Supplementary Material of the Paper: Learning in Distributed Contextual Linear Bandits Without Sharing the Context

A Proofs and Remarks for Section 3: Contextual Linear Bandits with Known Context Distribution

Remark. We note that, to reduce the multi-context problem to a single context problem in the case of known context distribution, the straightforward approach that replaces the actual context realizations \( \{ X_{t,a} \} \) with the fixed set \( \{ \mathbb{E}[X_{t,a}] \} \) and uses this latter set as \( X \) in \( \Lambda \), does not work and can lead to linear regret in some cases. For instance consider the case where \( d = 1, A = \{ 1, 2 \} \). \( X_{t,a} \in \{ -1, 1 \} \forall a \in A, \theta_* = 1 \) and \( X_{t,1} \) takes the value 1 with probability 3/4 and -1 otherwise, while \( X_{t,2} \) takes the values 1, -1 with probability 1/2. Then clearly \( \mathbb{E}[X_{t,1}] > \mathbb{E}[X_{t,2}] \), however, choosing \( a_t = 1 \forall t \in [T] \) leads to \( \mathbb{E}[R_T] \geq T/8 \) since it holds that \( X_{t,1} = -1, X_{t,2} = 1 \) with probability 1/8.

Downlink Communication. Note that in our setup we assume that the central learner does not have any communication constraints when communicating with the distributed agents. Yet our algorithm makes frugal use of this ability: the central learner only sends the updated parameter vector \( \theta_t \). We can quantize \( \theta_t \) without performance loss if the downlink were also communication constrained using \( \approx 5d \) bits and an approach similar to the one in Algorithm 2 - yet we do not expand on this in this paper, as our focus is in minimizing uplink communication costs.

A.1 Proof of Theorem 1

Theorem 1. Algorithm 1 uses 1 bit per reward and 0 bits per context. Under Assumption 2 it achieves a regret \( R_T = O_{\mathcal{F}}(\Lambda) + O(\sqrt{T \log T}) \) with probability at least \( 1 - \frac{1}{T} \).

Proof. It is obvious that the agent only sends 1 bit to the central learner to represent \( r_t \), using SQ1, hence, the algorithm uses 0 bits per context and 1 bit per reward. We next bound the regret of our algorithm as following. The regret can be expressed as

\[
R_T = \sum_{t=1}^{T} \max_{a \in A} \langle X_{t,a}, \theta_* \rangle - \langle X_{t,a}, \theta_* \rangle
\]

\[
= \sum_{t=1}^{T} (\arg \max_{X_{t,a}} \langle X_{t,a}, \theta_* \rangle, \theta_* ) - (\arg \max_{X_{t,a}} \langle X_{t,a}, \bar{\theta}_t \rangle, \theta_* )
\]

\[
= \sum_{t=1}^{T} \left( \langle \arg \max_{X_{t,a}} \langle X_{t,a}, \theta_* \rangle, \theta_* \rangle - \mathbb{E}[\langle \arg \max_{X_{t,a}} \langle X_{t,a}, \bar{\theta}_t \rangle, \theta_* \rangle] \right)
\]

\[
- \langle \arg \max_{X_{t,a}} \langle X_{t,a}, \bar{\theta}_t \rangle, \theta_* \rangle - \mathbb{E}[\langle \arg \max_{X_{t,a}} \langle X_{t,a}, \bar{\theta}_t \rangle, \theta_* \rangle] \rangle
\]

\[
+ \langle \mathbb{E}[\langle \arg \max_{X_{t,a}} \langle X_{t,a}, \theta_* \rangle, \theta_* \rangle], \theta_* \rangle - \langle \mathbb{E}[\langle \arg \max_{X_{t,a}} \langle X_{t,a}, \bar{\theta}_t \rangle, \theta_* \rangle], \theta_* \rangle \rangle. \tag{15}
\]

