Robust Budget Pacing with a Single Sample

Santiago Balseiro [1 2] **Rachitesh Kumar** [3] **Vahab Mirrokni** [2] **Balasubramanian Sivan** [2] **Di Wang** [2]

Abstract

Major Internet advertising platforms offer budget pacing tools as a standard service for advertisers to manage their ad campaigns. Given the inherent non-stationarity in an advertiser's value and also competing advertisers' values over time, a commonly used approach is to learn a target expenditure plan that specifies a target spend as a function of time, and then run a controller that tracks this plan. This raises the question: *how many historical samples are required to learn a good expenditure plan*? We study this question by considering an advertiser repeatedly participating in T second-price auctions, where the tuple of her value and the highest competing bid is drawn from an unknown time-varying distribution. The advertiser seeks to maximize her total utility subject to her budget constraint. Prior work has shown the sufficiency of $T \log T$ *samples per distribution* to achieve the optimal $O(\sqrt{T})$-regret. We dramatically improve this state-of-the-art and show that *just one sample per distribution* is enough to achieve the near-optimal $\tilde{O}(\sqrt{T})$-regret, while still being robust to noise in the sampling distributions.

1. Introduction

Online advertising is the economic engine of the internet. It allows platforms like Google, Facebook, and TikTok to fund services that are free at the point of delivery while providing businesses the ability to target their ads to relevant users. When a user logs onto these platforms, an auction is run among the interested advertisers to determine which ad is displayed to her. Given the scale of the internet, advertisers are typically interested in thousands (if not millions) of auctions every day. If they participate in all such auctions, they will likely spend far beyond their budget. This necessitates the need for budget management, which is our focus.

Target spend plans to address non-stationarities. Given the inherent non-stationarities that exist over time, in the volume of traffic, in the demographic of users that visit, the rates of conversion etc., an advertiser's own value and that of the competing advertisers' values are also non-stationary. To deal with this non-stationarity, budget management systems compute a target expenditure plan (Facebook-Guide; Kumar et al., 2022). The latter is a function of time that specifies the recommended amount of spend at each point of time. I.e., such a plan distributes the cumulative daily/weekly budget into smaller chunks of time, appropriately capturing the non-stationarity. A pacing algorithm like a controller is used to track the plan. This raises the question: how much data is required to come up with a good target spend plan?

Our model. We study this question by considering a budget-constrained advertiser participating in T second-price auctions. The advertiser seeks to maximize her utility while respecting her budget (formal model in Section 2). We model the tuple, of an advertiser's value and that of her highest competing bid, at time t, as being generated from unknown independent time-varying distributions $\{\mathcal{P}_t\}_{t=1}^T$. This is a major departure from the stochastic budget-management/resource-allocation literature, which for the most part assumes stationary distributions—an assumption that rarely holds in practice as daily traffic patterns are non-stationary (Zhou et al., 2019). If nothing is known about the distributions $\{\mathcal{P}_t\}_t$ and they can be chosen adversarially, Balseiro & Gur (2019) showed that every algorithm must incur linear regret against the hindsight-optimal benchmark. Their impossibility result also shows that budget management does not fall under the purview of online convex optimization (OCO) and is vastly different, given that one can achieve $O(\sqrt{T})$ regret for adversarially-chosen input in OCO. Thankfully, additional information is usually available in the form of historical samples. For example, both intraday and inter-day internet activity (and consequently the user traffic in which an advertiser is interested) tends to be similar week-on-week. However, this periodicity is never exact, i.e., it is too stringent to assume these historical samples are drawn from exactly the same distributions as $\{\mathcal{P}_t\}_t$. To reflect this reality, in this work, we assume that we have samples

Authors listed in alphabetical order. [1]DRO, Columbia Business School, New York, NY, USA [2]Google Research, New York, NY, USA [3]IEOR, Columbia University, New York, NY, USA. Work done as a Student Researcher at Google Research. Correspondence to: Rachitesh Kumar <rk3068@columbia.edu>.

from distributions $\{\tilde{\mathcal{P}}_t\}$, which are potentially different from $\{\mathcal{P}_t\}$, and develop algorithms that are robust to differences between them. We call an algorithm robust if its regret degrades smoothly as a function of the total Wasserstein distance between the two sequences of distributions. Crucially, classical training-based algorithms that learn a dual solution from samples and use it for bidding (see Algorithm 1, inspired by Devanur & Hayes 2009; Agrawal et al. 2014) are not robust (see Section 3).

Our result. Our main contribution is to show that Dual Follow-The-Regularized-Leader (FTRL) is robust and achieves the near optimal $\tilde{O}(\sqrt{T})$ regret even when it has access to *just one sample from each distribution* $\tilde{\mathcal{P}}_t$, dramatically improving over the prior $T \log T$ samples from each distribution (Jiang et al., 2020) to achieve $O(\sqrt{T})$ regret.

Key insights. Our algorithm uses the samples to estimate the ideal amount of expenditure to target in each auction and then uses Dual FTRL to follow these targets by shading the advertiser's values appropriately. The key insight driving the reduction in sample complexity from $T \log T$ to 1 per distribution is the following. Prior work by Jiang et al. (2020) first learns the sampling distributions $\{\tilde{\mathcal{P}}_t\}_t$ from the samples, then computes the optimal duals on the learned distributions, and finally uses those duals to compute the target expenditures. Our insight is that it is not necessary to learn the entire sampling distribution. Instead, it is far more efficient to directly learn the duals from the samples and construct target expenditure based on those duals (i.e., setting target expenditure at t to be what the dual-based solution consumes at t). Beyond being very efficient with samples, learning the duals from the samples also guarantees robustness to shifts between $\{\tilde{\mathcal{P}}_t\}_t$ and $\{\mathcal{P}_t\}_t$.

Practical implications. Consider T representing a week's worth of auctions. While a total of $T \log T$ samples may be possible to obtain by looking at the past few weeks (given that each week has about T samples to offer), requiring $T \log T$ samples *per distribution* calls for looking at many months into the past. This is because getting a sample for a distribution, where a distribution could correspond to say a particular hour (e.g. Monday 10 AM), entails looking at that same hour from the past week. Asking for $T \log T$ samples for any given hour, when T samples is what we get for the entire week, clearly requires looking at the numerous months into the past. Apart from posing huge storage and operational challenges, given that traffic pattern shifts over time, even gradual shifts would significantly degrade quality as one moves too much into the past. Our ask of one sample per distribution requires looking at just the past week and getting one sample for each hour.

Near optimality: $O(\sqrt{T})$ vs $\tilde{O}(\sqrt{T})$. Furthermore, our regret guarantee is near-optimal in light of the $\Omega(\sqrt{T})$ lower bound established in Arlotto & Gurvich (2019), which holds

even in the much easier case when the distributions $\{\mathcal{P}_t\}$ are identical (i.e., i.i.d. setting) and known ahead of time. Why is our $\tilde{O}(\sqrt{T})$ regret separated from the $O(\sqrt{T})$ regret in the lower bound (and the matching upper bound in Jiang et al. (2020))? Indeed, we can also obtain $O(\sqrt{T})$ regret when given $\log T$ samples per distribution, or in the other direction, the Jiang et al. (2020) paper also gets only $\tilde{O}(\sqrt{T})$ regret when given only T samples per distribution rather than $T \log T$ samples. In other words, we get a factor T reduction in samples: either from T to 1 sample while maintaining the same regret of $\tilde{O}(\sqrt{T})$ or from $T \log T$ to $\log T$ samples while maintaining the same regret of $O(\sqrt{T})$.

Primal vs dual approach. The recent works of Banerjee et al. (2020) and Banerjee & Freund (2020) show how to obtain a constant regret with a single historical trace by repeatedly solving the primal. We note that they crucially rely on the total number of types being small. When the number of types is very large as in Internet advertising where the type models user features, intent, time of day, location etc., a primal based algorithm is not even well defined because one would not have even seen all the types in the samples. Thus, a primal approach is out of question for us. More on this in Appendix A.

Technical contributions. We achieve our result by developing a novel dual-iterate coupling lemma (see Lemma 4.6) and leveraging it to analyze a leave-one-out thought experiment designed to break challenging correlations which arise from working with one sample per distribution (see Subsection 4.3 for details). Additionally, we also prove a novel regret decomposition for Dual FTRL (Theorem 4.1), which may be of independent interest. Finally, our algorithm does not require solving large linear programs and can be implemented efficiently (see Subsection 4.4), which is critical for online advertising since each auction runs in a few milliseconds. Due to space constraints, we directly move onto a formal description of our model next, and refer the reader to Appendix A for a discussion on other related work.

2. Model

Notation. We use $\mathbb{R}_+$ and $\mathbb{R}_{++}$ to denote the set of non-negative real numbers and the set of positive real numbers respectively. For $n \in \mathbb{N}$, we use $[n] = \{1, \ldots, n\}$ to denote the set of positive integers less than or equal to n. We use $\mathcal{W}(\cdot, \cdot)$ to denote the Wasserstein distance between two distributions under the metric with which the sample space is endowed.

Online Allocation with a Single Resource and Budget Management. For ease of exposition, we will prove our results for the more general single-resource online allocation problem with linear rewards/consumptions. It is well-known that bidding in repeated second-price auctions with budgets

can be modelled as an instance of this online allocation problem (e.g., see Balseiro et al. 2022b, or our Appendix B). It also captures the stochastic multi-secretary problem (Arlotto & Gurvich, 2019) as a special case.

Consider a decision maker with an initial budget $B \in \mathbb{R}_{++}$ of a resource, whose goal is to optimally spend it on T sequentially arriving requests. Each request $\gamma = (f, b)$ is comprised of a linear reward function $f : \mathcal{X} \to \mathbb{R}_+$ such that $f(x) = \mathrm{coeff}(f) \cdot x$, and a linear resource consumption function $b : \mathcal{X} \to \mathbb{R}_+$ such that $b(x) = \mathrm{coeff}(b) \cdot x$; where $\mathcal{X} \subseteq \mathbb{R}_+$ is a compact set which denotes the space of possible actions of the decision maker. We will use $\mathcal{S}$ to denote the set of all possible requests and $\Delta(S)$ to denote the set of distributions over $\mathcal{S}$. Moreover, we endow $\mathcal{S}$ with the following metric $d(\cdot, \cdot)$: For any two requests $\gamma = (f, b)$ and $\tilde{\gamma} = (\tilde{f}, \tilde{b})$:

$$d(\gamma, \tilde{\gamma}) = \sup_{x \in \mathcal{X}} \left| f(x) - \tilde{f}(x) \right| + \sup_{x \in \mathcal{X}} \left| b(x) - \tilde{b}(x) \right| .$$

We will assume that $0 \in \mathcal{X}$. This allows the decision maker to avoid spending the resource if she so chooses and ensures feasibility. Moreover, let $\bar{x} = \max_{x \in \mathcal{X}} x$. We will make standard regularity assumptions (Jiang et al., 2020; Balseiro et al., 2022a;b): there exist $\bar{f}, \bar{b} \in \mathbb{R}_+$ such that $f(x) \leq \bar{f}$ and $b(x) \leq \bar{b}$ for all $x \in \mathcal{X}$. Like Jiang et al. (2020), we will also assume that there exists $\kappa \in \mathbb{R}_+$ such that $f(x) \leq \kappa \cdot b(x)$ for all $x \in \mathcal{X}$, i.e., the maximum rate of return from spending the resource is bounded above by κ.

At time $t \in [T]$, the following sequence of events takes place: (i) a request $\gamma_t = (f_t, b_t)$ arrives; (ii) the decision maker observes γ_t and chooses an action $x_t \in \mathcal{X}$ based on the information observed so far; (iii) the request consumes $b_t(x_t)$ amount of the resource and generates a reward of $f_t(x_t)$. The decision maker aims to maximize her rewards subject to her budget constraint. A policy $\{x_t(\cdot)\}_t$ for the decision maker maps requests to actions $x_t : \mathcal{S} \to \mathcal{X}$ based on the available information at each time step, i.e., the action $x_t(\gamma_t)$ at time $t \in [T]$ can depend on the historical requests $\{\gamma_s\}_{s=1}^{t-1}$ and the current request γ_t, but not the future requests $\{\gamma_s\}_{s=t+1}^{T}$. Moreover, a policy is said to be budget-feasible if it respects the budget constraint by ensuring $\sum_{t=1}^{T} b_t(x_t(\gamma_t)) \leq B$ for every sequence $\{\gamma_t\}_t$.

The request γ_t at time t is drawn from a distribution $\mathcal{P}_t \in \Delta(\mathcal{S})$ unknown to the decision maker, independently of the requests at other time steps. We only require the requests $\{\gamma_t\}_t$ to be independent and allow the distributions $\mathcal{P}_t$ to vary arbitrarily across time. We will measure the performance of a policy against the fluid-optimal benchmark, which is defined as:

$$\mathrm{FLUID}(\{\mathcal{P}_t\}_t) \coloneqq \max \quad \sum_{t=1}^{T} \mathbb{E}[f_t(x_t(\gamma_t))]$$

$$\mathrm{s.t.} \quad \sum_{t=1}^{T} \mathbb{E}[b_t(x_t(\gamma_t))] \leq B$$

$$x_t : \mathcal{S} \to \mathcal{X} \quad \forall\, t \in [T] .$$

Another benchmark common in the literature on online resource allocation is the expected hindsight optimal solution, which is defined as $\mathbb{E}[\mathrm{OPT}(\{\gamma_t\}_t)]$ for

$$\mathrm{OPT}(\{\gamma_t\}_t) \coloneqq \max_{x \in \mathcal{X}^T} \sum_{t=1}^{T} f_t(x_t) \ \mathrm{s.t.} \ \sum_{t=1}^{T} b_t(x_t) \leq B .$$

It is well-known that $\mathrm{FLUID}(\{\mathcal{P}_t\}) \geq \mathbb{E}[\mathrm{OPT}(\{\gamma_t\}_t)]$, which makes our benchmark the stronger one (we provide a proof in Appendix C for completeness). Hence, our performance guarantees relative to the fluid-optimal benchmark also imply the same guarantees for the expected hindsight-optimal benchmark.

More concretely, we use $R(A|\{\gamma_t\}_t)$ to denote the total reward of a policy A on the request sequence $\{\gamma_t\}_t$, and the performance of an algorithm is measured using its expected regret against the fluid-optimal reward:

$$\mathrm{Regret}(A) \coloneqq \mathrm{FLUID}(\{\mathcal{P}_t\}_t) - \mathbb{E}\left[R(A|\{\gamma_t\}_t)\right] .$$

Now, if the distributions $\{\mathcal{P}_t\}_t$ are unknown and arbitrary, and no other information about $\{\mathcal{P}_t\}_t$ is available, then the requests $\{\gamma_t\}_t$ can be adversarial. This case has been addressed in Balseiro et al. (2022b), where the authors showed no policy can achieve sub-linear regret. In this work, we address the setting in which the decision maker has additional information in the form of historical samples. In particular, we focus on the setting where the decision maker has access to one independent sample $\tilde{\gamma}_t \sim \tilde{\mathcal{P}}_t$ for each $t \in [T]$. We will assume that the $\{\tilde{\gamma}_t\}$ samples are independent of the request sequence $\{\gamma_t\}_t$ and $\{\tilde{\mathcal{P}}_t\}$ are not known to the decision maker. We will show that when the sampling distributions $\{\tilde{\mathcal{P}}_t\}_t$ are not too far from the actual distributions $\{\mathcal{P}_t\}_t$, which is a minimal relaxation over the adversarial setting, it is possible to achieve sub-linear regret. We refer to the collection of samples $\{\tilde{\gamma}_t\}_t$ as a *trace* and allow the actions of the decision-maker to depend on it. Throughout this paper, we will use $\{\tilde{\gamma}_t\}_t$ to denote the trace and $\{\gamma_t\}_t$ to denote the (random) sequence of requests on which the decision maker wishes to maximize reward.

3. Warmup: Learning the Dual and Earning with It

First, let us focus on the simpler case when $\tilde{\mathcal{P}}_t = \mathcal{P}_t$ for all $t \in [T]$, i.e., the sampling distributions are the same as the request distributions. At first glance, it may appear that only having access to one sample from each of request

distributions $\mathcal{P}_t$ yields too little information to achieve near-optimal rewards. If one were to attempt to directly learn the optimal solution of $\text{FLUID}(\{\mathcal{P}_t\}_t)$, this initial impression would be accurate because of the high-dimensional nature of the space of all possible solutions $\{x_t(\cdot)\}_t$. Fortunately, we do not need to learn this high-dimensional information and can instead leverage the structure of the problem: the dual space is just one-dimensional and thereby amenable to learning. More precisely, the dual function $D(\mu|\{\mathcal{P}_t\}_t)$ of $\text{FLUID}(\{\mathcal{P}_t\}_t)$ at dual variable $\mu \geq 0$ is given by

$$\max_{\{x_t(\cdot)\}_t} \sum_{t=1}^{T} \mathbb{E}[f_t(x_t(\gamma_t))] + \mu\left(B - \sum_{t=1}^{T} \mathbb{E}[b_t(x_t(\gamma_t))]\right)$$

$$= \mu \cdot B + \sum_{t=1}^{T} \max_{x_t : \mathcal{S} \to \mathcal{X}} \mathbb{E}\left[f_t(x_t(\gamma_t)) - \mu \cdot b_t(x_t(\gamma_t))\right]$$

$$= \mu \cdot B + \sum_{t=1}^{T} \mathbb{E}\left[\max_{x_t \in \mathcal{X}} \{f_t(x_t) - \mu \cdot b_t(x_t)\}\right].$$

Throughout, we assume $\text{argmax}_{x \in \mathcal{X}} \{f(x) - \mu \cdot b(x)\}$ is non-empty for all requests $\gamma \in \mathcal{S}$ and dual solutions $\mu \geq 0$. If we treat the dual variable μ as the per-unit price of the resource, $\max_{x_t \in \mathcal{X}} \{f_t(x_t) - \mu \cdot b_t(x_t)\}$ captures the profit maximization problem. The following terminology would be helpful in working with the dual.

