A Simple Algorithm for Online Decision Making

Rui Chen[1], Oktay Günlük[2], Andrea Lodi[1], Guanyi Wang[3]

[1]Cornell Tech ({rui.chen, andrea.lodi}@cornell.edu)
[2]Cornell University (ong5@cornell.edu)
[3]National University of Singapore (guanyi.w@nus.edu.sg)

Abstract

Motivated by recent progress on online linear programming (OLP), we study the online decision making problem (ODMP) as a natural generalization of OLP. In ODMP, there exists a single decision maker who makes a series of decisions spread out over a total of T time stages. At each time stage, the decision maker makes a decision based on information obtained up to that point without seeing into the future. The task of the decision maker is to maximize the accumulated reward while overall meeting some predetermined m-dimensional long-term goal (linking) constraints. ODMP significantly broadens the modeling framework of OLP by allowing more general feasible regions (for local and goal constraints) potentially involving both discreteness and nonlinearity in each local decision making problem.

We propose a Fenchel dual-based online algorithm for ODMP. At each time stage, the proposed algorithm requires solving a potentially nonconvex optimization problem over the local feasible set and a convex optimization problem over the goal set. Under the uniform random permutation model, we show that our algorithm achieves $O(\sqrt{mT})$ constraint violation deterministically in meeting the long-term goals, and $O(\sqrt{m \log m}\sqrt{T})$ competitive difference in expected reward with respect to the optimal offline decisions. We also extend our results to the grouped random permutation model.

1 Introduction

We consider a very general class of online decision making problems (ODMPs) that includes online (packing) linear programming (OLP) as a special case. In ODMP, the decision maker has to make a sequence of decisions in T time stages without observing the future. The decision maker tries to maximize the accumulated reward while controlling the overall impact of all the decisions. For example, a manager may try to assign tasks to employees to minimize the makespan (maximize the negative of the makespan) of each project while controlling the workload of each employee in the long term. The *offline* version of the ODMP considered in this paper takes the following form:

$$\begin{aligned} z^* = \max_{\mathbf{r},\mathbf{x},\mathbf{y}} \ & \sum_{t=1}^{T} r^t \\ \text{s.t.} \ & (r^t, \mathbf{x}^t, \mathbf{y}^t) \in \Omega^t, \quad t \in \mathcal{T} := \{1, \dots, T\}, \\ & \sum_{t=1}^{T} \mathbf{y}^t \in T\Psi. \end{aligned} \tag{1}$$

The components of the online version of this problem are discussed below:

Local Constraints. In ODMP, at time stage t, the decision maker receives a feasible set Ω^t without observing $(\Omega^\tau)_{\tau>t}$, makes some decision $\mathbf{x}^t = \hat{\mathbf{x}}^t$ respecting the local feasible set Ω^t for variables $(r^t, \mathbf{x}^t, \mathbf{y}^t)$, obtains a reward $r^t = \hat{r}^t$ and incurs a long-term impact $\mathbf{y}^t = \hat{\mathbf{y}}^t$. The relationships among decision $\mathbf{x}^t$, reward r^t and impact $\mathbf{y}^t$ at time stage t are modeled by the set Ω^t, which can be very general to incorporate both discreteness and nonlinearity. In our setting, local constraints $(r^t, \mathbf{x}^t, \mathbf{y}^t) \in \Omega^t$ for $t \in \mathcal{T}$ are *hard constraints* that must be satisfied to ensure the decision at each stage is implementable.

Goal (Linking) Constraints. In ODMP, a convex "averaged goal set" Ψ of dimension m is specified at the very beginning. Throughout the paper, we assume $m \leq T$. The goal constraints state that the accumulated impact should satisfy $\sum_{t=1}^T \mathbf{y}^t \in T\Psi$, i.e., the averaged impact $\frac{1}{T}\sum_{t=1}^T \mathbf{y}^t$ should be in Ψ. In contrast to local constraints, the global goal constraints are *soft constraints* that should be "roughly satisfied" at time stage T.

Metrics. The quality of the solutions is measured by the following two metrics:

(i) **Reward** $= \sum_{t=1}^T \hat{r}^t$, which measures the accumulated reward.

(ii) **GoalVio** $= \mathrm{dist}_2\big(\sum_{t=1}^T \hat{\mathbf{y}}^t, T\Psi\big)$, which measures the final deviation from the goals. Here, $\mathrm{dist}_2(\cdot,\cdot)$ denotes the Euclidean distance from a point to a nonempty closed set.

1.1 Our Results

We propose a simple Fenchel dual-based algorithm for solving general ODMPs. On the primal side, at each time stage t, the proposed algorithm requires solving an optimization problem over the local feasible region Ω^t, and then a convex optimization problem over Ψ. On the dual side, the algorithm can be interpreted as an online projected gradient descent algorithm that updates the dual variables based on solutions computed from the primal side. Under some mild assumptions (see Section 2.2), the algorithm deterministically ensures $O(\sqrt{mT})$ goal violation. When comparing against the reward of the optimal offline solution, the algorithm achieves $O(\sqrt{m\log m}\sqrt{T})$ competitive difference in expectation under the uniform random permutation model (specified in Section 2.1). We also show that the results can be generalized to some grouped random permutation models if the grouping of time stages is "almost even" (see Section 4.3). It is worth mentioning that our algorithm does not require the knowledge of T, and therefore works in cases when T is not initially known as long as Ψ is provided.

1.2 Related Work

Online optimization problems have been receiving an increasing attention. Well-known ones include online bipartite matching [KVV90], online routing [AAP93], single-choice [Dyn63] and multiple-choice [Kle05] secretary problems, online advertising [MSVV07], the online knapsack problem [BIKK07] and OLP [BN09]. Most of the early studies focus on worst-case analysis. More recently, the focus has shifted to less pessimistic stochastic settings where the random permutation model is used. For example, the authors of [GM08] show that a better competitive ratio of $1-1/e$ can be achieved by the greedy algorithm in online bipartite matching under the uniform random permutation model in contrast to the pessimistic worst-case competitive ratio of $1/2$.

Recently there has been a stream of work on OLP under the random permutation model [FHK+10, MR14, AWY14, KTRV14, GM16, LSY20]. As a special case of ODMP, standard (multiple-choice packing) OLPs assume $\Psi = \{\mathbf{y} : \mathbf{y} \leq \mathbf{d}\}$ with $\mathbf{d} > \mathbf{0}$, and the local feasible set Ω^t takes the form

$$\Omega^t = \{(r^t, \mathbf{x}^t, \mathbf{y}^t) : r^t = (\pi^t)^\top \mathbf{x}^t,\ \mathbf{x}^t \in \Delta,\ \mathbf{y}^t = \mathbf{A}^t\mathbf{x}^t\},$$

where $\mathbf{A}^t$ and π^t are nonnegative and Δ is the standard simplex $\{\mathbf{x} \geq \mathbf{0} : \sum_j x_j \leq 1\}$ representing a multiple-choice setting (with a void choice of $\mathbf{x}^t = \mathbf{0}$). In the context of OLP, the authors of [LSY20] propose a simple LP dual-based OLP algorithm that achieves $\tilde{O}(m\sqrt{T})$ constraint violation and competitive difference in expected reward (they call these two metrics regrets). They also show that the algorithms proposed in [AWY14] and [KTRV14] achieve smaller $\tilde{O}(\sqrt{mT})$ regret bounds. It is claimed in [LSY20] that the algorithms in [AWY14] and [KTRV14] achieve stronger regret bounds by paying more computational cost as those algorithms require solving scaled LPs. In contrast, we show that strong bounds and computational efficiency can be achieved simultaneously for OLPs. Our algorithm reduces to the simple algorithm proposed in [LSY20] for solving OLPs, while we show that the regret bounds can be improved from $\tilde{O}(m\sqrt{T})$ to $\tilde{O}(\sqrt{mT})$ by taking different stepsizes.

Existing works [MR14, AWY14, AD14, KTRV14, GM16] in OLP have also shown that if $\mathbf{A}^t \geq \mathbf{0}$ for all t and $\mathbf{b} := T\mathbf{d}$ is sufficiently large, then there exist OLP algorithms achieving $1-\epsilon$ competitive ratio in expected reward with respect to the optimal offline solution. We do not apply competitive ratio analysis in this paper because the competitive difference may be more meaningful in the context of ODMP. In particular, we do not assume the existence of void decisions like $\mathbf{x}^t = \mathbf{0}$, which have zero reward in OLP. The optimal reward z^* in general ODMPs may even be negative, making the competitive ratio pointless.

Moreover, most papers on packing OLP do not put **GoalVio** into consideration since it is easy to obtain feasible solutions for packing OLP. Specifically, in packing OLP, even if $\sum_{\tau=1}^{t} \hat{\mathbf{y}}^\tau \not\leq t\mathbf{d}$ at time stage t, it is easy to "recover from failure" by choosing void decisions, i.e., $\hat{\mathbf{x}}^\tau = \mathbf{0}$, for all $\tau \geq t+1$ as long as $\sum_{\tau=1}^{t} \hat{\mathbf{y}}^\tau \leq \mathbf{b}$. Unfortunately, due to the absence of void decisions and our general goal set, such a property does not necessarily hold for general ODMPs, and therefore, it is necessary to consider **GoalVio** when dealing with general ODMPs. See [AD14, GM16] for examples when only almost feasible solutions can be guaranteed when the problem has general constraints other than packing constraints.

The settings of ODMP in this paper are most similar to the settings of online stochastic convex programming in [AD14]. Our algorithm is different from the algorithms proposed in [AD14] as our algorithm is not a multiplicative weights update method and does not require any prior knowledge of the dependence of the optimal objective value on the goal violation, which can be hard to obtain when sets Ω^t are nonconvex. Instead, we derive strong bounds for our algorithm by establishing strong duality using the additional strong Ψ-feasibility assumption, which is easy to verify and holds for a broad class of online optimization problems, including OLPs.

