ONLINE SEMI-INFINITE LINEAR PROGRAMMING VIA NONNEGATIVE FUNCTION APPROXIMATION

Anonymous authors
Paper under double-blind review

ABSTRACT

We study online semi-infinite linear programming, a sequential resource-allocation problem in which the decision space is finite dimensional but the constraints are indexed by a large, possibly infinite, set. This setting captures online allocation problems with continuously indexed safety, robustness, coverage, or moment constraints. Classical online linear-programming algorithms maintain one dual variable per constraint, and their regret guarantees typically scale with the number of constraints; this dependence becomes prohibitive in semi-infinite regimes. We propose a nonnegative function-approximation approach that represents the dual measure using q nonnegative basis functions. This yields a tractable projected problem whose regret bounds depend on the basis dimension q, rather than on the number of original constraints. We develop first-order primal–dual algorithms for stochastic and random-permutation input models. Under stochastic arrivals, our algorithms achieve $O(q\sqrt{T})$ regret, together with sublinear projected constraint violation. Under random permutation arrivals, we obtain a $\widetilde{O}(q\sqrt{T})$-type regret bound. We further give an accelerate-then-refine algorithm that improves the stochastic regret to $O(q \log T)$ under additional conditions. Experiments on large-scale instances show that the proposed basis-based methods remain stable as the number of constraints grows, while standard online LP methods degrade substantially.

1 INTRODUCTION

Many online decision problems involve a finite sequence of accept/reject or allocation decisions, while feasibility is governed by an extremely large family of constraints. Examples include robust online resource allocation, distributionally robust optimization with generalized moment constraints, spatial coverage over continuous regions, and safety-constrained control with constraints indexed by a continuous state space. These problems are naturally modeled as online semi-infinite linear programs: at each time t, the decision maker observes a reward r_t and a constraint coefficient function $a_t(\cdot)$, chooses $x_t \in [0, 1]$, and must satisfy aggregate constraints over all indices s in a constraint space $\mathcal{S}$.

The main challenge is dual dimensionality. In a standard online linear program with m constraints, a primal–dual algorithm maintains a vector of m dual prices. In a semi-infinite linear program, the corresponding dual object is a nonnegative measure or function over $\mathcal{S}$. Direct multiplicative-weights or gradient-based updates over such an infinite-dimensional cone are computationally infeasible. Even after discretization, existing regret guarantees typically scale with the number of discretized constraints, which can be large enough to dominate the regret bound.

We address this bottleneck through nonnegative function approximation. Instead of maintaining a separate dual variable for every constraint index, we approximate the dual function using q nonnegative basis functions:

$$u(s) \approx \Phi(s)^\top w, \qquad w \in \mathbb{R}^q_+.$$

The nonnegativity of the basis is essential: projecting w onto $\mathbb{R}^q_+$ automatically preserves dual nonnegativity, and the induced projected primal relaxation remains a valid upper-bounding device for the original offline optimum. As a result, the online algorithms operate in q dimensions and their regret guarantees depend on q, not on the cardinality of $\mathcal{S}$.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Table 1: Comparison with representative online LP results. Here m denotes the number of original/discretized constraints and q denotes the number of basis functions in our reduced dual representation. Prior online LP results control violation of the original finite constraint system. Our algorithms directly control violation of the projected q-dimensional constraints; exact pointwise violation can be recovered under the additional basis-approximation condition in Proposition 1.

Work	Input model	Regret dependence	Constraint violation
(Li & Ye, 2022)	stochastic	$O(m^2 \log T \log\log T)$	exact violation bound
(Li et al., 2023)	stochastic	$O(m\sqrt{T})$	exact violation bound
(Li et al., 2023)	random permutation	$O((m+\log T)\sqrt{T})$	exact violation bound
(Gao et al., 2025)	stochastic	$O(m \log T)$	exact violation bound
This work	stochastic	$O(q\sqrt{T})$, $O(q \log T)$	weaker projected bound
This work	random permutation	$\widetilde{O}(q\sqrt{T})$	weaker projected bound

Contributions. We make the following contributions. First, we introduce a regularized reduced-dual formulation for online semi-infinite LPs. The nonnegative dual function is represented by q nonnegative basis functions, yielding a projected q-dimensional problem. We prove a deterministic bridge showing that this projected relaxation upper bounds the original offline semi-infinite optimum, so it can be used as a tractable surrogate in the regret analysis.

Second, we emphasize the role of the dual regularizer. Unlike standard online LP analyses based on unregularized finite-dimensional dual prices, our analysis is carried out for a regularized dual objective. The regularizer stabilizes the online dual iterates and makes the exact infinite-dimensional dual problem strongly convex. Consequently, under continuity assumptions on the data, the regularized exact dual optimizer is unique and continuous, which supports uniform approximation by smooth nonnegative basis families under additional approximation conditions.

Third, we design projected-gradient and mirror-descent algorithms for the regularized projected dual. Under stochastic arrivals, the algorithms achieve $O(q\sqrt{T})$ regret. Under random permutation arrivals, we obtain a $\widetilde{O}(q\sqrt{T})$-type guarantee. Since the algorithms directly control only the projected constraints, we also provide a transfer result: under a uniform approximation condition on the basis class, exact violation is bounded by projected violation plus an explicit basis-approximation error.

Fourth, under additional assumptions for the regularized projected dual, we propose an accelerate-then-refine algorithm with improved stochastic regret $O(q \log T)$. Finally, experiments on large instances illustrate that the regularized basis-based methods remain stable as the number of constraints grows.

In summary, our work provides a regularized reduced-dual framework for online semi-infinite LPs. The algorithms operate on a q-dimensional nonnegative basis representation of the dual function, and their regret and projected-feasibility bounds depend on q rather than directly on the number of original or discretized constraints. Exact pointwise feasibility is not automatic from the projected updates; it follows under additional approximation conditions, with an explicit transfer bound depending on the basis approximation error.

Relation to prior work. Online linear programming and resource allocation have been extensively studied in ad allocation, revenue management, packing, and covering, with many algorithms following a dual-price paradigm such as sample-then-price, dynamic price learning, and primal–dual methods. These methods typically maintain one price per resource or constraint. Semi-infinite linear programming instead considers finite-dimensional decisions under infinitely many constraints, often addressed by discretization, cutting-plane, or exchange methods. Our framework connects these lines while controlling the dimension through the approximation architecture rather than the original constraint cardinality. We leave the full literature review to Appendix A.

2 Problem Formulation

Let $\mathcal{S}$ be a compact constraint-index set equipped with a reference measure σ. At each time $t = 1, \ldots, T$, the decision maker observes a reward $r_t \in \mathbb{R}_+$ and a continuous coefficient function

$a_t : \mathcal{S} \to \mathbb{R}_+$, and then chooses $x_t \in [0, 1]$. The offline semi-infinite linear program is

$$\begin{aligned} \mathrm{OPT}(I) = \max_{x \in [0,1]^T} \quad & \sum_{t=1}^{T} r_t x_t \\ \text{s.t.} \quad & \sum_{t=1}^{T} a_t(s) x_t \le T d(s), \qquad \forall s \in \mathcal{S}, \end{aligned} \tag{1}$$

where $d : \mathcal{S} \to \mathbb{R}_+$ is the normalized resource budget and $I = \{(r_t, a_t)\}_{t=1}^T$ denotes the realized instance. We study two standard input models.

Definition 1 (Input models)**.** The sequence $\{(r_t, a_t)\}_{t=1}^T$ is generated according to one of the following models.

1. **Stochastic input model:** (r_t, a_t) are drawn i.i.d. from an unknown distribution $\mathcal{P}$.
2. **Random permutation model:** An adversary fixes a multiset of T input pairs in advance, and the sequence is revealed in a uniformly random order.

For a policy π, let x_t^π be its decision at time t. We measure regret relative to the offline optimum:

$$\mathrm{Reg}_T(\pi) = \sup_{\mathcal{P} \in \Xi} \mathbb{E}\left[\mathrm{OPT}(I) - \sum_{t=1}^{T} r_t x_t^\pi \right],$$

where the expectation is over the input sequence and the algorithmic randomness. Furthermore, we consider a supplementary performance measure: *constraints violation*. For a policy π, define the residual function $g_\pi(s) := \sum_{t=1}^T a_t(s) x_t^\pi - T d(s), s \in \mathcal{S}$.. For notational simplicity, we write $A\boldsymbol{x}(\pi) - \boldsymbol{b}$ for the residual function g_π, i.e.,

$$(A\boldsymbol{x}(\pi) - \boldsymbol{b})(s) := \sum_{t=1}^{T} a_t(s) x_t^\pi - T d(s), \quad s \in \mathcal{S}.$$

Then the exact violation is

$$v(\pi) = \mathbb{E}\left[\left(\sup_{u \in \mathcal{U}_0} \{ \langle \boldsymbol{u}, A\boldsymbol{x}(\pi) - \boldsymbol{b} \rangle \} \right)^+ \right]. \tag{2}$$

Here, the operator $(\cdot)^+$ refers to take the maximum of the inner term and 0. The set $\mathcal{U}_0$ refers to the collection of all possible dual variables $\boldsymbol{u}$ such that it integrates to 1 over the index space $\mathcal{S}$, i.e. $\mathcal{U}_0 = \left\{ \boldsymbol{u} \ge 0 : \int_{s \in \mathcal{S}} \boldsymbol{u}(s) d\sigma(s) = 1 \right\}$. It is clear to see that the definition in equation 2 is equivalent to taking expectation over the maximum constraint violation, i.e., the largest positive element of $A\boldsymbol{x}(\pi) - \boldsymbol{b}$.

3 Nonnegative Dual Approximation

The dual of Equation (1) involves a nonnegative measure or density over $\mathcal{S}$. To obtain a finite-dimensional online algorithm, we choose q nonnegative basis functions $\phi_1, \ldots, \phi_q$, collect them as $\Phi(s) = (\phi_1(s), \ldots, \phi_q(s))^\top$, and approximate the dual function by

$$u(s) = \Phi(s)^\top w, \qquad w \in \mathbb{R}_+^q.$$

Define the projected coefficients

$$c_t := \int_{\mathcal{S}} \Phi(s) a_t(s) d\sigma(s) \in \mathbb{R}_+^q, \qquad d_\Phi := \int_{\mathcal{S}} \Phi(s) d(s) d\sigma(s) \in \mathbb{R}_+^q.$$

The induced projected primal problem is

$$\begin{aligned} \mathrm{OPT}_\Phi(I) = \max_{\boldsymbol{x} \in [0,1]^T} \quad & \sum_{t=1}^{T} r_t x_t \\ \text{s.t.} \quad & \sum_{t=1}^{T} c_t x_t \le T d_\Phi. \end{aligned} \tag{3}$$

The corresponding regularized empirical dual objective is

$$f^{\rho}_{T,\Phi}(w) = d_{\Phi}^{\top} w + \frac{1}{T}\sum_{t=1}^{T}\big(r_t - c_t^{\top} w\big)^{+} + \rho\|w\|_2^2, \qquad w \in \mathbb{R}^q_{+}. \tag{4}$$

In the stochastic input model, the population objective is

$$f^{\rho}_{\Phi}(w) = d_{\Phi}^{\top} w + \mathbb{E}_{(r,a)\sim\mathcal{P}}\left[\big(r - c^{\top} w\big)^{+}\right] + \rho\|w\|_2^2. \tag{5}$$

Lemma 1 (Deterministic bridge)**.** *For every realized instance I,*

$$\mathrm{OPT}(I) \le \mathrm{OPT}_{\Phi}(I) \le T \min_{w\in\mathbb{R}^q_{+}} f^{\rho}_{T,\Phi}(w).$$

Lemma 1 shows that the regularized projected dual remains an upper bound on the original offline optimum. Therefore, the regret analysis can be carried out directly through the empirical regularized objective $f^{\rho}_{T,\Phi}$, without requiring explicit knowledge of the underlying distribution $\mathcal{P}$.

Now we introduce our main assumption, which is common in relevant literature (Li et al., 2023; He et al., 2025; Gao et al., 2025).

Assumption 1 (main assumption)**.** The following conditions hold for all $t = 1, ..., T$.

1. (**Bounded Data**) There exist known constants $\overline{r} > 0$ such that $r_t \le \overline{r}$.
2. (**Finite Dual Measure**) With respect to the basis function Φ, there exist positive constants $0 < \underline{D} \le \overline{D}$ and $\overline{C} > 0$ such that,
$$\underline{D} \le (d_{\Phi})_j \le \overline{D}, \;\; (c_t)_j \le \overline{C}, \quad j = 1, \ldots, q.$$

3.1 Projected Constraint Violation

We introduce a weaker notion of the constraint violation. Instead of checking pointwise violation, we measure the magnitude of violation "tested" against an approximated set of dual variables. To be specific, let $\boldsymbol{m}_{\Phi} := \int_{\mathcal{S}} \Phi(s)\, d\sigma(s)$, $\underline{m}_{\Phi} := \min_{k\in[q]}(\boldsymbol{m}_{\Phi})_k > 0$ and

$$\mathcal{U}_{\Phi} := \left\{u(\cdot) = \Phi(\cdot)^{\top}\boldsymbol{w} : \boldsymbol{w} \ge 0,\ \boldsymbol{m}_{\Phi}^{\top}\boldsymbol{w} = 1\right\}. \tag{6}$$

denote the approximated set of dual variables and we then define the *projected constraints violation* as follows.

$$v(\pi, \mathcal{U}_{\Phi}) = \mathbb{E}\left[\left(\max_{u\in\mathcal{U}_{\Phi}} \langle u, A\boldsymbol{x}(\pi) - \boldsymbol{b}\rangle\right)^{+}\right]. \tag{7}$$

It is clear to see that $v(\pi) = v(\pi, \mathcal{U}_0)$.

3.2 From Projected to Exact Constraint Violation

The projected violation $v(\pi, \mathcal{U}_{\Phi})$ is the feasibility measure that is directly controlled by our q-dimensional dual update. Connecting this projected guarantee to the exact pointwise violation requires additional approximation structure. The dual regularizer plays an important role in this connection. In the absence of regularization, the exact dual problem may have nonunique, nonsmooth, or even measure-valued optimizers, and such objects are difficult to approximate uniformly using a finite smooth basis. By contrast, when $\rho > 0$, the exact regularized dual objective is strongly convex and hence admits a unique optimizer. This added regularity provides the motivation for approximating the regularized dual function by smooth nonnegative basis expansions.

We begin by introducing the regularized version of the empirical original dual problem:

$$\min_{u\in L^2_{+}(\mathcal{S})} \quad F^{\rho}_{T}(u) := \langle d, u\rangle + \frac{1}{T}\sum_{t=1}^{T}\big(r_t - \langle a_t, u\rangle\big)^{+} + \rho\|u\|^2_{L^2(\mathcal{S})}, \tag{8}$$

where $L^2_{+}(\mathcal{S}) := \{u \in L^2(\mathcal{S}) : u(s) \ge 0 \text{ a.e.}\}$, and

$$\langle a_t, u\rangle := \int_{\mathcal{S}} a_t(s)u(s)\, d\sigma(s), \qquad \langle d, u\rangle := \int_{\mathcal{S}} d(s)u(s)\, d\sigma(s).$$

Let $\mathcal{A} \subseteq C(\mathcal{S}; \mathbb{R}_+)$ be a fixed class of admissible coefficient functions, with $a_t \in \mathcal{A}$ almost surely for every t. We use two approximation conditions with different roles. The first concerns approximation of the optimizer of the regularized exact dual problem. The second is stronger and is used only to transfer projected violation to exact pointwise violation.

Assumption 2 (Basis approximation)**.** The following approximation conditions are imposed.
(a) Regularized dual approximation. Let $u^*_{T,\rho}$ be the optimizer of equation 8. There exists $u^*_\Phi \in \mathcal{U}_\Phi$ such that

$$\|u^*_{T,\rho} - u^*_\Phi\|_\infty \le \epsilon_{\text{dual}}.$$

(b) Test-function approximation. For every normalized nonnegative continuous test function $u \in \mathcal{U}_0$, there exists $u_\Phi \in \mathcal{U}_\Phi$ such that

$$\sup_{a \in \mathcal{A}} |\langle u - u_\Phi, a\rangle| \le \epsilon_{\text{test}}, \quad |\langle u - u_\Phi, d\rangle| \le \epsilon_{\text{test}}.$$

The next lemma shows that the optimizer of the regularized original dual is continuous.

Lemma 2 (Continuity of the regularized dual optimizer)**.** *Suppose $\rho > 0$ and $a_t(s)$, $d(s)$ are continuous over $s \in \mathcal{S}$, for each t. Then problem equation 8 admits a unique optimizer $u^*_{T,\rho}$. Moreover, there exist coefficients $\alpha_t \in [0,1]$, for each t, such that*

$$u^*_{T,\rho}(s) = \frac{1}{2\rho}\left[\frac{1}{T}\sum_{t=1}^{T} \alpha_t a_t(s) - d(s)\right]^+, \qquad s \in \mathcal{S}. \tag{9}$$

In particular, $u^*_{T,\rho}$ is continuous over $\mathcal{S}$.

Remark 1. Lemma 2 shows that the regularized exact dual optimizer belongs to the class of continuous functions. Hence, if Φ is chosen from a universal nonnegative family, for example suitable radial basis functions on a compact set, then $u^*_{T,\rho}$ can be approximated arbitrarily well in the uniform norm. This justifies Assumption 2(a), which is sufficient for approximating the regularized exact dual value.

However, projected-to-exact violation transfer requires the stronger condition in Assumption 2(b). The reason is that $v(\pi)$ is determined by the residual induced by policy π and can be represented as a supremum over normalized nonnegative test functions. The test function that witnesses violation generally depends on this residual and need not coincide with the regularized dual optimizer $u^*_{T,\rho}$. Therefore, approximation of the single optimizer $u^*_{T,\rho}$ does not by itself imply exact feasibility transfer.

Proposition 1 (Violation transfer)**.** *Under assumption 2(b), any policy π satisfies*

$$v(\pi) \le v(\pi, \mathcal{U}_\Phi) + 2T\epsilon_{\text{test}}.$$

Consequently, if $\epsilon_{\text{test}} = O(T^{-1/2})$, the exact violation has the same order as the projected violation up to constants.

4 A Projected First-Order Algorithm

We now present an online stochastic first-order method whose update is an unbiased sub-gradient step for the population regularized dual objective equation 5 under stochastic arrival. The algorithm maintains a dual estimate w_t, uses c_t as the reduced feature vector, and applies the threshold rule suggested by complementary slackness

$$x_t(w) = \mathbf{1}\{r_t > c_t^\top w\}.$$

The stochastic subgradient of Equation (5) at w_t is

$$\widehat{\zeta}^\rho_t = d_\Phi - c_t x_t(w_t) + 2\rho w_t. \tag{10}$$

The algorithm can also be interpreted as an online primal–dual method for the projected LP equation 3. The primal variable x_t greedily maximizes the instantaneous reduced Lagrangian under the current dual estimate w_t, while the dual variable w_t accumulates evidence of projected constraint violation and is simultaneously shrunk by the regularizer.