Definition 3.1. For a request $\gamma = (f, b)$ and dual variable $\mu \geq 0$, let $x^*(\gamma, \mu)$ be the optimal solution of $\max_{x \in \mathcal{X}} \{f(x) - \mu \cdot b(x)\}$ with the largest value of $f(x)$. If there are multiple such solutions, pick one which minimizes $b(x)$. Moreover, let $f^*(\mu) := f(x^*(\gamma, \mu))$ and $b^*(\mu) := b(x^*(\gamma, \mu))$ be the corresponding reward and resource consumption respectively.

We denote $D(\mu|\{\mathcal{P}_t\}_t) = \mu \cdot B + \sum_{t=1}^{T} \mathbb{E}[f_t^*(\mu) - \mu \cdot b_t^*(\mu)]$. Throughout this paper, we will repeatedly leverage weak duality, which is a central property of duals. We state the property here and refer the reader to any standard text on convex optimization (e.g., Bertsekas 2009) for a proof.

Proposition 3.2 (Weak Duality). *For all request distributions $\{\mathcal{P}_t\}_t$ and dual variables $\mu \geq 0$, we have $D(\mu|\{\mathcal{P}_t\}_t) \geq \text{FLUID}(\{\mathcal{P}_t\}_t)$, i.e.,*

$$\sum_{t=1}^{T} \mathbb{E}[f_t^*(\mu)] \geq \text{FLUID}(\{\mathcal{P}_t\}_t) - \mu \cdot \left(B - \sum_{t=1}^{T} \mathbb{E}[b_t^*(\mu)]\right).$$

Observe that $\sum_{t=1}^{T} \mathbb{E}[f_t^*(\mu)]$ is exactly the expected reward the decision maker would receive if she had an infinite budget and she took actions which maximized profit with μ being the per-unit price of the resource. Moreover, $B - \sum_{t=1}^{T} \mathbb{E}[b_t^*(\mu)]$ is the amount by which the decision maker would underspend her budget in expectation if she were to take actions using μ as the price. Suppose we can

find a dual variable $\mu \geq 0$ that satisfies approximate complementary slackness, i.e., μ satisfies at least one of the following statements: (1) $\mu = 0$ and maximizing profit with μ as the per-unit price results in total expenditure less than the budget B with high probability; (2) $\mu > 0$ and maximizing profit with μ as the per-unit price results in total expenditure close to the budget B with high probability. Then, if the decision maker were to use μ as the price and make decisions to maximize profit, she will not run out of budget too early and the complementary slackness term $\mu \cdot \left(B - \sum_{t=1}^{T} \mathbb{E}[b_t^*(\mu)]\right)$ would also be small. Therefore, such a μ would yield rewards that are close to $\text{FLUID}(\{\mathcal{P}_t\}_t)$, i.e., yield small regret, as required.

We next describe how such a μ can be learned from the sample trace $\{\tilde{\gamma}_t\}$ when $\tilde{\mathcal{P}}_t = \mathcal{P}_t$ for all $t \in [T]$. We will assume that the distributions satisfy the following mild and standard assumption (Devanur & Hayes, 2009; Agrawal et al., 2014) to exclude the degenerate case.

Assumption 3.3 (General Position). The request sequence $\{\gamma_t\}_t \sim \prod_t \mathcal{P}_t$ is in general position almost surely: For any $\mu \geq 0$, there is at most one request with multiple profit maximizers, i.e.,

$$\left|\left\{t \in [T] : |\text{argmax}_{x \in \mathcal{X}} \{f_t(x) - \mu \cdot b_t(x)\}| > 1\right\}\right| \leq 1.$$

Moreover, the sample trace $\{\tilde{\gamma}_t\}_t \sim \prod_t \tilde{\mathcal{P}}_t$ is also in general position almost surely.

Assumption 3.3 is made without any loss of generality because, as pointed out in Devanur & Hayes (2009) and Agrawal et al. (2014), adding an infinitesimally-small perturbation to the reward functions always results in perturbed distributions that satisfy Assumption 3.3 with only an infinitesimal change in the value of $\text{FLUID}(\{\mathcal{P}_t\}_t)$ (see Appendix D for a formal description). Assumption 3.3 ensures that there exists a dual solution $\tilde{\mu} \geq 0$ which spends close to the budget B on the trace $\{\tilde{\gamma}_t\}_t$ if it is possible to do so. In fact, as the following lemma shows, the optimal empirical dual solution satisfies this property.

Lemma 3.4. *Suppose the trace $\{\tilde{\gamma}\}_t \sim \prod_t \tilde{\mathcal{P}}_t$ is in general position, and consider*

$$\tilde{\mu} \in \text{argmin}_{\mu \geq 0} \left\{\mu \cdot B + \sum_{t=1}^{T} \max_{x \in \mathcal{X}} \left\{\tilde{f}_t(x) - \mu \cdot \tilde{b}_t(x)\right\}\right\}.$$

Then, at least one of the following statements holds:

1. *$\tilde{\mu} = 0$ and $\sum_{t=1}^{T} \tilde{b}_t^*(\tilde{\mu}) \leq B + \bar{b}$.*

2. *$\left|B - \sum_{t=1}^{T} \tilde{b}_t^*(\tilde{\mu})\right| \leq \bar{b}$.*

Recall that weak duality (Proposition 3.2) suggests that finding a dual solution which satisfies approximate complementary slackness with high probability would yield

Algorithm 1 Learning the Dual and Earning with It

Input: Trace $\{\tilde{\gamma}_t\} \sim \prod_t \tilde{\mathcal{P}}_t$, initial budget $B_1 = B$.
Compute an Optimal Empirical Dual Solution:

$$\tilde{\mu} \in \underset{\mu \geq 0}{\operatorname{argmin}} \left\{ \mu \cdot B + \sum_{t=1}^{T} \max_{x \in \mathcal{X}} \left\{ \tilde{f}_t(x) - \mu \cdot \tilde{b}_t(x) \right\} \right\} \quad (1)$$

for $t = 1, \ldots, T$ **do**

    Receive request $\gamma_t = (f_t, b_t) \sim \mathcal{P}_t$.
    Make the primal decision x_t and update the remaining resources B_t:

$$x'_t \in \underset{x \in \mathcal{X}}{\operatorname{argmax}} \left\{ f_t(x) - \tilde{\mu} \cdot b_t(x) \right\},$$

$$x_t = \begin{cases} x'_t & \text{if } b_t(x'_t) \leq B_t \\ 0 & \text{otherwise} \end{cases},$$

$$B_{t+1} = B_t - b_t(x_t).$$

end

reward close to $\text{FLUID}(\{\mathcal{P}_t\})$. Lemma 3.4 states that we can compute a dual variable $\tilde{\mu}$ which satisfies approximate complementary slackness on the trace. To finish the argument, we require a uniform convergence bound which shows that expenditure on the trace (or the sequence of requests) is concentrated close to the expected expenditure for all dual variables $\mu \geq 0$.

Theorem 3.5. *For $r(T) := 8\bar{b} \cdot \sqrt{T \log(T)}$ and request distributions $\{\tilde{\mathcal{P}}_t\}_t$, the following uniform convergence bound holds*

$$\Pr \left(\sup_{\mu \geq 0} \left| \sum_{t=1}^{T} \tilde{b}_t^*(\mu) - \sum_{t=1}^{T} \mathbb{E}_{\tilde{\gamma}_t \sim \tilde{\mathcal{P}}_t} \left[\hat{b}_t^*(\mu) \right] \right| \geq r(T) \right) \leq \frac{1}{T^2}.$$

With Theorem 3.5 in hand, we are now ready to state and prove the regret guarantee for Algorithm 1. It first learns an empirical optimal dual variable $\tilde{\mu}$ from the trace $\{\tilde{\gamma}_t\}_t$, and then uses it as the per-unit price of the resource to take profit-maximizing actions on the request sequence $\{\gamma_t\}_t$.

Theorem 3.6. *If $\mathcal{P}_t = \tilde{\mathcal{P}}_t$ for all $t \in [T]$, then Algorithm 1 (denoted by A) satisfies $\text{Regret}(A) \leq 12\kappa\bar{b} + 2\kappa r(T)$.*

Arlotto & Gurvich (2019) showed that every algorithm must incur a regret of $\Omega(\sqrt{T})$, even when the request distributions are identical (i.e., $\mathcal{P}_t = \mathcal{P}$ for all $t \in [T]$) and known to the decision-maker ahead of time. Thus, Theorem 3.6 shows the regret of Algorithm 1 achieves a near-optimal dependence on T with just a single sample per distribution, despite the request distributions being unknown and time-varying. However, as the following example demonstrates, this regret bound critically relies on the assumption that $\mathcal{P}_t = \tilde{\mathcal{P}}_t$ for all $t \in [T]$, and is fragile to even slight deviations from it.

This fact was also demonstrated in Jiang et al. (2020) in a related context which inspired the following example.

Example. Fix a small $\epsilon > 0$, an even horizon T and budget $B = T/2$. Assume actions are accept/reject decisions, i.e., $\mathcal{X} = \{0, 1\}$, and the reward/resource consumption functions are linear with $\text{coeff}(b) = 1$ for all $\gamma = (f, b) \in \mathcal{S}$. In this setting, a request is completely determined by the coefficient $\text{coeff}(f)$ of its reward function. We will overload notation and use γ to denote this coefficient. Set $\tilde{\mathcal{P}}_t = \text{Unif}([1 + \epsilon, 1 + 2\epsilon])$ for all $t \leq T/2 + 1$ and $\tilde{\mathcal{P}}_t = \text{Unif}([1 - \epsilon, 1])$ for all $t \geq T/2 + 2$. Moreover, set $\mathcal{P}_t = \text{Unif}([1 - \epsilon, 1])$ for all $t \in [T]$. Then, it is easy to see that $\mathcal{W}(\mathcal{P}_t, \tilde{\mathcal{P}}_t) \leq 3\epsilon$ for all $t \in [T]$. Also, observe that any trace $\{\tilde{\gamma}_t\}_t \sim \prod_t \tilde{\mathcal{P}}_t$ would satisfy $\tilde{\gamma}_t \geq 1 + \epsilon$ for all $t \leq T/2 + 1$ and $\tilde{\gamma}_t \leq 1$ for all $t \geq T/2 + 2$. Hence, we always have $\tilde{\mu} \geq 1 + \epsilon$. On the other hand, we also always have $\gamma_t \leq 1$ for all $t \in [T]$. Therefore, Algorithm 1 sets $x'_t = 0$ for all $t \in [T]$, yielding a reward of 0. Whereas, $\text{FLUID}(\{\mathcal{P}_t\}_t) \geq (1 - \epsilon) \cdot (T/2)$, thereby making the regret linear in T.

Since $\epsilon > 0$ was arbitrary in the above example, it shows that even infinitesimally-small differences between the sampling and request distributions can lead to linear regret for Algorithm 1. This is antithetical to our goal of developing robust online algorithms for pacing. Formally, we would like to develop online algorithms that achieve regret which is small and degrades smoothly as $\sum_{t=1}^{T} \mathcal{W}(\mathcal{P}_t, \tilde{\mathcal{P}}_t)$ grows large. Nonetheless, although Algorithm 1 falls short of this goal, it highlights the power of dual-based algorithms. Building on the intuition developed in this section, we next describe and analyze a Dual FTRL algorithm that achieves near-optimal regret while being robust to discrepancies between the sampling distributions $\{\tilde{\mathcal{P}}_t\}_t$ and the request distributions $\{\mathcal{P}_t\}_t$.

4. Dual FTRL with Target Rate Estimation

In this section, we will develop an algorithm based on Dual Follow-The-Regularized-Leader (FTRL) that achieves near-optimal regret with a single trace, and is robust to discrepancies between the sampling distributions and request distributions. Now, if one had complete knowledge of the sampling distributions $\{\tilde{\mathcal{P}}_t\}_t$, then one can solve $\text{FLUID}(\{\tilde{\mathcal{P}}_t\}_t)$ to find an optimal solution and run Dual Gradient Descent with the goal of spending the same as the optimal solution at each time step. It is known from Jiang et al. (2020) that this approach achieves $O(\max\{\sqrt{T}, \sum_{t=1}^{T} \mathcal{W}(\tilde{\mathcal{P}}_t, \mathcal{P}_t)\})$ regret, thereby making it rate optimal and robust to discrepancies. However, with just a single sample from each of distributions $\tilde{\mathcal{P}}_t$, we are far from having complete knowledge of $\{\tilde{\mathcal{P}}_t\}_t$. Despite this apparent lack of data, a careful analysis of Dual FTRL will allow us to show that it achieves near-optimal regret rate in a robust manner.

Algorithm 2 Dual Follow-The-Regularized-Leader

Input: Initial resource endowment $B_1 = B$, target consumption sequence $\{\lambda_t\}_{t=1}^T$, regularizer $h : \mathbb{R} \to \mathbb{R}$ and step-size η.

Set initial dual solution $\mu_1 = \mathrm{argmin}_{\mu \in [0,\kappa]} h(\mu)$.

for $t = 1, \ldots, T$ **do**

Receive request $\gamma_t = (f_t, b_t) \sim \mathcal{P}_t$.

Make the primal decision x_t and update the remaining resources B_t:

$$x_t' \in \underset{x \in \mathcal{X}_t}{\mathrm{argmax}} \left\{ f_t(x) - \mu_t \cdot b_t(x) \right\} , \qquad (2)$$

$$x_t = \begin{cases} x_t' & \text{if } b_t(x_t') \le B_t \\ 0 & \text{otherwise} \end{cases} ,$$

$$B_{t+1} = B_t - b_t(x_t).$$

Obtain a sample sub-gradient of the dual function $D(\mu|\mathcal{P}_t, \lambda_t)$: $g_t = \lambda_t - b_t(x_t')$.

Update the dual iterate with FTRL:

$$\mu_{t+1} = \underset{\mu \in [0,\kappa]}{\mathrm{argmin}} \left\{ \eta \sum_{r=1}^t g_r \cdot \mu + h(\mu) \right\} , \qquad (3)$$

end

4.1. Dual Follow-The-Regularized-Leader

The non-stationarity of the request distributions necessitates the need for Dual FTRL that can incorporate target resource consumptions (Algorithm 2). It takes as input a target sequence $\{\lambda_t\}_{t=1}^T$ which specifies $\lambda_t \ge 0$ to be the amount of resource Dual FTRL should attempt to consume at time t. Moreover, like FTRL (Shalev-Shwartz et al., 2012; Hazan et al., 2016), it also takes as input a regularizer $h(\cdot)$, an initial dual variable μ_1 and a step-size η. We will make the standard assumption that the regularizer $h(\cdot)$ is differentiable and is σ-strongly convex in the $\|\cdot\|_1$ norm.

Before stating the performance bound of Algorithm 2, we introduce some preliminaries. Given a budget of β_t for period $t \in [T]$, the optimal expected reward which can be collected in period t is captured by the following fluid optimization problem:

$$\mathrm{FLUID}(\mathcal{P}_t, \beta_t) := \max \quad \mathbb{E}[f_t(x_t(\gamma_t)]$$
$$\text{s.t.} \quad \mathbb{E}[b_t(x_t(\gamma_t))] \le \beta_t$$
$$x_t : \mathcal{S} \to \mathcal{X} .$$

The dual function of $\mathrm{FLUID}(\mathcal{P}_t, \beta_t)$ is given by

$$D(\mu|\mathcal{P}_t, \beta_t) := \mu \cdot \beta_t + \mathbb{E}\left[\max_{x \in \mathcal{X}} \left\{ f_t(x) - \mu \cdot b_t(x) \right\} \right] ,$$

for any $\mu \ge 0$. Then, by weak duality, we have

$\mathrm{FLUID}(\mathcal{P}_t, \beta_t) \le D(\mu|\mathcal{P}_t, \beta_t)$ for all $\mu \ge 0$. Moreover, since dual functions are always convex (they are the suprema of linear functions), the dual function $D(\cdot|\mathcal{P}_t, \beta_t)$ is convex.

Theorem 4.1 states a general regret bound for Algorithm 2 with an arbitrary target sequence $\{\lambda_t\}_t$ and against a general benchmark $\sum_{t=1}^T D(\mu_t|\mathcal{P}_t, \beta_t)$. Since $\mathrm{FLUID}(\mathcal{P}_t, \beta_t) \le D(\mu|\mathcal{P}_t, \beta_t)$ by weak duality, Theorem 4.1 also characterizes the performance against the weaker benchmark $\sum_{t=1}^T \mathrm{FLUID}(\mathcal{P}_t, \beta_t)$, which is simply the optimal expected reward the decision maker would collect if she spent β_t at time t.