2 Assumptions

We first discuss the main assumptions we make on the input data $(\Omega^t)_{t=1}^T$ and the goal set Ψ.

2.1 Stochastic Input Model

When comparing against the dynamic optimal decisions in hindsight, it is impossible to derive algorithms with a sublinear worst-case competitive difference in reward even for OLP (e.g., see [FHK$^+$10]). We adopt the convention in OLP for the underlying uncertainty by assuming that $(\Omega^t)_{t=1}^T$ follows a random permutation model. Specifically, there exists T (potentially unknown and adversarially chosen) deterministic feasible sets $Z^1, \dots, Z^T$, and for $t = 1, \dots, T$, the feasible set Ω^t observed at time stage t satisfies $\Omega^t = Z^{p(t)}$ for some sampled permutation function p of the set $\mathcal{T}$. Due to the symmetry in formulation (1), the optimal objective value z^* of (1) is invariant under any permutation p. This paper considers two random permutation models, namely the uniform random permutation model and the grouped random permutation model. The *uniform random permutation model*, also known as the random-order model, is widely studied in the context OLP. The model assumes that the permutation function p is sampled from all possible $T!$ permutations of $\{1, \dots, T\}$ with equal probability. It is known to be more general than the IID model in which each Ω^t is independently sampled from the same (potentially unknown) distribution. Due to practical concerns, we also consider the *grouped random permutation model*, which generalizes the uniform random permutation model by assuming that the set of time stages $\mathcal{T}$ is partitioned into K groups $(\mathcal{T}^k)_{k=1}^K$. For each group $\mathcal{T}^k$, $(\Omega^t)_{t\in\mathcal{T}^k}$ is a uniform random permutation of feasible sets $(Z^t)_{t\in\mathcal{T}^k}$. The grouped random permutation model is also more general than the grouped IID model in which for $t \in \mathcal{T}^k$, each Ω^t is independently sampled from a distribution associated with the group $\mathcal{T}^k$.

2.2 Other Assumptions

A1. Problem (1) is feasible. Sets $(Z^t)_{t=1}^T$ are compact. Set $\Psi \subseteq \mathbb{R}^m$ is a *full-dimensional closed convex set* satisfying $\Psi = V + C$ where V is a *compact convex set* and C is a *closed convex cone*. Here, $V + C$ denotes the Minkowski sum of sets V and C.

A2. **(Boundedness)** There exists constants $\bar{a} \in \mathbb{R}_+$ and $\bar{r} \in \mathbb{R}$ such that $\max_{\mathbf{v}\in V} \|\mathbf{y}^t - \mathbf{v}\|_\infty \le \bar{a}$ and $r^t \le \bar{r}$ for all $(r^t, \mathbf{x}^t, \mathbf{y}^t) \in Z^t$ and all $t \in \mathcal{T}$.

A3. **(Strong Ψ-Feasibility)** There exists constants $\underline{r} \in \mathbb{R}$, $\underline{d} > 0$ such that for $t \in \mathcal{T}$ there exists $(\tilde{r}^t, \tilde{\mathbf{x}}^t, \tilde{\mathbf{y}}^t) \in \mathrm{conv}(Z^t)$ satisfying $\{\tilde{\mathbf{y}}^t\} + \underline{d}B_2 \subseteq \Psi$ and $\tilde{r}^t \ge \underline{r}$. Here B_2 denotes the L_2-norm unit ball and $\mathrm{conv}(Z^t)$ denotes the convex hull of Z^t.

Note that the assumption on Ψ in **A1** covers the special cases when Ψ itself is compact (in which case $C = \{\mathbf{0}\}$) and when Ψ is polyhedral (in which case V and C are polyhedral). Assumption **A3** assumes the existence of strongly Ψ-feasible solutions $\tilde{\mathbf{y}}^t$. Note that $(\tilde{r}^t, \tilde{\mathbf{x}}^t, \tilde{\mathbf{y}}^t)$ can be picked in $\mathrm{conv}(Z^t)$ rather than Z^t. This allows us to model problems when $\mathbf{y}^t$ variables are binary indicator variables and Ψ represents some frequency requirements. Assumption **A3** naturally holds for OLP as void decisions $(\tilde{r}^t, \tilde{\mathbf{x}}^t, \tilde{\mathbf{y}}^t) = \mathbf{0}$ satisfy assumption **A3** with $\underline{r} = 0$ and $\underline{d} = \min_i d_i$. We treat $\bar{a}$, $\bar{r}$, $\underline{r}$, $\underline{d}$ as constants in this paper for simplicity of analysis. One can easily generalize our results to obtain bounds in terms of those parameters.

3 Fenchel Dual-Based Algorithm

In Algorithm 1, we describe the algorithm we use to solve ODMPs. In line 5 of Algorithm 1, C° denotes the polar cone of C, i.e., $C^\circ = \{\mathbf{u} : \mathbf{u}^\top \mathbf{v} \le 0 \text{ for all } \mathbf{v} \in C\}$, η^t is the algorithm stepsize at time stage t, and $\mathrm{proj}_{C^\circ}(\cdot)$ denotes the projection from a point onto the nonempty closed convex cone C°. In Algorithm 1 we assume that an optimization oracle which can efficiently solve local optimization problems of the form (2) is given. In this section, we describe how Algorithm 1 is derived and some intuitions from the dual side. Later in Section 4 we show some theoretical guarantees of our algorithm under the specific stepsizes $\eta^t = \min\{\frac{1}{m}, \frac{1}{\sqrt{mt}}\}$.

Algorithm 1 A Fenchel Dual-Based Algorithm for ODMP

1: **Initialize** $\mathbf{p}^1 = \mathbf{0}$
2: **for** $t = 1, \dots, T$ **do**
3: solve the following problem and implement the solution:

$$(\hat{r}^t, \hat{\mathbf{x}}^t, \hat{\mathbf{y}}^t) \in \arg\max \Big\{ r^t - (\mathbf{p}^t)^\top \mathbf{y}^t : (r^t, \mathbf{x}^t, \mathbf{y}^t) \in \Omega^t \Big\} \tag{2}$$

4: solve $\hat{\mathbf{v}}^t \in \arg\max_{\mathbf{v} \in V} (\mathbf{p}^t)^\top \mathbf{v}$
5: set $\mathbf{p}^{t+1} = \mathrm{proj}_{C^\circ}\Big(\mathbf{p}^t - \eta^t (\hat{\mathbf{v}}^t - \hat{\mathbf{y}}^t) \Big)$
6: **end for**

3.1 Fenchel Duality

Fenchel duality is a major tool we use for deriving the results in this paper. Consider functions $f^t : \mathbb{R}^m \to \mathbb{R} \cup \{-\infty\}$ defined by

$$f^t(\mathbf{y}^t) := \max_{r^t, \mathbf{x}^t} \{ r^t : (r^t, \mathbf{x}^t, \mathbf{y}^t) \in \mathrm{conv}(Z^t) \}, \ t \in \mathcal{T}.$$

Then for $t \in \mathcal{T}$, the conjugate $(f^t)^*$ of f^t is defined by

$$\begin{aligned}(f^t)^*(\mathbf{p}) &:= \min_{r^t, \mathbf{x}^t, \mathbf{y}^t} \{ \mathbf{p}^\top \mathbf{y}^t - r^t : (r^t, \mathbf{x}^t, \mathbf{y}^t) \in \mathrm{conv}(Z^t) \} \\ &= \min_{r^t, \mathbf{x}^t, \mathbf{y}^t} \{ \mathbf{p}^\top \mathbf{y}^t - r^t : (r^t, \mathbf{x}^t, \mathbf{y}^t) \in Z^t \}.\end{aligned}$$

Consider the following partial convexification of (1) with $(\Omega^t)_{t=1}^T = (Z^t)_{t=1}^T$:

$$\bar{z}^* := \max_{\mathbf{r}, \mathbf{x}, \mathbf{y}} \left\{ \sum_{t=1}^T r^t : (r^t, \mathbf{x}^t, \mathbf{y}^t)_{t=1}^T \in \prod_{t=1}^T \mathrm{conv}(Z^t), \ \sum_{t=1}^T \mathbf{y}^t \in T\Psi \right\}. \tag{3}$$

Then problem (3) can be reformulated as the following optimization problem:

$$\max_{\mathbf{y}} \ \sum_{t=1}^T f^t(\mathbf{y}^t) - \delta_{T\Psi}\Big(\sum_{t=1}^T \mathbf{y}^t \Big), \tag{4}$$

where $\delta_{T\Psi}$ is the indicator function of $T\Psi$ with $\delta_{T\Psi}(\mathbf{y}) = 0$ if $\mathbf{y} \in T\Psi$ and $\delta_{T\Psi}(\mathbf{y}) = +\infty$ otherwise. The Fenchel dual of (4) is

$$\inf_{\mathbf{p}} T h_\Psi(\mathbf{p}) - \sum_{t=1}^T (f^t)^*(\mathbf{p}), \tag{5}$$

where h_Ψ is the support function of Ψ defined by

$$h_\Psi(\mathbf{p}) = \sup_{\mathbf{v}} \{ \mathbf{p}^\top \mathbf{v} : \mathbf{v} \in \Psi \} = \begin{cases} \max_{\mathbf{v}} \{ \mathbf{p}^\top \mathbf{v} : \mathbf{v} \in V \} = h_V(\mathbf{p}) & \text{if } \mathbf{p} \in C^\circ, \\ +\infty & \text{otherwise.} \end{cases}$$

We present some strong duality results between (4) and (5) under our assumptions.

