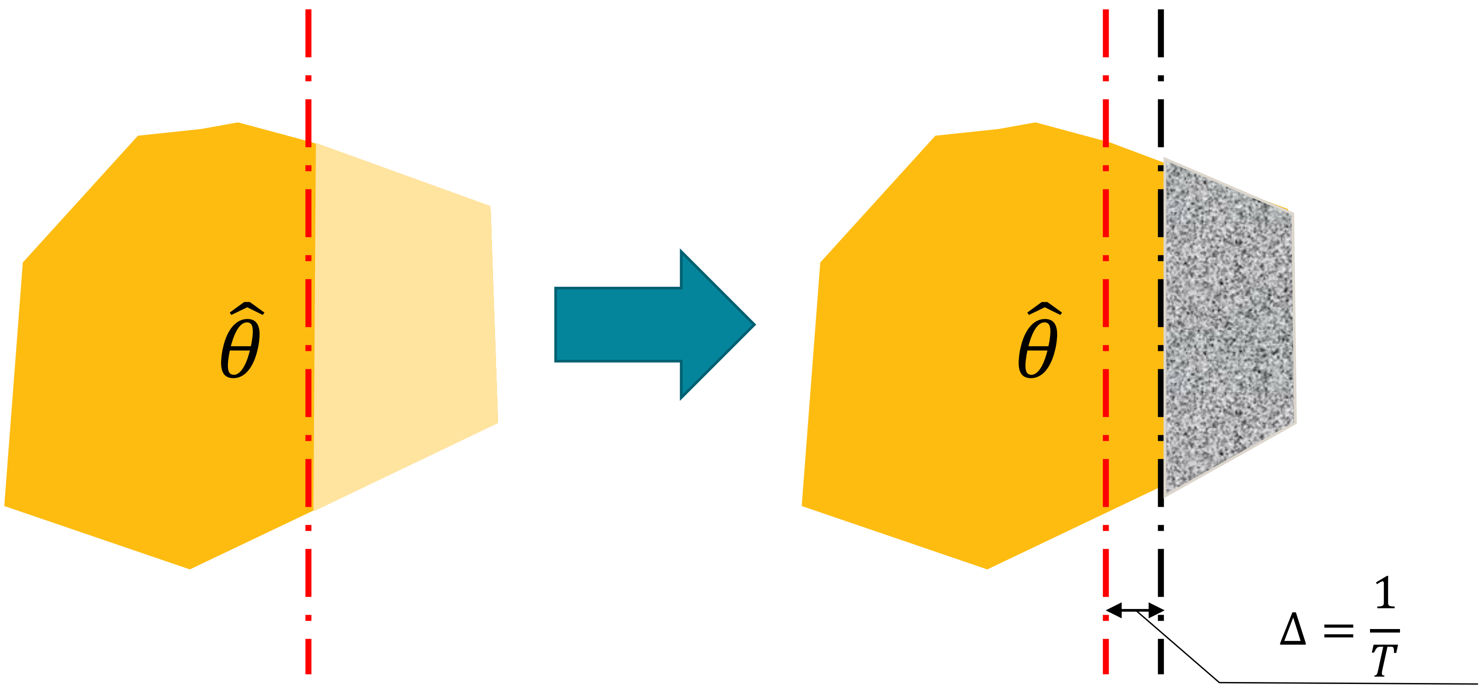
$$\sum_{t=1}^{n} \mathbb{E}_{N_t \sim \mathbb{D}}[p_t \cdot 1(p_t \leq x_t^\top \theta^* + N_t)]$$

Feature-based setting -- Regrets

	Deterministic	Noisy		
		Known distribution	Parametric distribution	Non-parametric distribution
Upper Bound	$O(\log\log T)$	$O(\log T)$	$O(\sqrt{T})$ for sub-Gaussian noise	$O\left(T^{\frac{3}{4}}\right)$
	[PLS 18]	[CLPL16] [JN19] [XW21]	[JN19]	
Lower Bound	$\Omega(\log\log T)$	$\Omega(\log T)$	$\Omega(\sqrt{T})$ for Gaussian noise	$\Omega\left(T^{\frac{2}{3}}\right)$
	[KL 03]	[JN19]	[BR12] [XW21]	[KL03]

Shallow Pricing: for small-variance noise [CLPL16]

- For noises with small ($O\left(\frac{1}{T\log T}\right)$) variance, we may directly apply binary search with compromission.



EMLP: for Known $\mathbb{D}$ and Stochastic x_t

- Since we know the noise distribution, we can use **Max Likelihood Estimate** to achieve a $\hat{\theta}$.
  - $||\hat{\theta}_t - \theta^*|| = O\left(\frac{1}{\sqrt{t}}\right)$
- We upper bound the regret with $C \cdot ||\hat{\theta}_t - \theta^*||^2$.
- The regret bound is then $= O\left(\sum_{t=1}^{T} \frac{1}{t}\right) = O(\log T)$

ONSP: for Known $\mathbb{D}$ and Adversarial x_t

- Since x_t is adversarial, for MLE we do not necessarily have $\hat{\theta} \to \theta^*$.
- However, we can make use of likelihood functions as *surrogation loss*, i.e. $Regret(\hat{\theta}) \leq C' \cdot \left(L(\theta^*) - L(\hat{\theta})\right)$
  - On the one hand, $L(\theta)$ is exp-concave (strongly convex).
  - On the other hand, we upper bound the regret with $C \cdot ||\hat{\theta}_t - \theta^*||^2$.
- We apply Online Newton Steps (ONS) to optimize $L(\theta)$.

For unknown noise distribution: why *not* a contextual bandits model?

- Action space
  - In contextual bandits, actions are discrete.
  - In dynamic pricings, actions are continuous. Extra discretization suffers an error-action trade-off.
- Policy set
  - In contextual bandits, we have a finite set of policies.
  - In feature-based dynamic pricing, we do not have a finite set of policies.

For unknown noise distribution: our approach

- Continuous action space
  - We discretize the action space at a rational scale.
- Unknown/infinite policy set
  - Define a policy as ``$\hat{\theta}$ and $\mathbb{D}$''.
  - Use covering number to finitely cover all policies.
  - Use Catalan number to count the finite subset.
  - Apply contextual bandits.
- There is still a huge gap.
  - Upper bound: $O\left(T^{\frac{3}{4}}\right)$; Lower bound: $\Omega\left(T^{\frac{2}{3}}\right)$

Reference

[KL03] Kleinberg, R., & Leighton, T. (2003, October). The value of knowing a demand curve: Bounds on regret for online posted-price auctions. In *44th FOCS, 2003. Proceedings*. (pp. 594-605). IEEE.

[BR12] Broder, J., & Rusmevichientong, P. (2012). Dynamic pricing under a general parametric choice model. *Operations Research*, *60*(4), 965-980.

[CLPL16] Cohen, Maxime C., Ilan Lobel, and Renato Paes Leme. "Feature-based dynamic pricing." *Management Science* 66.11 (2020): 4921-4943.

[PLS18] Leme, Renato Paes, and Jon Schneider. "Contextual search via intrinsic volumes." *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2018.

[JN19] Javanmard, Adel, and Hamid Nazerzadeh. "Dynamic pricing in high-dimensions." *The Journal of Machine Learning Research*20.1 (2019): 315-363.

[XW21] Xu, Jianyu, and Yu-Xiang Wang. "Logarithmic Regret in Feature-based Dynamic Pricing." *arXiv preprint arXiv:2102.10221* (2021).

UC SANTA BARBARA