New Results on Primal-Dual Methods for Online Allocation Problems: Nonconvexity, Nonstationarity, and Robustness

Haihao (Sean) Lu

University of Chicago

Stern School of Business, NYU, Oct. 2021

\(^1\)joint work with Santiago Balseiro (Columbia U), and Vahab Mirrokni (Google)
Based on a recent papers:

“The best of many worlds: dual mirror descent for online allocation problems”, with Santiago Balseiro (Columbia) and Vahab Mirrokni (Google), to appear at *Operations Research*
Online allocation problems with resource constraints are central in computer science and operations research.

- **Goal**: maximize revenue subject to resource constraints
- **Challenge**: information about future requests is not known in advance

Need to design data-driven algorithms!
Model (Online Allocation Problems)

- Resources $j = 1, \ldots, m$ with capacities $B_j$
- Finite horizon with $T$ time periods
- At time $t$, receive request $(f_t, b_t, X_t)$:
  - Receive an (expected) **reward function** $f_t : \mathbb{R}^d \to \mathbb{R}_+$, an (expected) **consumption function** $b_t : \mathbb{R}^d \to \mathbb{R}_m^+$, **action space** $X_t \subseteq \mathbb{R}^d$
  - Take action $x_t \in X_t$
  - Collect reward $f_t(x_t)$ and consume resources $b_t(x_t)$

A request $(f_t, b_t, X_t)$ arrives with revenue function $f_t$, consumption function $b_t$, action space $X_t$.

Choose action $x_t$ from action space $X_t$.

Generates a revenue $f_t(x_t)$.

Consumes resources $b_t(x_t)$. 
Online Algorithms and Offline Problem

An online algorithm

- Makes decision based on observed history
- Knows capacities $B$ and the length of horizon $T$
- Need to satisfy the resource constraints

The offline problem

$$\text{OPT} := \max_{x: x_t \in X_t} \sum_{t=1}^{T} f_t(x_t)$$

s.t. \[ \sum_{t=1}^{T} b_t(x_t) \leq B. \]

- Offline optimal value provides an upper bound on any online algorithms
Applications

- Network revenue management with applications in hotel/airline pricing
- Online linear programming
- Online matching
- Proportional Matching with High Entropy
- Personalized Assortment Optimization with Limited Inventories
- Jointly allocation and pricing
- **Bidding in Repeated Auctions with Budgets**
- ...
Motivating Application: Real-Time Bidding

Motivating Application: Bidding in Repeated Auctions with Budget

Figure from [Amar and Renegar, 2020]
Motivating Application: Bidding in Repeated Auctions

DSP (Demand-Side Platform) problem:

An advertiser bids in repeated auction to maximize total rewards subject to budget constraint.

Key assumptions

- The expected value (pCTR) $r_t$ is known at the time of bidding.
- The bidding landscape is known at the time of bidding:
  - probability of winning $w_t(p)$ with bid price $p$
  - expected payment $q_t(p)$ with bid price $p$ if winning

In practice

- The value and bidding landscape can be learnt using historical data (i.e., contextual information).
- The bidding landscape is usually stationary, but the traffic is non-stationary.
Bidding in Repeated Auctions, continued

- Resource $B$ is the total budget for one advertiser for one day
- $T$ incoming impressions ($T$ can be estimated)
- At time $t$, receive impression:
  - Two decision variables: $y_t \in \{0, 1\}$ denotes whether to bid, $0 \leq p_t \leq \bar{p}$ is the bidding price
  - (Expected) reward function: $f_t(y_t, p_t) = (r_t - q_t(p_t))w_t(p_t)y_t$
  - (Expected) consumption function: $b_t(y_t, p_t) = q_t(p_t)w_t(p_t)y_t$

The offline problem is

$$\max_{y_t \in \{0, 1\}, 0 \leq p_t \leq \bar{p}} \sum_{t=1}^{T} \left( r_t - q_t(p_t) \right) w_t(p_t) y_t$$

subject to

$$\sum_{t=1}^{T} q_t(p_t) y_t w_t(p_t) \leq B.$$
Why This is a Hard Problem?