To bound \( R_T \), we bound each of the three lines in the last expression. For the second term denoted by \( \Sigma_t = \sum_{i=1}^{T} \langle \arg \max_{X_{t,i,a}} \langle X_{t,i,a}, \bar{\theta}_t \rangle, \theta_* \rangle - \mathbb{E}[\langle \arg \max_{X_{t,i,a}} \langle X_{t,i,a}, \bar{\theta}_t \rangle, \theta_* \rangle] \rangle \), we have that

\[
\mathbb{E}[\Sigma_{t+1} | \Sigma_t] = \Sigma_t + \mathbb{E} \left[ \langle \arg \max_{X_{t,a}} \langle X_{t,a}, \bar{\theta}_t \rangle, \theta_* \rangle - \langle \mathbb{E}[\langle \arg \max_{X_{t,a}} \langle X_{t,a}, \bar{\theta}_t \rangle, \theta_* \rangle], \theta_* \rangle | \Sigma_t \right]
\]

\[
= \Sigma_t + \mathbb{E} \left[ \langle \arg \max_{X_{t,a}} \langle X_{t,a}, \bar{\theta}_t \rangle, \theta_* \rangle - \mathbb{E}[\langle \arg \max_{X_{t,a}} \langle X_{t,a}, \bar{\theta}_t \rangle, \theta_* \rangle | \Sigma_t] | \Sigma_t \right]
\]

\[
= \Sigma_t + \mathbb{E} \left[ \langle \mathbb{E}[\langle \arg \max_{X_{t,a}} \langle X_{t,a}, \bar{\theta}_t \rangle, \theta_* \rangle], \theta_* \rangle - \mathbb{E}[\langle \arg \max_{X_{t,a}} \langle X_{t,a}, \bar{\theta}_t \rangle, \theta_* \rangle | \Sigma_t] | \Sigma_t \right]
\]
We also have that
\[ |\Sigma_t - \Sigma_{t-1}| \leq \|\langle \arg\max_{X_t,a} \langle X_t,a, \hat{\theta}_t \rangle, \theta_* \rangle - \langle \mathbb{E}[\arg\max_{X_t,a} \langle X_t,a, \hat{\theta}_t \rangle], \theta_* \rangle \| \|\theta_*\| \]
\[ \leq \|\langle \arg\max_{X_t,a} \langle X_t,a, \hat{\theta}_t \rangle, \theta_* \rangle \| + \|\mathbb{E}[\arg\max_{X_t,a} \langle X_t,a, \hat{\theta}_t \rangle], \theta_* \rangle \|
\leq \| \arg\max_{X_t,a} \langle X_t,a, \hat{\theta}_t \rangle \| \|\theta_*\| + \|\mathbb{E}[\arg\max_{X_t,a} \langle X_t,a, \hat{\theta}_t \rangle] \| \|\theta_*\| \leq 2. \]  

(17)

Hence, \( \Sigma_t \) is a martingale with bounded difference. By Azuma–Hoeffding inequality [45], we have that \( \|\Sigma_T\| \leq C \sqrt{T \log T} \) with probability at least \( 1 - \frac{1}{2T} \). Similarly, the first line in (15) is a martingale with bounded difference, hence, the following holds with probability at least \( 1 - \frac{1}{2T} \)
\[ \sum_{t=1}^{T} \langle \arg\max_{X_t,a} \langle X_t,a, \theta_* \rangle, \theta_* \rangle - \langle \mathbb{E}[\arg\max_{X_t,a} \langle X_t,a, \theta_* \rangle], \theta_* \rangle \leq C \sqrt{T \log T}. \]

(18)

By substituting in (15) and using the union bound we get that the following holds with probability at least \( 1 - \frac{1}{T} \)
\[ R_T \leq C \sqrt{T \log T} + \sum_{t=1}^{T} \left( \langle \mathbb{E}[\arg\max_{X_t,a} \langle X_t,a, \theta_* \rangle], \theta_* \rangle - \langle \mathbb{E}[\arg\max_{X_t,a} \langle X_t,a, \hat{\theta}_t \rangle], \theta_* \rangle \right)
\]
\[ = C \sqrt{T \log T} + \sum_{t=1}^{T} (X^*(\theta_*), \theta_* - \langle X^*(\hat{\theta}_t), \theta_* \rangle
\]
\[ = C \sqrt{T \log T} + \sum_{t=1}^{T} (X^*(\theta_*), \theta_* - \langle X_t, \theta_* \rangle. \]  

(19)

