03/01/2022

Towards Agnostic Feature-based Dynamic Pricing: Linear Policies vs Linear Valuation with Unknown Noise

Jianyu Xu and Yu-Xiang Wang
University of California, Santa Barbara

UC SANTA BARBARA

Outline

- Basic Setting
- Problem Modeling
  - Linear Policy (LP)
  - Linear Valuation (LV)
- Summary of Results
- Algorithm Design
  - Linear-EXP4 for LP
  - D2-EXP4 for LV
  - Half-Lipschitzness
- Numerical Results
- Open Problem

Dynamic Pricing

Single-product Pricing

Feature-based Pricing

Basic Problem Setting

- An online-fashion sales:

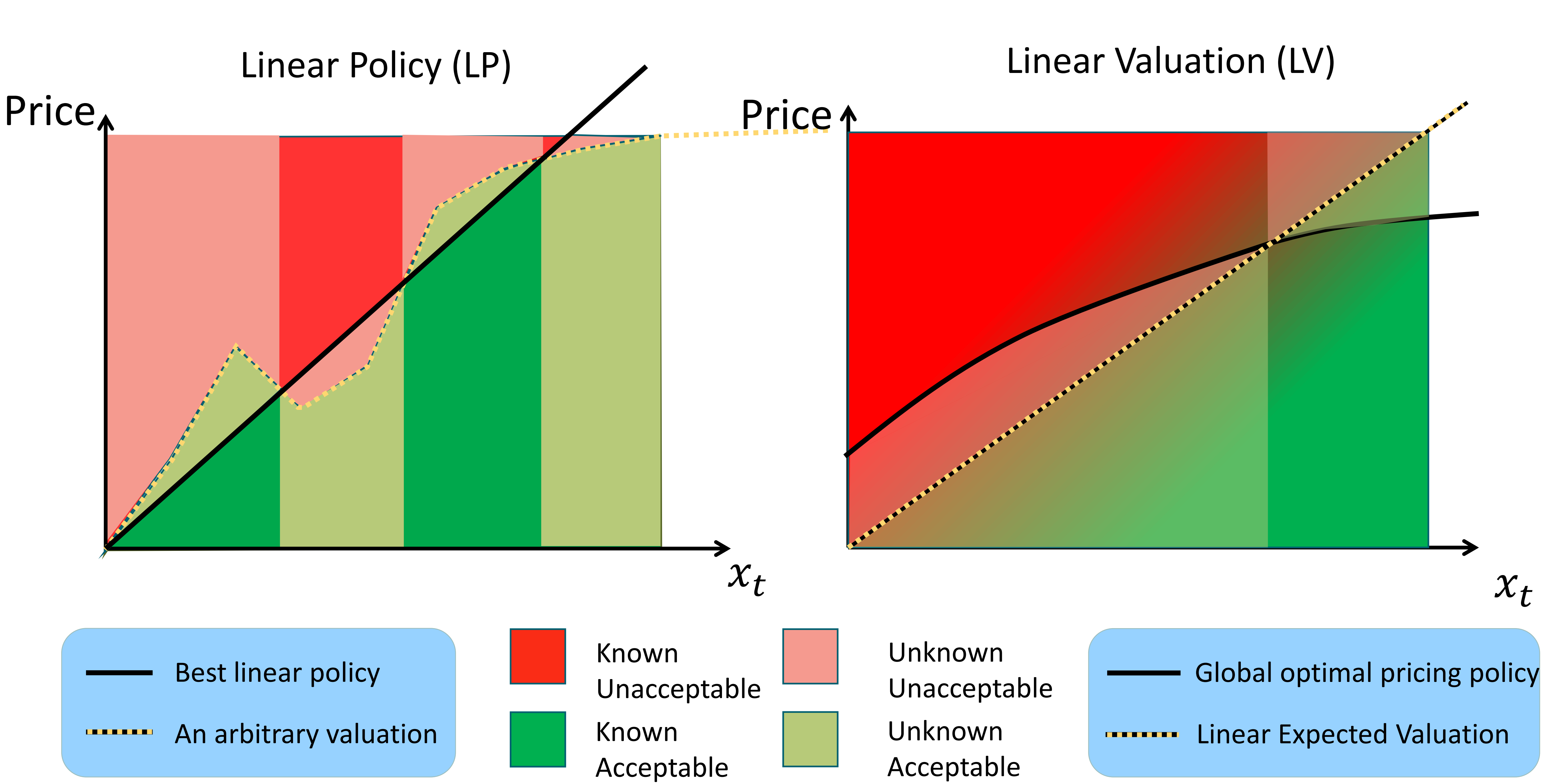
For $t = 1,2, \dots, T$:

- Feature $x_t \in \mathbb{R}^d$ is revealed;
- Customer generates a valuation y_t *secretly*;
- Seller (we) propose a price v_t;
- Customer makes a decision $1_t = 1[v_t \leq y_t]$;
- We get a reward $r_t = v_t \cdot 1_t$.

- Comparing with *contextual bandits*:
  - **Continuous** action and hypothesis spaces

Problem Modeling

- To make use of x_t, we consider two problem models:
  - *Linear Policy* (**LP**)
    - (x_t, y_t) are arbitrarily selected;
    - Compete with $v_t^* = x_t^\top \beta^*$ for a best fixed β^*.
  - *Linear Valuation* (**LV**)
    - $y_t = x_t^\top \theta^* + N_t$, where $\theta^* \in \mathbb{R}^d$ is fixed and $N_t \sim_{i.i.d.} \mathbb{D} \subseteq [-1,1]$;
    - Compete with $v_t^* = \arg\max_v v \cdot \Pr[v \leq y_t]$.
- **LP** models our *strategy*; **LV** models the *nature*.


Linear Policy (LP)
Price
x_t
Linear Valuation (LV)
Price
x_t
Best linear policy
An arbitrary valuation
Known Unacceptable
Known Acceptable
Unknown Unacceptable
Unknown Acceptable
Global optimal pricing policy
Linear Expected Valuation

LP versus LV: Regret

- LP compete with the *best fixed linear policy*:

$$Regret_{LP} \coloneqq \max_{\beta} \sum_{t=1}^{T} x_t^\top \beta \cdot \mathbb{E}[x_t^\top \beta \leq y_t] - x_t^\top \beta_t \cdot \mathbb{E}[x_t^\top \beta_t \leq y_t]$$

Max expected reward of a fixed linear policy — Expected reward of our (linear) prices

- LV compete with the *best price at each time*

$$Regret_{LV} \coloneqq \sum_{t=1}^{T} \max_{v} v \cdot \Pr[v \leq x_t^\top \theta^* + N_t | \theta^*, \mathbb{D}] - v_t \cdot \Pr[v_t \leq x_t^\top \theta^* + N_t | \theta^*, \mathbb{D}]$$

Max expected reward at time t — Expected reward of our prices

Existing Results

Problem	Linear Valuation				Linear Policy
Noise Assumption	Noise-free	Known, Log-concave	Parametric	Agnostic, Bounded	
Upper Regret Bound	$O(d \log\log T)$ [PLS18]	$O(d \log T)$ [XW21]	$\tilde{O}(d\sqrt{T})$ [WTL21]	$\tilde{O}(T^{\frac{3}{4}} + d^{\frac{1}{2}}T^{\frac{5}{8}})$ [This Work]	$\tilde{O}(d^{\frac{1}{3}}T^{\frac{2}{3}})$ [This Work]
Lower Regret Bound	$\Omega(d \log\log T)$ [KL03]	$\Omega(d \log T)$ [JN19]	$\Omega(d\sqrt{T})$ [BK21]	$\tilde{\Omega}(T^{\frac{2}{3}})$ [KL03, This Work]	$\tilde{\Omega}(d^{\frac{1}{3}}T^{\frac{2}{3}})$ [This Work]