The offline problem is

$$\max_{y_t \in \{0,1\}, 0 \leq p_t \leq \bar{p}} \sum_{t=1}^{T} \left( r_t - q_t(p_t) \right) w_t(p_t) y_t$$

subject to

$$\sum_{t=1}^{T} q_t(p_t) w_t(p_t) y_t \leq B.$$ 

- The reward function and consumption function are highly non-convex and the constraint is integral
- The samples are non-stationary with seasonality
- The algorithm needs to be finished in 100ms
- The estimation $r_t, w_t$ and $q_t$ can be noisy, and sometimes corrupted
We design algorithms so that ...

- **Nonconvex**: handle nonconvex objective, nonconvex consumption, integral constraints
- **Nonstationarity**: single algorithm working for multiple input models
- **Robust**: noisy input \((f_t, b_t)\), and adversarial corruption
- **Efficient**: make the decision within 100ms (thus no dynamic programming or large optimization problem using all observed data)
- **Simple**: easy to be implemented
- **Theoretical-justified**: worst-case guarantees
Input Models

How are requests \((f_t, b_t, X_t)\) drawn?
Input Models

How are requests \((f_t, b_t, X_t)\) drawn?

- **Stochastic i.i.d.:** from an i.i.d. but unknown distribution
Input Models

How are requests \((f_t, b_t, X_t)\) drawn?

- **Stochastic i.i.d.**: from an i.i.d. but unknown distribution
- **Adversarial**: worst-case given by an adversary
Input Models

How are requests \((f_t, b_t, X_t)\) drawn?

- **Stochastic i.i.d.**: from an i.i.d. but unknown distribution
- **Adversarial**: worst-case given by an adversary

- **Stochastic IID** from an unknown distribution
- **Markov/periodic input**
- **Independent input with low mean deviation**
- **Adversarial**
Input Models

How are requests \((f_t, b_t, X_t)\) drawn?

- **Stochastic i.i.d.**: from an i.i.d. but unknown distribution
- **Adversarial**: worst-case given by an adversary

Non-stationary input models:

- **Independent with low mean deviation**: from independent but not identical distribution (corrupted samples)
- **Markov/ergodic**: from a Markov/ergodic process
- **Periodic**: from a periodic process
Related Literature

Stochastic i.i.d. model

- Many works on **linear reward and linear consumption**
  - [Devanur and Hayes, 2009; Feldman et al., 2010] show $O(T^{2/3})$ regret for two-phases algorithms
  - [Agrawal, Wang, Ye, 2014] shows $O(T^{1/2})$ regret for a resolving scheme
  - [Li and Ye, 2019] shows $O(\log(T))$ regret for a resolving scheme if the dual is strongly convex
  - [Wang and Wang, 2020] shows constant regret for i.i.d. model with known dist.

- [Agrawal and Devanur 2014] shows $O(T^{1/2})$ regret for **concave reward and convex consumption**, soft constraint

- [Dughmi et al., 2021] shows $O(T^{1/2})$ regret for proportional matching with a simple algorithm

...
Related Literature, continued

**Adversarial model**

- Network Revenue Management: bid price control ([Simpson 1989; Williamson 1992; Talluri and van Ryzin 1998; Adelman 2007 ... ])

- AdWords problems
  - [Mehta et al., 2007, Buchbinder et al., 2007] shows \((1 - 1/e)\)-competitive ratio
  - Reward and consumption are linear and proportional

- [Buchbinder et al., 2014] shows a competitive ratio analysis for concave reward, linear consumption, bounded ratio

- [Balseiro and Gur, 2019] shows a competitive ratio for 1d linear reward, linear consumption, independent reward and consumption

