Supplementary material for "Efficient Algorithms for Adversarial Contextual Learning"

A. Omitted Proofs from Section 2

A.1. Proof of Theorem 1

We prove the theorem by analyzing a slightly modified algorithm, that only draws the perturbation once at the beginning of the learning process but is otherwise identical. The bulk of the proof is devoted to bounding this modified algorithm's regret against oblivious adversaries, i.e. an adversary that chooses the outcomes $y^{1:T}$ before the learning process begins. We use this regret bound along with a reduction due to Hutter and Poland (Hutter & Poland, 2005) (see their Lemma 12) to obtain a regret bound for Algorithm 1 against adaptive adversaries. We provide a proof of this reduction in Appendix A.2 and proceed here with the analysis of the modified algorithm.

To bound the regret of the modified algorithm, consider letting the algorithm observe y^t ahead of time, so that at each time step t, the algorithm plays $\pi^{t+1} = M\left(\{z\} \cup y^{1:t}\right)$. Notice trivially that the regret of the modified algorithm is,

$$\textsc{Regret} = \max_{\pi\in\Pi} \sum_{t=1}^{T} \ell(\pi^t, y^t) - \ell(\pi, y^t) = \max_{\pi\in\Pi} \sum_{t=1}^{T} \ell(\pi^t, y^t) - \ell(\pi^{t+1}, y^t) + \sum_{t=1}^{T} \ell(\pi^{t+1}, y^t) - \ell(\pi, y^t)$$

The first sum here is precisely the STABILITY term in the bound, so we must show that the second sum is bounded by ERROR. This is proved by induction in the following lemma.

Lemma 7 (Be-the-leader with fixed sample perturbations)**.** *For any realization of the sample sequence $\{z\}$ and for any policy π^*:*

$$\sum_{t=1}^{T} \left(\ell(\pi^{t+1}, y^t) - \ell(\pi^*, y^t)\right) \le \max_{\pi\in\Pi} \sum_{z^\tau\in\{z\}} \ell(\pi, z^\tau) - \min_{\pi\in\Pi} \sum_{z^\tau\in\{z\}} \ell(\pi, z^\tau) \tag{13}$$

Proof. Denote with k the length of sequence $\{z\}$. Consider the sequence $\{z\} \cup y^{1:T}$ and let $a^1 = M(\{z\})$. We will show that for any policy π^*:

$$\sum_{\tau=1}^{k} \ell(\pi^1, z^\tau) + \sum_{t=1}^{T} \ell(\pi^{t+1}, y^t) \le \sum_{\tau=1}^{k} \ell(\pi^*, z^\tau) + \sum_{t=1}^{T} \ell(\pi^*, y^t) \tag{14}$$

For $T = 0$, the latter trivially holds by the definition of a^1. Suppose it holds for some T, we will show that it holds for $T + 1$. Since the induction hypothesis holds for any π^*, applying it for a^{T+2}, i.e.,:

$$\begin{aligned}\sum_{\tau=1}^{k} \ell(\pi^1, z^\tau) + \sum_{t=1}^{T+1} \ell(\pi^{t+1}, y^t) \le & \sum_{\tau=1}^{k} \ell(\pi^{T+2}, z^\tau) + \sum_{t=1}^{T} \ell(\pi^{T+2}, y^t) + \ell(\pi^{T+2}, y^{T+1}) \\ = & \sum_{\tau=1}^{k} \ell(\pi^{T+2}, z^\tau) + \sum_{t=1}^{T+1} \ell(\pi^{T+2}, y^t)\end{aligned}$$

By definition of a^{T+2} the latter is at most: $\sum_{\tau=1}^{k} \ell(\pi^*, z^\tau) + \sum_{t=1}^{T+1} \ell(\pi^*, y^t)$ for any π^*. Which proves the induction step. Thus, by re-arranging Equation (14) we get:

$$\sum_{t=1}^{T} \left(\ell(\pi^{t+1}, y^t) - \ell(\pi^*, y^t)\right) \le \sum_{\tau=1}^{k} \left(\ell(\pi^*, z^\tau) - \ell(\pi^1, z^\tau)\right) \le \max_{\pi\in\Pi} \sum_{\tau=1}^{k} \ell(\pi, z^\tau) - \min_{\pi\in\Pi} \sum_{\tau=1}^{k} \ell(\pi, z^\tau)$$

■

Thus the regret of the modified algorithm against an oblivious adversary is bounded by STABILITY + ERROR. By applying the reduction of Hutter and Poland (Hutter & Poland, 2005) (see Appendix A.2 for a proof sketch), the regret of Algorithm 1 is bounded is bounded in the same way.

A.2. From adaptive to oblivious adversaries

We will utilize a generic reduction provided in Lemma 12 of (Hutter & Poland, 2005), which states that given that in Algorithm 1 we draw independent randomization at each iteration, it suffices to provide a regret bound only for oblivious adversaries, i.e., the adversary picks a fixed sequence $y^{1:T}$ ahead of time without observing the policies of the player. Moreover, for any such fixed sequence of an oblivious adversary, the expected loss of the algorithm can be easily shown to be equal to the expected loss if we draw a single random sequence $\{z\}$ ahead of time and use the same random vector all the time.

The proof is as follows: by linearity of expectation and the fact that each sequence $\{z\}^t$ drawn at each time-step t is identically distributed:

$$\begin{aligned}\mathbb{E}_{\{z\}^1,\ldots,\{z\}^t}\left[\sum_{t=1}^{T} u(M(\{z\}^t \cup y^{1:t-1}), y^t)\right] &= \sum_{t=1}^{T} \mathbb{E}_{\{z\}^t}\left[u(M(\{z\}^t \cup y^{1:t-1}), y^t)\right] \\ &= \sum_{t=1}^{T} \mathbb{E}_{\{z\}^1}\left[u(M(\{z\}^1 \cup y^{1:t-1}), y^t)\right] \\ &= \mathbb{E}_{\{z\}^1}\left[\sum_{t=1}^{T} u(M(\{z\}^1 \cup y^{1:t-1}), y^t)\right]\end{aligned}$$

The latter is equivalent to the expected reward if we draw a single random sequence $\{z\}$ ahead of time and use the same random vector all the time. Thus it is sufficient to upper bound the regret of this modified algorithm, which draws randomness only once.

Thus it is sufficient to upper bound the regret of this modified algorithm, which draws randomness only once.

B. Omitted Proofs from Section 3

B.1. Bounding the Laplacian Error

The upper bound on the ERROR term is identical in all settings, since it only depends on the input noise distribution, which is the same for all variants and for which it does not matter whether X is the set of contexts that will arrive or a separator. In subsequent sections we will upper bound the stability of the algorithm in each setting.