Lemma 3.1. *Suppose assumption **A1-A3** holds. Then there exists* $\mathbf{p}^* \in C^\circ$ *and* $(\bar{r}^t, \bar{\mathbf{x}}^t, \bar{\mathbf{y}}^t)_{t=1}^T \in$

$\prod_{t=1}^T$ *conv*(Z^t) *such that* $\mathbf{p}^*$ *optimizes* (5), $\sum_{t=1}^T \bar{\mathbf{y}}^t \in T\Psi$ *and*

$$\bar{z}^* = \sum_{t=1}^{T} \bar{r}^t = Th_\Psi(\mathbf{p}^*) - \sum_{t=1}^{T} (f^t)^*(\mathbf{p}^*).$$

The proof of Lemma 3.1 is presented in Appendix B together with other omitted proofs.

3.2 Algorithm Derivation

The basic idea of our Fenchel dual-based algorithm is straightforward: for any particular sequence of input data $(\Omega^t)_{t=1}^T = (Z^{p(t)})_{t=1}^T$, one learns some dual multiplier vector $\mathbf{p}^t$ using information observed up to time stage t and solves the following problem

$$-(f^{p(t)})^*(\mathbf{p}^t) = \max_{r^t, \mathbf{x}^t, \mathbf{y}^t} \left\{ r^t - (\mathbf{p}^t)^\top \mathbf{y}^t : (r^t, \mathbf{x}^t, \mathbf{y}^t) \in \Omega^t \right\} \tag{6}$$

to obtain the solution $(\hat{r}^t, \hat{\mathbf{x}}^t, \hat{\mathbf{y}}^t)$ which is implemented at time stage t. For $t = 1, \ldots, T$, define $\bar{z}^t(\mathbf{p}) = h_\Psi(\mathbf{p}) - (f^{p(t)})^*(\mathbf{p})$. Then the Fenchel dual problem (5) can be rewritten in the form $\inf_{\mathbf{p} \in C^\circ} \sum_{t=1}^T \bar{z}^t(\mathbf{p})$ since $\mathrm{dom}(\bar{z}^t) = C^\circ$ based on the definition of h_Ψ. We define the following regret for the dual multiplier learning problem:

$$\mathbf{Regret} = \sum_{t=1}^{T} \bar{z}^t(\mathbf{p}^t) - \inf_{\mathbf{p} \in C^\circ} \sum_{t=1}^{T} \bar{z}^t(\mathbf{p}) = \sum_{t=1}^{T} \bar{z}^t(\mathbf{p}^t) - \bar{z}^*. \tag{7}$$

Note that $\bar{z}^t$ is convex for all t. Therefore, the dual multiplier learning problem is a standard online convex optimization problem with regret defined by (7). Under such definition of regret for the dual problem, Algorithm 1 is essentially an online projected (sub)gradient descent algorithm for solving the dual multiplier learning problem. In fact, given solution $(\hat{r}^t, \hat{\mathbf{x}}^t, \hat{\mathbf{y}}^t) \in \arg\max \left\{ r^t - (\mathbf{p}^t)^\top \mathbf{y}^t : (r^t, \mathbf{x}^t, \mathbf{y}^t) \in \Omega^t \right\}$ and $\hat{\mathbf{v}}^t \in \arg\max\{(\mathbf{p}^t)^\top \mathbf{v}^t : \mathbf{v}^t \in V\}$, we have $\hat{\mathbf{v}}^t - \hat{\mathbf{y}}^t \in \partial \bar{z}^t(\mathbf{p}^t)$, where $\partial \bar{z}^t(\mathbf{p}^t)$ denotes the subdifferential of $\bar{z}^t$ at $\mathbf{p}^t$.

4 Algorithm Analysis

In this section, we present our main results regarding the performance of Algorithm 1 with specific stepsizes. In Section 4.1, we prove deterministic bounds on **GoalVio** and **Regret** based on Algorithm 1 independent of the permutation p. Then in Sections 4.2 and 4.3, we give bounds on **Reward** assuming $(\Omega^t)_{t=1}^T$ follows some random permutation model.

4.1 Deterministic Bounds

We first show present results that are independent of the particular permutation p that generates $(\Omega^t)_{t=1}^T$. The following lemma bounds the L_2-norm of the dual multipliers when stepsizes are small.

Lemma 4.1. *Suppose assumptions **A1**-**A3** hold and let* $\mathbf{p}^*$ *be an optimal solution to* (5)*. Then* $\|\mathbf{p}^*\|_2 \leq \frac{\bar{r} - \underline{r}}{\underline{d}} = O(1)$*. Moreover, if* $0 \leq \eta^t \leq \frac{1}{m}$ *for all* t*, then Algorithm 1 produces* $(\mathbf{p}^t)_{t=1}^{T+1}$ *with* $\max_t \|\mathbf{p}^t\|_2 \leq \frac{\bar{a}^2 + 2(\bar{r} - \underline{r})}{2\underline{d}} + \frac{\bar{a}}{\sqrt{m}} = O(1)$*.*

We next present deterministic bounds on **Regret** and **GoalVio** when certain stepsizes are used.

Theorem 4.2. *Suppose assumptions* ***A1-A3*** *hold and* $\eta_t = \min\{\frac{1}{m}, \frac{1}{\sqrt{mt}}\}$ *for all* t*. Then Algorithm 1 achieves* ***Regret*** $\leq O(\sqrt{mT})$ *and* ***GoalVio*** $\leq O(\sqrt{mT})$*.*

4.2 Uniform Random Permutation Model

We let $\mathbb{P}_p$ and $\mathbb{E}_p$ denote the probability measure and expectation with respect to the uniform random permutation p, respectively. As a benchmark we consider the optimal objective value $\bar{z}^*$ of the partial convexification (3), which is at least as large as z^*. We next establish an expected reward bound for the uniform random permutation model.

Theorem 4.3. *Suppose assumptions* ***A1-A3*** *hold and* $\eta_t = \min\{\frac{1}{m}, \frac{1}{\sqrt{mt}}\}$ *for all* t*. Under the uniform random permutation model Algorithm 1 achieves*

$$\mathbb{E}_p[\textbf{\textit{Reward}}] \geq \bar{z}^* - O(\sqrt{m \log m}\sqrt{T}).$$

Proof. Let $\mathbf{p}^*$ and $(\bar{r}^t, \bar{\mathbf{x}}^t, \bar{\mathbf{y}}^t)_{t=1}^T$ be such that Lemma 3.1 holds. Then for $(\Omega^t)_{t=1}^T = (Z^{p(t)})_{t=1}^T$ with p being a permutation of $\mathcal{T}$ we have

$$\begin{aligned}
\textbf{Reward} =& \sum_{t=1}^T (\mathbf{p}^t)^\top \hat{\mathbf{y}}^t + \sum_{t=1}^T \left(\hat{r}^t - (\mathbf{p}^t)^\top \hat{\mathbf{y}}^t\right) \\
=& \sum_{t=1}^T \left((\mathbf{p}^t)^\top \hat{\mathbf{y}}^t + \max\left\{ r^t - (\mathbf{p}^t)^\top \mathbf{y}^t : (r^t, \mathbf{x}^t, \mathbf{y}^t) \in \Omega^t \right\} \right) \\
=& \sum_{t=1}^T \left((\mathbf{p}^t)^\top \hat{\mathbf{y}}^t + \underbrace{\max\left\{ r^t - (\mathbf{p}^t)^\top \mathbf{y}^t : (r^t, \mathbf{x}^t, \mathbf{y}^t) \in \text{conv}(\Omega^t) \right\}}_{\geq \bar{r}^{p(t)} - (\mathbf{p}^t)^\top \bar{\mathbf{y}}^{p(t)}} \right) \\
\geq& \sum_{t=1}^T \bar{r}^{p(t)} + \sum_{t=1}^T (\mathbf{p}^t)^\top (\hat{\mathbf{y}}^t - \bar{\mathbf{y}}^{p(t)}) \\
=& \bar{z}^* + \sum_{t=1}^T (\mathbf{p}^t)^\top (\hat{\mathbf{y}}^t - \hat{\mathbf{v}}^t) + \sum_{t=1}^T (\mathbf{p}^t)^\top (\hat{\mathbf{v}}^t - \mathbb{E}_p[\bar{\mathbf{y}}^{p(t)}]) + \sum_{t=1}^T (\mathbf{p}^t)^\top (\mathbb{E}_p[\bar{\mathbf{y}}^{p(t)}] - \bar{\mathbf{y}}^{p(t)}).
\end{aligned}$$

Here the second equality is due to the optimality of $(\hat{r}^t, \hat{\mathbf{x}}^t, \hat{\mathbf{y}}^t)$, and the first inequality holds since $(\bar{r}^{p(t)}, \bar{\mathbf{x}}^{p(t)}, \bar{\mathbf{y}}^{p(t)}) \in \text{conv}(\Omega^t)$ from Lemma 3.1.