We next record that the regularized update preserves the same uniform upper bound on the dual iterates as in the unregularized case.

Algorithm 1 Projected dual gradient algorithm

1: **Input:** budget d, basis Φ, stepsizes $\{\gamma_t\}$, regularization ρ.
2: Initialize $w_1 = 0 \in \mathbb{R}^q_+$.
3: **for** $t = 1, \dots, T$ **do**
4: Observe (r_t, a_t) and compute $c_t = \int_{\mathcal{S}} \Phi(s) a_t(s) d\sigma(s)$.
5: Set $x_t = \mathbf{1}\{r_t > c_t^\top w_t\}$.
6: Update

$$w_{t+1} = \Pi_{\mathbb{R}^q_+} \left(w_t - \gamma_t \big(d_\Phi - c_t x_t + 2\rho w_t \big) \right).$$

7: **end for**
8: **Output:** $x_1, \dots, x_T$.

Lemma 3. *Suppose the stochastic input model and assumption 1 hold. If $\gamma_t \le 1$ for all $t = 1, \dots, T$, then the iterates generated by Algorithm 1 satisfy*

$$\|w_t\|_2 \le W_\Phi := \frac{q(\overline{C} + \overline{D})^2 + 2\overline{r}}{2\underline{D}} + \sqrt{q}(\overline{C} + \overline{D}), \qquad t = 1, \dots, T+1.$$

We are now ready to state the main performance guarantee for Algorithm 1. The feasibility guarantee is stated for the reduced dual class $\mathcal{U}_\Phi$ in equation 6. This is the natural notion controlled by the algorithm, since Algorithm 1 only updates the q-dimensional coefficients of the projected dual function. The key point is that the regret and projected-violation bounds scale with q, not with the number of original constraints.

Theorem 1 (Stochastic regret and projected feasibility)**.** *Suppose the stochastic input model and assumption 1 hold. Let Algorithm 1 use $\gamma_t = T^{-1/2}$ and $0 \le \rho \le T^{-1/2}$. Then*

$$\mathrm{Reg}_T(\pi_{\mathrm{GD}}) \le O(q\sqrt{T}), \qquad v(\pi_{\mathrm{GD}}, \mathcal{U}_\Phi) \le O(q^2\sqrt{T}),$$

where the constants depend on the boundedness parameters and basis normalization, but not on the number of original constraints.

5 Mirror Descent and Random Permutation Arrivals

The projected gradient method in the previous section uses the Euclidean geometry of the reduced dual space. While this already yields dimension-reduced regret guarantees, Euclidean updates are not always the most suitable choice when the dual vector has additional structure. Let $\psi : \mathbb{R}^q_+ \to \mathbb{R}$ be a differentiable mirror map and let

$$D_\psi(u \| v) = \psi(u) - \psi(v) - \nabla\psi(v)^\top (u - v)$$

be the associated Bregman divergence. Mirror descent replaces Euclidean projection by a projection in the geometry induced by D_ψ. When $\psi(u) = \frac{1}{2}\|u\|_2^2$, the update reduces to projected gradient descent. More general mirror maps allow the algorithm to adapt to the geometry of the dual approximation class $\{\Phi^\top w : w \in \mathbb{R}^q_+\}$. For example, the negative entropy potential leads to multiplicative-type updates and is often more appropriate when the dual weights are sparse or lie near the simplex.

We impose the following standard assumption on the mirror map and the generated iterates.

Assumption 3 (Mirror map)**.** The mirror map ψ is coordinate-wise separable and α-strongly convex with respect to a norm $\|\cdot\|$. The generated iterates satisfy $\|w_t\|_\infty \le \overline{W}$.

The regularized mirror-descent algorithm is presented in Algorithm 2. The only difference from the unregularized version is that the update uses the regularized stochastic subgradient equation 10. Thus, the algorithm preserves the same reduced-dimensional primal–dual structure as Algorithm 1, while allowing a more flexible update geometry.

The next theorem gives the regret and projected constraint-violation bounds under stochastic arrivals. It shows that the mirror-descent algorithm retains the main benefit of the nonnegative dual approximation: the performance depends on the number of basis functions q, rather than on the number of original constraints.

Algorithm 2 Regularized mirror-descent algorithm

1: **Input:** budget d, basis Φ, mirror map ψ, step sizes $\{\gamma_t\}$, regularization ρ.
2: Initialize $w_1 = 0 \in \mathbb{R}^q_+$.
3: **for** $t = 1, \dots, T$ **do**
4: Observe (r_t, a_t), compute c_t, and set $x_t = \mathbf{1}\{r_t > c_t^\top w_t\}$.
5: Form $\hat{\zeta}_t^\rho = d_\Phi - c_t x_t + 2\rho w_t$.
6: Update

$$w_{t+1} = \operatorname*{arg\,min}_{w \in \mathbb{R}^q_+} \left\{ \gamma_t (\hat{\zeta}_t^\rho)^\top w + D_\psi(w \| w_t) \right\}.$$

7: **end for**
8: **Output:** $x_1, \dots, x_T$.

Theorem 2 (Mirror descent under stochastic arrivals)**.** *Under the stochastic input model, Assumptions 1 and 3, Algorithm 2 with $\gamma_t = T^{-1/2}$ and $0 \le \rho \le T^{-1/2}$ satisfies*

$$\mathrm{Reg}_T(\pi_{\mathrm{MD}}) \le O(q\sqrt{T}), \qquad v(\pi_{\mathrm{MD}}, \mathcal{U}_\Phi) \le O(\sqrt{qT}).$$

Now we discuss a more general setting: Random Permutation Model. In this model, coefficient pair (r_t, a_t) arrives in a random order and their value can be chosen adversarially at the start. We present our formal result in Theorem 3. Since Random Permutation Model is a more challenging than under the Stochastic Input Model, our regret bound is weaker than prior results; however, these two settings exhibit comparable performance in terms of constraint violations.

Theorem 3 (Mirror descent under random permutation arrivals)**.** *Under the random permutation model, Assumptions 1 and 3, Algorithm 2 with $\gamma_t = T^{-1/2}$ and $\rho = 0$ satisfies*

$$\mathrm{Reg}_T(\pi_{\mathrm{MD}}) \le \widetilde{O}(q\sqrt{T}), \qquad v(\pi_{\mathrm{MD}}, \mathcal{U}_\Phi) \le O(\sqrt{qT}).$$

More explicitly, the regret bound scales as $O(q(1+\log T)\sqrt{T})$ up to norm-equivalence constants.

6 Improved Stochastic Regret via Accelerate-Then-Refine

In previous sections, we have introduced two first order algorithms achieving $O(q\sqrt{T})$ regret bound, which is optimal for general online convex optimization. Since the regularized projected dual problem is strongly convex over the dual variable, we know that the corresponding regularized projected dual has the unique optimizer $w^\star$. To obtain sharper regret guarantees, we need to exploit additional structure beyond convexity. We then impose a common assumption which is widely used in related literature (Zhang & Luo, 2022; Chen et al., 2024; Gao et al., 2025).

Assumption 4 (Linear error bound)**.** There exists $\lambda > 0$ such that, for all $w \in \mathbb{R}^q_+$,

$$f_\Phi^\rho(w) - f_\Phi^\rho(w^\star) \ge \lambda \operatorname{dist}(w, \{w^\star\}).$$

Remark 2. The above condition is a linear Hölder error bound, also known as a sharpness or weak sharp minima condition. It is stronger than quadratic growth, since it relates the objective gap to the distance, rather than the squared distance, from the optimal solution set. This assumption is used to translate objective decay into distance contraction in our convergence analysis. We evaluate the three algorithms from the perspectives of regret and projected constraint violation.

We present our formal algorithm in Algorithm 3. Our improved algorithm has two stages. The first stage runs an epoch-based projected stochastic-gradient method on the dual objective, with shrinking localization balls. It returns a point $\widehat{w}$ within $O(1/T)$ of the unique minimizer $w^\star$ of the population objective f_Φ^ρ. We illustrate the details of epoch-based projected SGD in Appendix D.

The second stage of Algorithm 3 (Refine Stage) is designed to exploit the localized dual estimate produced by the first stage. Unlike the projected-gradient and mirror-descent algorithms, which initialize the dual variable at 0, the Refine Stage starts from $\widehat{w}$, an $O(1/T)$-accurate approximation of the population dual optimizer $w^\star$. This warm start prevents the algorithm from spending a large number of rounds learning the correct dual price from scratch. Instead, the subsequent online updates only need to track local fluctuations around $w^\star$.

Algorithm 3 Accelerate-Then-Refine

1: **Input:** horizon T, budget d, basis Φ, regularization $\rho \in (0, 1/T]$, total budget $B_1 = Td_\Phi$.
2: **Accelerate:** Run epoch-based projected SGD for $T_{\text{fast}} = O(\log^2 T)$ rounds to obtain $\widehat{w}$ satisfying $\|\widehat{w} - w^\star\|_2 \le 1/T$ with decisions $\tilde{x}_1, \dots \tilde{x}_{T_{\text{fast}}}$ and $B_{T_{\text{fast}}+1}$.
3: Set $H = T - T_{\text{fast}}, d_{\text{ref}} = \frac{B_{T_{\text{fast}}+1}}{H}, w_{T_{\text{fast}}+1} = \widehat{w}, \gamma_{\text{ref}} = \frac{1}{T}$.
4: **Refine:** For each remaining round, use threshold decision $x_t = \mathbf{1}\{r_t > c_t^\top w_t\}$, enforce remaining budget feasibility by $\widetilde{x}_t = x_t \mathbf{1}\{B_t \ge c_t x_t\}$, and update

$$w_{t+1} = \Pi_{\mathbb{R}^q_+}\left(w_t - \gamma_{\text{ref}}(d_{\text{ref}} - c_t x_t + 2\rho w_t)\right), \quad B_{t+1} = B_t - c_t \widetilde{x}_t.$$

5: **Output:** decisions $\widetilde{x}_1, \dots, \widetilde{x}_T$.

To preserve this localization, the Refine Stage deliberately uses a $1/T$-scale step size, rather than the $T^{-1/2}$-scale step size used in Algorithms 1 and 2. The reason is that, after the first stage, the algorithm is already initialized within $O(1/T)$ of the population dual optimizer $w^\star$. Therefore, the Refine Stage is not intended to perform global dual learning. Instead, it only applies small local corrections based on the realized projected residuals. A larger $T^{-1/2}$ step size would introduce unnecessary dual movement and could move the iterates outside the neighborhood where the negative-drift property guarantees stability of the threshold decisions. The smaller $1/T$-scale update keeps dual sequence localized near $w^\star$, which is crucial for obtaining improved logarithmic regret bound.

We now present the end-to-end guarantee for Algorithm 3. The analysis decomposes the regret into the two stages of the algorithm. The Accelerate Stage uses epoch-based projected stochastic gradient updates to localize the dual variable near $w^\star$, contributing an $O(q \log T)$ regret bound. The Refine Stage achieves $O(1)$ regret bound, independent of the horizon T. In addition, the Refine Stage also enforces remaining projected budget feasibility, which leads to the final zero projected violation. The detailed proofs of the two stage-wise bounds are given in Appendix B.8. Combining them gives the end-to-end logarithmic regret guarantee stated below.

Theorem 4 (Logarithmic stochastic regret)**.** *Under the stochastic input model, Assumptions 1 and 4, Algorithm 3 with* $0 < \rho \le T^{-1}$ *satisfies,*

$$\text{Reg}_T(\pi_{\text{ATR}}) \le O\left(q \log T\right), \qquad v(\pi_{\text{ATR}}, \mathcal{U}_\Phi) = 0.$$

7 Experiments

We evaluate the proposed algorithms on large-scale online LP instances with many constraints. The goal is to test whether basis-based dual updates remain stable as the number of constraints grows.

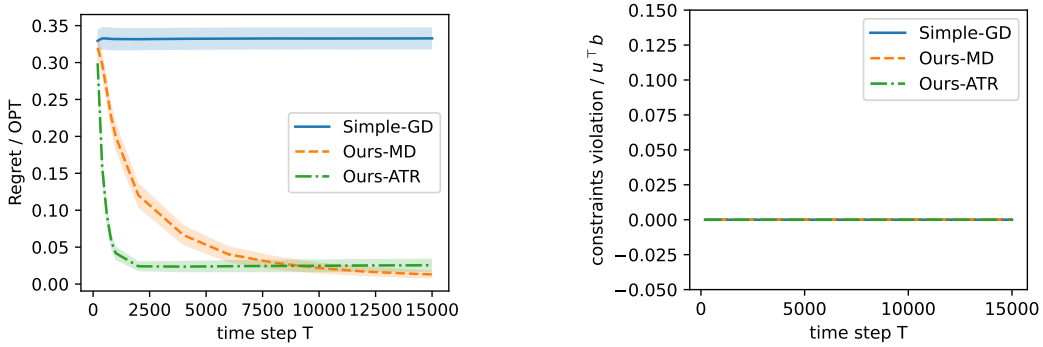
Basis construction. We use nonnegative radial basis functions. For a discretized constraint index $i = 1, \dots, m$, define $u_i = (i - 1/2)/m$. Each basis column is

$$\phi_k(u_i) = \exp\left(-\frac{(u_i - c_k)^2}{2\sigma_k^2}\right).$$

We use a two-resolution design: a coarse layer with broad bandwidths and a fine layer with narrower bandwidths. This gives a compact representation of smooth dual prices while preserving nonnegativity. We use fixed basis dimension $q = 200$ in our experiments.

Baselines. We compare three different algorithms: **Ours-MD**, Algorithm 2 with Euclidean mirror map; **Ours-ATR**, the proposed Algorithm 3; **Simple-GD**, the first-order online LP method of (Li et al., 2023).

Stochastic arrivals. For stochastic experiments, we test a bounded uniform distribution, and use Gaussian and Cauchy distributions as unbounded-input stress tests. Each setting is repeated over 100 trials, and we report averages with nominal 95% confidence intervals. Figure 1 presents the results of three algorithms under uniform distribution. We can observe that the ratio of regret and optimal value of both our Ours-MD and Ours-ATR can converge to 0, and Ours-ATR converges faster, which



(a) Regret, $m = 2000$.

(b) Constraint violation, $m = 2000$.

Figure 1: Stochastic uniform arrivals. Both proposed methods exhibit vanishing normalized regret, while the discretized baseline deteriorates when the number of constraints is large.

is consistent with our main results. In contrast, Simple-GD performs substantially worse, and its curve in Figure 1a suggests approximately linear regret growth over T. This poor performance is consistent with its $O(m\sqrt{T})$ regret bound having a linear dependence on the number of constraints, whose dimension-dependent term dominates the regret in our setting. In addition, Figure 1b further shows that all three algorithms achieve small normalized constraint violations. We leave additional results under Gaussian and Cauchy distributions to Appendix E.

Scaling with the number of constraints. We now consider fixed time horizion T and vary m. The proposed methods depend on the basis dimension q, which remains fixed, while the baseline maintains m-dimensional dual variables. For all three different distributions, Ours-MD and Ours-ATR perform well even with fewer than 500 constraints, whereas Simple-GD already degrades. As the number of constraints increases, Ours-ATR remains stable, while Simple-GD further deteriorates, especially under Cauchy distribution. We leave detailed experimental results under all three distributions in Appendix E.

Random permutation arrivals. For the random permutation model, we sample the input data a_{ij} from following five categories: i) Uniform distribution $U[0, 3]$; ii) Normal distribution $N[2, 1]$; iii) Normal distribution $N(1, 1)$ with different mean; iv) Normal distribution with different standard deviation $N(0, 1)$; v) an integer set $\{-1, 0, 1\}$. Besides, the input coefficients r_t are sampled from $U[0, 1]$ while the right-hand-side d are sampled from $U[\frac{2}{5}, \frac{4}{5}]$. We reveal the input data in a uniformly random order. The proposed Ours-MD maintains stable performance as the constraint dimension increases, converges faster than Simple-GD, and converges to zero, whereas Simple-GD does not. Both two algorithms have excellent performance of constraints violation. The constraints violation of Simple-GD is 0 and Ours-MD violates the constraints very small. Detailed results are provided in Appendix E.

8 Conclusion

In this work, we introduced an online framework for semi-infinite linear programming based on nonnegative function approximation of the dual cone. The approach transforms an infinite-dimensional dual update into a q-dimensional update while preserving a deterministic upper bound on the original offline optimum. This yields scalable primal–dual algorithms whose regret depends on the approximation dimension rather than on the number of constraints. Theoretical guarantees under stochastic and random-permutation arrivals, together with large-scale experiments, suggest that nonnegative dual approximation is a promising tool for online optimization with massive or continuous constraint families. However, the approximation quality depends on the chosen basis family. Projected feasibility is guaranteed directly, while exact pointwise feasibility requires an additional density or approximation condition. We leave addressing these limitations to future research.

AI USE STATEMENT

Generative AI tools were used in this work to assist with language polishing and experimental coding. Specifically, we used these tools to improve the clarity, grammar, and readability of author-written text, and to assist with writing, debugging, and refining code used in our experimental pipeline. The authors take full responsibility for the content of this paper and all associated artifacts.

REPRODUCIBILITY STATEMENT

The complete proofs of our theoretical results are provided in Appendix B. The code for reproducing the experiments is included in the supplementary materials, and details of the computational resources are provided in Appendix E.1.

REFERENCES

Shipra Agrawal and Nikhil R Devanur. Fast algorithms for online stochastic convex programming. In *Proceedings of the twenty-sixth annual ACM-SIAM symposium on Discrete algorithms*, pp. 1405–1424. SIAM, 2014.

Shipra Agrawal, Zizhuo Wang, and Yinyu Ye. A dynamic near-optimal algorithm for online linear programming. *Operations Research*, 62(4):876–890, 2014.

Santiago Balseiro, Haihao Lu, and Vahab Mirrokni. Dual mirror descent for online allocation problems. In *International Conference on Machine Learning*, pp. 613–628. PMLR, 2020.

Rémi Bardenet and Odalric-Ambrym Maillard. Concentration inequalities for sampling without replacement. 2015.

Omar Besbes, Yonatan Gur, and Assaf Zeevi. Non-stationary stochastic optimization. *Operations research*, 63(5):1227–1244, 2015.

Bruno Betro. An accelerated central cutting plane algorithm for linear semi-infinite programming. *Mathematical Programming*, 101(3):479–495, 2004.

Niv Buchbinder and Joseph Naor. Online primal-dual algorithms for covering and packing problems. In *European Symposium on Algorithms*, pp. 689–701. Springer, 2005.

Niv Buchbinder and Joseph Naor. Improved bounds for online routing and packing via a primal-dual approach. In *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*, pp. 293–304. IEEE, 2006.