Theorem 4.1. *Consider Algorithm 2 with target consumption sequence $\{\lambda_t\}_t$, regularizer $h(\cdot)$ and step-size η. Then, for a benchmark sequence $\{\beta_t\}_t$, we have*

$$\mathbb{E}\left[\left\{ \sum_{t=1}^T D(\mu_t|\mathcal{P}_t, \beta_t) \right\} - R(A|\{\gamma_t\}_t) \right] \le R_1 + R_2 + R_3 ,$$

where

- $R_1 = \kappa \bar{b} + \frac{2(\bar{b}+\bar{\lambda})^2}{\sigma} \cdot \eta T + \frac{d_R}{\eta}$, *for* $\bar{\lambda} = \max_t \lambda_t$ *and* $d_R = \max\{h(0) - h(\mu_1), h(\kappa) - h(\mu_1)\}$.

- $R_2 = \kappa \cdot \left(\left\{ \sum_{t=1}^T \lambda_t \right\} - B \right)^+$,

- $R_3 = \mathbb{E}\left[\sum_{t=1}^T \mu_t \cdot (\beta_t - \lambda_t) \right]$.

Theorem 4.1 decomposes the regret of Algorithm 2 into three terms, where (i) R_1 is simply the regret associated with the FTRL algorithm in the OCO setting (Hazan et al., 2016); (ii) R_2 captures the overspending error, which is large whenever the total target consumption $\sum_{t=1}^T \lambda_t$ is in excess of the budget B; (iii) R_3 captures the underestimation error, which is a weighted sum over the amounts by which the target sequence $\{\lambda_t\}_t$ underestimates the benchmark sequence $\{\beta_t\}_t$, with weights equal to the dual iterates μ_t. Observe that there is an inherent tension between the overspending error R_2 and the underestimation error R_3— R_2 can be made smaller by making the target consumptions $\{\lambda_t\}_t$ smaller, but this in turn makes R_3 bigger, and vice versa. To obtain the desired performance guarantees for Algorithm 2 (see Theorem 4.8), we need to carefully choose the benchmark sequence $\{\beta_t\}_t$ and the target sequence $\{\lambda_t\}_t$, which is what we do next (see (4)). We go on to show that

- Our choice of target sequence does not overspend too much. In particular, it satisfies $R_2 \le \kappa \cdot \bar{b}$ (see (5)).

- Our choice of benchmark sequence $\{\beta_t\}_t$ ensures

$$\mathrm{FLUID}(\{\mathcal{P}_t\}_t) - \sum_{t=1}^T D(\mu_t|\mathcal{P}_t, \beta_t)$$

$$\leq \tilde{O}\left(\max\left\{ \sqrt{T}, \sum_{t=1}^{T} \mathcal{W}(\mathcal{P}_t, \tilde{\mathcal{P}}_t) \right\} \right)$$

i.e., the benchmark in Theorem 4.1 is at most $\tilde{O}\left(\max\left\{ \sqrt{T}, \sum_{t=1}^{T} \mathcal{W}(\mathcal{P}_t, \tilde{\mathcal{P}}_t) \right\} \right)$ larger than our desired benchmark FLUID($\{\mathcal{P}_t\}_t$) (see Lemma 4.2 and the discussion that follows).

- Moreover, our choice of the sequences in combination with an intricate argument, consisting of a coupling lemma and a leave-one-out thought experiment, allows use to prove $R_3 = O(\sqrt{T})$ (see Subsection 4.3).

Finally in Subsection 4.4, we combine everything to prove the desired regret bound for Algorithm 2.

4.2. Choosing the Target and Benchmark Sequences

We define the target and benchmark sequences using the empirical optimal dual solution computed from the trace $\{\tilde{\gamma}_t\}_t$:

$$\tilde{\mu} \in \underset{\mu \geq 0}{\operatorname{argmin}} \left\{ \mu \cdot B + \sum_{t=1}^{T} \max_{x \in \mathcal{X}} \left\{ \tilde{f}_t(x) - \mu \cdot \tilde{b}_t(x) \right\} \right\}.$$

If there are multiple minimizers, set $\tilde{\mu}$ to be the smallest one. Given the empirical optimal dual solution $\tilde{\mu}$, the target and benchmark sequences are defined as

$$\beta_t = \mathbb{E}_{\hat{\gamma}_t \sim \tilde{\mathcal{P}}_t}\left[\hat{b}_t^*(\tilde{\mu}) \right] \quad \text{and} \quad \lambda_t = \tilde{b}_t^*(\tilde{\mu}), \qquad (4)$$

where $\hat{\gamma}_t = (\hat{f}_t, \hat{b}_t)$. In other words, the benchmark sequence is the expected consumption and the target sequence is the empirical consumption on the trace if we were to make profit-maximizing decisions using the empirical optimal dual solution $\tilde{\mu}$ as the price of the resource. Instead of learning the empirical optimal dual $\tilde{\mu}$ and directly making decisions with it like we did in Algorithm 1, we use $\tilde{\mu}$ to learn the empirical consumptions $\{\lambda_t\}_t$ and use Algorithm 2 to track this target. Unlike the former, we will show that the latter approach is robust to discrepancies between the sampling and request distributions, while maintaining the same $\tilde{O}(\sqrt{T})$-regret guarantee. Importantly, note that the benchmark sequence $\{\beta_t\}_t$ cannot be computed in practice because it requires full knowledge of the request distributions. Algorithm 2 respects this limitation and does not require knowledge of the benchmark sequence $\{\beta_t\}_t$; we only use it for our analysis.

Our choice of $\{\lambda_t\}_t$ and Lemma 3.4 immediately imply

$$R_2 = \kappa \cdot \left(\left\{ \sum_{t=1}^{T} \lambda_t \right\} - B \right)^+ \leq \kappa \cdot \bar{b}. \qquad (5)$$

Next, we show that, for our choice of the benchmark sequence $\{\beta_t\}_t$, the benchmark $\sum_{t=1}^{T} D(\mu_t | \mathcal{P}_t, \beta_t)$ of

Theorem 4.1 is not too far from the desired benchmark FLUID($\{\mathcal{P}_t\}$).

Lemma 4.2. *For any dual variable $\tilde{\mu} \geq 0$, dual iterates $\{\mu_t\}_t \in [0, \kappa]^T$ and benchmark sequence $\{\beta_t\}_t$ with $\beta_t = \mathbb{E}_{\hat{\gamma} \sim \tilde{\mathcal{P}}_t}[\hat{b}^*(\tilde{\mu})]$, we have*

$$\sum_{t=1}^{T} D(\mu_t | \mathcal{P}_t, \beta_t) \geq \text{FLUID}(\{\mathcal{P}_t\}_t) - \tilde{\mu} \cdot \left(B - \sum_{t=1}^{T} \beta_t \right)$$
$$- 2(1 + \kappa) \cdot \sum_{t=1}^{T} \mathcal{W}(\mathcal{P}_t, \tilde{\mathcal{P}}_t).$$

Observe that Theorem 3.5 implies that, with probability at least $1 - 1/T^2$, we have $\left| \sum_{t=1}^{T} \beta_t - \sum_{t=1}^{T} \lambda_t \right| \leq r(T)$. Combining this with Lemma 3.4 yields

$$\tilde{\mu} \cdot \left(B - \sum_{t=1}^{T} \beta_t \right) \leq \tilde{\mu} \cdot \left(r(T) + B - \sum_{t=1}^{T} \lambda_t \right)$$
$$= \tilde{\mu} \cdot r(T) + \tilde{\mu} \cdot \left(B - \sum_{t=1}^{T} \lambda_t \right)$$
$$\leq \tilde{\mu} \cdot \left(r(T) + \bar{b} \right)$$
$$\leq \kappa \cdot r(T) + \kappa \bar{b}, \qquad (6)$$

thereby showing that the benchmark $\sum_{t=1}^{T} D(\mu_t | \mathcal{P}_t, \beta_t)$ of Theorem 4.1 is not too far from the desired benchmark FLUID($\{\mathcal{P}_t\}$). In order to establish the desired regret and robustness guarantees for Algorithm 2, all that remains to show is that $R_3 \leq \tilde{O}(\sqrt{T})$. However, as we demonstrate in the next subsection, this step is rife with challenges.

4.3. Bounding R_3

We begin with a brief discussion of the challenges involved in bounding R_3. It is illuminating to consider the slightly more permissive setting in which the decision maker has access to two sample traces: suppose in addition to trace $\{\tilde{\gamma}_t\}_t \sim \prod_t \tilde{\mathcal{P}}_t$, we had access to an additional trace $\{\tilde{\tilde{\gamma}}_t\}_t \sim \prod_t \tilde{\mathcal{P}}_t$. Then, we could compute $\tilde{\mu}$ using $\{\tilde{\tilde{\gamma}}_t\}_t$ as follows

$$\tilde{\mu} \in \underset{\mu \geq 0}{\operatorname{argmin}} \left\{ \mu \cdot B + \sum_{t=1}^{T} \max_{x \in \mathcal{X}} \left\{ \tilde{f}_t(x) - \mu \cdot \tilde{b}_t(x) \right\} \right\},$$

making it completely independent of $\{\tilde{\gamma}_t\}_t$. With this modified $\tilde{\mu}$, we continue to define $\{\beta_t\}_t, \{\gamma_t\}_t$ as before (see (4)). As a consequence, we get that μ_s is completely determined by $\{\gamma_t\}_{t=1}^{s-1}$ and $\{\lambda_t\}_{t=1}^{s-1}$, with the latter being completely determined by $\tilde{\mu}$ and $\{\tilde{\gamma}_t\}_{t=1}^{s-1}$. This makes μ_s independent of λ_s conditional on $\tilde{\mu}$, and consequently yields

$$\mathbb{E}\left[\mu_s \cdot (\beta_s - \lambda_s) \mid \tilde{\mu}, \{\tilde{\gamma}_t, \gamma_t\}_{t=1}^{s-1} \right] = \mu_s \cdot \left(\beta_s - \mathbb{E}\left[\tilde{b}_s^*(\tilde{\mu}) \mid \tilde{\mu} \right] \right)$$
$$= 0.$$

Thus, we can apply the Tower Rule of conditional expectations to get

$$R_3 = \sum_{s=1}^{T} \mathbb{E}\left[\mu_s \cdot (\beta_s - \lambda_s)\right]$$

$$= \sum_{s=1}^{T} \mathbb{E}\left[\mathbb{E}\left[\mu_s \cdot (\beta_s - \lambda_s)\big|\, \tilde{\mu}, \{\tilde{\gamma}_t, \gamma_t\}_{t=1}^{s-1}\right]\right]$$

$$= 0\,.$$

It is straightforward to see that the bounds on R_1 and R_2 established in the previous subsection continue to hold in this two-trace setting. Therefore, two traces allow us to achieve the near-optimal $\tilde{O}(\sqrt{T})$-regret while being robust to discrepancies between $\tilde{\mathcal{P}}_t$ and $\mathcal{P}_t$.

Although moving from two traces to one trace might appear to be a minor change, it introduces correlations that make the proof much more difficult. Observe that Algorithm 2 determines μ_s using $\{\lambda_t\}_{t=1}^{s-1}$, all of which depend on $\tilde{\mu}$, which in turn is computed using the request $\tilde{\gamma}_s$. Furthermore, λ_s directly depends on $\tilde{\gamma}_s$. Thus, μ_s and λ_s are intricately correlated with each other, which breaks the aforementioned argument for the two-trace setting. Nonetheless, R_3 can still be shown to be small, as we note in the following lemma and prove in the remainder of this subsection.

Lemma 4.3. *For all $s \in [T]$, we have*

$$R_3 = \sum_{s=1}^{T} \mathbb{E}\left[\mu_s \cdot (\beta_s - \lambda_s)\right] \leq \frac{4\eta\bar{b}^2}{\sigma} \cdot T\quad.$$

We prove Lemma 4.3 in the remainder of this subsection. The following lemma will find repeated use in the proof. In keeping with economic intuition, it shows that increasing the price (dual variable) leads to smaller consumption under the profit-maximzing decision.

Lemma 4.4 (Monotonicity). *For $\mu > \mu'$, request $\gamma = (f, b) \in \mathcal{S}$, $x \in \operatorname{argmax}_{z\in\mathcal{X}}\{f(z) - \mu \cdot b(z)\}$ and $x' \in \operatorname{argmax}_{z\in\mathcal{X}}\{f(z) - \mu' \cdot b(z)\}$, we have $b(x) \leq b(x')$.*

Fix an $s \in [T]$. We will get around the correlation between μ_s and $\tilde{\gamma}_s$ by conducting the following leave-one-out thought experiment: suppose we remove the s-th sample $\tilde{\gamma}_s$, compute $\tilde{\mu}$ on the remaining trace $\{\tilde{\gamma}_t\}_{t\neq s}$, and run Algorithm 2 with the resulting target sequence. More precisely, in this thought experiment, we set $\tilde{\mathcal{P}}_s$ to be the distribution which always serves the request $\gamma = (f, b)$ with $f(x) = b(x) = 0$ for all $x \in \mathcal{X}$. Thus, $\tilde{f}_s(x) = \tilde{b}_s(x) = \tilde{f}^*(\mu) = \tilde{b}^*(\mu) = 0$ for all $x \in \mathcal{X}$ and $\mu \geq 0$. We will use the superscript $(-s)$ to denote the various variables in this thought experiment:

- $\tilde{\mu}^{(-s)} \in \operatorname{argmin}_{\mu\geq 0} \mu \cdot B + \sum_{t\neq s} \max_{x\in\mathcal{X}}\{\tilde{f}_t(x) - \mu \cdot \tilde{b}_t(x)\}$. If there are multiple minimizers, set $\tilde{\mu}^{(-s)}$ to be the smallest one amongst them.

- $\lambda_t^{(-s)} = \tilde{b}_t^*\left(\tilde{\mu}^{(-s)}\right)$ for all $t \in [T]$.

- $\mu_t^{(-s)}$ is the t-th iterate of Algorithm 2 with the target consumption sequence $\left\{\lambda_t^{(-s)}\right\}_t$.

We begin by characterizing the impact of this change on the target consumption sequence.

Lemma 4.5. *For every sample trace $\{\tilde{\gamma}_t\}_t$, we have $\tilde{\mu} \geq \tilde{\mu}^{(-s)}$ and $\lambda_t \leq \lambda_t^{(-s)}$ for all $t \neq s$. Moreover, $\sum_{t=1}^{s-1}\left|\lambda_t^{(-s)} - \lambda_t\right| \leq 3\bar{b}$.*

Lemma 4.5 shows that the target sequences $\{\lambda_t\}_t$ and $\{\lambda_t^{(-s)}\}_t$ are close to each other. Next, we couple the dual iterates μ_t and $\mu_t^{(-s)}$ generated by Algorithm 2 to show that they never stray too far from each other whenever the target sequences are close.

Lemma 4.6 (Dual Iterate Coupling). *Let $\{\mu_t\}_t$ and $\{\mu_t'\}_t$ denote the iterates generated by Algorithm 2 on the request sequence $\{\gamma_t\}_t$ for the target sequences $\{\lambda_t\}_t$ and $\{\lambda_t'\}_t$ respectively. Assume that the initial iterates are the same, i.e., $\mu_1 = \mu_1'$. Then, for all $s \in [T]$, we have*

$$|\mu_s - \mu_s'| \leq \frac{\eta}{\sigma}\cdot\left\{\sum_{t=1}^{s-1}|\lambda_t - \lambda_t'|\right\} + \frac{\eta}{\sigma}\cdot\bar{b}\,.$$

Applying Lemma 4.6 with $\lambda_t' = \lambda_t^{(-s)}$ and using Lemma 4.5 yields $|\mu_s - \mu_s'| \leq \frac{\eta}{\sigma}\cdot\{3\bar{b}\} + \frac{\eta}{\sigma}\cdot\bar{b} = \frac{4\eta\bar{b}}{\sigma}$.

Combining this with the fact that $|\beta_s - \lambda_s| \leq \bar{b}$, we get

$$\mathbb{E}\left[\mu_s \cdot (\beta_s - \lambda_s)\right] = \mathbb{E}\left[\left(\mu_s - \mu_s^{(-s)}\right)\cdot(\beta_s - \lambda_s)\right] + \mathbb{E}\left[\mu_s^{(-s)}\cdot(\beta_s - \lambda_s)\right]$$

$$\leq \frac{4\eta\bar{b}}{\sigma}\cdot\bar{b} + \mathbb{E}\left[\mu_s^{(-s)}\cdot(\beta_s - \lambda_s)\right]\,. \qquad (7)$$

The next lemma shows that the second term is non-positive. Its proof critically leverages the fact that the iterate $\mu_s^{(-s)}$ is independent of the s-th sample in the trace $\tilde{\gamma}_s$ (which is used to determine λ_s). This is in stark contrast to μ_s which depends on $\tilde{\gamma}_s$, and demonstrates the merit of our leave-one-out thought experiment.

