Sparse Linear Bandits with Fixed Sparsity Support: Adversarial and Stochastic Regimes

Kyoungseok Jang[*]
Chung-Ang University

Nam Phuong Tran[*]
University of Warwick

Nicolò Cesa-Bianchi
Università degli Studi di Milano

Abstract

We study sparse linear bandits in both adversarial and stochastic settings. While existing literature has extensively explored sparse linear bandits in the stochastic regime, the adversarial setting, particularly for general l_p-ball action sets ($p > 1$), remains poorly understood. Our work addresses this gap by showing that the curse of dimensionality in adversarial linear bandits can be broken under a natural fixed sparsity support assumption. Specifically, we design algorithms for the l_∞- and l_2-balls that integrate sparsity support identification with the OSMD algorithm, achieving regret bounds $O(s\sqrt{T}\log T)$ and $O(\sqrt{sT}\log T)$, respectively. These results nearly match the optimal results when the sparsity support is known, and significantly improve upon the $O(d\sqrt{T})$ regret of algorithms ignoring sparsity. Furthermore, in the stochastic setting, we show how the geometry of the l_p-ball action set influences both exploration and regret. Our work highlights fundamental contrasts between adversarial and stochastic regimes, and establishes the first regret guarantees for sparse adversarial linear bandits beyond the l_1-ball action set.

1 INTRODUCTION

Multi-armed bandit (MAB) is a fundamental model for sequential decision-making problems, where a learner selects an arm at each round and observes the associated loss. The goal of the learner is to minimize cumulative loss over T rounds. The linear bandit variant, where the loss at each round is the inner product of a feature vector and a hidden parameter vector in $\mathbb{R}^d$, is extensively studied (Lattimore and Szepesvári, 2020). However, in high-dimensional settings such as online advertising (Yan et al., 2014) or personalized healthcare (Razavian et al., 2015), linear bandits suffer from the curse of dimensionality, with regret scaling polynomially with the ambient dimension (Lattimore and Szepesvári, 2020). To address this challenge, a common strategy is to assume that the model exhibits a low-dimensional structure, with sparsity being a prominent form (Hao et al., 2020a), where only a small subset S of features (the sparsity support, $|S| = s \ll d$) significantly influences the loss. This setting is referred to as the sparse linear bandit.

Much of the existing literature on sparse linear bandits focuses on the stochastic regime, where the model's hidden parameters remain fixed, and the reward noise exhibits well-behaved stochastic properties. These stochastic assumptions enable the use of powerful tools from high-dimensional statistics and compressed sensing, such as Lasso-based estimation (Kim and Paik, 2019; Hao et al., 2020a), iterative hard thresholding methods (Ma et al., 2024), and Catoni's estimator (Jang et al., 2022a).

In contrast to the stochastic model, the adversarial bandit model provides a robustification approach, where the learner must handle any sequence of losses and compare its performance to the best possible action in hindsight. In the context of high-dimensional sparse linear bandits, an adversarial environment means the learner cannot leverage well-behaved stochastic properties of the loss, which prevents the use of the aforementioned tools from high-dimensional statistics. Given that adversarial bandits serve as a robust counterpart to their stochastic variants, it is surprising that sparse adversarial linear bandits have been rarely studied, except when the action set is the l_1-ball (i.e., the sparse K-armed bandits by Bubeck et al. (2018)). Generalising to other action sets, such as the l_2 or l_∞-balls, remains an open

Proceedings of the 29th International Conference on Artificial Intelligence and Statistics (AISTATS) 2026, Tangier, Morocco. PMLR: Volume 300.

[*] : Equal contributions.

challenge. Specifically, as we show later, when the sparsity support varies arbitrarily at each round, the regret for the l_1-ball can be upper-bounded by $\widetilde{O}(\sqrt{sT})$ (where $\widetilde{O}$ hides logarithmic factors), while for the l_∞-ball, the regret is lower-bounded by $\Omega(d\sqrt{T})$. This highlights a significant challenge when departing from the l_1-ball action set and emphasizes how the geometry of the action set plays a crucial role in determining the tractability of the problem.

Due to this fundamental limitation, in this paper, we study the sparse adversarial linear bandit problem beyond the l_1-ball, under the assumption of fixed sparsity support. We argue that the assumption of fixed sparsity support is reasonable, both for technical reasons as discussed earlier and for practical reasons. In many real-world problems, practitioners often collect a large number of potentially redundant features in the hope that the relevant ones are included. Over time, the irrelevant features remain irrelevant, and the sparsity support remains fixed. For instance, in online advertising, relevant features such as ad placement or content type consistently influence click-through rates, while irrelevant features, such as the user's browser version, have a negligible impact. Motivated by this, we ask:

Can we design an algorithm for the adversarial linear bandit on action sets other than the l_1-ball that leverages the sparsity structure to break the curse of dimensionality?

Besides the adversarial setting, we also investigate how the geometry of the action set, particularly the curvature of the l_p-ball, affects regret in the stochastic sparse linear bandit problem. Our motivation stems from a key observation made later in the paper: under the fixed-support assumption, it is possible to design algorithms for adversarial linear bandits with action sets beyond the l_1 ball, such as the l_2 and l_∞ balls, that avoid d-multiplicative factors in the main regret term, specifically in the $\sqrt{T}$ component. This suggests that, in the adversarial setting with fixed sparsity support, the geometry of the action set has limited impact on the achievable regret. In contrast, prior results in the stochastic setting indicate that the action set geometry significantly influences regret (Hao et al., 2020a; Jang et al., 2022a), even under fixed support. This contrast motivates a closer investigation into how the geometry of l_p-ball action sets, particularly for $p \in (2, \infty)$, affects regret in the stochastic sparse linear bandit problem. We address this question in later sections of the paper.

Main contributions. In this paper, we study the sparse linear bandit problems in both adversarial and stochastic settings, and make significant progress in both domains. Below, we summarize our main contributions. All proofs are provided in the appendix.

- We first present a positive result for the adversarial linear bandit on the hypercube $[-1, 1]^d$. We propose an algorithm that integrates sparse coordinate identification into the Online Stochastic Mirror Descent (OSMD) framework, achieving a *minimax optimal* regret of $O(s\sqrt{T}\log T)$ up to logarithmic factors. This algorithm is sparsity-agnostic, which means the algorithm does not require prior information about the sparsity level s.
- Next, for the l_2-ball action set, we modify OSMD's exploration strategy that balances identifying new coordinates and exploiting learnt sparse coordinates, yielding a regret bound of $O(\sqrt{sT}\log T)$ for sufficiently large T, which is *minimax optimal* up to logarithmic factor.
- Finally, we analyze the stochastic sparse linear bandit on the l_p-ball. We show that the action set's curvature affects coordinate identification and regret, and prove that for $p > \log T$, our algorithm achieves $O(s\sqrt{T}\log T)$ regret.

All of our results match the best-known regret bounds achievable by algorithms with access to the sparsity support, highlighting the optimality of our algorithms. To the best of our knowledge, our results are the first for sparse adversarial linear bandits beyond the case of l_1-ball action set. We also provide the first $\widetilde{O}(s\sqrt{T})$ regret bound for stochastic sparse linear bandits with l_p-ball action sets when $p \neq \infty$.

Technical challenges. Regarding the sparse adversarial linear bandit on the l_∞-ball (Section 2.1), the main challenge we address arises from adapting a coordinate-partitioning strategy, originally used in stochastic settings (Lattimore et al., 2015), to the adversarial environment. Selective Explore-then-Commit (SETC) algorithm in Lattimore et al. (2015) achieved an optimal regret bound in the stochastic setting by simultaneously searching sparsity support and exploiting supports that are already found at the same time. On the other hand, in the adversarial setting, where loss vectors can change arbitrarily, the impact of exploration on regret and the criterion for identifying coordinates as part of the sparsity support must be designed with more caution. Moreover, even after identifying several coordinates in the sparsity support, exploiting those coordinates is also a nontrivial challenge, mainly for two reasons. First, unlike the stochastic setting, identified supports still need additional exploration because of changing loss vectors in the adversarial setting. Second, our approach allows new coordinates to be added dynamically rather than in a fixed exploration phase.

This dynamic structure complicates the use of standard adversarial algorithms, such as OSMD, which are sensitive to changes in potential function and can incur significant additional regret.

Regarding the sparse adversarial linear bandit on the l_2-ball (Section 2.2), the main challenge is to adaptively recover the unknown sparsity support set without incurring the full cost of uniform exploration across the ambient dimension. Standard approaches explore all coordinates equally, leading to regret bounds of order $\sqrt{dT}$. The key difficulty is designing a dynamic exploration strategy that concentrates on coordinates already identified in S, while still exploring enough to uncover new support elements. This trade-off introduces a major technical challenge: controlling the variance of the loss estimator under adaptive exploration. The main innovation is a random stopping-time upper bound on the coordinate identification time, combined with carefully tuned coordinate-wise exploration weights to manage this variance.

Section 3 deals with the stochastic setting and shares structural similarities with Section 2.1, but presents a different challenge: how to allocate the norm budget between exploration and exploitation at each time step. In SETC, where the action set is a hypercube, the selection of one coordinate does not restrict the values of other coordinates. For instance, the i-th coordinate of the action at time t, $A_{t,i} \in [-1, 1]$, could be chosen freely, regardless of the value of $A_{t,1}$. However, in Section 3, where the action set is l_p norm ball with $p < \infty$, investing a large norm in one coordinate limits the possible values for the others, for example, if $A_{t,1} = 1$, then $A_{t,i} = 0$ for all $i = 2, \dots, d$. Yet, both the coordinate identifier (for a better signal-to-noise ratio) and the exploitation step (for a higher reward) benefit from larger norm allocations. Balancing this norm investment trade-off is a nontrivial problem. Section 3 addresses this challenge by designing a monotone decreasing exploration budget sequence B_t, which allows for effective interpolation between exploration and exploitation.

Notations. For any $k \in \mathbb{N}_+$, we denote $[k]$ as $\{1, \dots, k\}$. We let $\{e_1, \dots, e_d\}$ denote the standard basis vectors of $\mathbb{R}^d$. Let $\mathbf{0}_d$ denote the zero vector. Let $\mathbb{I}$ denote the indicator function. We let $\mathcal{A}$ denote the action set and $\mathcal{L}$ denote the space of losses. For any $p \geq 1$, let $\mathcal{B}_p(r) := \{x \in \mathbb{R}^d : \|x\|_p \leq r\}$ denote the l_p-ball of radius r. For brevity, we write $\mathcal{B}_p := \mathcal{B}_p(1)$. For a function F with effective domain $\mathcal{A}$, let F^* denote the Legendre-Fenchel dual of F, i.e., $F^*(v) := \sup_{x \in \mathcal{A}} x^\top v - F(x)$. For any index set $I \subset [d]$ and a vector $v \in \mathbb{R}^d$, we define $v^I := (v_i)_{i \in I} \in \mathbb{R}^{|I|}$. We define a generalized Rademacher distribution, $\mathrm{Rad}(p)$, as the distribution over $\{\pm 1\}$ that takes the value $+1$ with probability p and -1 with probability $1 - p$.

Related Work Table 1 presents a comparison of our results with those of previous works. For a more comprehensive review, please refer to Appendix A.

2 SPARSE ADVERSARIAL LINEAR BANDITS WITH FIXED SPARSITY SUPPORT

In this section, we address the adversarial linear bandit problem, which is described as follows. Before the game starts, the adversary chooses a sequence of loss vectors $(\ell_t)_{t=1}^T$, where for each $t \in [T]$, $\ell_t \in \mathcal{L} \subset \mathbb{R}^d$, where $\mathcal{L}$ is the space of loss vectors, and $\|\ell_t\|_0 \leq s$. Given an action set $\mathcal{A} \subset \mathbb{R}^d$, at each round t, the agent chooses (possibly at random) an action $A_t \in \mathcal{A}$ to play and observes the loss $\ell_t(A_t) := \langle \ell_t, A_t \rangle$. The performance of the learner is evaluated by the expected cumulative regret compared to the best fixed action in hindsight:

$$\mathrm{Reg}_T = \mathbb{E}\left[\sum_{t=1}^T \ell_t(A_t)\right] - \min_{a \in \mathcal{A}} \sum_{t=1}^T \ell_t(a). \tag{1}$$

As mentioned earlier, the geometry of the action set significantly affects the difficulty of the problem. While sparse adversarial bandits, particularly the case where $\mathcal{A}$ is a simplex, have been extensively studied and resulting in algorithms that achieve a tight regret bound of $\widetilde{O}(\sqrt{sT})$ even when the sparsity support changes over time (Bubeck et al., 2018), research on sparsity in other action sets remains unexplored. This is fundamentally because exploiting sparsity becomes technically more challenging with different action set geometries. An example is the hypercube, where, as we show through the lower bound presented below, it is impossible to leverage sparsity when the support changes at every round.

Proposition 2.1 (Impossible result with varying sparsity support)**.** *Let $\mathcal{A} = \{-1, 1\}^d$. For all $t \in [T]$, suppose $\|\ell_t\|_\infty \leq 1$, $\|\ell_t\|_0 = 1$, and assume that the sparsity support can vary each round. Suppose $T \geq 4d^2$. Then, the regret can be lower-bounded as $\mathrm{Reg}_T = \Omega(d\sqrt{T})$.*

This result suggests that in order to exploit sparsity for improved performance, additional assumptions are necessary. A natural assumption is to impose constraints on the adversary, such as requiring the sparsity support to remain fixed over time. In particular, in the remainder of this section, we assume that Assumption 1, stated below, holds.

Assumption 1. There exists $S \subset [d]$ with $|S| \leq s$ such that $\ell_{t,i} = 0$ for all $i \notin S$ and $t \in [T]$.

Model	Action set	Adv/Stoc	Sparsity /Agnostic	Regret bound
Bubeck et al. (2018)	$\mathcal{A} = \mathcal{B}_1$	Adv	Vary / ✗	$O(\sqrt{sT})$
Ours (Section 2.1)	$\mathcal{A} = \{-1, 1\}^d \subset \mathcal{B}_\infty$	Adv	Vary / ✗	$\Omega(d\sqrt{T})$
Bubeck et al. (2012)	$\mathcal{A} = \mathcal{B}_\infty$	Adv	Non-sparse	$O(d\sqrt{T})$
Bubeck et al. (2018)	$\mathcal{A} = \mathcal{B}_\infty$	Adv	Non-sparse	$O(d\sqrt{T})$
Ours (Section 2.1)	$\mathcal{A} = \mathcal{B}_\infty$	Adv	Fixed / ✓	$O(s\sqrt{T}\log T)$
Bubeck et al. (2012)	$\mathcal{A} = \mathcal{B}_2$	Adv	Non-sparse	$O(\sqrt{dT})$
Jacobsen et al. (2026)	$\mathcal{A} = \mathcal{B}_2$	Adv	Non-sparse	$\Omega(\sqrt{dT})$
Ours (Section 2.2)	$\mathcal{A} = \mathcal{B}_2$	Adv	Fixed / ✗	$O(s\sqrt{T}\log T)$
Abbasi-Yadkori et al. (2012)	Any $\mathcal{A} \subset \mathcal{B}_2$	Stoc	Fixed / ✗	$O(\sqrt{sdT})$
Lattimore and Szepesvári (2020)	$\mathcal{A} \subset \mathcal{B}_2(\sqrt{s})$	Stoc	Fixed / ✗	$\Omega(\sqrt{sdT})$
Lattimore and Szepesvári (2020)	$\mathcal{A} = \mathcal{B}_\infty$	Stoc	Fixed / ✗	$O(s\sqrt{T})$
Jin et al. (2024)	$\mathcal{A} \subset \mathcal{B}_2$	Stoc	Fixed / ✓	$O(s\sqrt{dT}\log T))$
Ours (Section 3)	$\mathcal{A} = \mathcal{B}_p$	Stoc	Fixed / ✗	$O(s^{\frac{p^2}{p^2-1}} d^{\frac{1}{p+1}} T^{\frac{p+2}{2(p+1)}} \log T)$
Ours (Section 3, Cor 3.2)	$\mathcal{A} = \mathcal{B}_p, T \leq e^p$	Stoc	Fixed / ✗	$48s\sqrt{T}\log T$

Table 1: Comparison with prior work on sparse linear bandits. The column **Sparsity/Agnostic** refers to the sparsity assumption on the loss/reward vector, and whether the algorithm is sparsity-agnostic (meaning that the algorithm does not require sparsity level information). "Fixed"/"Vary" means that the corresponding work assumes (resp. does not assume) the sparsity support of the loss/reward vector is fixed. ✓/✗ means that the corresponding algorithm is (resp. is not) sparsity-agnostic algorithm. "Non-sparse" means the corresponding algorithm does not exploit sparsity structure. Throughout this table, to ensure bounded rewards/losses, the loss/reward vector is assumed to satisfy the duality condition: if the $\mathcal{A} = \mathcal{B}_p$, then $\ell_t \in \mathcal{B}_q$ where $1/p + 1/q = 1$.