Pengyu Chen, Xu Shi, Rujun Jiang, and Jiulin Wang. Penalty-based methods for simple bilevel optimization under hölderian error bounds. *Advances in Neural Information Processing Systems*, 37:140731–140765, 2024.

Xiao Alison Chen and Zizhuo Wang. A dynamic learning algorithm for online matching problems with concave returns. *European Journal of Operational Research*, 247(2):379–388, 2015.

Nikhil R Devanur and Thomas P Hayes. The adwords problem: online keyword matching with budgeted bidders under random permutations. In *Proceedings of the 10th ACM conference on Electronic commerce*, pp. 71–78, 2009.

Nikhil R Devanur and Kamal Jain. Online matching with concave returns. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pp. 137–144, 2012.

Jon Feldman, Monika Henzinger, Nitish Korula, Vahab S Mirrokni, and Cliff Stein. Online stochastic packing applied to display ad allocation. In *European Symposium on Algorithms*, pp. 182–194. Springer, 2010.

Wenzhi Gao, Dongdong Ge, Chenyu Xue, Chunlin Sun, and Yinyu Ye. Beyond $\mathcal{O}(\sqrt{T})$ regret: Decoupling learning and decision-making in online linear programming. *arXiv preprint arXiv:2501.02761*, 2025.

Eric Hall and Rebecca Willett. Dynamical models and tracking regret in online convex programming. In *International Conference on Machine Learning*, pp. 579–587. PMLR, 2013.

Elad Hazan. Introduction to online convex optimization. *Foundations and Trends in Optimization*, 2 (3-4):157–325, 2016.

Siqi He, Yehua Wei, Jiaming Xu, and Sophie H Yu. Online resource allocation without re-solving: The effectiveness of primal-dual policies. 2025.

Ali Jadbabaie, Alexander Rakhlin, Shahin Shahrampour, and Karthik Sridharan. Online optimization: Competing with dynamic comparators. In *Artificial Intelligence and Statistics*, pp. 398–406. PMLR, 2015.

Stefanus Jasin and Sunil Kumar. A re-solving heuristic with bounded revenue loss for network revenue management with customer choice. *Mathematics of Operations Research*, 37(2):313–345, 2012.

Rodolphe Jenatton, Jim Huang, and Cédric Archambeau. Adaptive algorithms for online convex optimization with long-term constraints. In *International Conference on Machine Learning*, pp. 402–411. PMLR, 2016.

Jiashuo Jiang, Will Ma, and Jiawei Zhang. Degeneracy is ok: Logarithmic regret for network revenue management with indiscrete distributions. *Operations Research*, 73(6):3405–3420, 2025.

Daniel Jungen, Hatim Djelassi, and Alexander Mitsos. Adaptive discretization-based algorithms for semi-infinite programs with unbounded variables. *Mathematical Methods of Operations Research*, 96(1):83–112, 2022.

Thomas Kesselheim, Andreas Tönnis, Klaus Radke, and Berthold Vöcking. Primal beats dual on online packing lps in the random-order model. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pp. 303–312, 2014.

Xiaocheng Li and Yinyu Ye. Online linear programming: Dual convergence, new algorithms, and regret bounds. *Operations Research*, 70(5):2948–2966, 2022.

Xiaocheng Li, Chunlin Sun, and Yinyu Ye. Simple and fast algorithm for binary integer and online linear programming. *Mathematical Programming*, 200(2):831–875, 2023.

Sanjay Mehrotra and Dávid Papp. A cutting surface algorithm for semi-infinite convex programming with an application to moment robust optimization. *SIAM Journal on Optimization*, 24(4):1670–1697, 2014.

Aranyak Mehta, Amin Saberi, Umesh Vazirani, and Vijay Vazirani. Adwords and generalized online matching. *Journal of the ACM (JACM)*, 54(5):22–es, 2007.

Alexander Mitsos. Global optimization of semi-infinite programs via restriction of the right-hand side. *Optimization*, 60(10-11):1291–1308, 2011.

Marco Molinaro and Ramamoorthi Ravi. The geometry of online packing linear programs. *Mathematics of Operations Research*, 39(1):46–59, 2014.

Michael J Neely and Hao Yu. Online convex optimization with time-varying constraints. *arXiv preprint arXiv:1702.04783*, 2017.

Antoine Oustry and Martina Cerulli. Convex semi-infinite programming algorithms with inexact separation oracles. *Optimization Letters*, 19(3):437–462, 2025.

Rembert Reemtsen. Semi-infinite programming: discretization methods. In *Encyclopedia of Optimization*, pp. 1–8. Springer, 2025.

Georg Still. Discretization in semi-infinite programming: the rate of convergence. *Mathematical programming*, 91(1):53–69, 2001.

Kalyan T Talluri and Garrett J Van Ryzin. *The theory and practice of revenue management*, volume 68. Springer Science & Business Media, 2006.

Xinlei Yi, Xiuxian Li, Tao Yang, Lihua Xie, Tianyou Chai, and Karl Johansson. Regret and cumulative constraint violation analysis for online convex optimization with long term constraints. In *International conference on machine learning*, pp. 11998–12008. PMLR, 2021.

Jianjun Yuan and Andrew Lamperski. Online convex optimization for cumulative constraints. *Advances in Neural Information Processing Systems*, 31, 2018.

Jiawei Zhang and Zhi-Quan Luo. A global dual error bound and its application to the analysis of linearly constrained nonconvex optimization. *SIAM Journal on Optimization*, 32(3):2319–2346, 2022.

LiPing Zhang, Shu-Cherng Fang, and Soon-Yi Wu. An entropy based central cutting plane algorithm for convex min-max semi-infinite programming problems. *Science China Mathematics*, 56(1): 201–211, 2013.

A Related Literature

In this section, we introduce some related works to our paper, which are grouped into following four directions:

Semi-Infinite Linear Programming Semi-Infinite Linear Programming (SILP) addresses linear optimization problems with finite decision variables but infinite inequality constraints. Classical SILP algorithms typically follow a finite-approximation–correction paradigm: solve a master problem over a finite subset of constraints, then identify a new violating index and augment the active set. Several previous literature develops new algorithms based on the central cutting plane method (Betro, 2004; Mehrotra & Papp, 2014; Zhang et al., 2013), where any sufficiently informative violated or nearly tight constraint can generate an effective cut. Recently, many literature focuses on the adaptive discretization (Still, 2001; Mitsos, 2011; Jungen et al., 2022; Reemtsen, 2025). In addition, (Oustry & Cerulli, 2025) introduces a novel method that employing an inexact oracle to handle convex semi-infinite programmings.

Online Linear Programming/Resource Allocation Online Linear Programming/resource allocation has long been investigated and has various applications, such as displaying ad allocation (Mehta et al., 2007), revenue management (Talluri & Van Ryzin, 2006; Jasin & Kumar, 2012; Jiang et al., 2025), covering and packing problems (Buchbinder & Naor, 2005; 2006; Feldman et al., 2010) and more. A classical and prevalent framework for designing algorithm is the dual-based method, where the decision is made based on an estimated dual prices. Some literature follow the sample-then-price scheme, that is the algorithm samples an initial prefix to learn dual prices and then fixes them for the rest of the horizon (Devanur & Hayes, 2009; Feldman et al., 2010; Molinaro & Ravi, 2014; Devanur & Jain, 2012). In contrast, other literature adopt the dynamic price learning scheme, which re-solves and updates prices periodically, improving robustness and tightening dependence on horizon (Agrawal et al., 2014; Li & Ye, 2022; Chen & Wang, 2015). In addition, there are also some papers focusing on primal-guided methods (Kesselheim et al., 2014) and first-order method without repeating solving dual price (Agrawal & Devanur, 2014; Balseiro et al., 2020).

Role of dual regularization. A further distinction from standard online LP analyses is that our algorithms and guarantees are built around a regularized dual objective. Classical online LP methods typically analyze unregularized finite-dimensional dual prices associated with a fixed set of resources or constraints. In the semi-infinite setting, however, the unregularized dual object is a nonnegative measure or function over $\mathcal{S}$, which may be nonsmooth, nonunique, and difficult to approximate uniformly. The dual regularizer plays two roles in our framework. Algorithmically, it stabilizes the online dual updates and yields bounded reduced-dimensional iterates. Analytically, it makes the exact infinite-dimensional dual problem strongly convex, leading to a unique optimizer with additional regularity. This regularity is important for connecting the projected basis representation to the original semi-infinite constraint system.

Online Convex Optimization and Constrained Variants. Our work is also related to online convex optimization (OCO). In the standard OCO framework, the decision maker chooses the decision before observing the loss function, and then incurs loss. This model is mainly motivated by

machine learning applications such as online linear regression and online support vector machines (Hazan, 2016). Compared with our setting, standard OCO typically adopts a weaker static regret benchmark, where the comparator is fixed over time. By contrast, our benchmark is the offline optimum of the realized instance, which makes standard OCO algorithms and analyses not directly applicable. Several works study dynamic regret in non-stationary or adversarial environments (Besbes et al., 2015; Hall & Willett, 2013; Jadbabaie et al., 2015), but they focus on unconstrained problems. There is also a line of work on online convex optimization with constraints (OCOwC), where the constraint functions are either static (Jenatton et al., 2016; Yuan & Lamperski, 2018; Yi et al., 2021) or i.i.d. over time (Neely & Yu, 2017). To the best of our knowledge, existing OCOwC results do not allow for a large or infinite number of constraints, which is the problem we address in this paper.

B Omitted Proofs

B.1 Proof of Lemma 1

Fix any feasible x for the original semi-infinite LP. Since each basis function is nonnegative, for every $k \in [q]$,

$$\sum_{t=1}^{T} \left(\int_{\mathcal{S}} \phi_k(s) a_t(s) d\sigma(s) \right) x_t = \int_{\mathcal{S}} \phi_k(s) \left(\sum_{t=1}^{T} a_t(s) x_t \right) d\sigma(s) \le T \int_{\mathcal{S}} \phi_k(s) d(s) d\sigma(s).$$

Thus every feasible solution of the original problem is feasible for the projected primal problem, so $\mathrm{OPT}(I) \le \mathrm{OPT}_\Phi(I)$.

The dual of the projected primal is

$$f_\Phi(w) = \min_{w \in \mathbb{R}^q_+} \left\{ T d_\Phi^\top w + \sum_{t=1}^{T} (r_t - c_t^\top w)^+ \right\}.$$

By strong duality for finite-dimensional linear programs,

$$\mathrm{OPT}_\Phi(I) = T \min_{w \in \mathbb{R}^q_+} \left\{ d_\Phi^\top w + \frac{1}{T} \sum_{t=1}^{T} (r_t - c_t^\top w)^+ \right\}.$$

Adding the nonnegative regularizer $\rho \|w\|_2^2$ can only increase the minimum, yielding

$$\mathrm{OPT}_\Phi(I) \le T \min_{w \in \mathbb{R}^q_+} f^\rho_{T,\Phi}(w).$$

This proves the lemma.

B.2 Proof of Lemma 3

We prove by induction that

$$\|w_t\|_2 \le W_\Phi = \frac{q(\overline{C} + \overline{D})^2 + 2\overline{r}}{2\underline{D}} + \sqrt{q}(\overline{C} + \overline{D}), \qquad t = 1, \dots, T.$$

Base case. At $t = 1$, the claim holds by the initialization of the algorithm, since

$$w_1 = 0,$$

and hence $\|w_1\|_2 \le W_\Phi$.

Induction hypothesis. Assume that for some $t \in [T]$,

$$\|w_t\|_2 \le W_\Phi.$$

Induction step. We need to show that

$$\|w_{t+1}\|_2 \le W_\Phi.$$

To this end, define the auxiliary unregularized one-step update

$$\widehat{w}_{t+1} = \Pi_{\mathbb{R}^q_+}\big[w_t + \gamma_t(c_t x_t - d_\Phi)\big]. \tag{11}$$

and we will first establish that, under the induction hypothesis,

$$\|\widehat{w}_{t+1}\|_2 \le W_\Phi.$$

From the unregularized update equation 11, we have:

$$\begin{aligned} ||\widehat{w}_{t+1}||_2^2 &\le ||w_t + \gamma_t(c_t x_t - d_\Phi)||_2^2 \\ &= ||w_t||_2^2 - 2\gamma_t(d_\Phi - c_t x_t)w_t + \gamma_t^2||d_\Phi - c_t x_t||_2^2 \\ &\le ||w_t||_2^2 - 2\gamma_t d_\Phi w_t + 2\gamma_t c_t x_t w_t + \gamma_t^2 q(\overline{C}+\overline{D})^2 \\ &= ||w_t||_2^2 + \gamma_t^2 q(\overline{C}+\overline{D})^2 + 2\gamma_t c_t w_t \mathbb{I}\{r_t > c_t w_t\} - 2\gamma_t d_\Phi w_t \\ &\le ||w_t||_2^2 + \gamma_t^2 q(\overline{C}+\overline{D})^2 + 2\gamma_t r_t - 2\gamma_t d_\Phi w_t \end{aligned} \tag{12}$$

where the first inequality follows from the property of Euclidean projection, the second inequality relies on Assumption 1 and the third inequality utilizes the condition of indicator function $\mathbb{I}$. Next, we will show that when $||w_t||_2$ is large enough, there must have $||\widehat{w}_{t+1}||_2 \le ||w_t||_2$. To be specific, when $||w_t||_2 \ge \frac{q(\overline{C}+\overline{D})^2+2\overline{r}}{2\underline{D}}$, we have the following inequalities:

$$\begin{aligned} ||\widehat{w}_{t+1}||_2^2 - ||w_t||_2^2 &\le \gamma_t^2 q(\overline{C}+\overline{D})^2 + 2\gamma_t r_t - 2\gamma_t d_\Phi w_t \\ &\le \gamma_t^2 q(\overline{C}+\overline{D})^2 + 2\gamma_t r_t - 2\gamma_t \underline{D} w_t \\ &\le 0, \end{aligned} \tag{13}$$

which means $||\widehat{w}_{t+1}||_2 \le \|w_t\|_2 \le W_\Phi$. Otherwise, when $||w_t||_2 < \frac{q(\overline{C}+\overline{D})^2+2\overline{r}}{2\underline{D}}$, we have:

$$\begin{aligned} ||\widehat{w}_{t+1}||_2 &\le ||w_t + \gamma_t(c_t x_t - d_\Phi)||_2 \\ &\le ||w_t||_2 + ||\gamma_t(c_t x_t - d_\Phi)||_2 \\ &\le \frac{q(\overline{C}+\overline{D})^2+2\overline{r}}{2\underline{D}} + \sqrt{q}(\overline{C}+\overline{D}) \\ &= W_\Phi. \end{aligned} \tag{14}$$

Combining the two cases above, we know that

$$||\widehat{w}_{t+1}||_2 \le W_\Phi = \frac{q(\overline{C}+\overline{D})^2+2\overline{r}}{2\underline{D}} + \sqrt{q}(\overline{C}+\overline{D})$$

always holds.

Compare the regularized update with the unregularized one equation 11, since $-2\rho w_t \le 0$ coordinatewise and the projection onto $\mathbb{R}^q_+$ is monotone coordinatewise, we have

$$0 \le w_{t+1} \le \widehat{w}_{t+1} \qquad \text{coordinatewise.}$$

Therefore,

$$\|w_{t+1}\|_2 \le \|\widehat{w}_{t+1}\|_2 \le W_\Phi.$$

This completes the induction step.

Conclusion. By induction, we can conclude that

$$\|w_t\|_2 \le W_\Phi, \qquad \forall\, t = 1, \ldots, T.$$

B.3 Proof of Theorem 1

We first prove the expected regret bound under the stochastic input model. The potential inequality used below holds pathwise, whereas the comparison with the offline optimum is made in expectation.

Fix an arbitrary distribution $\mathcal{P} \in \Xi$ satisfying Assumption 1, and let $\mathcal{F}_{t-1}$ denote the history before the arrival of (r_t, a_t). The basis Φ and the budget d_Φ are fixed before the online process. Thus, w_t is $\mathcal{F}_{t-1}$-measurable, and (r_t, c_t) is independent of $\mathcal{F}_{t-1}$ under the i.i.d. input model. Write $\gamma = T^{-1/2}$.

Step 1: Population-dual comparison. Define the unregularized single-period dual loss and population dual objective by

$$g_t^0(w) = \langle d_\Phi, w\rangle + (r_t - c_t^\top w)^+, \qquad f_\Phi^0(w) = \langle d_\Phi, w\rangle + \mathbb{E}_{\mathcal{P}}\big[(r - c^\top w)^+\big], \qquad w \geq 0.$$

These quantities are used only for analysis; Algorithm 1 still uses its regularized update.

For every deterministic $w \geq 0$, weak duality for the projected LP and the nonnegativity of the basis give

$$\mathrm{OPT}(I) \leq \mathrm{OPT}_\Phi(I) \leq \sum_{t=1}^{T} g_t^0(w).$$

Taking expectations and using the identical distribution of the arrivals yields

$$\mathbb{E}_{\mathcal{P}}[\mathrm{OPT}(I)] \leq T f_\Phi^0(w).$$

Since this holds for every deterministic $w \geq 0$,

$$\mathbb{E}_{\mathcal{P}}[\mathrm{OPT}(I)] \leq T \inf_{w\geq 0} f_\Phi^0(w).$$

Moreover, by the predictability of w_t and the independence of the current arrival,

$$\mathbb{E}_{\mathcal{P}}\big[g_t^0(w_t) \mid \mathcal{F}_{t-1}\big] = f_\Phi^0(w_t).$$

Since $f_\Phi^0(w_t) \geq \inf_{w\geq 0} f_\Phi^0(w)$ almost surely, we obtain

$$\mathbb{E}_{\mathcal{P}}[\mathrm{OPT}(I)] \leq T \inf_{w\geq 0} f_\Phi^0(w) \leq \sum_{t=1}^{T} \mathbb{E}_{\mathcal{P}}[f_\Phi^0(w_t)] = \mathbb{E}_{\mathcal{P}}\left[\sum_{t=1}^{T} g_t^0(w_t)\right]. \tag{15}$$

Step 2: A pathwise potential inequality. Let

$$h_t := d_\Phi - c_t x_t.$$

The actual update of Algorithm 1 is

$$w_{t+1} = \Pi_{\mathbb{R}^q_+}\big(w_t - \gamma h_t - 2\gamma\rho w_t\big).$$

For comparison, define the auxiliary one-step point

$$\widetilde{w}_{t+1} := \Pi_{\mathbb{R}^q_+}(w_t - \gamma h_t).$$

This auxiliary point is computed from the actual iterate w_t and the actual decision x_t; it does not define a separate unregularized algorithm.

