JOINT PRICING + INVENTORY MANAGEMENT WITH DEMAND LEARNING

Yuan Zhou, *YMSC, Tsinghua University*

Joint work with Beryl Chen, David Simchi-Levi & Yining Wang

PRICING AND INVENTORY CONTROL

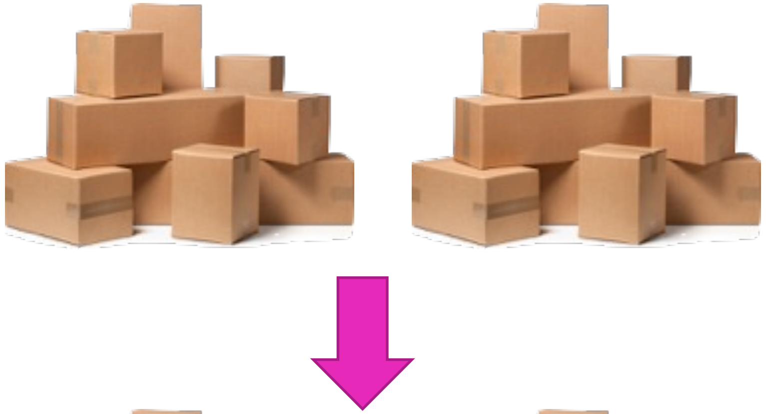
- Coordination of ***pricing*** and ***inventory control:*** two fundamental problems in operations management

- ***Pricing:*** the task of balance revenue and demand
  - ✓ The higher the price, the higher the revenue but also lower the expected demand: $\mathrm{E}[d_t|p_t] = D_0(p_t)$

- ***Inventory management:*** the question of re-ordering inventory stocks.
  - ✓ Need to balance ordering cost, holding cost and out-of-inventory cost (e.g., backlogging).

THE DECISION PROCESSES

- ***Step 1:*** inventory decisions.



At the beginning of time t, inventory level is x_t

Order-up-to level $y_t \geq x_t$

Ordering cost = $\underbrace{k \times 1[y_t > x_t]}_{\text{fixed cost}} + \underbrace{c(y_t - x_t)}_{\text{variable cost}}$

THE DECISION PROCESSES

- ***Step 2:*** pricing decisions.

Order-up-to level $y_t \geq x_t$

Price p_t, leading to realized demand d_t

The "additive" noisy demand model: $d_t = D_0(p_t) + \beta_t$

Remaining inventory: $x_{t+1} = y_t - d_t$

Sales revenue: $p_t(y_t - x_{t+1})$

"Censored" demand setting:

$x_{t+1} = \max\{0, y_t - d_t\}$

THE DECISION PROCESSES

- ***Step 3:*** holding/backlogging/lost-sales cost

Remaining inventory: $x_{t+1} = y_t - d_t$

"Censored" demand setting:

$x_{t+1} = \max\{0, y_t - d_t\}$

✓ $x_{t+1} > 0$: holding cost

✓ $x_{t+1} < 0$: backlogging/loss-of-good-will cost

✓ We use $h(\cdot)$ function to represent **both** costs.

THE DECISION PROCESSES

- Summary of the decision process:
  - ✓ **State**: x_t, the inventory level at the beginning of time t
  - ✓ **Decisions**: y_t (the order-up-to level), p_t (the price).
  - ✓ **State transition – backlogged**: $x_{t+1} = y_t - d_t = y_t - D_0(p_t) - \beta_t$
  - ✓ **State transition – censored**: $x_{t+1} = \max(0, y_t - d_t)$

Learning-while-Doing problem:
$D_0, \beta_t \sim P$ are unknown

- Immediate reward:
  - ✓ **Backlogging**:
    $-k \times 1\{y_t > x_t\} - c(y_t - x_t) + p_t(D_0(p_t) + \beta_t) - h(y_t - D_0(p_t) - \beta_t)$
  - ✓ **Censored demand:**
    $-k \times 1\{y_t > x_t\} - c(y_t - x_t) + p_t \min(y_t, D_0(p_t) + \beta_t) - h(y_t - D_0(p_t) - \beta_t)$

[1] Chen et al.' 20, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3632475
[2] Chen et al.'21, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3750413

COMPARISON WITH EXISTING RESULTS

✓ indicates optimal regret (up to poly-logarithmic terms)

	$k>0$?	*Pricing model*	*Censored demand?*	*Concavity?*	*Regret*
Yuan et al.'21	Yes	N/A	Yes	Implied	$\tilde{O}(\sqrt{T})$ ✓
[1]	**Yes**	**GLM**	**No**	**No**	$\tilde{\boldsymbol{O}}(\sqrt{\boldsymbol{T}})$ ✓
Huh & Rusmevichientong' 09	No	N/A	Yes	Implied	$O(\log T)$ ✓
Chen et al.'19	No	Non-param.	No	Implied	$\tilde{O}(\sqrt{T})$ ✓
Chen et al.'21	No	Non-param.	Yes	Assumed	$T^{\frac{1}{2}+o(1)}$
[2]	**No**	**Non-param.**	**Yes**	**No**	$\tilde{\boldsymbol{O}}(\boldsymbol{T}^{\frac{3}{5}})$ ✓

PART 1 (FIXED ORDERING COSTS)