Lemma 4.7. $\mathbb{E}\left[\mu_s^{(-s)}\cdot(\beta_s - \lambda_s)\right] \leq 0$ *for all $s \in [T]$.*

Lemma 4.7 in combination with (7) yields $\mathbb{E}\left[\mu_s \cdot (\beta_s - \lambda_s)\right] \leq 4\eta\bar{b}^2/\sigma$. Summing over all $s \in [T]$ finishes the proof of Lemma 4.3.

4.4. Putting It All Together

In the previous subsections, we bounded R_1, R_2 and R_3, and related the benchmark $\sum_{t=1}^{T} D(\mu_t|\mathcal{P}_t, \beta_t)$ from Theorem 4.1 to our desired benchmark $\text{FLUID}(\{\mathcal{P}_t\}_t)$. Combining everything yields the following performance gaurantee for Algorithm 2.

Theorem 4.8. *Let A be Algorithm 2 with target sequence $\{\lambda_t\}_t$, where $\lambda_t = \tilde{b}_t^*(\tilde{\mu})$ (as defined in (4)), regularizer $h(\cdot)$ and step-size $\eta = \sqrt{d_R/T}$, where $d_R = \max\{h(0) - h(\mu_1), h(\kappa) - h(\mu_1)\}$. Then,*

$$\text{Regret}(A) \leq C_1 \sqrt{T \log(T)} + C_2 \sum_{t=1}^{T} \mathcal{W}(\mathcal{P}_t, \tilde{\mathcal{P}}_t).$$

where $C_1 = \frac{12\bar{b}^2 \sqrt{d_R}}{\sigma} + \sqrt{d_R} + 12\kappa\bar{b}$ and $C_2 = 2(1 + \kappa)$.

Observe that the regret of Dual FTRL satisfies $\text{Regret}(A) = \tilde{O}(\sqrt{T})$ whenever $\sum_{t=1}^{T} \mathcal{W}(\mathcal{P}_t, \tilde{\mathcal{P}}_t) = \tilde{O}(\sqrt{T})$. In other words, Dual FTRL achieves near-optimal regret with a single trace as long as the total discrepancy $\sum_{t=1}^{T} \mathcal{W}(\mathcal{P}_t, \tilde{\mathcal{P}}_t)$ is not too large. Finally, we would also like to note that our algorithm is extremely efficient computationally. In particular, due to the equivalence of FTRL and "Lazy" Online Mirror Descent (OMD) (see Hazan et al. 2016), each dual update in (3) can be computed in constant time by running Lazy OMD. Moreover, given a trace $\{\tilde{\gamma}_t\}$ which is sorted in increasing order of bang-per-buck $\text{coeff}(\tilde{f}_t)/\text{coeff}(\tilde{b}_t)$, the target sequence $\{\lambda_t\}_t$ can be computed in $O(T)$ steps (see Appendix E for details).

References

Agrawal, S. and Devanur, N. R. 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.

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

Arlotto, A. and Gurvich, I. Uniformly bounded regret in the multisecretary problem. *Stochastic Systems*, 9(3): 231–260, 2019.

Azar, P. D., Kleinberg, R., and Weinberg, S. M. Prophet inequalities with limited information. In *Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms*, pp. 1358–1377. SIAM, 2014.

Balseiro, S., Kroer, C., and Kumar, R. Online resource allocation under horizon uncertainty. *arXiv preprint arXiv:2206.13606*, 2022a.

Balseiro, S. R. and Gur, Y. Learning in repeated auctions with budgets: Regret minimization and equilibrium. *Management Science*, 65(9):3952–3968, 2019.

Balseiro, S. R., Lu, H., and Mirrokni, V. The best of many worlds: Dual mirror descent for online allocation problems. *Operations Research*, 2022b.

Banerjee, S. and Freund, D. Uniform loss algorithms for online stochastic decision-making with applications to bin packing. In *Abstracts of the 2020 SIGMETRICS/Performance Joint International Conference on Measurement and Modeling of Computer Systems*, pp. 1–2, 2020.

Banerjee, S., Gurvich, I., and Vera, A. Constant regret in online allocation: On the sufficiency of a single historical trace, 2020. URL https://people.orie.cornell.edu/ig264/Online_Optimization_with_Samples.pdf.

Bertsekas, D. *Convex optimization theory*, volume 1. Athena Scientific, 2009.

Caramanis, C., Dütting, P., Faw, M., Fusco, F., Lazos, P., Leonardi, S., Papadigenopoulos, O., Pountourakis, E., and Reiffenhäuser, R. Single-sample prophet inequalities via greedy-ordered selection. In *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 1298–1325. SIAM, 2022.

Chen, X., Kroer, C., and Kumar, R. The complexity of pacing for second-price auctions. In *EC*, 2021. URL https://arxiv.org/abs/2103.13969.

Correa, J., Dütting, P., Fischer, F., and Schewior, K. Prophet inequalities for iid random variables from an unknown distribution. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, pp. 3–17, 2019.

Devanur, N. R. and Hayes, T. P. 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.

Devanur, N. R., Jain, K., Sivan, B., and Wilkens, C. A. Near optimal online algorithms and fast approximation algorithms for resource allocation problems. In *Proceedings of the 12th ACM conference on Electronic commerce*, pp. 29–38, 2011.

Facebook-Guide. Your guide to facebook bid strategy. https://www.facebook.com/gms_hub/share/biddingstrategyguide_final.pdf. Accessed: 2023-01-19.

Feldman, J., Henzinger, M., Korula, N., Mirrokni, V. S., and Stein, C. Online stochastic packing applied to display ad allocation. In *European Symposium on Algorithms*, pp. 182–194. Springer, 2010.

Gaitonde, J., Li, Y., Light, B., Lucier, B., and Slivkins, A. Budget pacing in repeated auctions: Regret and efficiency without convergence. *arXiv preprint arXiv:2205.08674*, 2022.

Gupta, A. and Molinaro, M. How the experts algorithm can help solve lps online. *Mathematics of Operations Research*, 41(4):1404–1431, 2016.

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

Jiang, J., Li, X., and Zhang, J. Online stochastic optimization with wasserstein based non-stationarity. *arXiv preprint arXiv:2012.06961*, 2020.

Kesselheim, T., Tönnis, A., Radke, K., and Vöcking, B. 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.

Kumar, B., Morgenstern, J., and Schrijvers, O. Optimal spend rate estimation and pacing for ad campaigns with budgets. *arXiv preprint arXiv:2202.05881*, 2022.

Rubinstein, A., Wang, J. Z., and Weinberg, S. M. Optimal single-choice prophet inequalities from samples. In *ITCS*, 2020.

Shalev-Shwartz, S. and Ben-David, S. *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.

Shalev-Shwartz, S. et al. Online learning and online convex optimization. *Foundations and Trends® in Machine Learning*, 4(2):107–194, 2012.

Zhou, Y., Chakrabarty, D., and Lukose, R. Budget constrained bidding in keyword auctions and online knapsack problems. In *International Workshop on Internet and Network Economics*, pp. 566–576. Springer, 2008.

Zhou, Y.-H., Liang, C., Li, N., Yang, C., Zhu, S., and Jin, R. Robust online matching with user arrival distribution drift. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pp. 459–466, 2019.

A. Related Work

Balseiro & Gur (2019) study budget pacing in repeated second-price auctions when the values and competing bids are either i.i.d. according to some unknown distribution or adversarially selected. They show that Dual Gradient Descent with the constant target $\lambda_t = B/T$ for all $t \in [T]$ attains the optimal regret of $O(\sqrt{T})$ in the i.i.d. stochastic setting, and the optimal parameter-dependent asymptotic competitive ratio (equal to ratio of the per-period budget to the maximum value) in the adversarial setting. Zhou et al. (2008) also study the adversarial setting and provide a pacing-based algorithm that achieves a differently-parameterized competitive ratio which scales as the logarithm of the ratio of the highest-to-lowest return-on-investment, and show that it is optimal. Kumar et al. (2022) study an episodic setting and provide a density-estimation-based algorithm for learning the target expenditures for each episode. Gaitonde et al. (2022) study the performance of the algorithm of Balseiro & Gur (2019) for the different objective of value maximization, and against the different benchmark comprised of pacing multipliers which spend the same amount B/T at each time period. Under the no-overbidding assumption, they show that the algorithm of Balseiro & Gur (2019) achieves $O(T^{3/4})$ regret. Recent years have also seen significant attention being given to pacing in multi-buyer settings, but we focus on the single-agent setting here and refer the reader to the recent works of Chen et al. (2021) and Gaitonde et al. (2022) for an overview.

More generally, budget pacing in second-price auctions is a special case of online linear packing, which in turn is a special case of the online resource allocation problem. Both these problems allow for multiple resources and have been studied extensively; we only provide a broad overview here. For the most part, these problems have also been studied in the i.i.d. stochastic model, or the slightly more general random arrival model (requests are selected by an adversary but arrive in a uniformly random order). Devanur & Hayes (2009) and Feldman et al. (2010) study online linear packing under the random arrival model, and show that learning the dual from the initial requests and then using it to make decisions yields $O(T^{2/3})$ regret. Agrawal et al. (2014) extended these results to show that repeatedly solving for the dual at geometrically increasing intervals yields the optimal $O(\sqrt{T})$ regret. Devanur et al. (2011) and Kesselheim et al. (2014) also achieve $O(\sqrt{T})$ regret but with a better dependence on the constants and the number of resources. Gupta & Molinaro (2016) give a dual-descent-based algorithm that also achieves $O(\sqrt{T})$ regret.

Agrawal & Devanur (2014) study online resource allocation with concave rewards and convex constraints, and give a dual-descent-based algorithm that achieves $O(\sqrt{T})$ regret. Balseiro et al. (2022b) give a Dual Mirror Descent algorithm which attempts to spend $\lambda_t = B/T$ at each time step and show that it achieves $O(\sqrt{T})$ regret for the general online allocation problem. Their results also hold for stochastic models that are close to i.i.d. like periodic, ergodic etc. Balseiro et al. (2022a) present and analyze a generalization of Dual Mirror Descent that can incorporate arbitrary target expenditures $\{\lambda_t\}_t$. When the horizon T is not known exactly and only assumed to lie in some known uncertainty window, they provide a procedure for computing target expenditures which yield a near-optimal asymptotic competitive ratio. Our Dual FTRL algorithm (Algorithm 2) is equivalent to the Lazy-update version of the algorithm of Balseiro et al. (2022a). Importantly, with the exception of Devanur et al. (2011), none of the aforementioned works provide algorithms which can achieve vanishing regret for the setting with time-varying distributions. Devanur et al. (2011) achieve $O(\sqrt{T})$ regret when the optimal expected reward for each distribution is known in advance. However, this quantity cannot be computed with a single sample for non-trivial distributions, and they do not provide guarantees for our sample-access setting.

Another line of work develops algorithms that beat $O(\sqrt{T})$ regret when the problem instance is well-structured. With the exception of Banerjee et al. (2020) and Banerjee & Freund (2020), all of these works assume complete knowledge of the distributions and/or assume that the distributions are identical. When the number of requests of each type satisfies a concentration property between the trace and the actual requests, Banerjee et al. (2020) and Banerjee & Freund (2020) show that a constant regret can be achieved for online resource allocation using one sample per distribution. For the budget pacing problem, a type corresponds to a value and competing bid pair. Since complex machine-learning models are typically used to estimate advertiser values to a high precision, this translates to an extremely large number of possible types. Far from concentrating, these large number of types imply that one is unlikely to even observe a type more than once, making their primal-based method ineffective for budget pacing. Moreover, neither Banerjee et al. (2020) nor Banerjee & Freund (2020) do not provide any robustness guarantees for possible discrepancies between the sampling and true distributions, and their algorithm requires knowledge of the competing bid. Finally, our results are meaningful when the budget is much larger than the maximum amount one can spend on an auction/request, as is the case for budget pacing. In contrast, the literature on prophet inequalities considers a unit-cost variable-reward online allocation problem where the budget is only large enough to accept one request. See Azar et al. (2014); Correa et al. (2019); Rubinstein et al. (2020); Caramanis et al. (2022) for a sample-driven treatment of prophet inequalities.

B. Application to Budget Pacing

Here we discuss how the budget pacing problem fits as a special case of the online resource allocation problem that we study in this paper. Consider the setting in which a budget-constrained advertiser repeatedly participates in T second-price auctions. For simplicity, assume that all ties are broken in favor of this advertiser. Let v_t and d_t denote her value and the highest competing bid in the t-th auction respectively. Moreover, let B denote her budget, which represents the maximum amount she is willing to spend over all T auctions.

We will assume that the tuple (v_t, d_t) is drawn from some distribution $\mathcal{P}_t$, independently of all other auctions. Now, observe that every bid of the advertiser results in one of two possible outcomes: (i) she bids greater than or equal to d_t, wins the auction, gains utility $v_t - d_t$ and pays d_t; (ii) she bids strictly less than d_t, loses the auction, gains zero value and pays zero. Thus, corresponding to the tuple (v_t, d_t), we can define a corresponding request γ_t with linear reward function $f_t(x) = (v_t - d_t) \cdot x$ and linear consumption function $b_t(x) = d_t \cdot x$, for the action space $x \in \{0, 1\} = \{\text{lose}, \text{win}\}$. Similarly, corresponding to the sample trace of tuples $\{(\tilde{v}_t, \tilde{d}_t)\}_t$, we can define a trace $\{\tilde{\gamma}_t\}_t$ for the online allocation problem. This defines a corresponding instance of the online allocation problem. Since every bid either results in either a win or loss, the maximum expected utility (value - payment) that the advertiser can earn subject to her budget constraint is bounded above by $\text{FLUID}(\{\mathcal{P}_t\}_t)$ for this instance. Finally, consider step t of Algorithm 2 on this instance. The decision x_t is calculated as $x_t \in \text{argmax}_{x \in \mathcal{X}_t} \{f_t(x) - \mu_t \cdot b_t(x)\}$. Therefore, $x_t = 1$ if $v_t - d_t \geq \mu d_t$, or equivalently $v_t/(1 + \mu_t) \geq d_t$, and $x_t = 0$ otherwise. Observe that, in a second price auction, if the advertiser bids $v_t/(1 + \mu_t)$, she will win ($x_t = 1$) if $v_t/(1 + \mu_t) \geq d_t$ and lose ($x_t = 0$) otherwise. Thus, by bidding $v_t/(1 + \mu_t)$, she can simulate the actions of Algorithm 2 for the online allocation instance. Moreover, she does not require knowledge of the competing bid d_t to compute her bid, which is crucial because d_t is not known in practice. Once the auction is over, the expenditure $b_t(x_t) = d_t \cdot x_t$ is revealed to the advertiser. She can then use it to update the dual iterate according to (3).

C. Fluid Benchmark is Stronger

Proposition C.1. *For any collection of request distributions $\{\mathcal{P}_t\}_t$, we have $\mathbb{E}_{\{\gamma_t\}_t}[\text{OPT}(\{\gamma_t\}_t)] \leq \text{FLUID}(\{\mathcal{P}_t\}_t)$.*

Proof. Fix any request sequence $\{\gamma_t\}_t$ and let $\{x_t^*\}_t$ be an optimal solution to the corresponding hindsight optimization problem $\text{OPT}(\{\gamma_t\}_t)$. Then, $x_t(\gamma) = x_t^*$ for all $\gamma \in \mathcal{S}$ is a feasible solution of $\text{FLUID}(\{\mathcal{P}_t\})$ and consequently, we have $\text{OPT}(\{\gamma_t\}_t) \leq \text{FLUID}(\{\mathcal{P}_t\}_t)$. Since the request sequence $\{\gamma_t\}_t$ was arbitrary, we have $\mathbb{E}_{\{\gamma_t\}_t}[\text{OPT}(\{\gamma_t\}_t)] \leq \text{FLUID}(\{\mathcal{P}_t\}_t)$, as required. $\square$

D. General Position

Given any collection of request distributions $\{\mathcal{P}_t\}_t$ (which may or may not satisfy Assumption 3.3), we can define perturbed distributions $\{\hat{\mathcal{P}}_t\}_t$ to capture the distribution of perturbed requests $\hat{\gamma}_t = (\hat{f}_t, \hat{b}_t)$ generated using the following two step procedure: (i) Draw a request $\gamma_t = (f_t, b_t)$ according to the unperturbed distributions $\mathcal{P}_t$; (ii) Add a perturbation by setting $\hat{f}_t(x) = f_t(x) + \epsilon_t \cdot x$ for all $x \in \mathcal{X}$, where $\epsilon_t \sim \text{Unif}([0, a])$, and leave the consumption function unchanged $\hat{b}_t(\cdot) = b_t(\cdot)$. Then, $\{\hat{\mathcal{P}}_t\}_t$ satisfy Assumption 3.3 and $\left|\text{FLUID}(\{\mathcal{P}_t\}_t) - \text{FLUID}(\{\hat{\mathcal{P}}_t\}_t)\right| \leq a \cdot T$, where $a > 0$ can be made arbitrarily small.

