Low-Rank Bandit Methods for High-Dimensional Dynamic Pricing
Jonas Mueller, Vasilis Syrgkanis, Matt Taddy
Amazon Microsoft Research Amazon jonasmueller@csail.mit.edu

Summary

- Dynamic pricing \( \approx \) online bandit optimization
- Bandit pricing can handle changing (adversarial) environments
- Number of pricing rounds until standard bandits find optimal prices (vanishing regret) depends on \# of products
- Observed product demands provide side information if they evolve in a low-rank fashion based on (latent) product features
- We develop bandit pricing algorithms that exploit this assumption, whose regret vanishes at a rate that only depends on \# of product features instead of \# of products

Objective

Choose prices for many products to maximize revenue/profit. Update prices over time to reflect changing demand curves.

Setup

- \( \mathbf{p}_t = \) vector of prices for \( N \) products sold during time-period \( t \)
- \( q_t \in \mathbb{R}^N = \) vector of demands for each product in time-period \( t \)
- \( R_t(\mathbf{p}_t) = \) total revenue over period \( t = (q_t, \mathbf{p}_t) \)
- Regret = \( \mathbb{E} \left[ \sum_{t=1}^{T} R_t(\mathbf{p}^* - R_t(\mathbf{p}_t)) \right] \)
- \( \mathbf{p}^* = \) optimal price vector (chosen in hindsight)

Standard Demand Model

- \( q_t = \mathbf{c}_t - \mathbf{B}_t \mathbf{p}_t + \mathbf{e}_t \)
  - \( \mathbf{B}_t \) describes how price of products affects demand for other products in round \( t \) (asymmetric, positive definite matrix)

- Can achieve \( O(T^{3/4}N^{1/2}) \) regret using standard online bandit method to select prices under this model
- Flaxman, Kalai, McMahan (2005): Online convex optimization in the bandit setting: Gradient descent without a gradient

Idea: Set price \( \mathbf{p} + \delta \mathbf{e} \) instead of \( \mathbf{p} \) with \( \mathbf{e} \) = random unit vector. For \( \hat{R}(\mathbf{p}) := R(\mathbf{p} + \delta \mathbf{e}) \): \( \nabla \hat{R}(\mathbf{p}) = \mathbb{E} [\nabla R(\mathbf{p} + \delta \mathbf{e})] \mathbf{e} / \delta \)

Product Features

- Let \( \mathbf{u}_i = d \)-dimensional features of product \( i \) \( (d \ll N) \)
- Product similarity = \( (\mathbf{u}_i, \mathbf{u}_j) \) for \( p_i = \mathbf{u}_i \mathbf{V} \mathbf{u}_j \cdot p_j \)

Low-Rank Demand Model

- \( q_t = \mathbf{U} \mathbf{c}_t - \mathbf{U} \mathbf{V}^T \mathbf{p}_t + \mathbf{e}_t \)
- \( \mathbf{U} = N \times d \) matrix, whose rows = product features

Known Product Features

- Algebra \( \implies R_t(\mathbf{p}) = f_t(\mathbf{x}) \) for concave \( f_t \) and \( \mathbf{x} := \mathbf{U}^T \mathbf{p} \in \mathbb{R}^d \)
- Strategy: use bandit algorithm to optimize \( \mathbf{x}_t \) w.r.t. \( f_t \), each time selecting prices \( \mathbf{p}_t \) via pseudo-inverse s.t. \( \mathbf{x}_t = \mathbf{U}^T \mathbf{p}_t \)
- Regret = \( O(T^{3/4}d^{1/2}) \) (does not depend on \( N \))

Unknown Product Features

- Assume orthonormal product features: \( \mathbf{U} \mathbf{x} = \mathbf{p} \) for \( \mathbf{x} = \mathbf{U}^T \mathbf{p} \)
- Previous strategy does not need known \( \mathbf{U} \), only need span(\( \mathbf{U} \))
- If \( \mathbf{e}_t = 0 \), then span(\( \mathbf{U} \)) = span(\( q_1, \ldots, q_d \))
- When \( \mathbf{e}_t \neq 0 \), we can estimate span(\( \mathbf{U} \)) via SVD of \( [q_1, \ldots, q_d] \)
- Run bandit algorithm to optimize \( \mathbf{x}_t = \hat{\mathbf{U}}^T \mathbf{p}_t \) w.r.t. \( f_t \)
  when span(\( \mathbf{U} \)) = current estimate of span(\( \mathbf{U} \))
- Regret = \( O(T^{3/4}d) \) (does not depend on \( N \))

Algorithm 1 Online Pricing Optimization with Latent Features

Input: \( \eta, \delta, \alpha > 0 \), rank \( d \in [1, N] \), initial prices \( \mathbf{p}_0 \in \mathcal{S} \)
Output: Prices \( \mathbf{p}_1, \ldots, \mathbf{p}_T \) to maximize overall revenue

- Initialize \( \hat{\mathbf{Q}} \) as \( N \times d \) matrix of zeros
- Initialize \( \hat{\mathbf{U}} \) as random \( N \times d \) orthogonal matrix
- Set prices to \( \mathbf{p}_0 \in \mathcal{S} \) and observe \( q_t(\mathbf{p}_0), R_t(\mathbf{p}_0) \)
- Define \( \mathbf{x}_0 = \hat{\mathbf{U}}^T \mathbf{p}_0 \)
- for \( t = 1, \ldots, T \):
  - \( \mathbf{x}_t := \hat{\mathbf{U}}^T \mathbf{x}_0 + \hat{\mathbf{U}}^T \mathbf{\xi}_t \), where \( \mathbf{\xi}_t \sim \text{Unif}(\{\mathbf{x} \in \mathbb{R}^d : ||\mathbf{x}||_2 = 1\}) \)
  - Set prices: \( \mathbf{p}_t = \hat{\mathbf{U}} \mathbf{x}_t \) and observe \( q_t(\mathbf{p}_t), R_t(\mathbf{p}_t) \)
  - \( \mathbf{\xi}_t \) = \( \text{projection}(\mathbf{x}_0 - \eta R_t(\mathbf{p}_t), \mathbf{\xi}_t) \)
  - \( \mathbf{Q}_{t+1} := \frac{1}{t} \theta_t + \mathbf{Q}_t \)
  - Set columns of \( \hat{\mathbf{U}} \) as top \( d \) left singular vectors of \( \mathbf{Q}_t \)

Evolving Demand Model