We first bound $\sum_{t=1}^T (\mathbf{p}^t)^\top (\hat{\mathbf{y}}^t - \hat{\mathbf{v}}^t) + \sum_{t=1}^T (\mathbf{p}^t)^\top (\hat{\mathbf{v}}^t - \mathbb{E}_p[\bar{\mathbf{y}}^{p(t)}])$. By rearranging the second inequality of Lemma A.1 (in Appendix A) with $\mathbf{p} = \mathbf{0}$, we have

$$(\mathbf{p}^t)^\top (\hat{\mathbf{y}}^t - \hat{\mathbf{v}}^t) \geq \frac{\|\mathbf{p}^{t+1}\|_2^2 - \|\mathbf{p}^t\|_2^2}{2\eta^t} - \frac{\eta^t m \bar{a}^2}{2}.$$

Then Lemma 4.1 implies that

$$\begin{aligned}
\sum_{t=1}^T (\mathbf{p}^t)^\top (\hat{\mathbf{y}}^t - \hat{\mathbf{v}}^t) \geq& \sum_{t=1}^T \frac{\|\mathbf{p}^{t+1}\|_2^2 - \|\mathbf{p}^t\|_2^2}{2\eta^t} - \sum_{t=1}^T \frac{\eta^t m \bar{a}^2}{2} \\
\geq& \sum_{t=2}^T \left(\frac{1}{2\eta^{t-1}} - \frac{1}{2\eta^t}\right) \|\mathbf{p}^t\|_2^2 - O(\sqrt{mT}) = -O(\sqrt{mT}).
\end{aligned}$$

Also note that $\mathbb{E}_p[\bar{\mathbf{y}}^{p(t)}] = T^{-1}\sum_{\tau=1}^T \bar{\mathbf{y}}^\tau \in \Psi$ for all t since p is a uniformly random permutation of $\mathcal{T}$. It implies that $(\mathbf{p}^t)^\top(\hat{\mathbf{v}}^t - \mathbb{E}_p[\bar{\mathbf{y}}^{p(t)}]) = h_\Psi(\mathbf{p}^t) - (\mathbf{p}^t)^\top\mathbb{E}_p[\bar{\mathbf{y}}^{p(t)}] \geq 0$ for all t by the definition of h_Ψ. Therefore,

$$\textbf{Reward} \geq \bar{z}^* - O(\sqrt{mT}) + 0 + \sum_{t=1}^T (\mathbf{p}^t)^\top(\mathbb{E}_p[\bar{\mathbf{y}}^{p(t)}] - \bar{\mathbf{y}}^{p(t)}). \tag{8}$$

Let $\mathcal{F}^{t-1}$ be the sigma algebra generated by the random events up to time stage $t-1$. Next we bound the term $(\mathbf{p}^t)^\top(\mathbb{E}_p[\bar{\mathbf{y}}^{p(t)}] - \bar{\mathbf{y}}^{p(t)})$ conditioned on $\mathcal{F}^{t-1}$,

$$\begin{aligned}\mathbb{E}_p[(\mathbf{p}^t)^\top(\mathbb{E}_p[\bar{\mathbf{y}}^{p(t)}] - \bar{\mathbf{y}}^{p(t)})|\mathcal{F}^{t-1}] = &\ (\mathbf{p}^t)^\top \left(\mathbb{E}_p[\bar{\mathbf{y}}^{p(t)}] - \mathbb{E}_p[\bar{\mathbf{y}}^{p(t)}|\mathcal{F}^{t-1}]\right) \\ = &\ (\mathbf{p}^t)^\top \left(\mathbb{E}_p[\bar{\mathbf{y}}^{p(t)}] - \frac{1}{T-t+1}\sum_{\tau=t}^T \bar{\mathbf{y}}^{p(\tau)}\right) \geq\ -\|\mathbf{p}^t\|_2 \left\|\mathbb{E}_p[\bar{\mathbf{y}}^{p(t)}] - \frac{1}{T-t+1}\sum_{\tau=t}^T \bar{\mathbf{y}}^{p(\tau)}\right\|_2,\end{aligned}$$

where the first equality holds since $\mathbf{p}^t$ is determined based on $(\Omega^\tau)_{\tau=1}^{t-1}$ (i.e., $\mathcal{F}^{t-1}$), the second equality is due to the fact that the conditional expectation of $\bar{\mathbf{y}}^{p(t)}$ is a sample from $\{\bar{\mathbf{y}}^t\}_{t\in\mathcal{T}} \setminus \{\bar{\mathbf{y}}^{p(\tau)}\}_{\tau=1}^{t-1} = \{\bar{\mathbf{y}}^{p(\tau)}\}_{\tau=t}^T$ with equal probability since p is a uniform random permutation.

As p is uniformly chosen at random, $\big(\sigma(t) = p(T-t+1)\big)_{t=1}^T$ is also a uniform random permutation of $\mathcal{T}$. By Hoeffding's inequality for sampling without replacement [Hoe63], for all $\epsilon > 0$ and $i \in \{1, \ldots, m\}$ we have

$$\mathbb{P}_p\left(\left|\left(\mathbb{E}_p[\bar{\mathbf{y}}^{p(t)}] - \frac{1}{T-t+1}\sum_{\tau=t}^T \bar{\mathbf{y}}^{p(\tau)}\right)_i\right| > \epsilon\right) \leq 2\exp\Big(-\frac{(T-t+1)\epsilon^2}{2\bar{a}^2}\Big).$$

In other words, for all $\rho \in (0,1]$, with probability at least $1-\rho/m$,

$$\left|\left(\mathbb{E}_p[\bar{\mathbf{y}}^{p(t)}] - \frac{1}{T-t+1}\sum_{\tau=t}^T \bar{\mathbf{y}}^{p(\tau)}\right)_i\right| \leq \sqrt{\frac{2\bar{a}^2\log(2m/\rho)}{T-t+1}}.$$

Then by taking the union bound over $i \in \{1, \ldots, m\}$, with probability at least $1-\rho$,

$$\left\|\mathbb{E}_p[\bar{\mathbf{y}}^{p(t)}] - \frac{1}{T-t+1}\sum_{\tau=t}^T \bar{\mathbf{y}}^{p(\tau)}\right\|_2 \leq \sqrt{\frac{2m\bar{a}^2\log(2m/\rho)}{T-t+1}} = O\Big(\sqrt{\frac{m\log m}{T-t+1}} + \sqrt{\frac{m\log(2/\rho)}{T-t+1}}\Big). \tag{9}$$

By Lemma 4.1 and integrating the quantile function, we have

$$\begin{aligned}\mathbb{E}_p[(\mathbf{p}^t)^\top(\mathbb{E}_p[\bar{\mathbf{y}}^{p(t)}] - \bar{\mathbf{y}}^{p(t)})] &= \mathbb{E}_p\left[\mathbb{E}_p[(\mathbf{p}^t)^\top(\mathbb{E}_p[\bar{\mathbf{y}}^{p(t)}] - \bar{\mathbf{y}}^{p(t)})|\mathcal{F}^{t-1}]\right] \\ &\geq -\mathbb{E}_p\left[\|\mathbf{p}^t\|_2\left\|\mathbb{E}_p[\bar{\mathbf{y}}^{p(t)}] - \frac{1}{T-t+1}\sum_{\tau=t}^T \bar{\mathbf{y}}^{p(\tau)}\right\|_2\right] \\ &\geq -\int_0^1 O\Big(\sqrt{\frac{m\log m}{T-t+1}} + \sqrt{\frac{m\log(2/\rho)}{T-t+1}}\Big)\, d\rho = -O\Big(\sqrt{\frac{m\log m}{T-t+1}}\Big).\end{aligned} \tag{10}$$

It then follows that

$$\mathbb{E}[\mathbf{Reward}] \geq \bar{z}^* + \sum_{t=1}^{T} \mathbb{E}\left[(\mathbf{p}^t)^\top(\mathbb{E}_p[\bar{\mathbf{y}}^{p(t)}] - \bar{\mathbf{y}}^{p(t)})\right] - O(\sqrt{mT})$$
$$\geq \bar{z}^* - O\big(\sqrt{m\log m}\sqrt{T}\big) - O(\sqrt{mT}) = \bar{z}^* - O(\sqrt{m\log m}\sqrt{T}).$$

□

We next present a high probability bound on the reward.

Corollary 4.4. *Suppose assumptions **A1-A3** hold and* $\eta_t = \min\{\frac{1}{m}, \frac{1}{\sqrt{mt}}\}$ *for all* t. *Under the uniform random permutation model, for all* $\rho \in (0, 1]$, *with probability at least* $1 - \rho$ *Algorithm 1 achieves **Reward*** $\geq \bar{z}^* - O(\sqrt{m\log m}\sqrt{T} + \sqrt{m}\sqrt{T\log(T/\rho)})$.

If T is known, one may apply a variant of Algorithm 1 with a restart-at-$T/2$ strategy similar to the one used in [GM16] to reduce $\log(T/\rho)$ to $\log(1/\rho)$ in the above probability bound using the maximal inequality in [Pru98].

4.3 Grouped Random Permutation Model

In a realistic setting, $(\Omega^t)_{t=1}^T$ may not follow a uniform random permutation of a single family of feasible sets. For example, weekday problems may be significantly different from weekend problems. We generalize the classic uniform random permutation model to what we call the grouped random permutation model. Formally speaking, we consider the case when the set of time stages $\mathcal{T} = \{1, \ldots, T\}$ is partitioned into K groups $(\mathcal{T}^k)_{k=1}^K$, and for each group $\mathcal{T}^k$, $(\Omega^t)_{t\in\mathcal{T}^k}$ is a uniform random permutation of feasible sets $(Z^t)_{t\in\mathcal{T}^k}$.

Example 1 (Half-Half Partition)**.** Consider an online knapsack problem where there are two phases of item arrivals, $\mathcal{T}^1 = \{1, \ldots, T/2\}$ and $\mathcal{T}^2 = \{T/2+1, \ldots, T\}$. Items arriving in phase 1 has weight 2 and reward 2, and items arriving in phase 2 has weight 0 and reward 0. Assume the decision maker has an average weight budget of 1 for each item arrival and has to decide whether to accept it or not when an item arrives. The optimal offline decisions would be to accept all phase 1 items and yield reward T. However, the decision maker does not know there is a phase 2 of item arrivals. Based on initial observations in phase 1, the best strategy for the decision maker is probably to accept half of the phase-1 items. This strategy only yields a reward of $T/2$.

Example 1 implies that to have a small (sublinear) competitive difference in reward, the partition $(\mathcal{T}^k)_{k=1}^K$ cannot be arbitrary as the decision maker cannot see into the future and can be biased by the observations made so far. We show that our results generalize to this more general grouped random permutation setting if $\mathcal{T}$ is partitioned "almost evenly".