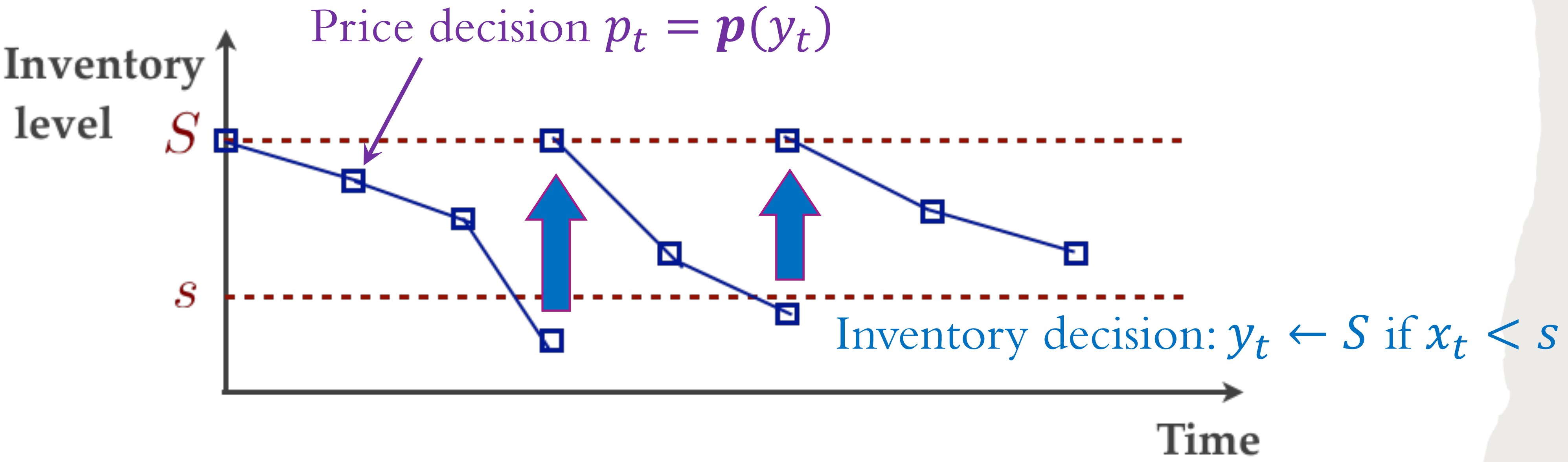
[1] Chen et al.' 20, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3632475

- Model primitives:
  - ✓ **Backlogging obs:** $o_t = d_t = D_0(p_t) + \beta_t$
  - ✓ **Fixed cost:** $k > 0$
  - ✓ **V-shaped costs**: $h(\cdot) = h\max(0,\cdot) - b\min(0,\cdot)$
  - ✓ **Linear demand:** $D_0(p) = \langle \phi(p), \theta \rangle$ (can be extended to GLM)

$$-k\times 1\{y_t > x_t\} - c(y_t - x_t) + p_t(D_0(p_t) + \beta_t) - h(y_t - D_0(p_t) - \beta_t)$$

FULL-INFORMATION SOLUTION

- [Chen and Simchi-Levi 2004a, 2004b] In the long run, the optimal policy is an $(s, S, \boldsymbol{p})$-policy
  - ✓ S: the order-up-to level
  - ✓ s: the inventory threshold below (or at) which ordering is initiated
  - ✓ $\boldsymbol{p}$: a pricing functions that maps x_t to p_t



FULL-INFORMATION SOLUTION

- [Chen and Simchi-Levi 2004a, 2004b] In the long run, the optimal policy is an $(s, S, \boldsymbol{p})$-policy. Given (s,S), the optimal $\boldsymbol{p}$ can be computed using DP:
- Let $\phi(x, r) = \sup_{\boldsymbol{p}}\{\mathrm{E}[\sum_{t=1}^{\tau}(r_t - r)]\}$ given initial inventory level x
- Recursion formula:

$$\phi(x; r) = \begin{cases} \sup_{p}\{H_0(x,p) - r + \mathrm{E}_\beta[\phi(x - D_0(p) - \beta; r)]\}, & x \geq s \\ -k, & x < s \end{cases}$$

  ✓ Immediate reward $H_0(x,p) = -\mathrm{E}_\beta[h(x - D_0(p) - \beta)] + (p - c)D_0(p)$

- Binary search of r: maximum r is the optimal per-period reward.
- Optimal $\boldsymbol{p}$ must satisfy $\phi(x, r) = 0$, where r is the per-period reward of $\boldsymbol{p}$

FULL-INFORMATION SOLUTION

- [Chen and Simchi-Levi 2004a, 2004b] The optimal policy is an $(s, S, \boldsymbol{p})$-policy
- Question: can we learn about the demand rate, and adopt near-optimal pricing + inventory control, at the same time?

  ✓ Also known as the "**Learning-While-Doing**" question.

  ✓ Has seen surging research interests in operations management recently.

