Appendix

A. Proof of Proposition 2.1

Proof. The proof is mainly about adapting the specific two-player game presented in (Mannor et al., 2009) to the general online convex programming setting with adversarial constraints. We closely follow the notations in the example from Proposition 4 in (Mannor et al., 2009).

Let us define the decision set $X = \Delta([1, 2])$, namely a 2-D simplex. We design two different loss functions: $\ell^1(x) = [-1, 0]x$, and $\ell^2(x) = [-1, -1]x$ (here $[a, b]$ stands for a 2-d row vector and hence $[a, b]x$ stands for the regular vector inner product). We also design two different constraints as: $f^1(x) = [-1, -1]x \leq 0$ and $f^2(x) = [1, -1]x \leq 0$. Note that both $\ell$ and $f$ are linear functions with respect $x$, hence they are convex loss functions and constraints with respect to $x$. The adversary picks loss functions among $\{\ell^1, \ell^2\}$ and constraints among $\{f^1, f^2\}$ and will generate the following sequence of loss functions and constraints. Initialize a counter $k = 0$, then:

1. while $k = 0$ or $\frac{1}{t-1} \sum_{i=1}^{t-1} x_i[1] > 3/4$, the adversary set $\ell_t = \ell^2(x)$ and $f_t = f^2(x)$, and set $k := k + 1$.
2. For next $k$ steps, the adversary set $\ell_t = \ell^1(x)$ and $f_t = f^1(x)$. Then reset $k = 0$ and go back to step 1.

For any time step $t$, let us define $\hat{q}_t = \frac{1}{t} \sum_{i=1}^{t} \mathbb{1}(f_i = f^2)$, namely the fraction of the adversary picking the second type of constraint. Let us define $\tilde{\alpha}_t = \sum_{i=1}^{t} x_i[1]/t$. Given any $\hat{q}_{ti}$, we see that $\mathcal{O}'$ can be defined as

$$\mathcal{O}' = \{x \in \Delta([1, 2]) : \hat{q}_t[1], -1]x + (1 - \hat{q}_t)[-1, 1]x \leq 0\}$$

and the minimum loss the learner can get in hindsight with decisions restricted to $\mathcal{O}'$ is:

$$r_{t_{min}} = \min_{x \in \mathcal{O}'} (1 - \hat{q}_t)[-1, 0]x + \hat{q}_t[-1, 1]x$$

The cumulative constraint violation at time step $t$ can be computed as $\sum_{i=1}^{t} f_i(x_i) = \sum_{i=1}^{t} \mathbb{1}(f_i = f^1)[-1, 1]x_i + \mathbb{1}(f_i = f^2)[-1, 1]x_i$. We want to show that no matter what strategy the learner uses, as long as $\frac{1}{t} \limsup_{t \to \infty} \sum_{i=1}^{t} f_i(x_i) \leq 0$, we will have $\lim_{t \to \infty} \sum_{i=1}^{t} f_i(x_i) - r_{t_{min}} > 0$.

Following a similar argument from (Mannor et al., 2009), we can show that Step 2 is entered an infinite number of times. To show this, assume that step 2 only enters finite number of times. Hence as the game keeps staying in Step 1, the fraction of the adversary picking the second constraint $f^2$ approaches to one ($\hat{q}_t \to 1$), we will have as $t$ approaches to infinity,

$$\lim_{t \to \infty} \frac{1}{t} \sum_{i=1}^{t} f_i(x_i) = \lim_{t \to \infty} \frac{1}{t} \sum_{i=1}^{t} \mathbb{1}(f_i = f^1)[-1, 1]x_i + \frac{1}{t} \sum_{i=1}^{t} \mathbb{1}(f_i = f^2)[-1, 1]x_i$$

$$= \lim_{t \to \infty} \frac{1}{t} \sum_{i=1}^{t} \mathbb{1}(f_i = f^2)[-1, 1]x_i = \lim_{t \to \infty} \frac{1}{t} \sum_{i=1}^{t} [-1, 1]x_i = \lim_{t \to \infty} [-1, 1]x_i = [1, 1]x_i.$$ (19)

Since $\sum_{i=1}^{t} x_i/t \in \Delta([1, 2])$, we must have $\hat{\alpha}_t = \sum_{i=1}^{t} x_i[1] / t \leq 1/2$ to ensure that the long-term constraint is satisfied: $\lim_{t \to \infty} \frac{1}{t} \sum_{i=1}^{t} f_i(x_i) \leq 0$. But when $\hat{\alpha}_t \leq 1/2$, the condition of entering Step 1 is violated and we must enter step 2. Hence step 2 is entered infinite number of times. In particular, there exist infinite sequences $t_i$ and $\ell_{t_i}$ such that $t_i < t_{i+1}$, and the adversary picks $f^2, \ell^2$ in $(t_i, t_{i+1})$ (Step 1) and the adversary picks $f^1, \ell^1$ in $(t_i, t_{i+1})$ (Step 2). Since step 1 and step 2 executes the same number of steps (i.e., using the counter $k$’s value), we must have $\hat{q}_{ti} = 1/2$ and $t_{i_{min}} = 1$. Furthermore, we must have $t_{i+1} \geq t_{i+1}/2$. Note that $\hat{\alpha}_{ti} \leq 3/4$ since otherwise the adversary would be in step 1 at time $t_{i+1}$. Thus, during the first $t_{i+1}$ steps, we must have:

$$\sum_{j=1}^{t_{i+1}} x_j[1] \leq \sum_{j=1}^{t_{i+1}} x_j[1] + \sum_{j=t_{i+1}+1}^{t_{i+1}} x_j[1] \leq \frac{1}{4} (t_{i+1} - t_i) = t_{i+1} - t_i/4 \leq \frac{7}{8} t_{i+1}. \quad (20)$$

The proof is mainly about adapting the specific two-player game presented in (Mannor et al., 2009) to the general online convex programming setting with adversarial constraints. We closely follow the notations in the example from Proposition 4 in (Mannor et al., 2009).
It is easy to verify that 
\[
\frac{1}{t+1} \sum_{t=1}^{t+1} \ell_t(x_t) \geq -\frac{1}{t+1} \sum_{t=1}^{t+1} x_t[1] \geq -\frac{7}{8}.
\]
Hence, simply let \( t \to \infty \), we have:

\[
\limsup_{t \to \infty} \left( \frac{1}{t} \sum_{i=1}^{t} \ell_i(x_i) - r_{i, \text{min}}^{\text{min}} \right) \geq -7/8 + 1 = 1/8.
\]  

(21)

Namely, we have shown that for cumulative regret, regardless what sequence of decisions \( x_1, \ldots, x_t \) the learner has played, as long as it needs to satisfy \( \limsup_{t \to \infty} \frac{1}{t} \sum_{i=1}^{t} f_i(x_i) \leq 0 \), we must have:

\[
\limsup_{t \to \infty} \left( \frac{1}{t} \sum_{i=1}^{t} \ell_i(x_i) - \min_{x \in \mathcal{O}'} \sum_{i=1}^{t} \ell_i(x) \right) \geq t/8 = \Omega(t).
\]  

(22)

Hence we cannot guarantee to achieve no-regret when competing against the decisions in \( \mathcal{O}' \) while satisfying the long-term constraint.

\[\Box\]