Definition 1. Let $\bar{\mu}$ denote the uniform distribution over $\mathcal{T}$, and let μ^k denote the uniform distribution over $\mathcal{T}^k$ for $k = 1, \ldots, K$. For $k = 1, \ldots, K$, the (1-) *Wasserstein distance* w^k between discrete distributions $\bar{\mu}$ and μ^k is defined as the optimal objective value of the following optimal transportation linear program [PC+19]:

$$\begin{aligned} w^k := \min_{\mathbf{q}\in\mathbb{R}_+^{\mathcal{T}^k\times\mathcal{T}}} \quad & \sum_{i\in\mathcal{T}^k}\sum_{j\in\mathcal{T}} d_{ij}q_{ij} \\ \text{s.t.} \quad & \sum_{i\in\mathcal{T}^k} q_{ij} = \frac{1}{T}, \quad j \in \mathcal{T}, \\ & \sum_{j\in\mathcal{T}} q_{ij} = \frac{1}{|\mathcal{T}^k|}, \quad i \in \mathcal{T}^k. \end{aligned} \tag{11}$$

We use the stepsizes to define the distance d_{ij} between two time stages i and j. Specifically, define

$$d_{ij} := \begin{cases} \sum_{t=i}^{j-1} \eta^t & \text{if } i < j, \\ \sum_{t=j}^{i-1} \eta^t & \text{otherwise.} \end{cases}$$

We use the weighted Wasserstein distance sum $W := \sum_{k=1}^{K} m|\mathcal{T}^k|w^k$ to measure the *unevenness* of the partition $(\mathcal{T}^k)_{k=1}^K$.

Example 2. Let $\eta_t = \min\{\frac{1}{m}, \frac{1}{\sqrt{mt}}\}$ for all t. Then for some special partitions of $\mathcal{T}$, the unevenness W is as follows.

	$(\mathcal{T}^k)_{k=1}^K$	W
Weekday-Weekend	$\mathcal{T}^1 = \bigcup_{\tau=1}^{5}\{t \in \mathcal{T} : t \equiv \tau \mod 7\}$, $\mathcal{T}^2 = \mathcal{T} \setminus \mathcal{T}^1$	$\Theta(\sqrt{mT})$
Half-Half	$\mathcal{T}^1 = \{t \in \mathcal{T} : t \leq T/2\}$, $\mathcal{T}^2 = \mathcal{T} \setminus \mathcal{T}^1$	$\Theta(\sqrt{m}T^{3/2})$
K-Even	$\mathcal{T}^k = \{t \in \mathcal{T} : t \equiv k \mod K\}$, $k = 1, \ldots, K$	$\Theta(K\sqrt{mT})$

Note that weekday-weekend and K-even partitions (with $K = o(\sqrt{T})$) have smaller unevenness W (sublinear in T) than the half-half partition does (not sublinear in T).

The following result that generalizes Theorem 4.3 is consistent with our intuition that an uneven partition $(\mathcal{T}^k)_{k=1}^K$ of $\mathcal{T}$ can yield a large competitive difference in reward.

Theorem 4.5. *Suppose assumptions **A1-A3** hold and $\eta_t = \min\{\frac{1}{m}, \frac{1}{\sqrt{mt}}\}$ for all t. Under the grouped random permutation model with groups $(\mathcal{T}^k)_{k=1}^K$, Algorithm 1 achieves*

$$\mathbb{E}_p[\textbf{\textit{Reward}}] \geq \bar{z}^* - O\Big(W + \sqrt{m \log m} \sum_{k=1}^{K} \sqrt{|\mathcal{T}^k|}\Big).$$

Corollary 4.4 also generalizes to the grouped random permutation setting.

Corollary 4.6. *Suppose assumptions **A1-A3** hold and $\eta_t = \min\{\frac{1}{m}, \frac{1}{\sqrt{mt}}\}$ for all t. Under the grouped random permutation model with groups $(\mathcal{T}^k)_{k=1}^K$, for all $\rho \in (0, 1]$, with probability at least $1 - \rho$ Algorithm 1 achieves **Reward** $\geq \bar{z}^* - O\Big(W + \big(\sqrt{m \log m} + \sqrt{m \log(T/\rho)}\big) \sum_{k=1}^{K} \sqrt{|\mathcal{T}^k|}\Big)$.*

5 Discussion

Optimization Tolerance. Solving the local optimization problem in line 3 of Algorithm 1 can be difficult. It is possible to generalize our algorithm by allowing implementing an approximate solution $(\hat{r}^t, \hat{\mathbf{x}}^t, \hat{\mathbf{y}}^t)$ of (2) as long as the absolute optimality gap δ^t at each time stage t is upper bounded by a constant, at the cost of having an additional $\sum_{t=1}^{T} \delta^t$ loss in reward bounds.
Other Stepsizes. For Algorithm 1, we recommend diminishing stepsizes $\eta_t = \min\{\frac{1}{m}, \frac{1}{\sqrt{mt}}\}$ that does not require a prior knowledge of T. If T is initially known, constant stepsizes like $\eta_t = \frac{1}{\sqrt{mT}}$ can be used and achieve the same bounds. Another advantage of using diminishing stepsizes $\eta_t = \min\{\frac{1}{m}, \frac{1}{\sqrt{mt}}\}$ is that, no matter T is known or not, we have $\text{dist}_2(\sum_{\tau=1}^{t} \hat{\mathbf{y}}^t, t\Psi) \leq O(\sqrt{mt})$ for all intermediate time stages $t \in [m, T]$ following arguments similar to the proof of Theorem 4.2.
Online Duality. It is known in nonlinear programming [AE76, UB16, BT20] that (offline) duality gap is small when one optimizes a sum of separable nonconvex objective functions subject to a small

amount of linear linking constraints. Comparing with these existing results, our online algorithm generalizes existing offline nonconvex duality to online settings under certain random permutation models, achieving strong primal and dual bounds simultaneously.

A Some Other Lemmas

Lemma A.1. *Suppose assumptions **A1**-**A3** hold. For all $\mathbf{p} \in C^\circ$ and $t = 1, \ldots, T$, we have*

1. $\|\mathbf{p}\|_2 \underline{d} \leq h_\Psi(\mathbf{p}) - \mathbf{p}^\top \tilde{\mathbf{y}}^t$;
2. $(\mathbf{p}^t - \mathbf{p})^T(\hat{\mathbf{v}}^t - \hat{\mathbf{y}}^t) \leq \frac{\|\mathbf{p}^t - \mathbf{p}\|_2^2 - \|\mathbf{p}^{t+1} - \mathbf{p}\|_2^2}{2\eta^t} + \frac{\eta^t}{2} m\bar{a}^2$.

Proof. Since $\tilde{\mathbf{y}}^t + \underline{d}B_2 \subseteq \Psi$ by assumption **A3**, we have $h_\Psi(\mathbf{p}) \geq h_{\tilde{\mathbf{y}}^t + \underline{d}B_2}(\mathbf{p}) = \mathbf{p}^\top \tilde{\mathbf{y}}^t + \|\mathbf{p}\|_2 \underline{d}$. The first conclusion then follows. For the second conslusion, note that since $\mathbf{p} \in C^\circ$ and $\mathbf{p}^{t+1} = \mathrm{proj}_{C^\circ}(\mathbf{p}^t - \eta^t(\hat{\mathbf{v}}^t - \hat{\mathbf{y}}^t))$, it follows that

$$\|\mathbf{p}^{t+1} - \mathbf{p}\|_2^2 \leq \|\mathbf{p}^t - \eta^t(\hat{\mathbf{v}}^t - \hat{\mathbf{y}}^t) - \mathbf{p}\|_2^2 = \|\mathbf{p}^t - \mathbf{p}^*\|_2^2 + (\eta^t)^2 \|\hat{\mathbf{v}}^t - \hat{\mathbf{y}}^t\|_2^2 - 2\eta^t(\mathbf{p}^t - \mathbf{p}^*)^\top(\hat{\mathbf{v}}^t - \hat{\mathbf{y}}^t),$$

where the first inequality is due to the fact that the projection operator is nonexpansive. By rearranging the inequality and applying assumption **A2**, we have

$$\begin{aligned} (\mathbf{p}^t - \mathbf{p})^T(\hat{\mathbf{v}}^t - \hat{\mathbf{y}}^t) &\leq \frac{\|\mathbf{p}^t - \mathbf{p}\|_2^2 - \|\mathbf{p}^{t+1} - \mathbf{p}\|_2^2}{2\eta^t} + \frac{\eta^t}{2}\|\hat{\mathbf{v}}^t - \hat{\mathbf{y}}^t\|_2^2 \\ &\leq \frac{\|\mathbf{p}^t - \mathbf{p}\|_2^2 - \|\mathbf{p}^{t+1} - \mathbf{p}\|_2^2}{2\eta^t} + \frac{\eta^t}{2} m\bar{a}^2. \end{aligned}$$

□

Lemma A.2. *Suppose assumptions **A1** holds. Then for $t = 1, \ldots, T$, we have $\mathbf{p}^{t+1} = \mathbf{p}^t - \eta^t(\hat{\mathbf{v}}^t - \hat{\mathbf{y}}^t) - \hat{\mathbf{u}}^t$ for some $\hat{\mathbf{u}}^t \in C$.*