Lemma 8 (Laplacian Error Bound)**.** *Let $\{z\}$ denote a sample from the random sequence of fake samples used by* CONTEXT-FTPL(X, ϵ)*. Then:*

$$\text{ERROR} = \mathbb{E}_{\{z\}}\left[\max_{\pi\in\Pi}\sum_{x\in X}\langle\pi(x), \ell_x\rangle\right] - \mathbb{E}_{\{z\}}\left[\min_{\pi\in\Pi}\sum_{x\in X}\langle\pi(x), \ell_x\rangle\right] \leq \frac{10}{\epsilon}\sqrt{dm}\log(N) \tag{15}$$

Proof. First we start by observing that each random variable $\ell_x(j)$ is distributed i.i.d. according to a Laplace(ϵ) distribution. Since a Laplace distribution is symmetric around 0, we get that $\ell_x(j)$ and $-\ell_x(j)$ are distributed identically. Thus we can write:

$$\mathbb{E}_{\{z\}}\left[\min_{\pi\in\Pi}\sum_{x\in X}\langle\pi(x), \ell_x\rangle\right] = \mathbb{E}_{\{z\}}\left[\min_{\pi\in\Pi}\sum_{x\in X}\langle\pi(x), -\ell_x\rangle\right] = -\mathbb{E}_{\{z\}}\left[\max_{\pi\in\Pi}\sum_{x\in X}\langle\pi(x), \ell_x\rangle\right]$$

Hence we get:

$$\text{ERROR} = 2\cdot\mathbb{E}_{\{z\}}\left[\max_{\pi\in\Pi}\sum_{x\in X}\langle\pi(x), \ell_x\rangle\right] \tag{16}$$

We now bound the latter expectation via a moment generating function approach. For any $\lambda \geq 0$:

$$\begin{aligned}\mathbb{E}_{\{z\}}\left[\max_{\pi\in\Pi}\sum_{x\in X}\langle\pi(x),\ell_x\rangle\right] &= \frac{1}{\lambda}\mathbb{E}_{\{z\}}\left[\max_{\pi\in\Pi}\lambda\sum_{x\in X}\langle\pi(x),\ell_x\rangle\right]\\ &= \frac{1}{\lambda}\log\left\{\exp\left\{\mathbb{E}_{\{z\}}\left[\max_{\pi\in\Pi}\lambda\sum_{x\in X}\langle\pi(x),\ell_x\rangle\right]\right\}\right\}\end{aligned}$$

By convexity and monotonicity of the exponential function:

$$\begin{aligned}\mathbb{E}_{\{z\}}\left[\max_{\pi\in\Pi}\sum_{x\in X}\langle\pi(x),\ell_x\rangle\right] &\leq \frac{1}{\lambda}\log\left\{\mathbb{E}_{\{z\}}\left[\max_{\pi\in\Pi}\exp\left\{\lambda\sum_{x\in X}\langle\pi(x),\ell_x\rangle\right\}\right]\right\}\\ &\leq \frac{1}{\lambda}\log\left\{\sum_{\pi\in\Pi}\mathbb{E}_{\{z\}}\left[\exp\left\{\lambda\sum_{x\in X}\langle\pi(x),\ell_x\rangle\right\}\right]\right\}\\ &\leq \frac{1}{\lambda}\log\left\{\sum_{\pi\in\Pi}\prod_{x\in X}\mathbb{E}\left[\exp\left\{\lambda\langle\pi(x),\ell_x\rangle\right\}\right]\right\}\\ &= \frac{1}{\lambda}\log\left\{\sum_{\pi\in\Pi}\prod_{x\in X}\mathbb{E}\left[\exp\left\{\lambda\sum_{j:\pi(x)(j)=1}\ell_x(j)\right\}\right]\right\}\\ &= \frac{1}{\lambda}\log\left\{\sum_{\pi\in\Pi}\prod_{x\in X}\prod_{j:\pi(x)(j)=1}\mathbb{E}\left[\exp\left\{\lambda\ell_x(j)\right\}\right]\right\}\end{aligned}$$

For any $j \in [K]$ and $x \in X$, $\ell_x(j)$ is a Laplace(ϵ) random variable. Hence, the quantity $\mathbb{E}\left[\exp\{\lambda\ell_x(j)\}\right]$ is the moment generating function of the Laplacian distribution evaluated at λ, which is equal to $\frac{1}{1-\frac{\lambda^2}{\epsilon^2}}$ provided that $\lambda < \epsilon$. Since $\sup_{x,\pi}|\{j\in[K]:\pi(x)(j)=1\}| \leq m$, we get:

$$\mathbb{E}_{\{z\}}\left[\max_{\pi\in\Pi}\sum_{x\in X}\langle\pi(x),\ell_x\rangle\right] \leq \frac{1}{\lambda}\log\left\{N\left(\frac{1}{1-\frac{\lambda^2}{\epsilon^2}}\right)^{dm}\right\} = \frac{1}{\lambda}\log(N) + \frac{dm}{\lambda}\log\left(\frac{1}{1-\frac{\lambda^2}{\epsilon^2}}\right)$$

By simple calculus, it is easy to derive that $\frac{1}{1-x} \leq e^{2x}$ for any $x \leq \frac{1}{4}$.[3] Thus as long as we pick $\lambda \leq \frac{\epsilon}{2}$, we get:

$$\mathbb{E}_{\{z\}}\left[\max_{\pi\in\Pi}\sum_{x\in X}\langle\pi(x),\ell_x\rangle\right] \leq \frac{1}{\lambda}\log(N) + \frac{dm}{\lambda}\log\left(\exp\left\{\frac{\lambda^2}{\epsilon^2}\right\}\right) = \frac{1}{\lambda}\log(N) + \frac{2dm\lambda}{\epsilon^2}$$

Picking $\lambda = \frac{\epsilon}{2\sqrt{dm}}$ and since $N \geq 2$:

$$\mathbb{E}_{\{z\}}\left[\max_{\pi\in\Pi}\sum_{x\in X}\langle\pi(x),\ell_x\rangle\right] \leq \frac{2\sqrt{dm}\log(N)}{\epsilon} + \frac{\sqrt{2dm}}{\epsilon} \leq \frac{5\sqrt{dm}\log(N)}{\epsilon}$$

■

B.2. Bounding Stability: Transductive Setting

We now turn to bounding the stability in the transductive combinatorial optimization setting. Combining the following lemma with the error bound in Lemma 8 and applying Theorem 1 proves the first claim of Theorem 2.

[3] Consider the function $f(x) = (1-x)e^{2x} - 1$. Then $f(0) = 0$ and $f'(x) = e^{2x}(1-2x)$, which is ≥ 0 for $0 \leq x \leq 1/2$.

Lemma 9 (Transductive Stability)**.** *For all $t \in [T]$ and for any sequence $y^{1:t}$ of contexts $x^{1:t}$ and loss functions $f^{1:t}$ with $f^i : \{0,1\}^K \to \mathbb{R}^K$, the stability of* CONTEXT-FTPL(X, ϵ) *is upper bounded by:*

$$\mathbb{E}_{\{z\}}\left[f^t(\pi^t(x^t)) - f^t(\pi^{t+1}(x^t))\right] \leq 4\epsilon K \cdot \|f^t\|_*^2$$

Proof. By the definition of $\|f^t\|_*$:

$$\mathbb{E}_{\{z\}}\left[f^t(\pi^t(x^t)) - f^t(\pi^{t+1}(x^t))\right] \leq 2\|f^t\|_* \Pr\left[\pi^t(x^t) \neq \pi^{t+1}(x^t)\right]$$

Now observe that:

$$\Pr\left[\pi^t(x^t) \neq \pi^{t+1}(x^t)\right] \leq \sum_{j \in K}\left(\Pr[j \in \pi^t(x^t), j \notin \pi^{t+1}(x^t)] + \Pr[j \notin \pi^t(x^t), j \in \pi^t(x^t)]\right)$$

We bound the probability $\Pr[j \in \pi^t(x^t), j \notin \pi^{t+1}(x^t)]$. We condition on all random variables of $\{z\}$ except for the random variable $\ell_{x^t}(j)$, i.e. the random loss placed at coordinate j on the sample associated with context x^t. Denote the event corresponding to an assignment of all these other random variables as $\mathcal{E}_{-x^t j}$. Let $\ell_{x^t j}$ denote a loss vector which is $\ell_{x^t}(j)$ on the j-th coordinate and zero otherwise. Also let:

$$\Phi(\pi) = \sum_{\tau=1}^{t-1} f^\tau(\pi(x^\tau)) + \sum_{x \in X - \{x^t\}} \langle \pi(x), \ell_x \rangle + \langle \pi(x^t), \ell_{x^t} - \ell_{x^t j} \rangle \tag{17}$$