Because $\rho \geq 0$ and $w_t \geq 0$, and projection onto $\mathbb{R}^q_+$ is coordinatewise monotone, we have

$$0 \leq w_{t+1} \leq \widetilde{w}_{t+1} \qquad \text{coordinatewise.}$$

Consequently,

$$\begin{aligned}\|w_{t+1}\|_2^2 &\leq \|\widetilde{w}_{t+1}\|_2^2 \\ &\leq \|w_t - \gamma h_t\|_2^2 \\ &= \|w_t\|_2^2 - 2\gamma\langle w_t, h_t\rangle + \gamma^2\|h_t\|_2^2.\end{aligned}$$

Rearranging gives

$$\langle w_t, h_t \rangle \leq \frac{\|w_t\|_2^2 - \|w_{t+1}\|_2^2}{2\gamma} + \frac{\gamma}{2}\|h_t\|_2^2. \tag{16}$$

Since

$$x_t = \mathbf{1}\{r_t > c_t^\top w_t\},$$

the threshold identity gives

$$(r_t - c_t^\top w_t)^+ = x_t(r_t - c_t^\top w_t).$$

Therefore,

$$g_t^0(w_t) - r_t x_t = \langle w_t, d_\Phi - c_t x_t \rangle = \langle w_t, h_t \rangle.$$

Combining this identity with equation 16, summing over t, and using $w_1 = 0$, we obtain the pathwise bound

$$\begin{aligned}\sum_{t=1}^{T}\big(g_t^0(w_t) - r_t x_t\big) &\leq -\frac{\|w_{T+1}\|_2^2}{2\gamma} + \frac{\gamma}{2}\sum_{t=1}^{T}\|h_t\|_2^2 \\ &\leq \frac{\gamma}{2}\sum_{t=1}^{T}\|h_t\|_2^2. \end{aligned}\tag{17}$$

Step 3: Expected regret bound. By Assumption 1, the nonnegativity of the projected coefficients, and $x_t \in \{0, 1\}$, we have

$$\|h_t\|_2^2 = \|d_\Phi - c_t x_t\|_2^2 \leq q(\overline{C} + \overline{D})^2 \qquad \text{for every } t.$$

Using equation 15 and taking expectations in equation 17, we obtain

$$\begin{aligned}\mathbb{E}_{\mathcal{P}}\left[\mathrm{OPT}(I) - \sum_{t=1}^{T} r_t x_t\right] &\leq \mathbb{E}_{\mathcal{P}}\left[\sum_{t=1}^{T}\big(g_t^0(w_t) - r_t x_t\big)\right] \\ &\leq \frac{\gamma}{2}\sum_{t=1}^{T}\mathbb{E}_{\mathcal{P}}[\|h_t\|_2^2] \\ &\leq \frac{\gamma T q}{2}(\overline{C} + \overline{D})^2 \\ &= \frac{q(\overline{C} + \overline{D})^2}{2}\sqrt{T}.\end{aligned}$$

The right-hand side is uniform over all $\mathcal{P} \in \Xi$ satisfying the same boundedness constants. Taking the supremum over $\mathcal{P} \in \Xi$ therefore gives

$$\mathrm{Reg}_T(\pi_{\mathrm{GD}}) \leq \frac{q(\overline{C} + \overline{D})^2}{2}\sqrt{T} = O(q\sqrt{T}).$$

This proves the regret part of the theorem.

Step 4: Projected constraint violation. For each coordinate $j \in [q]$, the update rule gives

$$(w_{t+1})_j = \Big[(w_t)_j + \gamma\big((c_t)_j x_t - (d_\Phi)_j - 2\rho(w_t)_j\big)\Big]^+.$$

Since $[y]^+ \geq y$ for every scalar y,

$$(c_t)_j x_t - (d_\Phi)_j \leq \frac{(w_{t+1})_j - (w_t)_j}{\gamma} + 2\rho(w_t)_j.$$

Summing over $t = 1, \ldots, T$, using $(w_1)_j = 0$, yields

$$\sum_{t=1}^{T}\big((c_t)_j x_t - (d_\Phi)_j\big) \leq \frac{(w_{T+1})_j}{\gamma} + 2\rho\sum_{t=1}^{T}(w_t)_j.$$

Since $(w_t)_j \le \|w_t\|_2 \le W_\Phi$ (Lemma 3),

$$\big[\sum_{t=1}^{T}\big((c_t)_j x_t - (d_\Phi)_j\big)\big]^+ \le \frac{W_\Phi}{\gamma} + 2\rho T W_\Phi.$$

Taking the maximum over $j \in [q]$, and using $\gamma = 1/\sqrt{T}$ and $\rho \le 1/\sqrt{T}$, we obtain

$$\big\|[\sum_{t=1}^{T} c_t x_t - T d_\Phi]^+\big\|_\infty \le 3W_\Phi\sqrt{T}.$$

Finally, by the definition of $\mathcal{U}_\Phi$,

$$\max_{u\in\mathcal{U}_\Phi}\langle u, A\boldsymbol{x}-\boldsymbol{b}\rangle = \max_{\substack{w\ge 0\\ \boldsymbol{m}_\Phi^\top w=1}} \langle w, \sum_{t=1}^{T} c_t x_t^\pi - T d_\Phi\rangle \le \frac{1}{\underline{m}_\Phi}\big\|[\sum_{t=1}^{T} c_t x_t - T d_\Phi]^+\big\|_\infty.$$

Therefore,

$$v(\pi_{\mathrm{GD}}, \mathcal{U}_\Phi) \le \frac{3W_\Phi}{\underline{m}_\Phi}\sqrt{T}.$$

Under Assumption 1, the prefactor is $O(q^2)$, which proves

$$v(\pi_{\mathrm{GD}}, \mathcal{U}_\Phi) \le O(q^2\sqrt{T}).$$

This completes the proof.

B.4 Proof of Theorem 2

We first prove the regret bound. By the optimality condition of the mirror-descent step, for any $u \in \mathbb{R}^q_+$,

$$\gamma\langle\hat{\zeta}_t^\rho, w_{t+1}-u\rangle \le D_\psi(u\|w_t) - D_\psi(u\|w_{t+1}) - D_\psi(w_{t+1}\|w_t). \tag{18}$$

Strong convexity and Young's inequality give

$$D_\psi(w_{t+1}\|w_t) \ge \frac{\alpha}{2}\|w_{t+1}-w_t\|^2,$$

$$\gamma\langle\hat{\zeta}_t^\rho, w_t - w_{t+1}\rangle \le \frac{\alpha}{2}\|w_{t+1}-w_t\|^2 + \frac{\gamma^2}{2\alpha}\|\hat{\zeta}_t^\rho\|_*^2.$$

Adding $\gamma\langle\hat{\zeta}_t^\rho, w_t - w_{t+1}\rangle$ to equation 18 therefore yields

$$\langle\hat{\zeta}_t^\rho, w_t - u\rangle \le \frac{D_\psi(u\|w_t) - D_\psi(u\|w_{t+1})}{\gamma} + \frac{\gamma}{2\alpha}\|\hat{\zeta}_t^\rho\|_*^2. \tag{19}$$

For the primal-dual comparison, let

$$f_\Phi^0(w) := d_\Phi^\top w + \mathbb{E}[(r - c^\top w)^+].$$

Nonnegativity of the basis and weak duality imply

$$\mathbb{E}[\mathrm{OPT}(I)] \le T\inf_{w\ge 0} f_\Phi^0(w) \le \sum_{t=1}^{T}\mathbb{E}[f_\Phi^0(w_t)].$$

Since w_t depends only on past arrivals, independence of the current arrival and the threshold identity $(r_t - c_t^\top w_t)^+ = x_t(r_t - c_t^\top w_t)$ give

$$\begin{aligned}\mathbb{E}\left[\mathrm{OPT}(I) - \sum_{t=1}^{T} r_t x_t\right] &\le \sum_{t=1}^{T}\mathbb{E}[\langle d_\Phi - c_t x_t, w_t\rangle]\\ &\le \sum_{t=1}^{T}\mathbb{E}[\langle\hat{\zeta}_t^\rho, w_t\rangle],\end{aligned}$$

where the last inequality uses $\hat{\zeta}_t^\rho = d_\Phi - c_t x_t + 2\rho w_t$ and $\rho \geq 0$. Applying equation 19 with $u = 0$, summing over t, and using $D_\psi(0\|w_1) = 0$ yields

$$\mathbb{E}\left[\mathrm{OPT}(I) - \sum_{t=1}^{T} r_t x_t\right] \leq \frac{\gamma}{2\alpha} \sum_{t=1}^{T} \mathbb{E}[\|\hat{\zeta}_t^\rho\|_*^2].$$

Under Assumptions 1 and 3, $\|\hat{\zeta}_t^\rho\|_2 \leq Z_\rho$, where

$$Z_\rho := \sqrt{q}\,(\overline{C} + \overline{D} + 2\rho\overline{W}). \tag{20}$$

Let K satisfy $\|z\|_*^2 \leq K\|z\|_2^2$ for all $z \in \mathbb{R}^q$. Taking the supremum over $P \in \Xi$, we obtain

$$\mathrm{Reg}_T(\pi_{\mathrm{MD}}) \leq \frac{K Z_\rho^2 \sqrt{T}}{2\alpha} = \frac{Kq}{2\alpha}(\overline{C} + \overline{D} + 2\rho\overline{W})^2\sqrt{T} = O(q\sqrt{T}).$$

We next prove the projected constraint-violation bound. Let $R_T := \sum_{t=1}^{T} c_t x_t - T d_\Phi$, and define the stopping time

$$\tau := \min\left(\left\{t \in \{0,\dots,T\} : \exists j \in [q], \quad \sum_{i=1}^{t}(c_i x_i)_j + \overline{C} \geq T(d_\Phi)_j\right\} \cup \{T\}\right),$$

with empty sums interpreted as zero. By minimality of τ and $(c_t x_t)_j \leq \overline{C}$, $\sum_{t=1}^{\tau} c_t x_t \leq T d_\Phi$ coordinatewise. Each remaining arrival contributes at most $\overline{C}$ per coordinate, so

$$\|[R_T]^+\|_2 \leq \overline{C}(T - \tau)\sqrt{q}. \tag{21}$$

For $\tau < T$, select a coordinate j satisfying the stopping condition. Write $\psi(w) = \sum_{j=1}^{q} \psi_j(w_j)$, and set

$$B_\psi := \max_{j \in [q]} \{\psi_j'(\overline{W}) - \psi_j'(0)\}.$$

The KKT condition for the mirror-descent update is

$$\gamma\hat{\zeta}_t^\rho + \nabla\psi(w_{t+1}) - \nabla\psi(w_t) + v_{t+1} = 0, \qquad v_{t+1} \in \mathcal{N}_{\mathbb{R}_+^q}(w_{t+1}).$$

Since $(v_{t+1})_j \leq 0$,

$$\gamma(d_\Phi - c_t x_t)_j \geq \psi_j'((w_t)_j) - \psi_j'((w_{t+1})_j) - 2\gamma\rho(w_t)_j.$$

Summing to τ, using $w_1 = 0$, monotonicity of ψ_j', and $\|w_t\|_\infty \leq \overline{W}$, gives

$$-\sum_{t=1}^{\tau}(d_\Phi - c_t x_t)_j \leq \frac{\psi_j'((w_{\tau+1})_j) - \psi_j'(0)}{\gamma} + 2\rho\sum_{t=1}^{\tau}(w_t)_j \leq \frac{B_\psi}{\gamma} + 2\rho\tau\overline{W}. \tag{22}$$

On the other hand, the stopping condition gives

$$(T - \tau)(d_\Phi)_j - \overline{C} \leq -\sum_{t=1}^{\tau}(d_\Phi - c_t x_t)_j.$$

Combining this with equation 22, $(d_\Phi)_j \geq \underline{D}$, and $\rho \leq T^{-1/2}$, we obtain

$$T - \tau \leq \frac{\overline{C}}{\underline{D}} + \frac{B_\psi}{\underline{D}\gamma} + \frac{2\rho T\overline{W}}{\underline{D}} \leq \frac{\overline{C} + (B_\psi + 2\overline{W})\sqrt{T}}{\underline{D}}.$$

This bound also holds for $\tau = T$.

Finally, let $\underline{m}_\Phi := \min_{j\in[q]}(m_\Phi)_j > 0$. By the definition of $\mathcal{U}_\Phi$ and equation 21,

$$\begin{aligned} v(\pi_{\mathrm{MD}}, \mathcal{U}_\Phi) &= \mathbb{E}\left[\max_{j\in[q]} \frac{((R_T)_j)^+}{(m_\Phi)_j}\right] \\ &\leq \frac{1}{\underline{m}_\Phi}\mathbb{E}[\|[R_T]^+\|_2] \\ &\leq \frac{\overline{C}\sqrt{q}}{\underline{m}_\Phi \underline{D}}\left[\overline{C} + (B_\psi + 2\overline{W})\sqrt{T}\right] = O(\sqrt{qT}). \end{aligned}$$

This completes the proof.

B.5 PROOF OF THEOREM 3

Let $w^\star$ be an optimal solution of equation 4 with $\rho = 0$, selected independently of the arrival order. For the auxiliary offline comparison, first consider rewards in general position: for every $w \geq 0$, at most q indices satisfy $r_i = c_i^\top w$. Appendix C.7 removes this convention from Proposition 2 by a vanishing perturbation of the offline rewards; the online algorithm always uses the original rewards. Define the associated offline threshold decisions by

$$x_i(w^\star) = \begin{cases} 1, & r_i > c_i^\top w^\star, \\ 0, & r_i \leq c_i^\top w^\star. \end{cases} \tag{23}$$

Lemma 4. *Denote x_i^* as the optimal solution for the projected primal problem equation 3 and $x_i(w^*)$ is defined in equation 23, then we have $x_i(w^*) \leq x_i^*$ for all $i = 1, \ldots, T$. Besides, under the general-position convention above, $x_i(w^*)$ and x_i^* differs for no more than q values of i.*

Lemma 4 tells us that the difference of $\sum_{i=1}^n r_i x_i(w^*)$ and $\mathrm{OPT}_\Phi(I)$ can be bounded by $q\overline{r}$. That is,

$$|\sum_{i=1}^{T} r_i x_i(w^*) - \mathrm{OPT}_\Phi(I)| \leq q\overline{r}$$

Then we consider a scaled version of the projected primal problem equation 3:

$$\max \sum_{i=1}^{s} r_i^\top x_i \ \text{ s.t. } \sum_{i=1}^{s} c_i x_i \leq s d_\Phi, 0 \leq \boldsymbol{x} \leq 1 \tag{24}$$

for $s = 1, .., T$. Denote the optimal value of scaled version of the projected primal problem equation 24 $\mathrm{OPT}_{\Phi,s}(I)$. The following proposition helps us relate $\mathrm{OPT}_{\Phi,s}(I)$ to $\mathrm{OPT}_\Phi(I)$:

Proposition 2. *For $s > \max\{16\overline{C}^2, e^{16\overline{C}^2}, e\}$,the following inequality holds:*

$$\frac{1}{s}\mathbb{E}[\mathrm{OPT}_{\Phi,s}(I)] \geq \frac{1}{T}\mathbb{E}[\mathrm{OPT}_\Phi(I)] - \frac{q\overline{r}}{T} - \frac{\overline{r}\log s}{\underline{D}\sqrt{s}} - \frac{q\overline{r}}{s} \tag{25}$$

for all $s \leq T \in \mathbb{N}^+$.

Now we can prove our formal result in Theorem 3. We know that

$$\begin{aligned}
\mathrm{Reg}_T(\pi_{\mathrm{MD}}) &= \mathbb{E}[\mathrm{OPT}(I) - \sum_{t=1}^{T} r_t x_t] \\
&\leq \mathrm{OPT}_\Phi(I) - \sum_{t=1}^{T} \mathbb{E}[r_t x_t] \\
&= \mathrm{OPT}_\Phi(I) - \sum_{t=1}^{T} \frac{1}{t}\mathbb{E}[\mathrm{OPT}_{\Phi,t}(I)] + \sum_{t=1}^{T} \frac{1}{t}\mathbb{E}[\mathrm{OPT}_{\Phi,t}(I)] - \sum_{t=1}^{n} \mathbb{E}[r_t x_t] \\
&= \sum_{t=1}^{T} (\frac{1}{T}\,\mathrm{OPT}_\Phi(I) - \frac{1}{t}\mathbb{E}[\mathrm{OPT}_{\Phi,t}(I)]) + \sum_{t=1}^{T} \mathbb{E}[\frac{1}{T-t+1}\tilde{R}^*_{T-t+1} - r_t x_t]
\end{aligned}$$

where x_t are specified in Algorithm 2 and $\tilde{R}^*_{T-t+1}$ is defined as the optimal value of following LP:

$$\begin{aligned}
&\max \sum_{j=t}^{T} r_j x_j \\
&\text{s.t.} \sum_{i=t}^{T} c_i x_i \leq (T-t+1) d_\Phi \\
&0 \leq \boldsymbol{x} \leq 1
\end{aligned}$$

For the first part of equation 26, we can apply Proposition 2. Meawhile, the analyses of the second part takes a similar form as the previous stochastic imput model. To be specific, we have

$$\mathbb{E}[\frac{1}{T-t+1}\tilde{R}^*_{T-t+1} - r_t x_t] \le \mathbb{E}[\langle d_\Phi - c_t x_t, w_t\rangle]\,.$$

Moreover, we know that

$$\langle d_\Phi - c_t x_t, w_t\rangle = \langle \hat{\zeta}^\rho_t, w_t\rangle - 2\rho\|w_t\|_2^2 \le \langle \hat{\zeta}^\rho_t, w_t\rangle \le \frac{1}{\gamma_t}\big(D_\psi(0\|w_t) - D_\psi(0\|w_{t+1})\big) + \frac{\gamma_t}{2\alpha}\|\hat{\zeta}^\rho_t\|_*^2.$$

Summing over $t = 1, \dots, T$, taking expectations, and using $\gamma_t = T^{-1/2}$ and $D_\psi(0\|w_1) = 0$, we obtain

$$\sum_{t=1}^{T}\mathbb{E}\left[\frac{\widetilde{R}^*_{T-t+1}}{T-t+1} - r_t x_t\right] \le \frac{1}{2\alpha\sqrt{T}}\sum_{t=1}^{T}\mathbb{E}\Big[\|\hat{\zeta}^0_t\|_*^2\Big]\,.$$

since $w_1 = 0$. Let

$$S_0 := \Big\lceil \max\{16\overline{C}^2, e^{16\overline{C}^2}, e\}\Big\rceil\,.$$

For $t \le S_0$, use

$$\frac{\mathrm{OPT}_\Phi(I)}{T} - \frac{\mathbb{E}[\mathrm{OPT}_{\Phi,t}(I)]}{t} \le \overline{r};$$

for $t > S_0$, apply Proposition 2. Combining the two bounds above with

$$\sum_{t=1}^{T}\frac{\log t}{\sqrt{t}} \le 2\sqrt{T}\log T, \qquad \sum_{t=1}^{T}\frac{1}{t} \le 1 + \log T,$$

and $\|\hat{\zeta}^0_t\|_*^2 \le Kq(\overline{C}+\overline{D})^2$, where K is as in Appendix C.4, gives

$$\begin{aligned}
\mathbb{E}\left[\mathrm{OPT}(I) - \sum_{t=1}^{T} r_t x_t\right] &\le \sum_{t=1}^{T}\left(\frac{\mathrm{OPT}_\Phi(I)}{T} - \frac{\mathbb{E}[\mathrm{OPT}_{\Phi,t}(I)]}{t}\right) + \sum_{t=1}^{T}\mathbb{E}\left[\frac{\widetilde{R}^*_{T-t+1}}{T-t+1} - r_t x_t\right] \\
&\le S_0\overline{r} + q\overline{r} + \frac{2\overline{r}}{\underline{D}}\sqrt{T}\log T + q\overline{r}(1+\log T) + \frac{1}{2\alpha\sqrt{T}}\sum_{t=1}^{T}\mathbb{E}\Big[\|\hat{\zeta}^0_t\|_*^2\Big] \\
&\le S_0\overline{r} + q\overline{r} + \frac{2\overline{r}}{\underline{D}}\sqrt{T}\log T + q\overline{r}(1+\log T) + \frac{Kq}{2\alpha}(\overline{C}+\overline{D})^2\sqrt{T}.
\end{aligned} \tag{26}$$

The bound is uniform over admissible input multisets. Taking their supremum yields

$$\mathrm{Reg}_T(\pi_{\mathrm{MD}}) \le O\Big(q(1+\log T)\sqrt{T}\Big) = \widetilde{O}(q\sqrt{T}).$$

For the constraints violation, the proof is exactly the same as it is in the stochastic input model in Appendix B.4. Our proof is thus completed.