One can estimate the best possible outcomes achievable under this assumption from the general results for adversarial linear bandits in d-dimensional settings. For example, Bubeck et al. (2018) proved that the regret bound for the unconstrained d-dimensional hypercube case is $\Theta(d\sqrt{T})$. Similarly, the classic adversarial linear bandit algorithm on the Euclidean ball Bubeck et al. (2012) yields a $\widetilde{O}(\sqrt{dT})$ regret. If the sparsity support S were known, the learner could restrict exploration to coordinates in S, reducing the regret to $O(s\sqrt{T})$ on the hypercube (and $O(\sqrt{sT})$ on the Euclidean ball, respectively). However, achieving those regret bounds without prior knowledge on S is a nontrivial challenge. In Section 2.1 and 2.2, we demonstrate that it is indeed possible to achieve $\widetilde{O}(s\sqrt{T})$ (and $\widetilde{O}(\sqrt{sT})$, respectively) regret without knowledge of the sparsity support.

Remark 1. In the $\mathcal{A} = B_\infty$ setting, when the sparsity support is fixed one might be tempted of running an independent two-armed adversarial bandit algorithm for each coordinate to learn whether each coordinate should be set to either $+1$ or -1. However, as the overall linear loss observed by all instances depends on their stochastic actions, the loss perceived by each instance—as a result of their chosen action in $\{-1, +1\}$—would be affected by non-zero mean noise, violating the assumptions required for adversarial bandit algorithms to operate correctly. See Appendix B for details.

2.1 Sparse Online Stochastic Mirror Descent on the Hypercube

In this section, we study the case where the action set is given by $\mathcal{A} = \mathcal{B}_\infty$. We propose an algorithm `MiDeSiCoE`, which naturally combines the OSMD technique with a coordinate identification procedure. `MiDeSiCoE` algorithm maintains, in each round t, a set S_t of coordinates that are believed, with high probability, to be part of the true sparsity support. When selecting an action, the learning agent treats S_t and its complement S_t^c differently. For the unidentified coordinates S_t^c, the agent employs uniform Rademacher exploration (implicitly in line 19, $A_{t+1}^{S_{t+1}^c}$), as used in SETC algorithm (Lattimore and Szepesvári, 2020) developed for the stochastic sparse linear bandit setting, in order to gather samples that help identify the sparsity support. If the magnitude of the coordinate estimator $\hat{\ell}_{t,i}$ for $i \in S_t^c$ exceeds a threshold W_t, that coordinate is added to the identified set S_{t+1}.

For the identified coordinates in S_t, the agent performs an exploitation phase using the Online Stochastic Mirror Descent (OSMD) algorithm (implicitly in line 20, $A_{t+1}^{S_{t+1}}$) of Bubeck et al. (2012). Given a subset $S \subseteq [d]$, the potential function F used for OSMD is defined as

$$F^S(x) := \sum_{i \in S} \int_0^{x_i} \tanh^{-1}(y)dy. \tag{2}$$

Algorithm 1 `MiDeSiCoE`: Mirror Descent with Simultaneous Coordinate Exploration

Input: Potential function F, learning rate η.
Initialize: $S_1 = \varnothing$, $A_1 = \mathbf{0}_d$, $\gamma_1 = 1$
for $t = 1$ to T **do**
  Draw $\xi_t \sim \text{Bernoulli}(\gamma_t)$
  if $\xi_t = 0$ **then**
    $p_t = \text{Unif}(\{-1, 1\}^d)$
  else
    Set p_t such that $p_{t,i} = \text{Rad}(\frac{1+A_{t,i}}{2})$ for $i \in [d]$.
  end if
  Sampling $\widetilde{A}_t \sim p_t$, receive $\ell_t(\widetilde{A}_t)$.
  Compute $\hat{\ell}_t = (P_t)^{-1}\widetilde{A}_t\widetilde{A}_t^\top \ell_t$, with $P_t = \mathbb{E}_{a\sim p_t}[aa^\top]$.
  Set the threshold $W_t := 4\sqrt{\frac{\log(T(t+1))}{t}}$.
  for $i \in S_t^c$ **do**
    if $|\frac{1}{t}\sum_{\tau=1}^t \hat{\ell}_{\tau,i}| > W_t$ **then**
      $S_{t+1} \leftarrow S_t \cup \{i\}$.
      $\gamma_{t+1} \leftarrow \sqrt{|S_{t+1}|/T}$
    end if
  end for
  Set A_{t+1} s.t. $A_{t+1}^{S_{t+1}^c} = (0, \cdots, 0)$ and $A_{t+1}^{S_{t+1}} = \nabla F^{S_{t+1},*}(-\eta\sum_{\tau=1}^t \hat{\ell}_\tau)$.
end for

We now present the following regret bound. The detailed proof can be found in Appendix C.

Theorem 2.2 (Regret upper bound of `MiDeSiCoE`). *Suppose that Assumption 1 holds. Let $\mathcal{A} = \mathcal{B}_\infty$ and $\mathcal{L} = \mathcal{B}_1$. Then, the expected regret of Algorithm 1 on $\mathcal{A}$, using the potential function in (2), setting $\eta = \frac{1}{\sqrt{T}}$ and $\gamma_t = \sqrt{\frac{|S_t|+1}{T}}$, can be bounded as follows:*

$$\text{Reg}_T = O(s\sqrt{T}\log T).$$

Remark 2. Using the lower bound of adversarial linear bandit in Bubeck et al. (2018), one can check that even if the learning agent knows the sparsity support S before the game starts, the regret lower bound is $\Omega(s\sqrt{T})$. This means our algorithm is *minimax optimal* up to logarithmic factor.

Proof sketch of Theorem 2.2. First, by the linearity of the loss, one can split the regret into two terms: the regret from identified coordinates S_t, and the regret from the remaining coordinates $S' := S \setminus S_t^c$, as follows. For any $a \in \mathcal{A}$,

$$\text{Reg}_T(a) = \mathbb{E}\left[\sum_{t=1}^T (\widetilde{A}_t^{S_t} - a^{S_t})^\top \ell_t^{S_t}\right] + \mathbb{E}\left[\sum_{t=1}^T (\widetilde{A}_t^{S\setminus S_t^c} - a^{S\setminus S_t^c})^\top \ell_t^{S\setminus S_t^c}\right].$$

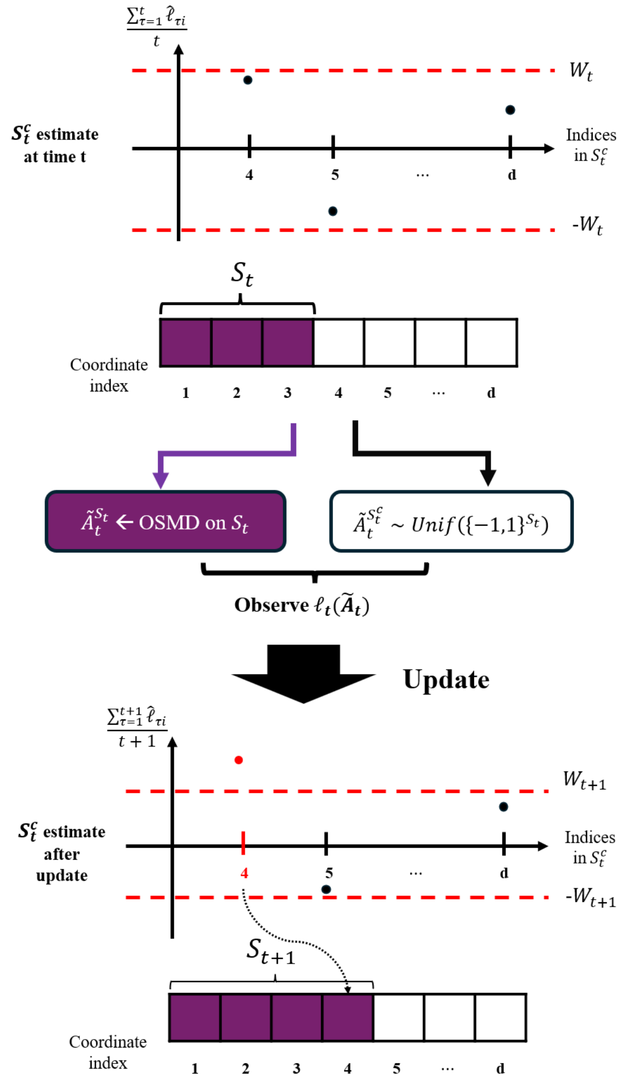


Figure 1: Algorithm 1 Diagram

First, one can prove that the second term can be bounded by $\widetilde{O}(s\sqrt{T})$ (Claim 1). Roughly speaking, it is because of the following reason: if a coordinate i is not being identified before time t, it means that i-th coordinate does not contribute much on the regret.

Let τ_i be the stopping time $\tau_i = \min\{t \in [T] : |S_t| \geq i\} \cup \{T\}$. Modifying the OSMD analysis in Bubeck et al. (2012), one can split regret into three main components.

$$\mathbb{E}\left[\sum_{t=1}^T (\widetilde{A}_t^{S_t} - a^{S_t})^\top \hat{\ell}_t^{S_t}\right] \leq \frac{F^S(a)}{\eta} + \underbrace{\eta\sum_{t=1}^T \langle a^{S_t^c}, \hat{\ell}_t\rangle}_{\text{(RA)}}$$

$$\underbrace{\mathbb{E}\left[\sum_{i=1}^{|S_T|}\left(F^{S_{\tau_i+1}}(-\eta\sum_{s=1}^{\tau_i}\hat{\ell}_t) - F^{S_{\tau_i}}(-\eta\sum_{s=1}^{\tau_i}\hat{\ell}_t)\right)\right]}_{\text{(PC)}}$$

$$+\,\eta\,\underbrace{\mathbb{E}\left[\sum_{i=1}^{|S_T|}\sum_{t=\tau_i+1}^{\tau_{i+1}}\sum_{j\in S_t}\mathbb{E}[(1-A_{tj}^2)\hat{\ell}_{t,j}^2])\right]}_{\text{(V)}}.$$

For the first term, the regret due to Regret Adjustment (RA) on S_t, we can apply similar logic in Claim 1 and obtain a bound of $\widetilde{O}(s\sqrt{T})$. For the second term, the regret caused by the Potential function Change (PC), we showed in Proposition C.1 that our Rademacher exploration ensures this term is bounded by $\widetilde{O}(s\sqrt{T})$. Finally, for the third term, denoted Variance (V), we restrict the analysis to coordinates in S_t and adapt the structure from Bubeck et al. (2012). This yields a bound of $\widetilde{O}(s\sqrt{T})$ for the variance term as well. Combining the three bounds, the overall regret is thus bounded by $\widetilde{O}(s\sqrt{T})$.

Technical novelties. The main technical challenge lies in bounding the difference in the potential function when a new coordinate is added to S_t. In our setting, where the support size increases over time, the key difficulty is controlling how much the potential function changes as the support expands (term PC). By carefully designing the Rademacher exploration scheme and analyzing its properties, we can effectively bound the potential difference, as established in Proposition C.1. Moreover, with the Rademacher exploration scheme and an appropriate choice of W_t, we can properly bound the regret caused by unidentified coordinates (term RA).

Remark 3. For consistency of presentation, the main result is stated in the oblivious adversary setting. In fact, however, Algorithm 1 also works against an **adaptive adversary**. See Appendix C for details. Moreover, Algorithm 1 is **sparsity-agnostic**, in the sense that it does not require prior knowledge of the sparsity level s in order to achieve its guarantee. Indeed, as seen from Theorem 2.2, the hyperparameter choices do not depend on s.

2.2 Sparse Online Stochastic Mirror Descent on the Euclidean Ball

In this section, we address sparse adversarial linear bandit problems where both $\mathcal{A}$ and $\mathcal{L}$ are $\mathcal{B}_2$.

We adapt the OSMD framework for adversarial linear bandits on the l_2-ball, originally proposed by Bubeck et al. (2012), and introduce a modification to the exploration procedure that reduces the total variance of the loss estimation from dT to sT. Specifically, we propose the Online Stochastic Mirror Descent with Selective Exploration (`OSMDSE`) algorithm, whose pseudocode is presented in Algorithm 2.

Algorithm 2 `OSMDSE`: Online Stochastic Mirror Descent with Selective Exploration

1: **Input**: Potential function F, learning rate η, perturbed action set $\mathcal{A}'$, coordinate-priority weight H.
2: **Initialize**: $A_1 = \mathbf{0}_d$, $S_1 = \varnothing$, $\hat{\ell}_1 = \mathbf{0}_d$, $\rho_1(i) = 1/d$ for all $i \in [d]$.
3: **for** $t = 1$ to T **do**
4:   Draw $\xi_t \sim \mathrm{Bernoulli}(\|A_t\|_2)$.
5:   **if** $\xi_t = 1$ **then**
6:     Set $\widetilde{A}_t = A_t/\|A_t\|_2$.
7:   **else**
8:     Sample $I_t \sim \rho_t$.
9:     Sample $\widetilde{A}_t \sim \mathrm{Rad}(\frac{1}{2}) \cdot e_{I_t}$.
10:   **end if**
11:   Play $\widetilde{A}_t$, observe loss $\ell_t(\widetilde{A}_t)$.
12:   Compute $\hat{\ell}_t = (1-\xi_t)\frac{\widetilde{A}_t \ell_t(\widetilde{A}_t)}{\rho_t(I_t)(1-\|A_t\|_2)}$.
13:   Update $A_{t+1} = \Pi_{\mathcal{A}'}(\nabla F^*(-\eta \sum_{\tau=1}^{t} \hat{\ell}_\tau))$.
14:   **if** $\xi_t = 0$ and $\ell_{t,I_t} \neq 0$ **then**
15:     Set $S_{t+1} = S_t \cup \{I_t\}$.
16:   **end if**
17:   Update $w_{t+1}(i) = \begin{cases} H, & \text{if } i \in S_{t+1} \\ 1, & \text{otherwise.} \end{cases}$
18:   Compute $\rho_{t+1}(i) = \frac{w_{t+1}(i)}{\sum_{j=1}^{d} w_{t+1}(j)}$.
19: **end for**

At each round $t \in [T]$, the algorithm draws $\xi_t \sim \mathrm{Bernoulli}(\|A_t\|_2)$ to decide between exploitation and exploration. If $\xi_t = 1$ (exploitation), the algorithm exploits by playing $\widetilde{A}_t = A_t/\|A_t\|_2$. If $\xi_t = 0$ (exploration), the algorithm samples a coordinate $I_t \sim \rho_t$, where ρ_t is a probability distribution over the coordinates. Once I_t is drawn, the algorithm plays an exploratory action (line 9). The algorithm then estimates the loss (line 12), which is equivalent to

$$\hat{\ell}_t = \begin{cases} \dfrac{\ell_{t,I_t} e_{I_t}}{\rho_t(I_t)(1-\|A_t\|_2)} & \text{if } \xi_t = 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

One can verify that the loss estimator is unbiased, i.e., $\mathbb{E}[\hat{\ell}_t \mid A_t] = \ell_t$. Next, the algorithm updates A_{t+1} using mirror descent (line 13), where $\Pi_{\mathcal{A}'}$ is the projection operator on to $\mathcal{A}'$ w.r.t. $\|\cdot\|_2$. During exploration rounds, if the algorithm detects that coordinate I_t incurs a nonzero loss, it adds I_t to the sparsity support S_{t+1} (lines 14–16). It then updates the weights $w_{t+1}(i)$, setting them to a fixed parameter $H > 0$ (called the *coordinate-priority weight*) for $i \in S_{t+1}$ and to 1 otherwise, in order to bias future exploration toward identified coordinates (line 17). The coordinate-exploring distribution is then computed as $\rho_{t+1}(i) = \frac{w_{t+1}(i)}{\sum_{j=1}^{d} w_{t+1}(j)}$.

