A PROOFS: STRICT ROS CONSTRAINT

The goal of this section is to prove Proposition 3.1, Proposition 4.1, and Proposition 4.2.

The proof of Proposition 3.1 essentially follows the outline in [Balseiro et al. 2020]. We need two components: an upper bound on the expected optimal value followed by a lower bound on the expected reward of Algorithm 3.1. To prove the former, we go through the primal-dual framework, for which need the following technical definitions.

Definition A.1. *We define the following dual variables.*

- *Let $f_t(b) := v_t \cdot x_t(b)$, recall g_t from Equation (2.3), and for some fixed $\lambda \geq 0$, define*
$$f^\star_{t,\text{RoS}}(\lambda) := \max_{b\geq 0}\left[f_t(b) + \lambda \cdot g_t(b)\right]. \tag{A.1}$$
- *Let $f^\star$ be defined as in Equation (A.1). Then we define the following dual variable parametrized by the input distribution $\mathcal{P}$.*
$$\overline{\mathcal{D}}_{\text{RoS}}(\lambda|\mathcal{P}) := \mathbb{E}_{(v,p)\sim\mathcal{P}}\left[f^\star_{\text{RoS}}(\lambda)\right]. \tag{A.2}$$

We state and prove the following upper bound on the expected optimal reward and lower bound on the expected reward of our algorithm.

Proposition A.1. *Recall $\overline{\mathcal{D}}_{\text{RoS}}(\lambda|\mathcal{P})$ as defined in Equation (A.2). Then the optimum value $\text{Reward}(\text{Opt}, \overrightarrow{\gamma})$ for Problem 2.2 defined for a sequence $\overrightarrow{\gamma}_\ell \sim \mathcal{P}^\ell$ of ℓ requests satisfies the inequality*
$$\mathbb{E}_{\overrightarrow{\gamma}_\ell\sim\mathcal{P}^\ell}\left[\text{Reward}(\text{Opt}, \overrightarrow{\gamma}_\ell)\right] \leq \ell \cdot \min_{\lambda\geq 0}\overline{\mathcal{D}}_{\text{RoS}}(\lambda|\mathcal{P}).$$

Proof. We have, by the definition of $\text{Reward}(\text{Opt}, \overrightarrow{\gamma})$ in Equation (2.7) and the definition of g_t in Equation (2.3) that
$$\begin{aligned}\text{Reward}(\text{Opt}, \overrightarrow{\gamma}_\ell) &= \max_{\{b_t\}:\sum_{t=1}^{\ell} g_t(b_t)\geq 0}\sum_{t=1}^{\ell} f_t(b_t)\\ &= \max_{\{b_1,b_2,\dots,b_\ell\}}\left\{\sum_{t=1}^{\ell} f_t(b_t) + \min_{\lambda\geq 0}\lambda\cdot\left[\sum_{t=1}^{\ell} g_t(b_t)\right]\right\}\\ &= \max_{\{b_1,b_2,\dots,b_\ell\}}\min_{\lambda\geq 0}\left\{\sum_{t=1}^{\ell}\left[f_t(b_t) + \lambda\cdot g_t(b_t)\right]\right\}.\end{aligned}$$

By Sion's min-max theorem and by the definition of $f^\star_{t,\text{RoS}}$ in Equation (A.1), we have that
$$\begin{aligned}\text{Reward}(\text{Opt}, \overrightarrow{\gamma}_\ell) &\leq \min_{\lambda\geq 0}\max_{\{b_1,b_2,\dots,b_\ell\}}\left\{\sum_{t=1}^{\ell}\left[f_t(b_t) + \lambda\cdot g_t(b_t)\right]\right\}\\ &\leq \min_{\lambda\geq 0}\left\{\sum_{t=1}^{\ell}\max_{b_t}\left[f_t(b_t) + \lambda\cdot g_t(b_t)\right]\right\}\\ &= \min_{\lambda\geq 0}\left\{\sum_{t=1}^{\ell} f^\star_{t,\text{RoS}}(\lambda)\right\} \leq \sum_{t=1}^{\ell} f^\star_{t,\text{RoS}}(\lambda'),\ \text{for any } \lambda'\geq 0.\end{aligned} \tag{A.3}$$

Now taking expectations[5] on both sides of Inequality (A.3) and using the definition of $\overline{\mathcal{D}}_{\text{RoS}}$ from Equation (A.2) gives, for any fixed $\lambda' \geq 0$,
$$\begin{aligned}\mathbb{E}_{\overrightarrow{\gamma}_\ell\sim\mathcal{P}^\ell}\left[\text{Reward}(\text{Opt}, \overrightarrow{\gamma}_\ell)\right] &\leq \mathbb{E}_{\overrightarrow{\gamma}_\ell\sim\mathcal{P}^\ell}\left[\sum_{t=1}^{\ell} f^\star_{t,\text{RoS}}(\lambda')\right]\\ &= \sum_{t=1}^{\ell}\mathbb{E}_{\overrightarrow{\gamma}_\ell\sim\mathcal{P}^\ell}\left[f^\star_{t,\text{RoS}}(\lambda')\right]\\ &= \ell\cdot\overline{\mathcal{D}}_{\text{RoS}}(\lambda'|\mathcal{P}).\end{aligned}$$

where the final equation crucially uses the fact that all t requests are drawn i.i.d. from the same distribution $\mathcal{P}$ and the argument $\lambda' > 0$ is fixed. Therefore, in particular, the preceding inequality holds for the specific λ' that minimizes the right-hand side, thus finishing the proof. □

[5]The expectation is over the randomness in the sequence of requests $\overrightarrow{\gamma}_\ell$ over the time horizon ℓ.

