Appendix A. Proof to Theorem 1

In this section, follow the proof idea of (Badanidiyuru, Kleinberg, and Slivkins 2013), we will show the regret analysis of **Algorithm 1** in the deterministic case.

Theorem 1. *Consider an instance of PDG-BwK with J upper bound constraints and K lower bound constraints. The number of arms is N. Assume the value of constraints bound is U and L. The regret of algorithm Primal Dual G-BwK with parameter*

$$\begin{aligned} \epsilon_U &= \sqrt{\ln(J)/\|U\|_1} \\ \epsilon_L &= \sqrt{\ln(K)/\|L\|_1}. \end{aligned} \tag{17}$$

satisfies

$$\begin{aligned} regret(\tau) \leq & \mathcal{O}(\sqrt{\|U\|_1 \ln J} - \sqrt{\|L\|_1 \ln K} \\ & + N) \cdot \frac{OPT_{LP}}{\|U\|_1 - \|L\|_1} \end{aligned} \tag{18}$$

Proof. According to (Arora, Hazan, and Kale 2012), the payoff vector in any round t of the multiplicative-weights update method is given by C_t and D_t respectively. The goal is to optimize the total expected payoff of τ rounds

$$W = \sum_{t=N+1}^{\tau} \sum_{i \in \mathcal{N}} C_t(i) \cdot \eta^t - \sum_{t=N+1}^{\tau} \sum_{i \in \mathcal{N}} D_t(i) \cdot \sigma^t.$$

W is total cost consumed by **Algorithm 1**. Next, we will prove W can be related to the total reward REW by the following way

$$REW \geq W \cdot OPT_{LP}/(\|U\|_1 - \|L\|_1) \tag{19}$$

For each round t, let $z_t = e^N_{x_t}$ denote the x_t-th coordinate vector. We use $C_t \in [0,1]^{N \times J}$ and $D_t \in [0,1]^{N \times K}$ as the matrix consists of $C_t(x)$ and $D_t(x)$, as well as $r_t \in [0,1]^N$ as the vector consists of $r_t(x)$. That is

$$\begin{aligned} C_t &= (C_t(1), C_t(2), \cdots, C_t(J))^\top \\ D_t &= (D_t(1), D_t(2), \cdots, D_t(K))^\top \end{aligned}$$

and

$$r_t = (r_t(1), r_t(2), \cdots, r_t(N))^\top$$

Thus we have the decision of t-th round z_t satisfies

$$z_t \in \arg\max_z \{ \frac{r_t^\top z}{z^\top C_t \cdot \eta^t - z^\top D_t \cdot \sigma^t} \}. \tag{20}$$

That is to say, z_t was chosen by **Algorithm 1** to maximize the ratio (15) among all distributions over arms. Let ρ be the corresponding max value of ratio, thus we have the inequality holds

$$\rho(z^\top C_t \eta^t - z^\top D_t \sigma^t) \geq r_t^\top z. \tag{21}$$

By the definition of OPT_{LP}, it follows that

$$z_t^\top C_t \eta^t - z_t^\top D_t \sigma^t \leq r_t^\top ((v^*)^\top C_t \eta^t - (v^*)^\top D_t \sigma^t)/OPT_{LP}. \tag{22}$$

Thus,

$$\begin{aligned} W &= \sum_t C_t \eta^t \cdot \mathbf{1} - \sum_t D_t \sigma^t \cdot \mathbf{1} \\ &\leq \frac{1}{OPT_{LP}} \sum_t r_t (C \eta^t v_t^* - D \sigma^t v_t^*) \\ &= \frac{1}{OPT_{LP}} \left((\sum_t r_t \eta_t^\top) C v^* - (\sum_t r_t \sigma_t^\top) D v^* \right) \end{aligned} \tag{23}$$

Let

$$\bar{\eta} = \frac{1}{REW} \sum_t r_t \eta^t \in [0,1]^J$$

and

$$\bar{\sigma} = \frac{1}{REW} \sum_t r_t \sigma^t \in [0,1]^K,$$

it has

$$\begin{aligned} W &\leq \frac{REW}{OPT_{LP}} \left(\bar{\eta}^\top C v^* - \bar{\sigma}^\top D v^* \right) \\ &\leq \frac{REW}{OPT_{LP}} (\sum_j U_J - \sum_k L_k) \\ &= \frac{REW}{OPT_{LP}} (\|U\|_1 - \|L\|_1). \end{aligned} \tag{24}$$

Thus we proved (19).

By using **Proposition 5.4** in (Badanidiyuru, Kleinberg, and Slivkins 2013) and follow the same idea, we obtain

$$\begin{aligned} & REW \geq W \cdot OPT_{LP}/(\|U\|_1 - \|L\|_1) \\ & \geq \left[(1-\epsilon_U) \sum_t \eta^\top C_t - \frac{\ln J}{\epsilon_U} - (1-\epsilon_L) \sum_t \sigma_t^\top D_t \right. \\ & \left. + \frac{\ln K}{\epsilon_L} \right] \cdot \frac{OPT_{LP}}{\|U\|_1 - \|L\|_1} \\ & \geq \left[(1-\epsilon_U)(\|U\|_1 - m - 1) - (1-\epsilon_L)\|L\|_1 - \frac{\ln J}{\epsilon_U} \right. \\ & \left. + \frac{\ln K}{\epsilon_L} \right] \cdot \frac{OPT_{LP}}{\|U\|_1 - \|L\|_1} \\ & \geq OPT_{LP} - \left[\epsilon_U \|U\|_1 + \epsilon_L \|L\|_1 + \frac{\ln J}{\epsilon_U} \right. \\ & \left. - \frac{\ln K}{\epsilon_L} + m + 1 \right] \cdot \frac{OPT_{LP}}{\|U\|_1 - \|L\|_1} \\ & = OPT_{LP} - \mathcal{O}(\sqrt{\|U\|_1 \ln J} - \sqrt{\|L\|_1 \ln K} + m) \\ & \cdot \frac{OPT_{LP}}{\|U\|_1 - \|L\|_1}. \end{aligned}$$

Rearrange this inequality, we have

$$\begin{aligned} & OPT_{LP} - REW \leq \\ & \mathcal{O}(\sqrt{\|U\|_1 \ln J} - \sqrt{\|L\|_1 \ln K} + m) \cdot \frac{OPT_{LP}}{\|U\|_1 - \|L\|_1}. \end{aligned} \tag{25}$$

This completes the regret analysis for deterministic case. Note that the regret bound of **Algorithm 1** in the non-deterministic case can be obtained by combining the similar technique from above and (Badanidiyuru, Kleinberg, and Slivkins 2013).

□