We now present the main regret upper bound of

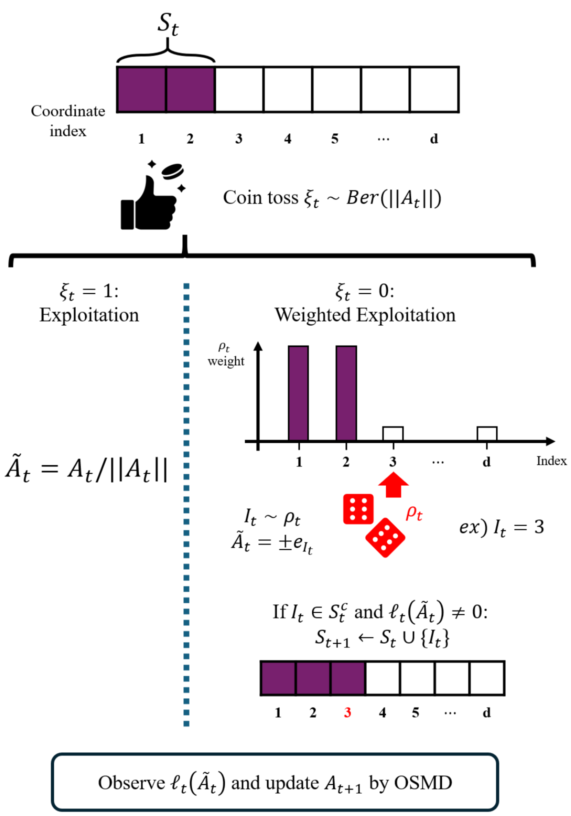


Figure 2: Algorithm 2 Diagram

`OSMDSE`.

Theorem 2.3 (Regret upper-bound of `OSMDSE` Algorithm)**.** *Suppose that Assumption 1 holds. Let $\mathcal{A}$ and $\mathcal{L}$ be $\mathcal{B}_2$. Let $\mathcal{A}' = \{x \in \mathbb{R}^d \mid \|x\|_2 \leq 1-\gamma\}$. Let $F(x) = -\log(1-\|x\|_2) - \|x\|_2$. Choose $H = d$, $\gamma = 1/\sqrt{T}$, and $\eta = \frac{1}{4\sqrt{sT}}$. Suppose $T \geq sd^2$. Then, the expected regret of Algorithm 2 can be bounded as:*

$$\mathrm{Reg}_T \leq 5\sqrt{sT}\log T + s^{2.5}d^2.$$

Proof sketch of Theorem 2.3. We provide a proof sketch of Theorem 2.3 here and defer the full proof to Appendix D. First, we choose the potential function $F(x) = -\log(1-\|x\|_2) - \|x\|_2$, similar to the seminal work (Bubeck et al., 2012). Therefore, as in Bubeck et al. (2012), with suitable choices of η and H, we obtain the following regret decomposition for OSMD:

Proposition 2.4. *Let $F(x) = -\log(1-\|x\|_2) - \|x\|_2$. Choose $H = d$, $\eta = \frac{1}{4\sqrt{sT}}$, $\gamma = 1/\sqrt{T}$. Suppose $T \geq sd^2$, then*

$$\mathrm{Reg}_T \leq \gamma T + \frac{\log(\gamma^{-1})}{\eta} + \eta \sum_{t=1}^{T} \mathbb{E}\left[(1-\|A_t\|_2)\|\hat{\ell}_t\|_2^2\right].$$

The regret decomposition in Proposition 2.4 suggests that, to achieve regret of order $\sqrt{sT}$, we need to show that $\sum_{t=1}^T \mathbb{E}[(1-\|A_t\|_2)\|\hat{\ell}_t\|_2^2] = O(sT)$. In the next lemma, we show that this can be achieved by using a coordinate-weight updating procedure, which is a key ingredient in our analysis.

Lemma 2.5. *Choose $H = d$, $\gamma = 1/\sqrt{T}$, then $\sum_{t=1}^T \mathbb{E}\left[(1-\|A_t\|_2)\|\hat{\ell}_t\|_2^2\right] \leq 2sT + 4s^3d^2\sqrt{T}$.*

Combining Lemma 2.5 together with Proposition 2.4, we obtain the upper-bound as in Theorem 2.3.

Remark 4. We note that, as T grows, the first term in Theorem 2.3 dominates, resulting in a regret of order $\widetilde{O}(\sqrt{sT})$. Moreover, the lower bound in the recent work on adversarial linear bandits over $\mathcal{B}_2$ (Jacobsen et al., 2026) proved that, even when the learner knows the sparsity support S before the game starts, the regret is at least $\Omega(\sqrt{sT})$. Hence, our algorithm is *minimax optimal* up to logarithmic factors.

Technical novelties. The main challenge lies in bounding the variance term, as shown in Lemma 2.5. We address this by carefully handling the effect of identifying entries in S on the total variance. Specifically, we control this effect by employing a random stopping-time upper bound argument and by selecting the coordinate-exploring weight H with great care.

Regarding the choice of H, intuitively it should balance between exploiting already identified coordinates and exploring ones that have not yet been identified. In particular, H must be large enough so that if a coordinate i is identified prior to round t, the probability of selecting i among the d coordinates should be on the order of $\Omega(1/s)$, leading to a long-term regret of $O(\sqrt{sT})$. In other words, for $i \in S_t$,

$$\rho_t(i) = \frac{H}{H|S_t| + (d-|S_t|)} = \Omega\left(\frac{1}{s}\right),$$

which holds if one chooses $H = \Omega(d)$. Moreover, a closer inspection of the proof of Lemma 2.5 reveals that the total variance incurred from identifying a coordinate $i \in S$ is at most $(sH)^2\sqrt{T}$. To ensure a regret bound of order $\sqrt{sT}$, we require that $H = O(s^{-1}\sqrt{T})$. When T is sufficiently large, setting $H = d$ satisfies both conditions and ensures the regret bound stated in Theorem 2.3.

Remark 5. An additional advantage worth highlighting is that both `MiDeSiCoE` and `OSMDSE` are computationally efficient algorithms. This point is far from trivial: in this line of work, designing algorithms that not only attain optimal regret guarantees but are also implementable with reasonable computational cost is a challenging task, as already noted by Bubeck et al. (2012). We provide the time and space complexity analyses of `MiDeSiCoE` and `OSMDSE` in Appendices C.3 and D.4.

Algorithm 3 `OFULSiCoE`: OFUL with Simultaneous Coordinate Exploration

1: **Input**: Decreasing exploration budget $\{B_t\}_{t=1}^T$, confidence radius parameter β.
2: **Initialize**: $S_1 = \varnothing$.
3: **for** $t = 1$ to T **do**
4: Set A_t such that
$A_{t,i} \sim B_t \cdot \mathrm{Rad}(1/2)$ if $i \in S_t^c$,
and $A_t^{S_t} := \underset{a \in \mathcal{B}_p^{S_t}(R_t)}{\arg\max} \langle \hat{\theta}_t, a\rangle + \sqrt{\beta}\|a\|_{V_t^{-1}}$ where
$R_t = (1 - (d - |S_t|)B_t^p)^{1/p}$ and $\mathcal{B}_p^{S_t}(r) := \{v \in \mathbb{R}^{|S_t|} : \|v\|_p \le r\}$.
5: Sample A_t and observe the reward r_t.
6: Compute the threshold
$W_t := 4\sqrt{\frac{1}{\sum_{\tau=1}^t B_\tau^2}\log(t)}$.
7: **for** $i \in S_t^c$ **do**
8: Compute $\tilde{\theta}_{t,i} := \frac{\sum_{\tau=1}^t A_{\tau,i} r_\tau}{\sum_{\tau=1}^t B_\tau^2}$.
9: **if** $|\tilde{\theta}_{t,i}| > W_t$ **then**
10: $S_{t+1} \leftarrow S_t \cup \{i\}$.
11: **end if**
12: **end for**
13: $V_{t+1} = \lambda I_{|S_{t+1}|} + \sum_{\tau=1}^t A_\tau^{S_{t+1}}(A_\tau^{S_{t+1}})^\top$.
14: $\hat{\theta}_{t+1} := V_{t+1}^{-1}\sum_{\tau=1}^t A_\tau^{S_{t+1}} r_\tau$.
15: **end for**

3 SPARSE STOCHASTIC LINEAR BANDITS ON l_p-BALLS

In this section, we investigate the stochastic sparse linear bandit problem in the intermediate regime where $p \in (2, \infty)$. Let q be the Hölder conjugate of p. The setting is as follows: at each round t, the learning agent selects an action $A_t \in \mathcal{B}_p$, and observes the reward $r_t = \langle A_t, \theta^*\rangle + \epsilon_t$, where θ^* is a fixed vector satisfying $\|\theta^*\|_0 \le s$ and $\theta^* \in \mathcal{B}_q$, and ϵ_t is a 1-subgaussian noise variable that is conditionally independent of the past, i.e., independent of the σ-field $\mathcal{F}_t := \sigma\left(\{A_i\}_{i=1}^t, \{r_i\}_{i=1}^{t-1}\right)$, which represents the information available just before r_t is observed. Note that θ^* is fixed over timesteps. The performance of learner is evaluated using expected regret as

$$\mathrm{Reg}_T = \max_{a \in \mathcal{A}} \langle a, \theta^*\rangle - \mathbb{E}\left[\sum_{t=1}^T \langle A_t, \theta^*\rangle\right].$$

The main objective of this section is to examine whether the `MiDeSiCoE` introduced in Section 2.1 can be effectively extended to this stochastic setting. A key challenge is that the poking strategy used in the adversarial case (Section 2.2) is no longer applicable. In adversarial settings, where noise is absent, any nonzero loss value confirms the inclusion of a coordinate in the sparsity support. However, in the stochastic setting, the presence of noise makes such identification unreliable. Indeed, it has been shown that there exists an arm set $\mathcal{A} \subset \mathcal{B}_2(\sqrt{s})$ such that, any algorithm must incur a regret of at least $\Omega(\sqrt{sdT})$, indicating that achieving bounds independent of dimensionality is impossible (Lattimore and Szepesvári, 2020). In contrast, the SETC algorithm, originally proposed as the stochastic counterpart to `MiDeSiCoE`, achieves a regret of $\widetilde{O}(s\sqrt{T})$ under the stochastic setting. This naturally leads to the intriguing question of whether there exists a smooth interpolation between these two regimes as p varies within $(2, \infty)$, and to what extent the structure of the action set influences the possibility of exploiting sparsity efficiently.

We propose an algorithm named `OFULSiCoE`, whose pseudocode is provided in Algorithm 3. Algorithm `OFULSiCoE` follows a two-stage structure similar to `MiDeSiCoE`, consisting of a coordinate identification phase and an exploitation phase. However, it differs in two key aspects. First, for coordinate identification, each unidentified coordinate $i \in S_t^c$ is assigned an exploration weight $A_{t,i} = B_t$, such that the signal gradually decreases as t increases. Second, in the exploitation phase, instead of using online mirror descent, the algorithm uses OFUL (Abbasi-Yadkori et al., 2011), a standard algorithm for stochastic linear bandits. It is important to note that the estimator for identifying the coordinate $\tilde{\theta}_{t,i}$ might differ from the one in the estimator for OFUL, $\hat{\theta}_{t,i}$. Moreover, by convention, when $S_t = \varnothing$, we simply skip the step of updating $\hat{\theta}_t$ and V_t (lines 13–14). This algorithm achieves the following regret bound:

Theorem 3.1 (Regret upper bound of `OFULSiCoE`)**.** *Let $\mathcal{A} = \mathcal{B}_p$ and $\mathcal{L} = \mathcal{B}_q$. For $T > \max(d, e^4 s^8)$, the expected regret of Algorithm 3 on $\mathcal{A}$ with exploration budget $B_t = \min\left((s^q d^{-1} t^{-\frac{1}{2}})^{\frac{1}{p+1}}, d^{-1/p}\right)$, $\beta = s\log T$ can be bounded as*

$$\mathrm{Reg}_T = O\left(s^{\frac{p^2}{(p^2-1)}} d^{\frac{1}{p+1}} T^{\frac{1}{2(p+1)}} \sqrt{T}\log T\right).$$

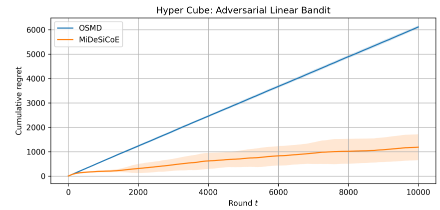
Theorem 3.1 reveals two distinct regimes for the regret of Algorithm 3.

Corollary 3.2. *Under the same assumptions as Theorem 3.1, the expected regret of Algorithm 3 satisfies*

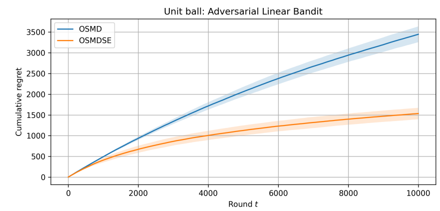
$$\mathrm{Reg}_T \le \begin{cases} 48s\sqrt{T}\log T & \text{if } T \le e^p \\ O(T^{\frac{1}{2}+\frac{1}{2(p+1)}}\log T) & \text{otherwise.} \end{cases}$$

The significance of this result is showing that the dimension-free regret bound, previously known only for $\mathcal{A} = \mathcal{B}_\infty$, can be extended to more general action sets whenever an appropriate relationship between the curvature of the action set and the time horizon T is satisfied. In this sense, it also provides a useful indication that regret bounds may interpolate according to the curvature of the action set.

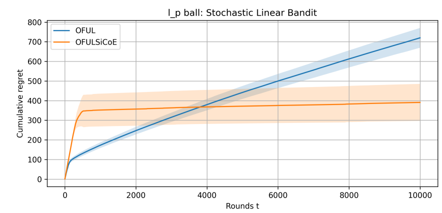
Technical novelties. The main technical challenge in this setting lies in the careful allocation of the exploration budget B_t. In the $\mathcal{B}_\infty$ case, the selection of one coordinate does not interfere with the selection of others, allowing independent exploration. However, for $\mathcal{B}_p$ with $p < \infty$, selecting a large value for one coordinate imposes restrictions on the values assignable to others due to the global norm constraint. For instance, if $A_{t,1} = 1$, then $A_{t,i} = 0$ for all $i \in \{2, \ldots, d\}$. This coupling effect necessitates a fundamental shift from the `MiDeSiCoE` algorithm: at each time step, the learner must determine how much of the available norm to allocate to exploration versus exploitation. Allocating too little norm to coordinate identification leads to poor signal-to-noise ratio and slow tightening of confidence bounds, while allocating too much reduces the capacity for exploitation, ultimately increasing regret. We identify a suitable trade-off by setting the exploration budget as $B_t \approx t^{-\frac{1}{2(p+1)}}$, which enables a form of interpolation across different regimes. Establishing this balance and proving that it leads to desirable regret guarantees constitutes one of the key contributions of our work.



(a) Hypercube, $d = 100$, $s = 3$



(b) l_2 ball, $d = 100$, $s = 3$



(c) l_p ball, $d = 100$, $s = 3$, $p = 30$

Figure 3: Expected cumulative regret for adversarial and stochastic linear bandits.

4 EXPERIMENTS

In this section, we present simulation results for the adversarial setting. We ran 100 independent trials of adversarial and stochastic linear bandits with dimension $d = 100$, sparsity $s = 3$, and horizon $T = 10{,}000$, with randomly generated loss vectors. In both the hypercube and unit Euclidean ball settings, we compared our algorithms (MiDeSiCoE and OSMDSE) with OSMD baselines, recording cumulative regret curves. In the stochastic case, we compared our algorithm with the classic OFUL. Across all settings, our algorithms (blue) consistently achieved lower regret than the baselines (orange), which do not exploit sparsity. Results are shown in Figure 3. We refer readers to Appendix F for additional results and a detailed description of the experiments.