We also have by definition of \( X^*(\theta_*) \) that for any given \( \theta, \theta_* \)
\[ \langle X^*(\theta_*), \theta_* \rangle = \mathbb{E}[\max_{X_t,a} \langle X_t,a, \theta_* \rangle]
\]
\[ \geq \mathbb{E}[\langle \arg\max_{X_t,a} \langle X_t,a, \theta_* \rangle \| \theta_* \rangle = \langle X^*(\theta), \theta_* \rangle. \]  

(20)

Hence, we have that \( \max_{X_t \subseteq X} \langle X, \theta_* \rangle = \langle X^*(\theta_*), \theta_* \rangle \). By substituting in (19), we get that
\[ R_T \leq C \sqrt{T \log T} + \sum_{t=1}^{T} \max_{X_t \subseteq X} \langle X_t, \theta_* \rangle - \langle X_t, \theta_* \rangle = C \sqrt{T \log T} + R_T(\Lambda), \]

(21)

where \( R_T(\Lambda) \) is the regret of the subroutine \( \Lambda \).

\[ \square \]

A.2 Proof of Corollary [1]

Corollary 1. Suppose we are given \( \tilde{X}^* \) that satisfies [12]. Then, there exists an algorithm \( \Lambda \) for which Algorithm [1] achieves \( R_T = O(d\sqrt{T} + \epsilon \sqrt{T}) \) with probability at least \( 1 - \frac{1}{T} \).

Proof. \( \Lambda \) assumes that the reward \( r_t \) is generated according to \( \langle \tilde{X}^*(\hat{\theta}_t), \theta_* \rangle + \eta_t \), while it is actually generated according to
\[ r_t = \langle X^*(\hat{\theta}_t), \theta_* \rangle + \eta_t = \langle \tilde{X}^*(\hat{\theta}_t), \theta_* \rangle + \eta_t + f(\hat{\theta}_t), \]

(22)

where \( f(\hat{\theta}_t) = \langle X^*(\hat{\theta}_t) - \tilde{X}^*(\hat{\theta}_t), \theta_* \rangle \). We have that
\[ |f(\hat{\theta}_t)| \leq \| X^*(\hat{\theta}_t) - \tilde{X}^*(\hat{\theta}_t) \| \|\theta_*\| \leq \epsilon. \]

(23)

Hence, the rewards follow a misspecified linear bandit model [24]. It was shown in [24] that for the single context case, there is an algorithm \( \Lambda \) that achieves \( R_T(\Lambda) = O(d\sqrt{T} + \epsilon T) \) with probability at least \( 1 - \frac{1}{T} \). The corollary follows from Theorem [1] by noting that \( R_T(\Lambda) \) is defined based on the true \( X^* \) as in [5].

\[ \square \]
B Proofs of Section 4: Contextual Linear Bandits with Unknown Context Distribution

B.1 Proof of Theorem 2

Theorem 2. Algorithm 2 satisfies that for all $t$: $X_t \in Q$, and $B_t \leq 1 + \log_2(2d + 1) + 5.03d$ bits. Under assumptions 1, it achieves a regret $R_T = O(d \sqrt{T \log T})$ with probability at least $1 - \frac{1}{T}$.

Proof. We start by proving some properties about the quantized values $\hat{X}_t, \hat{r}_t, \hat{X}_t^2$. We first note that by definition of SQ, we have that

$$m|\hat{X}_t - X_{t,a_t}| \leq 1. \quad (24)$$

Hence,

$$|(\hat{X}_t^2 - X_{t,a_t}^2)_j| = |(\hat{X}_t^2 - (\hat{X}_t + X_{t,a_t} - \hat{X}_t)^2)_j| = |(2(X_{t,a_t} - \hat{X}_t) \hat{X}_t + (X_{t,a_t} - \hat{X}_t)^2)_i| \leq 2|(X_{t,a_t} - \hat{X}_t)_i|(|\hat{X}_t|) + |(|(X_{t,a_t} - \hat{X}_t)^2)_i| \leq \frac{2}{m} + \frac{1}{m^2} \leq \frac{3}{m}. \quad (25)$$

We also have that

$$E[\hat{X}_t^D|X_{t,a_t}^2] = E[X_t^2 + \varepsilon_t^2|X_{t,a_t}] = E[E[X_t^2 + \varepsilon_t^2|X_{t,a_t}, \hat{X}_t]|X_{t,a_t}] \leq E[X_t^2 + \varepsilon_t^2|X_{t,a_t}, \hat{X}_t] = E[X_t^2] = X_{t,a_t}^2. \quad (26)$$