B.6 Proof of Lemma 4

Consider the projected primal problem equation 3 and its dual. Let w denote the dual variables associated with the resource constraints, and let z_i denote the dual variables associated with the constraints $x_i \le 1$. Then the dual problem takes the form

$$\begin{aligned}
&\text{minimize} && \text{(dual objective)} \\
&\text{subject to} && w^\top a_i + z_i \ge r_i, \qquad i = 1, \dots, T, \\
&&& w \ge 0,\ z_i \ge 0, \qquad i = 1, \dots, T,
\end{aligned}$$

where r_i denotes the objective coefficient of x_i in the projected primal problem.

Let x^* be an optimal solution to equation 3, and let (w^*, z^*) be an optimal solution to its dual. By complementary slackness, for every $i = 1, \dots, T$, we have

$$x^*_i \cdot \big((w^*)^\top a_i + z^*_i - r_i\big) = 0, \qquad (1 - x^*_i)z^*_i = 0.$$

If $x_i(w^*) = 1$, then by the definition in equation 23, we have

$$r_i > (w^*)^\top a_i.$$

Together with the dual constraint

$$(w^*)^\top a_i + z_i^* \geq r_i,$$

this implies $z_i^* > 0$. Hence, by the complementary slackness condition

$$(1 - x_i^*) z_i^* = 0,$$

we obtain $x_i^* = 1$. Therefore, $x_i(w^*) \leq x_i^*$ for all $i = 1, \ldots, T$.

On the other hand, if

$$r_i < (w^*)^\top a_i,$$

then by equation 23, we have $x_i(w^*) = 0$. Moreover, since

$$(w^*)^\top a_i + z_i^* \geq (w^*)^\top a_i > r_i,$$

the corresponding dual constraint is slack. Hence, by complementary slackness, we must have $x_i^* = 0$. Therefore,

$$x_i(w^*) = x_i^*$$

whenever $(w^*)^\top a_i \neq r_i$.

Under the general-position convention above, there are at most q values of i such that

$$(w^*)^\top a_i = r_i.$$

Consequently, $x_i(w^*)$ and x_i^* differ for no more than q values of i.

B.7 Proof of Proposition 2

Fix the multiset of T projected input pairs, and take expectations with respect to its uniformly random arrival order. All dual objectives in this proof are unregularized, i.e., $\rho = 0$. We first consider rewards in general position, so that at most q equalities $r_i = c_i^\top w$ hold for any $w \geq 0$, and remove this convention at the end of the proof. Choose a full-instance optimal dual solution $w^\star$ independently of the arrival order, and write

$$y_i := x_i(w^\star) = \mathbb{I}\{r_i > c_i^\top w^\star\},$$

so ties are assigned zero. By Lemma 4,

$$\sum_{i=1}^{T} c_i y_i \leq T d_\Phi, \qquad \sum_{i=1}^{T} r_i y_i \geq \mathrm{OPT}_\Phi(I) - q\overline{r}.$$

Here and below, indices refer to the randomly ordered items.

For $d_{\Phi,0} \geq 0$, define SLP$(s, d_{\Phi,0})$ by

$$\begin{aligned} \max \quad & \sum_{i=1}^{s} r_i x_i \\ \text{s.t.} \quad & \sum_{i=1}^{s} c_i x_i \leq s(d_\Phi + d_{\Phi,0}), \\ & 0 \leq x_i \leq 1, \qquad i = 1, \ldots, s. \end{aligned} \tag{27}$$

Denote its optimal value by $\mathrm{OPT}(s, d_{\Phi,0})$; thus $\mathrm{OPT}(s, 0) = \mathrm{OPT}_{\Phi,s}(I)$. Fix an integer $s \leq T$ satisfying

$$s > s_0 := \max\{16\overline{C}^2, e^{16\overline{C}^2}, e\}, \qquad \epsilon_s := \frac{\log s}{\sqrt{s}}.$$

We establish the following two claims:

(i) The vector $(y_1, \ldots, y_s)$ is feasible for SLP$(s, \epsilon_s \mathbf{1})$ with probability at least $1 - q/s$.

(ii) For every arrival order,

$$\mathrm{OPT}_{\Phi,s}(I) \geq \mathrm{OPT}(s, \epsilon_s \mathbf{1}) - \frac{\overline{r}\sqrt{s}\log s}{\underline{D}}.$$

For the first claim, let $c_{ji} := (c_i)_j$ and $\alpha_{ji} := c_{ji} y_i$. Since $w^\star$ is fixed before the permutation, $(\alpha_{j1}, \ldots, \alpha_{js})$ is a sample without replacement from a fixed population of size T. By Assumption 1 and feasibility of $(y_1, \ldots, y_T)$,

$$\begin{aligned} c_\alpha &:= \max_{j,i} \alpha_{ji} - \min_{j,i} \alpha_{ji} \leq 2\overline{C}, \\ \bar{\alpha}_j &:= \frac{1}{T}\sum_{i=1}^{T} \alpha_{ji} \leq (d_\Phi)_j, \\ \sigma_j^2 &:= \frac{1}{T}\sum_{i=1}^{T} (\alpha_{ji} - \bar{\alpha}_j)^2 \leq 4\overline{C}^2. \end{aligned} \tag{28}$$

Bernstein's inequality for sampling without replacement (Bardenet & Maillard, 2015, Proposition 1.4) therefore gives, for every $j \in [q]$,

$$\begin{aligned} \mathbb{P}\left(\sum_{i=1}^{s} \alpha_{ji} - s(d_\Phi)_j \geq \sqrt{s}\log s \right) &\leq \mathbb{P}\left(\sum_{i=1}^{s} \alpha_{ji} - s\bar{\alpha}_j \geq \sqrt{s}\log s \right) \\ &\leq \exp\left(-\frac{s(\log s)^2}{8s\overline{C}^2 + 2\overline{C}\sqrt{s}\log s} \right) \\ &\leq \frac{1}{s}. \end{aligned} \tag{29}$$

Indeed, $s > s_0$ implies $8\overline{C}^2 \leq (\log s)/2$ and $2\overline{C}/\sqrt{s} \leq 1/2$, which justify the last inequality. Define

$$E_j := \left\{ \sum_{i=1}^{s} \alpha_{ji} - s(d_\Phi)_j < \sqrt{s}\log s \right\}, \qquad E := \bigcap_{j=1}^{q} E_j.$$

A union bound gives $\mathbb{P}(E^c) \leq q/s$. On E, the vector $(y_1, \ldots, y_s)$ is feasible for $\mathrm{SLP}(s, \epsilon_s \mathbf{1})$, proving claim (i).

For the second claim, let w_s^* and $\widetilde{w}_s^*$ be optimal dual solutions to $\mathrm{SLP}(s, 0)$ and $\mathrm{SLP}(s, \epsilon_s \mathbf{1})$, respectively. Strong duality and optimality of $\widetilde{w}_s^*$ yield

$$\begin{aligned} \mathrm{OPT}(s, \epsilon_s \mathbf{1}) &= s(d_\Phi + \epsilon_s \mathbf{1})^\top \widetilde{w}_s^* + \sum_{i=1}^{s} (r_i - c_i^\top \widetilde{w}_s^*)^+ \\ &\leq s(d_\Phi + \epsilon_s \mathbf{1})^\top w_s^* + \sum_{i=1}^{s} (r_i - c_i^\top w_s^*)^+ \\ &= \mathrm{OPT}_{\Phi,s}(I) + s\epsilon_s \|w_s^*\|_1 \\ &\leq \mathrm{OPT}_{\Phi,s}(I) + \frac{\overline{r}\sqrt{s}\log s}{\underline{D}}. \end{aligned} \tag{30}$$

The last inequality follows from the following bound.

Lemma 5. *For any $n \in \{1, \ldots, T\}$, every optimal solution w_n^* of*

$$\min_{w \geq 0} f_{n,\Phi}^0(w), \qquad f_{n,\Phi}^0(w) := d_\Phi^\top w + \frac{1}{n}\sum_{i=1}^{n} (r_i - c_i^\top w)^+,$$

satisfies

$$\|w_n^*\|_1 \leq \frac{\overline{r}}{\underline{D}}.$$

In particular, $\|w_n^*\|_2 \leq \overline{r}/\underline{D} \leq \sqrt{q}\,\overline{r}/\underline{D}$.

Proof. Since $w = 0$ is feasible, nonnegativity of w_n^* and the hinge terms gives

$$\underline{D}\|w_n^*\|_1 \le d_\Phi^\top w_n^* \le f_{n,\Phi}^0(w_n^*) \le f_{n,\Phi}^0(0) = \frac{1}{n}\sum_{i=1}^n r_i \le \overline{r}. \tag{31}$$

The Euclidean bound follows from $\|w_n^*\|_2 \le \|w_n^*\|_1$. □

Rearranging equation 30 proves claim (ii).

Finally, let $\mathbb{I}_E$ be the indicator of E, and write $Y_s := \sum_{i=1}^s r_i y_i$. Combining the two claims and dividing by s, we obtain

$$\begin{aligned}\frac{1}{s}\mathbb{E}[\mathbb{I}_E \,\mathrm{OPT}_{\Phi,s}(I)] &\ge \frac{1}{s}\mathbb{E}[\mathbb{I}_E \,\mathrm{OPT}(s, \epsilon_s \mathbf{1})] - \frac{\overline{r}\log s}{\underline{D}\sqrt{s}} \\ &\ge \frac{1}{s}\mathbb{E}[\mathbb{I}_E Y_s] - \frac{\overline{r}\log s}{\underline{D}\sqrt{s}}.\end{aligned} \tag{32}$$

Since $0 \le Y_s \le s\overline{r}$ and $\mathbb{P}(E^c) \le q/s$,

$$\frac{1}{s}\mathbb{E}[\mathbb{I}_E Y_s] \ge \frac{1}{s}\mathbb{E}[Y_s] - \frac{q\overline{r}}{s}.$$

Also, each item is included in the first s arrivals with probability s/T, so

$$\frac{1}{s}\mathbb{E}[Y_s] = \frac{1}{T}\sum_{i=1}^T r_i y_i.$$

Using $\mathrm{OPT}_{\Phi,s}(I) \ge 0$ and Lemma 4, we conclude that

$$\begin{aligned}\frac{1}{s}\mathbb{E}[\mathrm{OPT}_{\Phi,s}(I)] &\ge \frac{1}{s}\mathbb{E}[Y_s] - \frac{\overline{r}\log s}{\underline{D}\sqrt{s}} - \frac{q\overline{r}}{s} \\ &= \frac{1}{T}\sum_{i=1}^T r_i y_i - \frac{\overline{r}\log s}{\underline{D}\sqrt{s}} - \frac{q\overline{r}}{s} \\ &\ge \frac{1}{T}\,\mathrm{OPT}_\Phi(I) - \frac{q\overline{r}}{T} - \frac{\overline{r}\log s}{\underline{D}\sqrt{s}} - \frac{q\overline{r}}{s}.\end{aligned} \tag{33}$$

The full-instance value $\mathrm{OPT}_\Phi(I)$ is invariant under permutation, so this is the claimed inequality.

To remove the general-position convention, for $\eta \in (0,1)$ replace the rewards, only for this offline comparison, by $r_i^{(\eta)} = (1-\eta)r_i + \eta U_i$, where the U_i are independent and uniform on $[0,\overline{r}]$, independently of the arrival order. These rewards are in general position almost surely, remain in $[0,\overline{r}]$, and satisfy $\max_i |r_i^{(\eta)} - r_i| \le \eta\overline{r}$. For every subset of n items with the same projected budget, the corresponding optimal value changes by at most $n\eta\overline{r}$, since $0 \le x_i \le 1$. Apply equation 33 conditionally on the perturbed rewards, average over the perturbations, and let $\eta \downarrow 0$. This proves the same bound for arbitrary rewards, without altering the online algorithm, and completes the proof.

B.8 Proof of Theorem 4

Algorithm 3 comprises two phases: the Accelerate Stage and the Refine Stage. We analyze these stages separately and combine their bounds after accounting for the resources consumed during Accelerate. Let $N = T_{\mathrm{fast}} = LJ$, $H = T - N$, and let $\mathcal{F}_t$ denote the history through arrival t. All norms are Euclidean unless specified otherwise. For a nonnegative budget-rate vector b, define

$$F_b(w) := b^\top w + \mathbb{E}[(r - c^\top w)^+] + \rho\|w\|_2^2, \qquad P(b) := \min_{w \in \mathbb{R}_+^q} F_b(w). \tag{34}$$

Thus $F_{d_\Phi} = f_\Phi^\rho$ and $P(d_\Phi) = f_\Phi^\rho(w^\star)$. The quantity $P(b)$ is the regularized population-dual value. The minima in equation 34 are attained since $\rho > 0$. Write

$$\Delta_N := \sum_{t=1}^N c_t \widetilde{x}_t - N d_\Phi, \qquad d_{\mathrm{ref}} := \frac{B_{N+1}}{H} = d_\Phi - \frac{\Delta_N}{H}, \qquad \mathcal{E} := \{\|\widehat{w} - w^\star\|_2 \le T^{-1}\}.$$

In both stages, x_t^D denotes the unfiltered actual decision candidate and $\widetilde{x}_t$ denotes the executed action. The virtual candidate x_t^L is used only by the virtual trajectory. The actual decision-price update uses x_t^D, and the virtual update uses x_t^L.

For the Refine Stage, introduce the constants

$$b_0 := \frac{\underline{D}}{2}, \qquad G_{\mathrm{R},0} := \sqrt{q}\left(\overline{C} + \overline{D} + \frac{\underline{D}}{2}\right), \qquad A_{\mathrm{R}} := \frac{G_{\mathrm{R},0}^2}{2b_0} + G_{\mathrm{R},0},$$
$$R_{\mathrm{R}} := \frac{\overline{r}}{b_0} + 1 = 2R_0 + 1, \qquad G := \max\{1, \lambda, G_{\mathrm{R},0} + 2\rho R_{\mathrm{R}}\}. \tag{35}$$

The sufficient horizon conditions used below are

$$N < T, \qquad T \geq A_{\mathrm{R}}, \qquad \frac{NG_0}{T-N} \leq \frac{1}{2}\min\{\lambda, \underline{D}\}. \tag{36}$$

These conditions are imposed in addition to Assumptions 1 and 4. Also set

$$h := \max\left\{1, \frac{2G^2}{\lambda}\right\}, \qquad \kappa := h + G, \qquad \beta := \frac{\lambda}{4G^2}, \qquad M := \kappa + \frac{2G^2}{\lambda}. \tag{37}$$

The two stage-wise bounds are stated in the following lemmas.

Lemma 6. *Under the stochastic input model, Assumptions 1 and 4, and equation 36, the output of Algorithm 4 satisfies $\mathbb{P}(\mathcal{E}) \geq 1 - \delta$. Moreover, writing $R_{\mathrm{e}}^{\mathrm{act}} := \sum_{t=1}^{N} r_t \widetilde{x}_t$,*

$$\mathbb{E}\big[NP(d_\Phi) - R_{\mathrm{e}}^{\mathrm{act}} + (w^\star)^\top \Delta_N\big] \leq C_{\mathrm{e}} := \frac{R_0^2}{2\alpha_{\mathrm{e}}} + \frac{\alpha_{\mathrm{e}} N Z_\rho^2}{2} + \rho N R_0^2. \tag{38}$$

Lemma 7. *Under the same conditions, let $R_{\mathrm{R}}^{\mathrm{act}} := \sum_{t=N+1}^{T} r_t \widetilde{x}_t$. For almost every prefix history in $\mathcal{E}$,*

$$HP(d_{\mathrm{ref}}) - \mathbb{E}[R_{\mathrm{R}}^{\mathrm{act}} \mid \mathcal{F}_N] \leq C_{\mathrm{R}}, \tag{39}$$

where

$$C_{\mathrm{R}} := \frac{HG^2}{2T} + R_{\mathrm{R}}(1+M) + \overline{r}\left[2 + \frac{\kappa + 1 + 2\rho H R_{\mathrm{R}} + (1+\log 2)/\beta}{b_0}\right]. \tag{40}$$

In particular, C_{R} is bounded independently of T when the problem parameters are fixed.