EXISTING APPROACHES

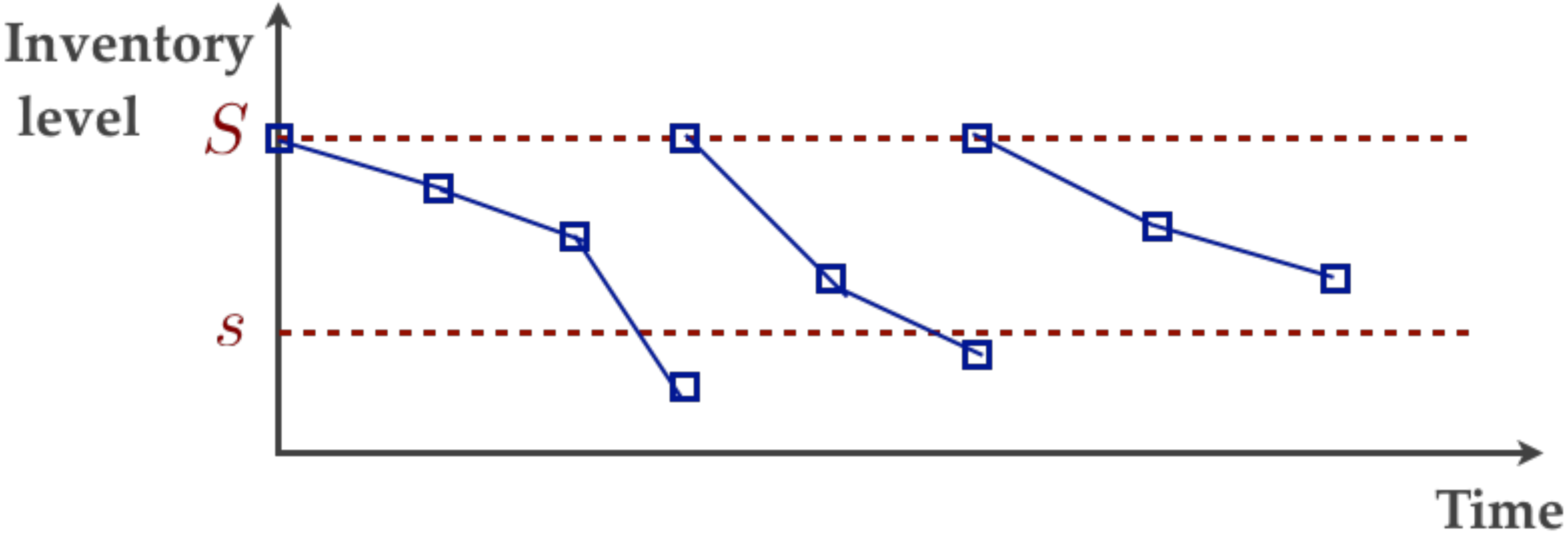
- Explore-then-exploit: [Chen et al., 2019, 2020] and more
  - ✓ Completely separates learning and optimization.
  - ✓ Only successful with strong ***convexity/concavity*** structures; otherwise leading to sub-optimal $O(T^{2/3})$ regret.

- Stochastic gradient descent: [Yuan et al., 2021], [Ban, 2020], and more
  - ✓ Using (noisy) optimization methods to find good policies
  - ✓ Also require ***convexity/concavity*** structures.
  - ✓ Very difficult to handle ***infinite-dimensional*** objects, such as the price function $\boldsymbol{p}: [s, S] \to \mathrm{R}^{+}$

1., assuming $\beta_t \sim P$ is ***known***.

JOINT LEARNING AND OPTIMIZING

- Divide T periods into (variable-length) ***epochs***



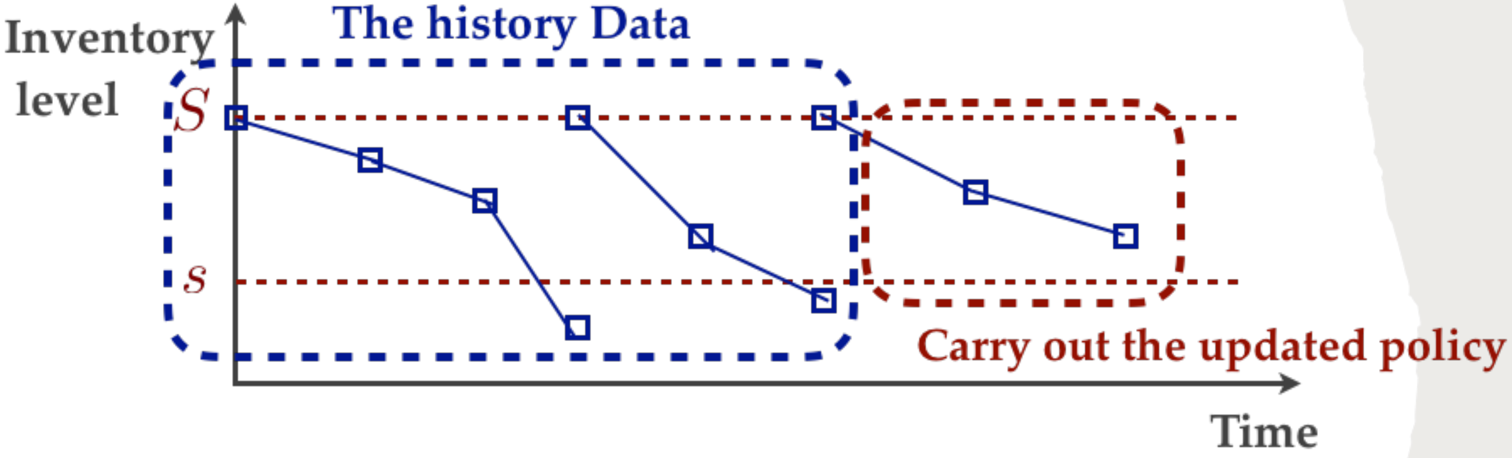
- Epochs start with order-up-to S and ends with $x_t < s$
- Update $(s, S, \boldsymbol{p})$ at the end of each epoch

1., assuming $\beta_t \sim P$ is ***known***.