In summary, from this and the definition of SQ, we get that

$$m|\hat{X}_t - X_{t,a_t}| \leq 1, E[\hat{X}_t|X_{t,a_t}] = X_{t,a_t},$$

$$m|\hat{X}_t^D - X_{t,a_t}^2| \leq 3, E[\hat{X}_t^D|X_{t,a_t}^2] = X_{t,a_t}^2,$$

$$|\hat{r}_t - r_t| \leq 1, E[|\hat{r}_t|] = r_t \quad (27)$$

We next show that $X_t \in \text{dom}(h)$. By definition of SQ, we have that $X_t \in \mathbb{N}^d$. We also have that

$$\|X_t\|_1 = \sum_{i=1}^{d} m|X_{t,a_t}_i| \leq 1 + |\sqrt{d}|(X_{t,a_t}_i) \leq d + \sum_{i=1}^{d} |\sqrt{d}|(X_{t,a_t})_i \leq 2d. \quad (28)$$

Therefore, we have that $X_t \in Q = \text{dom}(h)$.

We next show that $X_t$ is the upper bound on the number of bits $B_t$. We have that $\hat{r}_t$ uses 1 bit, the sign vector $s_t$ uses $d$ bits, $\varepsilon_t^2$ uses $d$ bits and $X_t$ uses $\log(|Q|)$ bits. We bound $|Q|$ as follows. The number of non-negative solutions for the equation $\|a\|_1 = x$ for $a \in \mathbb{N}^d, x \in \mathbb{N}$ is $(d+x-1) \leq \binom{d+x}{d}$, hence,

$$|Q| \leq (2d + 1)^{\binom{3d}{d}} \leq (2d + 1)^{\frac{3d}{d}}. \quad (29)$$

Hence, we have that

$$B_t \leq 1 + \log(2d + 1) + (2 + \log(3e))d. \quad (30)$$

We next show the regret bound. We start by bounding the regret in iteration $t$ by the distance between $\theta_*, \hat{\theta}_{t-1}$. From step 7 of Algorithm 2, we have that $\langle X_{t,a_t}, \hat{\theta}_{t-1} \rangle \geq \langle X_{t,a_t}, \hat{\theta}_{t-1} \rangle \forall a \in \mathcal{A}$, hence, we have that

$$\max_{a \in \mathcal{A}} \langle X_{t,a}, \theta_* \rangle - \langle X_{t,a_t}, \theta_* \rangle \leq \max_{a \in \mathcal{A}} \langle X_{t,a} - X_{t,a_t}, \theta_* \rangle - \max_{a \in \mathcal{A}} \langle X_{t,a} - X_{t,a_t}, \hat{\theta}_{t-1} \rangle \leq \max_{a \in \mathcal{A}} \|X_{t,a} - X_{t,a_t}\| \|\theta_* - \hat{\theta}_{t-1}\| \leq 2\|\theta_* - \hat{\theta}_{t-1}\|. \quad (31)$$

We next bound the distance $\|\theta_* - \hat{\theta}_{t-1}\|$. Let us denote $e_t = \hat{X}_t - X_{t,a_t}, \hat{\eta}_t = \eta_t + (\hat{r}_t - r_t), E_t = X_{t,a_t}X_{t,a_t}^T - (V_t - V_{t-1})$. We have that

$$\|\theta_* - \hat{\theta}_{t-1}\| = \|\theta_* - V_t^{-1} \sum_{i=1}^{t} \hat{r}_i \hat{X}_i\| = \|\theta_* - V_t^{-1} \sum_{i=1}^{t} (X_{t,a_t}X_{t,a_t}^T \theta_*) + \hat{r}_i e_i + \hat{\eta}_t X_{t,a_t} + \hat{\eta}_te_i\|$$
\[
= \left\| V_t^{-1} \sum_{i=1}^t \left( E_i \theta_i^* + r_i e_i + \hat{\eta}_i X_{i,a_i} + \tilde{\eta}_i e_i \right) \right\|
\leq \left\| V_t^{-1} \right\| \left( \| \sum_{i=1}^t E_i \| + (|r_i| + |\eta_i|) \sum_{i=1}^t e_i \| + \| \sum_{i=1}^t \hat{\eta}_i X_{i,a_i} \| \right)
\leq \left\| V_t^{-1} \right\| \left( \| \sum_{i=1}^t E_i \| + (1 + |\eta_i|) \sum_{i=1}^t e_i \| + \| \sum_{i=1}^t \hat{\eta}_i X_{i,a_i} \|. \right.
\]  