Proof. Let $\hat{\mathbf{u}}^t := \mathbf{p}^t - \eta^t(\hat{\mathbf{v}}^t - \hat{\mathbf{y}}^t) - \mathbf{p}^{t+1}$ and $\mathbf{p}^{t+1/2} := \mathbf{p}^t - \eta^t(\hat{\mathbf{v}}^t - \hat{\mathbf{y}}^t)$. Then $\mathbf{p}^{t+1} = \mathrm{proj}_{C^\circ}(\mathbf{p}^{t+1/2})$ and $\hat{\mathbf{u}}^t = \mathbf{p}^{t+1/2} - \mathbf{p}^{t+1} = \mathbf{p}^{t+1/2} - \mathrm{proj}_{C^\circ}(\mathbf{p}^{t+1/2})$. By the Bourbaki-Cheney-Goldstein inequality [HUL13, Chapter III Theorem 3.1.1], for all $\mathbf{z} \in C^\circ$, we have

$$\left(\mathrm{proj}_{C^\circ}(\mathbf{p}^{t+1/2}) - \mathbf{p}^{t+1/2}\right)^\top \left(\mathbf{z} - \mathrm{proj}_{C^\circ}(\mathbf{p}^{t+1/2})\right) \geq 0.$$

Since C° is a closed cone, we have $\lambda\mathbf{z} \in C^\circ$ for all $\lambda \geq 0$. Therefore, all $\mathbf{z} \in C^\circ$ and all $\lambda \geq 0$, we have

$$\left(\mathrm{proj}_{C^\circ}(\mathbf{p}^{t+1/2}) - \mathbf{p}^{t+1/2}\right)^\top \left(\lambda\mathbf{z} - \mathrm{proj}_{C^\circ}(\mathbf{p}^{t+1/2})\right) \geq 0. \tag{12}$$

It implies that for all $\mathbf{z} \in C^\circ$,

$$\left(\mathrm{proj}_{C^\circ}(\mathbf{p}^{t+1/2}) - \mathbf{p}^{t+1/2}\right)^\top \mathbf{z} \geq 0.$$

Otherwise, (12) does not hold for some $\mathbf{z} \in C^\circ$ with large enough λ. By the definition of polar cone, we have $\hat{\mathbf{u}}^t = \mathbf{p}^{t+1/2} - \mathrm{proj}_{C^\circ}(\mathbf{p}^{t+1/2}) \in C^{\circ\circ}$. Since C is a nonemtpy closed convex cone, we have $C^{\circ\circ} = C$ by [Roc70, Theorem 14.1]. The conclusion then follows. □

B Omitted Proofs

B.1 Proof of Lemma 3.1

Proof. Note that by definition of $(f^t)^*$ and compactness of Z^t from assumption **A1**, $\mathrm{dom}(\sum_{t=1}^T (f^t)^*) = \mathbb{R}^m \supseteq \mathrm{dom}(Th_\Psi)$. Here $\mathrm{dom}(\cdot)$ denotes the effective domain of a function. By [Roc70, Corollary 31.2.1], it suffices to show that there exists $(\mathbf{y}^t)_{t=1}^T \in \mathrm{ri}(\mathrm{proj}_{\mathbf{y}}(\prod_{t=1}^T \mathrm{conv}(Z^t)))$ such that $\sum_{t=1}^T \mathbf{y}^t \in \mathrm{ri}(T\Psi)$, where $\mathrm{ri}(\cdot)$ denotes the relative interior of a set. By assumption **A3**, we have

$$\sum_{t=1}^T \tilde{\mathbf{y}}^t + \frac{T\bar{d}}{2} B_2 \subseteq \mathrm{ri}(T\Psi) \tag{13}$$

Pick $(\check{\mathbf{y}}^t)_{t=1}^T \in \mathrm{ri}(\mathrm{proj}_{\mathbf{y}}(\prod_{t=1}^T \mathrm{conv}(Z^t)))$. Then we have

$$(\mathbf{y}^t_\lambda)_{t=1}^T := (1-\lambda)(\tilde{\mathbf{y}}^t)_{t=1}^T + \lambda(\check{\mathbf{y}}^t)_{t=1}^T \in \mathrm{ri}\Big(\mathrm{proj}_{\mathbf{y}}\Big(\prod_{t=1}^T \mathrm{conv}(Z^t)\Big)\Big)$$

for all $\lambda \in (0,1]$ by [Roc70, Theorem 6.1], and $\sum_{t=1}^T \mathbf{y}^t_\lambda \in \mathrm{ri}(T\Psi)$ for some small enough $\lambda > 0$ due to (13). The conclusion then follows. □

B.2 Proof of Lemma 4.1

Proof. Let p denote the permutation of $\mathcal{T}$ such that $(\Omega^t)_{t=1}^T = (Z^{p(t)})_{t=1}^T$. Let $(\bar{r}^t, \bar{\mathbf{x}}^t, \bar{\mathbf{y}}^t)_{t=1}^T$ and $\mathbf{p}^*$ be such that Lemma 3.1 is satisfied, and $(\tilde{r}^t, \tilde{\mathbf{x}}^t, \tilde{\mathbf{y}}^t)_{t=1}^T$ be such that assumption **A3** is assumed. Then $\mathbf{p}^*$ satisfies

$$T\underline{r} + T\|\mathbf{p}^*\|_2\underline{d} \le \sum_{t=1}^T \left(\tilde{r}^t + h_\Psi(\mathbf{p}^*) - (\mathbf{p}^*)^\top \tilde{\mathbf{y}}^t\right) \le \sum_{t=1}^T \left(h_\Psi(\mathbf{p}^*) - (f^t)^*(\mathbf{p}^*)\right) = \bar{z}^* \le T\bar{r},$$

where the first inequality holds since $T\underline{r} \le \sum_{t=1}^T \tilde{r}^t$ by assumption **A3** and $T\|\mathbf{p}^*\|_2\underline{d} \le \sum_{t=1}^T h_\Psi(\mathbf{p}^*) - (\mathbf{p}^*)^\top \tilde{\mathbf{y}}^t$ by the first inequality of Lemma A.1 with $\mathbf{p} = \mathbf{p}^*$, the second inequality holds due to the definition of $(f^t)^*$. Therefore, $\|\mathbf{p}^*\|_2 \le \frac{\bar{r}-\underline{r}}{\underline{d}}$. Similarly,

$$\underline{r} + \|\mathbf{p}^t\|_2\underline{d} \le \tilde{r}^{p(t)} + h_\Psi(\mathbf{p}^t) - (\mathbf{p}^t)^\top \tilde{\mathbf{y}}^{p(t)} \le \hat{r}^t + h_\Psi(\mathbf{p}^t) - (\mathbf{p}^t)^\top \hat{\mathbf{y}}^t \le \bar{r} + (\mathbf{p}^t)^\top(\hat{\mathbf{v}}^t - \hat{\mathbf{y}}^t), \tag{14}$$

where the second inequality holds due to optimality of $(\hat{r}^t, \hat{\mathbf{x}}^t, \hat{\mathbf{y}}^t)$ in line 3 of Algorithm 1, and the third inequality holds due to assumption **A2** and the optimality of $\hat{\mathbf{v}}^t$ in line 4 of Algorithm 1. By rearranging the second inequality in Lemma A.1 with $\mathbf{p} = \mathbf{0}$, we have

$$\|\mathbf{p}^{t+1}\|_2^2 \le \|\mathbf{p}^t\|_2^2 + (\eta^t)^2 m\bar{a}^2 + 2\eta^t(\mathbf{p}^t)^\top(\hat{\mathbf{y}}^t - \hat{\mathbf{v}}^t). \tag{15}$$

Combining (14) and (15), we have

$$\|\mathbf{p}^{t+1}\|_2^2 \le \|\mathbf{p}^t\|_2^2 + (\eta^t)^2 m\bar{a}^2 + 2\eta^t\left(\bar{r} - \underline{r} - \|\mathbf{p}^t\|_2\underline{d}\right).$$

Suppose $0 \le \eta^t \le \frac{1}{m}$. We have $\|\mathbf{p}^{t+1}\|_2 \le \|\mathbf{p}^t\|_2$ when $\|\mathbf{p}^t\|_2 \ge \frac{\bar{a}^2 + 2(\bar{r}-\underline{r})}{2\underline{d}}$. On the other hand, when $\|\mathbf{p}^t\|_2 \le \frac{\bar{a}^2 + 2(\bar{r}-\underline{r})}{2\underline{d}}$, we have $\|\mathbf{p}^{t+1}\|_2 \le \|\mathbf{p}^t\|_2 + \eta^t\|\hat{\mathbf{y}}^t - \hat{\mathbf{v}}^t\|_2 \le \frac{\bar{a}^2 + 2(\bar{r}-\underline{r})}{2\underline{d}} + \frac{\bar{a}}{\sqrt{m}}$. Since $\mathbf{p}^1 = \mathbf{0}$, the conclusion follows from induction on $\|\mathbf{p}^t\|_2$. □

B.3 Proof of Theorem 4.2

Proof. Let $\mathbf{p}^*$ be an optimal dual multiplier that satisfies Lemma 3.1. For $t \geq 1$, by the second inequality of Lemma A.1 with $\mathbf{p} = \mathbf{p}^*$, we have

$$(\mathbf{p}^t - \mathbf{p}^*)^\top (\hat{\mathbf{v}}^t - \hat{\mathbf{y}}^t) \leq \frac{\|\mathbf{p}^t - \mathbf{p}^*\|_2^2 - \|\mathbf{p}^{t+1} - \mathbf{p}^*\|_2^2}{2\eta^t} + \frac{\eta^t}{2} m\bar{a}^2. \tag{16}$$

By optimality of $(\hat{r}^t, \hat{\mathbf{x}}^t, \hat{\mathbf{y}}^t)$ in line 3 of Algorithm 1,

$$\bar{z}^t(\mathbf{p}^t) - \bar{z}^t(\mathbf{p}^*) \leq \big(\hat{r}^t - (\mathbf{p}^t)^\top \hat{\mathbf{y}}^t + h_\Psi(\mathbf{p}^t)\big) - \big(\hat{r}^t - (\mathbf{p}^*)^\top \hat{\mathbf{y}}^t + h_\Psi(\mathbf{p}^*)\big) \leq (\mathbf{p}^t - \mathbf{p}^*)^\top (\hat{\mathbf{v}}^t - \hat{\mathbf{y}}^t).$$