Proposition A.2. *For some fixed number r, let $\overline{\lambda}_r := \frac{1}{r}\sum_{t=1}^{r}\lambda_t$, where λ_t are the dual iterates in Algorithm 3.1. Then, the reward (see Equation (2.6)) of Algorithm 3.1 is lower bounded as*
$$\begin{aligned}&\text{Reward}(\textit{Algorithm } 3.1, \overrightarrow{\gamma}_r)\\ &\geq \mathbb{E}_{\overrightarrow{\gamma}_r\sim\mathcal{P}^r}\left[r\cdot\overline{\mathcal{D}}_{\text{RoS}}(\overline{\lambda}_r|\mathcal{P})\right] - \mathbb{E}_{\overrightarrow{\gamma}_r\sim\mathcal{P}^r}\left[\sum_{t=1}^{r}\lambda_t\cdot g_t(b_t)\right].\end{aligned}$$

Proof. Recall we have by Line 4 in Algorithm 3.1 and by the definition of $f^\star_{t,\text{RoS}}$ in Equation (A.1),
$$f_t(b_t) = f^\star_{t,\text{RoS}}(\lambda_t) - \lambda_t\cdot g_t(b_t).$$

Taking expectations on both sides by conditioning on the randomness until (and including) iteration $t-1$, we have,
$$\begin{aligned}\mathbb{E}\left[f_t(b_t)|\sigma_{t-1}\right] &= \mathbb{E}\left[f^\star_{t,\text{RoS}}(\lambda_t)|\sigma_{t-1}\right] - \mathbb{E}\left[\lambda_t\cdot g_t(b_t)|\sigma_{t-1}\right]\\ &= \mathbb{E}_{(v,p)\sim\mathcal{P}}\left[f^\star_{\text{RoS}}(\lambda_t)\right] - \mathbb{E}\left[\lambda_t\cdot g_t(b_t)|\sigma_{t-1}\right]\\ &= \overline{\mathcal{D}}_{\text{RoS}}(\lambda_t|\mathcal{P}) - \mathbb{E}\left[\lambda_t\cdot g_t(b_t)|\sigma_{t-1}\right],\end{aligned}$$

where the second step used the fact that in Algorithm 3.1, the λ_t is fixed when σ_{t-1} is, and the final step used the definition of $\overline{\mathcal{D}}_{\text{RoS}}$ in Equation (A.2). Summing over $t = 1, 2, \dots, r$ and taking expectations (this is a valid operation since r is a fixed number) gives,
$$\begin{aligned}&\mathbb{E}_{\overrightarrow{\gamma}_r\sim\mathcal{P}^r}\left[\sum_{t=1}^{r} f_t(b_t)\right]\\ &= \mathbb{E}_{\overrightarrow{\gamma}_r\sim\mathcal{P}^r}\left[\sum_{t=1}^{r}\overline{\mathcal{D}}_{\text{RoS}}(\lambda_t|\mathcal{P})\right] - \mathbb{E}_{\overrightarrow{\gamma}_r\sim\mathcal{P}^r}\left[\sum_{t=1}^{r}\lambda_t\cdot g_t(b_t)\right],\end{aligned}$$

and we may finish the proof by invoking convexity of $f^\star_{\text{RoS}}$. □

Proposition 3.1. *With i.i.d. inputs from a distribution $\mathcal{P}$ over a time horizon T, the regret of Algorithm 3.1 on Problem 2.2 is bounded by*
$$\text{Regret}(\textit{Algorithm } 3.1, \mathcal{P}^T) \leq \mathbb{E}_{\overrightarrow{\gamma}\sim\mathcal{P}^T}\left[\sum_{t\in[T]}\lambda_t\cdot g_t(b_t)\right],$$

where g_t and λ_t are as defined in Line 5 and Line 6 of Algorithm 3.1. We note that the bound on the right-hand side can be negative since Algorithm 3.1 does not guarantee $\sum_{t=1}^{T} g_t(b_t) \geq 0$, but we do show a bound on the worst-case constraint violation in Lemma 3.1.