### B. Analysis of Alg. 1 and Proof Of Theorem 3.1

**Proof of Theorem 3.1.** Since the algorithm runs online mirror descent on the sequence of loss \( \{\mathcal{L}_t(x, \lambda_t)\} \) with respect to \( x \), using the existing results of online mirror descent (Theorem 4.2 and Eq. 4.10 from (Bubeck, 2015)), we know that for the sequence of \( \{x_t\} \):

\[
\sum_{t=1}^{T} (\mathcal{L}_t(x_t, \lambda_t) - \mathcal{L}_t(x, \lambda_t)) \leq \frac{D_R(x, x_1)}{\mu} + \frac{\mu}{2\alpha} \sum_{t=1}^{T} \|\nabla x \mathcal{L}_t(x, \lambda_t)\|^2.
\]  

(23)

Also, we know that the algorithm runs online gradient ascent on the sequence of loss \( \{\mathcal{L}_t(x_t, \lambda_t)\} \) with respect to \( \lambda \), using the existing analysis of online gradient descent (Zinkevich, 2003), we have for the sequence of \( \lambda_t \):

\[
\sum_{t=1}^{T} \mathcal{L}_t(x_t, \lambda_t) - \sum_{t=1}^{T} \sum_{t=1}^{T} \mathcal{L}_t(x_t, \lambda_t) \leq \frac{1}{\mu} \lambda^2 + \frac{\mu}{2} \sum_{t=1}^{T} \left( \frac{\partial \mathcal{L}_t(\omega_t, \lambda_t)}{\partial \lambda_t} \right)^2,
\]  

(24)

Note that for \( (\nabla_x \mathcal{L}_t(x_t, \lambda_t))^2 = (f_t(x_t) - \delta \mu \lambda_t)^2 \leq 2f_t^2(x_t) + 2\delta^2 \mu^2 \lambda_t^2 \leq 2D^2 + \delta^2 \mu^2 \lambda_t^2 \). Similarly for \( \|\nabla_x \mathcal{L}_t(x_t, \lambda_t)\|^2 \), we also have:

\[
\|\nabla_x \mathcal{L}_t(x_t, \lambda_t)\|^2 \leq 2\|\nabla \ell_t(x_t)\|^2 + 2\|\lambda_t \nabla f_t(x_t)\|^2 \leq 2G^2(1 + \lambda_t^2),
\]  

(25)

where we first used triangle inequality for \( \|\nabla_x \mathcal{L}_t(x_t, \lambda_t)\|_\ast \) and then use the inequality of \( 2ab \leq a^2 + b^2 \), \( \forall a, b \in \mathcal{R}^+ \). We also assume that the norm of the gradients are bounded as \( \max(\|\nabla \ell_t(x_t)\|_\ast, \|\nabla f_t(x_t)\|_\ast) \leq G \in \mathcal{R}^+ \). Now sum Inequality 23 and 24 from \( t = 1 \) to \( T \), we get:

\[
\sum_{t=1}^{T} \mathcal{L}_t(x_t, \lambda_t) - \mathcal{L}_t(x, \lambda_t) \leq \frac{2D_R(x, x_0) + \lambda^2}{2\mu} + \sum_{t=1}^{T} \mu(D^2 + \delta^2 \mu^2 \lambda_t^2) + \sum_{t=1}^{T} \frac{\mu G^2}{\alpha}(1 + \lambda_t^2)
\]

\[
= \frac{2D_R(x, x_0) + \lambda^2}{2\mu} + T\mu(D^2 + \frac{G^2}{\alpha}) + \mu(\delta^2 \mu^2 + \frac{G^2}{\alpha}) \sum_{t=1}^{T} \lambda_t^2.
\]  

(26)

Using the saddle-point convex and concave formation for \( \mathcal{L}_t \), we have:

\[
\sum_{t=1}^{T} \mathcal{L}_t(x_t, \lambda) - \mathcal{L}_t(x_t, \lambda_t) = \sum_{t=1}^{T} (\ell_t(x_t) - \ell_t(x_t)) + \sum_{t=1}^{T} (\lambda f_t(x_t) - \lambda_t f_t(x)) + \frac{\delta \mu}{2} \sum_{t=1}^{T} \lambda_t^2 - \frac{\delta \mu T}{2} \lambda^2
\]

\[
\leq \frac{2B + \lambda^2}{2\mu} + T\mu(D^2 + \frac{G^2}{\alpha}) + \mu(\delta^2 \mu^2 + \frac{G^2}{\alpha}) \sum_{t=1}^{T} \lambda_t^2.
\]  

(27)
Note that based on the setting of $\delta$ and $\mu$, we can show that $\delta \geq \delta^2 \mu^2 + G^2/\alpha$. This is because $\delta^2 \mu^2 + G^2/\alpha = \frac{4G^2B}{\sigma^2T(D^2 + G^2/\alpha)} + G^2/\alpha \leq \frac{4G^2B}{\sigma^2T} + G^2/\alpha \leq 2G^2/\alpha$, where we assume that $T$ is large enough such that $T \geq 4B$. Since we have $\delta \geq \delta^2 \mu^2 + G^2/\alpha$, we can remove the term $\sum_t \lambda_t^2$ in the above inequality.

Now set $x = x^*$, and set $\lambda = 0$, since $f_t(x^*) \leq 0$ for all $t$, we get:

$$\sum_t (\ell_t(x_t) - \ell_t(x^*)) \leq \frac{2B}{2\mu} + T\mu(D^2 + G^2/\alpha) \leq 2\sqrt{BT(D^2 + G^2/\alpha)},$$

(29) where we set $\mu = \sqrt{B/(T(D^2 + G^2/\alpha))}$.

To upper bound $\sum_t f_t(x_t)$, we first note that we can lower bound $\sum_{t=1}^T (\ell_t(x_t) - \ell_t(x))$ as $\sum_{t=1}^T (\ell_t(x_t) - \ell_t(x)) \geq -2FT$. Now let us assume that $\sum_t f_t(x_t) > 0$ (otherwise we are done). We set $\lambda = (\sum_t f_t(x_t)) / (\delta \mu T + 1/\mu)$, we have:

$$\left(\sum_t f_t(x_t)\right)^2 \leq \frac{2B}{2\mu} T\mu(D^2 + G^2/\alpha) + \sum_t (\ell_t(x^*) - \ell_t(x_t)) \leq 2\sqrt{BT(D^2 + G^2/\alpha)} + 2FT$$

(30) Substitute $\mu = \sqrt{B/(T(D^2 + G^2/\alpha))}$ into the above inequality, we have:

$$\left(\sum_{t=1}^T f_t(x_t)\right)^2 \leq 2\sqrt{BT(D^2 + G^2/\alpha)}(2\delta \mu T + 1/\mu) + 2FT(2\delta \mu T + 1/\mu) \leq \frac{8G^2}{\alpha} BT + 2T(D^2 + \frac{D^2}{\alpha}) + 2T(D^2 + \frac{G^2}{\alpha}) + T^{3/2}\sqrt{8F^2G^2/\alpha}.$$  

(31) Take the square root on both sides of the above inequality and observe that $T^{3/2}\sqrt{8F^2G^2/\alpha}$ dominates the RHS of the above inequality, we prove the theorem.

\section{C. Analysis of EXP4.R}

In this section we provide the full proof of theorem 4.2.