JOINT LEARNING AND OPTIMIZING

UCB for D_0 during epoch b:

$$\overline{D}_b(p) = \langle \phi(p), \hat{\theta}_b \rangle + \Delta_b(p)$$



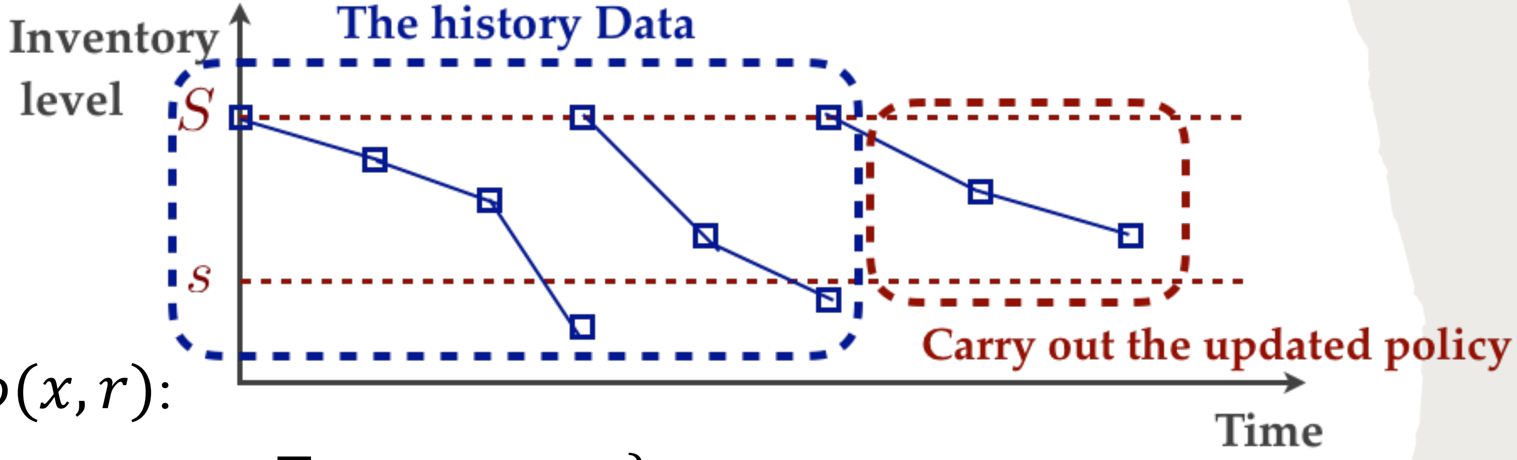
OLS with LinUCB:

- ✓ $\hat{\theta}_b = \arg\min_{\theta} \sum_{\tau<t} |d_\tau - \langle \phi(p_\tau), \theta \rangle|^2 + \|\theta\|_2^2$
- ✓ $\Delta_b(p) = C\sqrt{\phi(p)^T \Lambda_b^{-1} \phi(p)}$, where $\Lambda_b = I + \sum_{\tau<t} \phi(p_\tau)\phi(p_\tau)^T$
- ✓ Satisfies $\overline{D}_b(p) \geq D_0(p) \geq \overline{D}_b(p) - 2\Delta_b(p)$

1., assuming $\beta_t \sim P$ is ***known***.

JOINT LEARNING AND OPTIMIZING

UCB for D_0 during epoch b:



$$\bar{D}_b(p) = \langle \phi(p), \hat{\theta}_b \rangle + \Delta_b(p)$$

Use $\bar{D}_b(p)$ to calculate the DP $\phi(x, r)$:

$$\phi(x; r) = \begin{cases} \sup_p \{ \bar{H}_b(x, p) - r + \mathrm{E}_\beta[\phi(x - \bar{D}_b(p) - \beta; r)] \}, & x \geq s \\ -k, & x < s \end{cases}$$

Estimated immediate reward $\bar{H}_b(p) = -\mathrm{E}_\beta[h(x - \bar{D}_b(p) - \beta)] + (p - c)\bar{D}_b(p)$