Proof. We begin by restating the result from Proposition A.1 for an input sequence of length T:

$$\mathbb{E}_{\overrightarrow{\gamma}\sim\mathcal{P}^T}\left[\mathrm{Reward}(\mathrm{Opt},\overrightarrow{\gamma})\right] \leq T\cdot\min_{\lambda\geq 0}\overline{\mathcal{D}}_{\mathrm{RoS}}(\lambda|\mathcal{P}). \tag{A.4}$$

The minimum on the right-hand side of the preceding inequality can be further bounded as follows for any $\lambda(\overrightarrow{\gamma})\geq 0$.

$$T\cdot\min_{\lambda\geq 0}\overline{\mathcal{D}}_{\mathrm{RoS}}(\lambda|\mathcal{P}) \leq T\cdot\mathbb{E}_{\overrightarrow{\gamma}\sim\mathcal{P}^T}\left[\overline{\mathcal{D}}_{\mathrm{RoS}}(\lambda(\vec{\gamma})|\mathcal{P})\right]. \tag{A.5}$$

In particular, then, we may choose $\lambda(\overrightarrow{\gamma})=\frac{1}{T}\sum_{t=1}^{T}\lambda_t := \overline{\lambda}_T$ on the right-hand side of Inequality (A.5) and combine with Inequality (A.4) to obtain

$$\mathbb{E}_{\overrightarrow{\gamma}\sim\mathcal{P}}\left[\mathrm{Reward}(\mathrm{Opt},\overrightarrow{\gamma})\right] \leq \mathbb{E}_{\overrightarrow{\gamma}\sim\mathcal{P}^T}\left[T\cdot\overline{\mathcal{D}}_{\mathrm{RoS}}(\overline{\lambda}_T|\mathcal{P})\right]. \tag{A.6}$$

Combining this with Proposition A.2 applied to an input sequence of length T finishes the proof. ◻

Proposition 4.1. *Under Assumption 4.1 for the distribution $\mathcal{P}$, let $K(\overrightarrow{\gamma})$ be the number of iterations in the first phase of Algorithm 4.1 for some input sequence $\overrightarrow{\gamma}$. Then, we have*

$$\mathbb{E}_{\overrightarrow{\gamma}\sim\mathcal{P}^T}[K(\overrightarrow{\gamma})] \leq O(\sqrt{T}\log T).$$

Proof. Let $z_t=\max(0, v_t\cdot x_t(b_t)-p_t(b_t))$ be the reward collected at iteration t (in the first phase). Let $z'_t := \frac{\beta}{2}\mathbf{1}_{z_t\geq\beta/2}$. By construction, $z'_t\leq z_t$ for all t. For any given sequence $\overrightarrow{\gamma}$, let $K(\overrightarrow{\gamma})$ and $K'(\overrightarrow{\gamma})$, respectively, be the first time such that $\sum_{t=1}^{K(\overrightarrow{\gamma})} z_t\geq R$ and $\sum_{t=1}^{K'(\overrightarrow{\gamma})} z'_t\geq R$ for some reward R. Then, for every $\overrightarrow{\gamma}$, we have $K(\overrightarrow{\gamma})\leq K'(\overrightarrow{\gamma})$. By the boundedness assumption on z_t, we have

$$\begin{aligned}\mathrm{Prob}(z'_t=\beta/2) &= \mathrm{Prob}(z_t\geq\beta/2)\\ &\geq \mathbb{E}\left[z_t\right]-\mathrm{Prob}(z_t\leq\beta/2)\cdot\mathbb{E}\left[z_t|z_t\leq\beta/2\right]\geq\beta/2.\end{aligned}$$

Then, by Hoeffding bound,

$$\mathrm{Prob}(K(\overrightarrow{\gamma})\geq q)\leq\mathrm{Prob}(K'(\overrightarrow{\gamma})\geq q)\leq e^{-O(q\beta^2)}. \tag{A.7}$$

Picking $q=O(R/\beta^2)$ for $R=2\sqrt{T}\log T$ finishes the claim. ◻

Proposition 4.2. *Let $\overrightarrow{\gamma}_\ell\sim\mathcal{P}^\ell$ and $\overrightarrow{\gamma}_r\sim\mathcal{P}^r$ be sequences of lengths ℓ and r, respectively, with $\ell\leq r$, of i.i.d. requests each from a distribution $\mathcal{P}$. Then the following inequality holds.*

$$\begin{aligned}&\mathbb{E}_{\overrightarrow{\gamma}_\ell\sim\mathcal{P}^\ell}\left[\mathrm{Reward}(\mathit{Algorithm}\ 3.1,\overrightarrow{\gamma}_\ell)\right]\\ &\geq\frac{\ell}{r}\mathbb{E}_{\overrightarrow{\gamma}_r\sim\mathcal{P}^r}\left[\mathrm{Reward}(\mathrm{Opt},\overrightarrow{\gamma}_r)\right]-O(\sqrt{r}).\end{aligned}$$

Proof. By Proposition A.1, Proposition A.2, and Inequality (3.6),

$$\begin{aligned}&\mathrm{Reward}(\mathit{Algorithm}\ 3.1,\overrightarrow{\gamma}_\ell)\\ &\geq\mathbb{E}_{\overrightarrow{\gamma}_\ell\sim\mathcal{P}^\ell}\left[\ell\cdot\overline{\mathcal{D}}_{\mathrm{RoS}}(\overline{\lambda}_\ell|\mathcal{P})\right]-\mathbb{E}_{\overrightarrow{\gamma}_\ell\sim\mathcal{P}^\ell}\left[\sum_{t=1}^{\ell}\lambda_t\cdot g_t(b_t)\right]\\ &\geq\ell\cdot\min_{\lambda\geq 0}\overline{\mathcal{D}}_{\mathrm{RoS}}(\lambda|\mathcal{P})-\mathbb{E}_{\overrightarrow{\gamma}_\ell\sim\mathcal{P}^\ell}\left[\sum_{t=1}^{\ell}\lambda_t\cdot g_t(b_t)\right]\\ &\geq\frac{\ell}{r}\cdot\mathbb{E}_{\overrightarrow{\gamma}_r\sim\mathcal{P}^r}\left[\mathrm{Reward}(\mathrm{Opt},\overrightarrow{\gamma}_r)\right]-O(\sqrt{r}).\end{aligned}$$

◻

B PROOFS: BOTH STRICT CONSTRAINTS

The goal of this section is proving Theorem 5.1 and Theorem 5.2.