\textbf{Proof of Theorem 4.2.} We first present several known facts. First we have that for $w_t^T \hat{s}_t$:

$$w_t^T \hat{s}_t = \mathbb{E}_{i \sim w_t} \hat{s}_t[i] = \mathbb{E}_{i \sim w_t} \pi_i(s_t)T \hat{r}_t = \mathbb{E}_{i \sim w_t} \mathbb{E}_{j \sim \pi_i(s_t)} \hat{r}_t[j] = \mathbb{E}_{j \sim p_t} \hat{r}_t[j] = r_t[a_t] \leq 1.$$  

(32) For $w_t^T \hat{y}_t$, we have:

$$w_t^T \hat{y}_t = \mathbb{E}_{i \sim w_t} \hat{y}_t[i] = \mathbb{E}_{i \sim w_t} \pi_i(s_t)T \hat{c}_t = \mathbb{E}_{j \sim p_t} \hat{c}_t[j] = c_t[a_t] \leq 1.$$  

(33) For $\mathbb{E}_{a_t \sim p_t}(w_t^T \hat{s}_t - \beta)^2$, we then have:

$$\mathbb{E}_{a_t \sim p_t}(w_t^T \hat{s}_t - \beta)^2 = \mathbb{E}_{a_t \sim p_t}(r_t[a_t] - \beta)^2 \leq \mathbb{E}_{a_t} 2r_t[a_t]^2 + 2\beta^2 \leq 4.$$  

(34) For $\mathbb{E}_{a_t \sim p_t} \hat{y}_t[a_t]$, we have:

$$\mathbb{E}_{a_t \sim p_t} \hat{y}_t[j] = \pi_j(s_t) T \mathbb{E}_{a_t \sim p_t} \hat{c}_t = \pi_j(s_t) T c_t = y_t[j],$$  

(35) Note that here for analysis simplicity we consider asymptotic property of the algorithm and assume $T$ is large enough and particularly larger than any constant. We don’t necessarily have to assume $T \geq 4B$ here because we can explicitly solve the inequality $\delta \geq \delta^2 \mu^2 + G^2/\alpha$ to find the valid range of $\delta$, as (Mahdavi et al., 2012) did.
which gives us \( \mathbb{E}_{a_t \sim p_t} \hat{y}_t = y_t \). Similarly we can easily verify that \( \mathbb{E}_{a_t \sim p_t} \hat{e}_t = z_t \).

For \( \sum_{i=1}^{||I||} w_t[i] \hat{y}_t[i]^2 \), we have:

\[
\sum_{i=1}^{||I||} w_t[i] \hat{y}_t[i]^2 = \mathbb{E}_{i \sim w_t} \hat{y}_t[i]^2 = \mathbb{E}_{i \sim w_t} (\pi_j(s_i)^T \hat{c}_i)^2 = \mathbb{E}_{i \sim w_t} (\mathbb{E}_{j \sim \pi_i(s_i)} \hat{c}_i[j])^2 \\
\leq \mathbb{E}_{i \sim w_t} \mathbb{E}_{j \sim \pi_i(s_i)} (\hat{c}_i[j])^2 = \mathbb{E}_{j \sim p_t} (\hat{c}_i[j])^2 = \frac{c_t[a_i]^2}{p_t[a_i]}.
\]

Hence, for \( \sum_{i=1}^{||I||} w_t[i] \hat{y}_t[i]^2 \) we have:

\[
\mathbb{E}_{a_t \sim p_t} \sum_{i=1}^{||I||} w_t[i] \hat{y}_t[i]^2 \leq \mathbb{E}_{a_t \sim p_t} \frac{c_t[a_i]^2}{p_t[a_i]} = \sum_{k=0}^{K} c_t[k]^2 \leq K.
\]  

(36)

Similarly, for \( \sum_{i=1}^{||I||} w_t[i] \hat{z}_t[i]^2 \), we have:

\[
\sum_{i=1}^{||I||} w_t[i] \hat{z}_t[i]^2 = \mathbb{E}_{i \sim w_t} (\pi_i(s_i)^T \hat{r}_i)^2 \leq \mathbb{E}_{j \sim p_t} (\hat{r}_i[j])^2 = \frac{r_t[a_i]^2}{p_t[a_i]},
\]

and

\[
\mathbb{E}_{a_t \sim p_t} \sum_{i=1}^{||I||} w_t[i] \hat{z}_t[i]^2 \leq K.
\]  

(39)

Now we are going to take expectation with respect to the randomized decisions \( \{a_i\} \) on both sides of Inequality 11. Fix time step \( t \), conditioned on \( a_1, ..., a_{t-1} \), we have:

\[
\mathbb{E}_{a_t} \left[ \mathcal{L}_t(w_t, \lambda) - \mathcal{L}_t(w, \lambda_t) \right] \\
= \mathbb{E}_{a_t} \left[ c_t[a_t] + \lambda r_t[a_t] - \beta \right] - \frac{\delta \mu}{2} \lambda^2 - \hat{y}_t^T w - \lambda_t (\hat{z}_t^T w - \beta) + \frac{\delta \mu}{2} \lambda_t^2 \\
= \mathbb{E}_{a_t} c_t[a_t] + \lambda (\mathbb{E}_{a_t} r_t[a_t] - \beta) - \frac{\delta \mu}{2} \lambda^2 - \hat{y}_t^T w - \lambda_t (\hat{z}_t^T w - \beta) + \frac{\delta \mu}{2} \lambda_t^2.
\]

(Used fact that \( \mathbb{E}_{a_t \sim p_t} \hat{y}_t = y_t \) and \( \mathbb{E}_{a_t \sim p_t} \hat{z}_t = z_t \))

Take the expectation with respect to \( a_1, ..., a_T \) on the LHS of Inequality 11, we have:

\[
\mathbb{E}_{(a_t)_T} \sum_{t=1}^{T} \left[ \mathcal{L}_t(w_t, \lambda) - \mathcal{L}_t(w, \lambda_t) \right] = \sum_{t=1}^{T} \mathbb{E}_{a_1, ..., a_{t-1}} \mathbb{E}_{a_t|a_1, ..., a_{t-1}} \left[ \mathcal{L}_t(w_t, \lambda) - \mathcal{L}_t(w, \lambda_t) \right] \\
= \sum_{t=1}^{T} \left[ \mathbb{E}_{c_t[a_t]} + \lambda (\mathbb{E}_{r_t[a_t]} - \beta) - \hat{y}_t^T w - \lambda_t (\hat{z}_t^T w - \beta) + \frac{\delta \mu}{2} \lambda_t^2 \right] - \frac{\delta \mu T}{2} \lambda^2
\]

(37)

(40)

Now take the expectation with respect to \( a_1, ..., a_T \) on the RHS of Inequality 11 (we use \( \mathbb{E}_{a_t|a_{t-1}} \) to represent the expectation over the distribution of \( a_t \) conditioned on \( a_1, ..., a_{t-1} \)), we have:

\[
\frac{\lambda^2}{\mu} + \frac{\ln(||I||)}{\mu} + \mu \sum_{t=1}^{T} \left( \mathbb{E}_{a_t|a_{t-1}} \left( \sum_{i=1}^{||I||} w_t[i] \hat{y}_t[i]^2 + \lambda_t^2 w_t[i] \hat{z}_t[i]^2 \right) + \mathbb{E}_{a_t|a_{t-1}} \left( w_t^T \hat{z}_t - \beta \right)^2 + \delta^2 \mu^2 \lambda_t^2 \right) \\
\leq \frac{\lambda^2}{\mu} + \frac{\ln(||I||)}{\mu} + \mu \sum_{t=1}^{T} \left( K + \lambda_t^2 K + 4 + \delta^2 \mu^2 \lambda_t^2 \right) \\
\leq \frac{\lambda^2}{\mu} + \frac{\ln(||I||)}{\mu} + \mu T (K + 4) + \mu (K + \delta^2 \mu^2) \sum_{t=1}^{T} \lambda_t^2.
\]

(Used Eq. 37 and 39)

(41)
Note that based on the setting of $\delta$ and $\mu$, we can show that $\delta \geq 2K + 2\delta^2 \mu^2$. This is because $2K + 2\delta^2 \mu^2 = 2K + 18K^2 \ln(|\Pi|)/(T(K + 4)) \leq 2K + 18K \ln(|\Pi|)/T \leq 3K$, where for simplicity we assume that $T$ is large enough ($T \geq 18 \ln(|\Pi|)$).