5 CONCLUSIONS

In this paper, we study sparse linear bandits in both adversarial and stochastic regimes beyond the l_1-ball action set. In the adversarial case, we first show that when $\mathcal{A} = \{-1, 1\}^d$ and the sparsity support varies arbitrarily, any algorithm must suffer $\Omega(d\sqrt{T})$ regret, highlighting the challenge beyond the canonical l_1-ball. However, under fixed sparsity support, we design algorithms that achieve $\widetilde{O}(s\sqrt{T})$ and $\widetilde{O}(\sqrt{sT})$ regret, which are minimax optimal up to logs. These are the first such results for adversarial linear bandits beyond l_1-balls. In the stochastic setting, we show that $\widetilde{O}(s\sqrt{T})$ regret is achievable for sufficiently large p.

Despite these advances, our work has certain limitations. In the adversarial case, we focus only on two special cases of the l_p-ball action sets, namely $p = 2$ and $p = \infty$. The achievable regret for general values of p remains unclear. In particular, it is an open question whether one can avoid the curse of dimensionality without assuming fixed sparsity support in the range $p \in (1, 2)$. Therefore, in future work, we aim to extend our l_2 and l_∞ results to general l_p-balls and identifying the threshold p where $\widetilde{O}(s\sqrt{T})$ remains achievable under varying sparsity. In the stochastic setting, while our result is minimax optimal when $p \geq \log T$, it becomes suboptimal when $p \to 2$. A key direction for future work is to design algorithms that achieve $\widetilde{O}(s\sqrt{T})$ regret as $p \to \infty$, while smoothly interpolating to the known $O(\sqrt{dsT})$ bound as $p \to 2$.

Acknowledgements

KJ was supported by the Institute of Information & Communications Technology Planning & Evaluation

(IITP) grant funded by the Korea government (MSIT) [RS-2021-II211341, Artificial Intelligence Graduate School Program (Chung-Ang University)]. NCB acknowledges the financial support from the EU Horizon CL4-2022-HUMAN-02 research and innovation action under grant agreement 101120237, project ELIAS (European Lighthouse of AI for Sustainability) and from the EU Horizon CL4-2021-HUMAN-01 research and innovation action under grant agreement 101070617, project ELSA (European Lighthouse on Secure and Safe AI).

References

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. *Advances in neural information processing systems*, 24, 2011.

Yasin Abbasi-Yadkori, David Pal, and Csaba Szepesvari. Online-to-confidence-set conversions and application to sparse stochastic bandits. In *Artificial Intelligence and Statistics*, pages 1–9. PMLR, 2012.

Kaito Ariu, Kenshi Abe, and Alexandre Proutière. Thresholded lasso bandit. In *International Conference on Machine Learning*, pages 878–928. PMLR, 2022.

Sébastien Bubeck, Nicolo Cesa-Bianchi, and Sham M Kakade. Towards minimax policies for online linear optimization with bandit feedback. In *Conference on Learning Theory*, 2012.

Sébastien Bubeck, Michael Cohen, and Yuanzhi Li. Sparsity, variance and curvature in multi-armed bandits. In *Algorithmic Learning Theory*, pages 111–127. PMLR, 2018.

Sunrit Chakraborty, Saptarshi Roy, and Ambuj Tewari. Thompson sampling for high-dimensional sparse linear contextual bandits. In *International Conference on Machine Learning*, pages 3979–4008. PMLR, 2023.

Yi Chen, Yining Wang, Ethan X Fang, Zhaoran Wang, and Runze Li. Nearly dimension-independent sparse linear bandit over small action spaces via best subset selection. *Journal of the American Statistical Association*, 119(545):246–258, 2024.

Yan Dai, Ruosong Wang, and Simon Shaolei Du. Variance-aware sparse linear bandits. In *The Eleventh International Conference on Learning Representations*, 2023.

Varsha Dani, Sham M Kakade, and Thomas Hayes. The price of bandit information for online optimization. In *Advances in Neural Information Processing Systems*, 2007.

Varsha Dani, Thomas P Hayes, and Sham M Kakade. Stochastic linear optimization under bandit feedback. 2008.

Botao Hao, Tor Lattimore, and Mengdi Wang. High-dimensional sparse linear bandits. In *Advances in Neural Information Processing Systems*, 2020a.

Botao Hao, Tor Lattimore, and Mengdi Wang. High-dimensional sparse linear bandits. *Advances in Neural Information Processing Systems*, 33:10753–10763, 2020b.

Botao Hao, Tor Lattimore, and Wei Deng. Information directed sampling for sparse linear bandits. *Advances in Neural Information Processing Systems*, 34:16738–16750, 2021.

Andrew Jacobsen, Dorian Baudry, Shinji Ito, and Nicolò Cesa-Bianchi. A perturbation approach to unconstrained linear bandits, 2026. URL https://arxiv.org/abs/2603.28201.

Kyoungseok Jang, Chicheng Zhang, and Kwang-Sung Jun. Popart: Efficient sparse regression and experimental design for optimal sparse linear bandits. In *Advances in Neural Information Processing Systems*, 2022a.

Kyoungseok Jang, Chicheng Zhang, and Kwang-Sung Jun. Popart: Efficient sparse regression and experimental design for optimal sparse linear bandits. *Advances in Neural Information Processing Systems*, 35:2102–2114, 2022b.

Tianyuan Jin, Kyoungseok Jang, and Nicolò Cesa-Bianchi. Sparsity-agnostic linear bandits with adaptive adversaries. *Advances in Neural Information Processing Systems*, 37:42015–42047, 2024.

Gi-Soo Kim and Myunghee Cho Paik. Doubly-robust lasso bandit. In *Advances in Neural Information Processing Systems*, 2019.

Joon Kwon and Vianney Perchet. Gains and losses are fundamentally different in regret minimization: The sparse case. *Journal of Machine Learning Research*, 2016.

Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

Tor Lattimore, Koby Crammer, and Csaba Szepesvári. Linear multi-resource allocation with semi-bandit feedback. In *Advances in Neural Information Processing Systems*, volume 2015-January, 2015.

Ke Li, Yun Yang, and Naveen N Narisetty. Regret lower bound and optimal algorithm for high-dimensional contextual linear bandit. *Electronic Journal of Statistics*, 15(2):5652–5695, 2021.

Wenjie Li, Adarsh Barik, and Jean Honorio. A simple unified framework for high dimensional bandit problems. In *International Conference on Machine Learning*, pages 12619–12655. PMLR, 2022.

Wanteng Ma, Dong Xia, and Jiashuo Jiang. High-dimensional linear bandits with knapsacks. In *Proceedings of the 41st International Conference on Machine Learning*, Proceedings of Machine Learning Research, 2024.

Quan Nguyen, Shinji Ito, Junpei Komiyama, and Nishant A. Mehta. Data-dependent bounds with t-optimal best-of-both-worlds guarantees in multi-armed bandits using stability-penalty matching. *arXiv preprint arXiv:2502.08143*, 2025.

Min-Hwan Oh, Garud Iyengar, and Assaf Zeevi. Sparsity-agnostic lasso bandit. In *International Conference on Machine Learning*, pages 8271–8280. PMLR, 2021.

Narges Razavian, Saul Blecker, Ann Marie Schmidt, Aaron Smith-Mclallen, Somesh Nigam, and David Sontag. Population-level prediction of type 2 diabetes from claims data and analysis of risk factors. *Big Data*, 3, 2015.

Taira Tsuchiya, Shinji Ito, and Junya Honda. Stability-penalty-adaptive follow-the-regularized-leader: Sparsity, game-dependency, and best-of-both-worlds. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

Ling Yan, Wu Jun Li, Gui Rong Xue, and Dingyi Han. Coupled group lasso for web-scale ctr prediction in display advertising. In *31st International Conference on Machine Learning, ICML 2014*, volume 3, 2014.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes]
   - We specified the problem setting in Sections 2 and 3, and stated our additional assumptions explicitly, such as in Assumption 1. In Sections 2.1, 2.2 and 3, we also included algorithm pseudocodes and detailed explanations to help readers gain a clear understanding.

   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes]
   - In bandit algorithms, the most crucial complexity is the sample complexity, which directly determines the regret. We discuss this in detail in Appendices C, D, and E, where we provide the proofs of Theorems 2.2, 2.3, and 3.1. At the end of each appendix, we also include the time and space complexity to provide additional information for interested readers.

   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Yes]
   - We have included the anonymized source code, which can be found in the Appendix F.

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. [Yes]
   - We clearly stated our main assumptions of theoretical results in Sections 2 and 3, and stated our additional assumptions explicitly, such as in Assumption 1.

   (b) Complete proofs of all theoretical results. [Yes]
   - We include complete proofs of our theoretical results, Theorems 2.2, 2.3, and 3.1, in Appendices C, D, and E.

   (c) Clear explanations of any assumptions. [Yes]
   - In Section 2, we provide a clear explanation of Assumption 1. We further demonstrate the necessity of this additional assumption through Proposition 2.1.

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Yes]
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Yes]
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Yes]
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Yes]
   - All the experimental details, including anonymized code URL, are in Appendix F.

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. [Not Applicable]
   (b) The license information of the assets, if applicable. [Not Applicable]

(c) New assets either in the supplemental material or as a URL, if applicable. [Not Applicable]
(d) Information about consent from data providers/curators. [Not Applicable]
(e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]

- We used a synthetic dataset and did not rely on existing assets for our experiments.

5. If you used crowdsourcing or conducted research with human subjects, check if you include:

(a) The full text of instructions given to participants and screenshots. [Not Applicable]
(b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]
(c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

- We did not use crowdsourcing or conduct research involving human subjects.

Sparse Linear Bandits with Fixed Sparsity Support: Adversarial and Stochastic Regimes

Contents of Appendix

Additional notations

For a distribution P, denote $\mathrm{supp}(P)$ as its support. Given $F : \mathcal{D} \to \mathbb{R}$, a continuously-differentiable, strictly convex function defined in a domain $\mathcal{D}$, define the Bergman divergence between two points $x, y \in \mathcal{D}$ w.r.t. F as $D_F(x, y) := F(x) - F(y) - \langle \nabla F(y), x - y\rangle$. Let Tr denote the matrix trace operator.

A Related Works

Sparse adversarial MABs. Sparsity in adversarial linear bandits has been primarily studied when the action set is the l_1-ball and the sparsity support varies over time. Kwon and Perchet (2016) first established $\widetilde{O}(\sqrt{sT})$ regret bounds for gains, and later Bubeck et al. (2018); Tsuchiya et al. (2023); Nguyen et al. (2025) extended these results to both gains and losses. However, all prior work is restricted to the l_1-ball setting. In contrast, we investigate sparse adversarial linear bandits with different action sets, such as the l_2- and l_∞-balls. We show that departing from the l_1-ball cases introduces new challenges: for example, in the l_∞-ball setting, the regret can be as large as $\Omega(d\sqrt{T})$ even when the loss vectors are 1-sparse and sparsity support can vary each round. This underscores the difficulty of extending sparsity-aware algorithms beyond the l_1-ball action set.

Adversarial linear bandits. Adversarial linear bandit problems have been extensively studied for various action sets. For the l_∞-ball, it is known that the regret upper bound is $\widetilde{O}(d\sqrt{T})$ (Bubeck et al., 2012), and the lower bound is $\Omega(d\sqrt{T})$ (Bubeck et al., 2018; Dani et al., 2007). For the l_2-ball, Bubeck et al. (2012) proposed a mirror descent-based algorithm that achieves an upper bound of $\widetilde{O}(\sqrt{dT})$. In the case of l_p-balls with $p \in (2, \infty]$, the lower bound is $\Omega(d\sqrt{T})$ (Bubeck et al., 2012). However, these works do not consider any sparsity structure in the problem, which is the main focus of this paper. Our work builds on the mirror descent algorithm (Bubeck et al., 2012), combining it with a coordinate identification procedure to achieve a regret of $O(s\sqrt{T})$ for the l_∞-ball and $O(\sqrt{sT})$ for the l_2-ball, explicitly leveraging the sparsity structure.

Sparse stochastic linear bandits. Since the theoretical formulation of the stochastic linear bandit (Dani et al., 2008; Abbasi-Yadkori et al., 2011), research on the sparse stochastic linear bandit has progressed much more actively than in the adversarial setting. In particular, when the action set is bounded in the l_2-norm, it has been shown that both the upper and lower bounds of the regret are $\Theta(\sqrt{sdT})$ (Abbasi-Yadkori et al., 2012; Lattimore and Szepesvári, 2020), implying that dimensionality dependence is unavoidable in such environments. In the case where the learning agent is unaware of the sparsity level s it is even worse—the most recent result is the algorithm in Jin et al. (2024), which achieves a regret bound of $O(s\sqrt{dT})$. Due to this limitation, a line of work has focused on avoiding dimensional dependence by imposing alternative curvature conditions (Dai et al., 2023; Hao et al., 2020a; Jang et al., 2022a; Li et al., 2022) or stochastic assumptions on the action set (Ariu et al., 2022; Chakraborty et al., 2023; Chen et al., 2024; Li et al., 2021; Oh et al., 2021). A representative result closely related to our work is that, when the arm set is a hypercube, the regret can be reduced to $O(s\sqrt{T})$, thereby eliminating the dimensional dependence (Lattimore and Szepesvári, 2020). On the other hand, the intermediate case of the l_p-norm ball has received little attention, except for some $T^{2/3}$-order regret bounds developed under high-dimensional settings (Jang et al., 2022b; Hao et al., 2020b, 2021).

B Additional Explanation about Remark 1

When interpreting the adversarial linear bandit problem in Section 2.1, one might imagine the following scenario:

- Each coordinate-wise agent $i \in [d]$ selects $A_{t,i} \in -1, 1$ at each round. Each agent will regard other agents as part of the environment noise.
- Afterward, all agents observe the common loss $\langle \ell_t, A_t \rangle$, where ℓ_t satisfies Assumption 1.

Under this interpretation, one may mistakenly assume that the problem simply reduces to d independent adversarial bandits, and since only s agents truly influence the loss, one might expect a regret of order $O(s\sqrt{T})$.

However, this scenario does not lead to the claimed bound in Section 2.1. Depending on the algorithm used by each coordinate-wise agent, the behavior can be far more complex. For instance, if each coordinate runs a randomized algorithm such as Exp3, the update performed by the i-th agent induces non-zero mean noise in the feedback observed by the j-th agent, and this interaction makes the analysis nontrivial. If this argument were truly trivial, then the d-dimensional adversarial linear bandit on the hypercube studied in Bubeck et al. (2012) could also have been reduced to a simple combination of d independent agents, which is clearly not the case.

C Proof of Subsection 2.1

C.1 Proof of Theorem 2.2

In this analysis, note that the coordinates that are not on the support won't affect the regret. So all vectors are in $\mathbb{R}^S$.

First, we need to check how our loss estimates change when we add a new coordinate so that $S_t \neq S_{t+1}$. For any index set $I \subset [d]$, let $P_t^{(I)} := \mathbb{E}_{a \sim p_t}\left[a^I (a^I)^\top\right] \in \mathbb{R}^{|I|\times|I|}$ and $z_t^{(I)} := (P_t^{(I)})^{-1} \widetilde{A}_t^I \ell_t(\widetilde{A}_t)$.

Proposition C.1. *For all t, the following holds:*

- *For any $S' \supset S_t$, $z_t^{(S')} = \hat{\ell}_t^{S'}$. In particular, $\hat{\ell}_t^{S_t} = z_t^{(S_t)}$.*

- *For any* $i \in S_t^c$, $\hat{\ell}_{t,i} = \widetilde{A}_{t,i}\ell(\widetilde{A}_{t,i})$.