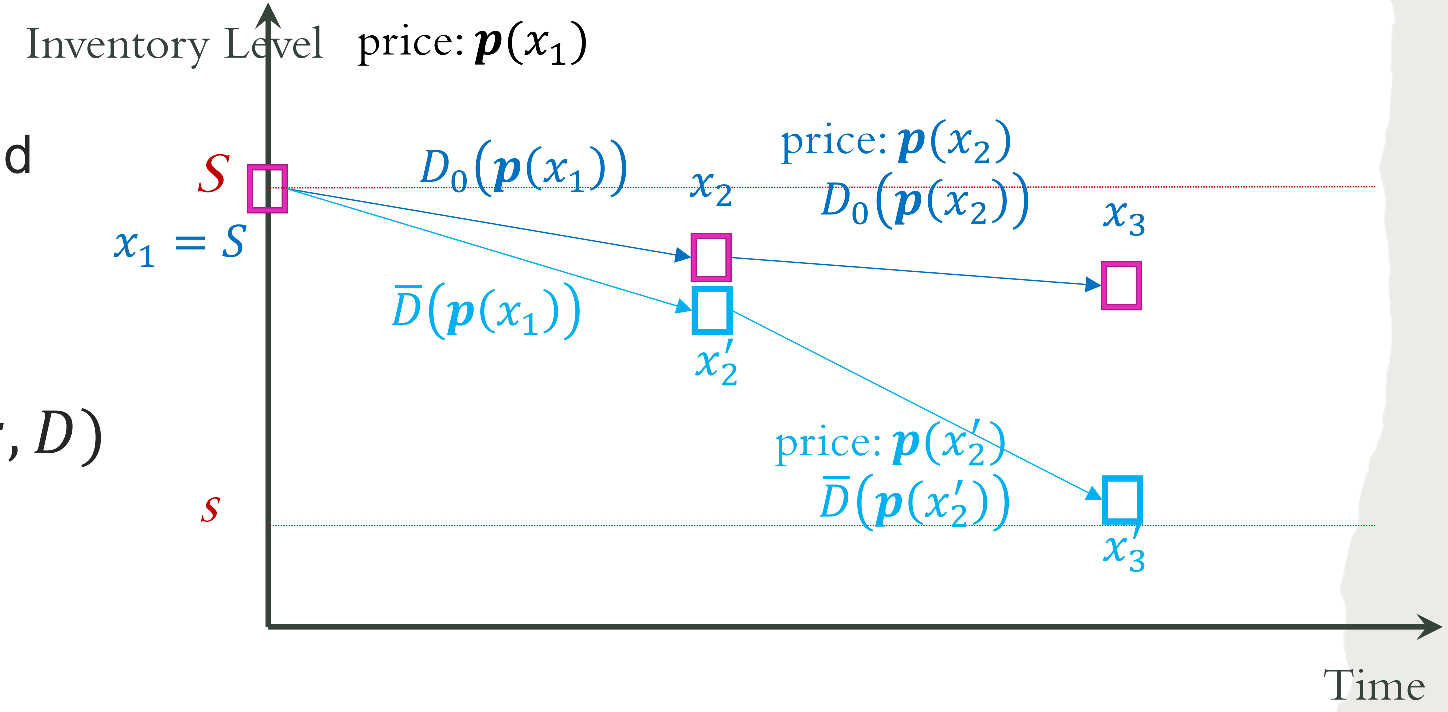
Key technical challenge: prove that

$$E^{\hat{\pi}}[\Sigma_{t \in E_b} \bar{r}_b - r_t] \leq O(1) \times E^{\hat{\pi}}[\Sigma_{t \in E_b} \Delta_b(p_t)]$$

1., assuming $\beta_t \sim P$ is ***known***.

JOINT LEARNING AND OPTIMIZING

- Objective: prove $E\left[\Sigma_{t\in E_b} \bar{r}_b - r_t\right] \leq O(1)\times E\left[\Sigma_{t\in E_b} \Delta_b(p_t)\right]$
- Plan: unroll the trajectory under D_0 and $\bar{D}$, and compare them.
- Challenge:
  - ✓ $|x_2 - x_2'| \leq \Delta(x_1)$
  - ✓ $|\boldsymbol{p}(x_2) - \boldsymbol{p}(x_2')|$ unbounded
  - ✓ $|x_3 - x_3'|$ unbobunded



- Solution: ***stability*** of $\phi(\cdot; r, D)$

1., assuming $\beta_t \sim P$ is ***known***.

JOINT LEARNING AND OPTIMIZING

- Objective: prove $E\left[\sum_{t\in E_b} \bar{r}_b - r_t\right] \leq O(1)\times E\left[\sum_{t\in E_b} \Delta_b(p_t)\right]$
- For *any* pricing function $\boldsymbol{p}(\cdot)$ and demand function D**,** define

$$\psi(x;r,D,\boldsymbol{p}) = \begin{cases} H(x,\boldsymbol{p}(x);D) - r + E_\beta[\psi(x - D(\boldsymbol{p}(x) - \beta; r, D, \boldsymbol{p})], & x \geq s \\ -k, & x < s \end{cases}$$

✓ Easy to verify that $\phi(x;r,D) = \psi(x;r,D,\boldsymbol{p}^*)$ where $\boldsymbol{p}^*$ solves ϕ

- Key stability lemma: for $\boldsymbol{p}$ which solves $\boldsymbol{\phi}(\cdot;\bar{r},\bar{D})$,

$$|\psi(x;r,\bar{D},\boldsymbol{p}) - \psi(x;r,D,\boldsymbol{p})| \leq O(1)\times \mathrm{E_D}\left[\sum_{t=1}^{\tau} \Delta\big(\boldsymbol{p}(x_t)\big)\right]$$

✓ Implies the objective, because $\psi(x;\bar{r},\bar{D},\boldsymbol{p}) = 0$ and $\psi(x;\bar{r},D_0,\boldsymbol{p}) = E\left[\sum_{t\in E_b} r_t - \bar{r}\right]$

ESTIMATION OF NOISE DISTRIBUTION

- Use the empirical distribution to estimate $\beta_t \sim P$
- Two technical challenges:
  - ✓ **Error propagation**: estimation quality of *P* also depends on estimation quality of D_0
  - ✓ **Data correlation**: the $\{\beta_t\}_t$ samples are actually *not* independent and identically distributed.

ESTIMATION OF NOISE DISTRIBUTION

- ***Error propagation***: estimation quality of P also depends on estimation quality of D_0
  - ✓ How to obtain samples of noises? $\hat{\beta}_t = d_t - \langle \phi(p_t), \hat{\theta}_t \rangle$
  - ✓ The quality of $\hat{\beta}_t$ depends on the quality of $\hat{\theta}_t$
  - ✓ The estimation is **not** accurate on *all* prices

$$|\bar{D}(p) - D_0(p)| \leq 2\Delta(p) \leq 2C\sqrt{\phi(p)^T \Lambda^{-1} \phi(p)}$$

- ***Solution.*** Only use those periods with accurate demand predictions.

$$\tilde{E}_{<b} = \left\{t \in B_1 \cup \cdots \cup B_{b-1} \colon \Delta_{b(t)}(p_t) \leq \kappa/\sqrt{b}\right\}$$

ESTIMATION OF NOISE DISTRIBUTION

- ***Data correlation***: the $\{\beta_t\}_t$ samples are actually *not* independent and identically distributed.
  - ✓ β_t depends on the (s, S, p) policy used during that time period
  - ✓ The (s, S, p) policy further depends on noises from previous periods.
- ***Solution.*** Uniform concentration via *Wasserstein's distance:*