Together with (16) and boundedness results of dual multipliers in Lemma 4.1, we have

$$\begin{aligned}
\textbf{Regret} &= \sum_{t=1}^{T} \big(\bar{z}^t(\mathbf{p}^t) - \bar{z}^t(\mathbf{p}^*)\big) \\
&\leq \sum_{t=1}^{T} (\mathbf{p}^t - \mathbf{p}^*)^\top (\hat{\mathbf{v}}^t - \hat{\mathbf{y}}^t) \\
&\leq \sum_{t=1}^{T} \frac{\|\mathbf{p}^t - \mathbf{p}^*\|_2^2 - \|\mathbf{p}^{t+1} - \mathbf{p}^*\|_2^2}{2\eta^t} + \sum_{t=1}^{T} \frac{\eta^t}{2} m\bar{a}^2 \\
&\leq \frac{\|\mathbf{p}^*\|_2^2}{2\eta^1} + \sum_{t=2}^{T} \Big(\frac{1}{2\eta^t} - \frac{1}{2\eta^{t-1}}\Big) \|\mathbf{p}^t - \mathbf{p}^*\|_2^2 + \sum_{t=1}^{T} \frac{\eta^t}{2} m\bar{a}^2 \\
&\leq O(m) + O(\sqrt{mT}) + O(\sqrt{mT}) = O(\sqrt{mT}).
\end{aligned} \tag{17}$$

On the other hand, by Lemma A.2, there exists $\hat{\mathbf{u}}^t \in C$ such that $\mathbf{p}^{t+1} = \mathbf{p}^t - \eta^t(\hat{\mathbf{v}}^t - \hat{\mathbf{y}}^t) - \hat{\mathbf{u}}^t$ for $t = 1, \dots, T$. Therefore,

$$\sum_{t=1}^{T} \hat{\mathbf{y}}^t - \sum_{t=1}^{T} \Big(\hat{\mathbf{v}}^t + \frac{\hat{\mathbf{u}}^t}{\eta^t}\Big) = \sum_{t=1}^{T} \frac{\mathbf{p}^{t+1} - \mathbf{p}^t}{\eta^t} = \sum_{t=2}^{T} \Big(\frac{1}{\eta^{t-1}} - \frac{1}{\eta^t}\Big) \mathbf{p}^t + \frac{1}{\eta^T} \mathbf{p}^{T+1}.$$

Note that $\hat{\mathbf{v}}^t + \frac{\hat{\mathbf{u}}^t}{\eta^t} \in V + C = \Psi$ for all $t = 1, \dots, T$. Then by Lemma 4.1 we have

$$\begin{aligned}
\textbf{GoalVio} = \mathrm{dist}_2\Big(\sum_{t=1}^{T} \hat{\mathbf{y}}^t, T\Psi\Big) &\leq \Big\| \sum_{t=1}^{T} \hat{\mathbf{y}}^t - \sum_{t=1}^{T} \Big(\hat{\mathbf{v}}^t + \frac{\hat{\mathbf{u}}^t}{\eta^t}\Big) \Big\|_2 \\
&\leq \sum_{t=2}^{T} \Big(\frac{1}{\eta^{t-1}} - \frac{1}{\eta^t}\Big) \|\mathbf{p}^t\|_2 + \frac{1}{\eta^T} \|\mathbf{p}^{T+1}\|_2 = \sum_{t=2}^{T} O\Big(\frac{\sqrt{m}}{\sqrt{t}}\Big) + O(\sqrt{mT}) = O(\sqrt{mT}).
\end{aligned}$$

□

B.4 Proof of Corollary 4.4

Proof. Replacing ρ by $\rho/2T$ in (9) and taking the union bound over t, with probability at least $1 - \rho/2$ we have

$$\sum_{t=1}^{T}(\mathbf{p}^t)^\top\left(\mathbb{E}_p[\bar{\mathbf{y}}^{p(t)}]-\mathbb{E}_p[\bar{\mathbf{y}}^{p(t)}|\mathcal{F}^{t-1}]\right)\geq-\sum_{t=1}^{T}\|\mathbf{p}^t\|_2\left\|\mathbb{E}_p[\bar{\mathbf{y}}^{p(t)}]-\frac{1}{T-t+1}\sum_{\tau=t}^{T}\bar{\mathbf{y}}^{p(\tau)}\right\|_2$$
$$\geq-O(\sqrt{m\log m}\sqrt{T}+\sqrt{m}\sqrt{T\log(T/\rho)}). \quad (18)$$

Now define random variables $Z^0=0$ and $Z^t=(\mathbf{p}^t)^\top\left(\mathbb{E}_p[\bar{\mathbf{y}}^{p(t)}|\mathcal{F}^{t-1}]-\bar{\mathbf{y}}^{p(t)}\right)$ for $t=1,\ldots,T$. Note that $\mathbb{E}_p[Z^t|\mathcal{F}^{t-1}]=0$. Therefore, $(\sum_{\tau=1}^{t}Z_\tau)_{t=0}^T$ is a martingale with respect to the filtration $(\mathcal{F}^t)_{t=0}^T$. Also note that $|Z^t|\leq2\sqrt{m}\bar{a}\|\mathbf{p}^t\|_2=O(\sqrt{m})$ for all t. By Azuma-Hoeffding inequality [Azu67], for all $\epsilon>0$ we have

$$\mathbb{P}\left(\sum_{t=1}^{T}Z_t\leq-\epsilon\right)\leq\exp\left(\frac{-\epsilon^2}{O(mT)}\right).$$

Let $\epsilon=\sqrt{\log(2/\rho)O(mT)}=O(\sqrt{m}\sqrt{T\log(1/\rho)})$. Then with probability at least $1-\rho/2$ we have

$$\sum_{t=1}^{T}(\mathbf{p}^t)^\top\left(\mathbb{E}_p[\bar{\mathbf{y}}^{p(t)}|\mathcal{F}^{t-1}]-\bar{\mathbf{y}}^{p(t)}\right)=\sum_{t=1}^{T}Z_t\geq-O(\sqrt{m}\sqrt{T\log(1/\rho)}). \quad (19)$$

Taking the union bound of (18) and (19), the conclusion then follows from (8). □

B.5 Proof of Theorem 4.5

Proof. Let p denote the grouped random permutation of $(Z^t)_{t=1}^T$ that such that $(\Omega^t)_{t=1}^T=(Z^{p(t)})_{t=1}^T$. Let $\mathbf{p}^*$ and $(\bar{r}^t,\bar{\mathbf{x}}^t,\bar{\mathbf{y}}^t)_{t=1}^T$ be such that Lemma 3.1 holds. Following arguments similar to the proof of Theorem 4.3, we have the following deterministic bound

$$\textbf{Reward}\geq\bar{z}^*+\sum_{t=1}^{T}(\mathbf{p}^t)^\top(\hat{\mathbf{v}}^t-\mathbb{E}_p[\bar{\mathbf{y}}^{p(t)}])+\sum_{t=1}^{T}(\mathbf{p}^t)^\top(\mathbb{E}_p[\bar{\mathbf{y}}^{p(t)}]-\bar{\mathbf{y}}^{p(t)})-O(\sqrt{mT}). \quad (20)$$

Under the grouped random permutation model, we have $\mathbb{E}_p[\bar{\mathbf{y}}^{p(t)}]=|\mathcal{T}^k|^{-1}\sum_{\tau\in\mathcal{T}^k}\bar{\mathbf{y}}^\tau$ for all $t\in\mathcal{T}^k$ and $k=1,\ldots,K$. Let $\bar{\mathbf{p}}=T^{-1}\sum_{t=1}^T\mathbf{p}^t$. Since $T^{-1}\sum_{\tau\in\mathcal{T}}\bar{\mathbf{y}}^\tau\in\Psi$, then $(\mathbf{p}^t)^\top\hat{\mathbf{v}}^t=h_\Psi(\mathbf{p}^t)\geq(\mathbf{p}^t)^\top(T^{-1}\sum_{\tau\in\mathcal{T}}\bar{\mathbf{y}}^\tau)$. Fix an arbitrary $\bar{\mathbf{v}}\in V$. We have

$$\sum_{t=1}^{T}(\mathbf{p}^t)^\top(\hat{\mathbf{v}}^t-\mathbb{E}_p[\bar{\mathbf{y}}^{p(t)}])\geq\Big(\underbrace{\sum_{t=1}^{T}\mathbf{p}^t}_{=T\bar{\mathbf{p}}}\Big)^\top\Big(T^{-1}\underbrace{\sum_{\tau\in\mathcal{T}}\bar{\mathbf{y}}^\tau}_{\sum_{k=1}^K\sum_{\tau\in\mathcal{T}^k}\bar{\mathbf{y}}^\tau}\Big)-\sum_{k=1}^{K}\Big(\sum_{t\in\mathcal{T}^k}\mathbf{p}^t\Big)^\top\Big(|\mathcal{T}^k|^{-1}\sum_{\tau\in\mathcal{T}^k}\bar{\mathbf{y}}^\tau\Big)$$
$$=\sum_{k=1}^{K}\Big(|\mathcal{T}^k|\bar{\mathbf{p}}-\sum_{t\in\mathcal{T}^k}\mathbf{p}^t\Big)^\top\Big(|\mathcal{T}^k|^{-1}\sum_{\tau\in\mathcal{T}^k}\bar{\mathbf{y}}^\tau\Big)$$
$$=\sum_{k=1}^{K}\Big(|\mathcal{T}^k|\bar{\mathbf{p}}-\sum_{t\in\mathcal{T}^k}\mathbf{p}^t\Big)^\top\Big(|\mathcal{T}^k|^{-1}\sum_{\tau\in\mathcal{T}^k}(\bar{\mathbf{y}}^\tau-\bar{\mathbf{v}})\Big)$$
$$\geq-\sum_{k=1}^{K}|\mathcal{T}^k|\Big\|\bar{\mathbf{p}}-|\mathcal{T}^k|^{-1}\sum_{t\in\mathcal{T}^k}\mathbf{p}^t\Big\|_2\cdot O(\sqrt{m})$$