We now combine the two lemmas. By Lemma 8, $w^\star$ minimizes both F_{d_Φ} and $F_{d_{\mathrm{ref}}}$, and

$$HP(d_{\mathrm{ref}}) = HP(d_\Phi) - (w^\star)^\top \Delta_N.$$

Consequently,

$$\begin{aligned} TP(d_\Phi) - R_{\mathrm{e}}^{\mathrm{act}} - R_{\mathrm{R}}^{\mathrm{act}} = NP(d_\Phi) - R_{\mathrm{e}}^{\mathrm{act}} + (w^\star)^\top \Delta_N \\ + HP(d_{\mathrm{ref}}) - R_{\mathrm{R}}^{\mathrm{act}}. \end{aligned} \tag{41}$$

The learning event $\mathcal{E}$ is $\mathcal{F}_N$-measurable. Since $0 \leq P(d_{\mathrm{ref}}) \leq F_{d_{\mathrm{ref}}}(0) \leq \overline{r}$ and the rewards are nonnegative, Lemma 7 gives

$$\begin{aligned} &\mathbb{E}[HP(d_{\mathrm{ref}}) - R_{\mathrm{R}}^{\mathrm{act}}] \\ &\quad = \mathbb{E}\big[\mathbf{1}_{\mathcal{E}}\big(HP(d_{\mathrm{ref}}) - \mathbb{E}[R_{\mathrm{R}}^{\mathrm{act}} \mid \mathcal{F}_N]\big)\big] + \mathbb{E}\big[\mathbf{1}_{\mathcal{E}^c}(HP(d_{\mathrm{ref}}) - R_{\mathrm{R}}^{\mathrm{act}})\big] \\ &\quad \leq C_{\mathrm{R}} + H\overline{r}\,\mathbb{P}(\mathcal{E}^c) \leq C_{\mathrm{R}} + H\overline{r}\delta. \end{aligned}$$

For any deterministic $w \geq 0$, nonnegativity of the basis and weak duality imply

$$\mathrm{OPT}(I) \leq \mathrm{OPT}_\Phi(I) \leq Td_\Phi^\top w + \sum_{t=1}^{T}(r_t - c_t^\top w)^+.$$

Taking expectations and choosing $w = w^\star$ yields

$$\mathbb{E}[\mathrm{OPT}(I)] \leq T\big(P(d_\Phi) - \rho\|w^\star\|_2^2\big) \leq TP(d_\Phi). \tag{42}$$

Combining equation 41, equation 42, and Lemma 6, we obtain

$$\mathbb{E}\left[\mathrm{OPT}(I) - \sum_{t=1}^{T} r_t \widetilde{x}_t\right] \leq C_{\mathrm{e}} + C_{\mathrm{R}} + H\overline{r}\delta. \tag{43}$$

To obtain the stated order, take $\delta = T^{-2}$ and then

$$Z_\rho^2 = O(q), \qquad G^2 = O(q), \qquad L = O(\log T), \qquad J = O(q \log T), \qquad N = O(q \log^2 T).$$

The choice of α_{e} gives

$$C_{\mathrm{e}} \le \frac{R_0^2}{2}(1 + A_0) + R_0 Z_\rho \sqrt{N} + \rho N R_0^2 = O(q \log T),$$

where $\rho N \le 1$ follows from $N < T$ and $\rho \le T^{-1}$. Moreover, κ, M, and β^{-1} are $O(q)$, whereas R_{R} and b_0^{-1} are independent of q, T. Since $H/T \le 1$ and $\rho H \le 1$, we have $C_{\mathrm{R}} = O(q)$. Finally, $H\overline{r}\delta \le \overline{r}/T$. Taking the supremum over the admissible input distributions in equation 43 proves

$$\mathrm{Reg}_T(\pi_{\mathrm{ATR}}) = O(q \log T).$$

For projected feasibility, the budget filter ensures $B_t \ge 0$ on every sample path. Thus

$$\sum_{t=1}^{T} c_t \widetilde{x}_t - T d_\Phi = -B_{T+1} \le 0.$$

Testing this coordinatewise inequality against any nonnegative coefficient vector in the definition of $\mathcal{U}_\Phi$ gives $v(\pi_{\mathrm{ATR}}, \mathcal{U}_\Phi) = 0$. This completes the proof of Theorem 4.

B.9 Proof of Lemma 6

We first establish the bounds on the iterates and the remaining budget that will be used in both stages.

Lemma 8. *Under equation 36, $w^\star \in \mathcal{Q}$, the virtual iterates and their epoch averages belong to $\mathcal{Q}$, and the Accelerate actual decision prices have norm at most R_{e}. Every actual decision candidate is executed during the first N arrivals. For every prefix history,*

$$\|d_{\mathrm{ref}} - d_\Phi\|_2 \le \frac{N G_0}{H}, \qquad b_0 \le (d_{\mathrm{ref}})_j \le \overline{D} + \frac{\underline{D}}{2}. \tag{44}$$

The same $w^\star$ is the unique minimizer of $F_{d_{\mathrm{ref}}}$, and

$$F_{d_{\mathrm{ref}}}(w) - F_{d_{\mathrm{ref}}}(w^\star) \ge \frac{\lambda}{2}\|w - w^\star\|_2, \qquad w \in \mathbb{R}_+^q. \tag{45}$$

All Refine prices have norm at most R_{R}. The sampled gradients have norm at most Z_ρ during Accelerate and at most G during Refine.

Proof. By optimality and nonnegativity of the terms in F_{d_Φ},

$$\underline{D}\|w^\star\|_2 \le d_\Phi^\top w^\star \le F_{d_\Phi}(w^\star) \le F_{d_\Phi}(0) \le \overline{r}.$$

Hence $w^\star \in \mathcal{Q}$. Projection and averaging preserve membership in the closed convex set $\mathcal{Q}$.

To bound the actual prices, consider a threshold update with a fixed rate b, step size $\alpha > 0$, and $\rho \ge 0$:

$$u_{t+1} = [u_t + \alpha(c_t s_t - b - 2\rho u_t)]^+, \qquad s_t = \mathbf{1}\{r_t > c_t^\top u_t\}.$$

Suppose $b_j \ge b_{\min} > 0$ and $\|c_t s_t - b\|_2 \le K_0$. The auxiliary point $v_{t+1} := [u_t + \alpha(c_t s_t - b)]^+$ satisfies $0 \le u_{t+1} \le v_{t+1}$ coordinatewise. Since $(c_t^\top u_t)s_t \le \overline{r}$,

$$\|v_{t+1}\|_2^2 \le \|u_t\|_2^2 + \alpha^2 K_0^2 + 2\alpha\overline{r} - 2\alpha b_{\min}\|u_t\|_2.$$

Thus the next norm does not increase when $\|u_t\|_2 \ge \overline{r}/b_{\min} + \alpha K_0^2/(2b_{\min})$. Otherwise, $\|u_{t+1}\|_2 \le \|u_t\|_2 + \alpha K_0$. Induction yields

$$\sup_t \|u_t\|_2 \le \max\left\{\|u_1\|_2, \frac{\overline{r}}{b_{\min}} + \alpha\left(\frac{K_0^2}{2b_{\min}} + K_0\right)\right\}. \tag{46}$$

Applying this bound with $b = d_\Phi$, $K_0 = G_0$, and $\alpha = \alpha_{\mathrm{e}}$ gives $\|w_t\|_2 \le R_0 + 1 = R_{\mathrm{e}}$, since $\alpha_{\mathrm{e}} A_0 \le 1$.

For each resource, any prefix of at most N candidate actual decisions consumes at most $N\overline{C}$. The horizon condition gives

$$N\overline{C} \le NG_0 \le H\underline{D}/2 < T\underline{D} \le T(d_\Phi)_j.$$

Consequently, no candidate is rejected during Accelerate. Also, $\|c_t\widetilde{x}_t - d_\Phi\|_2 \le G_0$ implies $\|\Delta_N\|_2 \le NG_0$, which proves equation 44. For every $w \ge 0$, Assumption 4 now gives

$$\begin{aligned} F_{d_{\mathrm{ref}}}(w) - F_{d_{\mathrm{ref}}}(w^\star) &= F_{d_\Phi}(w) - F_{d_\Phi}(w^\star) + (d_{\mathrm{ref}} - d_\Phi)^\top (w - w^\star) \\ &\ge (\lambda - \|d_{\mathrm{ref}} - d_\Phi\|_2)\,\|w - w^\star\|_2 \ge \frac{\lambda}{2}\|w - w^\star\|_2. \end{aligned}$$

This proves equation 45 and uniqueness. In particular,

$$P(d_{\mathrm{ref}}) = P(d_\Phi) + (w^\star)^\top (d_{\mathrm{ref}} - d_\Phi). \tag{47}$$

For Refine, apply equation 46 with $b_{\min} = b_0$, $K_0 = G_{\mathrm{R},0}$, $\alpha = T^{-1}$, and $\|u_1\|_2 = \|\widehat{w}\|_2 \le R_0$. The condition $T \ge A_{\mathrm{R}}$ gives $\|u_t\|_2 \le R_{\mathrm{R}}$. Finally, the gradient bounds follow from

$$\|d_\Phi - c_t x + 2\rho w\|_2 \le G_0 + 2\rho R_{\mathrm{e}} = Z_\rho, \qquad \|d_{\mathrm{ref}} - c_t x + 2\rho w\|_2 \le G_{\mathrm{R},0} + 2\rho R_{\mathrm{R}} \le G$$

on their respective iterate domains, for $x \in \{0,1\}$. □

For the learning guarantee, we use the following one-epoch estimate.

Lemma 9. *Fix epoch l and condition on its starting history. If $f^\rho_\Phi(\widetilde{w}_{l-1}) - f^\rho_\Phi(w^\star) \le \epsilon_{l-1}$, then with conditional probability at least $1 - \widehat{\delta}$,*

$$f^\rho_\Phi(\widetilde{w}_l) - f^\rho_\Phi(w^\star) \le \frac{\eta_l Z_\rho^2}{2} + \frac{\epsilon_{l-1}^2}{2\eta_l J\lambda^2} + \frac{4\epsilon_{l-1}Z_\rho}{\lambda}\sqrt{\frac{2\log(1/\widehat{\delta})}{J}}. \tag{48}$$

For the first claim of Lemma 6, define

$$\epsilon_l := \epsilon_0/2^l, \qquad \mathcal{A}_l := \{f^\rho_\Phi(\widetilde{w}_l) - f^\rho_\Phi(w^\star) \le \epsilon_l\}.$$

The event $\mathcal{A}_0$ holds since $f^\rho_\Phi(0) - f^\rho_\Phi(w^\star) \le \overline{r} = \epsilon_0$. With $\eta_l = \epsilon_{l-1}/(3Z_\rho^2)$, the three terms on the right-hand side of equation 48 are at most $\epsilon_{l-1}/6$, $\epsilon_{l-1}/6$, and $\epsilon_{l-1}/6$, respectively, by the prescribed lower bounds on J. Thus, for every starting history in $\mathcal{A}_{l-1}$,

$$\mathbb{P}(\mathcal{A}_l^c \mid \mathcal{F}_{(l-1)J}) \le \widehat{\delta}.$$

A union bound over the first unsuccessful epoch gives $\mathbb{P}(\bigcap_{l=0}^{L}\mathcal{A}_l) \ge 1 - L\widehat{\delta} = 1 - \delta$. On this event, Assumption 4 yields

$$\|\widehat{w} - w^\star\|_2 \le \frac{f^\rho_\Phi(\widetilde{w}_L) - f^\rho_\Phi(w^\star)}{\lambda} \le \frac{\epsilon_0}{\lambda 2^L} \le \frac{1}{T}.$$

This proves $\mathbb{P}(\mathcal{E}) \ge 1 - \delta$.

We next prove equation 38. By Lemma 8, $\widetilde{x}_t = x_t^D$ for $t \le N$. Set

$$h_t := d_\Phi - c_t x_t^D, \qquad \widehat{\zeta}_t^\rho := h_t + 2\rho w_t, \qquad \zeta_t^\rho := \mathbb{E}[\widehat{\zeta}_t^\rho \mid \mathcal{F}_{t-1}].$$

The threshold identity gives

$$d_\Phi^\top w_t + (r_t - c_t^\top w_t)^+ + \rho\|w_t\|_2^2 = r_t x_t^D + \langle h_t, w_t\rangle + \rho\|w_t\|_2^2.$$

Taking the conditional expectation and rearranging, we obtain

$$\begin{aligned} &\mathbb{E}\big[P(d_\Phi) - r_t x_t^D + (w^\star)^\top (c_t x_t^D - d_\Phi) \mid \mathcal{F}_{t-1}\big] \\ &\quad = P(d_\Phi) - f^\rho_\Phi(w_t) + \langle \zeta_t^\rho, w_t - w^\star\rangle - \rho\|w_t - w^\star\|_2^2 + \rho\|w^\star\|_2^2 \\ &\quad \le \langle \zeta_t^\rho, w_t - w^\star\rangle + \rho R_0^2. \end{aligned}$$

The projection inequality with comparator $w^\star$ implies

$$\langle \widehat{\zeta}_t^\rho, w_t - w^\star\rangle \le \frac{\|w_t - w^\star\|_2^2 - \|\Pi_{\mathbb{R}^q_+}(w_t - \alpha_{\mathrm{e}}\widehat{\zeta}_t^\rho) - w^\star\|_2^2}{2\alpha_{\mathrm{e}}} + \frac{\alpha_{\mathrm{e}} Z_\rho^2}{2}.$$

Let w^{e}_{N+1} denote the last exploration update, before the Refine initialization $w_{N+1} = \widehat{w}$. Taking expectations and summing the preceding inequalities gives

$$\begin{aligned}
&\mathbb{E}\big[NP(d_\Phi) - R^{\mathrm{act}}_{\mathrm{e}} + (w^\star)^\top \Delta_N\big] \\
&\quad \le \frac{\|w^\star\|_2^2 - \mathbb{E}\|w^{\mathrm{e}}_{N+1} - w^\star\|_2^2}{2\alpha_{\mathrm{e}}} + \frac{\alpha_{\mathrm{e}} N Z_\rho^2}{2} + \rho N R_0^2 \\
&\quad \le \frac{R_0^2}{2\alpha_{\mathrm{e}}} + \frac{\alpha_{\mathrm{e}} N Z_\rho^2}{2} + \rho N R_0^2.
\end{aligned}$$

This proves the second claim without conditioning on $\mathcal{E}$, and completes the proof of Lemma 6.

B.10 Proof of Lemma 9

Fix epoch l and condition on a starting history in $\mathcal{A}_{l-1}$. Sharpness and Lemma 8 imply

$$\|\widetilde{w}_{l-1} - w^\star\|_2 \le \epsilon_{l-1}/\lambda = V_l, \qquad w^\star \in \mathcal{K}_l.$$

For $t = (l-1)J + j$, define

$$\widetilde{\zeta}^{\rho}_{j,l} := d_\Phi - c_t x_t^L + 2\rho \widetilde{w}^l_j, \qquad \zeta^\rho_{j,l} := \mathbb{E}[\widetilde{\zeta}^{\rho}_{j,l} \mid \mathcal{F}_{t-1}].$$

Since $x_t^L = \mathbf{1}\{r_t > c_t^\top \widetilde{w}^l_j\}$ and the current arrival is independent of $\mathcal{F}_{t-1}$, $\zeta^\rho_{j,l} \in \partial f^\rho_\Phi(\widetilde{w}^l_j)$. This remains a valid subgradient selection at threshold ties. The scalar random variables

$$X_j := \left\langle \zeta^\rho_{j,l} - \widetilde{\zeta}^{\rho}_{j,l}, \widetilde{w}^l_j - w^\star \right\rangle$$

form a martingale difference sequence. By the gradient bound and $\widetilde{w}^l_j, w^\star \in \mathcal{K}_l$,

$$|X_j| \le 2Z_\rho \|\widetilde{w}^l_j - w^\star\|_2 \le 4 Z_\rho V_l = \frac{4\epsilon_{l-1} Z_\rho}{\lambda}.$$

The one-sided Azuma–Hoeffding inequality gives, with conditional probability at least $1 - \widehat{\delta}$,

$$\frac{1}{J}\sum_{j=1}^J X_j \le \frac{4\epsilon_{l-1} Z_\rho}{\lambda} \sqrt{\frac{2\log(1/\widehat{\delta})}{J}}. \tag{49}$$

By nonexpansiveness of projection onto $\mathcal{K}_l$,

$$\left\langle \widetilde{\zeta}^{\rho}_{j,l}, \widetilde{w}^l_j - w^\star \right\rangle \le \frac{\|\widetilde{w}^l_j - w^\star\|_2^2 - \|\widetilde{w}^l_{j+1} - w^\star\|_2^2}{2\eta_l} + \frac{\eta_l Z_\rho^2}{2}.$$

Summing over j and using $\widetilde{w}^l_1 = \widetilde{w}_{l-1}$ yields

$$\frac{1}{J}\sum_{j=1}^J \left\langle \widetilde{\zeta}^{\rho}_{j,l}, \widetilde{w}^l_j - w^\star \right\rangle \le \frac{\epsilon_{l-1}^2}{2\eta_l J \lambda^2} + \frac{\eta_l Z_\rho^2}{2}. \tag{50}$$

Finally, convexity and the definition of the epoch average give

$$\begin{aligned}
f^\rho_\Phi(\widetilde{w}_l) - f^\rho_\Phi(w^\star) &\le \frac{1}{J}\sum_{j=1}^J \big(f^\rho_\Phi(\widetilde{w}^l_j) - f^\rho_\Phi(w^\star)\big) \\
&\le \frac{1}{J}\sum_{j=1}^J \langle \zeta^\rho_{j,l}, \widetilde{w}^l_j - w^\star \rangle.
\end{aligned}$$

Combining this display with equation 49 and equation 50 proves equation 48.

B.11 Proof of Lemma 7

Condition on a prefix history in $\mathcal{E}$. The vector d_{ref} and the initialization $\widehat{w}$ are now fixed, while the future arrivals remain i.i.d. with the original law. The actual reward loss can be decomposed as

$$\begin{aligned} & HP(d_{\mathrm{ref}}) - \mathbb{E}\left[\sum_{t=N+1}^{T} r_t \widetilde{x}_t \,\middle|\, \mathcal{F}_N\right] \\ &= \underbrace{HP(d_{\mathrm{ref}}) - \mathbb{E}\left[\sum_{t=N+1}^{T} r_t x_t^D \,\middle|\, \mathcal{F}_N\right]}_{\textbf{BOUND I}} + \underbrace{\mathbb{E}\left[\sum_{t=N+1}^{T} r_t (x_t^D - \widetilde{x}_t) \,\middle|\, \mathcal{F}_N\right]}_{\textbf{BOUND II}}. \end{aligned} \tag{51}$$

For **BOUND I**, put $\widehat{\zeta}_t^{\rho} = d_{\mathrm{ref}} - c_t x_t^D + 2\rho w_t$. By the threshold identity and $P(d_{\mathrm{ref}}) \le F_{d_{\mathrm{ref}}}(w_t)$,

$$\begin{aligned} & P(d_{\mathrm{ref}}) - \mathbb{E}[r_t x_t^D \mid \mathcal{F}_{t-1}] \\ & \quad \le \mathbb{E}[\langle d_{\mathrm{ref}} - c_t x_t^D, w_t \rangle \mid \mathcal{F}_{t-1}] + \rho \|w_t\|_2^2 \\ & \quad = \mathbb{E}[\langle \widehat{\zeta}_t^{\rho}, w_t \rangle \mid \mathcal{F}_{t-1}] - \rho \|w_t\|_2^2. \end{aligned}$$

The projection inequality with comparator zero and step size $1/T$ gives

$$\langle \widehat{\zeta}_t^{\rho}, w_t \rangle \le \frac{T}{2}(\|w_t\|_2^2 - \|w_{t+1}\|_2^2) + \frac{G^2}{2T}.$$

Taking conditional expectations, summing, and dropping the nonpositive regularization terms yields

$$\textbf{BOUND I} \le \frac{HG^2}{2T} + \frac{T}{2}\left(\|\widehat{w}\|_2^2 - \mathbb{E}[\|w_{T+1}\|_2^2 \mid \mathcal{F}_N]\right). \tag{52}$$

Rewrite the difference of squares as

$$\|\widehat{w}\|_2^2 - \|w_{T+1}\|_2^2 = \langle \widehat{w} + w_{T+1}, \widehat{w} - w_{T+1} \rangle.$$

By Lemma 8, its absolute value is bounded by

$$2R_{\mathrm{R}}\left(\|\widehat{w} - w^\star\|_2 + \|w_{T+1} - w^\star\|_2\right).$$

The initialization error is at most $1/T$ on $\mathcal{E}$. To control the subsequent iterates, we use the following lemma.