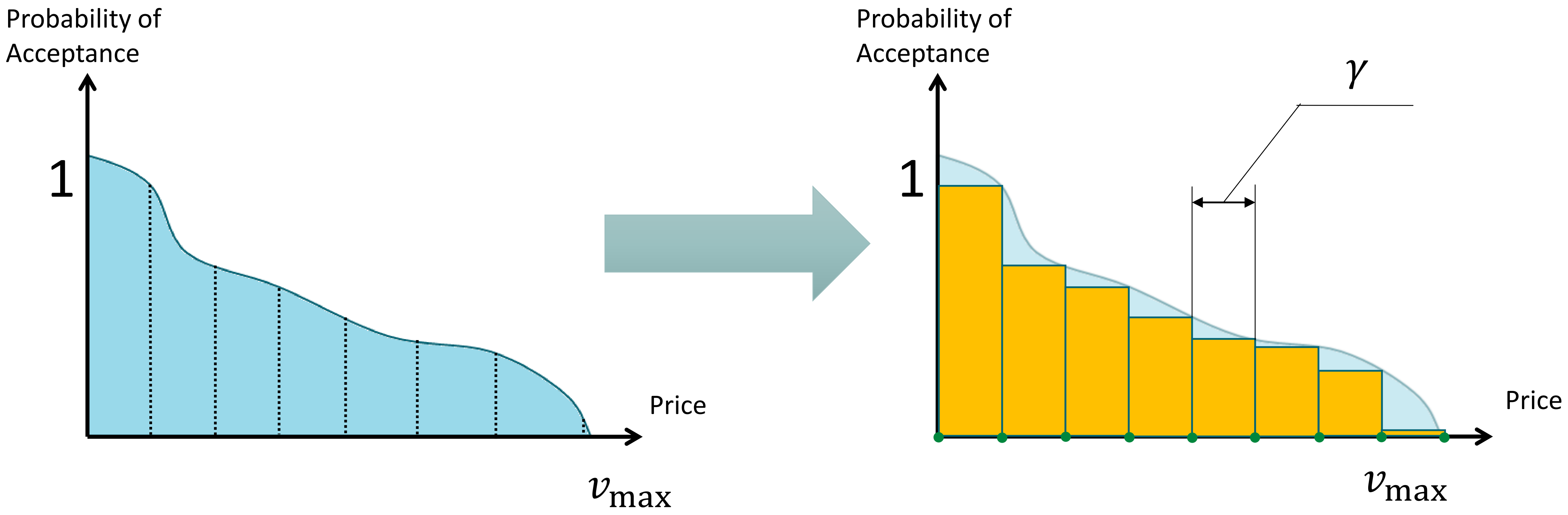
EXP-4 [ACBFS02]: a Contextual Bandit Algorithm

for $t = 1$ **to** T **do**
  Set probability $p_j(t)$ for each action j according to weights of all policies;
  Get a_t by Thompson sampling the action set A according to current probability $\{p_j(t)\}$;
  Receive a reward r_t;
  Construct an *Inverse Propensity Scoring (IPS)* estimator $\hat{r}_i(t)$ for the reward of each action i.
  Update weights w_i's according to $\hat{r}_i(t)$.
end for

- **Input:** Time horizon T, action set $\mathcal{K}$, policy set Π; features x_t at each time
- **Output:** action a_t at each time
  - approaching optimal policy π^*
  - with $O\left(\sqrt{T|\mathcal{K}|\log|\Pi|}\right)$ regret
- Only works for **finite** action/policy sets.
  - Discretize the price/hypothesis space.