Proof. For any $i \in S_t^c$, $\widetilde{A}_{t,i} \sim \mathrm{Rad}(\frac{1}{2})$ independently from the coordinates in S_t. This means that for any $j \neq i$, $\mathbb{E}\big[\widetilde{A}_{t,i}\widetilde{A}_{t,j}\big] = 0$ and $\mathbb{E}\big[\widetilde{A}_{t,i}\widetilde{A}_{t,i}\big] = 1$. Therefore,

$$\left(P_t^{(S')}\right)_{i,j\in S'} := \begin{cases} P_t^{(S_t)} & \text{if } i,j \in S_t \\ 1 & \text{if } i = j \in S_t^c \\ 0 & \text{Otherwise} \end{cases}.$$

To simplify, let $R \in \mathbb{R}^{S' \times S'}$ be a permutation matrix that rearranges the indices in S' such that indices in S_t come first and indices in S_t^c come last. Then one can write matrices $P_t^{S'}$ as

$$RP_t^{(S')}R^\top = \begin{bmatrix} P_t^{(S_t)} & \mathbf{0}_{|S_t|\times|S'\setminus S_t|} \\ \mathbf{0}_{|S'\setminus S_t|\times|S_t|} & I_{|S'\setminus S_t|} \end{bmatrix}.$$

From the block matrix inversion, one can derive that

$$\left((P_t^{(S')})^{-1}\right)_{i,j\in S'} := \begin{cases} (P_t^{(S_t)})^{-1} & \text{if } i,j \in S_t \\ 1 & \text{if } i = j \in S_t^c \\ 0 & \text{Otherwise} \end{cases}.$$

This implies

$$\begin{aligned} z_t^{(S')} &= (P_t^{(S')})^{-1}\widetilde{A}_t^{S'}\ell_t(\widetilde{A}_t) \\ &= R^\top R (P_t^{(S')})^{-1} R^\top R \widetilde{A}_t^{S'}\ell_t(\widetilde{A}_t) && (R^\top R = I_{|S'|}) \\ &= R^\top \begin{bmatrix} (P_t^{(S_t)})^{-1} & \mathbf{0}_{|S_t|\times|S'\setminus S_t|} \\ \mathbf{0}_{|S'\setminus S_t|\times|S_t|} & I_{|S'\setminus S_t|} \end{bmatrix} (R\widetilde{A}_t^{S'})\ell_t(\widetilde{A}_t) \\ &= R^\top \begin{bmatrix} z_t^{(S_t)} \\ \ell_t(\widetilde{A}_t)\cdot \widetilde{A}_t^{S'\setminus S_t} \end{bmatrix}. \end{aligned}$$

which further implies

$$(z_t^{(S')})^{S_t} = z_t^{(S_t)} \tag{4}$$

$$\text{and for any } i \in S'\setminus S_t, (z_t^{(S')})_i = \ell_t(\widetilde{A}_t)\cdot \widetilde{A}_{t,i}. \tag{5}$$

This is true even when $S' = [d]$, and we can check that $z_t^{[d]} = \hat{\ell}_t$. Therefore, one gets $z_t^{(S')} = \hat{\ell}_t^{S'}$ which proves the first statement. The second statement immediately follows from (5). □

Remark 6. This is one of the core results of Section 2.1. Checking the original OSMD algorithm Bubeck et al. (2012), one might notice that if we add an extra coordinate at time t and change the shape of the potential function from F^{S_t} to $F^{S_{t+1}}$, then the learner needs to recompute all the ellements of $\{z_\tau^{(S_{t+1})}\}_{\tau=1}^{t}$ to get A_{t+1}. This can induce a nontrivial potential change, which will be discussed in Claim 1. Thanks to Proposition C.1, we need not recompute all loss vectors again, and may simply set $z_t^{(S_t)} = \hat{\ell}_t^{S_t}$.

Next, we prove that $\hat{\ell}_t^{S_t}$ is indeed an unbiased estimator of $\ell_t^{S_t}$.

Proposition C.2. $\mathbb{E}\big[\hat{\ell}_t\big] = \ell_t$ *and* $\mathbb{E}\big[\hat{\ell}_t^{S_t} \mid S_t\big] = \ell_t^{S_t}$

Proof. From the definition of $P_t = \mathbb{E}_{A\sim p_t}[AA^\top]$ and $\widetilde{A}_t \sim p_t$, $\mathbb{E}[\hat{\ell}_t] = \ell_t$. From Proposition C.1, we can see that $\hat{\ell}_t^{S_t} = z_t^{S_t}$. Therefore,

$$\mathbb{E}\left[\hat{\ell}_t^{(S_t)} \mid S_t\right] = \mathbb{E}\left[z_t^{(S_t)} \mid S_t\right] = \mathbb{E}\left[\mathbb{E}\big[z_t^{(S_t)} \mid A_t, S_t\big] \mid S_t\right] \qquad \text{(Law of total expectation)}$$

$$
\begin{aligned}
&= \mathbb{E}\left[\mathbb{E}\big[(P_t^{(S_t)})^{-1}\widetilde{A}_t^{S_t}\widetilde{A}_t^\top \ell_t \mid A_t, S_t\big] \mid S_t\right] \\
&= \mathbb{E}\left[\mathbb{E}\Big[(P_t^{(S_t)})^{-1}\widetilde{A}_t^{S_t}\big(\langle \widetilde{A}_t^{S_t}, \ell_t^{S_t}\rangle + \langle \widetilde{A}_t^{S_t^c}, \ell_t^{S_t^c}\rangle\big) \mid A_t, S_t\Big] \mid S_t\right] \\
&= \mathbb{E}\left[\mathbb{E}\Big[(P_t^{(S_t)})^{-1}\widetilde{A}_t^{S_t}\langle \widetilde{A}_t^{S_t}, \ell_t^{S_t}\rangle \mid A_t, S_t\Big] \mid S_t\right] && (A_{t,i} \text{ for } i \in S_t^c \text{ indep. and zero mean}) \\
&= \ell_t^{S_t}
\end{aligned}
$$

where in the last step we used that $\mathbb{E}\big[\widetilde{A}_t^{S_t}(\widetilde{A}_t^{S_t})^\top \mid A_t, S_t\big] = P_t^{(S_t)}$. □

Next, we check the confidence of the coordinate identifier. We will show that if a coordinate i has not been included by time t, then the i-th components of the loss vectors have not incurred sufficiently large regret. Let $\bar{\ell}_{t,i} = \frac{1}{t}\sum_{\tau=1}^t \ell_{\tau,i}$. Let

$$
T(i) = \min\Big(\{t \in [T] : i \in S_t\} \cup \{T\}\Big)
$$

be the stopping time when the learner first include i in S_t (the union with $\{T\}$ makes $T(i)$ well defined when $i \notin S_t$ for all $t \in [T]$). Define

$$
F_i := \left\{\exists t \in [T(i)] : \left|\bar{\ell}_{t,i} - \frac{1}{t}\sum_{\tau=1}^t \hat{\ell}_{\tau,i}\right| > W_t\right\} .
$$

Proposition C.3. *For all i, $\mathbb{P}(F_i) \le \frac{1}{T^2}$.*

Proof. From Proposition C.1, $\hat{\ell}_{t,i} = \widetilde{A}_{t,i}\ell_t(\widetilde{A_t})$ if $i \notin S_t$.

Let $M_{t,i} = \sum_{j\neq i}\widetilde{A}_{t,j}\ell_{t,j}$ and $Z_{t,i} = \widetilde{A}_{t,i}M_{t,i}$. Then,

$$
\begin{aligned}
\frac{1}{t}\sum_{\tau=1}^t \hat{\ell}_{\tau,i} - \bar{\ell}_{t,i} &= \frac{1}{t}\sum_{\tau=1}^t \widetilde{A}_{\tau,i}\ell_\tau(\widetilde{A}_\tau) - \frac{1}{t}\sum_{\tau=1}^t \ell_{\tau,i} \\
&= \frac{1}{t}\left(\sum_{\tau=1}^t \widetilde{A}_{\tau,i}\Big(\sum_{j=1}^d \ell_{\tau,j}\widetilde{A}_{\tau,j}\Big) - \ell_{t,i}\right) \\
&= \frac{1}{t}\sum_{\tau=1}^t\left(\widetilde{A}_{\tau,i}\Big(\sum_{j\neq i}\ell_{\tau,j}\widetilde{A}_{\tau,j}\Big)\right) && (\widetilde{A}_{t,i}^2 = 1) \\
&= \frac{1}{t}\sum_{\tau=1}^t \widetilde{A}_{\tau,i}M_{\tau,i} && (\text{definition of } M_{\tau,i}.)
\end{aligned}
$$

Next, we will prove that $\widetilde{A}_{t,i}M_{t,i}$ is conditionally $\sqrt{2}$-subgaussian for $t \le T(i)$[1]. To do that, let $M_{t,i}^{(1)} = \sum_{j\in S_t}\widetilde{A}_{t,j}\ell_{t,j}$ and $M_{t,i}^{(2)} = \sum_{j\in S_t^c, j\neq i}\widetilde{A}_{t,j}\ell_{t,j}$. Note that $M_{t,i}^{(2)}$ is conditionally 1-subgaussian since all $\widetilde{A}_{t,j}$ are drawn independently and $\big|\sum_{j\in S_t^c, j\neq i}\widetilde{A}_{t,j}\ell_{t,j}\big| \le \|\widetilde{A}_t\|_\infty\|\ell_t\|_1 \le 1$ by Holder's inequality and $\ell_t \in \mathcal{L} = \mathcal{B}_1$. For convenience, let's assume that for drawing $\widetilde{A}_t$, we first draw $\widetilde{A}_t^{S_t}$ and then we draw $\widetilde{A}_t^{S_t^c}$. Define $\mathcal{H}_t = \sigma\big(\{\widetilde{A}_\tau\}_{\tau=1}^{t-1}, \{A_\tau\}_{\tau=1}^t, \{S_\tau\}_{\tau=1}^t\big)$ and $\mathcal{G}_t = \sigma(\widetilde{A}_t^{S_t}, \mathcal{H}_{t-1})$. Then, by closely following the derivation of (Lattimore and Szepesvári, 2020, Proof of Lemma 23.3)

$$
\begin{aligned}
\mathbb{E}\left[\exp(\lambda Z_{t,i}) \mid \mathcal{H}_{t-1}\right] &= \mathbb{E}\Big[\mathbb{E}\left[\exp\left(\lambda Z_{t,i}\right) \mid \mathcal{G}_{t-1}\right] \mid \mathcal{H}_{t-1}\Big] \\
&= \mathbb{E}\left[\exp\big(\lambda\widetilde{A}_{t,i}M_{t,i}^{(1)}\big)\mathbb{E}\left[\exp\big(\lambda\widetilde{A}_{t,i}M_{t,i}^{(2)}\big) \mid \mathcal{G}_{t-1}\right] \mid \mathcal{H}_{t-1}\right]
\end{aligned}
$$

[1]This step closely follows the statement used in (Lattimore and Szepesvári, 2020, Proof of Lemma 23.3). Alternatively, one can state it as: "Given $\mathcal{H}_{t-1}$, if $T(i) > t$, then $\widetilde{A}_{t,i}M_{t,i}$ is $\sqrt{2}$-subgaussian."

$$
\begin{aligned}
&\le \mathbb{E}\left[\exp\left(\lambda \widetilde{A}_{t,i} M_{t,i}^{(1)}\right)\exp\left(\frac{\lambda^2}{2}\right) \mid \mathcal{H}_{t-1}\right] && (M_{t,i}^{(2)} \text{ is 1-subgaussian}) \\
&= \exp\left(\frac{\lambda^2}{2}\right)\mathbb{E}\left[\mathbb{E}\left[\exp\left(\lambda \widetilde{A}_{t,i} M_{t,i}^{(1)}\right) \mid \mathcal{H}_{t-1}, M_{t,i}^{(1)}\right] \mid \mathcal{H}_{t-1}\right] \\
&\le \exp\left(\frac{\lambda^2}{2}\right)\mathbb{E}\left[\exp\left(\frac{\lambda^2 (M_{t,i}^{(1)})^2}{2}\right) \mid \mathcal{H}_{t-1}\right] && (\widetilde{A}_{t,i} \text{ is conditionally Rademacher for } t \le T(i)) \\
&\le \exp(\lambda^2). && (|M_{t,i}^{(1)}| = |\textstyle\sum_{j\in S_t} \widetilde{A}_{t,j}\ell_{t,j}| \le \|\widetilde{A}_t\|_\infty \|\ell_t\|_1 \le 1)
\end{aligned}
$$

Note that for the last inequality, we used the fact that $\ell_t \in \mathcal{L} = \mathcal{B}_1$. Therefore, one can use the Exercise 20.8 of Lattimore and Szepesvári (2020) to get the desired result. □

Remark 7. The proof above in fact extends directly to the adaptive adversary setting, since $\ell_t \in \mathcal{H}_{t-1}$.

Proposition C.4. *Under the event F_i^c, $\left|\sum_{t=1}^{T(i)} \ell_{t,i}\right| \le 8\sqrt{T(i)\log T(i)} + 1$.*

Proof. Consider the moment just before the i-th coordinate is included in S_t, namely $t = T(i) - 1$. Under F_i^c,

$$
\left|\sum_{t=1}^{T(i)-1} \ell_{t,i} - \sum_{t=1}^{T(i)-1} \hat{\ell}_{t,i}\right| < (T(i)-1) W_{T(i)-1}.
$$

Moreover, by line 14 of Algorithm 1,

$$
\frac{1}{T(i)-1}\sum_{t=1}^{T(i)-1} \hat{\ell}_{t,i} < W_{T(i)-1}.
$$

Applying the triangle inequality yields

$$
\left|\sum_{t=1}^{T(i)-1} \ell_{t,i}\right| < 2(T(i)-1)W_{T(i)-1} = 8\sqrt{(T(i)-1)\log(T(i)-1)}.
$$

Adding the final step increases the magnitude by at most 1, and thus

$$
\left|\sum_{t=1}^{T(i)} \ell_{t,i}\right| \le \left|\sum_{t=1}^{T(i)-1} \ell_{t,i}\right| + 1 < 8\sqrt{T(i)\log T(i)} + 1,
$$

which is $O\left(\sqrt{T(i)\log T(i)}\right)$ after hiding constant factors. □

Proof of Theorem 2.2. Now let's compute the regret. First, note that all zero coordinates, $[d]\backslash S$, are irrelevant for the regret computation, and we will only consider coordinates in S.

We can split the regret as follows: the first term (P1), the regret caused by identified coordinates, and the second term (P2), the regret caused by unidentified coordinates.

$$
\mathrm{Reg}_T(a) = \underbrace{\mathbb{E}\left[\sum_{t=1}^{T}\left(\widetilde{A}_t^{S_t} - a^{S_t}\right)^\top \hat{\ell}_t^{S_t}\right]}_{(P1)} + \underbrace{\mathbb{E}\left[\sum_{t=1}^{T}\left(\widetilde{A}_t^{S\backslash S_t^c} - a^{S\backslash S_t^c}\right)^\top \hat{\ell}_t^{S\backslash S_t^c}\right]}_{(P2)}.
$$

First, one can bound (P2) by $O(s\sqrt{T}\log T)$. Roughly speaking, this claim means that if a coordinate is not identified until time t, it means that the $\left|\sum_{\tau=1}^{t} \ell_{\tau,i}\right|$ is negligibly small, so the regret caused by unidentified coordinates will not contribute much on the regret bound.

Claim 1. $(P2) = O(s\sqrt{T\log T})$.