Lemma 10. *For almost every prefix history in $\mathcal{E}$ and each $t = N+1, \ldots, T+1$, define $Y_t := T\|w_t - w^\star\|_2$. Then*

$$\mathbb{E}[Y_t \mid \mathcal{F}_N] \le M, \qquad \mathbb{E}[e^{\beta(Y_t - \kappa)} \mid \mathcal{F}_N] \le 2. \tag{53}$$

In particular, $\mathbb{E}[\|w_{T+1} - w^\star\|_2 \mid \mathcal{F}_N] \le M/T$ on $\mathcal{E}$, and unconditionally,

$$\mathbb{E}\|w_{T+1} - w^\star\|_2 \le \frac{M}{T} + (R_{\mathrm{R}} + R_0)\delta.$$

Combining equation 52 and Lemma 10, we obtain

$$\textbf{BOUND I} \le \frac{HG^2}{2T} + R_{\mathrm{R}}(1 + M).$$

For **BOUND II**, let

$$S_t := \sum_{i=N+1}^{t} c_i x_i^D, \qquad I_t := \mathbf{1}\{x_t^D = 1,\ \widetilde{x}_t = 0\}.$$

Since $[z]^+ \ge z$ coordinatewise, the Refine update gives

$$c_t x_t^D - d_{\mathrm{ref}} \le T(w_{t+1} - w_t) + 2\rho w_t.$$

Summing from $N+1$ to t, we have

$$S_t - (t-N)d_{\mathrm{ref}} \le T(w_{t+1} - \widehat{w}) + 2\rho \sum_{i=N+1}^{t} w_i. \tag{54}$$

If $I_t = 1$, some coordinate j satisfies

$$\sum_{i=N+1}^{t-1} (c_i)_j \widetilde{x}_i + (c_t)_j > H(d_{\mathrm{ref}})_j.$$

All resource coefficients are nonnegative and $x_i^D \ge \widetilde{x}_i$, so $(S_t)_j > H(d_{\mathrm{ref}})_j$. Combining this with equation 54 yields

$$I_t = 1 \quad \Longrightarrow \quad (T-t)b_0 < Y_{t+1} + a_0, \qquad a_0 := 1 + 2\rho H R_{\mathrm{R}}. \tag{55}$$

Here we used $T\|\widehat{w} - w^\star\|_2 \le 1$ and $\|w_i\|_2 \le R_{\mathrm{R}}$.

By Lemma 10 and Markov's inequality,

$$\mathbb{P}(Y_{t+1} > \kappa + u \mid \mathcal{F}_N) \le 2e^{-\beta u}, \qquad u \ge 0.$$

Define

$$n_0 := \left\lceil \frac{a_0 + \kappa + \beta^{-1}\log 2}{b_0} \right\rceil.$$

For $n = T - t \ge n_0$, equation 55 implies

$$\mathbb{P}(I_t = 1 \mid \mathcal{F}_N) \le e^{-\beta b_0 (n - n_0)}.$$

For $n < n_0$, this probability is at most one. Hence

$$\begin{aligned} \mathbb{E}\left[\sum_{t=N+1}^{T} I_t \,\middle|\, \mathcal{F}_N \right] &\le n_0 + \sum_{m=0}^{\infty} e^{-\beta b_0 m} \\ &\le 2 + \frac{a_0 + \kappa + (1+\log 2)/\beta}{b_0}, \end{aligned} \tag{56}$$

where we used $\lceil x \rceil \le x + 1$ and $(1 - e^{-x})^{-1} \le 1 + x^{-1}$ for $x > 0$. Each rejection loses at most $\overline{r}$, and therefore

$$\textbf{BOUND II} \le \overline{r} \left[2 + \frac{\kappa + 1 + 2\rho H R_{\mathrm{R}} + (1+\log 2)/\beta}{b_0} \right].$$

Combining the two bounds in equation 51 proves Lemma 7. Conditional weak duality also bounds the expected offline optimum for the remaining H arrivals with budget B_{N+1} by $HP(d_{\mathrm{ref}})$. Since $H/T \le 1$ and $\rho H \le 1$, the resulting bound is $O(1)$ in T for fixed problem parameters.

B.12 Proof of Lemma 10

Condition on a prefix history in $\mathcal{E}$. For $t = N+1, \dots, T$, write the Refine update as

$$w_{t+1} = \Pi_{\mathbb{R}^q_+}\left(w_t - \frac{1}{T}\widehat{\zeta}_t^{\rho} \right), \qquad \widehat{\zeta}_t^{\rho} = d_{\mathrm{ref}} - c_t x_t^D + 2\rho w_t.$$

Independence of the current arrival gives

$$\mathbb{E}[\widehat{\zeta}_t^{\rho} \mid \mathcal{F}_{t-1}] \in \partial F_{d_{\mathrm{ref}}}(w_t).$$

Since $w^\star \in \mathbb{R}^q_+$, nonexpansiveness of projection implies

$$\|w_{t+1} - w^\star\|_2^2 \le \|w_t - w^\star\|_2^2 + \frac{\|\widehat{\zeta}_t^{\rho}\|_2^2}{T^2} - \frac{2}{T}\langle w_t - w^\star, \widehat{\zeta}_t^{\rho} \rangle.$$

The conditional subgradient inequality and equation 45 give

$$\left\langle w_t - w^\star, \mathbb{E}[\widehat{\zeta}_t^{\rho} \mid \mathcal{F}_{t-1}] \right\rangle \ge F_{d_{\mathrm{ref}}}(w_t) - F_{d_{\mathrm{ref}}}(w^\star) \ge \frac{\lambda}{2}\|w_t - w^\star\|_2.$$

Taking conditional expectations and using the gradient bound, we obtain

$$\begin{aligned}&\mathbb{E}[\|w_{t+1}-w^\star\|_2^2 \mid \mathcal{F}_{t-1}]\\ &\quad\le \|w_t-w^\star\|_2^2+\frac{G^2}{T^2}-\frac{\lambda}{T}\|w_t-w^\star\|_2.\end{aligned} \tag{57}$$

Moreover,

$$\|w_{t+1}-w_t\|_2 \le \frac{1}{T}\|\widehat{\zeta}_t^{\rho}\|_2 \le \frac{G}{T}.$$

For $Y_t = T\|w_t - w^\star\|_2$, these two inequalities become

$$\mathbb{E}[Y_{t+1}^2 \mid \mathcal{F}_{t-1}] \le Y_t^2 + G^2 - \lambda Y_t, \qquad |Y_{t+1}-Y_t| \le G.$$

Whenever $Y_t \ge h \ge 2G^2/\lambda$,

$$Y_t^2 + G^2 - \lambda Y_t \le (Y_t - \lambda/4)^2.$$

Since $G \ge \lambda$, the quantity $Y_t - \lambda/4$ is nonnegative on this event. Jensen's inequality therefore yields

$$\mathbb{E}[Y_{t+1} - Y_t \mid \mathcal{F}_{t-1}] \le -\lambda/4 \qquad \text{whenever } Y_t \ge h.$$

The initialization satisfies $Y_{N+1} \le 1 \le h$ on $\mathcal{E}$. Apply Lemma 12 under the conditional law given $\mathcal{F}_N$, with increment bound $K = G$ and drift $a = \lambda/4$. The filtration for this application is $\mathcal{H}_n = \mathcal{F}_{N+n}$ and its adapted process is Y_{N+1+n}. The constants in equation 37 give

$$\mathbb{E}[Y_t \mid \mathcal{F}_N] \le M, \qquad \mathbb{E}[e^{\beta(Y_t-\kappa)} \mid \mathcal{F}_N] \le 2$$

for every $t = N+1, \dots, T+1$. This proves equation 53. Finally, integrating over successful prefix histories and using $\|w_{T+1} - w^\star\|_2 \le R_{\mathrm{R}} + R_0$ on $\mathcal{E}^c$ gives

$$\mathbb{E}\|w_{T+1}-w^\star\|_2 \le \frac{M}{T} + (R_{\mathrm{R}}+R_0)\mathbb{P}(\mathcal{E}^c) \le \frac{M}{T} + (R_{\mathrm{R}}+R_0)\delta.$$

This completes the proof.

B.13 Proof of Lemma 2

Since $\rho > 0$, the mapping $u \mapsto \rho\|u\|_{L^2}^2$ is 2ρ-strongly convex on $L^2(\mathcal{S})$. The remaining terms in F_T^ρ are convex and continuous. Hence F_T^ρ is strongly convex, proper, lower semi-continuous, and coercive on the closed convex cone $L_+^2(\mathcal{S})$. Therefore, equation 8 admits a unique minimizer $u_{T,\rho}^*$.

For each t, choose

$$\alpha_t \in \partial\big(r_t - \langle a_t, u_{T,\rho}^*\rangle\big)^+,$$

so that $\alpha_t \in [0,1]$, and more explicitly,

$$\alpha_t = \begin{cases} 1, & r_t > \langle a_t, u_{T,\rho}^*\rangle, \\ 0, & r_t < \langle a_t, u_{T,\rho}^*\rangle, \\ \text{an arbitrary value in } [0,1], & r_t = \langle a_t, u_{T,\rho}^*\rangle. \end{cases}$$

The first-order optimality condition for minimizing F_T^ρ over $L_+^2(\mathcal{S})$ is

$$0 \in d - \frac{1}{T}\sum_{t=1}^T \alpha_t a_t + 2\rho u_{T,\rho}^* + N_{L_+^2(\mathcal{S})}(u_{T,\rho}^*),$$

where $N_{L_+^2(\mathcal{S})}(u_{T,\rho}^*)$ denotes the normal cone of $L_+^2(\mathcal{S})$ at $u_{T,\rho}^*$.

Equivalently, $u_{T,\rho}^*$ is the projection of $\frac{1}{2\rho}\big(\frac{1}{T}\sum_{t=1}^T \alpha_t a_t - d\big)$ onto the nonnegative cone $L_+^2(\mathcal{S})$. Since the projection onto $L_+^2(\mathcal{S})$ is given point-wise by the positive-part operator, we obtain

$$u_{T,\rho}^*(s) = \frac{1}{2\rho}\left[\frac{1}{T}\sum_{t=1}^T \alpha_t a_t(s) - d(s)\right]^+, \qquad s \in \mathcal{S}.$$

This proves equation 9. Because each $a_t(\cdot)$ and $d(\cdot)$ is continuous on the compact set $\mathcal{S}$, the function inside the bracket is continuous, and so is its positive part. Hence $u_{T,\rho}^* \in C(\mathcal{S})$.

B.14 Proof of Proposition 1

Fix an arbitrary realization of the input sequence and define the residual function

$$g_\pi := A\boldsymbol{x}(\pi) - \boldsymbol{b}.$$

By continuity of a_t and d, $g_\pi(\cdot)$ is continuous on the compact set $\mathcal{S}$. Hence

$$\sup_{u\in\mathcal{U}_0} \langle u, g_\pi\rangle = \sup_{s\in\mathcal{S}} g_\pi(s).$$

Moreover, by approximating a Dirac mass at a maximizer of g_π with continuous nonnegative densities, we also have

$$\sup_{u\in\mathcal{U}_0} \langle u, g_\pi\rangle = \sup_{u\in\mathcal{U}_0^c} \langle u, g_\pi\rangle.$$

Let $\delta > 0$. Choose $u^\delta \in \mathcal{U}_0^c$ such that

$$\langle u^\delta, g_\pi\rangle \geq \sup_{u\in\mathcal{U}_0} \langle u, g_\pi\rangle - \delta.$$

By Assumption 2, there exists $u_\Phi^\delta \in \mathcal{U}_\Phi$ satisfying

$$\sup_{a\in\mathcal{A}} \left|\langle u^\delta - u_\Phi^\delta, a\rangle\right| \leq \epsilon_{\text{test}}, \quad \left|\langle u^\delta - u_\Phi^\delta, d\rangle\right| \leq \epsilon_{\text{test}}.$$

Therefore,

$$\langle u^\delta, g_\pi\rangle = \langle u_\Phi^\delta, g_\pi\rangle + \langle u^\delta - u_\Phi^\delta, g_\pi\rangle \leq \max_{u_\Phi\in\mathcal{U}_\Phi} \langle u_\Phi, g_\pi\rangle + \epsilon_{\text{test}} \left(\sum_{t=1}^T x_t + T\right),$$

which implies

$$\sup_{u\in\mathcal{U}_0} \langle u, g_\pi\rangle \leq \max_{u_\Phi\in\mathcal{U}_\Phi} \langle u_\Phi, g_\pi\rangle + \epsilon_{\text{test}} \left(\sum_{t=1}^T x_t + T\right) + \delta.$$

Letting $\delta \downarrow 0$, we obtain

$$\sup_{u\in\mathcal{U}_0} \langle u, g_\pi\rangle \leq \max_{u_\Phi\in\mathcal{U}_\Phi} \langle u_\Phi, g_\pi\rangle + \epsilon_{\text{test}} \left(\sum_{t=1}^T x_t + T\right).$$

Taking positive parts and using $(a+b)^+ \leq a^+ + b$ for $b \geq 0$, we have

$$\left(\sup_{u\in\mathcal{U}_0} \langle u, g_\pi\rangle\right)^+ \leq \left(\max_{u_\Phi\in\mathcal{U}_\Phi} \langle u_\Phi, g_\pi\rangle\right)^+ + \epsilon_{\text{test}} \left(\sum_{t=1}^T x_t + T\right). \tag{58}$$

Finally, since $0 \leq x_t \leq 1$, we can obtain

$$\sum_{t=1}^T x_t + T \leq 2T.$$

Taking expectations in equation 58 therefore yields

$$v(\pi) \leq v(\pi, \mathcal{U}_\Phi) + 2T\epsilon_{\text{test}},$$

which completes the proof.

C Auxiliary Lemmas

Lemma 11 (Azuma–Hoeffding inequality)**.** *Let $(X_k, \mathcal{F}_k)_{k=1}^n$ be a martingale difference sequence with $|X_k| \leq c_k$ almost surely. Then, for every $t > 0$,*

$$\Pr\left(\left|\sum_{k=1}^n X_k\right| \geq t\right) \leq 2\exp\left(-\frac{t^2}{2\sum_{k=1}^n c_k^2}\right).$$

Lemma 12 (Negative drift)**.** *Let $(Y_n, \mathcal{H}_n)_{n\geq 0}$ be a nonnegative adapted process. Suppose that $K > 0$, $h \geq 0$, $0 < a < K$, $Y_0 \leq h$ almost surely, and*

$$|Y_{n+1} - Y_n| \leq K, \qquad \mathbb{E}[Y_{n+1} - Y_n \mid \mathcal{H}_n] \leq -a \quad \text{on } \{Y_n \geq h\}.$$

Set $b = h + K$ and $\theta = a/K^2$. Then, for every $n \geq 0$,

$$\mathbb{E}[Y_n] \leq b + \frac{K^2}{2a}, \qquad \mathbb{E}[e^{\theta(Y_n - b)}] \leq 2. \tag{59}$$

The same bounds hold conditionally on an initial sigma-field whenever the hypotheses hold under that conditional law with the stated constants.

Proof. Let

$$p := \frac{1 - a/K}{2}, \qquad \nu := \frac{p}{1-p} = \frac{K-a}{K+a}.$$

Consider the reflected random walk

$$\Gamma_0 = b, \qquad \Gamma_{n+1} = \max\{b, \Gamma_n + K\xi_{n+1}\},$$

where the signs ξ_n are i.i.d. with $\mathbb{P}(\xi_n = 1) = p$ and $\mathbb{P}(\xi_n = -1) = 1 - p$. Thus $\mathbb{E}[K\xi_n] = -a$.

We compare (Y_n) and (Γ_n) using increasing convex test functions. For such a function $\varphi : \mathbb{R} \to \mathbb{R}$, define

$$(\mathcal{M}\varphi)(y) := p\varphi(\max\{b, y + K\}) + (1-p)\varphi(\max\{b, y - K\}).$$

The function $\mathcal{M}\varphi$ is also increasing and convex. If $Y_n < h$, then $Y_{n+1} < h + K = b$, and hence

$$\mathbb{E}[\varphi(Y_{n+1}) \mid \mathcal{H}_n] \leq \varphi(b) \leq (\mathcal{M}\varphi)(Y_n).$$

If $Y_n \geq h$, write $D_{n+1} = Y_{n+1} - Y_n$. For any $z \in [-K, K]$, convexity gives

$$\varphi(Y_n + z) \leq \frac{K - z}{2K}\varphi(Y_n - K) + \frac{K + z}{2K}\varphi(Y_n + K).$$

Taking conditional expectations, using $\mathbb{E}[D_{n+1} \mid \mathcal{H}_n] \leq -a$, and using monotonicity of φ, we obtain

$$\begin{aligned}\mathbb{E}[\varphi(Y_{n+1}) \mid \mathcal{H}_n] &\leq p\varphi(Y_n + K) + (1-p)\varphi(Y_n - K) \\ &\leq (\mathcal{M}\varphi)(Y_n).\end{aligned}$$

Since $Y_0 \leq b = \Gamma_0$ and $\mathcal{M}$ preserves increasing convexity, induction shows

$$\mathbb{E}[\varphi(Y_n)] \leq \mathbb{E}[\varphi(\Gamma_n)]$$

for every increasing convex φ for which these expectations are finite.

The chain $(\Gamma_n - b)/K$ has stationary distribution

$$\mathbb{P}(Z_\infty = j) = (1-\nu)\nu^j, \qquad j \geq 0.$$

Indeed, the detailed-balance equations are $(1-\nu)\nu^j p = (1-\nu)\nu^{j+1}(1-p)$. Coupling a chain started at zero and a stationary chain with the same signs preserves their order, because the transition is monotone. It follows that

$$\mathbb{E}[Y_n] \leq b + K\mathbb{E}[Z_\infty] = b + \frac{K\nu}{1-\nu} = b + \frac{K(K-a)}{2a} \leq b + \frac{K^2}{2a}.$$

For the exponential bound, set $s = a/K \in (0, 1)$. Since $\nu = (1-s)/(1+s) \leq e^{-2s}$, the geometric series converges and

$$\begin{aligned}\mathbb{E}[e^{\theta(Y_n - b)}] \leq \mathbb{E}[e^{sZ_\infty}] &= \frac{1-\nu}{1-\nu e^s} \\ &\leq \frac{1 - e^{-2s}}{1 - e^{-s}} = 1 + e^{-s} \leq 2.\end{aligned}$$

Here $(1-v)/(1-ve^s)$ is increasing in v for $0 \leq v < e^{-s}$. The argument also applies under the conditional law given an initial history, which proves the conditional statement and completes the proof. □

D EPOCH-BASED PROJECTED SGD ALGORITHM

In this section, we present the full epoch-based projected SGD algorithm used in the Accelerate Stage of Algorithm 3. The procedure is given in Algorithm 4. Fix the basis Φ before observing the arrivals, and let $T \geq 3$, $\delta \in (0,1)$, and $0 < \rho \leq T^{-1}$. The boundedness parameters and a valid constant $\lambda > 0$ in Assumption 4 are assumed known.