Definition B.1. *We need the following definitions of dual variables.*

- *For some $\lambda\geq 0$ and $\mu\geq 0$, let $f_t(b):=v_t\cdot x_t(b)$, define g_t as in Equation (2.3), and define*

$$f^\star_{t,\mathrm{combined}}(\mu,\lambda):=\max_b\left[f_t(b)+\lambda\cdot g_t(b))-\mu\cdot p_t(b)\right]. \tag{B.1}$$

- *The following dual variable parametrized by ρ and $\mathcal{P}$; the quantity $f^\star$ is defined in the same way as in Equation (B.1).*

$$\overline{\mathcal{D}}_{\mathrm{combined}}(\mu,\lambda|\mathcal{P},\rho):=\mu\cdot\rho+\mathbb{E}_{(v,p)\sim\mathcal{P}}\left[f^\star_{\mathrm{combined}}(\mu,\lambda)\right]. \tag{B.2}$$

Proposition B.1. *For some $\rho'\geq 0$, let $\overline{\mathcal{D}}_{\mathrm{combined}}(\mu,\lambda|\mathcal{P},\rho')$ be as defined in Equation (B.2). Then the optimum value $\mathrm{Reward}(\mathrm{Opt},\overrightarrow{\gamma}_\ell,\rho)$ for Problem 2.5 with a total initial budget of $\rho\ell$ over a sequence $\overrightarrow{\gamma}_\ell\sim\mathcal{P}^\ell$ of ℓ requests satisfies the inequality*

$$\begin{aligned}&\mathbb{E}_{\overrightarrow{\gamma}_\ell\sim\mathcal{P}^\ell}\left[\mathrm{Reward}(\mathrm{Opt},\overrightarrow{\gamma}_\ell,\rho)\right]\\ &\leq\ell\cdot\min_{\mu\geq 0,\lambda\geq 0}\left[\overline{\mathcal{D}}_{\mathrm{combined}}(\mu,\lambda|\mathcal{P},\rho')+(\rho-\rho')\cdot\mu\right].\end{aligned}$$

Proof. In this proof, we essentially repeat the ideas in the proof of Proposition A.1. First, by definition, we have

$$\begin{aligned}&\mathrm{Reward}(\mathrm{Opt},\overrightarrow{\gamma}_\ell,\rho)\\ &:=\max_{\{b_t\}:\sum_{t=1}^{\ell}g_t(b_t)\geq 0,\sum_{t=1}^{\ell}p_t(b_t)\leq\rho\ell}\left[\sum_{t=1}^{\ell}v_t\cdot x_t(b_t)\right]\\ &=\max_{\{b_t\}}\min_{\lambda\geq 0,\mu\geq 0}\left[\sum_{t=1}^{\ell}v_t\cdot x_t(b_t)+\lambda\cdot\sum_{t=1}^{\ell}g_t(b_t)+\mu\cdot(\ell\rho-\sum_{t=1}^{\ell}p_t(b_t))\right]\end{aligned}$$

By application of Sion's minimax theorem and the definition of $f^\star_{t,\mathrm{combined}}$, we get

$$\begin{aligned}&\mathrm{Reward}(\mathrm{Opt},\overrightarrow{\gamma}_\ell,\rho) && \text{(B.3)}\\ &\leq\min_{\lambda\geq 0,\mu\geq 0}\sum_{t=1}^{\ell}\mu\cdot\rho+\max_{b_t}\left[v_t\cdot x_t(b_t)+\lambda\cdot g_t(b_t)-\mu\cdot p_t(b_t)\right]\\ &=\min_{\lambda\geq 0,\mu\geq 0}\sum_{t=1}^{\ell}\left[\rho\cdot\mu+f^\star_{t,\mathrm{combined}}(\mu,\lambda)\right] && \text{(B.4)}\\ &\leq\sum_{t=1}^{\ell}\left[\rho\cdot\mu'+f^\star_{t,\mathrm{combined}}(\mu',\lambda')\right], && \text{(B.5)}\end{aligned}$$

for some fixed $\mu'\geq 0$ and $\lambda'\geq 0$. Then, taking the expectations on both sides of Inequality (B.5) and using the fact that v_t, x_t, p_t are all drawn from i.i.d. distributions, we get

$$\begin{aligned}&\mathbb{E}_{\overrightarrow{\gamma}_\ell\sim\mathcal{P}^\ell}[\mathrm{Reward}(\mathrm{Opt},\overrightarrow{\gamma}_\ell,\rho)]\\ &\leq\sum_{t=1}^{\ell}\left[\rho\cdot\mu'+\mathbb{E}_{(v,p)\sim\mathcal{P}}\left[f^\star_{\mathrm{combined}}(\mu',\lambda')\right]\right]\\ &=\ell\cdot\overline{\mathcal{D}}_{\mathrm{combined}}(\mu',\lambda'|\mathcal{P},\rho')+\ell(\rho-\rho')\cdot\mu'.\end{aligned}$$