E. Efficiently Computing the Target Sequence

In this section, we describe an efficient procedure for computing the empirical optimal dual solution $\tilde{\mu}$ and the target sequence $\{\lambda_t\}_t$. Consider a trace $\{\tilde{\gamma}_t\}_t$ and set $\tilde{\gamma}_0 = (f_0, b_0)$ such that $f_0(x) = b_0(x) = 0$ for all $x \in \mathcal{X}$. Without loss of generality, we will assume that $\{\tilde{\gamma}_t\}_t$ is sorted in increasing order of $\text{coeff}(\tilde{f}_t)/\text{coeff}(\tilde{b}_t)$ (assume $0/0 = 0$), i.e.,

$$\frac{\text{coeff}(\tilde{f}_s)}{\text{coeff}(\tilde{b}_s)} \leq \frac{\text{coeff}(\tilde{f}_t)}{\text{coeff}(\tilde{b}_t)} \quad \forall s \leq t.$$

This can be easily achieved by maintaining a sorted array with $O(\log(T))$ insertion time or sorting the array with $O(T \log(T))$ processing time. Moreover, since the trace $\{\tilde{\gamma}_t\}_t$ is in general position by Assumption 3.3, all the $\text{coeff}(\tilde{f}_t)/\text{coeff}(\tilde{b}_t)$ are distinct for $t \in [T]$.

Algorithm 3 Learning the Dual from the Trace

Input: Trace $\{\tilde{\gamma}_t\}_t$ in general position and sorted in increasing order of $\operatorname{coeff}(\tilde{f}_t)/\operatorname{coeff}(\tilde{b}_t)$.
Initialize: Total payment $P = 0$ and target sequence $\lambda_t = 0$ for all $t \in [T]$.
for $t = T, \ldots, 0$ **do**
 if $P + \tilde{b}_t(\bar{x}) > B$ **then**
  Set $\lambda_t \leftarrow \tilde{b}_t(\bar{x})$, and set $\tilde{\mu} = \operatorname{coeff}(\tilde{f}_t)/\operatorname{coeff}(\tilde{b}_t)$. **Break**.
 end
 else
  Update total payment $P \leftarrow P + \tilde{b}_t(\bar{x})$ and set $\lambda_t \leftarrow \tilde{b}_t(\bar{x})$.
 end
end
return *Dual variable $\tilde{\mu}$, target sequence $\{\lambda_t\}_t$*

Theorem E.1. *$\tilde{\mu}$ returned by Algorithm 3 is the smallest element in $\operatorname{argmin}_{\mu \geq 0} \mu \cdot B + \sum_{t=1}^{T} \max_{x \in \mathcal{X}} \left\{ \tilde{f}_t(x) - \mu \cdot \tilde{b}_t(x) \right\}$. Moreover, $\lambda_t = b_t^*(\tilde{\mu})$ for all $t \in [T]$.*

Proof. Set $q(\mu) = \mu \cdot B + \sum_{t=1}^{T} \max_{x \in \mathcal{X}} \left\{ \tilde{f}_t(x) - \mu \cdot \tilde{b}_t(x) \right\}$. First, we show that the dual variable $\tilde{\mu}$ is smallest element in $\operatorname{argmin}_{\mu \geq 0} q(\mu)$. To do so, we consider the following two cases:

- $\tilde{\mu} = 0$. In this case, the 'If' condition implies that there exists $s \in [T]$ such that $\sum_{t=s}^{T} \tilde{b}_t(\bar{x}) \leq B$ and $\operatorname{coeff}(\tilde{f}_t) = 0$ for all $t < s$. Moreover, we have $0 \in \operatorname{argmax}_{x \in \mathcal{X}} \left\{ \tilde{f}_t(x) - \mu \cdot \tilde{b}_t(x) \right\}$ for all $t < s$ and $\bar{x} \in \operatorname{argmax}_{x \in \mathcal{X}} \left\{ \tilde{f}_t(x) - \mu \cdot \tilde{b}_t(x) \right\}$ for all $t \geq s$. Now, note that the set of sub-gradients of the maximum of a collection of linear functions is equal to convex hull of gradients of all the linear functions which are binding (for example, see Chapter 5 of Bertsekas 2009). Therefore, $B - \sum_{t=s}^{T} \tilde{b}(\bar{x}) \in \partial q(0)$. Since $B - \sum_{t=s}^{T} \tilde{b}(\bar{x}) \geq 0$, the definition of a subgradient implies that $q(0) \leq q(\mu)$ for all $\mu \geq 0$. Hence, we have shown that $\tilde{\mu}$ is the smallest minimizer of $q(\cdot)$, as required.

- $\tilde{\mu} > 0$. In this case, the 'If' condition implies that there exists $s \in [T]$ such that $\sum_{t=s+1}^{T} \tilde{b}_t(\bar{x}) < B$, $\sum_{t=s}^{T} \tilde{b}_t(\bar{x}) > B$ and $\tilde{f}_s(x) - \tilde{\mu} \cdot \tilde{b}_s(x) = 0$ for all $x \in \mathcal{X}$. Moreover, we have $0 \in \operatorname{argmax}_{x \in \mathcal{X}} \left\{ \tilde{f}_t(x) - \mu \cdot \tilde{b}_t(x) \right\}$ for all $t < s$, $\bar{x} \in \operatorname{argmax}_{x \in \mathcal{X}} \left\{ \tilde{f}_t(x) - \mu \cdot \tilde{b}_t(x) \right\}$ for all $t > s$ and $\{0, \bar{x}\} \subseteq \operatorname{argmax}_{x \in \mathcal{X}} \left\{ \tilde{f}_s(x) - \mu \cdot \tilde{b}_s(x) \right\}$. Now, select a $\lambda \in [0, 1]$ such that

$$B - \lambda \cdot \tilde{b}_s(0) + (1 - \lambda) \cdot \tilde{b}_s(\bar{x}) + \sum_{t=s+1}^{T} \tilde{b}_t(\bar{x}) = 0 \,.$$

Now, note that the set of sub-gradients of the maximum of a collection of linear functions is equal to convex hull of gradients of all the linear functions which are binding (for example, see Chapter 5 of Bertsekas 2009). Therefore, $0 = B - \lambda \cdot \tilde{b}_s(0) + (1 - \lambda) \cdot \tilde{b}_s(\bar{x}) + \sum_{t=s+1}^{T} \tilde{b}_t(\bar{x}) \in \partial q(\tilde{\mu})$. Consequently, the definition of a subgradient implies that $q(0) \leq q(\mu)$ for all $\mu \geq 0$. Finally, consider any $\mu < \tilde{\mu}$. Then, $\tilde{f}_s(x) - \mu \cdot \tilde{b}_s(x) > 0$, which further implies $\{\bar{x}\} = \operatorname{argmax}_{x \in \mathcal{X}} \left\{ \tilde{f}_t(x) - \mu \cdot \tilde{b}_t(x) \right\}$ for all $t \geq s$. Therefore, for any $\{x_t\}_t$ such that $x_t \in \operatorname{argmax}_{x \in \mathcal{X}} \left\{ \tilde{f}_t(x) - \mu \cdot \tilde{b}_t(x) \right\}$, we have $B - \sum_{t=1}^{T} \tilde{b}_t(x_t) > 0$. Hence, $v > 0$ for all $v \in \partial q(\mu)$ and consequently μ is a minimizer of $q(\cdot)$. Hence, we have shown that $\tilde{\mu}$ is the smallest minimizer of $q(\cdot)$, as required.

Finally, we show that $\lambda_t = \tilde{b}_t^*(\tilde{\mu})$ for all $t \in [T]$. Let s be the value of t at which the 'For' loop terminates. From the definition of $\tilde{\mu}$, we have $\{\bar{x}\} = \operatorname{argmax}_{x \in \mathcal{X}} \left\{ \tilde{f}_t(x) - \tilde{\mu} \cdot \tilde{b}_t(x) \right\}$ for all $t > s$, $\mathcal{X} = \operatorname{argmax}_{x \in \mathcal{X}} \left\{ \tilde{f}_s(x) - \tilde{\mu} \cdot \tilde{b}_s(x) \right\}$ and $\{0\} = \operatorname{argmax}_{x \in \mathcal{X}} \left\{ \tilde{f}_t(x) - \tilde{\mu} \cdot \tilde{b}_t(x) \right\}$ for all $t < s$. Therefore, $\tilde{b}_t^*(\tilde{\mu}) = \tilde{b}_t(\bar{x}) = \lambda_t$ for all $t \geq s$ and $\tilde{b}_t^*(\tilde{\mu}) = \tilde{b}_t(0) = \lambda_t$ for all $t < s$. $\qquad \square$

F. Missing Proofs from Section 3

F.1. Proof of Lemma 3.4

Proof of Lemma 3.4. Define $q(\mu) = \mu \cdot B + \sum_{t=1}^{T} \max_{x \in \mathcal{X}} \left\{ \tilde{f}_t(x) - \mu \cdot \tilde{b}_t(x) \right\}$. Then, $q(\cdot)$ is a convex function of μ because the maximum of a collection of linear function is convex and the sum of convex function is also convex (Bertsekas, 2009). Since $\tilde{\mu} \in \operatorname{argmin}_{\mu \geq 0} q(\mu)$, first-order condition of optimality (Proposition 5.4.7 of Bertsekas 2009) implies that one of the following statements holds:

(i) $\tilde{\mu} = 0$ and there exists $v \in \partial q(0)$ such that $v \geq 0$.

(ii) Zero is a sub-differential of q at $\tilde{\mu}$, i.e., $0 \in \partial q(\tilde{\mu})$.

Recall that the trace $\{\tilde{\gamma}_t\}_t$ is assumed to be in general position with probability one. Therefore, there is at most one time step for which $\operatorname{argmax}_{x \in \mathcal{X}} f_t(x) - \tilde{\mu} \cdot b_t(x)$ is not unique. Let s be that time step. Now, note that the set of sub-gradients of the maximum of a collection of linear functions is equal to convex hull of gradients of all the linear functions which are binding (for example, see Chapter 5 of Bertsekas 2009). Hence, $v \in \partial q(\tilde{\mu})$ implies the existence of $\mathcal{D}_s \in \Delta(\mathcal{X})$ such that

$$\text{Support}(\mathcal{D}_s) \subseteq \operatorname*{argmax}_{x \in \mathcal{X}} \tilde{f}_s(x) - \tilde{\mu} \cdot \tilde{b}_s(x) \quad \text{and} \quad v = B - \mathbb{E}_{x \sim \mathcal{D}_s}[\tilde{b}_s(x)] - \sum_{t \neq s}^{T} \tilde{b}_t^*(\tilde{\mu}) \,.$$

where $x_t^*(\gamma_t, \tilde{\mu})$ is the optimal solution to $\max_{x \in \mathcal{X}} \tilde{f}_t(x) - \tilde{\mu} \cdot \tilde{b}_t(x)$ as described in Definition 3.1. Since $0 \leq b_s(x) \leq \bar{b}$ for all $x \in \mathcal{X}$, we get

$$\left| B - v - \sum_{t=1}^{T} \tilde{b}_t^*(\tilde{\mu}) \right| \leq \bar{b} \,,$$

where we have used $0 \leq b_t(x) \leq \bar{b}$ for all $x \in \mathcal{X}$ and $t \in [T]$. Therefore, statements (i) and (ii) imply that either $\tilde{\mu} = 0$ and $\bar{b} + B - \sum_{t=1}^{T} \tilde{b}_t^*(\tilde{\mu}) \geq v \geq 0$, or $\left| B - \sum_{t=1}^{T} \tilde{b}_t^*(\tilde{\mu}) \right| \leq \bar{b}$, as required. $\square$

F.2. Proof of Theorem 3.5

Proof of Theorem 3.5. Define the hypothesis class

$$\mathcal{F} := \{(f, b) \mapsto b^*(\mu) \mid \mu \geq 0\} \,.$$

Let $\text{Rad}(\cdot)$ denote Radmacher complexity. Then, we know from PAC learning theory (for example see Chapter 26 of Shalev-Shwartz & Ben-David 2014) that

$$\Pr_{\{\tilde{\gamma}\}_t \sim \prod_t \tilde{\mathcal{P}}_t} \left(\sup_{\mu \geq 0} \left| \sum_{t=1}^{T} \tilde{b}_t^*(\mu) - \sum_{t=1}^{T} \mathbb{E}_{\hat{\gamma}_t \sim \tilde{\mathcal{P}}_t} \left[\hat{b}_t^*(\mu) \right] \right| \geq r(T) \right) \leq \frac{1}{T^2} \,. \tag{8}$$

for

$$r(T) \geq 2T \cdot \mathbb{E}_{\{\hat{\gamma}_t\}_t \sim \prod_t \tilde{\mathcal{P}}_t} [\text{Rad}(\mathcal{F} \circ \{\hat{\gamma}_t\}_t)] + 2\bar{b} \cdot \sqrt{T \log(2T)} \,.$$

Let $H(\{\hat{\gamma}_t\}_t) = \left\{ \{\hat{b}_t^*(\mu)\}_t \mid \mu \geq 0 \right\}$ denote the set of all possible resource expenditure vectors that can be generated from a trace $\{\hat{\gamma}_t\}_t$, then

$$\mathbb{E}_{\{\hat{\gamma}_t\}_t \sim \prod_t \tilde{\mathcal{P}}_t} [\text{Rad}(\mathcal{F} \circ \{\hat{\gamma}_t\}_t)] = \mathbb{E}_{\{\hat{\gamma}_t\}_t \sim \prod_t \tilde{\mathcal{P}}_t} [\text{Rad}(H(\{\hat{\gamma}_t\}_t))] = \frac{1}{T} \cdot \mathbb{E}_{\{\hat{\gamma}_t\}_t \sim \prod_t \tilde{\mathcal{P}}_t} \mathbb{E}_{\vec{\sigma}} \left[\sup_{\mu \geq 0} \sum_{t=1}^{T} \sigma_t \cdot \hat{b}_t^*(\mu) \right], \tag{9}$$

where $\{\sigma_t\}_t$ are independent Radmacher random variables.

For a linear function $f : \mathbb{R} \to \mathbb{R}$, let $\text{coeff}(f)$ denote its coefficient. Moreover, let $\bar{x} = \max_{x \in \mathcal{X}} x$. Then, observe that for a request $\gamma = (f, b)$ and dual variable $\mu \geq 0$, we have

$$x^*(\gamma, \mu) = \begin{cases} \bar{x} & \text{if } \text{coeff}(f) - \mu \cdot \text{coeff}(b) \geq 0 \text{ and } \text{coeff}(f) \neq 0 \\ 0 & \text{otherwise} \end{cases}.$$

Therefore, for any request $\gamma = (f, b)$ with $\text{coeff}(f) \neq 0$ and $\text{coeff}(b) \neq 0$, there exists a critical $\mu^* = \text{coeff}(f)/\text{coeff}(b)$ such that

$$b^*(\mu) = \begin{cases} b(\bar{x}) & \text{if } \mu \leq \mu^* \\ 0 & \text{if } \mu > \mu^* \end{cases}.$$

For $\gamma = (f, b)$ with $\text{coeff}(f) = 0$ or $\text{coeff}(b) \neq 0$, we have $b^*(\mu) = 0$ for all $\mu \geq 0$. Consider the trace $\{\hat{\gamma}_t\}_t$, and let μ_t^* be the critical dual solution for request $\hat{\gamma}_t$ as defined above. Then, the assumption that the trace is in general position (Assumption 3.3) implies that the critical points $\{\mu_t^*\}_t$ are distinct. Consequently, we get that $\{\hat{b}_t^*(\mu)\}_t$ remains constant whenever μ lies between any two critical points. Since the total number of critical points is T, we get that $|H(\{\hat{\gamma}_t\}_t)| \leq T$. Therefore, Massart Lemma applies and we get

$$\text{Rad}(H(\{\hat{\gamma}_t\}_t)) \leq \bar{b} \cdot \sqrt{\frac{2 \log(T)}{T}}.$$

Combining this with (8) and (9) yields the theorem. □

F.3. Proof of Theorem 3.6

Proof of Theorem 3.6. By Assumption 3.3, the request sequence $\{\gamma_t\}$ is in general position almost surely. Therefore, there is at most 1 time step such that $f_t(x_t') \neq f_t^*(\tilde{\mu})$, call it s. Let ζ_A be the first time step t in which $B_{t+1} \leq \bar{b}$, i.e., $\sum_{t=1}^{\zeta_A} b_t^*(\tilde{\mu}) \geq B - \bar{b}$. Then,

$$\mathbb{E}\left[R(A|\{\gamma_t\}_t)\right] = \mathbb{E}\left[\sum_{t=1}^{\zeta_A} f_t(x_t')\right]$$

$$\geq \mathbb{E}\left[\sum_{t=1}^{\zeta_A} f_t^*(\tilde{\mu})\right] - |f_s^*(\tilde{\mu}) - f_s(x_s')|$$

$$\geq \mathbb{E}\left[\sum_{t=1}^{T} f_t^*(\tilde{\mu})\right] - \mathbb{E}\left[\sum_{t=\zeta_A+1}^{T} f_t^*(\tilde{\mu})\right] - \bar{f}$$

$$\geq \mathbb{E}\left[\sum_{t=1}^{T} f_t^*(\tilde{\mu})\right] - \mathbb{E}\left[\kappa \cdot \sum_{t=\zeta_A+1}^{T} b_t^*(\tilde{\mu})\right] - \bar{f}$$

$$\geq D(\tilde{\mu}|\{\mathcal{P}_t\}_t) - \mathbb{E}\left[\tilde{\mu} \cdot \left(B - \sum_{t=1}^{T} b_t^*(\tilde{\mu})\right)\right] - \mathbb{E}\left[\kappa \cdot \sum_{t=\zeta_A+1}^{T} b_t^*(\tilde{\mu})\right] - \bar{f}$$

$$\geq \text{FLUID}(\{\mathcal{P}_t\}_t) - \mathbb{E}\left[\tilde{\mu} \cdot \left(B - \sum_{t=1}^{T} b_t^*(\tilde{\mu})\right)\right] - \mathbb{E}\left[\kappa \cdot \sum_{t=\zeta_A+1}^{T} b_t^*(\tilde{\mu})\right] - \bar{f}.$$

Therefore,

$$\text{Regret}(A) \leq \mathbb{E}\left[\kappa \cdot \sum_{t=\zeta_A+1}^{T} b_t^*(\tilde{\mu})\right] + \mathbb{E}\left[\tilde{\mu} \cdot \left(B - \sum_{t=1}^{T} b_t^*(\tilde{\mu})\right)\right] + \bar{f}.$$

In the remainder of the proof, we bound the first two terms on the RHS.