Let $\pi^* = \operatorname{argmin}_{\pi \in \Pi : j \in \pi(x^t)} \Phi(\pi)$ and $\tilde{\pi} = \min_{\pi \in \Pi : j \notin \pi(x^t)} \Phi(\pi)$. The event that $\{j \in \pi^t(x^t)\}$ happens only if:

$$\Phi(\pi^*) + \ell_{x^t}(j) \leq \Phi(\tilde{\pi}) \tag{18}$$

Let and $\nu = \Phi(\tilde{\pi}) - \Phi(\pi^*)$. Thus $j \in \pi^t(x^t)$ only if:

$$\ell_{x^t}(j) \leq \nu \tag{19}$$

Now if:

$$\ell_{x^t}(j) < \nu - 2\|f^t\|_* \tag{20}$$

then it is easy to see that $\{j \in \pi^{t+1}(x^t)\}$, since an extra loss of $f^t(a) \in [0,1]$ cannot push j out of the optimal solution. More elaborately, for any other policy $\pi \in \Pi$, such that $j \notin \pi(x^t)$, the loss of π^* including time-step t is bounded as:

$$\begin{aligned}
\Phi(\pi^*) + \ell_{x^t}(j) + f^t(\pi^*(x^t)) &< \Phi(\pi) - 2\|f^t\|_* + f^t(\pi^*(x^t)) \\
&< \Phi(\pi) - \|f^t\|_* \\
&< \Phi(\pi) + f^t(\pi(x^t))
\end{aligned}$$

Thus any policy π, such that $j \notin \pi(x^t)$ is suboptimal after seeing the loss at time-step t. Thus

$$\Pr[j \in \pi^t(x^t), j \notin \pi^{t+1}(x^t) \mid \mathcal{E}_{-x^t j}] \leq \Pr[\ell_{x^t}(j) \in [\nu - 2\|f^t\|_*, \nu] \mid \mathcal{E}_{-x^t j}]$$

Since all other random variables are independent of $\ell_{x^t}(j)$ and $\ell_{x^t}(j)$ is a Laplacian with parameter ϵ:

$$\begin{aligned}
\Pr[\ell_{x^t}(j) \in [\nu - 2\|f^t\|_*, \nu] \mid \mathcal{E}_{-x^t j}] &= \Pr[\ell_{x^t}(j) \in [\nu - 2\|f^t\|_*, \nu]] \\
&= \frac{\epsilon}{2} \int_{\nu - 2\|f^t\|_*}^{\nu} e^{-\epsilon|z|} dz \leq \frac{\epsilon}{2} \int_{\nu - 2\|f^t\|_*}^{\nu} dz \leq \epsilon \|f^t\|_*
\end{aligned}$$

Similarly it follows that that: $\Pr[j \notin \pi^t(x^t)$ and $j \in \pi^{t+1}(x^t)] \leq \epsilon\|f^t\|_*$. To sum we get that:

$$\mathbb{E}_{\{z\}}\left[f^t(\pi^t(x^t)) - f^t(\pi^{t+1}(x^t))\right] \leq 2\|f^t\|_* \Pr\left[\pi^t(x^t) \neq \pi^{t+1}(x^t)\right] \leq 4\epsilon K \|f^t\|_*^2$$

■

B.3. Bounding Stability: Transductive Setting with Linear Losses

In the transductive setting with linear losses, we provide a significantly more refined stability bound, which enables applications to partial information or bandit settings. As before, combining this stability bound with the error bound in Lemma 8 and applying Theorem 1 gives the second claim of Theorem 2.

Lemma 10 (Multiplicative Stability)**.** *For any sequence $y^{1:T}$ for all $t \in [T]$ of contexts and non-negative linear loss functions, the stability of* CONTEXT-FTPL(X, ϵ) *in the transductive setting, is upper bounded by:*

$$\mathbb{E}_{\{z\}}\left[\langle \pi^t(x^t), \ell^t\rangle - \langle \pi^{t+1}(x^t), \ell^t\rangle\right] \leq \epsilon \cdot \mathbb{E}\left[\langle \pi^t(x^t), \ell^t\rangle^2\right]$$

Proof. To prove the result we first must introduce some additional terminology. For a sequence of parameters $y^{1:t}$, let $\phi^t \in \mathbb{R}^{dK}$ be a vector with $\phi^t_{x,j} = \sum_{\tau \leq t: x^\tau = x} \ell^\tau(j)$. The component of this vector corresponding to context $x \in X$ and coordinate $j \in [K]$ is the cumulative loss associated with that coordinate on the subset of time points when context x appeared. Note that this vector ϕ^t is a sufficient statistic, since for any fixed policy π:

$$\sum_{\tau=1}^{t} \ell(\pi, y^\tau) = \sum_{x \in X} \sum_{\tau \leq t: x^\tau = x} \langle \pi(x), \ell^\tau\rangle = \sum_{x \in X} \langle \pi(x), \phi^t_x\rangle \tag{21}$$

where $\phi^t_x = \sum_{\tau \leq t: x^\tau = x} \ell^\tau$.

We denote with $z \in \mathbb{R}^{dK}$ the sufficient statistic that corresponds to the fake sample sequence $\{z\}$ and with ϕ^t the sufficient statistics for the parameter sequence $y^{1:t}$. Observe that the sufficient statistic for the augmented sequence $\{z\} \cup y^{1:t}$ is simply $z + \phi^t$. For any sequence of parameters $y^{1:T}$ we will be denoting with $\phi^{1:T}$ the sequence of $d \cdot K$ dimensional cumulative loss vectors. We will also overload notation and denote with $M(\phi^t) = M(y^{1:t})$ the best policy on a sequence $y^{1:t}$ with statistics ϕ^t.

Consider a specific sequence $y^{1:T}$ and a specific time step t. Define, for each $\pi \in \Pi$, a sparse tuple $y^t_\pi = (x^t, \ell^t_\pi)$ where $\ell^t_\pi(j) = \ell^t(j)$ if $\pi(x^t)(j) = 1$ and zero otherwise, i.e. we zero out coordinates of the true loss vector that were not picked by the policy π. Moreover, define with ϕ^t_π the sufficient statistic of the sequence $\phi(y^{1:t-1} \cup y^t_\pi)$ for each π. We define $1 + |\Pi|$ distributions over $|\Pi|$, via their probability density functions, as follows:

$$\begin{aligned} p^t(\pi) &= \Pr[M(z + \phi^{t-1}) = \pi] \\ \forall \pi^* \in \Pi: p^{t+1}_{\pi^*}(\pi) &= \Pr\left[M(z + \phi^t_{\pi^*}) = \pi\right] \end{aligned}$$

At the end of this proof, we will show that $p^{t+1}_\pi(\pi) \leq p^{t+1}(\pi)$. Moreover, we denote for convenience:

$$\begin{aligned} \texttt{FTPL}^t &= \mathbb{E}_z[\langle \pi^t(x^t), \ell^t\rangle] = \mathbb{E}_{\pi \sim p^t}\left[\langle \pi(x^t), \ell^t\rangle\right] \\ \texttt{BTPL}^t &= \mathbb{E}_z[\langle \pi^{t+1}(x^t), \ell^t\rangle] = \mathbb{E}_{\pi \sim p^{t+1}}\left[\langle \pi(x^t), \ell^t\rangle\right] \end{aligned}$$

We will construct a mapping $\mu_\pi : \mathbb{R}^{dK} \to \mathbb{R}^{dK}$ such that for any $z \in \mathbb{R}^{dK}$,

$$M(z + \phi^t_\pi) = M(\mu_\pi(z) + \phi^{t-1})$$

Notice that $\mu_\pi(z) = z + \phi^t_\pi - \phi^{t-1}$. Now,

$$\begin{aligned} p^t(\pi) &= \int_z \mathbf{1}[\pi = M(z + \phi^{t-1})] f(z) dz \\ &= \int_z \mathbf{1}[\pi = M(\mu_\pi(z) + \phi^{t-1})] f(\mu_\pi(z)) dz \\ &= \int_z \mathbf{1}[\pi = M(z + \phi^t_\pi)] f(\mu_\pi(z)) dz \end{aligned}$$