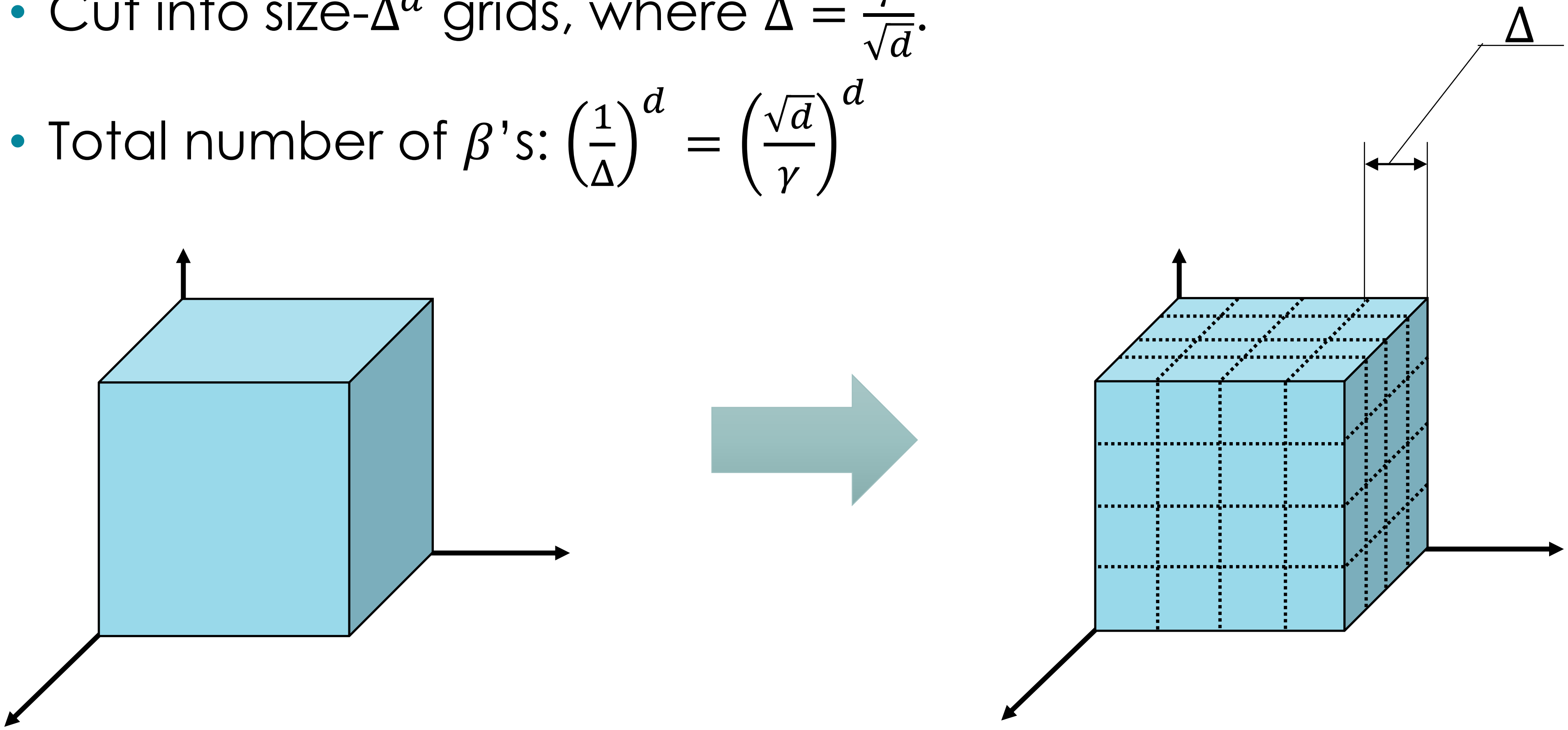
Discretization: Action Space

- Split the price range into size-γ segments.
- Action set consists of all end points



Discretization: Policy Space (1) --Vector

- Discretize the policy vector space into grids:
  - Cut into size-Δ^d grids, where $\Delta = \frac{\gamma}{\sqrt{d}}$.
  - Total number of β's: $\left(\frac{1}{\Delta}\right)^d = \left(\frac{\sqrt{d}}{\gamma}\right)^d$



Linear-EXP4: Algorithm for LP

- Action $\mathcal{K} = \{k \cdot \gamma, k = 0,1, \dots, \lfloor 1 \rfloor_\gamma\}$
- Policy $\Pi = \{\pi_\beta : \pi_\beta(x) = \lfloor x^\top \beta \rfloor_\gamma\}$, with $\beta \in$ size-Δ^d grids.
  - "γ-flooring": $\lfloor a \rfloor_\gamma = \left\lfloor \frac{a}{\gamma} \right\rfloor \cdot \gamma$.
- Let $\gamma = d^{\frac{1}{3}} T^{-\frac{1}{3}}$, and the regret = $\tilde{O}\left(d^{\frac{1}{3}} T^{\frac{2}{3}}\right)$.
  - Notice that $|\mathcal{K}| = O\left(\frac{1}{\gamma}\right), |\Pi| = \left(\frac{\sqrt{d}}{\gamma}\right)^{\mathrm{d}}$,
  - matching the discretization error $O(T\gamma)$.