Therefore, in particular, the preceding inequality holds for the $\lambda,\mu\geq 0$ that minimize the bound, thus finishing the proof. ◻

Next, to analyze the reward collected by Algorithm 5.1, similar to [Balseiro et al. 2020], which gives an algorithm for a strict budget constraint, we need the following notion of stopping time.

Definition B.2. *The stopping time τ of Algorithm 5.1, with a total initial budget of B is the first time τ at which*

$$\sum_{t=1}^{\tau} p_t(b_t) + 1 \geq B.$$

Intuitively, this is the first time step at which the total price paid almost exceeds the total budget.

Our main regret bound in Theorem 5.1 is proved via Proposition B.2 and Proposition B.3. The proof follows along the lines of that in [Balseiro et al. 2020] and Proposition A.2.

Proposition B.2. *Let τ be a stopping time as in Definition B.2 for some initial budget $\rho' k$. Let $\overline{\mu}_\tau = \frac{1}{\tau}\sum_{i=1}^{\tau}\mu_i$ and $\overline{\lambda}_\tau = \frac{1}{\tau}\sum_{i=1}^{\tau}\lambda_i$. Then the expected reward (see Equation (2.6)) of Algorithm 5.1 over a sequence of length k with i.i.d. input requests from distribution $\mathcal{P}^k$ is lower bounded as*

$$\begin{aligned}
&\mathbb{E}_{\overrightarrow{\gamma}_k \sim \mathcal{P}^k}\left[\text{Reward}(\textit{Algorithm } 5.1, \overrightarrow{\gamma}, \rho')\right] \\
&\geq \mathbb{E}_{\overrightarrow{\gamma}_k \sim \mathcal{P}^k}\left[\tau \cdot \overline{\mathcal{D}}_{\text{combined}}(\overline{\mu}_\tau, \overline{\lambda}_\tau | \mathcal{P}, \rho')\right] \\
&\quad - \mathbb{E}_{\overrightarrow{\gamma}_k \sim \mathcal{P}^k}\left[\sum_{t=1}^{\tau}\mu_t \cdot (\rho' - p_t(b_t)) - \sum_{t=1}^{\tau}\lambda_t \cdot g_t(b_t)\right].
\end{aligned}$$

Proof. By Line 4 in Algorithm 5.1, we have, until the stopping time $t = \tau$,

$$f^\star_{t,\text{combined}}(\mu_t, \lambda_t) := v_t \cdot x_t(b_t) + \lambda_t \cdot g_t(b_t) - \mu_t \cdot p_t(b_t).$$

Rearranging the terms and taking expectations conditioned on the randomness up to step $t-1$,

$$\mathbb{E}_{\overrightarrow{\gamma}_k \sim \mathcal{P}^k}\left[v_t \cdot x_t(b_t) | \sigma_{t-1}\right] \tag{B.6}$$

$$= \mathbb{E}_{\overrightarrow{\gamma}_k \sim \mathcal{P}^k}\left[f^\star_{t,\text{combined}}(\mu_t, \lambda_t) + \mu_t \cdot p_t(b_t) - \lambda_t \cdot g_t(b_t) | \sigma_{t-1}\right]. \tag{B.7}$$

Per Line 7 in Algorithm 5.1 that once we fix the randomness up to $t-1$, the dual variables are all fixed, which gives us

$$\mathbb{E}_{\overrightarrow{\gamma}_k \sim \mathcal{P}^k}\left[f^\star_{t,\text{combined}}(\mu_t, \lambda_t) | \sigma_{t-1}\right] = \mathbb{E}_{(v,p)\sim\mathcal{P}}\left[f^\star_{\text{combined}}(\mu_t, \lambda_t)\right] \tag{B.8}$$

Combining Equation (B.8) with the definition of $\overline{\mathcal{D}}_{\text{combined}}$ in Equation (B.2) and plugging back into Equation (B.7) then gives

$$\begin{aligned}
&\mathbb{E}_{\overrightarrow{\gamma}_k \sim \mathcal{P}^T}\left[v_t \cdot x_t(b_t) | \sigma_{t-1}\right] \\
&= \overline{\mathcal{D}}_{\text{combined}}(\mu_t, \lambda_t | \mathcal{P}, \rho') \\
&\quad - \mathbb{E}_{\overrightarrow{\gamma}_k \sim \mathcal{P}^k}\left[\mu_t \cdot (\rho' - p_t(b_t)) + \lambda_t \cdot g_t(b_t) | \sigma_{t-1}\right]
\end{aligned}$$