Let $R_0 := \overline{r}/\underline{D}$ and $\mathcal{Q} := \{w \in \mathbb{R}^q_+ : \|w\|_2 \leq R_0\}$, which contains $w^\star$. Define $R_{\mathrm{e}} := R_0 + 1$ and $G_0 := \sqrt{q}\,(\overline{C} + \overline{D})$. The sampled gradients during Accelerate are bounded by $Z_\rho := G_0 + 2\rho R_{\mathrm{e}}$, as shown in Appendix B.9.

Set $\epsilon_0 := \overline{r}$, so that $f^\rho_\Phi(0) - f^\rho_\Phi(w^\star) \leq \epsilon_0$. The number of epochs is

$$L = \max\left\{1, \left\lceil \log_2\left(\frac{\epsilon_0 T}{\lambda}\right)\right\rceil\right\}. \tag{60a}$$

Each epoch is allocated a failure probability of $\widehat{\delta} := \delta/L$ and contains J projected-gradient steps, where

$$J = \left\lceil \max\left\{1, \frac{9Z_\rho^2}{\lambda^2}, \frac{1152 Z_\rho^2}{\lambda^2}\log\frac{1}{\widehat{\delta}}\right\}\right\rceil. \tag{60b}$$

Thus the total number of virtual samples is $N = T_{\mathrm{fast}} = LJ$, which must satisfy $N < T$. We use natural logarithms except where $\log_2$ is indicated.

At epoch l, set

$$\epsilon_l = \frac{\epsilon_0}{2^l}, \qquad \eta_l = \frac{\epsilon_{l-1}}{3Z_\rho^2}, \qquad V_l = \frac{\epsilon_{l-1}}{\lambda}. \tag{60c}$$

The virtual iterates are projected onto $\mathcal{K}_l = \mathcal{Q} \cap \mathcal{B}(\widetilde{w}_{l-1}, V_l)$, where $\mathcal{B}$ denotes a closed Euclidean ball. The epoch starts at $\widetilde{w}^l_1 = \widetilde{w}_{l-1}$ and returns the average $\widetilde{w}_l = J^{-1}\sum_{j=1}^J \widetilde{w}^l_j$. Both the step size and the localization radius are halved between epochs, yielding the final estimate $\widehat{w} = \widetilde{w}_L$. The actual price trajectory instead uses projection onto $\mathbb{R}^q_+$ and the constant step size

$$\alpha_{\mathrm{e}} = \min\left\{1, \frac{1}{A_0}, \frac{R_0}{Z_\rho\sqrt{N}}\right\}, \qquad A_0 := \frac{G_0^2}{2\underline{D}} + G_0. \tag{60d}$$

At arrival $t = (l-1)J + j$, the two trajectories share the same sample (r_t, c_t) but compute their own threshold candidates x^L_t and x^D_t. The virtual update uses x^L_t, while the actual decision-price update uses x^D_t. Only the actual decision candidate can be executed: $\widetilde{x}_t = x^D_t \mathbf{1}\{c_t x^D_t \leq B_t\}$. The virtual candidate consumes no resources, and neither price update uses the budget-filtered action $\widetilde{x}_t$. All budget comparisons are coordinatewise.

Under the stochastic input model and Assumptions 1 and 4, the conditional mean of the learning gradient belongs to $\partial f^\rho_\Phi(\widetilde{w}^l_j)$. Conditioned on an epoch-starting history satisfying $f^\rho_\Phi(\widetilde{w}_{l-1}) - f^\rho_\Phi(w^\star) \leq \epsilon_{l-1}$, the linear error bound ensures $w^\star \in \mathcal{K}_l$. The centered gradient, paired with $\widetilde{w}^l_j - w^\star$, forms scalar martingale differences bounded in absolute value by $4Z_\rho V_l$. The choice of J controls the martingale deviation and the projected-SGD error, giving $f^\rho_\Phi(\widetilde{w}_l) - f^\rho_\Phi(w^\star) \leq \epsilon_l$ with conditional probability at least $1 - \widehat{\delta}$. Induction and a union bound over the first unsuccessful epoch therefore give

$$\|\widehat{w} - w^\star\|_2 \leq \frac{\epsilon_0}{\lambda 2^L} \leq \frac{1}{T}$$

with probability at least $1 - \delta$. The detailed proof is given in Appendix B.9 and the proof of Lemma 9.

Algorithm 4 returns the executed decisions, the estimate $\widehat{w}$, and the remaining budget $B_{T_{\mathrm{fast}}+1}$. Algorithm 3 then initializes Refine at $\widehat{w}$ with the fixed budget rate $d_{\mathrm{ref}} = B_{T_{\mathrm{fast}}+1}/(T - T_{\mathrm{fast}})$, using the same ρ. Combining the two stages as in Appendix B.8 yields the $O(q \log T)$ expected regret and zero projected constraint violation stated in Theorem 4.

Algorithm 4 Epoch-based Projected SGD

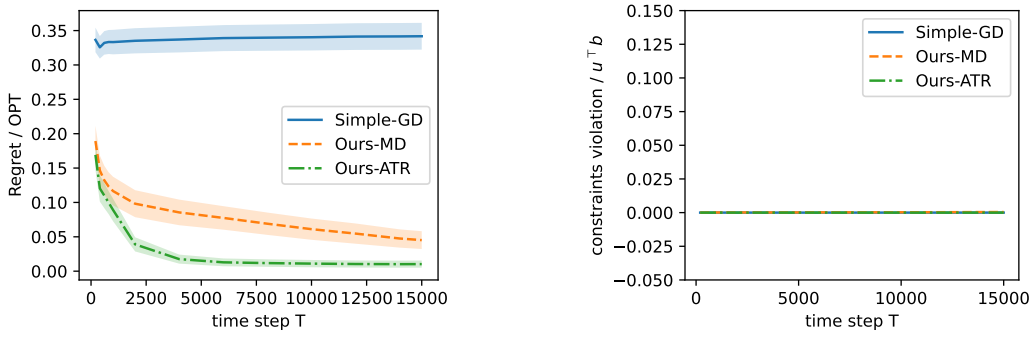
Input: projected budget rate d_Φ, fixed basis Φ, horizon $T \geq 3$, failure probability $\delta \in (0,1)$, regularization $\rho \in (0, 1/T]$, initial budget $B_1 = Td_\Phi$, known boundedness parameters, and a valid constant $\lambda > 0$.
Set the parameters in equation 60, with $N = LJ < T$.
Initialize $w_1 = 0$ and $\widetilde{w}_0 = 0$.
for $l = 1, \ldots, L$ **do**
  Set $\widetilde{w}_1^l = \widetilde{w}_{l-1}$ and $\mathcal{K}_l = \mathcal{Q} \cap \mathcal{B}(\widetilde{w}_{l-1}, V_l)$.
  for $j = 1, \ldots, J$ **do**
    Set $t = (l-1)J + j$.
    Observe (r_t, a_t) and compute c_t.
    Set $x_t^L = \mathbf{1}\{r_t > c_t^\top \widetilde{w}_j^l\}$ and $x_t^D = \mathbf{1}\{r_t > c_t^\top w_t\}$.
    Execute $\widetilde{x}_t = x_t^D \mathbf{1}\{c_t x_t^D \leq B_t\}$.
    Update $B_{t+1} = B_t - c_t \widetilde{x}_t$.
    Update $w_{t+1} = \Pi_{\mathbb{R}^q_+}[w_t - \alpha_e(d_\Phi - c_t x_t^D + 2\rho w_t)]$.
    Update $\widetilde{w}_{j+1}^l = \Pi_{\mathcal{K}_l}[\widetilde{w}_j^l - \eta_l(d_\Phi - c_t x_t^L + 2\rho \widetilde{w}_j^l)]$.
  end for
  Set $\widetilde{w}_l = \frac{1}{J}\sum_{j=1}^J \widetilde{w}_j^l$.
  Set $\eta_{l+1} = \eta_l/2$ and $V_{l+1} = V_l/2$.
end for
Output: $(\widetilde{x}_1, \ldots, \widetilde{x}_N)$, $\widehat{w} = \widetilde{w}_L$, T_{fast}, and $B_{T_{\text{fast}}+1}$.

E Additional Experiments

While the main text reports results under bounded uniform arrivals, this section provides additional experiments assessing the empirical robustness of our approach. Specifically, we evaluate three methods' performance under stochastic arrivals with Gaussian and Cauchy distribution as unbounded-input stress tests, analyze scaling behavior with respect to the number of constraints across all three distributions, and report results under random-permutation arrivals.

Performance under stochastic arrivals Figure 2 shows the performance of the three algorithms under the Gaussian distribution. We observe that, for both Ours-MD and Ours-ATR, the ratio of regret to the optimal value converges to 0, with Ours-ATR converging more rapidly, showing the same qualitative trend as in the bounded uniform setting. In contrast, Simple-GD performs poorly; its curve in Figure 2a suggests an $O(T)$ regret. This behavior stems from its $O(\sqrt{T})$ regret bound combined with a linear dependence on the number of constraints, which is large in our setting and dominates the overall regret. Moreover, Figure 2b indicates that all methods achieve quite small normalized projected violations. Similar trends are also observed in Figure 3, further demonstrating the effectiveness and robustness of the proposed methods.

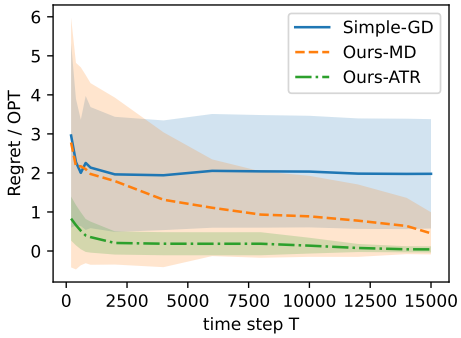
Figure 2: Gaussian Distribution with $M = 2000$



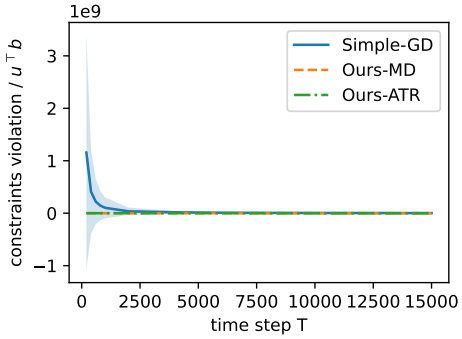
(a) Regret

(b) Constraints violation

Figure 3: Cauchy Distribution with $M = 2000$



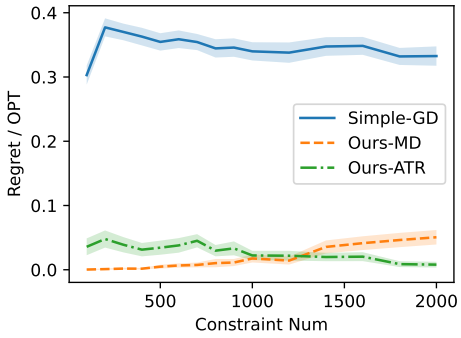
(a) Regret



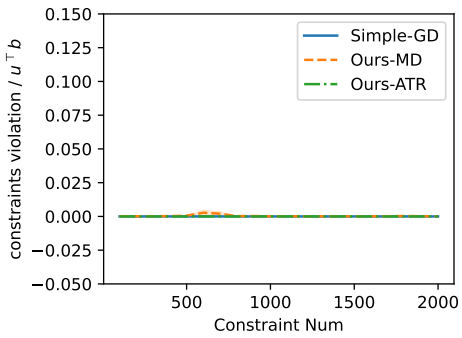
(b) Constraints violation

Scaling with the number of constraints We conduct experiments under fixed time horizon $T = 5000$ and observe both regret and constraint violation under different number of constraints. We sample input data $\{(r_t, a_t)\}_{t=1}^{T}$ in the same way of previous fixed number of constraints experiments. We can observe that for the uniform and normal distribution experiment, even when the number of constraints is relative small (less than 500), both Ours-MD and Ours-ATR enjoys good performance, whereas Simple-GD already has poor performance. When the number of constraints become larger, Ours-ATR is quite robust with stable performance. In addition, Simple-GD is prone to severe performance degradation when faced with extreme data as shown in Figure 6.

Figure 4: Uniform Distribution with $T = 5000$



(a) Regret



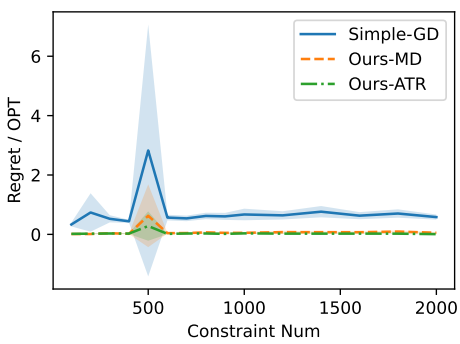
(b) Constraints violation

Random permutation arrivals Figure 7a shows that both Ours-MD and Simple-GD have $O(\sqrt{T})$ convergence rate, while our Ours-MD converges quite faster and can converge to 0. Figure 7b further shows that both two algorithms have excellent performance of projected constraints violation. The constraints violation of Simple-GD is 0 and Ours-MD violates the constraints very small. We also conduct experiments with fixed time horizon T and various number of constraints. Figure 8 shows that Ours-MD can maintain good performance when the number of constraints is not that large, while Simple-GD's performance deteriorates steadily as the number of constraints increases.

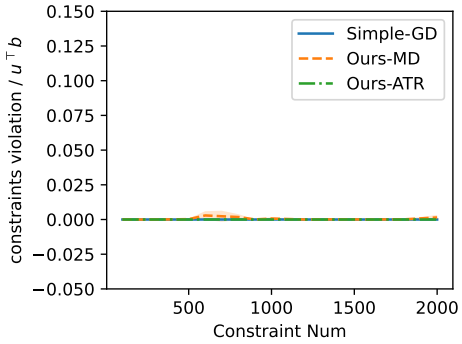
E.1 Experiments compute resources

The experimental environment is based on the Windows 11 Home Edition (Version 26200) operating system. The specific hardware platform is a MECHREVO imini Pro, equipped with an AMD Ryzen 7 8845H processor (8 cores, 3.81 GHz) and 24 GB of RAM. This processor integrates an AMD Radeon 780M graphics card for computational acceleration.

Figure 5: Normal Distribution with $T = 5000$

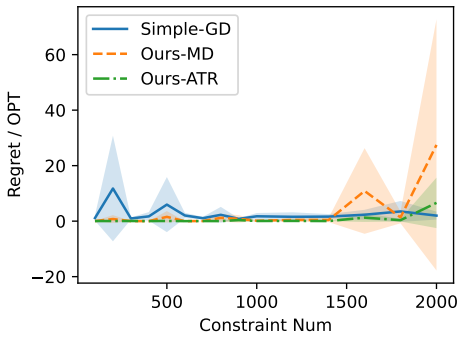


(a) Regret

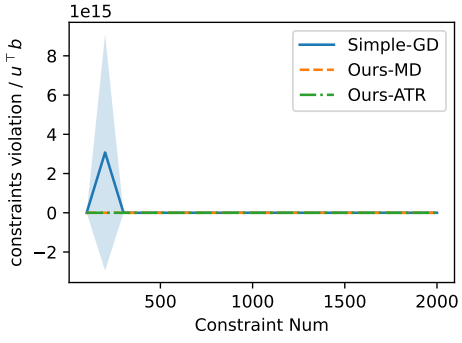


(b) Constraints violation

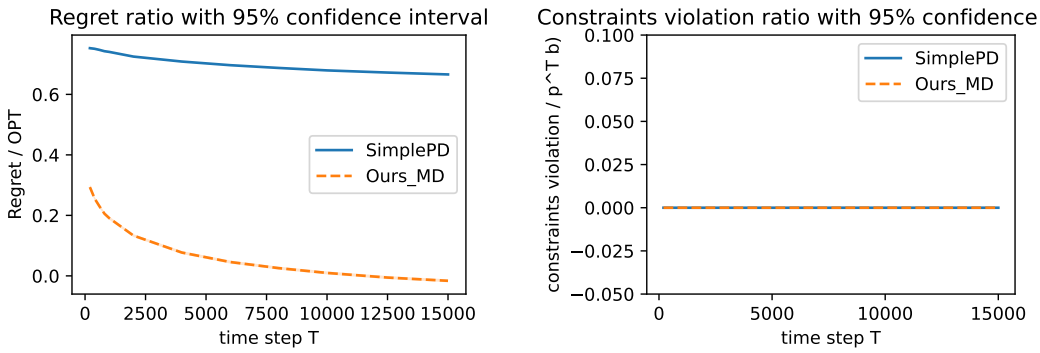
Figure 6: Cauchy Distribution with $T = 5000$



(a) Regret



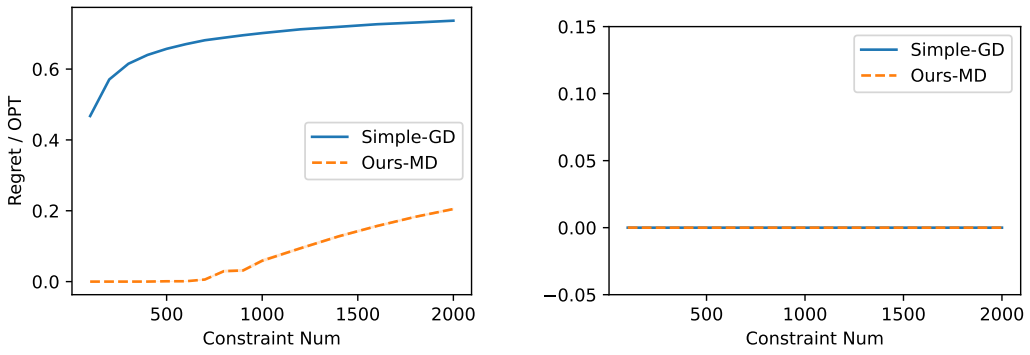
(b) Constraints violation



(a) Regret, $m = 2000$.

(b) Constraint violation, $m = 2000$.

Figure 7: Random permutation model.



(a) Regret, $T = 5000$.

(b) Constraint violation, $T = 5000$.

Figure 8: Random permutation model.