**Non-stationary model**

- [Ciocan and Farias, 2012]: Competitive ratio for correlated Gaussian process

- [Jiang, Li, Zhang, 2020]: Concurrent work with Wasserstein based non-stationarity
Algorithm
The offline problem

$$\text{OPT} := \max_{x : x_t \in X_t} \sum_{t=1}^{T} f_t(x_t)$$

s.t. $$\sum_{t=1}^{T} b_t(x_t) \leq B.$$
Introducing **Lagrange multipliers** $\mu \geq 0$ for resource constraints, we obtain

$$D(\mu) = \max_{x : x_t \in X_t} \sum_{t=1}^{T} f_t(x_t) + \mu^\top \left( B - \sum_{t=1}^{T} b_t(x_t) \right)$$

$$= \sum_{t=1}^{T} \max_{x_t \in X_t} \left\{ f_t(x_t) - \mu^\top b_t(x_t) \right\} + \mu^\top \frac{T}{B} \quad \text{weight-adjust reward}$$

$$= \sum_{t=1}^{T} f_t^*(\mu) + \mu^\top \frac{T}{B}$$

- **Key insight:** decisions decompose across time
- We call $f_t(x_t) - \mu^\top b_t(x_t)$ weight-adjusted reward
- We define $f_t^*(\mu) := \max_{x_t \in X_t} \left\{ f_t(x_t) - \mu^\top b_t(x_t) \right\}$ as the conjugate function of request $(f_t, b_t, X_t)$
Lagrangian Duality to the Rescue, continued

**Challenge #1:** How do we make decisions?

- If “optimal” dual variable $\mu^*$ is known, we can take action by maximizing the weight-adjusted reward

$$x_t = \arg\max_{x \in X_t} \{ f_t(x) - (\mu^*)^\top b_t(x) \}$$

- ... but we DO NOT know $\mu^*$ in advance

**Challenge #2:** How do we compute good dual variables?

- At time $t$, we face the dual objective

$$\max_{x_t \in X_t} \{ f_t(x_t) - \mu^\top b_t(x_t) \} + \mu^\top \frac{B}{T}$$

- **Key observation:** We can obtain a **subgradient** of the dual function

$$g_t = \frac{B}{T} - b_t(x_t^*)$$

- **Approach:** Update the dual variables using Online Mirror Descent
Online Dual Mirror Descent Algorithm

**Algorithm:** Dual Mirror Descent for Regularized Online Allocation

**Inputs:** initial dual solution $\mu_0 \in \mathbb{R}^m$ and initial budget $B_0 = B$

For each $t = 1, \ldots, T$ and while resources are available:

1. **Observe request:** $(f_t, b_t, X_t)$
2. **Determine an action:** $x_t = \arg\max_{x \in X : b_t x_t \leq B_t} \left\{ f_t(x) - \mu_t^\top b_t() x \right\}$
   - weight-adjust revenue
3. **Update resources:** $B_{t+1} = B_t - b_t x_t$
4. **Update dual variable** by online mirror descent:
   
   \[
   \arg\min_{\mu \geq 0} \left\{ \left( \frac{B}{T} - b_t x_t \right)^\top \mu + \frac{1}{\eta} \ g_t \mu + \frac{1}{\eta} \ V_h(\mu, \mu_t) \right\}
   \]

- We can use multiple sample to update the dual variables
- The argmax problem can be solved approximately
Examples of Mirror Descent

- **Online Gradient Descent**: if reference function is $h(\mu) = \frac{1}{2} \|\mu\|^2$, dual update is

\[
\mu_{t+1} = \max \left( \mu_t - \eta \left( \frac{B}{T} - b_t(x_t) \right) , 0 \right)
\]

- **Multiplicative Weights**: if reference function is negative entropy, the update is

\[
\mu_{t+1} = \mu_t \cdot \exp \left( -\eta \left( \frac{B}{T} - b_t(x_t) \right) \right)
\]
Intuitions on Non-Convexity

Q: How can dual algorithms work for nonconvex problems?