For the first term, observe that

$$\sum_{t=\zeta_A+1}^{T} b_t^*(\tilde{\mu}) \le \left(\sum_{t=1}^{T} b_t^*(\tilde{\mu})\right) - (B - \bar{b})$$

$$= \left(\sum_{t=1}^{T} b_t^*(\tilde{\mu}) - \sum_{t=1}^{T} \tilde{b}_t^*(\tilde{\mu})\right) - \left(B - \sum_{t=1}^{T} \tilde{b}_t^*(\tilde{\mu})\right) + \bar{b}$$

$$\le \left|\sum_{t=1}^{T} b_t^*(\tilde{\mu}) - \sum_{t=1}^{T} \tilde{b}_t^*(\tilde{\mu})\right| + 2 \cdot \bar{b}, \tag{10}$$

where the first inequality follows from the definition of ζ_A and the last inequality follows from Lemma 3.4.

For the second term, observe that Lemma 3.4 implies

$$\tilde{\mu} \cdot \left(B - \sum_{t=1}^{T} b_t^*(\tilde{\mu})\right) = \tilde{\mu} \cdot \left(B - \sum_{t=1}^{T} \tilde{b}_t^*(\tilde{\mu})\right) + \tilde{\mu} \cdot \left(\sum_{t=1}^{T} \tilde{b}_t^*(\tilde{\mu}) - \sum_{t=1}^{T} b_t^*(\tilde{\mu})\right)$$

$$\le \tilde{\mu} \cdot \bar{b} + \tilde{\mu} \cdot \left|\sum_{t=1}^{T} \tilde{b}_t^*(\tilde{\mu}) - \sum_{t=1}^{T} b_t^*(\tilde{\mu})\right|.$$

Note that $\tilde{\mu} \le \kappa$. This is because the definition of κ implies that $\max_{x \in \mathcal{X}} f(x) - \mu \cdot b(x) = 0$ for all $\gamma = (f, b) \in \mathcal{S}$ and $\mu \ge \kappa$. Hence,

$$\mu \cdot B + \sum_{t=1}^{T} \max_{x \in \mathcal{X}} \left\{\tilde{f}_t(x) - \mu \cdot \tilde{b}_t(x)\right\} = \mu \cdot B < \kappa \cdot B = \kappa \cdot B + \sum_{t=1}^{T} \max_{x \in \mathcal{X}} \left\{\tilde{f}_t(x) - \kappa \cdot \tilde{b}_t(x)\right\}$$

for all $\mu > \kappa$. Therefore, we get

$$\tilde{\mu} \cdot \left(B - \sum_{t=1}^{T} b_t^*(\tilde{\mu})\right) \le \kappa \cdot \bar{b} + \kappa \cdot \left|\sum_{t=1}^{T} \tilde{b}_t^*(\tilde{\mu}) - \sum_{t=1}^{T} b_t^*(\tilde{\mu})\right|. \tag{11}$$

Define G to be the good event to be one in which the total expenditures under the trace and the requests sequence are close:

$$\sup_{\mu \ge 0} \left|\sum_{t=1}^{T} \tilde{b}_t^*(\mu) - \sum_{t=1}^{T} b_t^*(\mu)\right| \le r(T).$$

Then, Theorem 3.5 and Union Bound imply that $\Pr(G^c) \le 2/T^2$ and $\Pr(G) \ge 1/2/T^2$. Finally, we can put it all together to get the required bound:

$$\text{Regret}(A) \le \mathbb{E}\left[\kappa \cdot \sum_{t=\zeta_A+1}^{T} b_t^*(\tilde{\mu})\right] + \mathbb{E}\left[\tilde{\mu} \cdot \left(B - \sum_{t=1}^{T} b_t^*(\tilde{\mu})\right)\right] + \bar{f}$$

$$\le \mathbb{E}\left[\kappa \cdot \left|\sum_{t=1}^{T} \tilde{b}_t^*(\tilde{\mu}) - \sum_{t=1}^{T} b_t^*(\tilde{\mu})\right|\right] + 2\kappa\bar{b} + \mathbb{E}\left[\kappa \cdot \left|\sum_{t=1}^{T} \tilde{b}_t^*(\tilde{\mu}) - \sum_{t=1}^{T} b_t^*(\tilde{\mu})\right|\right] + \kappa\bar{b} + \kappa\bar{b}$$

$$= 2\kappa \cdot \mathbb{E}\left[\left|\sum_{t=1}^{T} \tilde{b}_t^*(\tilde{\mu}) - \sum_{t=1}^{T} b_t^*(\tilde{\mu})\right| \,\Big|\, G\right] \Pr(G) + 2\kappa \cdot \mathbb{E}\left[\left|\sum_{t=1}^{T} \tilde{b}_t^*(\tilde{\mu}) - \sum_{t=1}^{T} b_t^*(\tilde{\mu})\right| \,\Big|\, G^c\right] \Pr(G^c) + 4\kappa\bar{b}$$

$$\le 2\kappa \cdot r(T) + 2\kappa \cdot 2T\bar{b} \cdot \frac{2}{T^2} + 4\kappa\bar{b}$$

$$\le 12\kappa\bar{b} + 2\kappa r(T). \qquad \square$$

G. Missing Proofs from Section 4

G.1. Proof of Theorem 4.1

Proof of Theorem 4.1. Let ζ_A be the first time less than T for which $\sum_{t=1}^{\zeta_A} b_t(x_t) + \bar{b} \geq B$. Set $\zeta_A = T$ if this inequality is never satisfied. Then, $x_t = x'_t$ for all $t \leq \zeta_A$ and $\sum_{t=1}^{\zeta_A} b_t(x'_t) \geq B - \bar{b}$.

First, observe that

$$R(A|\{\gamma_t\}_t) \geq \sum_{t=1}^{\zeta_A} f_t(x'_t) = \sum_{t=1}^{T} f_t(x'_t) - \sum_{t=\zeta_A+1}^{T} f_t(x'_t) \geq \sum_{t=1}^{T} f_t(x'_t) - \kappa \cdot \sum_{t=\zeta_A+1}^{T} b_t(x'_t). \tag{12}$$

Next observe that, for all $t \in [T]$, μ_t is independent of γ_t because μ_t is completely determined by $\{\gamma_1, \ldots, \gamma_{t-1}\}$. Hence,

$$\mathbb{E}_{\gamma_t}\left[f_t(x'_t) \mid \mu_t\right] = \mathbb{E}_{\gamma_t}\left[f_t(x'_t) + \mu_t \cdot (\beta_t - b_t(x'_t)) \mid \mu_t\right] - \mathbb{E}_{\gamma_t}\left[\mu_t \cdot (\lambda_t - b_t(x'_t)) \mid \mu_t\right] - \mathbb{E}_{\gamma_t}\left[\mu_t \cdot (\beta_t - \lambda_t) \mid \mu_t\right]$$
$$= \mathbb{E}_{\gamma_t}\left[D(\mu_t|\mathcal{P}_t, \beta_t)|\mu_t\right] - \mathbb{E}_{\gamma_t}\left[\mu_t \cdot (\lambda_t - b_t(x'_t)) \mid \mu_t\right] - \mathbb{E}_{\gamma_t}\left[\mu_t \cdot (\beta_t - \lambda_t) \mid \mu_t\right].$$

Taking unconditional expectations on both sides and applying the tower rule yields

$$\mathbb{E}\left[f_t(x'_t)\right] = \mathbb{E}\left[D(\mu_t|\mathcal{P}_t, \beta_t)\right] - \mathbb{E}\left[\mu_t \cdot (\lambda_t - b_t(x'_t))\right] - \mathbb{E}\left[\mu_t \cdot (\beta_t - \lambda_t)\right].$$

Summing over $t \in [T]$, we get

$$\sum_{t=1}^{T} \mathbb{E}[f_t(x'_t)] = \sum_{t=1}^{T} \mathbb{E}\left[D(\mu_t|\mathcal{P}_t, \beta_t)\right] - \sum_{t=1}^{T} \mathbb{E}\left[\mu_t \cdot (\lambda_t - b_t(x'_t))\right] - \sum_{t=1}^{T} \mathbb{E}\left[\mu_t \cdot (\beta_t - \lambda_t)\right]. \tag{13}$$

Therefore, (12) and (13) together imply

$$\mathbb{E}\left[\left\{\sum_{t=1}^{T} D(\mu_t|\mathcal{P}_t, \beta_t)\right\} - R(A|\{\gamma_t\}_t)\right] \leq \mathbb{E}\left[\sum_{t=1}^{T} \mu_t \cdot (\lambda_t - b_t(x'_t)) + \kappa \cdot \sum_{t=\zeta_A+1}^{T} b_t(x'_t)\right] + \sum_{t=1}^{T} \mathbb{E}[\mu_t \cdot (\beta_t - \lambda_t)]. \tag{14}$$

FTRL Regret Bound. Define $w_t(\mu) \coloneqq \mu \cdot (\lambda_t - b_t(x'_t))$. Then, Algorithm 2 can be seen as running FTRL with linear losses $\{w_t(\cdot)\}_t$. The gradients of these loss functions are given by $\nabla w_t(\mu) = \lambda_t - b_t(x'_t)$, which satisfy $\|\nabla w_t(\mu)\|_\infty \leq \|b_t(x'_t)\|_\infty + \|\lambda_t\|_\infty \leq \bar{b} + \bar{\lambda}$. Therefore, the regret bound for FTRL implies that for all $\mu \geq 0$:

$$\sum_{t=1}^{T} w_t(\mu_t) - w_t(\mu) \leq E(T, \mu), \tag{15}$$

where $E(T, \mu) = \frac{2(\bar{b}+\bar{\lambda})^2}{\sigma} \eta \cdot T + \frac{h(\mu) - h(\mu_1)}{\eta}$ is the regret bound of FTRL after T iterations (Hazan et al., 2016). Equivalently, we can write

$$\sum_{t=1}^{T} \mu_t \cdot (\lambda_t - b_t(x'_t)) \leq E(T, \mu) + \sum_{t=1}^{T} \mu \cdot (\lambda_t - b_t(x'_t)) \qquad \forall \mu \geq 0.$$

Now, consider the following two cases:

- Case 1: $\zeta_A = T$. Here, setting $\mu = 0$ yields

$$\sum_{t=1}^{T} \mu_t \cdot (\lambda_t - b_t(x'_t)) + \kappa \cdot \sum_{t=\zeta_A+1}^{T} b_t(x'_t) \leq E(T, 0).$$

- Case 2: $\zeta_A < T$. Then, $\sum_{t=1}^{\zeta_A} b_t(x'_t) \geq B - \bar{b}$. Hence, setting $\mu = \kappa$ yields

$$\sum_{t=1}^{T} \mu_t \cdot (\lambda_t - b_t(x'_t)) + \kappa \cdot \sum_{t=\zeta_A+1}^{T} b_t(x'_t) \leq E(T, \kappa) + \sum_{t=1}^{T} \kappa \cdot (\lambda_t - b_t(x'_t)) + \kappa \cdot \sum_{t=\zeta_A+1}^{T} b_t(x'_t)$$

$$= E(T, \kappa) + \kappa \cdot \left(\sum_{t=1}^{T} \lambda_t - \sum_{t=1}^{\zeta_A} b_t(x'_t) \right)$$

$$\leq E(T, \kappa) + \kappa \cdot \left(\left\{ \sum_{t=1}^{T} \lambda_t \right\} - (B - \bar{b}) \right)$$

$$= E(T, \kappa) + \kappa \bar{b} + \kappa \cdot \left(\left\{ \sum_{t=1}^{T} \lambda_t \right\} - B \right).$$

Combining the two cases implies that for all values of ζ_A we have

$$\sum_{t=1}^{T} \mu_t \cdot (\lambda_t - b_t(x'_t)) + \kappa \cdot \sum_{t=\zeta_A+1}^{T} b_t(x'_t) \leq \max\{E(T, 0), E(T, \kappa)\} + \kappa \bar{b} + \kappa \cdot \left(\left\{ \sum_{t=1}^{T} \lambda_t \right\} - B \right)^+. \quad (16)$$

Finally, combining (14) and (16) yields

$$\mathbb{E}\left[\left\{ \sum_{t=1}^{T} D(\mu_t | \mathcal{P}_t, \beta_t) \right\} - R(A | \{\gamma_t\}_t) \right] \leq \max\{E(T, 0), E(T, \kappa)\} + \kappa \bar{b} + \kappa \cdot \left(\left\{ \sum_{t=1}^{T} \lambda_t \right\} - B \right)^+$$

$$+ \sum_{t=1}^{T} \mathbb{E}[\mu_t \cdot (\beta_t - \lambda_t)].$$

Plugging in the definition of $E(T, \mu)$ finishes the proof. $\qquad \square$

G.2. Proof of Lemma 4.2

Proof of Lemma 4.2. Consider any two request distributions $\mathcal{P}, \tilde{\mathcal{P}} \in \Delta(\mathcal{S})$. Then, by the definition of the Wasserstein metric, there exists a joint probability distribution Q, with marginals $\mathcal{P}$ and $\tilde{\mathcal{P}}$ on the first and second factors respectively, such that

$$\mathcal{W}(\mathcal{P}, \tilde{\mathcal{P}}) = \mathbb{E}_{(\gamma, \tilde{\gamma}) \sim Q}\left[\sup_x |f(x) - \tilde{f}(x)| + \sup_x |b(x) - \tilde{b}(x)| \right].$$

Let $x^*(\gamma, \mu)$ be the optimal solution of $\max_{x \in X} f(x) - \mu \cdot b(x)$ for request $\gamma = (f, b)$ as described in Definition 3.1. Then, for any $\mu \in [0, \kappa]$ and $x : \mathcal{S} \to \mathcal{X}$, we have

$$\mathbb{E}_{(\gamma, \tilde{\gamma}) \sim Q}\left[\left| f(x(\gamma)) - \mu \cdot b(x(\gamma)) - \left\{ \tilde{f}(x(\gamma)) - \mu \cdot \tilde{b}(x(\gamma)) \right\} \right| \right]$$

$$\leq \mathbb{E}_{(\gamma, \tilde{\gamma}) \sim Q}\left[\left| f(x(\gamma)) - \tilde{f}(x(\gamma)) \right| + \mu \cdot \left| b(x(\gamma)) - \tilde{b}(x(\gamma)) \right| \right]$$

$$\leq \mathcal{W}(\mathcal{P}, \tilde{\mathcal{P}}) + \kappa \cdot \mathcal{W}(\mathcal{P}, \tilde{\mathcal{P}})$$

$$= (1 + \kappa) \cdot \mathcal{W}(\mathcal{P}, \tilde{\mathcal{P}}). \quad (17)$$

Now, for $t \in [T]$, we can use (17) to write

$$D(\mu_t | \tilde{\mathcal{P}}, \beta_t) - D(\mu_t | \mathcal{P}, \beta_t)$$

$$= \mathbb{E}_{\tilde{\gamma} \sim \tilde{\mathcal{P}}}[\tilde{f}(x^*(\tilde{\gamma}, \mu_t)) - \mu_t \cdot \tilde{b}(x^*(\tilde{\gamma}, \mu_t)) + \mu_t \cdot \beta_t] - \mathbb{E}_{\gamma \sim \mathcal{P}}[f(x^*(\gamma, \mu_t)) - \mu_t \cdot b(x^*(\gamma, \mu_t)) + \mu_t \cdot \beta_t]$$

$$\leq \mathbb{E}_{\tilde{\gamma} \sim \tilde{\mathcal{P}}, \gamma \sim \mathcal{P}}[\tilde{f}(x^*(\tilde{\gamma}, \mu_t)) - \mu_t \cdot \tilde{b}(x^*(\tilde{\gamma}, \mu_t)) + \mu_t \cdot \beta_t - \{f(x^*(\tilde{\gamma}, \mu_t)) - \mu_t \cdot b(x^*(\tilde{\gamma}, \mu_t)) + \mu_t \cdot \beta_t\}]$$

18

$$\leq \mathbb{E}_{(\gamma,\tilde{\gamma})\sim Q}\left[\left|f(x^*(\tilde{\gamma},\mu_t)) - \mu_t \cdot b(x^*(\tilde{\gamma},\mu_t)) - \left\{\tilde{f}(x^*(\tilde{\gamma},\mu_t)) - \mu_t \cdot \tilde{b}(x^*(\tilde{\gamma},\mu_t))\right\}\right|\right]$$

$$\leq (1+\kappa) \cdot \mathcal{W}(\mathcal{P},\tilde{\mathcal{P}}),$$

where the first inequality follows from the definition of $x^*(\gamma,\mu_t)$ and the second inequality follows from the fact that $(\mathcal{P},\tilde{\mathcal{P}})$ are the marginals of Q. As a consequence, we get

$$\sum_{t=1}^{T} D(\mu_t|\mathcal{P}_t,\beta_t) = \sum_{t=1}^{T} D(\mu_t|\tilde{\mathcal{P}}_t,\beta_t) - \sum_{t=1}^{T}\left\{D(\mu_t|\tilde{\mathcal{P}}_t,\beta_t) - D(\mu_t|\mathcal{P}_t,\beta_t)\right\}$$

$$\geq \sum_{t=1}^{T} D(\mu_t|\tilde{\mathcal{P}}_t,\beta_t) - (1+\kappa) \cdot \sum_{t=1}^{T} \mathcal{W}(\mathcal{P}_t,\tilde{\mathcal{P}}_t)$$

$$\geq \sum_{t=1}^{T} \mathrm{FLUID}(\tilde{\mathcal{P}}_t,\beta_t) - (1+\kappa) \cdot \sum_{t=1}^{T} \mathcal{W}(\mathcal{P}_t,\tilde{\mathcal{P}}_t), \tag{18}$$

where the second inequality follows from weak duality.