Chain Eq. 40 and 41 together and get rid of the terms that have $\lambda_i$ (due to the fact that $\delta \geq 2K + 2\delta^2 \mu^2$) and rearrange terms, we get:

$$\mathbb{E}\left(\sum_{t=1}^{T} c_t [a_t] - y_T^T w + \sum_{t=1}^{T} \left(\lambda (\mathbb{E} r_t[a_t] - \beta) - \lambda_i (z_t^T w - \beta)\right) - \left(\frac{\delta \mu T}{2} + \frac{1}{\mu}\right)\lambda^2\right) \leq \frac{\ln(|\Pi|)}{\mu} + \mu T(K + 4).$$  (42)

The above inequality holds for any $w$. Substitute $w^*$ into Eq. 42, we get:

$$\mathbb{E}\left(\sum_{t=1}^{T} c_t [a_t] - \sum_{t=1}^{T} y_t^T w^* + \sum_{t=1}^{T} \lambda (\mathbb{E} r_t[a_t] - \beta) - \left(\frac{\delta \mu T}{2} + \frac{1}{\mu}\right)\lambda^2\right) \leq \frac{\ln(|\Pi|)}{\mu} + \mu T(K + 4).$$

Now let us set $\lambda = 0$, for regret, we get:

$$\mathbb{E}\left(\sum_{t=1}^{T} c_t [a_t] - \sum_{t=1}^{T} y_t^T w^*\right) \leq \frac{\ln(|\Pi|)}{\mu} + \mu T(K + 4) \leq 2\sqrt{\ln(|\Pi|)T(K + 4)} = O(\sqrt{TK \ln(|\Pi|)}),$$  (43)

where $\mu = \sqrt{\ln(|\Pi|)/T(K + 4)}$.

For constraints $\sum (\mathbb{E} r_t[a_t] - \beta)$, let us assume that $\sum \mathbb{E} (r_t[a_t] - \beta) > 0$ (otherwise we are done), and substitute $\lambda = (\sum \mathbb{E} r_t[a_t] - \beta)/(\delta \mu T + 2/\mu)$ into inequality 43 (note that $\lambda > 0$). Using the fact that $\mathbb{E}\left(\sum_{t=1}^{T} c_t [a_t] - \sum_{t=1}^{T} y_t^T w^*\right) \geq -2T$, we get:

$$(\sum_{t=1}^{T} (\mathbb{E} r_t[a_t] - \beta))^2 \leq (2\delta \mu T + 4/\mu) (2T + 2\sqrt{\ln(|\Pi|)T(K + 2 + 2\beta^2)})$$  (44)

Substitute $\mu = \sqrt{\ln(|\Pi|)/T(K + 4)}$ and $\delta = 3K$ back to the above equation, it is easy to verify that:

$$(\sum_{t=1}^{T} (\mathbb{E} r_t[a_t] - \beta))^2 \leq 12K \sqrt{\ln(|\Pi|)/K + 4} T^{3/2} + 12K \ln(|\Pi|)T + 8T^{3/2} + 8T(K + 4).$$  (45)

Since we consider the asymptotic property when $T \rightarrow \infty$, we can see that the LHS of the above inequality is dominated by $\sqrt{K \ln(|\Pi|)T^{3/2}}$. Hence,

$$(\sum_{t=1}^{T} (\mathbb{E} r_t[a_t] - \beta))^2 \leq O(\sqrt{K \ln(|\Pi|)T^{3/2}}).$$  (46)

Take the square root on both sides of the above inequality, we prove the theorem.

**D. Algorithm and Analysis of EXP4.P.R**

**D.1. Algorithm**

We present the EXP4.P.R algorithm in Alg. 3.
Algorithm 3 Exp4.P with Risk Constraints (EXP4.P.R)

1: **Input:** Policy Set Π
2: Initialize $w_1 = [1/N, ..., 1/N]^T$ and $\lambda_1 = 0$.
3: for $t = 1$ to $T$ do
4:   Receive context $s_t$.
5:   Query each experts to get the sequence of advice $\{\pi_i(s_t)\}_{i=1}^N$.
6:   Draw action $a_t$ randomly according to probability $p_t$.
7:   For each expert $j$, set:
8:      $\hat{y}_t[j] = \pi_j(s_t)^T \hat{c}_t$,  $\hat{z}_t[j] = \pi_j(s_t)^T \hat{r}_t$, $\forall j \in \{1, 2, ..., N\}$. (48)
9: Set the cost vector $\hat{c}_t \in R^K$ and the risk vector $\hat{r}_t \in R^K$ as:
10: $\hat{c}_t[i] = \frac{c_t[i]1(a_t = i)}{p_t[i]}$, $\hat{r}_t[i] = \frac{r_t[i]1(a_t = i)}{p_t[i]}$, $\forall i \in \{1, 2, ..., K\}$. (47)
11: Set $\tilde{x}_t = \hat{y}_t + \lambda_t \hat{z}_t$.
12: Update $w_{t+1}$ as:
13: $w_{t+1}[i] = \frac{w_t[i] \exp(-\mu(\tilde{x}_t[i] - \kappa \sum_{k=1}^K \frac{\pi_i(s_t[k])}{p_t[k]}))}{\sum_{j=1}^{|\Pi|} w_t[j] \exp(-\mu(\tilde{x}_t[j] - \kappa \sum_{k=1}^K \frac{\pi_j(s_t[k])}{p_t[k]}))}$,
14: Update $\lambda_{t+1}$ as:
15: $\lambda_{t+1} = \max\{0, \lambda_t + \mu(w_t^T \hat{z}_t - \beta - \delta \mu \lambda_t)\}$. 
16: end for

We give detailed regret analysis of EXP4.P.R in this section. Let us define \( \hat{x}_t(\lambda) \) as \( \hat{x}_t(\lambda)[i] = \hat{y}_t[i] + \lambda \hat{z}_t[i] - \kappa \sum_{k=1}^K \frac{s_k[i]}{p_k[i]} \), \( \forall i \in [N] \) and \( \mathcal{L}_t(w, \lambda) = w^T \hat{x}_t - \lambda \beta - \frac{\delta}{2} \lambda^2 \). As we can see that Line 3 is essentially running Weighted Majority algorithm on the sequence of functions \( \{ \mathcal{L}_t(w, \lambda) \}_t \) while Line 3 is running Online Gradient Ascent on the sequence of functions \( \{ \mathcal{L}_t(w_t, \lambda) \}_t \). Applying the classic analysis of Weighted Majority and analysis of Online Gradient Descent, we can show that:

**Lemma D.1.** The sequences \( \{ w_t \}_t \) and \( \{ \lambda_t \}_t \) generated from Lines 3 and 3 in EXP4.P.R has the following property:

\[
\sum_{t=1}^{T} \mathcal{L}_t(w_t, \lambda) - \sum_{t=1}^{T} \mathcal{L}_t(w, \lambda_t) \leq \frac{\lambda^2}{\mu} + \frac{\ln([\Pi])}{\mu} + \frac{\mu}{2} \sum_{t=1}^{T} \left( \sum_{i=1}^{[\Pi]} w_t[i](\hat{x}_t(\lambda_t)[i])^2 + 2(w_t^T \hat{z}_t - \beta)^2 + 2\delta^2 \mu^2 \lambda_t^2 \right). \tag{49}
\]

**Proof.** Using the classic analysis of Weighted Majority algorithm, we can get that for the sequence of loss \( \{ \mathcal{L}_t(w, \lambda_t) \}_t \):

\[
\sum_{t=1}^{T} \mathcal{L}_t(w_t, \lambda_t) - \sum_{t=1}^{T} \mathcal{L}_t(w, \lambda_t) \leq \frac{\ln([\Pi])}{\mu} + \frac{\mu}{2} \sum_{t=1}^{T} \sum_{i=1}^{[\Pi]} w_t[i](\hat{x}_t(\lambda_t)[i])^2,
\]

for any \( w \in \mathcal{B} \). On the other hand, we know that we compute \( \lambda_t \) by running Online Gradient Ascent on the loss functions \( \{ \mathcal{L}_t(w_t, \lambda) \}_t \). Applying the classic analysis of Online Gradient Descent, we can get:

\[
\sum_{t=1}^{T} \mathcal{L}_t(w_t, \lambda) - \sum_{t=1}^{T} \mathcal{L}_t(w_t, \lambda_t) \leq \frac{1}{\mu} \lambda^2 + \frac{\mu}{2} \sum_{t=1}^{T} \left( \frac{\partial \mathcal{L}_t(w_t, \lambda_t)}{\partial \lambda_t} \right)^2,
\]

for any \( \lambda \geq 0 \).

We know that \( \partial \mathcal{L}_t(w_t, \lambda) / \partial \lambda_t = w_t^T \hat{z}_t - \beta - \delta \mu \lambda_t \). Substitute these gradient and derivatives back to the above two inequalities, and then sum the above two inequality together we get:

\[
\sum_{t=1}^{T} \mathcal{L}_t(w_t, \lambda) - \sum_{t=1}^{T} \mathcal{L}_t(w, \lambda_t) \leq \frac{\lambda^2}{\mu} + \frac{\ln([\Pi])}{\mu} + \frac{\mu}{2} \sum_{t=1}^{T} \left( \sum_{i=1}^{[\Pi]} w_t[i](\hat{x}_t(\lambda_t)[i])^2 + (w_t^T \hat{z}_t - \beta - \delta \mu \lambda_t)^2 \right)
\]

\[
\leq \frac{\lambda^2}{\mu} + \frac{\ln([\Pi])}{\mu} + \frac{\mu}{2} \sum_{t=1}^{T} \left( \sum_{i=1}^{[\Pi]} w_t[i](x_t(\lambda_t)[i])^2 + 2(w_t^T \hat{z}_t - \beta)^2 + 2\delta^2 \mu^2 \lambda_t^2 \right),
\]

where in the last inequality we use the fact that \( (a + b)^2 \leq 2a^2 + 2b^2 \), for any \( a, b \in \mathbb{R} \). \( \square \)

We first show that the Lagrangian dual parameter \( \lambda_t \) can be upper bounded:

**Lemma D.2.** Assume that \( \delta \leq 1/\mu^2 \). For any \( t \in [T] \), we have \( \lambda_t \leq \frac{|\beta|}{\delta \mu} \).

**Proof.** Remember that the update rule for \( \lambda_t \) is defined as:

\[
\lambda_{t+1} = \max\{0, \lambda_t + \mu(w_t^T \hat{z}_t - \beta - \delta \mu \lambda_t)\}. \tag{50}
\]

We prove the lemma by induction. For \( t = 0 \), since we set \( \lambda_0 = 0 \), we have \( \lambda_0 \leq (|\beta|/\delta \mu) \). Now let us consider time step \( t \) and assume that that \( \lambda_t \leq (|\beta|/\delta \mu) \) for \( t \leq t \). Note that \( w_t^T \hat{z}_t = r_t[a_t] \leq 0 \) and from the update rule of \( \lambda \), we have:

\[
\lambda_{t+1} \leq \max\{0, \lambda_t + \mu(|\beta| - \delta \mu \lambda_t)\} \tag{51}
\]
For the case when \( \lambda_t = 0 \), we have \( \lambda_{t+1} = \mu |\beta| \). Since we assume that \( \delta \leq 1/\mu^2 \), we can easily verify that \( \lambda_{t+1} \leq \mu |\beta| \leq |\beta|/(\delta \mu) \).

For the case when \( \lambda_t \geq 0 \), since we see that \( \lambda_t + \mu (|\beta| - \delta \mu \lambda_t) \geq 0 \) from the induction hypothesis that \( \lambda_t \leq |\beta|/(\delta \mu) \), we must have:

\[
\lambda_{t+1} = \lambda_t + \mu (|\beta| - \delta \mu \lambda_t).
\]

Subtract \( |\delta|/\mu \beta \) on both sides of the above inequality, we get:

\[
\lambda_{t+1} - \frac{|\beta|}{\delta \mu} = (1 - \delta \mu^2)(\lambda_t - \frac{|\beta|}{\delta \mu})
\]

Since we have \( \lambda_t \leq |\beta|/(\delta \mu) \) and \( \delta \leq 1/\mu^2 \), it is easy to see that we have for \( \lambda_{t+1} \):

\[
\lambda_{t+1} - \frac{|\beta|}{\delta \mu} \leq 0.
\]

Hence we prove the lemma. \( \square \)

For notation simplicity, let us denote \( \frac{|\beta|}{\delta \mu} \) as \( \lambda_m \).

We now show how to relate \( \sum_t \hat{y}_t[i] + \lambda_t \hat{z}_t[i] - \kappa \sum_{j=1}^K \frac{\Pi(s_t)[j]}{p_t[j]} \) to \( \sum_t y_t[i] + \lambda_t z[i] \) for any \( i \in [|\Pi|] \):

**Lemma D.3.** In \( \text{EXP4.P.R} \) (Alg. 3), with probability at least \( 1 - \delta \), for any \( w \in \Delta \Pi \), we have:

\[
\sum_{t=1}^T \sum_{i=1}^{|\Pi|} w[i](\hat{y}_t[i] + \lambda_t \hat{z}_t[i] - \kappa \sum_{j=1}^K \frac{\Pi(s_t)[j]}{p_t[j]})
\leq \sum_{t=1}^T \sum_{i=1}^{|\Pi|} (w[i](y_t[i] + \lambda_t z[i]) + (1 + \lambda_m) \frac{\ln(|\Pi|/\delta)}{\kappa}).
\]

We use similar proof strategy as shown in the proof of Lemma 3.1 in (Bubeck et al., 2012) with three additional steps:(1) union bound over all polices in \( \Pi \), (2) introduction of a distribution \( w \in \Delta \Pi \), (3) taking care of \( \lambda_t \) by using its upper bound from Lemma D.2.

**Proof.** Let us set \( \delta' = \delta/|\Pi| \) and fix \( i \in [|\Pi|] \). Define \( \hat{x}_t(\lambda_t) = \hat{y}_t + \lambda_t \hat{z}_t \) and we denote \( \hat{x}_t(\lambda_t)[i] = \hat{x}_t(\lambda_t)[i] - \kappa \sum_{j=1}^K \frac{\Pi(s_t)[j]}{p_t[j]}(\Pi(s_t)[j]/p_t[j]) \).