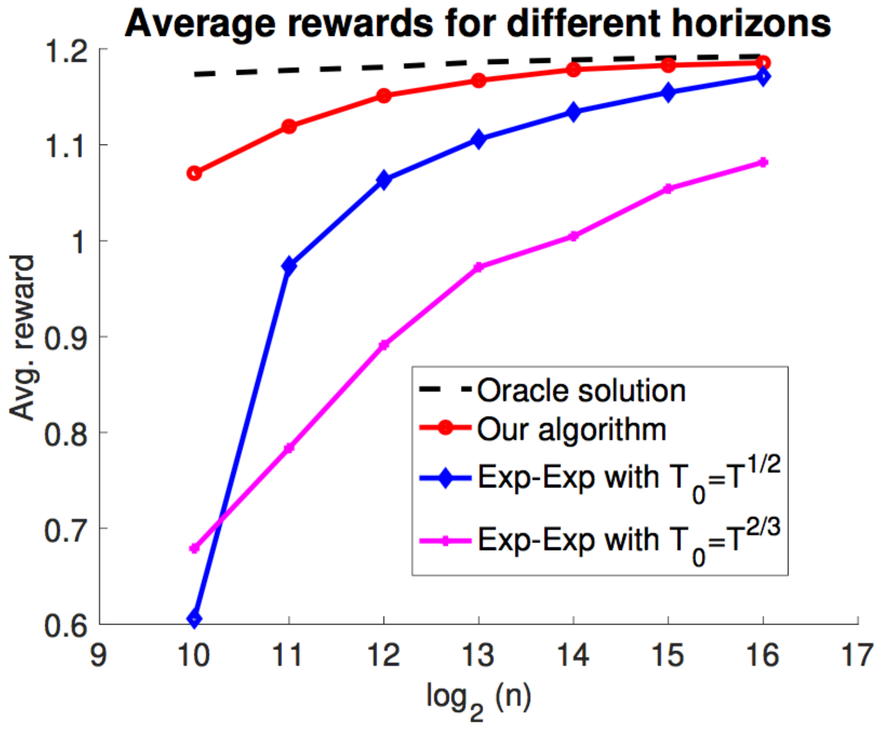
$$W_1(P, \hat{P}) = \inf_{\xi \in \Xi(P, \hat{P})} \int |x - y| \mathrm{d}\xi(x, y)$$

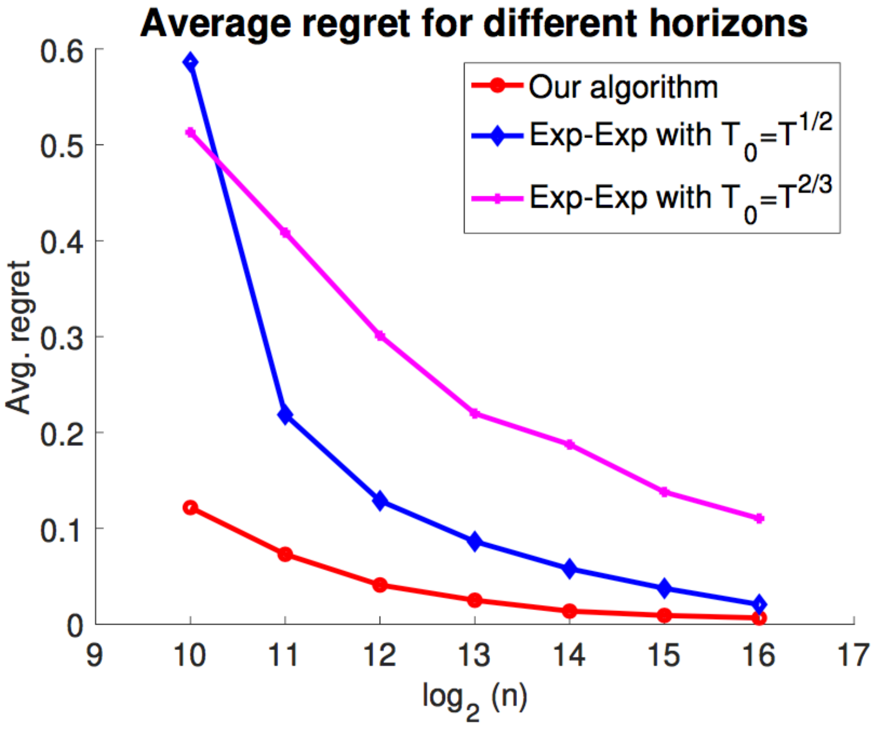
  - ✓ For **any** function *f* that is *L*-Lipschitz continuous,
    $|E_P[f(x)] - E_{\hat{P}}[f(x)]| \leq W_1(P, \hat{P})$

$D_0(p) = 18 - 15p, h(x) = 0.05\max\{x, 0\} - \min\{x, 0\}, k = 10$

NUMERICAL RESULTS

- Summary: $\tilde{O}(\sqrt{T})$ regret, which is optimal
- Numerical results: compare with Explore-Then-Commit baseline:





[1] Chen et al.' 20, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3632475
[2] Chen et al.'21, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3750413

COMPARISON WITH EXISTING RESULTS

✓ indicates optimal regret (up to poly-logarithmic terms)

	$k > 0$?	*Pricing model*	*Censored demand?*	*Concavity?*	*Regret*
Yuan et al.'21	Yes	N/A	Yes	Implied	$\tilde{O}(\sqrt{T})$ ✓
[1]	**Yes**	**GLM**	**No**	**No**	$\boldsymbol{\tilde{O}(\sqrt{T})}$ ✓
Huh & Rusmevichientong' 09	No	N/A	Yes	Implied	$O(\log T)$ ✓
Chen et al.'19	No	Non-param.	No	Implied	$\tilde{O}(\sqrt{T})$ ✓
Chen et al.'21	No	Non-param.	Yes	Assumed	$T^{\frac{1}{2}+o(1)}$
[2]	**No**	**Non-param.**	**Yes**	**No**	$\boldsymbol{\tilde{O}(T^{\frac{3}{5}})}$ ✓