Next, observe that the definition of $\beta_t = \mathbb{E}_{\hat{\gamma}\sim\tilde{\mathcal{P}}_t}[\hat{b}^*(\tilde{\mu})]$ implies that $x^*(\tilde{\gamma}_t,\tilde{\mu})$ is a feasible to solution to the optimization problem which defines $\mathrm{FLUID}(\tilde{\mathcal{P}}_t,\beta_t)$. Hence,

$$\sum_{t=1}^{T}\mathrm{FLUID}(\tilde{\mathcal{P}}_t,\beta_t) \geq \sum_{t=1}^{T}\mathbb{E}_{\tilde{\gamma}_t\sim\tilde{\mathcal{P}}_t}\left[\tilde{f}_t(x^*(\tilde{\gamma}_t,\tilde{\mu}))\right]$$

$$= \sum_{t=1}^{T}\mathbb{E}_{\tilde{\gamma}_t\sim\tilde{\mathcal{P}}_t}\left[\tilde{f}_t(x^*(\tilde{\gamma}_t,\tilde{\mu})) - \tilde{\mu}\cdot\tilde{b}_t(x^*(\tilde{\gamma}_t,\tilde{\mu}))\right] + \tilde{\mu}\cdot\sum_{t=1}^{T}\mathbb{E}_{\tilde{\gamma}_t\sim\tilde{\mathcal{P}}_t}\left[\tilde{b}_t(x^*(\tilde{\gamma}_t,\tilde{\mu}))\right].$$

Let $\{x_t(\cdot)\}_t$ be an optimal solution for $\mathrm{FLUID}(\{\mathcal{P}_t\})$. Then, for all $t \in [T]$, we have

$$\mathbb{E}_{\tilde{\gamma}_t\sim\tilde{\mathcal{P}}_t}\left[\tilde{f}_t(x^*(\tilde{\gamma}_t,\tilde{\mu})) - \tilde{\mu}\cdot\tilde{b}_t(x^*(\tilde{\gamma}_t,\tilde{\mu}))\right] = \mathbb{E}_{(\gamma_t,\tilde{\gamma}_t)\sim Q}\left[\tilde{f}_t(x^*(\tilde{\gamma}_t,\tilde{\mu})) - \tilde{\mu}\cdot\tilde{b}_t(x^*(\tilde{\gamma}_t,\tilde{\mu}))\right]$$

$$\geq \mathbb{E}_{(\gamma_t,\tilde{\gamma}_t)\sim Q}\left[\tilde{f}_t(x_t(\gamma_t)) - \tilde{\mu}\cdot\tilde{b}_t(x(\gamma_t))\right]$$

$$\geq \mathbb{E}_{(\gamma_t,\tilde{\gamma}_t)\sim Q}\left[f_t(x_t(\gamma_t)) - \tilde{\mu}\cdot b_t(x(\gamma_t))\right] - (1+\kappa)\cdot\mathcal{W}(\mathcal{P}_t,\tilde{\mathcal{P}}_t)$$

$$= \mathbb{E}_{\gamma_t\sim\mathcal{P}_t}\left[f_t(x_t(\gamma_t)) - \tilde{\mu}\cdot b_t(x(\gamma_t))\right] - (1+\kappa)\cdot\mathcal{W}(\mathcal{P}_t,\tilde{\mathcal{P}}_t),$$

where the first inequality follows from the definition of $x^*(\tilde{\gamma}_t,\tilde{\mu})$ and the second inequality follows from (17). Therefore,

$$\sum_{t=1}^{T}\mathrm{FLUID}(\tilde{\mathcal{P}}_t,\beta_t)$$

$$\geq \sum_{t=1}^{T}\mathbb{E}_{\gamma_t}\left[f_t(x_t(\gamma_t)) - \tilde{\mu}\cdot b_t(x(\gamma_t))\right] - (1+\kappa)\cdot\sum_{t=1}^{T}\mathcal{W}(\mathcal{P}_t,\tilde{\mathcal{P}}_t) + \tilde{\mu}\cdot\sum_{t=1}^{T}\mathbb{E}_{\tilde{\gamma}_t}\left[\tilde{b}_t(x^*(\tilde{\gamma}_t,\tilde{\mu}))\right]$$

$$= \sum_{t=1}^{t}\mathbb{E}_{\gamma_t\sim\mathcal{P}_t}[f_t(x(\gamma_t))] - \tilde{\mu}\cdot\left(\sum_{t=1}^{T}\mathbb{E}_{\gamma_t}[b_t(x(\gamma_t))] - \sum_{t=1}^{T}\mathbb{E}_{\tilde{\gamma}_t}[\tilde{b}_t^*(\tilde{\mu})]\right) - (1+\kappa)\cdot\sum_{t=1}^{T}\mathcal{W}(\mathcal{P}_t,\tilde{\mathcal{P}}_t)$$

$$\geq \mathrm{FLUID}(\{\mathcal{P}_t\}_t) - \tilde{\mu}\cdot\left(B - \sum_{t=1}^{T}\beta_t\right) - (1+\kappa)\cdot\sum_{t=1}^{T}\mathcal{W}(\mathcal{P}_t,\tilde{\mathcal{P}}_t),$$

where the second inequality follows from the feasibility of the optimal solution $\{x_t(\cdot)\}_t$, i.e., $\sum_{t=1}^{T}\mathbb{E}_{\gamma_t}[b_t(x(\gamma_t))] \leq B$. Combining this with (18) yields

$$\sum_{t=1}^{T} D(\mu_t|\mathcal{P}_t,\beta_t) \geq \mathrm{FLUID}(\{\mathcal{P}_t\}_t) - \tilde{\mu}\cdot\left(B - \sum_{t=1}^{T}\beta_t\right) - 2(1+\kappa)\cdot\sum_{t=1}^{T}\mathcal{W}(\mathcal{P}_t,\tilde{\mathcal{P}}_t),$$

as required. $\square$

G.3. Proof of Lemma 4.4

Proof of Lemma 4.4. The definition of x and x' implies

$$f(x) - \mu \cdot b(x) \geq f(x') - \mu \cdot b(x') \quad \text{and} \quad f(x') - \mu' \cdot b(x') \geq f(x) - \mu' \cdot b(x) .$$

Combining the two inequalities, we get

$$f(x) - \mu \cdot b(x) - \{f(x) - \mu' \cdot b(x)\} \geq f(x') - \mu \cdot b(x') - \{f(x') - \mu' \cdot b(x')\}$$
$$\implies (\mu - \mu') \cdot (b(x') - b(x)) \geq 0 .$$

The lemma follows from the last inequality because $\mu - \mu' > 0$. $\qquad\square$

G.4. Proof of Lemma 4.5

Proof of Lemma 4.5. Define

$$q(\mu) := \mu \cdot B + \sum_{t=1}^{T} \max_{x \in \mathcal{X}} \left\{ \tilde{f}_t(x) - \mu \cdot \tilde{b}_t(x) \right\} \quad \text{and} \quad q^{(-s)}(\mu) := \mu \cdot B + \sum_{t \neq s} \max_{x \in \mathcal{X}} \left\{ \tilde{f}_t(x) - \mu \cdot \tilde{b}_t(x) \right\} .$$

We start by proving $\tilde{\mu} \geq \tilde{\mu}^{(-s)}$. For contradiction, suppose $\tilde{\mu} < \tilde{\mu}^{(-s)}$. Consider the following two cases:

- Case I: $0 \in \operatorname{argmax}_{x \in \mathcal{X}} \tilde{f}_s(x) - \tilde{\mu} \cdot \tilde{b}_s(x)$. Then, we must have $0 \in \operatorname{argmax}_{x \in \mathcal{X}} \tilde{f}_s(x) - \mu \cdot \tilde{b}_s(x)$ for all $\mu \geq \tilde{\mu}$. This is because, for $\mu \geq \tilde{\mu}$, Lemma 4.4 implies that $\tilde{b}_s(x') \leq \tilde{b}_s(0) = 0$ for all $x' \in \operatorname{argmax}_{x \in \mathcal{X}} \tilde{f}_s(x) - \mu \cdot \tilde{b}_s(x)$, and $\tilde{f}_s(x) \leq \kappa \cdot \tilde{b}_s(x)$ for all $x \in \mathcal{X}$. Therefore, $q(\mu) = q^{(-s)}(\mu)$ for all $\mu \geq \tilde{\mu}$. Since $\tilde{\mu}$ is a minimizer of $q(\cdot)$ and $\tilde{\mu}^{(-s)} > \tilde{\mu}$, we get that

$$q^{(-s)}(\tilde{\mu}) = q(\tilde{\mu}) \leq q(\tilde{\mu}^{(-s)}) = q^{(-s)}\left(\tilde{\mu}^{(-s)}\right) .$$

  On the other hand, $\tilde{\mu}^{(-s)}$ is a minimizer of $q^{(-s)}(\cdot)$, which implies $q^{(-s)}\left(\tilde{\mu}^{(-s)}\right) \leq q^{(-s)}(\tilde{\mu})$. Therefore, $q^{(-s)}\left(\tilde{\mu}^{(-s)}\right) \leq q^{(-s)}(\tilde{\mu})$, which contradicts the fact that $\tilde{\mu}^{(-s)}$ is the smallest minimizer of $q^{(-s)}(\cdot)$.

- Case II: $0 \notin \operatorname{argmax}_{x \in \mathcal{X}} \tilde{f}_s(x) - \tilde{\mu} \cdot \tilde{b}_s(x)$. Since $f_s(x) \leq \kappa \cdot b_s(x)$ for all $x \in \mathcal{X}$, we get that $\tilde{b}_s(x') > 0$ for all $x' \in \operatorname{argmax}_{x \in \mathcal{X}} \tilde{f}_s(x) - \tilde{\mu} \cdot \tilde{b}_s(x)$. Consider any sequences of optimal action sequences $\{x_t\}_t$ and $\{x_t^{(-s)}\}_t$ such that for all $t \in [T]$, we have

$$x_t \in \operatorname*{argmax}_{x \in \mathcal{X}} \tilde{f}_t(x) - \tilde{\mu} \cdot \tilde{b}_t(x) \quad \text{and} \quad x_t^{(-s)} \in \operatorname*{argmax}_{x \in \mathcal{X}} \tilde{f}_t(x) - \tilde{\mu}^{(-s)} \cdot \tilde{b}_t(x) .$$

  Then, Lemma 4.4 implies that $\tilde{b}_t(x_t) \geq \tilde{b}_t(x_t^{(-s)})$ for all $t \neq s$. Therefore,

$$B - \sum_{t=1}^{T} \tilde{b}_t(x_t) = \left\{ B - \sum_{t \neq s} \tilde{b}_t(x_t) \right\} - b_t(x_t) < B - \sum_{t \neq s} \tilde{b}_t(x_t) \leq B - \sum_{t \neq s} \tilde{b}_t(x_t^{(-s)}) . \tag{19}$$

  Now observe that, since $q(\cdot)$ (and $q^{(-s)}(\cdot)$) are the maxima of a collection of linear functions, its sub-gradient is given by the convex hull of gradients of all the linear functions which are binding (for example, see Chapter 5 of Bertsekas 2009). Therefore, $\partial q(\tilde{\mu})$ (and $\partial q^{(-s)}\left(\tilde{\mu}^{(-s)}\right)$) is a convex hull of terms of the form $B - \sum_{t=1}^{T} \tilde{b}_t(x_t)$ for some optimal action sequence $\{x_t\}_t$ (and $\{x_t^{(-s)}\}_t$). Since $\tilde{\mu}^{(-s)} > \tilde{\mu} \geq 0$, first-order optimality conditions imply that $0 \in \partial q^{(-s)}\left(\tilde{\mu}^{(-s)}\right)$. Therefore, (19) implies that $v < 0$ for all $v \in \partial q(\tilde{\mu})$. This contradicts the optimality of $\tilde{\mu}$ for $q(\cdot)$.

As we have obtained a contradiction in both cases, we get that $\tilde{\mu} \geq \tilde{\mu}^{(-s)}$, as required. Moreover, $\lambda_t \leq \lambda_t^{(-s)}$ for all $t \neq s$ follows immediately from Lemma 4.4. Hence, to finish the proof, it suffices to show the final inequality in the following chain:

$$\sum_{t=1}^{s-1} \left| \lambda_t^{(-s)} - \lambda_t \right| \leq \sum_{t \neq s} \left| \lambda_t^{(-s)} - \lambda_t \right| = \sum_{t \neq s} \lambda_t^{(-s)} - \sum_{t \neq s} \lambda_t \leq 3\bar{b} . \tag{20}$$

Note that, Lemma 3.4 implies that at least one of the following conditions hold

1. $\tilde{\mu} = 0$ and $\sum_{t=1}^{T} \lambda_t \le B + \bar{b}$.

2. $\left| B - \sum_{t=1}^{T} \tilde{\lambda}_t \right| \le \bar{b}$.

If $\tilde{\mu} = 0$, then $\tilde{\mu}^{(-s)} = 0$ because $\tilde{m}u^{(-s)} \le \tilde{\mu}$. Therefore, in that case $\lambda_t^{(-s)} = \lambda_t = \tilde{b}_t^*(0)$ for all $t \ne s$ and (20) follows.

Suppose $\left| B - \sum_{t=1}^{T} \tilde{\lambda}_t \right| \le \bar{b}$. Observe that Lemma 3.4 applied to the trace $\{\hat{\gamma}_t\}_t$, where $\hat{\gamma}_t = \tilde{\gamma}_t$ for all $t \ne s$ and $\hat{\gamma}_s = (0, 0)$, implies that at least one of the following conditions hold:

1. $\tilde{\mu}^{(-s)} = 0$ and $\sum_{t \ne s} \lambda_t^{(-s)} \le B + \bar{b}$.

2. $\left| B - \sum_{t \ne s} \tilde{\lambda}_t^{(-s)} \right| \le \bar{b}$.

Therefore, $\sum_{t \ne s} \lambda_t^{(-s)} - \sum_{t \ne s} \lambda_t \le B + \bar{b} - \sum_{t \ne s} \lambda_t \le \bar{b} + \bar{b} + \lambda_s \le 3\bar{b}$, as required to establish (20). $\qquad \square$

G.5. Proof of Lemma 4.6

Proof of Lemma 4.6. It is known that FTRL is equivalent to Lazy Online Mirror Descent (for example, see Hazan et al. 2016). In particular, if we let $V_h(x, y) = h(x) - h(y) - \nabla h(y)^\top (x - y)$ denote the Bregman divergence w.r.t. $h(\cdot)$, then the FTRL update (3) of Algorithm 2 can be equivalently written as:

$$\theta_s = \nabla h(\mu_s)$$
$$\theta_{s+1} = \theta_t - \eta \cdot g_s = \theta_t - \eta \cdot (\lambda_s - b_s(x'_t))$$
$$\mu_{t+1} = \operatorname*{argmin}_{\mu \in [0,\kappa]} V_h(\mu, (\nabla h)^{-1}(\theta_{s+1})).$$

where $x'_t \in \operatorname{argmax}_{x \in \mathcal{X}} f_s(x) - \mu_s \cdot b_s(x)$. We will use $\{\mu'_t\}_t$ and $\{\theta'_t\}_t$ to represent the dual and mirror iterates of Algorithm 2 with target sequence $\{\lambda'_t\}_t$:

$$\theta'_s = \nabla h(\mu'_s)$$
$$\theta'_{s+1} = \theta'_t - \eta \cdot g_s = \theta'_t - \eta \cdot (\lambda'_s - b_s(y'_t))$$
$$\mu'_{t+1} = \operatorname*{argmin}_{\mu \in [0,\kappa]} V_h(\mu, (\nabla h)^{-1}(\theta'_{s+1})).$$

where $y'_t \in \operatorname{argmax}_{x \in \mathcal{X}} f_s(x) - \mu'_s \cdot b_s(x)$.