Summing over $t = 1, 2, \ldots, \tau$, and using the Optional Stopping Theorem, we get

$$\begin{aligned}
&\mathbb{E}_{\overrightarrow{\gamma}_k \sim \mathcal{P}^k}\left[\sum_{t=1}^{\tau} v_t \cdot x_t(b_t)\right] \\
&= \mathbb{E}_{\overrightarrow{\gamma}_k \sim \mathcal{P}^k}\left[\sum_{t=1}^{\tau}\overline{\mathcal{D}}_{\text{combined}}(\mu_t, \lambda_t | \mathcal{P}, \rho')\right] \\
&\quad - \mathbb{E}_{\overrightarrow{\gamma}_k \sim \mathcal{P}^k}\left[\mu_t \cdot (\rho' - p_t(b_t)) - \lambda_t \cdot g_t(b_t)\right].
\end{aligned}$$

We finish the proof by using the convexity of $\overline{\mathcal{D}}_{\text{combined}}$ in the preceding equation. □

Our bound on regret requires the following technical result bounding one of the terms arising in Proposition B.2.

Proposition B.3. *Consider a run of Algorithm 5.1 with initial total budget $\rho\ell$ and the total time horizon ℓ. We define the corresponding stopping time (as defined in Definition B.2) as the time τ at which $\sum_{t=1}^{\tau} p_t(b_t) \geq \rho\ell - 1$. Then, the dual variable $\{\mu_t\}$ that evolves as per Line 8 in Algorithm 5.1 satisfies the following inequality.*

$$\sum_{t=1}^{\tau}\mu_t \cdot (\rho - p_t(b_t)) \leq (\tau - \ell) + 1/\rho + O(\sqrt{\ell}).$$

Proof. To bound $\sum_{t=1}^{\tau}\mu_t \cdot (\rho - p_t(b_t))$, we observe that the mirror descent guarantee of Lemma C.1 applies to give, for any $\mu \geq 0$,

$$\sum_{t=1}^{\tau}\mu_t \cdot (\rho - p_t(b_t)) \leq \sum_{t=1}^{\tau}\mu \cdot (\rho - p_t(b_t)) + \text{Err}(\tau, \eta), \tag{B.9}$$

where $\text{Err}(R, \eta) := \frac{1}{\sigma}(1+\rho^2)\eta R + \frac{1}{\eta}V_h(\mu, \mu_1)$, where $h(u) = \frac{1}{2}u^2$, $\sigma = 1$, and $\mu_1 = 0$. To finish the proof, we choose $\mu = 1/\rho$, use $\sum_{t=1}^{\tau} p_t(b_t) \geq \rho\ell - 1$, and choose $\eta = \frac{1}{(1+\rho^2)\sqrt{\ell}}$ in $\text{Err}(\tau, \eta)$ □

Theorem 5.1. *With i.i.d. inputs from a distribution $\mathcal{P}$ over a time horizon T, the regret of Algorithm 5.1 on Problem 2.5 is bounded by*

$$\text{Regret}(\textit{Algorithm } 5.1, \mathcal{P}^T) \leq O(\sqrt{T}).$$

Further, Algorithm 5.1 incurs a violation of at most $O(\sqrt{T}\log T)$ of the RoS constraint and no violation of the budget constraint.

Proof. Recall that τ is the stopping time of Algorithm 5.1 as defined in Definition B.2. Then,

$$\begin{aligned}
&\mathbb{E}_{\overrightarrow{\gamma} \sim \mathcal{P}^T}\left[\text{Reward}(\text{Opt}, \overrightarrow{\gamma}, \rho)\right] - \frac{\tau}{T} \cdot \mathbb{E}_{\overrightarrow{\gamma} \sim \mathcal{P}^T}\left[\text{Reward}(\text{Opt}, \overrightarrow{\gamma}, \rho)\right] \\
&= \frac{T-\tau}{T} \cdot \mathbb{E}_{\overrightarrow{\gamma} \sim \mathcal{P}^T}\left[\text{Reward}(\text{Opt}, \overrightarrow{\gamma}, \rho)\right] \\
&\leq T - \tau,
\end{aligned} \tag{B.10}$$

where the final step is because $\text{Reward}(\text{Opt}, \overrightarrow{\gamma}, \rho) \leq T$ (due to the value capped at one). Combining Inequality (B.10), Proposition B.1, and the lower bound on $\text{Reward}(\textit{Algorithm } 5.1, \overrightarrow{\gamma}, \rho)$ from Proposition B.2 in the definition of Regret in Equation (2.8) gives

$$\text{Regret}(\textit{Algorithm } 5.1, \overrightarrow{\gamma}) \tag{B.11}$$

$$\leq \mathbb{E}_{\overrightarrow{\gamma} \sim \mathcal{P}^T}\left[(T-\tau) + \sum_{t=1}^{\tau}\mu_t \cdot (\rho - p_t(b_t)) + \sum_{t=1}^{\tau}\lambda_t \cdot g_t(b_t)\right]. \tag{B.12}$$

We invoke Proposition B.3 with a total initial budget ρT and total time horizon T to get

$$\mathbb{E}_{\overrightarrow{\gamma} \sim \mathcal{P}^T}\left[\sum_{t=1}^{\tau}\mu_t(\rho - p_t(b_t))\right] \leq \mathbb{E}_{\overrightarrow{\gamma} \sim \mathcal{P}^T}\left[\tau - T\right] + \frac{1}{\rho} + O(\sqrt{T}). \tag{B.13}$$

Finally, we invoke Lemma 3.2 to conclude $\sum_{t=1}^{R}\lambda_t g_t \leq O(\sqrt{R})$, which lets us conclude

$$\mathbb{E}_{\overrightarrow{\gamma} \sim \mathcal{P}^T}\left[\sum_{t=1}^{\tau}\lambda_t g_t\right] \leq O(\sqrt{T}). \tag{B.14}$$

Combining Inequality (B.12), Inequality (B.13), and Inequality (B.14) yields the claimed bound.