Proof. One can rewrite the regret in a coordinate-wise fashion:

$$\mathbb{E}\left[\sum_{t=1}^{T}\left(\widetilde{A}_t^{S\setminus S_t^c} - a^{S\setminus S_t^c}\right)^\top \hat{\ell}_t^{S\setminus S_t^c}\right] = \mathbb{E}\left[\sum_{i\in S}\sum_{t=1}^{T(i)}(\widetilde{A}_{t,i} - a_{t,i})\ell_{t,i}\right]$$

For each $i \in S$, by Proposition C.4, under F_i^c, we have that $\left|\sum_{t=1}^{T(i)}\ell_{t,i}\right| = O(\sqrt{T\log T})$. Therefore for $i \in S$ and for any $a \in \mathcal{A}$,

$$\begin{aligned}
\mathbb{E}\left[\mathbf{1}\left\{F_i^c\right\}\sum_{t=1}^{T(i)}(\widetilde{A}_{t,i} - a_i)\ell_{t,i}\right] &\le \mathbb{E}\left[\mathbf{1}\left\{F_i^c\right\}\sum_{t=1}^{T(i)}\widetilde{A}_{t,i}\ell_{t,i}\right] + \left|\mathbb{E}\left[\mathbf{1}\left\{F_i^c\right\}\sum_{t=1}^{T(i)}a_i\ell_{t,i}\right]\right| \\
&\le \mathbb{E}\left[\mathbf{1}\left\{F_i^c\right\}\sum_{t=1}^{T(i)}\widetilde{A}_{t,i}\ell_{t,i}\right] + \mathbb{E}\left[\mathbf{1}\left\{F_i^c\right\}\left|\sum_{t=1}^{T(i)}\ell_{t,i}\right|\right] \\
&\qquad \left(\left|\textstyle\sum_{t=1}^{T(i)}a_i\ell_{t,i}\right| = \left|a_i\textstyle\sum_{t=1}^{T(i)}\ell_{t,i}\right| \le \left|\textstyle\sum_{t=1}^{T(i)}\ell_{t,i}\right|\right) \\
&\le \mathbb{E}\left[\mathbf{1}\left\{F_i^c\right\}\sum_{t=1}^{T(i)}\widetilde{A}_{t,i}\ell_{t,i}\right] + O(\sqrt{T\log T}) && \text{(Proposition C.4)} \\
&\le \mathbb{E}\left[\mathbf{1}\left\{F_i^c\right\}\left((T(i)-1)W_{T(i)-1}+1\right)\right] + O(\sqrt{T\log T}) && \text{(Definition of } T(i) \text{ and } F_i\text{)} \\
&= O(\sqrt{T\log T}) + O(\sqrt{T\log T}) && \text{(Proposition C.4)} \\
&= O(\sqrt{T\log T}). && (6)
\end{aligned}$$

Therefore, $\mathbb{E}\left[\sum_{i\in S}\sum_{t=1}^{T(i)}(\widetilde{A}_{t,i} - a_{t,i})\ell_{t,i}\right] = O(s\sqrt{T\log T})$ and the claim follows. □

Now, let's compute (P1) by the standard Stochastic Online Mirror Descent analysis in Bubeck et al. (2012). Let $\hat{L}_t = \sum_{\tau=1}^{t}\hat{\ell}_\tau$. For any convex function f, let D_f be the Bregman Divergence induced by f. Then,

$$\begin{aligned}
&-\eta\sum_{t=1}^{T}a^\top\hat{\ell}_t \le F(a) + F^*(-\eta\sum_{t=1}^{T}\hat{\ell}_t) \\
&\le F(a) + F^{\emptyset,*}(0) + \sum_{i=0}^{|S_T|}\left(F^{S_{\tau_{i+1}},*}(-\eta\hat{L}_{\tau_i}) - F^{S_{\tau_i},*}(-\eta\hat{L}_{\tau_i}) + \sum_{t=\tau_i+1}^{\tau_{i+1}}\left(F^{S_t,*}(-\eta\hat{L}_t) - F^{S_t,*}(-\eta\hat{L}_{t-1})\right)\right) \\
&= F(a) + \underbrace{\sum_{i\in[|S_T|]}F^{S_{\tau_{i+1}},*}(-\eta\hat{L}_{\tau_i}) - F^{S_{\tau_i},*}(-\eta\hat{L}_{\tau_i})}_{(PC)} + \sum_{i=0}^{|S_T|}\sum_{t=\tau_i+1}^{\tau_{i+1}}\left(F^{S_t,*}(-\eta\hat{L}_t) - F^{S_t,*}(-\eta\hat{L}_{t-1})\right) \quad (F^{\emptyset,*}(0)=0) \\
&= F(a) + (PC) + \sum_{i=0}^{|S_T|}\sum_{t=\tau_i+1}^{\tau_{i+1}}\left(-\eta A_t^{S_t}\hat{\ell}_t + D_{F^{S_t,*}}(-\eta\hat{L}_t, -\eta\hat{L}_{t-1})\right). \qquad \text{(Bergman Divergence)}
\end{aligned}$$

By reordering, one can achieve the following estimate for $(P1)$.

$$\eta\sum_{t=1}^{T}(A_t^{S_t} - a^{S_t})^\top\hat{\ell}_t \le F(a) + \underbrace{\eta\sum_{t=1}^{T}\langle a^{S_t^c}, \hat{\ell}_t\rangle}_{(RA)} + (PC) + \underbrace{\sum_{i=0}^{|S_T|}\sum_{t=\tau_i+1}^{\tau_{i+1}}\left(D_{F^{S_t,*}}(-\eta\hat{L}_t, -\eta\hat{L}_{t-1})\right)}_{(V)} \qquad (7)$$

Calculating (RA) By Proposition C.3 and Claim 1, under event $\overline{F}_i$ we have that $|\sum_{t=1}^{T(i)} \ell_{t,i}| = O(\sqrt{T(i)\log T(i)})$. Therefore,

$$\begin{aligned}
\mathbb{E}[(RA)] &= \mathbb{E}\left[\sum_{i\in S}\sum_{t=1}^{T(i)} \eta a_i \hat{\ell}_{t,i}\right] && \text{(Reordering the sum)} \\
&= \mathbb{E}\left[\sum_{i\in S}\sum_{t=1}^{T(i)} \eta a_i \ell_{t,i}\right] + \mathbb{E}\left[\sum_{i\in S}\sum_{t=1}^{T(i)} \eta a_i (\hat{\ell}_{t,i} - \ell_{t,i})\right] \\
&\le \mathbb{E}\left[\sum_{i\in S}\left|\sum_{t=1}^{T(i)} \eta \ell_{t,i}\right|\right] + \mathbb{E}\left[\sum_{i\in S}\left|\sum_{t=1}^{T(i)} \eta (\hat{\ell}_{t,i} - \ell_{t,i})\right|\right] && (|a_i| \le 1) \\
&\le \mathbb{E}\left[\mathbf{1}\left\{\cap_{i\in S} F_i^c\right\}\sum_{i\in S} 8\eta\sqrt{T(i)\log T(i)}\right] + \eta\mathbb{E}\left[\mathbf{1}\left\{\cap_{i\in S} F_i^c\right\}\sum_{i\in S}\sum_{t=1}^{T(i)} W_t\right] + O(\eta s) \\
& && \text{(Claim 1 and Proposition C.3)} \\
&\le \mathbb{E}\left[\mathbf{1}\left\{\cap_{i\in S} F_i^c\right\}\sum_{i\in S} 8\eta\sqrt{T(i)\log T(i)}\right] + 4\eta\mathbb{E}\left[\mathbf{1}\left\{\cap_{i\in S} F_i^c\right\}\sum_{i\in S} \sqrt{T(i)\log T}\right] + O(\eta s) \\
& && \left(\textstyle\sum_{t=1}^{T(i)} W_t \le \sum_{t=1}^{T(i)} 4\sqrt{\frac{2\log T}{t}} \le 8\sqrt{T(i)\log T}\right) \\
&= O(\eta s\sqrt{T\log T}). && (T(i) \le T)
\end{aligned}$$

Calculating (PC) By some computation, one can check that

$$\begin{aligned}
F^S(x) &= \sum_{i\in S}\int_0^{x_i} \tanh^{-1}(y)dy \\
F^{S,*}(x) &= \sum_{i\in S}\int_0^{x_i} \tanh(y)dy.
\end{aligned}$$

which means $F^{S,*}$ has a coordinatewisely independent form. Therefore by Proposition C.1,

$$F^{S_{\tau_i+1},*}(-\eta\hat{L}_{\tau_i}) - F^{S_{\tau_i},*}(-\eta\hat{L}_{\tau_i}) = F^{\{c_i\},*}(-\eta\hat{L}_{\tau_i}) = F_0^*(-\eta\hat{L}_{\tau_i,c_i}).$$

Now let's see how large $(-\eta\hat{L}_{\tau_i,c_i})$ will be. Again,

$$\hat{L}_{\tau_i,c_i} = \sum_{t=1}^{\tau_i}\hat{\ell}_{t,c_i} \le \tau_i \cdot W_{\tau_i} = O(\sqrt{\tau_i\log\tau_i})$$

This means $|\eta\hat{L}_{\tau_i,c_i}| \le O(\eta\sqrt{\tau_i\log\tau_i})$. If $\eta \le \frac{1}{\sqrt{T}}$, $|\eta\hat{L}_{\tau_i,c_i}| \le O(\sqrt{\log T})$, and therefore

$$\begin{aligned}
F_0^*(-\eta\hat{L}_{t,c_i}) &\le \int_0^{|\eta\hat{L}_{\tau_i,c_i}|}\tanh(x)dx \\
&\le |\eta\hat{L}_{\tau_i,c_i}| = O(\sqrt{\log T}). && (|\tanh(x)| \le 1)
\end{aligned}$$

which implies $(PC) \le O(s\sqrt{\log T})$.

Calculating (V) As proven in Bubeck et al. (2012), we have that $D_{F^{S_t},*}(u,v) = \sum_{i\in S_t}(1-\tanh^2(v_i))(u_i-v_i)^2$ when $\|u-v\|_\infty \le 1$. Now substitute $u=-\eta\hat{L}_t$ and $v=-\eta\hat{L}_{t-1}$,

$$\begin{aligned} D_{F^{S_t},*}(-\eta\hat{L}_t,-\eta\hat{L}_{t-1}) &\le \sum_{i\in S_t}(1-\tanh^2(\eta\hat{L}_{t-1,i}))(-\eta\hat{\ell}_{t,i})^2 \\ &= \eta^2\sum_{i\in S_t}(1-A_{t,i}^2)\hat{\ell}_{t,i}^2. && (A_t = \nabla F^{S_t,*}(-\eta\hat{L}_{t-1}) = \tanh(-\eta\hat{L}_{t-1})) \end{aligned} \tag{8}$$

Now, to simplify even more, note that

$$\begin{aligned} P_t^{(S_t)} &= \mathbb{E}_{\widetilde{A}_t\sim p_t}\left[\widetilde{A}_t^{S_t}\widetilde{A}_t^{S_t\top}\right] = \gamma_t I_{|S_t|} + (1-\gamma_t)\sum_{i,j\in S_t}\mathbb{E}[\widetilde{A}_{t,i}\widetilde{A}_{t,j}]e_ie_j^\top \\ &= \gamma_t I_{|S_t|} + (1-\gamma_t)I_{|S_t|} + (1-\gamma_t)\sum_{i\neq j, i,j\in S_t}\mathbb{E}[\widetilde{A}_{t,i}\widetilde{A}_{t,j}]e_ie_j^\top \\ &= \gamma_t I_{|S_t|} + (1-\gamma_t)I_{|S_t|} + (1-\gamma_t)\sum_{i\neq j, i,j\in S_t}A_{t,i}A_{t,j}e_ie_j^\top \\ &= \gamma_t I_{|S_t|} + (1-\gamma_t)A_t^{S_t}(A_t^{S_t})^\top + (1-\gamma_t)\sum_{i\in S_t}(1-A_{t,i}^2)e_ie_i^\top. \end{aligned}$$

This means

$$P_t^{(S_t)} \succeq \gamma_t I_{|S_t|}. \tag{9}$$

$$P_t^{(S_t)} \succeq (1-\gamma_t)\sum_{i\in S_t}(1-A_{t,i}^2)e_ie_i^\top. \tag{10}$$

First, let's check $\|\hat{\ell}_t^{S_t}\|_\infty$,

$$\begin{aligned} ||\hat{\ell}_t^{S_t}||_\infty &= \|(P_t^{(S_t)})^{-1}\widetilde{A}_t^{S_t}\widetilde{A}_t^\top\ell_t\|_\infty \\ &\le \|(P_t^{(S_t)})^{-1}\widetilde{A}_t^{S_t}\|_\infty && (|\widetilde{A}_t\ell_t|\le 1) \\ &\le \max_{i\in[|S_t|]}|e_i^\top(P_t^{(S_t)})^{-1}\widetilde{A}_t^{S_t}| \\ &\le \frac{\sqrt{|S_t|}}{\lambda_{\min}(P_t^{(S_t)})} && (\|\widetilde{A}_t^{S_t}\|_2\le\sqrt{|S_t|}) \\ &\le \frac{\sqrt{|S_t|}}{\gamma_t}. && \text{(Eq. (9))} \end{aligned}$$

To satisfy the conditions for $D_{F^{S_t},*}(u,v)$ before, one should satisfy $\|\eta\hat{\ell}_t^{S_t}\|_\infty \le 1$, and setting $\gamma_t = \sqrt{\frac{|S_t|}{T}}$ will satisfy the condition.

After that, from Eq.(10) one can get

$$(1-\gamma)\sum_{j\in S_t}\mathbb{E}[(1-A_{t,j}^2)\hat{\ell}_{t,j}^2] \le \mathbb{E}[(\hat{\ell}_t^{S_t})^\top P_t^{(S_t)}\hat{\ell}_t^{S_t}].$$

To bound $(\hat{\ell}_t^{S_t})^\top P_t^{(S_t)}\hat{\ell}_t^{S_t}$,

$$\begin{aligned} \mathbb{E}[\hat{\ell}_t^{(S_t)\top}P_t^{(S_t)}\hat{\ell}_t^{(S_t)}|S_t] &= \mathbb{E}[\ell_t^\top\widetilde{A}_t\widetilde{A}_t^{S_t\top}(P_t^{(S_t)})^{-1}P_t^{(S_t)}(P_t^{(S_t)})^{-1}\widetilde{A}_t^{S_t}\widetilde{A}_t^\top\ell_t|S_t] \\ &\le \mathbb{E}[\widetilde{A}_t^{S_t\top}(P_t^{(S_t)})^{-1}\widetilde{A}_t^{S_t}|S_t] && (|a^\top\ell_t|\le 1 \text{ for all } a\in\mathcal{A}) \\ &= \mathbb{E}[\mathrm{Tr}((P_t^{(S_t)})^{-1}\widetilde{A}_t^{S_t\top}\widetilde{A}_t^{S_t})|S_t] && \text{(Commutativity of Trace)} \end{aligned}$$

$$= \mathrm{Tr}(\mathbb{E}[(P_t^{(S_t)})^{-1}\widetilde{A}_t^{S_t\top}\widetilde{A}_t^{S_t}|S_t]) \qquad \text{(Linearity of Trace)}$$
$$= \mathbb{E}[I_{|S_t|}|S_t] = |S_t| < s. \qquad (P_t^{(S_t)} = \mathbb{E}[A_t^{(S_t)}A_t^{(S_t)}])$$

This leads to the following upper-bound for (V):

$$\begin{aligned}
\mathbb{E}[(\mathrm{V})] &= \mathbb{E}\left[\sum_{i=1}^{|S_T|}\sum_{t=\tau_i+1}^{\tau_{i+1}} D_{F^{S_t},*}(-\eta\hat{L}_t, -\eta\hat{L}_{t-1})\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T} D_{F^{S_t},*}(-\eta\hat{L}_t, -\eta\hat{L}_{t-1})\right] \\
&\le \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i\in S_t}\eta^2(1-A_{t,i}^2)\hat{\ell}_{t,i}^2\right] && \text{(Eq. (8))} \\
&= \mathbb{E}\left[\sum_{t=1}^{T}(\hat{\ell}_t^{(S_t)})^\top[\eta^2(\sum_{i\in S_t}(1-A_{t,i}^2)e_ie_i^\top]\hat{\ell}_t^{(S_t)}\right] \\
&\le \mathbb{E}\left[\sum_{t=1}^{T}(\hat{\ell}_t^{(S_t)})^\top[\frac{\eta^2}{1-\gamma_t}P_t^{(S_t)}]\hat{\ell}_t^{(S_t)}\right] && \text{(Eq. (10))} \\
&\le \mathbb{E}\left[\sum_{t=1}^{T}(\hat{\ell}_t^{(S_t)})^\top[2P_t^{(S_t)}]\hat{\ell}_t^{(S_t)}\right] && (T \gg d > |S_t|) \\
&\le 2\eta^2 sT
\end{aligned}$$

Putting things together Overall, we can simplify Eq. (7) as

$$\mathbb{E}\left[\eta\sum_{t=1}^{T}(A_t^{S_t} - a^{S_t})\hat{\ell}_t\right] \le F(a) + O(\eta s\sqrt{T\log T}) + O(s\sqrt{\log T}) + 2\eta^2 sT$$

Now setting $\eta = \frac{1}{\sqrt{T}}$, using $|S_t| \le s$ and $F(a) = 2s\ln 2$ will lead

$$\mathbb{E}\left[\sum_{t=1}^{T}(A_t^{S_t} - a^{S_t})\hat{\ell}_t^{S_t}\right] = O(s\sqrt{T\log T}). \tag{11}$$

By Proposition C.2, $\hat{\ell}_t$ is an unbiased estimator of ℓ_t, so

$$\mathbb{E}\left[\sum_{t=1}^{T}(A_t^{S_t} - a^{S_t})\hat{\ell}_t^{S_t}\right] = \mathbb{E}\left[\sum_{t=1}^{T}(A_t^{S_t} - a^{S_t})\ell_t^{S_t}\right].$$

Taking into $\gamma_t < \sqrt{\frac{s}{T}}$ uniform random selection into account, one can check

$$\mathbb{E}\left[\sum_{t=1}^{T}(\widetilde{A}_t^{S_t} - a^{S_t})\hat{\ell}_t^{S_t}\right] \le \sqrt{\frac{s}{T}}T + \mathbb{E}\left[\sum_{t=1}^{T}(A_t^{S_t} - a^{S_t})\ell_t^{S_t}\right] = O(s\sqrt{T\log T}).$$

Now, summing up Eq. (11) with Eq. (6) leads to

$$\begin{aligned}
\mathrm{Reg}_T(a) &= \mathbb{E}\left[\sum_{t=1}^{T}(\widetilde{A}_t^{S_t} - a^{S_t})\ell_t^{S_t}\right] + \mathbb{E}\left[\sum_{t=1}^{T}(\widetilde{A}_t^{S_t^c} - a^{S_t^c})\ell_t^{S_t^c}\right] \\
&= O(s\sqrt{T\log T}).
\end{aligned}$$

□

Remark 8. Combining the fact that Online Mirror Descent is in general applicable to adaptive adversaries with the observation that the key concentration step, Proposition C.3, also remains valid in the adaptive adversary setting, we conclude that Algorithm 1 achieves the same regret bound against adaptive adversaries.