Now observe that for any $z \in \mathbb{R}^{dK}$:

$$\begin{aligned} f(\mu_\pi(z)) &= \exp\{-\epsilon\left(\|z + \phi^t_\pi - \phi^{t-1}\|_1 - \|z\|_1\right)\} f(z) \\ &\leq \exp\{-\epsilon\left(\|z + \phi^t_\pi - \phi^{t-1}\|_1 - \|z + \phi^t_\pi - \phi^{t-1}\|_1 - \|\phi^{t-1} - \phi^t_\pi\|_1\right)\} f(z) \\ &\leq \exp\{\epsilon \|\phi^t_\pi - \phi^{t-1}\|_1\} f(z) \\ &= \exp\{\epsilon \langle \pi(x^t), \ell^t\rangle\} f(z) \end{aligned}$$

Substituting in this bound, we have,

$$p^t(\pi) \leq \exp\{\epsilon\langle\pi(x^t), \ell^t\rangle\} \cdot p_\pi^{t+1}(\pi) \leq \exp\{\epsilon\langle\pi(x^t), \ell^t\rangle\} \cdot p^{t+1}(\pi)$$

Re-arranging and lower bounding $\exp\{-x\} \geq (1-x)$:

$$p^{t+1}(\pi) \geq \exp\{-\epsilon\langle\pi(x^t), \ell^t\rangle\} \cdot p^t(\pi) \geq (1 - \epsilon\langle\pi(x^t), \ell^t\rangle) \cdot p^t(\pi) \tag{22}$$

Using the definition of $\texttt{FTPL}^t$ and $\texttt{BTPL}^t$, this gives,

$$\begin{aligned}
\texttt{BTPL}^t &= \sum_\pi p^{t+1}(\pi)\langle\pi(x^t), \ell^t\rangle \geq \sum_\pi (1 - \epsilon\langle\pi(x^t), \ell^t\rangle) p^t(\pi)\langle\pi(x^t), \ell^t\rangle \\
&= \texttt{FTPL}^t - \epsilon \sum_\pi p^t(\pi)\langle\pi(x^t), \ell^t\rangle^2 \\
&= \texttt{FTPL}^t - \epsilon\mathbb{E}\left[\langle\pi(x^t), \ell^t\rangle^2\right]
\end{aligned}$$

We will finish the proof by showing that $p_\pi^{t+1}(\pi) \leq p^{t+1}(\pi)$ for all $\pi \in \Pi$. For succinctness we drop the dependence on t. Notice that for any other policy $\pi' \neq \pi$

$$\mathcal{L}(\pi, z + \phi_\pi^t) \leq \mathcal{L}(\pi', z + \phi_\pi^t) \Rightarrow \mathcal{L}(\pi, z + \phi^t) \leq \mathcal{L}(\pi', z + \phi^t).$$

And similarly for strict inequalities. This follows since the loss of π remains unchanged, but the loss of π' can only go up, since $\ell_\pi^t(j) \leq \ell^t(j)$ (as losses are non-negative). For simplicity assume that π always wins in case of ties, though the argument goes through if we assume a deterministic tie-breaking rule based on some global ordering of policies. Thus,

$$p^{t+1}(\pi) = \mathbf{P}\left[\bigcap_{\pi'} \mathcal{L}(\pi, z + \phi^t) \leq \mathcal{L}(\pi', z + \phi^t)\right] \leq \mathbf{P}\left[\bigcap_{\pi'} \mathcal{L}(\pi, z + \phi_\pi^t) \leq \mathcal{L}(\pi', z + \phi_\pi^t)\right] = p_\pi^{t+1}(\pi)$$

as claimed. ■

B.4. Bounding Stability: Small Separator Setting

Finally, we prove the third claim in Theorem 2. This involves a new stability bound for the small separator setting.

Lemma 11 (Stability for small separator)**.** *For any $t \in [T]$ and any sequence $y^{1:t}$ of contexts $x^{1:t}$ and losses $f^{1:t}$ with $f^i : \{0,1\}^K \to \mathbb{R}^K$, the stability of* CONTEXT-FTPL(ϵ)*, when X is a separator, is upper bounded by:*

$$\mathbb{E}_{\{z\}}\left[f^t(\pi^t(x^t)) - f^t(\pi^{t+1}(x^t))\right] \leq 4\epsilon K d \cdot \|f^t\|_*^2$$

Proof. By the definition of $\|f^t\|_*$:

$$\mathbb{E}_{\{z\}}\left[f^t(\pi^t(x^t)) - f^t(\pi^{t+1}(x^t))\right] \leq 2\|f^t\|_* \Pr\left[\pi^t(x^t) \neq \pi^{t+1}(x^t)\right] \leq 2\|f^t\|_* \Pr[\pi^t \neq \pi^{t+1}]$$

Since X is a separator, $\pi^t \neq \pi^{t+1}$ if and only if there exists a context $x \in X$, such that $\pi^t(x) \neq \pi^{t+1}(x)$. Otherwise the two policies are identical. Thus we have by two applications of the union bound:

$$\begin{aligned}
\Pr[\pi^t \neq \pi^{t+1}] &\leq \sum_{x \in X} \Pr[\pi^t(x) \neq \pi^{t+1}(x)] \\
&\leq \sum_{x \in X} \sum_{j \in K} \left(\Pr[j \in \pi^t(x), j \notin \pi^{t+1}(x)] + \Pr[j \notin \pi^t(x), j \in \pi^{t+1}(x)]\right)
\end{aligned}$$

We bound the probability $\Pr[j \in \pi^t(x), j \notin \pi^{t+1}(x)]$. We condition on all random variables of $\{z\}$ except for the random variable $\ell_x(j)$, i.e. the random loss placed at coordinate j on the sample associated with context x. Denote the event corresponding to an assignment of all these other random variables as $\mathcal{E}_{-xj}$. Let ℓ_{xj} denote a loss vector which is $\ell_x(j)$ on the j-th coordinate and zero otherwise. Also let:

$$\Phi(\pi) = \sum_{\tau=1}^{t-1} f^\tau(\pi(x^\tau)) + \sum_{x' \neq x} \langle\pi(x'), \ell_{x'}\rangle + \langle\pi(x), \ell_x - \ell_{xj}\rangle \tag{23}$$

Let $\pi^* = \operatorname{argmin}_{\pi\in\Pi: j\in\pi(x)}\Phi(\pi)$ and $\tilde{\pi} = \min_{\pi\in\Pi: j\notin\pi(x)}\Phi(\pi)$. The event that $\{j \in \pi^t(x)\}$ happens only if:

$$\Phi(\pi^*) + \ell_x(j) \leq \Phi(\tilde{\pi}) \tag{24}$$

Let and $\nu = \Phi(\tilde{\pi}) - \Phi(\pi^*)$. Thus $j \in \pi^t(x)$ only if:

$$\ell_x(j) \leq \nu \tag{25}$$

Now if:

$$\ell_x(j) < \nu - 2\|f^t\|_* \tag{26}$$

then it is easy to see that $\{j \in \pi^{t+1}(x)\}$, since an extra loss of $f^t(a) \leq \|f^t\|_*$ cannot push j out of the optimal solution. More elaborately, for any other policy $\pi \in \Pi$, such that $j \notin \pi(x)$, the loss of π^* including time-step t is bounded as:

$$\begin{aligned}\Phi(\pi^*) + \ell_x(j) + f^t(\pi^*(x^t)) &< \Phi(\pi) - 2\|f^t\|_* + f^t(\pi^*(x^t)) \\ &< \Phi(\pi) - \|f^t\|_* \\ &< \Phi(\pi) + f^t(\pi(x^t))\end{aligned}$$