- Weak duality always holds: \( \text{OPT} \leq D(\mu) \) for any \( \mu \geq 0 \)
- Strong duality does not hold: \( \text{OPT} < \min_{\mu \geq 0} D(\mu) \)
Intuitions on Non-Convexity

**Q**: How can dual algorithms work for nonconvex problems?

- Weak duality always holds: $\text{OPT} \leq D(\mu)$ for any $\mu \geq 0$
- Strong duality does not hold: $\text{OPT} < \min_{\mu \geq 0} D(\mu)$

**A**: Shapley-Folkman Theorem! The optimality gap is independent of $T$ if $f, b$ are uniformly bounded

$$\min_{\mu \geq 0} D(\mu) - \text{OPT} \leq O(m)$$

Figure from [Kerdreux et al., 2019]
Theoretical Guarantees
Assumptions

Assumption (on the requests \((f, b, X)\))

There exists \(\bar{f} > 0\) and \(\bar{b} > 0\) such that for all requests \((f, b, X)\)

1. \(0 \in X\).
2. \(f(0) = 0\) and \(0 \leq f(x) \leq \bar{f}\) for all \(x \in X\).
3. \(b(0) = 0\) and \(b(x) \geq 0\) and \(\|b(x)\|_{\infty} \leq \bar{b}\) for all \(x \in X\).
4. The weight adjust reward maximization problems admit a finite optimal solution.
Non-stationary input models:

- **Independent with low mean deviation**: from independent but not identical distribution (corrupted samples)
- **Markov/ergodic**: from a Markov/ergodic process
- **Periodic**: from a periodic process
Performance Measurement and Benchmarks

- **Cumulative reward** of an online algorithm $A$ for a request sequence $\vec{\gamma} = \{(f_1, b_1, X_1), \ldots (f_T, b_T, X_T)\}$:

  \[
  R(A|\vec{\gamma}) = \sum_{t=1}^{T} f_t(x_t)
  \]

- **Benchmark**: offline optimal for a request sequence $\vec{\gamma}$:

  \[
  \text{OPT}(\vec{\gamma}) = \max_{x: x_t \in X_t} \sum_{t=1}^{T} f_t(x_t)
  \]

  \[
  \text{s.t. } \sum_{t=1}^{T} b_t(x_t) \leq B = T\rho
  \]

- **Regret** of an algorithm

  \[
  \text{Regret}(A|\mathcal{C}) = \sup_{\mathcal{P} \in \mathcal{C}} \{\mathbb{E}_{\vec{\gamma} \sim \mathcal{P}} [\text{OPT}(\vec{\gamma}) - R(A|\vec{\gamma})]\}
  \]

- **$\alpha$-asymptotic competitive ratio** of an algorithm

  \[
  \lim_{T \to \infty} \sup_{\vec{\gamma} \in \mathcal{S}} \left\{ \frac{1}{T} \left( \text{OPT}(\vec{\gamma}) - \alpha R(A|\vec{\gamma}) \right) \right\} \leq 0.
  \]
Stochastic IID Model

Let $\bar{P}$ be the unknown i.i.d. distribution of request

**Theorem: Regret with Stochastic i.i.d. Inputs**

Suppose the requests come from an i.i.d. model with unknown distribution, then

$$\operatorname{Regret}(A|C) = \sup_{\bar{P} \in C} \left\{ \mathbb{E}_{\gamma_t \sim \bar{P}} \left[ \operatorname{OPT}(\tilde{\gamma}) - R(A|\tilde{\gamma}) \right] \right\} \leq O(T^{1/2})$$

- Extend the existing results to nonconvex objective and nonconvex consumption
- Match with minimax regret bound: no online algorithm can achieve better than $O(T^{1/2})$ regret for this class of problems
Inexact Solutions (Noisy Inputs)

In practice, we may not be able to solve exactly the weight-adjust revenue maximization problem

\[ x_t = \arg\max_{x \in X_t} \left\{ f_t(x) - \mu^\top_t b_t(x) \right\} \]

- The problem can be non-convex thus solving exactly may be expensive
- The input \((f_t, b_t)\) usually come from ML model and can be noisy

Extension: Regret of Inexact Solutions

Suppose in algorithm \(\tilde{A}\), we solve the maximization problem inexactly with tolerance \(\epsilon_t\).

\[ f_t(x_t) - \mu^\top_t b_t(x_t) \geq \max_{x \in X_t} \left\{ f_t(x) - \mu^\top_t b_t(x) \right\} - \epsilon_t \]

Then

\[ \text{Regret}(\tilde{A}|C) \leq \text{Regret}(A|C) + \sum_{t=1}^{T} \epsilon_t \]