We now finish with the proof of maximum constraint violation. By design of Algorithm 5.1, the budget constraint is never violated throughout the run of the algorithm. To see the claimed maximum violation of the RoS constraint, we note by Proposition 3.2 that the gradient g_t satisfies $g_t(b_t) \geq -1/\lambda_t$, as a result of which, Lemma 3.1 applies. □

Theorem 5.2. *With i.i.d. inputs from a distribution $\mathcal{P}$ over a time horizon T, the regret of Algorithm 5.2 on Problem 2.5 is bounded by*

$$\text{Regret}(\textit{Algorithm } 5.2, \mathcal{P}^T) \leq O(\sqrt{T}\log T).$$

Further, Algorithm 5.2 suffers no constraint violation of either the RoS or budget constraint.

Proof. The RoS constraint is not violated because the first phase accumulates the buffer that is the guaranteed cap on violation in the second phase. The budget constraint is respected by design: the first phase pays at most ρT, followed by the second phase, which, as guaranteed by Algorithm 5.1, strictly respects the budget. To bound the regret, we note that the total expected reward is at least as much as is collected in the second phase

$$\begin{aligned}&\mathbb{E}_{\overrightarrow{\gamma}\sim\mathcal{P}^T}\left[\text{Reward}(\textit{Algorithm } 5.2, \overrightarrow{\gamma}, \rho)\right]\\ &\geq \mathbb{E}_k\left[\mathbb{E}_{\overrightarrow{\gamma}_{k+1:T}\sim\mathcal{P}^{T-k}}\left(\text{Reward}(\textit{Algorithm } 5.1, \overrightarrow{\gamma}_{k+1:T}, \widehat{\rho})\right)\right],\end{aligned} \tag{B.15}$$

where $\widehat{\rho} = \frac{\rho T - K(\overrightarrow{\gamma})}{T - K(\overrightarrow{\gamma})}$ and the right-hand side captures the reduced time horizon $T - K(\overrightarrow{\gamma})$ and reduced initial budget $\rho T - K(\overrightarrow{\gamma})$ for Algorithm 5.1. Conditioning on the high-probability event that $k \leq \rho T$ (by Inequality (A.7) coupled with the assumption that ρ is a fixed constant) and letting $R = \text{Reward}(\textit{Algorithm } 5.1, \overrightarrow{\gamma}_{k+1:T}, \widehat{\rho})$:

$$\begin{aligned}&\mathbb{E}_k\left[\mathbb{E}_{\overrightarrow{\gamma}_{k+1:T}\sim\mathcal{P}^{T-k}}(R)\right]\\ &\geq (1 - e^{-O(T)})\mathbb{E}_k\left[\mathbb{E}_{\overrightarrow{\gamma}_{k+1:T}\sim\mathcal{P}^{T-k}}(R) \mid k \leq \rho T\right].\end{aligned} \tag{B.16}$$

Applying Proposition B.2 with the reduced budget and time horizon:

$$\begin{aligned}\mathbb{E}_{\overrightarrow{\gamma}_{k+1:T}\sim\mathcal{P}^{T-k}}[R] \geq\ &\mathbb{E}_{\overrightarrow{\gamma}_{k+1:T}\sim\mathcal{P}^{T-k}}\left[\tau\cdot\overline{\mathcal{D}}_{\text{combined}}(\overline{\mu}_\tau, \overline{\lambda}_\tau \mid \mathcal{P}, \widehat{\rho})\right]\\ &- \mathbb{E}_{\overrightarrow{\gamma}_{k+1:T}\sim\mathcal{P}^{T-k}}\left[\sum_{t=1}^{\tau}\mu_t\cdot(\widehat{\rho} - p_t(b_t))\right]\\ &- \mathbb{E}_{\overrightarrow{\gamma}_{k+1:T}\sim\mathcal{P}^{T-k}}\left[\sum_{t=1}^{\tau}\lambda_t\cdot g_t(b_t)\right].\end{aligned} \tag{B.17}$$

Next, by Proposition B.1, we have:

$$\overline{\mathcal{D}}_{\text{combined}}(\overline{\mu}_\tau, \overline{\lambda}_\tau \mid \mathcal{P}, \widehat{\rho}) + (\rho - \widehat{\rho})\cdot\overline{\mu}_\tau \geq \frac{1}{T}\mathbb{E}_{\overrightarrow{\gamma}\sim\mathcal{P}^T}\left[\text{Reward}(\text{Opt}, \overrightarrow{\gamma}, \rho)\right]. \tag{B.18}$$