We next bound each of the values in the last expression. As \( \hat{\eta}_i \) is subgaussian we have that with probability at least \( 1 - \frac{1}{\sqrt{T}} \), we have that \( |\hat{\eta}_i| \leq C \log T \| e_i \| \in [T] \). We also have that, using [27], \( S_t^i = \sum_{i=1}^t e_i \) is a martingale with bounded difference, hence, by Azuma–Hoeffding inequality, we get that with probability at least \( 1 - \frac{1}{\sqrt{T}} \) we have that \( |(S_t^i)_j| \leq \frac{C \sqrt{t \log(dT)}}{d} \); note that \( |(e_i)_j| \leq \frac{1}{\sqrt{d}} \). Hence, by the union bound we get that with probability at least \( 1 - \frac{1}{\sqrt{T}} \) we have that \( \| \sum_{i=1}^t e_i \| \leq C \sqrt{t \log(dT)} \). Similarly, conditioned on \( X_{1,a_1}, \ldots, X_{t,a_t} \), \( t_i = 1 \\sum_{i=1}^t \hat{\eta}_i X_{i,a_i} \) is a martingale with bounded difference, hence, with probability at least \( 1 - \frac{1}{\sqrt{T}} \) we have that \( \| \sum_{i=1}^t \hat{\eta}_i X_{i,a_i} \| \leq C \sqrt{\sum_{i=1}^t \| X_{i,a_i} \|^2 \log(dT)} \). Summing up, we get that with probability at least \( 1 - \frac{1}{\sqrt{T}} \)

\[
\| \theta_t - \hat{\theta}_t \| \leq \left\| V_t^{-1} \right\| \left( \| \sum_{i=1}^t E_i \| + C \sqrt{t \log(dT)} \right). 
\]  

It remains to bound \( \| V_t^{-1} \| \| \sum_{i=1}^t E_i \| \) which we do in the following by starting with \( \| \sum_{i=1}^t E_i \| \).

We have that

\[
E_i = X_{i,a_i} X_{i,a_i}^T - \hat{X}_i \hat{X}_i^T + \text{diag}(\hat{X}_i) - \text{diag}(X_{i,D}^D)
= \text{diag}(\hat{X}_i) - 2X_{i,a_i} e_i^T - e_i e_i^T \text{diag}(X_{i,D}^D)
= 2\text{diag}(X_{i,a_i} e_i^T) - 2X_{i,a_i} e_i^T - (e_i e_i^T - \text{diag}(e_i e_i^T)) \text{diag}(X_{i,D}^D) - X_{i,a_i}^2 .
\]  

Hence, we have that

\[
\| \sum_{i=1}^t E_i \| \leq 2 \left\| \sum_{i=1}^t \text{diag}(X_{i,a_i} e_i^T) \right\| + 2 \left\| \sum_{i=1}^t X_{i,a_i} e_i^T \right\|
+ \left\| \sum_{i=1}^t e_i e_i^T - \text{diag}(e_i e_i^T) \right\| + \left\| \sum_{i=1}^t \text{diag}(X_{i,D}^D) - X_{i,a_i}^2 \right\|.
\]  

We have that, using [27], conditioned on \( X_{1,a_1}, \ldots, X_{t,a_t} \), \( \sum_{i=1}^t \text{diag}(X_{i,a_i} e_i^T) \) is a martingale with bounded difference, hence, similar to what we did before using Azuma–Hoeffding inequality and the union bound we get that with probability at least \( 1 - \frac{1}{\sqrt{T}} \), we have that \( \| \sum_{i=1}^t \text{diag}(X_{i,a_i} e_i^T) \| \leq \frac{C \sqrt{t \log(dT)}}{d} \). Similarly, with probability at least \( 1 - \frac{1}{\sqrt{T}} \), we have that \( \| \sum_{i=1}^t \text{diag}(X_{i,D}^D) - X_{i,a_i}^2 \| \leq \frac{C \sqrt{t \log(dT)}}{d} \). We next turn to bounding \( \| \sum_{i=1}^t X_{i,a_i} e_i^T \| \).