- The algorithm is robust to noise
- Similar story happens in other input models
Theorem: Asymptotic Competitive Ratio with Adversarial Inputs

Suppose the requests are chosen by an adversary. Then, online mirror descent is

$$\alpha^* = \max \left\{ \sup_{(f,b,X) \in S} \sup_{x \in X, j \in [m]} b_j(x)/\rho_j, 1 \right\}$$

asymptotically competitive, i.e.,

$$\lim_{T \to \infty} \sup_{\tilde{\gamma} \in S} \left\{ \frac{1}{T} \left( \text{OPT}(\tilde{\gamma}) - \alpha^* R(A|\tilde{\gamma}) \right) \right\} \leq 0.$$ 

- The competitive ratio involves consumption budget ratio
- Match with minimax competitive ratio: no online algorithm can achieve better than $\alpha^*$-asymptotically competitive.
- **Disclaimer**: One can potentially get better competitive ratio if there are known correlation between $f$ and $b$
- The stopping time in the worst case is $T - O(\sqrt{T})$
Independent Inputs with Low Mean Deviation

- Let $\mathcal{P}_t$ be the independent distribution of the sample received at time $t$
- Let $\bar{\mathcal{P}} = \frac{1}{T}(\mathcal{P}_1 + \cdots + \mathcal{P}_T)$ be the average stationary distribution
- Define mean deviation of the distribution sequence as
  $$MD(\bar{\mathcal{P}}) = \sum_{t=1}^{T} \| \mathcal{P}_t - \bar{\mathcal{P}} \|_{TV}$$

Theorem: Independent Inputs with Low Mean Deviation

Suppose the requests are chosen from an independent sequence of distributions with the total mean deviation $MD(\mathcal{P})$, then

$$\text{Regret}(A|C) = \sup_{\mathcal{P} \in C} \left\{ \mathbb{E}_{\tilde{\gamma} \sim \bar{\mathcal{P}}} [\text{OPT}(\tilde{\gamma})] - \mathbb{E}_{\gamma_t \sim \mathcal{P}_t} [R(A|\tilde{\gamma})] \right\} \leq O\left( T^{1/2} \right) + \bar{f} MD(\mathcal{P}).$$

- [Jiang, Li, Zhang] considers a similar setup where $MD(\bar{\mathcal{P}})$ is defined by Wasserstein norm instead of total variation norm.
Suppose the requests are chosen from an independent sequence of distributions with the total mean deviation $MD(P)$. Then it holds for any online algorithm $A$ that

$$\text{Regret}(A|C) \geq O\left( T^{1/2} + MD(P) \right).$$

The regret is tight.
Adversarial Corruption

- Suppose an adversary can corrupt at most $\delta$ instances (equivalently there are $\delta$ outliers)

**Corollary: Adversarial Corruption**

Suppose the requests come from a stochastic i.i.d. model with at most $\delta$ corrupted instances, then

$$\text{Regret}(A|C) = \sup_{P \in C} \left\{ \mathbb{E}_{\tilde{\gamma} \sim \tilde{P}} \left[ \text{OPT}(\tilde{\gamma}) \right] - \mathbb{E}_{\tilde{\gamma} \sim P_t} \left[ R(A|\tilde{\gamma}) \right] \right\} \leq O(T^{1/2}) + O(\delta).$$

- This bound is optimal, i.e., no algorithm can achieve better than $O(T^{1/2}) + O(\delta)$ regret
- If less than $T^{1/2}$ samples are corrupted, the regret is $O(T^{1/2})$
- Recover the result of stochastic i.i.d. input with $\delta = 0$ and adversarial input with $\delta = T$
Markov/Ergodic Inputs

The distribution $P_{t+1}$ may depend on the realization in time $1, \ldots, t$, for example,

- In dynamic asset allocation, the asset price is correlated with its previous price
- In online pricing, the demand of one item is correlated with the demand in near past

**Theorem: Markov/Ergodic Inputs**

Suppose the requests come from an irreducible and aperiodic Markov process (or an ergodic process), then

$$\text{Regret}(A|\mathcal{C}) = \sup_{\mathcal{P} \in \mathcal{C}} \left\{ \mathbb{E}_{\gamma \sim \mathcal{P}} [\text{OPT}(\gamma) - R(A|\gamma)] \right\} \leq O\left( T^{1/2} \log T \right).$$

- This bound is almost tight (up to a log term)
- The Markov process is ergodic, and the mixing time plays a role in the regret bound
Consider periodic inputs $P_{1:q} = P_{q+1:2q} = \cdots = P_{T-q+1:T}$.