For notation simplicity, we are going to use \( \hat{x}_t \) and \( \tilde{x}_t \) to represent \( \hat{x}_t(\lambda_t)[i]/(1 + \lambda_m) \) and \( \tilde{x}_t(\lambda_t)[i]/(1 + \lambda_m) \) respectively in the rest of the proof.

Let us also define \( x_t = (y_t[i] + \lambda_t z_t[i])/(1 + \lambda_m) \). It is also straightforward to check that \( \kappa (\tilde{x}_t - x_t) \leq 1 \) since \( \tilde{x}_t \leq 0 \), \( -x_t \leq 1 \) and \( 0 \leq \kappa \leq 1 \). Note that it is straightforward to show that \( \mathbb{E}_t(\tilde{x}_t) = x_t \), where we denote \( \mathbb{E}_t \) as the expectation conditioned on randomness from \( a_1, \ldots, a_{t-1} \).

Following the same strategy in the proof of Lemma 3.1 in (Bubeck et al., 2012), we can show that:

\[
\mathbb{E}_t[\exp(\kappa(\tilde{x}_t - x_t)) = \mathbb{E}_t[\exp(\kappa(\tilde{x}_t - \kappa \sum_{j=1}^K \frac{\Pi(s_t)[j]}{p_t[j]} - x_t))]
\leq (1 + \mathbb{E}_t(\tilde{x}_t - x_t) + \kappa^2 \mathbb{E}_t(\tilde{x}_t - x_t)^2) \exp(-\kappa^2 \sum_{j=1}^K \frac{\Pi(s_t)[j]}{p_t[j]})
\leq (1 + \kappa^2 \mathbb{E}_t(\tilde{x}_t^2)) \exp(-\kappa^2 \sum_{j=1}^K \frac{\Pi(s_t)[j]}{p_t[j]})
\]

\( (55) \)
We can upper bound $\mathbb{E}_t(\hat{x}_t^2)$ as follows:

$$
\mathbb{E}_t(\hat{x}_t^2) = \mathbb{E}_t\left[\left(\sum_{j=1}^{K} \pi_t(s_t)\frac{c_t[j]1(a_t = j)}{p_t[j]} + \lambda m \sum_{j=1}^{K} \pi_t(s_t)\frac{r_t[j]1(a_t = j)}{p_t[j]}\right)^2\right]
\leq \mathbb{E}_{t,j \sim \pi_t(s_t)}\left((\hat{c}_t[j]/p_t[j] + \lambda m \hat{r}_t[j]/p_t[j])/(1 + \lambda m)\right)^2
= \mathbb{E}_{j \sim \pi_t(s_t)} ((c_t[j] + \lambda m r_t[j])/(1 + \lambda m))^2/p_t[j] \leq \mathbb{E}_{j \sim \pi_t(s_t)} (1/p_t[j]) = \sum_{j=1}^{K} \pi_t(s_t)[j] / p_t[j]$$

(56)

where the first inequality comes from Jensen’s inequality and the last inequality comes from the fact that $|c_t[j]| \leq 1$ and $|\lambda m r_t[j]| \leq \lambda m$. Substitute the above results in Eq. 55, we get:

$$
\mathbb{E}_t\left[\exp(\kappa(\hat{x}_t - x_t))\right] \leq (1 + \kappa^2 \sum_{j=1}^{K} \pi_t(s_t)[j]/p_t[j]) \exp(-\kappa^2 \sum_{j=1}^{K} \pi_t(s_t)[j]/p_t[j])
\leq \exp(\kappa^2 \sum_{j=1}^{K} \pi_t(s_t)[j]/p_t[j]) \exp(-\kappa^2 \sum_{j=1}^{K} \pi_t(s_t)[j]/p_t[j]) \leq 1.
$$

(57)

Hence, we have:

$$
\mathbb{E}\exp(\kappa \sum_{t=1}^{T} (\hat{x}_t - x_t)) \leq 1.
$$

(58)

Now from Markov inequality we know $P(X \geq \ln(\delta^{-1})) \leq \delta \mathbb{E}(e^X)$. Hence, this gives us that with probability least $1 - \delta$:

$$
\kappa \sum_{t=1}^{T} (\hat{x}_t - x_t) \leq \ln(1/\delta).
$$

(59)

Substitute the representation of $\hat{x}_t, x_t$ in, we get for $i$, with probability $1 - \delta'$:

$$
\sum_{t=1}^{T} \hat{y}_t[i] + \lambda m \hat{z}_t[i] - \kappa \sum_{j=1}^{K} (\pi_t(s_t)[j]/p_t[j]) \leq \sum_{t=1}^{T} y_t[i] + \lambda m z_t[i] + (1 + \lambda m) \frac{\ln(1/\delta')}{\kappa}.
$$

Now apply union bound over all policies in $\Pi$, it is straightforward to show that for any $i \in \|\Pi\|$, with probability at least $1 - \delta$, we have:

$$
\sum_{t=1}^{T} \hat{y}_t[i] + \lambda m \hat{z}_t[i] - \kappa \sum_{j=1}^{K} (\pi_t(s_t)[j]/p_t[j]) \leq \sum_{t=1}^{T} y_t[i] + \lambda m z_t[i] + (1 + \lambda m) \frac{\ln(\|\Pi\|/\delta)}{\kappa}.
$$

To prove the lemma, now let us fix any $w \in \Delta(\|\Pi\|)$, we can simply multiple $w[i]$ on the both sides of the above inequality, and then sum over from $i = 1$ to $\|\Pi\|$.

Let us define $\hat{w} \in \Delta(\Pi)$ as:

$$
\hat{w} = \arg \min_{w \in \Delta(\Pi)} \sum_{t=1}^{T} \sum_{i=1}^{\|\Pi\|} w[i] (\hat{y}_t[i] + \lambda m \hat{z}_t[i] - \kappa \sum_{j=1}^{K} \pi_t(s_t)[j]/p_t[j]),
$$

(60)

and $\hat{w}^\ast \in \Delta(\Pi)$ as:

$$
\hat{w}^\ast = \arg \min_{w \in \Delta(\Pi)} \sum_{t=1}^{T} \sum_{i=1}^{\|\Pi\|} w[i] (y_t[i] + \lambda m z_t[i])
$$

(61)

Now we turn to prove Theorem 4.3.
Proof of Theorem 4.3. We prove the asymptotic property of Alg. 3 when $T$ approaches to infinity. Since we set $\mu = \sqrt{\frac{\ln(\Pi)}{3K + 1}}$ and $\delta = T^{\theta + 1/2}K$, we can first verify the condition $\delta \leq 1/\mu^2$ in Lemma D.2. This condition holds since $\delta = O(T^{0.5})$ while $1/\mu^2 = \Theta(T)$.