Thus any policy π, such that $j \notin \pi(x)$ is suboptimal after seeing the loss at time-step t. Thus

$$\Pr[j \in \pi^t(x), j \notin \pi^{t+1}(x) \mid \mathcal{E}_{-xj}] \leq \Pr[\ell_x(j) \in [\nu - 2\|f^t\|_*, \nu] \mid \mathcal{E}_{-xj}]$$

Since all other random variables are independent of $\ell_x(j)$ and $\ell_x(j)$ is a Laplacian with parameter ϵ:

$$\begin{aligned}\Pr[\ell_x(j) \in [\nu - 2\|f^t\|_*, \nu] \mid \mathcal{E}_{-xj}] &= \Pr[\ell_x(j) \in [\nu - 2\|f^t\|_*, \nu]] \\ &= \frac{\epsilon}{2}\int_{\nu-2\|f^t\|_*}^{\nu} e^{-\epsilon|z|}dz \leq \frac{\epsilon}{2}\int_{\nu-2\|f^t\|_*}^{\nu} dz \leq \epsilon\|f^t\|_*\end{aligned}$$

Similarly it follows that that: $\Pr[j \notin \pi^t(x), j \in \pi^{t+1}(x)] \leq \epsilon\|f^t\|_*$. To sum we get that:

$$\mathbb{E}_{\{z\}}\left[f^t(\pi^t(x^t)) - f^t(\pi^{t+1}(x^t))\right] \leq 2\|f^t\|_* \Pr\left[\pi^t \neq \pi^{t+1}\right] \leq 4\epsilon K d \cdot \|f^t\|_*^2$$

■

C. Omitted Proofs from Section 4

C.1. Proof of Theorem 3: Transductive Setting

Consider the expected loss of the bandit algorithm at time-step t, conditional on $\mathcal{H}^{t-1}$:

$$\mathbb{E}[\langle \pi^t(x^t), \ell^t\rangle \mid \mathcal{H}^{t-1}] = \sum_{j=1}^{K} q^t(j)\cdot\ell^t(j) \leq \sum_{j=1}^{K} q^t(j)\cdot\mathbb{E}\left[\hat{\ell}^t(j) \mid \mathcal{H}^{t-1}\right] + \sum_{j=1}^{K}\ell^t(j)q^t(j)\cdot(1-q^t(j))^L \tag{27}$$

As was observed by (Neu & Bartók, 2013), the second quantity can be upper bounded by $\frac{K}{eL}\|\ell^t\|_*$, since $q(1-q)^L \leq qe^{-Lq} \leq \frac{1}{eL}$.

Now observe that: $\sum_{j\in K} q^t(j)\cdot\mathbb{E}\left[\hat{\ell}^t(j) \mid \mathcal{H}^{t-1}\right]$ is the expected loss of the full feedback algorithm on the sequence of losses it observed and conditional on the history of play. By the regret bound of CONTEXT-FTPL(X,ϵ), given in case 2 of Theorem 2, we have that for any policy π^*:

$$\mathbb{E}\left[\sum_{t=1}^{T}\sum_{j=1}^{K} q^t(j)\cdot\hat{\ell}^t(j)\right] \leq \mathbb{E}\left[\sum_{t=1}^{T}\langle\pi^*(x^t),\hat{\ell}^t\rangle\right] + \epsilon\mathbb{E}\left[\sum_{t=1}^{T}\sum_{\pi\in\Pi}p^t(\pi)\langle\pi(x^t),\hat{\ell}^t\rangle^2\right] + \frac{10}{\epsilon}\sqrt{dm}\log(N)$$

Using the fact that expected estimates $\hat{\ell}$ are upper bounded by true losses:

$$\mathbb{E}\left[\sum_{t=1}^{T}\sum_{j=1}^{K} q^t(j)\hat{\ell}^t(j)\right] \leq \min_{\pi^*\in\Pi}\mathbb{E}\left[\sum_{t=1}^{T}\langle\pi^*(x^t),\ell^t\rangle\right] + \epsilon\mathbb{E}\left[\sum_{t=1}^{T}\sum_{\pi\in\Pi}p^t(\pi)\langle\pi(x^t),\hat{\ell}^t\rangle^2\right] + \frac{10}{\epsilon}\sqrt{dm}\log(N)$$

Combining the two upper bounds, we get that the expected regret of the bandit algorithm is upper bounded by:

$$\textsc{Regret} \leq \epsilon\mathbb{E}\left[\sum_{t=1}^{T}\sum_{\pi\in\Pi}p^t(\pi)\langle\pi(x^t),\hat{\ell}^t\rangle^2\right] + \frac{10}{\epsilon}\sqrt{dm}\log(N) + \frac{K}{eL}\sum_{t=1}^{T}\mathbb{E}\left[\|\ell^t\|_*\right]$$

Now observe that, by a simple norm inequality and re-grouping:

$$\sum_{\pi\in\Pi}p^t(\pi)\langle\pi(x^t),\hat{\ell}^t\rangle^2 = \sum_{\pi\in\Pi}p^t(\pi)\left(\sum_{j\in\pi(x^t)}\hat{\ell}^t(j)\right)^2 \leq m\sum_{\pi\in\Pi}p^t(\pi)\sum_{j\in\pi(x^t)}\hat{\ell}^t(j)^2 = m\sum_{j\in[K]}q^t(j)\hat{\ell}^t(j)^2$$

Thus we get:

$$\textsc{Regret} \leq \epsilon m\sum_{t=1}^{T}\mathbb{E}\left[\sum_{j\in[K]}q^t(j)\hat{\ell}^t(j)^2\right] + \frac{10}{\epsilon}\sqrt{dm}\log(N) + \frac{K}{eL}\sum_{t=1}^{T}\mathbb{E}\left[\|\ell^t\|_*\right]$$

Now we bound each of the terms in the first summation, conditional on any history of play:

$$\sum_{j\in[K]}q^t(j)\mathbb{E}\left[\hat{\ell}^t(j)^2 \mid \mathcal{H}^{t-1}\right] = \sum_{j\in[K]}q^t(j)q^t(j)\ell^t(j)^2\mathbb{E}\left[J^t(j)^2 \mid \mathcal{H}^{t-1}, j\in\pi^t(x^t)\right]$$

Each $J^t(j)$ conditional on $\mathcal{H}^{t-1}$ and $j \in \pi^t(x^t)$ is distributed according to a geometric distribution with mean $q^t(j)$ truncated at L. Hence, it is stochastically dominated by a geometric distribution with mean $q^t(j)$. By known properties, if X is a geometrically distributed random variable with mean q, then $\mathbb{E}[X^2] = Var(X)+(\mathbb{E}[X])^2 = \frac{1-q}{q^2}+\frac{1}{q^2} = \frac{2-q}{q^2} \leq \frac{2}{q^2}$. Thus we have:

$$\sum_{j\in[K]}q^t(j)\mathbb{E}\left[\hat{\ell}^t(j)^2 \mid \mathcal{H}^{t-1}\right] \leq \sum_{j\in[K]}q^t(j)^2\ell^t(j)^2\frac{2}{q^t(j)^2} = 2\sum_{j=1}^{K}\ell^t(j)^2 \leq 2K\|\ell^t\|_\infty^2$$

Combining all the above we get the theorem.