Discretization: Policy Space (2) -- Distribution

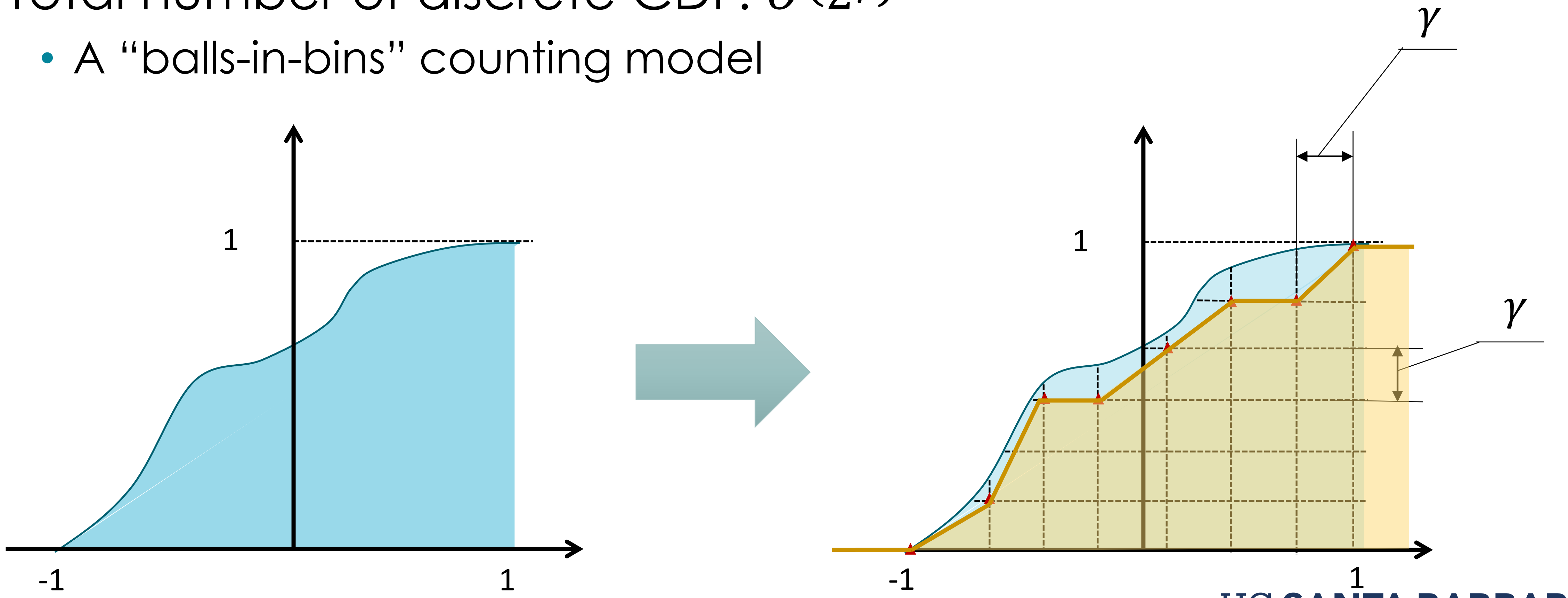
- In LV, recall: $y_t = x_t^\top \theta^* + N_t$
  - with $\theta^* \in \mathbb{R}^d$ fixed and $N_t \sim \mathbb{D}$.
- If we know θ^* and $\mathbb{D}$, then …

$$\pi^*(x) = \arg\max_v v \cdot \Pr[v \le y_t] = \arg\max_v v \cdot \left(1 - F_{\mathbb{D}}(v - x_t^\top \theta^*)\right)$$

  - $F_{\mathbb{D}}$ is the CDF of $\mathbb{D}$.
- Idea: policy built on both $\hat{\theta}$ and $\hat{F}_{\mathbb{D}}$.