Let us first compute some facts. For $w_t^T \hat{x}_t$, we have:

$$w_t^T \hat{x}_t(\lambda_t) = E_{i \sim u_t}(\hat{y}_t[i] + \lambda_t \hat{z}_t[i] - \kappa \sum_{i=1}^{K} \frac{\pi_i(s_t)[i]}{p_t[i]}) = E_{i \sim u_t}(\hat{y}_t[i] + \lambda_t \hat{z}_t[i] - \kappa \sum_{i=1}^{K} \frac{\pi_i(s_t)[i]}{p_t[i]})$$

$$= c_t[a_t] + \lambda r_t[a_t] - \kappa K. \tag{62}$$

For $\sum_{i=1}^{[\Pi]} w_t[i](\hat{x}_t(\lambda_t)[i])^2$, we have:

$$\sum_{i=1}^{[\Pi]} w_t[i](\hat{x}_t(\lambda_t)[i])^2 = E_{i \sim u_t}(\hat{x}_t(\lambda_t)[i])^2 = E_{i \sim u_t}(\hat{y}_t[i] + \lambda_t \hat{z}_t[i] - \kappa \sum_{i=1}^{K} \frac{\pi_i(s_t)[i]}{p_t[i]})^2$$

$$\leq E_{i \sim u_t, j \sim \pi_t(s_t)}(\hat{c}_t[j] + \lambda r_t[j] - \kappa/p_t[j])^2 = E_{i \sim u_t}(\hat{c}_t[j] + \lambda r_t[j] - \kappa/p_t[j])^2$$

$$= \sum_{i=1}^{K} p_t[i] \left( c_t[i] \mathbb{1}(a_t = i) + \lambda r_t[i] \mathbb{1}(a_t = i) - \kappa \right)^2$$

$$\leq \sum_{i=1}^{K} (1 - \lambda_t - \kappa)(\hat{c}_t[i] + \lambda r_t[i] - \kappa/p_t[i])$$

$$= K(-1 - \lambda_t - \kappa) \sum_{i=1}^{[\Pi]} \hat{w}[i](\hat{y}_t[i] + \lambda t \hat{z}_t[i] - \kappa \sum_{j=1}^{K} \frac{\pi_i(s_t)[j]}{p_t[j]}), \tag{63}$$

where the first inequality comes from Jesen’s inequality and the last inequality uses the assumption that the $\Pi$ contains the uniform policy (i.e., the policy that assign probability $1/K$ to each action). Consider the RHS of Eq. 49, we have:

$$\frac{\lambda^2}{\mu} + \frac{\ln([\Pi])}{\mu} + \frac{\mu T}{2} \sum_{i=1}^{[\Pi]} \sum_{t=1}^{T} w_t[i](\hat{x}_t(\lambda_t)[i])^2 + \frac{\mu T}{2} \sum_{t=1}^{T} (w_t^T \hat{z}_t - \beta - \delta \mu \lambda_t)^2$$

$$\leq \frac{\lambda^2}{\mu} + \frac{\ln([\Pi])}{\mu} + \frac{\mu T}{2} \sum_{t=1}^{T} K(-1 - \lambda_t - \kappa) \left( \sum_{i=1}^{[\Pi]} \hat{w}[i](\hat{y}_t[i] + \lambda t \hat{z}_t[i] - \kappa \sum_{j=1}^{K} \frac{\pi_i(s_t)[j]}{p_t[j]}) \right)$$

$$+ \mu \sum_{t=1}^{T} ((w_t^T \hat{z}_t - \beta)^2 + \delta^2 \mu^2 \lambda_t^2)$$

$$= \frac{\lambda^2}{\mu} + \frac{\ln([\Pi])}{\mu} + \frac{\mu T}{2} \sum_{t=1}^{T} K(-1 - \lambda_t - \kappa) \left( \sum_{i=1}^{[\Pi]} \hat{w}[i](\hat{y}_t[i] + \lambda t \hat{z}_t[i] - \kappa \sum_{j=1}^{K} \frac{\pi_i(s_t)[j]}{p_t[j]}) \right)$$

$$+ \mu \sum_{t=1}^{T} (r_t[a_t] - \beta)^2 + \delta^2 \mu^2 \lambda_t^2). \tag{64}$$

Consider the LHS of Eq. 49, set $w = \hat{w}$, we have:

$$\sum_{t=1}^{T} \left[ L_t(w_t, \lambda) - L_t(\hat{w}, \lambda_t) \right]$$

$$= \sum_{t=1}^{T} \left[ c_t[a_t] + \lambda r_t[a_t] - \kappa K - \lambda \beta - \delta \mu \lambda^2/2 - \left( \sum_{i=1}^{[\Pi]} \hat{w}[i](\hat{y}_t[i] + \lambda t \hat{z}_t[i] - \kappa \sum_{j=1}^{K} \frac{\pi_i(s_t)[j]}{p_t[j]}) \right) + \lambda \beta + \delta \mu \lambda_t^2/2 \right]. \tag{65}$$
Chaining Eq. 64 and Eq. 65 together and rearrange terms, we will get:

\[
\sum_{t=1}^{T} \left[ c_t[a_t] + \lambda(r_t[a_t] - \beta) + \lambda_t\beta + \delta\mu \lambda^2_t/2 \right] - T\delta\mu \lambda^2/2 \\
\leq T\kappa K + \frac{\lambda^2 + \ln(|\Pi|)}{\mu} + \sum_{t=1}^{T} (1 - \frac{\mu K}{2} (1 + \lambda_t + \kappa)) \left( \sum_{i=1}^{|\Pi|} \hat{w}[i](\hat{y}_t[i] + \lambda_t z_t[i]) \right) \\
+ \mu \sum_{t=1}^{T} (2 + 2\beta^2 + \delta^2 \mu \lambda^2_t).
\]  

(66)

Since we have \( \delta \geq \frac{\beta^2}{2|\Pi| - \mu - \kappa} \), we can show that \( 1 - \frac{\mu K}{2} (1 + \lambda_t + \kappa) \geq 0 \).

Now back to Eq. 66, using Lemma. D.3, we have with probability \( 1 - \nu \):

\[
\sum_{t=1}^{T} \left[ c_t[a_t] + \lambda(r_t[a_t] - \beta) + \lambda_t\beta + \delta\mu \lambda^2_t/2 \right] - T\delta\mu \lambda^2/2 \\
\leq T\kappa K + \frac{\lambda^2 + \ln(|\Pi|)}{\mu} + \sum_{t=1}^{T} (1 - \frac{\mu K}{2} (1 + \lambda_t + \kappa)) \left( \sum_{i=1}^{|\Pi|} \hat{w}^*[i](\hat{y}_t[i] + \lambda_t z_t[i]) \right) \\
+ (1 + \lambda_m) \frac{\ln(|\Pi|/\nu)}{\kappa} + (2 + 2\beta^2)T\mu + \mu^3\delta^2 \sum_{t} \lambda^2_t \\
\leq T\kappa K + \frac{\lambda^2 + \ln(|\Pi|)}{\mu} + \sum_{t=1}^{T} (1 - \frac{\mu K}{2} (1 + \lambda_t + \kappa)) \left( \sum_{i=1}^{|\Pi|} w^*[i](\hat{y}_t[i] + \lambda_t z_t[i]) \right) \\
+ (1 + \lambda_m) \frac{\ln(|\Pi|/\nu)}{\kappa} + (2 + 2\beta^2)T\mu + \mu^3\delta^2 \sum_{t} \lambda^2_t.
\]  