PART 2 (CENSORED DEMANDS)

[2] Chen et al.'21, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3750413

- Model primitives:
  - ✓ **Censored demands:** $o_t = \min\{y_t, d_t\} = \min\{y_t, D_0(p_t) + \beta_t\}$
  - ✓ **No fixed cost:** $k = 0$
  - ✓ **V-shaped costs**: $h(\cdot) = h\max(0,\cdot) - b\min(0,\cdot)$
  - ✓ **Nonparametric demand:** $D_0(p)$ is strictly monotonically decreasing and twice continuously differentiable

$$-k\times 1\{y_t > x_t\} - c(y_t - x_t) + p_t\min\{y_t, D_0(p_t) + \beta_t\} - h(y_t - D_0(p_t) - \beta_t)$$

FULL-INFORMATION SOLUTION

- [Sobel 1981] In the long run, the optimal policy is ***stationary*** and ***myopic***
  - ✓ Define $r(p) = (p - c)D(p)$ and

$$Q(p,y) \coloneqq r(p) - (b+p)E[(D_0(p) + \beta - y)^+] - hE[(y - D_0(p) - \beta)^+]$$

  - ✓ Value of the optimal policy $\leq T \times \max_{p,y} Q(p,y)$
  - ✓ Static policy committing to $p^*, y^* = \arg\max_{p,y} Q(p,y)$ has $O(\sqrt{T})$ regret.

Learning-while-Doing problem:
$D_0, \beta_t \sim P$ are unknown

HIGH-LEVEL IDEA

- Fix p, finding $y^*(p) = \arg\max_y Q(p, y)$ is easy:
  - ✓ $Q(p,\cdot)$ is concave in y, and $E\left[\partial_y Q(p,y)\right] = (b+p)\mathbf{1}\{d \geq y\} - h\mathbf{1}\{d < y\}$
  - ✓ Can use either SGD [Huh & Rusmevichientong' 09] or bisection search.
- Discretize into $T^{0.2}$ prices and run Multi-Armed bandit
  - ✓ Strong smoothness of $Q(\cdot,\cdot)$ implies an $\tilde{O}(T^{0.6})$ regret
- Where's the catch?

$$Q(p,y) = E[(p-c)\min\{y, D_0(p)+\beta\}] - hE[(y - D_0(p) - \beta)^+] - bE[(D_0(p)+\beta-y)^+]$$

COMPARISON OF ORACLES

Let $r(a)$ be the expected immediate reward with action a:

✓ **0^{th}-order oracle:** $E[s|a] = r(a)$

✓ **1^{st}-order oracle:** $E[s|a] = r'(a)$

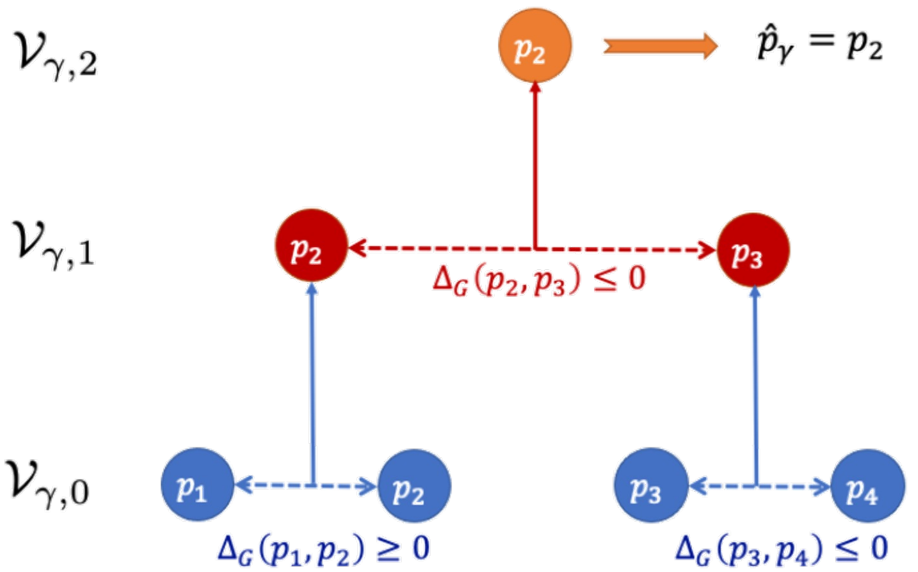
	Pricing?	*Inventory replenishment?*	*0^{th}-order oracle?*	*1^{st}-order oracle?*
Huh et al.'09	No	Yes	No	Yes
Wang et al.'10	Yes	No	Yes	No
This paper	Yes	Yes	No	No

PAIRWISE COMPARISON ORACLE

- Let $G(p) = \max_y Q(p, y)$
- For $p < p'$, let $y^*(p), y^*(p')$ be the y's that maximize Q, which are easy to obtain as explained in the previous slides.
- Can we estimate "***pairwise comparison***" objective $G(p') - G(p)$, using censored demands?