Conditioned on \( X_{1,a_1}, \ldots, X_{t,a_t} \), we have that by Azuma–Hoeffding, with probability at least \( 1 - \frac{1}{\sqrt{T}} \), we have

\[
\left\| \sum_{i=1}^t X_{i,a_i} e_i^T \right\| \leq \frac{C \sqrt{t \log(dT)}}{d} \left\| \left[ \sum_{i=1}^t (X_{i,a_i})^2, \ldots, \sum_{i=1}^t (X_{i,a_i})^2 \right] \right\|
\]  

We notice that taking the absolute value of all elements of a matrix does not decrease its maximum eigenvalue, hence, by the union bound we have that with probability at least \( 1 - \frac{1}{20t^2} \) we have that

\[
\left\| \sum_{i=1}^t X_{i,a_i} e_i^T \right\| \leq C \sqrt{t \log(dT)} \left\| \left[ \sum_{i=1}^t (X_{i,a_i})^2, \ldots, \sum_{i=1}^t (X_{i,a_i})^2 \right] \right\|
\]
We next turn to bounding \( \sum_{i=1}^{t} e_i e_i^T - \text{diag}(e_i e_i^T) \). We notice that for all elements except the diagonal we have that \( \mathbb{E}[(e_i)_j(e_i)_k] = \mathbb{E}[(e_i)_j X_{i,a_i}] = \mathbb{E}[(e_i)_j X_{i,a_i}|X_{i,a_i}] = 0, j \neq k \). Hence, it can be shown that \( \sum_{i=1}^{t} (e_i)_j(e_i)_k \) is a martingale with bounded difference for \( j \neq k \), hence, with probability at least \( 1 - \frac{1}{20T^2} \), we have that \( |\sum_{i=1}^{t} (e_i)_j(e_i)_k| \leq C \sqrt{t \log(dT)} \). Hence, we have that with probability at least \( 1 - \frac{1}{20T^2} \)

\[
\| \sum_{i=1}^{t} e_i e_i^T - \text{diag}(e_i e_i^T) \| \leq C \sqrt{t \log(dT) \log(dT)} \leq C \log(dT).
\]

Hence, from (35) and the union bound we have that with probability at least \( 1 - \frac{1}{20T^2} \)

\[
\| \sum_{i=1}^{t} E_i \| \leq C \sqrt{t \log(dT)}.
\]

We next turn to bounding \( |V_{t-1}^{-1}| \). We have from (39), and Assumption 2 and the union bound, the following holds with probability at least \( 1 - \frac{1}{2T^2} \)

\[
\| V_t \| = \| \sum_{i=1}^{t} X_{i,a_i} X_{i,a_i}^T - E_t \| \geq \| \sum_{i=1}^{t} X_{i,a_i} X_{i,a_i}^T \| - \| E_t \| \geq C \left( \frac{t}{d} - \sqrt{t \log(dT)} \right).
\]

Hence for \( t \geq 4 \log(dT) \), we have that with probability at least \( 1 - \frac{1}{2T^2} \), it holds that \( \| V_t \| \geq C \frac{t}{2T} \), and hence,

\[
\| V_{t-1}^{-1} \| \leq C \frac{d}{t}. \tag{41}
\]

Hence, from (32) and the union bound, the following holds with probability at least \( 1 - \frac{1}{2T} \)

\[
\| \theta - \hat{\theta}_t \| \leq C t \frac{\sqrt{\log(dT)}}{\sqrt{t}}. \tag{42}
\]

Therefore, from (31) and the union bound again we have that the following holds with probability at least \( 1 - \frac{1}{T} \)

\[
R_T \leq t \sum_{i=1}^{T} C d \frac{\sqrt{\log(dT)}}{\sqrt{t}} \leq C d \frac{\sqrt{\log(dT)}}{\sqrt{t}} (1 + \int_{t=1}^{T} \frac{1}{\sqrt{t}} dt) \leq 2C d \frac{\sqrt{T \log(dT)}}{\sqrt{t}}. \tag{43}
\]

\[
\square
\]

References