We can now repeat the trick in the proof of Theorem 5.1:

$$\begin{aligned}&\mathbb{E}_{\overrightarrow{\gamma}\sim\mathcal{P}^T}\left[\text{Reward}(\text{Opt}, \overrightarrow{\gamma}, \rho)\right]\\ &\leq \frac{\tau}{T}\cdot\mathbb{E}_{\overrightarrow{\gamma}\sim\mathcal{P}^T}\left[\text{Reward}(\text{Opt}, \overrightarrow{\gamma}, \rho)\right] + (T - \tau),\end{aligned} \tag{B.19}$$

Combining Inequality (B.15), Inequality (B.16), Inequality (B.17), Inequality (B.18), and Inequality (B.19) then gives

$$\begin{aligned}&\mathbb{E}_{\overrightarrow{\gamma}\sim\mathcal{P}^T}\left[\text{Regret}(\textit{Algorithm } 5.2, \overrightarrow{\gamma})\right]\\ &\leq \mathbb{E}_{\overrightarrow{\gamma}\sim\mathcal{P}^T}(T - \tau) + \frac{T}{e^{O(T)}}\\ &\quad + \mathbb{E}_k\left[\mathbb{E}_{\overrightarrow{\gamma}_{k+1:T}\sim\mathcal{P}^{T-k}}\left[\sum_{t=1}^{\tau}\mu_t\cdot(\widehat{\rho} - p_t(b_t))\right] \mid k \leq \rho T\right]\\ &\quad + \mathbb{E}_k\left[\mathbb{E}_{\overrightarrow{\gamma}_{k+1:T}\sim\mathcal{P}^{T-k}}\left[\sum_{t=1}^{\tau}\lambda_t\cdot g_t(b_t)\right] \mid k \leq \rho T\right]\\ &\quad + \mathbb{E}_k\left[\mathbb{E}_{\overrightarrow{\gamma}_{k+1:T}\sim\mathcal{P}^{T-k}}\left[\tau\overline{\mu}_\tau(\rho - \widehat{\rho})\right] \mid k \leq \rho T\right].\end{aligned} \tag{B.20}$$

By applying Proposition B.3 and Proposition 4.1, we have

$$\begin{aligned}&\mathbb{E}_k\left[\mathbb{E}_{\overrightarrow{\gamma}_{k+1:T}\sim\mathcal{P}^{T-k}}\left[\sum_{t=1}^{\tau}\mu_t\cdot(\widehat{\rho} - p_t(b_t))\right] \mid k \leq \rho T\right]\\ &\leq \mathbb{E}_{\overrightarrow{\gamma}\sim\mathcal{P}^T}\left[(\tau - T) + K(\overrightarrow{\gamma})\right] + O(\sqrt{T}) + \mathbb{E}_k\left[\mathbb{E}_{\overrightarrow{\gamma}_{k+1:T}}(1/\widehat{\rho}) \mid k \leq \rho T\right]\\ &\leq \mathbb{E}_{\overrightarrow{\gamma}\sim\mathcal{P}^T}(\tau - T) + O(\sqrt{T}\log T).\end{aligned} \tag{B.21}$$

We invoke Lemma 3.2 to conclude $\sum_{t=1}^{R}\lambda_t g_t \leq O(\sqrt{R})$ for all $R \leq T$, which lets us conclude

$$\mathbb{E}_{\overrightarrow{\gamma}\sim\mathcal{P}^T}\left[\sum_{t=1}^{\tau}\lambda_t g_t\right] \leq O(\sqrt{T}). \tag{B.22}$$

To bound the final term in Inequality (B.20), we observe that $\rho - \widehat{\rho} = \frac{(1-\rho)K(\overrightarrow{\gamma})}{T - K(\overrightarrow{\gamma})}$ by definition of $\widehat{\rho}$. Combining this with $\tau \leq T$, the bound on $\sum_{i=1}^{\tau}\mu_i$ from Algorithm 5.1, the result of Proposition 4.1, and the conditional expectation, we get

$$\mathbb{E}_k\left[\mathbb{E}_{\overrightarrow{\gamma}_{k+1:T}\sim\mathcal{P}^{T-k}}\left[\tau\overline{\mu}_\tau(\rho - \widehat{\rho})\right] \mid k \leq \rho T\right] \leq O(\sqrt{T}). \tag{B.23}$$

Combining Inequality (B.20), Inequality (B.21), Inequality (B.22), and Inequality (B.23) finishes the proof. □

C ONLINE MIRROR DESCENT

Lemma C.1 ([Bubeck et al. 2015], Theorem 4.2). *Let h be a mirror map which is ρ-strongly convex on $\mathcal{X}\cap\mathcal{D}$ with respect to a norm $\|\cdot\|$. Let f be convex and L-Lipschitz with respect to $\|\cdot\|$. Then, mirror descent with step size α satisfies*

$$\sum_{s=1}^{t}(f(x_s) - f(x)) \leq \frac{1}{\alpha}V_h(x, x_1) + \alpha\frac{L^2 t}{2\rho}.$$