C.2. Proof of Theorem 3: Small Separator Setting

Consider the expected loss of the bandit algorithm at time-step t, conditional on $\mathcal{H}^{t-1}$:

$$\mathbb{E}[\langle\pi^t(x^t),\ell^t\rangle \mid \mathcal{H}^{t-1}] = \sum_{j=1}^{K}q^t(j)\cdot\ell^t(j) \leq \sum_{j=1}^{K}q^t(j)\cdot\mathbb{E}\left[\hat{\ell}^t(j) \mid \mathcal{H}^{t-1}\right] + \sum_{j=1}^{K}\ell^t(j)q^t(j)\cdot(1-q^t(j))^L \qquad (28)$$

As was observed by (Neu & Bartók, 2013), the second quantity can be upper bounded by $\frac{K}{eL}\|\ell^t\|_*$, since $q(1-q)^L \leq qe^{-Lq} \leq \frac{1}{eL}$.

Now observe that: $\sum_{j\in K}q^t(j)\cdot\mathbb{E}\left[\hat{\ell}^t(j) \mid \mathcal{H}^{t-1}\right]$ is the expected loss of the full feedback algorithm on the sequence of losses it observed and conditional on the history of play. By the regret bound of $\textsc{Context-FTPL}(X,\epsilon)$, given in case 3 of Theorem 2, we have that for any policy π^*:

$$\begin{aligned}\mathbb{E}\left[\sum_{t=1}^{T}\sum_{j=1}^{K}q^t(j)\cdot\hat{\ell}^t(j)\right] &\leq \mathbb{E}\left[\sum_{t=1}^{T}\langle\pi^*(x^t),\hat{\ell}^t\rangle\right] + 4\epsilon Kd\cdot\sum_{t=1}^{T}\mathbb{E}\left[\|\hat{f}^t\|_*^2\right] + \frac{10}{\epsilon}\sqrt{dm}\log(N)\\ &\leq \sum_{t=1}^{T}\langle\pi^*(x^t),\hat{\ell}^t\rangle + 4\epsilon Kd\cdot\sum_{t=1}^{T}\mathbb{E}\left[\|\hat{\ell}^t\|_1^2\right] + \frac{10}{\epsilon}\sqrt{dm}\log(N)\end{aligned}$$

Using the fact that expected estimates $\hat{\ell}$ are upper bounded by true losses:

$$\mathbb{E}\left[\sum_{t=1}^{T}\sum_{j=1}^{K}q^t(j)\hat{\ell}^t(j)\right] \leq \min_{\pi^*\in\Pi}\mathbb{E}\left[\sum_{t=1}^{T}\langle\pi^*(x^t),\hat{\ell}^t\rangle\right] + 4\epsilon Kd\cdot\sum_{t=1}^{T}\mathbb{E}\left[\|\hat{\ell}^t\|_1^2\right] + \frac{10}{\epsilon}\sqrt{dm}\log(N)$$

Combining the two upper bounds, we get that the expected regret of the semi-bandit algorithm is upper bounded by:

$$\textsc{Regret} \leq \ 4\epsilon K d \cdot \sum_{t=1}^{T} \mathbb{E}\left[\|\hat{\ell}^t\|_1^2\right] + \frac{10}{\epsilon}\sqrt{dm}\log(N) + \frac{K}{eL}\sum_{t=1}^{T}\mathbb{E}\left[\|\ell^t\|_*\right]$$

Now we bound each of the terms in the first summation, conditional on any history of play:

$$\mathbb{E}\left[\|\hat{\ell}^t\|_1^2 \mid \mathcal{H}^{t-1}\right] \leq m\mathbb{E}\left[\|\hat{\ell}^t\|_2^2\right] = m\sum_{j\in[K]}\mathbb{E}\left[\hat{\ell}^t(j)^2\right] = m\sum_{j\in[K]} q^t(j)\ell^t(j)^2\mathbb{E}\left[J^t(j)^2 \mid \mathcal{H}^{t-1}, j\in\pi^t(x^t)\right]$$

Each $J^t(j)$ conditional on $\mathcal{H}^{t-1}$ and $j \in \pi^t(x^t)$ is distributed according to a geometric distribution with mean $q^t(j)$ truncated at L. Hence, it is stochastically dominated by a geometric distribution with mean $q^t(j)$. By known properties, if X is a geometrically distributed random variable with mean q, then $\mathbb{E}[X^2] = Var(X)+(\mathbb{E}[X])^2 = \frac{1-q}{q^2}+\frac{1}{q^2} = \frac{2-q}{q^2} \leq \frac{2}{q^2}$. Moreover, trivially $\mathbb{E}[X^2] \leq L^2$, since X is truncated at L. Thus we have:

$$\mathbb{E}\left[\|\hat{\ell}^t\|_1^2 \mid \mathcal{H}^{t-1}\right] \leq m\sum_{j\in[K]} q^t(j)\ell^t(j)^2\min\left\{\frac{2}{q^t(j)^2}, L^2\right\} \leq m\|\ell^t\|_*^2\sum_{j\in[K]}\min\left\{\frac{2}{q^t(j)}, q^t(j)L^2\right\}$$

Now observe that: $\min\left\{\frac{2}{q^t(j)}, q^t(j)L^2\right\} \leq 2L$, since either $\frac{1}{q^t(j)} \leq L$ or otherwise, $q^t(j)L^2 \leq \frac{1}{L}L \leq L$. Thus we get:

$$\mathbb{E}\left[\|\hat{\ell}^t\|_1^2 \mid \mathcal{H}^{t-1}\right] \leq 2LKm\|\ell^t\|_*^2$$

Combining all the above we get the theorem.

D. Omitted Proofs from Section 5

Proof of Lemma 5. Oracle $\tilde{M}$ must compute the best sequence of policies $\pi^1, \ldots, \pi^T$, such that $\pi^t \neq \pi^{t-1}$ at most k times. Let $R(t,q)$ denote the loss of the optimal sequence of policies up to time-step t and with at most q switches. Then it is easy to see that:

$$R(t,q) = \min_{\tau\leq t} R(\tau, q-1) + \mathcal{L}\left(M(y^{\tau+1:t}), y^{\tau+1:t}\right), \tag{29}$$

i.e. compute the best sequence of policies up till some time step $\tau \leq t$ with at most $q-1$ switches and then augment it with the optimal fixed policy for the period $(\tau+1, t)$. Then take the best over possible times $\tau \leq t$.

This can be implemented by first invoking oracle M for every possible period $[\tau_1, \tau_2]$. Then filling up iteratively all the entries $R(t,q)$. For $q=0$, the problem $R(t,0)$ corresponds to exactly the original oracle problem M, hence for each t, we can solve the problem $R(t,0)$. Computing all values of $R(t,q)$ then takes time Tk in total. ■

E. Omitted Proofs from Section 6

E.1. Proof of Theorem 6

Similar to the analysis in Section 3, the proof of the Theorem is broken apart in two main Lemmas. The first lemma is an analogue of Theorem 1 for algorithms that use a predictor. This lemma can be phrased in the general online learning setting analyzed in Section 2. The second Lemma is an analogue of our additive stability Lemma 9.

Let

$$\rho^t = M(\{z\} \cup y^{1:t}) \tag{30}$$

denote the policy that would have been played at time-step t if the predictor was equal to the actual loss vector that occured at time-step t. Moreover, for succinctness we will denote with $a^t = \pi^t(x^t)$ and with $b^t = \rho^t(x^t)$.

Lemma 12 (Follow vs Be the Leader with Predictors)**.** *The regret of a player under the optimistic FTPL and with respect to any $\pi^* \in \Pi$ is upper bounded by:*

$$\textsc{Regret} \leq \sum_{t=1}^{T} \mathbb{E}\left[\Delta Q^t(a^t) - \Delta Q^t(b^t)\right] + \mathbb{E}[\textsc{Error}] \tag{31}$$

where $\Delta Q^t(a) = f^t(a) - Q^t(a)$ and $\textsc{Error} = \max_{\pi\in\Pi} \sum_{x\in X} \langle \pi(x), \ell_x\rangle - \min_{\pi\in\Pi} \sum_{x\in X} \langle \pi(x), \ell_x\rangle$.