(67)

where the last inequality follows from the definition of \( \hat{w}^* \) and \( w^* \). Rearrange terms, we get:

\[
\sum_{t=1}^{T} \left[ (c_t[a_t] - w^*\hat{y}_t) + \lambda(r_t[a_t] - \beta) - \lambda_t(w^*\hat{z}_t - \beta) \right] - T\delta\mu \lambda^2/2 + \sum_{t=1}^{T} \delta\mu \lambda^2_t/2 \\
\leq T\kappa K + \frac{\lambda^2 + \ln(|\Pi|)}{\mu} + \sum_{t=1}^{T} \frac{\mu K}{2} (1 + \lambda_t + \kappa)(1 + \lambda_t) + (1 + \lambda_m) \frac{\ln(|\Pi|/\nu)}{\kappa} + (2 + 2\beta^2)T\mu + \mu^3\delta^2 \sum_{t} \lambda^2_t \\
\leq T\kappa K + \frac{\lambda^2 + \ln(|\Pi|)}{\mu} + \sum_{t=1}^{T} \frac{\mu K}{2} (1 + (2 + \kappa)\lambda_t + \kappa) + (1 + \lambda_m) \frac{\ln(|\Pi|/\nu)}{\kappa} + (2 + 2\beta^2)T\mu + \left( \frac{K\mu}{2} + \mu^3\delta^2 \right) \sum_{t} \lambda^2_t \\
= T\kappa K + \frac{\lambda^2 + \ln(|\Pi|)}{\mu} + \sum_{t=1}^{T} \frac{\mu K}{2} (1 + (2 + \kappa)\lambda_t + \kappa) + \left( \frac{\beta^2}{\delta\mu} \right) \frac{\ln(|\Pi|/\nu)}{\kappa} + (2 + 2\beta^2)T\mu + \left( \frac{K\mu}{2} + \mu^3\delta^2 \right) \sum_{t} \lambda^2_t.
\]  

(68)

Note that under the setting of \( \delta \) and \( \mu \) we have \( \frac{\delta\mu}{\nu^2} \geq \frac{K\mu}{2} + \mu^3\delta^2 \) (we will verify it at the end of the proof), we can drop the terms that relates to \( \lambda^2_t \) in the above inequality. Note that we have \( \delta\mu = T^{-\epsilon} \sqrt{K\ln(|\Pi|)} \geq T^{-\epsilon} \), where \( \epsilon \in (0, 1/2) \). Substitute \( \delta\mu \geq T^{-\epsilon} \) into the above inequality and rearrange terms, we get:

\[
\sum_{t=1}^{T} c_t[a_t] - w^*\hat{y}_t + \lambda(r_t[a_t] - \beta) - \lambda_t(w^*\hat{z}_t - \beta) - T\delta\mu \lambda^2/2 \\
= \frac{\lambda^2 + \ln(|\Pi|)}{\mu} + T\kappa K + (K + 2 + 2\beta^2 + 2K|\beta|)T\mu + (1 + |\beta|) \frac{\ln(|\Pi|/\nu)}{\kappa}
\]  

(69)
Now let us set $\lambda = 0$ and since we have that $\sum_{t=1}^{\infty} \lambda_t (w^* z_t - \beta) \leq 0$, we get:

$$
\sum_{t=1}^{\infty} c_t [a_t] - w^* y_t \leq \frac{\ln(||\Pi||)}{\mu} + T \kappa K + (K + 2 + \beta^2 + 2K\|\beta\|)T \mu + (1 + \|\beta\|T^c) \frac{\ln(||\Pi||/\nu)}{\kappa}
$$

$$
\leq \frac{\ln(||\Pi||)}{\mu} + T \kappa K + (3K + 4)T \mu + (1 + T^c) \frac{\ln(||\Pi||/\nu)}{\kappa}
$$

$$
\leq 2\sqrt{T(\ln(||\Pi||)(3K + 4))} + 2\sqrt{TK(1 + T^c) \ln(||\Pi||/\nu)} = O(\sqrt{T^{1+\epsilon} K \ln(||\Pi||/\nu)})
$$

(70)

where we set $\mu$ and $\kappa$ as:

$$
\mu = \sqrt{\frac{\ln(||\Pi||)}{(K + 3)T}}, \quad \kappa = \sqrt{\frac{(1 + T^c) \ln(||\Pi||/\nu)}{TK}}.
$$

(71)

Now let us consider $\sum_t (r_t [a_t] - \beta)$. Let us assume $\sum_t (r_t [a_t] - \beta) \geq 0$, otherwise we prove the theorem already. Note that $\sum_{t=1}^{\infty} c_t [a_t] - w^* y_t \geq -2T$. Hence we have:

$$
\lambda \sum_{t=1}^{\infty} (r_t [a_t] - \beta) - \lambda^2 (\mu T/2 + 1/\mu)
$$

$$
\leq 2T + 2\sqrt{T(\ln(||\Pi||)(3K + 4))} + 2\sqrt{TK(1 + T^c) \ln(||\Pi||/\nu)}.
$$

To maximize the LHS of the above inequality, we set $\lambda = \frac{\sum_{t=1}^{\infty} (r_t [a_t] - \beta)}{\delta T^2/\mu}$. Substitute $\lambda$ into the above inequality, we get:

$$
(T \sum_{t=1}^{\infty} (r_t [a_t] - \beta))^2 \leq (2\delta T + \frac{4}{\mu})(2T + 2\sqrt{T(\ln(||\Pi||)(3K + 4))} + 2\sqrt{TK(1 + T^c) \ln(||\Pi||/\nu)})
$$

$$
\leq (2T^{1-\epsilon} \sqrt{\ln(||\Pi||)} + \frac{4}{\mu})(2T + 2\sqrt{TK(1 + T^c) \ln(||\Pi||/\nu)})
$$

$$
= 24(T^{2-\epsilon} \sqrt{K \ln(||\Pi||)} + T^{1.5-\epsilon} K \ln(||\Pi||) + T^{1.5-0.5\epsilon} K \ln(||\Pi||) + T^{1.5} \sqrt{K + TK + T^{1+\epsilon} K \ln(1/\delta)})
$$

$$
= O(T^{2-\epsilon} K \ln(||\Pi||)).
$$

(72)

Hence we have:

$$
\sum_{t=1}^{\infty} (r_t [a_t] - \beta) = O(T^{1-\epsilon/2} \sqrt{K \ln(||\Pi||)}).
$$

(73)

Note that for $\delta$, we have $\delta = KT^{-\epsilon+0.5}$. To verify that $\delta \geq \frac{1}{2TK-\mu-\kappa\mu}$ we can see that as long as $\epsilon \in (0, 1/2)$, we have $\delta = \Theta(T^{0.5-\epsilon})$ while $||\beta||/(2/K - \mu - \kappa\mu) = O(1)$. Hence when $T$ is big enough, we can see that it always holds that $\delta \geq \frac{1}{2TK-\mu-\kappa\mu}$. For the second condition that $\delta \geq K + 2\mu^2 \delta^2 = K + 2\ln(||\Pi||)K T^{-2\epsilon}$. Note that again as long as $\epsilon \in (0, 1/2)$, we have $\delta = \Theta(T^{0.5-\epsilon})$, and $K + 2\ln(||\Pi||)K T^{-2\epsilon} = O(1)$. Hence we have $\delta \geq K + 2\ln(||\Pi||)K T^{-2\epsilon}$. Hence, we have shown that when $\mu = \sqrt{\frac{\ln(||\Pi||)}{(3K + 4)T}}, \kappa = \sqrt{\frac{(1 + T^c) \ln(||\Pi||/\nu)}{TK}},$ and $\delta = T^{-\epsilon+0.5} K$, we have that as $T \to \infty$:

$$
\sum_{t=1}^{\infty} (c_t [a_t] - w^* y_t) = O(\sqrt{T^{1+\epsilon} \ln(||\Pi||/\nu)}),
$$

$$
\sum_{t=1}^{\infty} (r_t [a_t] - \beta) \leq O(T^{1-\epsilon/2} \sqrt{K \ln(||\Pi||)}).
$$

(74)