$$= -\sum_{k=1}^{K} |\mathcal{T}^k| \Big\| T^{-1} \sum_{t=1}^{T} \mathbf{p}^t - |\mathcal{T}^k|^{-1} \sum_{t \in \mathcal{T}^k} \mathbf{p}^t \Big\|_2 \cdot O(\sqrt{m}),$$

where the second inequality holds since $\|\bar{\mathbf{y}}^\tau - \bar{\mathbf{v}}\|_2 \le \sqrt{m}\bar{a} = O(\sqrt{m})$ from assumption **A2**. Note that for $1 \le i < j \le T$, we have $\|\mathbf{p}^i - \mathbf{p}^j\|_2 \le \sum_{t=i}^{j-1} \eta^t \|\hat{\mathbf{v}}^t - \hat{\mathbf{y}}^t\|_2 = \sum_{t=i}^{j-1} \eta^t \cdot O(\sqrt{m})$. Let $\mathbf{q}^*$ denote an optimal solution of (11). It then follows that

$$\begin{aligned}
\Big\| T^{-1} \sum_{t=1}^{T} \mathbf{p}^t - |\mathcal{T}^k|^{-1} \sum_{t \in \mathcal{T}^k} \mathbf{p}^t \Big\|_2 &= \Big\| \sum_{t \in \mathcal{T}} \sum_{i \in \mathcal{T}^k} q^*_{it} \mathbf{p}^t - \sum_{t \in \mathcal{T}^k} \sum_{j \in \mathcal{T}} q^*_{tj} \mathbf{p}^t \Big\| = \Big\| \sum_{i \in \mathcal{T}^k} \sum_{j \in \mathcal{T}} q^*_{ij} (\mathbf{p}^i - \mathbf{p}^j) \Big\| \\
&\le \sum_{(i,j) \in \mathcal{T}^k \times \mathcal{T} : i<j} q^*_{ij} \sum_{t=i}^{j-1} \eta^t \cdot O(\sqrt{m}) + \sum_{(i,j) \in \mathcal{T}^k \times \mathcal{T} : j<i} q^*_{ij} \sum_{t=j}^{i-1} \eta^t \cdot O(\sqrt{m}) \\
&\le \Big(\sum_{i \in \mathcal{T}^k} \sum_{j \in \mathcal{T}} d_{ij} q^*_{ij} \Big) \cdot O(\sqrt{m}) = w^k \cdot O(\sqrt{m}).
\end{aligned}$$

Therefore,

$$\sum_{t=1}^{T} (\mathbf{p}^t)^\top (\hat{\mathbf{v}}^t - \mathbb{E}_p[\bar{\mathbf{y}}^{p(t)}]) \ge -\sum_{k=1}^{K} |\mathcal{T}^k| w^k \cdot O(m) = -O(W). \tag{21}$$

Similar to (10), for $k = 1, \ldots, K$, we have

$$\sum_{t \in \mathcal{T}^k} \mathbb{E}_p[(\mathbf{p}^t)^\top (\mathbb{E}_p[\bar{\mathbf{y}}^{p(t)}] - \bar{\mathbf{y}}^{p(t)})] \ge -\sum_{i=1}^{|\mathcal{T}^k|} O\Big(\sqrt{\frac{m \log m}{i}}\Big) = -\sqrt{m \log m} \sqrt{|\mathcal{T}^k|}. \tag{22}$$

The conclusion follows by combining (20), (21) and (22). □

B.6 Proof of Corollary 4.6

Proof. Note that by the proof of Theorem 4.5, inequality $\sum_{t=1}^{T} (\mathbf{p}^t)^\top (\mathbf{v}^t - \mathbb{E}_p[\bar{\mathbf{y}}^{p(t)}]) \ge -O(W)$ holds deterministically for any grouped random permutation p. Following arguments similar to the proof of Corollary 4.4, with probability at least $1 - \rho/2$ we have

$$\begin{aligned}
\sum_{t=1}^{T} (\mathbf{p}^t)^\top \Big(\mathbb{E}_p[\bar{\mathbf{y}}^{p(t)}] - \mathbb{E}_p[\bar{\mathbf{y}}^{p(t)} | \mathcal{F}^{t-1}] \Big) &= \sum_{k=1}^{K} \sum_{t \in \mathcal{T}^k} (\mathbf{p}^t)^\top \Big(\mathbb{E}_p[\bar{\mathbf{y}}^{p(t)}] - \mathbb{E}_p[\bar{\mathbf{y}}^{p(t)} | \mathcal{F}^{t-1}] \Big) \\
&\ge -O\Big(\big(\sqrt{m \log m} + \sqrt{m \log(T/\rho)} \big) \sum_{k=1}^{K} \sqrt{\mathcal{T}^k} \Big),
\end{aligned} \tag{23}$$

and with probability at least $1 - \rho/2$ we have

$$\sum_{t=1}^{T} (\mathbf{p}^t)^\top \Big(\mathbb{E}_p[\bar{\mathbf{y}}^{p(t)} | \mathcal{F}^{t-1}] - \bar{\mathbf{y}}^{p(t)} \Big) \ge -O(\sqrt{m} \sqrt{T \log(1/\rho)}). \tag{24}$$

The conclusion follows by taking the union bound of (23) and (24). □

References

[AAP93] Baruch Awerbuch, Yossi Azar, and Serge Plotkin. Throughput-competitive on-line routing. In *Proceedings of 1993 IEEE 34th Annual Foundations of Computer Science*, pages 32–40. IEEE, 1993.

[AD14] Shipra Agrawal and Nikhil R Devanur. Fast algorithms for online stochastic convex programming. In *Proceedings of the 26th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1405–1424. SIAM, 2014.

[AE76] Jean-Pierre Aubin and Ivar Ekeland. Estimates of the duality gap in nonconvex optimization. *Mathematics of Operations Research*, 1(3):225–245, 1976.

[AWY14] Shipra Agrawal, Zizhuo Wang, and Yinyu Ye. A dynamic near-optimal algorithm for online linear programming. *Operations Research*, 62(4):876–890, 2014.

[Azu67] Kazuoki Azuma. Weighted sums of certain dependent random variables. *Tohoku Mathematical Journal, Second Series*, 19(3):357–367, 1967.

[BIKK07] Moshe Babaioff, Nicole Immorlica, David Kempe, and Robert Kleinberg. A knapsack secretary problem with applications. In *Approximation, randomization, and combinatorial optimization. Algorithms and techniques*, pages 16–28. Springer, 2007.

[BN09] Niv Buchbinder and Joseph Naor. Online primal-dual algorithms for covering and packing. *Mathematics of Operations Research*, 34(2):270–286, 2009.

[BT20] Yingjie Bi and Ao Tang. Duality gap estimation via a refined Shapley–Folkman lemma. *SIAM Journal on Optimization*, 30(2):1094–1118, 2020.

[Dyn63] Evgenii Borisovich Dynkin. The optimum choice of the instant for stopping a Markov process. *Soviet Mathematics*, 4:627–629, 1963.

[FHK+10] 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*, pages 182–194. Springer, 2010.

[GM08] Gagan Goel and Aranyak Mehta. Online budgeted matching in random input models with applications to Adwords. In *Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 982–991. Citeseer, 2008.

[GM16] Anupam Gupta and Marco Molinaro. How the experts algorithm can help solve LPs online. *Mathematics of Operations Research*, 41(4):1404–1431, 2016.

[Hoe63] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

[HUL13] Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Convex analysis and minimization algorithms I: Fundamentals*, volume 305. Springer science & business media, 2013.

[Kle05] Robert Kleinberg. A multiple-choice secretary algorithm with applications to online auctions. In *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 630–631. Citeseer, 2005.

[KTRV14] 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 46th Annual ACM Symposium on Theory of Computing*, pages 303–312, 2014.

[KVV90] Richard M Karp, Umesh V Vazirani, and Vijay V Vazirani. An optimal algorithm for on-line bipartite matching. In *Proceedings of the twenty-second annual ACM symposium on Theory of computing*, pages 352–358, 1990.

[LSY20] Xiaocheng Li, Chunlin Sun, and Yinyu Ye. Simple and fast algorithm for binary integer and online linear programming. *Advances in Neural Information Processing Systems*, 33:9412–9421, 2020.

[MR14] Marco Molinaro and Ramamoorthi Ravi. The geometry of online packing linear programs. *Mathematics of Operations Research*, 39(1):46–59, 2014.

[MSVV07] Aranyak Mehta, Amin Saberi, Umesh Vazirani, and Vijay Vazirani. Adwords and generalized online matching. *Journal of the ACM*, 54(5):22–es, 2007.

[PC+19] Gabriel Peyré, Marco Cuturi, et al. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.

[Pru98] Alexander Pruss. A maximal inequality for partial sums of finite exchangeable sequences of random variables. *Proceedings of the American Mathematical Society*, 126(6):1811–1819, 1998.

[Roc70] R Tyrrell Rockafellar. *Convex analysis*, volume 18. Princeton university press, 1970.

[UB16] Madeleine Udell and Stephen Boyd. Bounding duality gap for separable problems with linear constraints. *Computational Optimization and Applications*, 64(2):355–378, 2016.