Proof. Consider the augmented sequence $(x^1, Q^1), (x^1, f^1 - Q^1), (x^2, Q^2), (x^2, f^2 - Q^2), \ldots$, where each observation (x^t, f^t) is replaced by two observations (x^t, Q^t) followed by $(x^t, f^t - Q^t)$. Observe that by linearity of the objective, the two observations cancell out each other at the end, to give the same effect as a single observation of (x^t, f^t). Moreover, the leader after observing (x^t, Q^t) is equal to a^t, whilst after observing $(x^t, f^t - Q^t)$ is equal to b^t. Thus by applying Lemma 7 to this augmented sequence we get:

$$\begin{aligned}\sum_{t=1}^{T} \left(Q^t(a^t) + f^t(b^t) - Q^t(b^t)\right) \leq & \sum_{t=1}^{T} \left(Q^t(\pi^*(x^t)) + f^t(\pi^*(x^t)) - Q^t(\pi^*(x^t))\right) + \textsc{Error} \\ = & \sum_{t=1}^{T} f^t(\pi^*(x^t)) + \textsc{Error}\end{aligned}$$

Let $\texttt{BTPL}^t_Q = Q^t(a^t) + f^t(b^t) - Q^t(b^t)$ and $\texttt{FTPL}^t = f^t(a^t)$. Then, observe that:

$$\texttt{FTPL}^t - \texttt{BTPL}^t_Q = f^t(a^t) - Q^t(a^t) - (f^t(b^t) - Q^t(b^t)) = \Delta Q^t(a^t) - \Delta Q^t(b^t) \tag{32}$$

Combining the two properties we get that for any policy π^*:

$$\begin{aligned}\sum_{t=1}^{T} \texttt{FTPL}^t \leq & \sum_{t=1}^{T} \left(\Delta Q^t(a^t) - \Delta Q^t(b^t)\right) + \sum_{t=1}^{T} \texttt{BTPL}^t_Q \\ \leq & \sum_{t=1}^{T} \left(\Delta Q^t(a^t) - \Delta Q^t(b^t)\right) + \sum_{t=1}^{T} f^t(\pi^*(x^t)) + \textsc{Error}\end{aligned}$$

Re-arranging and taking expectation concludes the proof. ∎

Lemma 13 (Stability with Predictors)**.** *In the transductive setting:*

$$\mathbb{E}\left[\Delta Q^t(a^t) - \Delta Q^t(b^t)\right] \leq 4\epsilon K \|f^t - Q^t\|_*^2 \tag{33}$$

In the small separator setting:

$$\mathbb{E}\left[\Delta Q^t(a^t) - \Delta Q^t(b^t)\right] \leq 4\epsilon K d \|f^t - Q^t\|_*^2 \tag{34}$$

Proof. We prove the first part of the Lemma. The second follows along identical arguments. By the definition of $\|f^t - Q^t\|_* = \|\Delta Q^t\|_* = \max_{a\in\mathcal{A}} |\Delta Q^t(a)|$, we have:

$$\mathbb{E}_{\{z\}}\left[\Delta Q^t(a^t) - \Delta Q^t(b^t)\right] \leq 2\|\Delta Q^t\|_* \Pr\left[a^t \neq b^t\right]$$

Now observe that:

$$\Pr[a^t \neq b^t] \leq \sum_{j\in K} \left(\Pr[j \in a^t, j \notin b^t] + \Pr[j \notin a^t, j \in b^t]\right)$$

We bound the probability $\Pr[j \in a^t, j \notin b^t]$. We condition on all random variables of $\{z\}$ except for the random variable $\ell_{x^t}(j)$, i.e. the random loss placed at coordinate j on the sample associated with context x^t. Denote the event corresponding

to an assignment of all these other random variables as $\mathcal{E}_{-x^t j}$. Let $\ell_{x^t j}$ denote a loss vector which is $\ell_{x^t}(j)$ on the j-th coordinate and zero otherwise. Also let:

$$\Phi(\pi) = \sum_{\tau=1}^{t-1} f^\tau(\pi(x^\tau)) + Q^t(\pi(x^t)) + \sum_{x \in X - \{x^t\}} \langle \pi(x), \ell_x \rangle + \langle \pi(x^t), \ell_{x^t} - \ell_{x^t j} \rangle \tag{35}$$

Let $\pi^* = \operatorname{argmin}_{\pi \in \Pi: j \in \pi(x^t)} \Phi(\pi)$ and $\tilde{\pi} = \min_{\pi \in \Pi: j \notin \pi(x^t)} \Phi(\pi)$. The event that $\{j \in a^t\}$ happens only if:

$$\Phi(\pi^*) + \ell_{x^t}(j) \leq \Phi(\tilde{\pi}) \tag{36}$$

Let and $\nu = \Phi(\tilde{\pi}) - \Phi(\pi^*)$. Thus $j \in a^t$ only if:

$$\ell_{x^t}(j) \leq \nu \tag{37}$$

Now if:

$$\ell_{x^t}(j) < \nu - 2\|\Delta Q^t\|_* \tag{38}$$

then it is easy to see that $\{j \in b^t\}$, since an extra loss of $f^t(a) - Q^t(a) \leq \|\Delta Q^t\|_*$ cannot push j out of the optimal solution. More elaborately, for any other policy $\pi \in \Pi$, such that $j \notin \pi(x^t)$, the loss of π^* including time-step t is bounded as:

$$\begin{aligned}
\Phi(\pi^*) + \ell_{x^t}(j) + f^t(\pi^*(x^t)) - Q^t(\pi^*(x^t)) <&\ \Phi(\pi) - 2\|\Delta Q^t\|_* + f^t(\pi^*(x^t)) - Q^t(\pi^*(x^t)) \\
<&\ \Phi(\pi) - \|\Delta Q\|_* \\
<&\ \Phi(\pi) + f^t(\pi(x^t)) - Q^t(\pi(x^t))
\end{aligned}$$

Thus any policy π, such that $j \notin \pi(x^t)$ is suboptimal after seeing the loss at time-step t. Thus

$$\Pr[j \in a^t, j \notin b^t \mid \mathcal{E}_{-x^t j}] \leq \Pr[\ell_{x^t}(j) \in [\nu - 2\|\Delta Q^t\|_*, \nu] \mid \mathcal{E}_{-x^t j}]$$

Since all other random variables are independent of $\ell_{x^t}(j)$ and $\ell_{x^t}(j)$ is a Laplacian with parameter ϵ:

$$\begin{aligned}
\Pr[\ell_{x^t}(j) \in [\nu - 2\|\Delta Q^t\|_*, \nu] \mid \mathcal{E}_{-x^t j}] =&\ \Pr[\ell_{x^t}(j) \in [\nu - 2\|\Delta Q^t\|_*, \nu]] \\
=&\ \frac{\epsilon}{2} \int_{\nu - 2\|\Delta Q^t\|_*}^{\nu} e^{-\epsilon|z|} dz \leq \frac{\epsilon}{2} \int_{\nu - 2\|\Delta Q^t\|_*}^{\nu} dz \leq \epsilon \cdot \|\Delta Q^t\|_*
\end{aligned}$$

Similarly it follows that that: $\Pr[j \notin \pi^t(x^t) \text{ and } j \in \pi^{t+1}(x^t)] \leq \epsilon \cdot \|\Delta Q^t\|_*$. To sum we get that:

$$\mathbb{E}_{\{z\}} \left[\Delta Q^t(a^t) - \Delta Q^t(b^t)\right] \leq\ 2\|\Delta Q^t\|_* \Pr\left[\pi^t(x^t) \neq \pi^{t+1}(x^t)\right] \leq 4\epsilon K \|\Delta Q^t\|_*^2$$

■

The expected error term is identical to the expected error that we upper bounded in Lemma 8, hence the same bound carries over. Combining the above Lemmas with this observation, yields Theorem 6.