C.2 Proof of Proposition 2.1

The proof of Proposition 2.1 largely follows the approach of Dani et al. (2007), with a slight modification. We begin by stating the following lemma, which is required for our calculation.

Lemma C.5. *Let P and Q be two Bernoulli distributions, where $p = P(1)$ and $q = Q(1)$. Then, we can compute the chi-square distance between P and Q as*

$$\chi^2(Q, P) = (p - q)^2 \left(\frac{1}{p} + \frac{1}{1-p}\right).$$

Now, we proceed with the proof of Proposition 2.1.

Consider a uniformly random subset $S \subseteq [d]$, and let $\zeta \in \{-1, 1\}$ be 1 with probability $\frac{1+\varepsilon}{2}$ and -1 with probability $\frac{1-\varepsilon}{2}$. Choose i uniformly at random. Define the loss for each round as:

$$\ell_t = \begin{cases} \zeta e_i, & i \in S; \\ -\zeta e_i, & i \notin S. \end{cases} \tag{12}$$

Let $P_{S,a}$ represent the distribution of $\ell_t(a)$ given set S. Note that for all $S \subset [d]$ and $a \in \mathcal{A}$, we have $\operatorname{supp}(P_{S,a}) = \{-1, 1\}$. For a set $S \subseteq [d]$ and element $i \in S$, we have:

$$\begin{cases} P_{S,a}(\ell_t(a) = 1 \mid i_t \neq i) - P_{S\setminus\{i\},a}(\ell_t(a) = 1 \mid i_t \neq i) & = 0; \\ P_{S,a}(\ell_t(a) = 1 \mid i_t = i) - P_{S\setminus\{i\},a}(\ell_t(a) = 1 \mid i_t = i) & = \varepsilon. \end{cases} \tag{13}$$

Therefore,

$$P_{S,a}(\ell_t(a) = 1) - P_{S\setminus\{i\},a}(\ell_t(a) = 1) = \frac{\varepsilon}{d}.$$

Moreover, suppose $\varepsilon \leq \frac{1}{2}$, then for all S and a, $1/4 \leq P_{S,a}(\ell_t(a) = 1) \leq 3/4$. Using Lemma C.5, we have

$$\chi^2(P_{S,a}, P_{S\setminus\{i\},a}) \leq 8\left(\frac{\varepsilon}{d}\right)^2.$$

Let $\mathbb{P}_S$ denote the distribution induced by a chosen algorithm and environment defined by S. We have:

$$\mathrm{KL}(\mathbb{P}_S, \mathbb{P}_{S\setminus\{i\}}) = \mathbb{E}_S\left[\sum_{t=1}^{T} \mathrm{KL}(P_{S,a_t}, P_{S\setminus\{i\},a_t})\right] \leq \mathbb{E}_S\left[\sum_{t=1}^{T} \chi^2(P_{S,a_t}, P_{S\setminus\{i\},a_t})\right] \leq 8T\frac{\varepsilon^2}{d^2}. \tag{14}$$

Let $\mathbf{1}_S \in \mathbb{R}^d$ be the signed indicator vector for a subset $S \subseteq [d]$, defined by $\mathbf{1}_{S,i} = 1$ if $i \in S$ and $\mathbf{1}_{S,i} = -1$ if $i \notin S$. Define $p_S^i = \mathbb{P}_S\left(\sum_{t=1}^{T} \mathbb{I}\{\operatorname{sign}(a_{t,i}) \neq \operatorname{sign}(\mathbf{1}_{S,i})\} \geq T/2\right)$. Choose $\varepsilon = \frac{d}{\sqrt{T}}$, then by the Bretagnolle–Huber inequality, we have

$$p_S^i + p_{S\setminus\{i\}}^i \geq \frac{1}{2}\exp\left(-\mathrm{KL}(\mathbb{P}_S, \mathbb{P}_{S\setminus\{i\}})\right) \geq \frac{1}{2}\exp\left(-T\frac{\varepsilon^2}{d^2}\right) \geq \frac{1}{2}\exp(-8).$$

We have that

$$\sum_{S\subseteq[d]} \frac{1}{2^d} \sum_{i=1}^{d} p_S^i = \sum_{i=1}^{d} \frac{1}{2^d} \sum_{S\subseteq[d]} p_S^i \geq \frac{d}{4}\exp(-8). \tag{15}$$

Therefore, there must exist $S \subseteq [d]$ such that $\sum_{i=1}^{d} p_S^i = \Omega(d)$. Note that if $\text{sign}(a_{t,i}) \neq \text{sign}(\mathbf{1}_{S,i})$, the algorithm incurs regret $\frac{\varepsilon}{d} = \frac{1}{\sqrt{T}}$. As a result, similar as Lattimore and Szepesvári (2020) (Theorem 24.1), we can lower bound the regret for the bandit instance defined by S as follows:

$$\begin{aligned}
\text{Reg}_T &= \mathbb{E}_S\left[\sum_{t=1}^{T}\sum_{i=1}^{d}(\text{sign}(\mathbf{1}_{S,i}) - a_{t,i})\right] \\
&= \frac{1}{\sqrt{T}}\sum_{i=1}^{d}\mathbb{E}_S\left[\sum_{t=1}^{T}\mathbb{I}\{\text{sign}(\mathbf{1}_{S,i}) \neq a_{t,i}\}\right] \\
&\geq \frac{\sqrt{T}}{2}\sum_{i=1}^{d}\mathbb{P}_S\left(\sum_{t=1}^{T}\mathbb{I}\{\text{sign}(\mathbf{1}_{S,i}) \neq a_{t,i}\} \geq T/2\right) && \text{(By Markov inequality)} \\
&\geq \frac{\sqrt{T}}{2}\sum_{i=1}^{d} p_S^i \\
&\geq \frac{d\sqrt{T}}{8}\exp(-8).
\end{aligned}$$

□

C.3 Complexity analysis

Time complexity At each iteration, the most computationally demanding step is the matrix inversion of (P_t^{-1}) (Line 11 in Algorithm 1). When a new coordinate is added to S_t (i.e., when $S_{t+1} \neq S_t$, one can avoid recomputing a full matrix inversion for $(P_t^{(S_t)})^{-1}$ by simply extracting the corresponding submatrix. Similarly, when computing the loss vector, it is unnecessary to recompute it at every round; instead, by using Proposition C.1, we can maintain only $\sum_t \hat{\ell}_t$ across time and, whenever S_{t+1} is updated, project $\sum_t \hat{\ell}_t$ onto $\mathbb{R}^{S_{t+1}}$ to compute $A_{t+1}^{S_{t+1}}$ efficiently. Consequently, only $O(d^3)$ operation is required at each round for the inversion of P_t, leading to an overall computational complexity of at most $O(d^3T)$.

Space complexity Again, thanks to the Proposition C.1, the temporary storage of the inverse matrix, which requires $O(d^2)$, and the storage of $\sum_t z_t^{[d]}$, which requires $O(d)$, are the main space complexity we need to spend. Hence, the total space complexity is $O(d^2)$. (Note that, by Proposition , P_t is used only once for computing $\hat{\ell}_t$, and thus it does not need to be stored across time; the only quantity that must be maintained is $\sum_t \hat{\ell}_t$).

D Proof of Subsection 2.2

D.1 Proof of Proposition 2.4

The proof is essentially similar as Bubeck et al. (2012), with slight modification due to the coordinate-weight update. F is strictly convex, differentiable and,

$$\nabla F(x) = \frac{x}{1 - \|x\|_2}.$$

First, one can bound the regret of mirror descent on $\mathcal{A}'$ Bubeck et al. (2012) as

$$\text{Reg}_T \leq \gamma T + \frac{\sup_{a \in \mathcal{A}'} F(a) - F(A_1)}{\eta} + \frac{1}{\eta}\sum_{t=1}^{T}\mathbb{E}\left[D_{F^*}\left(\nabla F(A_t) - \eta\hat{\ell}_t, \nabla F(A_t)\right)\right].$$

The second term is clearly bounded by $\frac{\log(\gamma^{-1})}{\eta}$.

Now we bound the third term. With this choice of F, we have that

$$\begin{aligned}\nabla F^*(u) &= \frac{u}{1+\|u\|_2}\\ F^*(u) &= -\log(1+\|u\|_2)+\|u\|_2\\ D_{F^*}(u,v) &= \frac{1}{1+\|v\|_2}\underbrace{\left(\|u\|_2-\|v\|_2+\|u\|_2\|v\|_2-v^\top u-(1+\|v\|_2)\log\left(1+\frac{\|u\|_2-\|v\|_2}{1+\|v\|_2}\right)\right)}_{=:G(u,v)}.\end{aligned} \tag{16}$$

Now, we need to show that $G(u,v) \le \|u-v\|_2^2$, for $(u,v) = (-\eta\sum_{\tau=1}^{t}\hat{\ell}_\tau, -\eta\sum_{\tau=1}^{t-1}\hat{\ell}_\tau)$. The above is immediately true if $\frac{\|u\|_2-\|v\|_2}{1+\|v\|_2} \ge -\frac{1}{2}$. Note that,

$$\frac{1}{1+\|v\|_2} = \frac{1}{1+\|-\eta\sum_{\tau=1}^{t-1}\hat{\ell}_\tau\|_2} = \frac{1}{1+\|\nabla F(A_t)\|_2} = (1-\|A_t\|_2).$$

Below, we prove that it is indeed true for the pairs under consideration.

$$\frac{\|u\|_2-\|v\|_2}{1+\|v\|_2} \ge -\frac{\eta\hat{\ell}_t}{1+\|v\|_2} = -\eta(1-\|A_t\|_2)\hat{\ell}_t. \tag{17}$$

Until now, the proof is identical as that of Bubeck et al. (2012). Below is the only difference. Consider the case where $\xi_t = 0$ (otherwise $\hat{\ell}_t = 0$).

$$\hat{\ell}_t = \frac{1}{1-\|A_t\|_2}\frac{\ell_{t,I_t}e_{I_t}}{\rho_t(I_t)}.$$

There are two cases:

1. If index $\ell_{t,I_t} = 0$, then $\hat{\ell}_t = 0$, we have $-\eta\|\hat{\ell}_t\|_2 \ge -1/2$.

2. Consider the case index $\ell_{t,I_t} \ne 0$. Note that $w_t(I_t) \le 2sd$ and $\rho_t(I_t) \ge \frac{1}{2sd}$. Therefore, with the choice of $\eta = \frac{1}{4\sqrt{sT}}$, and note that $\sqrt{T} \ge d$,

$$-\eta(\|1-A_t\|_2)\|\hat{\ell}_t\|_2 = \frac{-\eta}{\rho_t(I_t)}\ell_{t,I_t} \ge -\frac{2sd}{4\sqrt{sT}} \ge -1/2.$$

Consequently, $G(u,v) \le \|u-v\|_2$ for the pairs under consideration. With it, we conclude the proof.

□

D.2 Proof of Lemma 2.5

The proof of Lemma 2.5 requires the following result.

Lemma D.1. *Let x be Bernoulli random variable with probability p of getting 1. We sample x i.i.d., and denote N_x as a random variable indicating the number of samples until one get $x = 1$ the first time. Then, $\mathbb{E}[N_x] = 1/p$.*

Proof.

$$\begin{aligned}\mathbb{E}[N_x] &= \sum_{k=1}^{\infty} k\mathbb{P}(N_x = k) = \sum_{k=1}^{\infty} k(1-p)^{k-1}p\\ &= p\sum_{k=1}^{\infty} k(1-p)^{k-1} = p\frac{1}{(1-(1-p))^2} = \frac{1}{p}.\end{aligned} \tag{18}$$

□

Now we proceed the proof of Lemma 2.5. First, we have that

$$\hat{\ell}_t = (1-\xi_t)\frac{1}{1-\|A_t\|_2}\frac{1}{\rho_t(I_t)}\widetilde{A}_t\ell_t(\widetilde{A}_t).$$

Therefore,

$$\begin{aligned}
\mathbb{E}\left[\|\hat{\ell}_t\|^2 \mid A_t\right] &= \mathbb{E}\left[(1-\xi_t)^2\right]\mathbb{E}\left[\frac{\ell_{t,I_t}^2}{(1-\|A_t\|_2)^2\rho_t^2(I_t)} \mid A_t\right] && \text{(As } \xi_t \text{ is sampled independently)}\\
&= (1-\|A_t\|_2)\mathbb{E}\left[\frac{\ell_{t,I_t}^2}{(1-\|A_t\|_2)^2\rho_t^2(I_t)} \mid A_t\right]\\
&= \mathbb{E}\left[\sum_{i=1}^d \frac{\rho_t(i)\ell_{t,i}^2}{(1-\|A_t\|_2)\rho_t^2(i)} \mid A_t\right]\\
&= \mathbb{E}\left[\sum_{i=1}^d \frac{\ell_{t,i}^2}{(1-\|A_t\|_2)\rho_t(i)} \mid A_t\right].
\end{aligned}$$

Therefore, we have that

$$\mathbb{E}\left[\mathbb{E}\left[(1-\|A_t\|_2)\|\hat{\ell}_t\|_2^2 \mid A_t\right]\right] = \mathbb{E}\left[\sum_{i=1}^d \frac{\ell_{t,i}^2}{\rho_t(i)}\right]. \tag{19}$$

Thus,

$$\sum_{t=1}^T \mathbb{E}\left[(1-\|A_t\|_2)\|\hat{\ell}_t\|_2^2\right] = \mathbb{E}\left[\sum_{t=1}^T\sum_{i=1}^d \frac{\ell_{t,i}^2}{\rho_t(i)}\right] = \mathbb{E}\left[\sum_{t=1}^T\sum_{i\in S} \frac{\ell_{t,i}^2}{\rho_t(i)}\right]. \tag{20}$$

Suppose that the adversary has fixed an infinite sequence of losses $(\ell_t)_{t=1}^{\infty}$, but we only use the first T losses to calculate the regret. Let T_i be the first round coordinate $i \in S$ is recognised, that is, $T_i = \min\{t \in \mathbb{N}^+ \mid \xi_t = 0 \text{ and } I_t = i \text{ and } \ell_{t,i} \neq 0\}$. Define the total weight $Z_t = \sum_{i=1}^d w_t(i)$. With the choice of $H = d$, we can upper-bound $Z_{T_i} \leq 2sd$ for all $i \in [S]$.