Discretization: Policy Space (2) -- Distribution

- 3 steps to discretize $F_{\mathbb{D}}$: Griding, Flooring, Connecting
- Total number of discrete CDF: $O\left(2^{\frac{3}{\gamma}}\right)$
  - A "balls-in-bins" counting model



D2-EXP4: Algorithm for LV

- Still, we play the EXP-4:
  - Action $\mathcal{K} = \{k \cdot \gamma, k = 0,1, \dots, \lfloor 1 \rfloor_\gamma\}$
  - Policy $\Pi = \{\pi(x; \hat{\theta}, \hat{F}) := \arg\max_v v \cdot \left(1 - \hat{F}(v - x^\top \hat{\theta})\right) - (B+1)\gamma\}$
    - $\hat{\theta} \in$ size-Δ^d grids, $\hat{F} \in$ discrete CDF family.
    - Subtracting $(B+1)\gamma$ for more chance to succeed.
- Choose $\gamma = T^{-\frac{1}{4}}$, and $Regret = \tilde{O}(T^{\frac{3}{4}} + d^{\frac{1}{2}} T^{\frac{5}{8}})$.
  - Matching the discretization error $O(T\gamma)$.
- Prove a $\tilde{\Omega}(T^{\frac{2}{3}})$ lower bound.

Property: Half-Lipschitzness

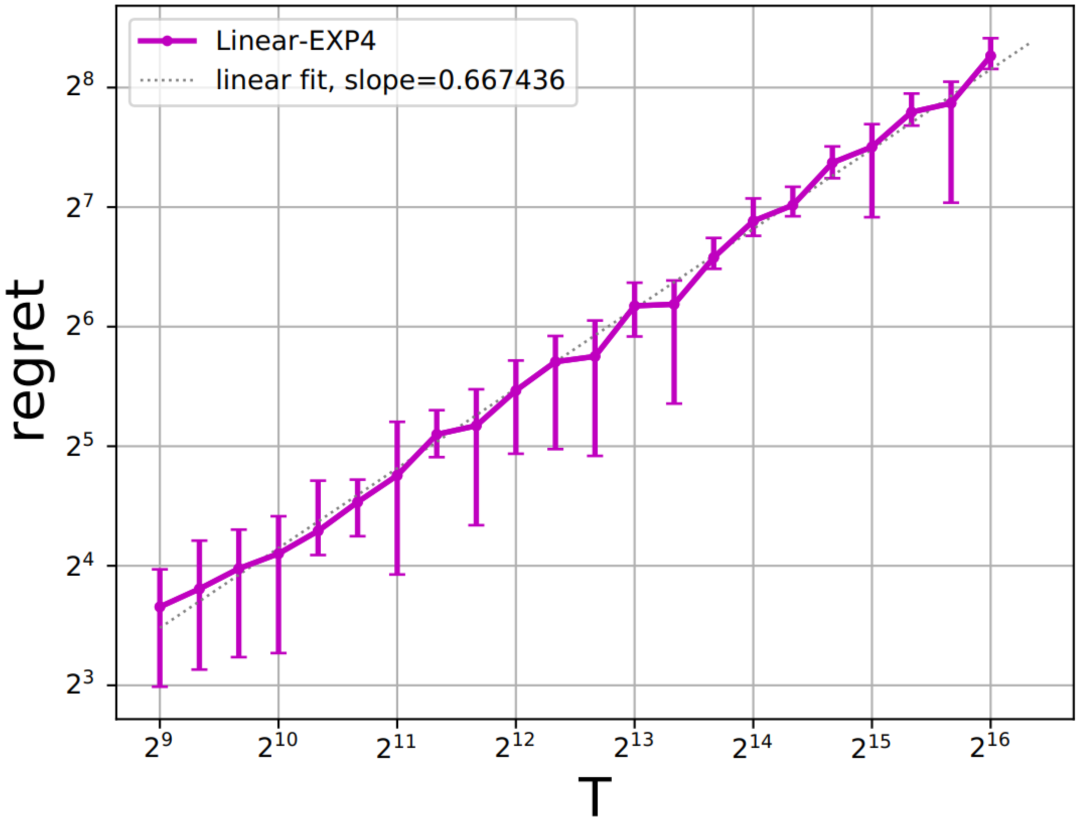
- Noise distribution is not necessarily Lipschitz.
- Probability of acceptance never increases