F. Infinite Policy Classes

In this section we focus on the contextual experts problem but consider infinite policy classes. Recall that in this setting, in each round t, the adversary picks a context $x^t \in \mathcal{X}$ and a loss function $\ell^t \in \mathbb{R}^K_{\geq 0}$, the learner, upon seeing the context x^t, chooses an action $a^t \in [K]$, and then suffers loss $\ell^t(a^t)$. We showed that as a simple consequence of Theorem 2, that when competing with a set of policies $\Pi \subset (\mathcal{X} \to [K])$ with $|\Pi| = N$ and against an adaptive adversary, CONTEXT-FTPL has regret at most $O(d^{1/4}\sqrt{T \log(N)})$ in the transductive setting and regret at most $O(d^{3/4}\sqrt{KT \log(N)})$ in the non-transductive setting with small separator.

Here we consider the situation where the policy class Π is infinite in size, but has small Natarajan dimension, which generalizes VC-dimension to multiclass problems. Specifically, we prove two results in this section: First we show that in the transductive case, CONTEXT-FTPL can achieve low regret relative to a policy class with bounded Natarajan dimension. Then we show that in the non-transductive case, it is hard in an information-theoretic sense to achieve sublinear regret

relative to a policy class with constant Natarajan dimension. Together, these results show that finite Natarajan or VC dimension is sufficient for sublinear regret in the transductive setting, but it is *insufficient* for sublinear regret in the fully online setting.

Before proceeding with the two results, we must introduce the notion of Natarajan dimension, which requires some notation. For a class of functions $\mathcal{F}$ from $\mathcal{X} \to [K]$ and for a sequence $X = (x_1, \ldots, x_n) \in \mathcal{X}^n$, define $\mathcal{F}_X = \{(f(x_1), \ldots, f(x_n)) \in [K]^n : f \in \mathcal{F}\}$ be the restriction of the functions to the points. Let Ψ be a family of mappings from $[K] \to \{0, 1, \star\}$ where $\star$ is any new symbol that we use to denote irrelevant classes. Let $\bar{\psi} = (\psi_1, \ldots, \psi_n) \in \Psi^n$ be a fixed sequence of such mappings and for a sequence $(s_1, \ldots, s_n) \in [K]^N$ define $\bar{\psi}(s) = (\psi_1(s_1), \ldots, \psi_1(s_n)) \in \{0, 1, \star\}^n$. We say a sequence $X \in \mathcal{X}^n$ is Ψ-shattered by $\mathcal{F}$ if there exists $\bar{\psi} \in \Psi^n$ such that:

$$\{0, 1\}^n \subseteq \{\bar{\psi}(s) : s \in \mathcal{F}_X\}$$

The Ψ-dimension of a function class $\mathcal{F}$ is the largest n such that there exist a sequence $X \in \mathcal{X}^n$ that is Ψ-shattered by $\mathcal{F}$. Notice that if $K = 2$ and Ψ contains only the identity map, then the Ψ-dimension is exactly the VC dimension.

The **Natarajan dimension** is the Ψ dimension for the class $\Psi_N = \{\psi_{N,i,j}, i, j \in [K], j \neq i\}$ where $\psi_{N,i,j}(a) = 1$ if $a = i$, $\psi_{N,i,j}(a) = 0$ if $a = j$ and $\psi_{N,i,j}(a) = \star$ otherwise. Notice that Natarajan dimension is a strict generalization of VC-dimension as Ψ_N contains only the identity map if $K = 2$. Thus our result also applies to VC-classes in the two-action case. The main property we will use about function classes with bounded Natarajan Dimension is the following analog of the Sauer-Shelah Lemma:

Lemma 14 (Sauer-Shelah for Natarajan Dimension (Haussler & Long, 1995; Ben-David et al., 1995))**.** *Suppose that $\mathcal{F}$ has Ψ_N dimension at most ν. Then for any set $X \in \mathcal{X}^n$, we have:*

$$|\mathcal{F}_X| \leq \left(\frac{ne(K+1)^2}{2\nu}\right)^{\nu}$$

Our positive result for transductive learning with a Natarajan class is the following regret bound for CONTEXT-FTPL,

Corollary 15. *Consider running* CONTEXT-FTPL(X, ϵ) *in the transductive contextual experts setting with $|X| = d$ and with a policy class Π with Natarajan dimension at most ν. Then the algorithm achieves regret against an adaptive and adversarially chosen sequence of contexts and loss functions,*

$$\epsilon \sum_{t=1}^{T} \mathbb{E}[\langle \pi^t(x^t), \ell^t \rangle^2] + \frac{10}{\epsilon} \sqrt{d\nu \log(K) \log\left(\frac{de(K+1)^2}{2\nu}\right)}.$$

When ϵ is set optimally and losses are in $[0, 1]^K$, this is $O((d\nu \log(K) \log(dK/\nu))^{1/4}\sqrt{T})$.

Proof. The result is a consequence of the second clause of Theorem 2, using the additional fact that any sequence of contexts $X = (x_1, \ldots, x_d)$ induce a finite policy class $\Pi_X \subseteq [K]^d$. The fact that Π has Natarajan dimension at most ν means that $|\Pi_X| \leq \left(\frac{de(K+1)^2}{2\nu}\right)^{\nu}$ by Lemma 14. Therefore, once the d contexts are fixed, as they are in the transductive setting, we are back in the finite policy case and can apply Theorem 2 with N replaced by $|\Pi_X|$. ■

Thus we see that CONTEXT-FTPL has sublinear regret relative to policy classes with bounded Natarajan dimension, even against adaptive adversaries. The second result in this section shows that this result cannot be lifted to the non-transductive setting. Specifically, we prove the following theorem in the section, which shows that no algorithm, including inefficient ones, can achieve sublinear regret against a VC class in the non-transductive setting.

Theorem 16. *Consider an online binary classification problem in one dimension with $\mathcal{F} \subset [0, 1] \to \{0, 1\}$ denoting the set of all threshold functions. Then there is no learning algorithm that can guarantee $o(T)$ expected regret against an adaptive adversary. In particular, there exists a policy class of VC dimension one such that no learning algorithm can achieve sublinear regret against an adaptive adversary in the contextual experts problem.*

Proof. We define an adaptive adversary and argue that it ensures at least $1/2$ expected regret per round. While the adversary does not have access to the random coins of the learner, it can compute the probability that the learner would label any point

as $\{0, 1\}$. At round t, let $p_t(x)$ denote the probability that the learner would label a point $x \in [0, 1]$ as 1, and note that this quantity is conditioned on the entire history of interaction. At each round t, the adversary will have played a set of points X_t^+ with positive label and X_t^- with negative label and she will maintain the invariant that $\min_{x \in X_t^+} x > \max_{x \in X_t^-} x$ for all t. At every time t, the adversary will play context $x_t \in (\max_{x \in X_t^-} x, \min_{x \in X_t^+} x)$. The adversary, knowing the learning algorithm, will compute $p_t(x_t)$ and assign label $y_t = 1$ if $p_t(x) < 1/2$ and 0 otherwise. The adversary will then update the sets $X_{t+1}^+ \leftarrow X_t^+ \cup \{x_t\}$ if $y_t = 1$ and $X_{t+1}^+ \leftarrow X_t^+$ otherwise. X_{t+1}^- is updated analogously.

Clearly this sequence of contexts maintains the appropriate invariant for the adversary, namely there is always an interval between the positive and negative examples in which he can pick a context. This implies that on the sequence, there is a threshold $f^\star \in \mathcal{F}$ that perfectly classifies the points, so its cumulative reward is T. Moreover, by the choice of label selected by the adversary, the expected reward of the learner at round t is at most $1/2$, which means the cumulative expected reward of the learner is at most $T/2$. Thus the regret of the learner is at least $T/2$. ■