**Theorem: Periodic Inputs**

Suppose the requests come from a periodic process with period $q$, then

$$\text{Regret}(A|C) = \sup_{P \in C} \mathbb{E}_{\gamma \sim P} \left[ \text{OPT}(\gamma) - R(A|\gamma) \right] \leq O((qT)^{1/2}).$$

- $q = 1$ recovers stochastic i.i.d. model and $O(T^{1/2})$ regret
- $q = T$ recovers adversarial model and $O(T)$ regret
- The first online allocation algorithm on periodic inputs with theoretical guarantees
Extensions

- The algorithm can be extended beyond online mirror descent.
- If \( \bar{f} \) is known to the user, one can constrain the dual variable in the standard simplex, and obtain \( \log m \) dependence in the regret bound, where \( m \) is the number of resource.
- In practice, the total horizon \( T \) is a random variable. We can instead set \( \rho = B/ET \), and there is an extra term \( \bar{f}\text{SD}(T) \) in the bound.
Additional Slides
Summary: Primal-dual methods for online allocation problems

- **Nonconvex**: handle nonconvex objective, nonconvex consumption, integral constraints
- **Nonstationarity**: single algorithm working for multiple input models
- **Robust**: noisy input \((f_t, b_t)\), and adversarial corruption
- **Efficient**: make the decision within 100ms (thus no dynamic programming or large optimization problem using all observed data)
- **Simple**: easy to be implemented
- **Theoretical-justified**: worst-case guarantees

Not covered in the talk but in the paper:
- Presents a *simplified and unified* analysis
- Opens up *new areas of applications* with nonconvex objective

Thank you!
Why Non-Stationary Input Models?

- There is always noise in real data
- Real data has seasonality
Online Convex Optimization (OCO)

In online convex optimization:

- the player chooses an action $x_t$
- the adversary chooses a convex loss function $f_t$

The goal is to ensure that the total loss is not much larger than the smallest total loss

$$\text{Regret} = \sum_{t=1}^{T} f_t(x_t) - \min_x \sum_{t=1}^{T} f_t(x)$$

Online first-order methods: use sub-gradient information to determine actions

- Examples: online gradient descent, online mirror descent, regularized follow-the-leaders, ADAM, ADAGRAD, ...
- $O(\sqrt{T})$ regret
Independent Inputs with Low Mean Deviation

- Let $P_t$ be the independent distribution of the sample received at time $t$
- Let $\bar{P}$ be any stationary distribution
  - $\bar{P} = \frac{1}{T}(P_1 + \cdots + P_T)$
  - $\bar{P}$ is the underlining stochastic i.i.d. model ignoring outliers
- Define the deviation at time $t$ from the stationary distribution as
  $$\Delta_t = \sup_{\mu \geq 0} \left\{ \mathbb{E}_{(f,b,X) \sim \bar{P}} [f^*(\mu)] - \mathbb{E}_{(f,b,X) \sim P_t} [f^*(\mu)] \right\}$$
- Recall the conjugate function $f^*(\mu) := \max_{x \in X} \{ f_t(x) - \mu^\top b_t(x) \}$

Theorem: Independent Inputs with Low Mean Deviation

Suppose the requests are chosen from an independent sequence of distributions with the total mean deviation $MD(P)$, then

$$\text{Regret}(A|C) = \sup_{P \in C} \left\{ \mathbb{E}_{\tilde{\gamma} \sim \bar{P}} [\text{OPT}(\tilde{\gamma})] - \mathbb{E}_{\gamma_t \sim P_t} [R(A|\gamma)] \right\} \leq O(T^{1/2}) + \sum_{t=1}^{T} \Delta_t.$$