$$
\begin{aligned}
R(v) &= v \cdot \Pr[v \le y_t] \\
&\ge v \cdot \Pr[(v+\delta) \le y_t] \\
&\ge (v+\delta) \cdot \Pr[(v+\delta) \le y_t] - \delta \\
&= R(v+\delta) - \delta
\end{aligned}
$$

- Expected reward $R(v)$ increament $\le$ Price v increament
  - We call this property a "**half-Lipschitz**".
- Therefore, we only suffer δ-more regret (*discretization error*) by choosing a δ-conservative price, i.e. choosing $(\hat{v} - \delta)$.
  - This enables us to discretizing the action/policy spaces

Numerical Experiments on Linear-EXP4

- A log-log plot of regret
  - r-slope indicates $O(T^r)$ regret.
- 2/3 in theory, 0.67 in practice
- D2-EXP4 consumes EXP-time!
  - Code released



Open Problem: Regret Gap of LV

- Our $\tilde{O}(T^{\frac{3}{4}})$ result holds for **any** noise CDF, w/ or w/o continuity.

- For m^{th}-order smooth CDF, [FGY21] shows a $\tilde{O}(T^{\frac{2m+1}{4m-1}})$ regret.
  - Non-trivial for $m \geq 2$.
  - Still unmatched with the $\tilde{O}(T^{\frac{m+1}{2m+1}})$ lower bound presented in [WCSL21]
- [LS21] achieves a $\tilde{O}(T^{\frac{2}{3} \vee (1-\alpha)})$ regret by assuming a good estimator $\hat{\theta}_t$: $||\hat{\theta}_t - \theta^*||_2 = O(t^{-\alpha})$ with logged data.
  - The existence is unknown and is highly non-trivial.

References

- [ACBFS02] Auer, P., Cesa-Bianchi, N., Freund, Y., & Schapire, R. E. (2002). The nonstochastic multiarmed bandit problem. SIAM journal on computing, 32(1), 48-77.
- [KL03] Kleinberg, R. and Leighton, T. (2003). The value of knowing a demand curve: Bounds on regret for online posted-price auctions. In FOCS-03, pages 594-605. IEEE.
- [CLPL16] Cohen, M. C., Lobel, I., and Paes Leme, R. (2016). Feature-based dynamic pricing. In EC-16, pages 817-817.
- [PLS18] Leme, R. P. and Schneider, J. (2018). Contextual search via intrinsic volumes. In FOCS-18, pages 268-282. IEEE.
- [JN19] Javanmard, A. and Nazerzadeh, H. (2019). Dynamic pricing in high-dimensions. The Journal of Machine Learning Research, 20(1):315-363.
- [XW21] Xu, J., & Wang, Y. X. (2021). Logarithmic regret in feature-based dynamic pricing. In NeurIPS-21, 34.
- [WCSL21] Wang, Y., Chen, B., & Simchi-Levi, D. (2021). Multimodal dynamic pricing. Management Science, 67(10), 6136-6152.
- [LS21] Luo, Y., Sun, W. W., et al. (2021). Distribution-free contextual dynamic pricing. arXiv preprint arXiv:2109.07340.
- [FGY21] Fan, J., Guo, Y., & Yu, M. (2021). Policy Optimization Using Semiparametric Models for Dynamic Pricing. Available at SSRN 3922825.
- [WTL21] Wang, H., Talluri, K., & Li, X. (2021). On Dynamic Pricing with Covariates. arXiv preprint arXiv:2112.13254.
- [BK21] Ban, G. Y., & Keskin, N. B. (2021). Personalized dynamic pricing with machine learning: High-dimensional features and heterogeneous elasticity. Management Science, 67(9), 5549-5568.

UC SANTA BARBARA