MAB WITH PAIRWISE COMPARISON

- For any p, p', we can estimate $\Delta(p, p') = G(p') - G(p)$ with error decaying at $\sim 1/\sqrt{n}$, where n is the # of samples involved
- How to use this "pairwise comparison" oracle to do MAB?
- ***Solution.*** Tournament + elimination



Uses the winner of the tournament $\hat{p}_\gamma = p_2$

$\Delta_\gamma = 0.2$

- Price p_1: $\Delta_G(\hat{p}_\gamma, p_1) = -0.4 < -\Delta_\gamma$ ✗
- Price p_3: $\Delta_G(\hat{p}_\gamma, p_3) = -0.3 < -\Delta_\gamma$ ✗
- Price p_4: $\Delta_G(\hat{p}_\gamma, p_4) = -0.15 \geq -\Delta_\gamma$ ✓

Update: $S_{\gamma+1} \leftarrow \{p_2, p_4\}$

⋆LOWER BOUND

- How to prove lower bounds for noise distributions P that are
  - ✓ Bounded a.s. with pdf $\geq c_0 > 0$ uniformly;
  - ✓ Do not change with actions.
- The classical arguments based on KL-divergence doesn't work
  - ✓ Supports of observables shift with different actions.
  - ✓ The KL-divergence would be infinity!
- ***Solution.*** Generalized square Hellinger's distance (s=2: std Hellinger)

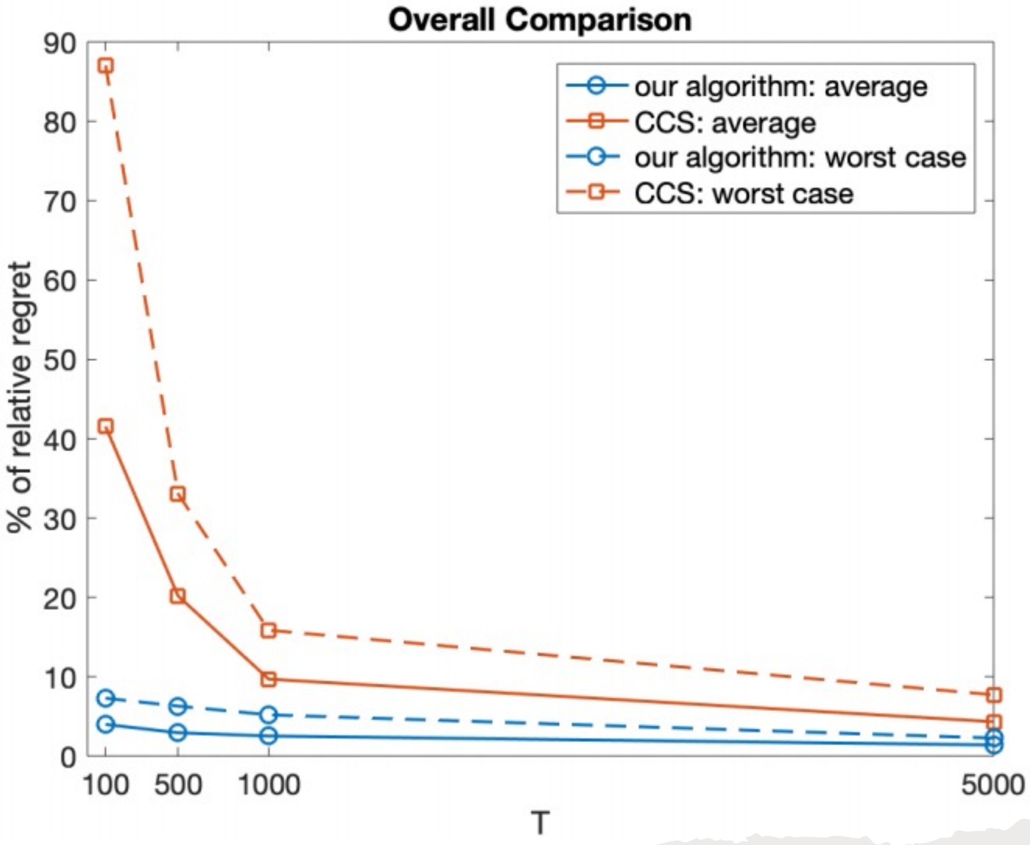
$$H_s^2(P,Q) \coloneqq 1 - \int_{-\infty}^{+\infty} p(x)^{1-\frac{1}{s}} q(x)^{\frac{1}{s}} \mathrm{d}x$$

  - ✓ Behaves "like" KL with $s \to +\infty$ in MAB type environments

$$H_s^2(P_0, P_j) \leq \{E_0[T_j]\}^{1-\frac{1}{s}} T^{\frac{1}{s}} \times \sup_p H_s^2(P_0(\cdot\,|p), P_j(\cdot\,|p))$$

NUMERICAL RESULTS

- Summary: $\tilde{O}(T^{0.6})$ regret, which is optimal
- Numerical results: comparison with an Explore-Then-Commit (ETC) baseline



FUTURE DIRECTIONS

- ***Open question 1.*** Fixed ordering cost + censored demand

  ✓ The parametric case is already difficult. Censored generalized linear models.

  ✓ How do we estimate the noise distribution is also a challenge. Unlikely the algorithm/analysis in the no-fixed-cost setting can be applied, because the optimal solution is not myopic and there is no easy characterization of the **p** function.

FUTURE DIRECTIONS

- ***Open question 2.*** Multiplicative demand noises.

$$d_t = \alpha_t D_0(p_t) + \beta_t, \qquad E[\alpha_t] = 1, E[\beta_t] = 0$$

✓ Parametric setting with fixed ordering costs: $(s, S, \boldsymbol{p})$ still optimal **asymptotically**, but difficult to reproduce $\psi(x; r, D, \boldsymbol{p})$ stability analysis.

✓ Nonparametric setting with censored costs: difficult to reproduce the **pairwise comparison** estimator. The observables are not shifts of the same distributions any more.

Thank you! Questions?

References:

Dynamic Pricing and Inventory Control with Fixed Ordering Cost and Incomplete Demand Information
Boxiao Chen, David Simchi-Levi, Yining Wang, Yuan Zhou
Management Science

Optimal Policies for Dynamic Pricing and Inventory Control with Nonparametric Censored Demands
Boxiao Chen, Yining Wang, Yuan Zhou
Management Science, to appear