We will first use induction on s to prove the following statement,

$$|\theta_s - \theta'_s| \le \eta \cdot \left\{ \sum_{t=1}^{s-1} |\lambda_t - \lambda'_t| \right\} + \eta \cdot \bar{b}. \tag{21}$$

The base case $s = 1$ follows directly from our assumption that the initial iterates $\theta_1 = \nabla h(\mu_1) = \nabla h(\mu'_1) = \theta_2$ are the same.

Suppose (21) holds for $s \in [T - 1]$ (Induction Hypothesis). Define $\theta_{s+1/2} = \theta_s + \eta \cdot b_s(x'_t)$ and $\theta'_{s+1/2} = \theta'_s + \eta \cdot b_s(y'_t)$. W.l.o.g. assume that $\theta_s \ge \theta'_s$. Due to the invertibility of ∇h, we get that $\mu_s \ge \mu'_s$, and consequently Lemma 4.4 implies $b(x'_t) \le b(y'_t)$. Consider the following cases:

- Case I: $\theta'_{s+1/2} \le \theta_{s+1/2}$. Then, $\theta_{s+1/2} - \theta'_{s+1/2} = \theta_s - \theta'_s + \eta \cdot (b(x'_t) - b(y'_t)) \le \theta_s - \theta'_s$ because $b(x'_t) \le b(y'_t)$.

- Case II: $\theta'_{s+1/2} \ge \theta_{s+1/2}$. Then, $\theta'_{s+1/2} - \theta_{s+1/2} = \theta'_s - \theta_s + \eta \cdot (b(x'_t) - b(y'_t)) \le \eta \cdot \bar{b}$ because $\theta'_s \le \theta_s$ and $b(y'_t) - b(x'_t) \le \bar{b}$.

Therefore, in both cases we have

$$|\theta_{s+1/2} - \theta'_{s+1/2}| \le \max\{\theta_s - \theta'_s, \eta \cdot \bar{b}\} \le \eta \cdot \left\{ \sum_{t=1}^{s-1} |\lambda_t - \lambda'_t| \right\} + \eta \cdot \bar{b}.$$

where we used the induction hypothesis in the second inequality. Consequently, we can write

$$\begin{aligned}
|\theta_{s+1} - \theta'_{s+1}| &= \left| \theta_{s+1/2} - \eta \cdot \lambda_s + (\theta'_{s+1/2} - \eta \cdot \lambda'_s) \right| \\
&\le |\theta_{s+1/2} - \theta'_{s+1/2}| + \eta \cdot |\lambda_s - \lambda'_s| \\
&\le \eta \cdot \left\{ \sum_{t=1}^{s} |\lambda_t - \lambda'_t| \right\} + \eta \cdot \bar{b}.
\end{aligned}$$

Hence, we have established (21) for all $s \in [T]$. Now, since h is σ-strongly convex and differentiable, we have

$$\nabla h(x) - \nabla h(y) \ge \sigma \cdot (x - y) \qquad \forall \, x \ge y.$$

Therefore, we have

$$\left| (\nabla h)^{-1}(\theta_s) - (\nabla h)^{-1}(\theta'_s) \right| \le \frac{1}{\sigma} \cdot |\theta_s - \theta'_s|. \tag{22}$$

To finish the proof, we will use the fact that Bregman projections are contractions in one dimensions, which we prove next. Consider any $x < 0$, then for any $\mu \in [0, \kappa]$, we have

$$V_h(\mu, x) - V_h(0, x) = h(\mu) - h(0) - \nabla h(x)^\top (\mu - 0) \ge h(\mu) - h(0) - \nabla h(0)^\top (\mu - 0) \ge 0,$$

where the inequality follows from $\nabla h(0) \ge \nabla h(x)$ (convexity of $h(\cdot)$). Therefore, $\operatorname{argmin}_{\mu \in [0,\kappa]} V_h(\mu, x) = 0 = \operatorname{argmin}_{\mu \in [0,\kappa]} |\mu - x|$. Similarly, for $x > \kappa$ and $\mu \in [0, \kappa]$, we have

$$V_h(\mu, x) - V_h(\kappa, x) = h(\mu) - h(\kappa) - \nabla h(x)^\top (\mu - \kappa) \ge h(\mu) - h(\kappa) - \nabla h(\kappa)^\top (\mu - \kappa) \ge 0,$$

where the inequality follows from $\nabla h(x) \ge \nabla h(\kappa)$ (convexity of $h(\cdot)$). Therefore, $\operatorname{argmin}_{\mu \in [0,\kappa]} V_h(\mu, x) = \kappa = \operatorname{argmin}_{\mu \in [0,\kappa]} |\mu - x|$. Consequently, we have shown that $\operatorname{argmin}_{\mu \in [0,\kappa]} V_h(\mu, x) = \operatorname{argmin}_{\mu \in [0,\kappa]} |\mu - x|$, i.e., the Bregman project is identical to the Euclidean projection in one dimension. Since Euclidean projection is a contraction, we get

$$\begin{aligned}
|\mu_s - \mu'_s| &= \left| \operatorname*{argmin}_{\mu \in [0,\kappa]} V_h(\mu, (\nabla h)^{-1}(\theta_{s+1})) - \operatorname*{argmin}_{\mu \in [0,\kappa]} V_h(\mu, (\nabla h)^{-1}(\theta'_{s+1})) \right| \\
&= \left| \operatorname*{argmin}_{\mu \in [0,\kappa]} |\mu - (\nabla h)^{-1}(\theta_{s+1})| - \operatorname*{argmin}_{\mu \in [0,\kappa]} |\mu - (\nabla h)^{-1}(\theta'_{s+1})| \right| \\
&\le \left| (\nabla h)^{-1}(\theta_s) - (\nabla h)^{-1}(\theta'_s) \right|.
\end{aligned}$$

Finally, combining this with (21) and (22), we get

$$|\mu_s - \mu'_s| \le \frac{\eta}{\sigma} \cdot \left\{ \sum_{t=1}^{s-1} |\lambda_t - \lambda'_t| \right\} + \frac{\eta}{\sigma} \cdot \bar{b},$$

as required. $\qquad \square$

G.6. Proof of Lemma 4.7

Proof of Lemma 4.7. Using the definitions of λ_s and β_s, we can write

$$\mathbb{E}\left[\lambda_s \big| \tilde{\mu}^{(-s)} \right] = \mathbb{E}\left[\tilde{b}_s^*(\tilde{\mu}) \big| \tilde{\mu}^{(-s)} \right] \quad \text{and} \quad \mathbb{E}\left[\beta_s \big| \tilde{\mu}^{(-s)} \right] = \mathbb{E}\left[\mathbb{E}_{\hat{\gamma} \sim \tilde{\mathcal{P}}_t}\left[\hat{b}_s^*(\tilde{\mu}) \right] \big| \tilde{\mu}^{(-s)} \right].$$

Fix a trace $\{\tilde{\gamma}_t\}_t$. Observe that for any request

$$x^*(\tilde{\gamma}_s, \tilde{\mu}) = \begin{cases} \bar{x} & \text{if } \operatorname{coeff}(\tilde{f}_s) - \tilde{\mu} \cdot \operatorname{coeff}(\tilde{b}_s) \geq 0 \text{ and } \operatorname{coeff}(\tilde{f}_s) \neq 0 \\ 0 & \text{otherwise} \end{cases},$$

and

$$x^*(\tilde{\gamma}_s, \tilde{\mu}^{(-s)}) = \begin{cases} \bar{x} & \text{if } \operatorname{coeff}(\tilde{f}_s) - \tilde{\mu}^{(-s)} \cdot \operatorname{coeff}(\tilde{b}_s) \geq 0 \text{ and } \operatorname{coeff}(\tilde{f}_s) \neq 0 \\ 0 & \text{otherwise} \end{cases}.$$

From Lemma 4.5, we know that $\tilde{\mu}^{(-s)} \leq \tilde{\mu}$. Now, if $\operatorname{coeff}(\tilde{f}_s) = 0$, then $x^*(\tilde{\gamma}_s, \tilde{\mu}) = x^*(\tilde{\gamma}_s, \tilde{\mu}^{(-s)}) = 0$. Assume that $\operatorname{coeff}(\tilde{f}_s) > 0$ (and thus $\operatorname{coeff}(\tilde{b}_s) > 0$ because $f_s(x) \leq \kappa \cdot b(x)$), let $A := \{\mu \geq 0 \mid \operatorname{coeff}(\tilde{f}_s) - \mu \cdot \operatorname{coeff}(\tilde{b}_s) < 0\}$ be the set of all dual variables that lead to $x^*(\tilde{\gamma}_s, \mu) = 0$.

Define the dual functions:

$$q(\mu) := \mu \cdot B + \sum_{t=1}^{T} \max_{x \in \mathcal{X}} \left\{ \tilde{f}_t(x) - \mu \cdot \tilde{b}_t(x) \right\} \quad \text{and} \quad q^{(-s)}(\mu) := \mu \cdot B + \sum_{t \neq s} \max_{x \in \mathcal{X}} \left\{ \tilde{f}_t(x) - \mu \cdot \tilde{b}_t(x) \right\}.$$

For contradiction, suppose $\tilde{\mu} \in A$ and $\tilde{\mu}^{(-s)} \notin A$. Since A is open, there exists a point $\mu \in A$ such that $\mu = \alpha \cdot \tilde{\mu} + (1 - \alpha) \cdot \tilde{\mu}^{(-s)}$ for some $\alpha \in (0, 1)$. Moreover, observe that the minimality of $\tilde{\mu}^{(-s)}$ implies $q^{(-s)}\left(\tilde{\mu}^{(-s)}\right) \leq q^{(-s)}(\mu)$ and $q^{(-s)}\left(\tilde{\mu}^{(-s)}\right) \leq q^{(-s)}(\tilde{\mu})$. Therefore, as $q^{(-s)}$ is convex, we get

$$q^{(-s)}(\mu) \leq \alpha \cdot q^{(-s)}(\tilde{\mu}) + (1 - \alpha) \cdot q^{(-s)}\left(\tilde{\mu}^{(-s)}\right) \leq \alpha \cdot q^{(-s)}(\tilde{\mu}) + (1 - \alpha) \cdot q^{(-s)}(\tilde{\mu}) = q^{(-s)}(\tilde{\mu}).$$

Now, observe that $q(\mu) = q^{(-s)}(\mu)$ for all $\mu \in A$. Therefore, $q(\mu) \leq q(\tilde{\mu})$, which contradicts the fact that $\tilde{\mu}$ is the smallest minimizer of $q(\cdot)$. Hence, either $\tilde{\mu}, \tilde{\mu}^{(-s)} \in A$ or $\tilde{\mu}, \tilde{\mu}^{(-s)} \in A^c$, and as a consequence, we get $\tilde{b}_s^*(\tilde{\mu}) = \tilde{b}_s^*\left(\tilde{\mu}^{(-s)}\right)$. Furthermore, combining $\tilde{\mu} \geq \tilde{\mu}^{(-s)}$ (from Lemma 4.5) and Lemma 4.4, we also get $\hat{b}_s^*\left(\tilde{\mu}^{(-s)}\right) \geq \hat{b}_s^*(\tilde{\mu})$ for every $\hat{\gamma}_s \in \mathcal{S}$. Therefore,

$$\begin{aligned} \mathbb{E}\left[\lambda_s \big| \tilde{\mu}^{(-s)}\right] &= \mathbb{E}\left[\tilde{b}_s^*(\tilde{\mu}) \Big| \tilde{\mu}^{(-s)}\right] \\ &= \mathbb{E}\left[\tilde{b}_s^*\left(\tilde{\mu}^{(-s)}\right) \Big| \tilde{\mu}^{(-s)}\right] \\ &= \mathbb{E}_{\hat{\gamma}_s \sim \tilde{\mathcal{P}}_s}\left[\hat{b}_s^*\left(\tilde{\mu}^{(-s)}\right) \Big| \tilde{\mu}^{(-s)}\right] \\ &\geq \mathbb{E}\left[\mathbb{E}_{\hat{\gamma}_s \sim \tilde{\mathcal{P}}_s}\left[\hat{b}_s^*(\tilde{\mu}) \Big| \tilde{\mu}^{(-s)}\right]\right] \\ &= \mathbb{E}\left[\beta_s \big| \tilde{\mu}^{(-s)}\right], \end{aligned}$$

where the third equality follows from the fact that $\tilde{\gamma}_s$ and $\tilde{\mu}^{(-s)}$ are independent of each other, which allows us to rename the variable from $\tilde{\gamma}_s \sim \tilde{\mathcal{P}}_s$ to $\hat{\gamma}_s \sim \tilde{\mathcal{P}}_s$. We combine this with the Tower Property of conditional expectations to finish the proof:

$$\mathbb{E}\left[\mu_s^{(-s)} \cdot (\beta_s - \lambda_s)\right] = \mathbb{E}\left[\mu_s^{(-s)} \cdot \mathbb{E}\left[(\beta_s - \lambda_s)\Big| \mu_s^{(-s)}\right]\right] = \mathbb{E}\left[\mu_s^{(-s)} \cdot \left(\mathbb{E}\left[\beta_s \Big| \mu_s^{(-s)}\right] - \mathbb{E}\left[\lambda_s \Big| \mu_s^{(-s)}\right]\right)\right] \leq 0.$$

$\square$

G.7. Proof of Theorem 4.8

Proof of Theorem 4.8. Theorem 4.1 and (6) together imply that, with probability at least $1 - 1/T^2$, we have

$$\operatorname{Regret}(A) = \operatorname{FLUID}(\{\mathcal{P}_t\}_t) - \mathbb{E}_{\{\gamma_t\}_t \sim \prod_t \mathcal{P}_t}[R(A|\{\gamma_t\}_t)] \leq R_1 + R_2 + R_3 + \kappa \cdot r(T) + \kappa \bar{b} + 2(1 + \kappa) \cdot \sum_{t=1}^{T} \mathcal{W}(\mathcal{P}_t, \tilde{\mathcal{P}}_t).$$

23

From our choice of step size $\eta = \sqrt{d_R/T}$ and the observation that $\bar{\lambda} = \max_t \lambda_t \leq \bar{b}$, we get that

$$R_1 = \kappa\bar{b} + \frac{2(\bar{b} + \bar{\lambda})^2}{\sigma} \cdot \eta T + \frac{d_R}{\eta} \leq \kappa\bar{b} + \left(\frac{8\bar{b}^2}{\sigma} + 1\right) \cdot \sqrt{d_r T} \,.$$

From (10), we know that

$$R_2 = \kappa \cdot \left(\left\{\sum_{t=1}^{T} \lambda_t\right\} - B\right)^+ \leq \kappa\bar{b} \,.$$

Moreover, from Lemma 4.7 and $\eta = \sqrt{d_R/T}$, we know that

$$R_3 = \sum_{s=1}^{T} \mathbb{E}\left[\mu_s \cdot (\beta_s - \lambda_s)\right] \leq \frac{4\eta\bar{b}^2}{\sigma} \cdot T = \frac{4\bar{b}^2}{\sigma} \cdot \sqrt{d_r T} \quad .$$

Combining the above inequalities and plugging in $r(T) = 8\bar{b}\sqrt{T\log(T)}$, we get that

$$\text{Regret}(A) \leq 3\kappa\bar{b} + \left(\frac{12\bar{b}^2}{\sigma} + 1\right) \cdot \sqrt{d_r T} + 8\kappa\bar{b}\sqrt{T\log(T)} + 2(1+\kappa) \cdot \sum_{t=1}^{T} \mathcal{W}(\mathcal{P}_t, \tilde{\mathcal{P}}_t)$$

$$\leq$$

with probability at least $1 - 1/T^2$. On the other than, we always have $\text{Regret}(T) \leq \bar{f}T \leq \kappa\bar{b}T$. Hence, we get

$$\text{Regret}(A) \leq \left(1 - \frac{1}{T^2}\right) \cdot \left[3\kappa\bar{b} + \left(\frac{12\bar{b}^2}{\sigma} + 1\right) \cdot \sqrt{d_r T} + 8\kappa\bar{b}\sqrt{T\log(T)} + 2(1+\kappa) \cdot \sum_{t=1}^{T} \mathcal{W}(\mathcal{P}_t, \tilde{\mathcal{P}}_t)\right] + \frac{1}{T^2} \cdot \kappa\bar{b}T$$

$$\leq 4\kappa\bar{b}\sqrt{T\log(T)} + \left(\frac{12\bar{b}^2}{\sigma} + 1\right) \cdot \sqrt{d_r} \cdot \sqrt{T\log(T)} + 8\kappa\bar{b}\sqrt{T\log(T)} + 2(1+\kappa) \cdot \sum_{t=1}^{T} \mathcal{W}(\mathcal{P}_t, \tilde{\mathcal{P}}_t)$$

$$= C_1 \cdot \sqrt{T\log(T)} + C_2 \cdot \sum_{t=1}^{T} \mathcal{W}(\mathcal{P}_t, \tilde{\mathcal{P}}_t)$$

where $C_1 = \frac{12\bar{b}^2\sqrt{d_R}}{\sigma} + \sqrt{d_R} + 12\kappa\bar{b}$ and $C_2 = 2(1+\kappa)$. $\qquad\square$