For any $i \in S$, we first show that

$$\mathbb{E}\left[\sum_{t=1}^T \frac{\ell_{t,i}^2}{\rho_t(i)}\right] \leq \underbrace{\mathbb{E}\left[\sum_{t=1}^{T_i-1} \frac{\ell_{t,i}^2}{\rho_t(i)}\right]}_{(\star)} + 2s\sum_{t=1}^T \ell_{t,i}^2. \tag{21}$$

To see this, consider two cases. The first case is that $T < T_i$, which means that coordinate i is not identified, it is clear that

$$\mathbb{E}\left[\sum_{t=1}^T \frac{\ell_{t,i}^2}{\rho_t(i)}\right] \leq \mathbb{E}\left[\sum_{t=1}^{T_i-1} \frac{\ell_{t,i}^2}{\rho_t(i)}\right].$$

Therefore, (21) holds. The second case is that $T \geq T_i$, and note that $\rho_t(i) \geq 1/2s$ for $t \geq T_i$. We have that

$$\mathbb{E}\left[\sum_{t=1}^T \frac{\ell_{t,i}^2}{\rho_t(i)}\right] = \mathbb{E}\left[\sum_{t=1}^{T_i-1} \frac{\ell_{t,i}^2}{\rho_t(i)}\right] + \mathbb{E}\left[\sum_{t=T_i}^T \frac{\ell_{t,i}^2}{\rho_t(i)}\right] \leq \mathbb{E}\left[\sum_{t=1}^{T_i-1} \frac{\ell_{t,i}^2}{\rho_t(i)}\right] + 2s\sum_{t=1}^T \ell_{t,i}^2.$$

Therefore, (21) holds for any value of T_i.

Now, we bound the term $(\star)$. Let $N_i(t) = \sum_{\tau=1}^t \mathbb{I}(\ell_{\tau,i} \neq 0)$ be the number of time index i has non-zero loss, and

$\tilde{T}_j(i) = \min\{t \mid N_i(t) \geq j\}$.

$$\mathbb{E}\left[\sum_{t=1}^{T_i-1} \frac{\ell_{t,i}^2}{\rho_t(i)}\right] = \mathbb{E}\left[\sum_{j=1}^{N_i(T_i-1)} \frac{\ell_{\tilde{T}_j(i),i}^2}{\rho_{\tilde{T}_j(i)}(i)}\right] \leq \mathbb{E}\left[\sum_{j=1}^{N_i(T_i-1)} Z_{\tilde{T}_j(i)}\right] \leq \mathbb{E}[N_i(T_i-1)](2sd). \tag{22}$$

Note that, for $t \leq T_i$, the probability of choosing arm $\widetilde{A}_t = e_i$ is

$$\mathbb{P}(I_t = i \,\cap\, \xi_t = 0) \geq \frac{\gamma}{Z_{T_i}} = \frac{1}{\sqrt{T} Z_{T_i}}.$$

Consequently, By Lemma D.1, we can bound $\mathbb{E}[N_i(t)] \leq \sqrt{T} Z_{T_i} \leq 2sd\sqrt{T}$. As a result, we have that

$$\mathbb{E}\left[\sum_{t=1}^{T_i-1} \frac{\ell_{t,i}^2}{\rho_t(i)}\right] \leq \mathbb{E}[N_i(T_i-1)](2sd) \leq 4(sd)^2\sqrt{T}. \tag{23}$$

Summing up, we have that,

$$\mathbb{E}\left[\sum_{t=1}^{T} \frac{\ell_{t,i}^2}{\rho_t(i)}\right] \leq 4(sd)^2\sqrt{T} + 2s\sum_{t=1}^{T} \ell_{t,i}^2. \tag{24}$$

Putting things together, we can bound the total variance term as follows.

$$\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i\in S} \frac{\ell_{t,i}^2}{\rho_t(i)}\right] \leq 4s^3d^2\sqrt{T} + 2s\sum_{t=1}^{T}\sum_{i\in S} \ell_{t,i}^2 \leq 4s^3d^2\sqrt{T} + 2sT. \tag{25}$$

□

D.3 Proof of Theorem 2.3

Combining Proposition 2.4 and Lemma 2.5, we obtain the regret upper bound as

$$\mathrm{Reg}_T \leq \gamma T + \frac{\log(\gamma^{-1})}{\eta} + 4\eta s^3 d^2\sqrt{T} + 2\eta sT.$$

Choose γ and η as stated in theorem, we obtain the regret bound

$$\mathrm{Reg}_T \leq \sqrt{T} + 4\sqrt{sT}\log(T) + s^{2.5}d^2 + \sqrt{sT} \leq 5\sqrt{sT}\log(T) + s^{2.5}d^2.$$

□

D.4 Complexity analysis

Time complexity `OSMDSE` is even more efficient than `MiDeSiCoE`. Since $\mathcal{A}'$ is a Euclidean ball with radius $1-\gamma$, projection can be done efficiently by simply dividing the input vector by its norm. Except that, all computations can be done within $O(d)$ time, so overall time complexity for each round is $O(d)$, and the overall time complexity of `OSMDSE` is $O(dT)$.

Space complexity The parameter `OSMDSE` need to store at each round are d-dimensional vectors, such as $\tilde{A}$, ρ_t, $\hat{\ell}_t$ and $\sum_t \hat{\ell}_t$. These parameter only spend $O(d)$ spaces, and we only need to store the most recent parameter (for example, we don't need to store ρ_3 when $t > 3$). Overall, the space complexity of this algorithm is $O(d)$.

E Proof of Section 3

Proof. Before we proceed, we need to check the confidence width of $\tilde{\theta}_{t,i}$.

Proposition E.1. *Let* $W_t = 4\sqrt{\frac{1}{\sum_{\tau=1}^t B_\tau^2}\log(t+1)}$ *and define* $F_i := \{\exists t \in [T(i)] \quad s.t. \quad |\tilde{\theta}_{t,i} - \theta_i^*| > W_t\}$*. Then,* $\mathbb{P}(F_i) \leq \frac{1}{T}$.

Remark 9. This proposition implies that under $\cap_{i=1}^T F_i^c$, $S_t \subset S$, since for $i \in S^c$, under the event F_i^c, $i \notin S_T$.

Proof. The proof structure is quite similar to Proposition C.3. Let $M_{t,i} = \sum_{j\neq i} \widetilde{A}_{t,j}\theta_j^*$ and $Z_{t,i} = \widetilde{A}_{t,i}(M_{t,i} + \epsilon_t)$. Then,

$$\begin{aligned}
\tilde{\theta}_{t,i} - \theta_i^* &= \frac{\sum_{\tau=1}^t \widetilde{A}_{\tau,i} r_t}{\sum_{\tau=1}^t B_\tau^2} - \theta_i^* \\
&= \frac{\sum_{\tau=1}^t \left(\widetilde{A}_{\tau,i}(\sum_{j=1}^d \theta_j^* \widetilde{A}_{\tau,j} + \epsilon_\tau)\right) - \sum_{\tau=1}^t B_\tau^2 \theta_i^*}{\sum_{\tau=1}^t B_\tau^2} \\
&= \frac{\sum_{\tau=1}^t \left(\widetilde{A}_{\tau,i}(\sum_{j\neq i} \theta_j^* \widetilde{A}_{\tau,j})\right)}{\sum_{\tau=1}^t B_\tau^2} + \frac{\sum_{\tau=1}^t \widetilde{A}_{\tau,i}\epsilon_\tau}{\sum_{\tau=1}^t B_\tau^2} && (\widetilde{A}_{t,i}^2 = 1) \\
&= \frac{\sum_{\tau=1}^t Z_{\tau,i}}{\sum_{\tau=1}^t B_\tau^2}.
\end{aligned}$$

Next, we will prove that $Z_{t,i}$ is conditionally $\sqrt{2}B_t$-subgaussian for $t \leq T(i)$. Define $\mathcal{H}_t = \sigma(\{\widetilde{A}_\tau\}_{\tau=1}^t, \{r_\tau\}_{\tau=1}^t)$ and $\mathcal{G}_t = \sigma(\widetilde{A}_t, \mathcal{H}_{t-1})$. Then,

$$\begin{aligned}
\mathbb{E}\left[\exp(\lambda Z_{t,i}) \mid \mathcal{H}_{t-1}\right] &= \mathbb{E}\left[\mathbb{E}\left[\exp(\lambda Z_{t,i}) \mid \mathcal{G}_{t-1}\right] \mid \mathcal{H}_{t-1}\right] \\
&= \mathbb{E}\left[\exp(\lambda \widetilde{A}_{t,i} M_{t,i})\mathbb{E}\left[\exp(\lambda \widetilde{A}_{t,i}\epsilon_t) \mid \mathcal{G}_{t-1}\right] \mid \mathcal{H}_{t-1}\right] \\
&\leq \mathbb{E}\left[\exp(\lambda \widetilde{A}_{t,i} M_{t,i}) \exp\left(\frac{\lambda^2 B_t^2}{2}\right) \mid \mathcal{H}_{t-1}\right] \\
&\qquad (\epsilon_t \text{ is conditionally 1-subgaussian, and } |\widetilde{A}_{t,i}| = B_t) \\
&= \exp\left(\frac{\lambda^2 B_t^2}{2}\right)\mathbb{E}\left[\mathbb{E}\left[\exp(\lambda \widetilde{A}_{t,i} M_{t,i}) \mid \mathcal{H}_{t-1}, M_{t,i}\right] \mid \mathcal{H}_{t-1}\right] \\
&\leq \exp\left(\frac{\lambda^2 B_t^2}{2}\right)\mathbb{E}\left[\exp\left(\frac{\lambda^2 B_t^2 M_{t,i}^2}{2}\right) \mid \mathcal{H}_{t-1}\right] \\
&\qquad (A_{t,i} \text{ is conditionally } B_t \cdot \mathrm{Rad}(1/2) \text{ when } t \leq T(i)) \\
&\leq \exp(B_t^2\lambda^2). \qquad (|M_{t,i}| = |\textstyle\sum_{j\in S_t} \widetilde{A}_{t,j}\theta_j^*| \leq \|\widetilde{A}_t\|_\infty \|\theta_*\|_1 \leq 1)
\end{aligned}$$

Therefore, one can use a modified version of Exercise 20.8 of Lattimore and Szepesvári (2020) to get the desired result. ∎

Let a_* be the optimal action. Now note that each instantaneous regret can be split into three terms as follows:

$$\begin{aligned}
\langle \theta^*, a_* - A_t \rangle &= \langle \theta_*^{S_t}, a_*^{S_t} - A_t^{S_t}\rangle + \langle \theta_*^{S_t^c}, a_*^{S_t^c} - A_t^{S_t^c}\rangle \\
&= \underbrace{\langle \theta_*^{S_t}, a_*^{S_t} - b(S_t)\rangle}_{(A)} + \underbrace{\langle \theta_*^{S_t}, b(S_t) - A_t^{S_t}\rangle}_{(B)} + \underbrace{\langle \theta_*^{S_t^c}, A_*^{S_t^c} - A_t^{S_t^c}\rangle}_{(C)}.
\end{aligned}$$

where $b(S_t) := \arg\max_{a \in \mathcal{B}_p^{S_t}(R_t)} \langle \theta_*^{S_t}, a\rangle$, the best action with norm restriction on S_t.

Analysis for (A) For the first part, it is an inevitable regret caused by the exploration budget. Simply, the $b(S_t) = R_t \cdot a_*^{S_t}$ and therefore

$$(A) = \langle \theta_*^{S_t}, a_*^{S_t} - b(S_t)\rangle = \left(1 - (1 - (d - |S_t|) \cdot B_t^p)^{1/p}\right)\langle \theta_*^{S_t}, a_*^{S_t}\rangle \leq (1 - R_t).$$

Analysis for (B) In this part we follow the traditional OFUL approach Abbasi-Yadkori et al. (2011). First, we have that

$$A_t^{S_t} := \operatorname*{arg\,max}_{a\in\mathcal{B}_p^{S_t}(R_t)} \langle\hat{\theta}_t, a\rangle + \sqrt{\beta}\|a\|_{V_t^{-1}} = \operatorname*{arg\,max}_{a\in\mathcal{B}_p^{S_t}(R_t)} \max_{\theta\in\mathcal{C}_t}\langle\theta, a\rangle,$$

where $R_t = (1-(d-|S_t|)B_t^p)^{1/p}$ and $\mathcal{C}_t := \{\theta\in\mathbb{R}^{|S_t|} : \|\theta-\hat{\theta}_t\|_{V_t}^2 \le \beta\}$ (Please refer to Section 19, 20 of Lattimore and Szepesvári (2020) for more details). If we let $\bar{\theta}_t := \arg\max_{\theta\in\mathcal{C}_t}\langle\theta, A_t^{S_t}\rangle$, then

$$\begin{aligned}
(B) &= \langle\theta_*^{S_t}, b(S_t)\rangle - \langle\theta_*^{S_t}, A_t^{S_t}\rangle \\
&\le \langle\bar{\theta}_t, A_t^{S_t}\rangle - \langle\theta_*^{S_t}, A_t^{S_t}\rangle && \text{(Definition of } A_t^{S_t}) \\
&\le \|\bar{\theta}_t - \theta_*^{S_t}\|_{V_t}\|A_t^{S_t}\|_{V_t^{-1}} && \text{(Cauchy)} \\
&\le \beta\|A_t^{S_t}\|_{V_t^{-1}}. && \text{(Lemma E.2)}
\end{aligned}$$

Lemma E.2. *With probability* $1-\frac{2}{T}$, $\|\bar{\theta}_t - \theta_*^{S_t}\|_{V_t}^2 \le \beta$.

Proof. For any $x\in\mathbb{R}^{|S_t|}$,

$$\begin{aligned}
|\langle\hat{\theta}_t - \theta_*^{S_t}, x\rangle| &= |\langle V_t^{-1}\sum_{\tau=1}^t A_\tau^{S_t} r_\tau - \theta_*^{S_t}, x\rangle| \\
&= \left|\langle V_t^{-1}\sum_{\tau=1}^t A_\tau^{S_t}(\langle\theta_*^{S_t}, A_\tau^{S_t}\rangle + \langle\theta_*^{S_t^c}, A_\tau^{S_t^c}\rangle + \epsilon_\tau) - \theta_*^{S_t}, x\rangle\right| \\
&\le \left|\langle x, V_t^{-1}\sum_{\tau=1}^t A_\tau^{S_t}(\langle\theta_*^{S_t^c}, A_\tau^{S_t^c}\rangle + \epsilon_\tau)\rangle\right| + \|\lambda\theta_*\|_{V_t^{-1}}.
\end{aligned}$$

Note that $\langle\theta_*^{S_t}, A_\tau^{S_t^c}\rangle + \epsilon_\tau$ is $\sqrt{2}$-subgaussian random variable, since $A_\tau^{S_t^c}$ are independent random Rademacher sampling with $|\langle\theta_*^{S_t}, A_\tau^{S_t^c}\rangle| \le 1$ and ϵ_τ is a conditionally independent 1-subgaussian random variable. Therefore, by the technique in (Abbasi-Yadkori et al., 2011, Theorem 2),

$$\mathbb{P}\left(\langle\hat{\theta}_t - \theta_*^{S_t}, x\rangle \ge \|x\|_{V_t^{-1}}\sqrt{\ln\frac{1}{T}} + \lambda^{1/2}\|\theta_*\|\right) \le \frac{1}{T}.$$

and using the similar covering argument in (Lattimore and Szepesvári, 2020, Section 20) (or the original version (Abbasi-Yadkori et al., 2011, Theorem 2)) with Remark 9, $\mathbb{P}\left(\|\hat{\theta}_t - \theta_*^{S_t}\|_{V_t}^2 \ge \beta\right) \le \frac{2}{T}$. ☐

Now the summation of (B), or $\mathrm{Reg}_T^{(B)} := \sum_{t=1}^T\langle\theta_*^{S_t}, b(S_t) - A_t^{S_t}\rangle$, can be bounded by $\beta\sum_{t=1}^T\|A_t^{S_t}\|_{V_t^{-1}}$ by Lemma E.2. Also, since $\theta_*\in\mathcal{B}_q$, by Holder's inequality

$$\mathrm{Reg}_T^{(B)} \le \|\theta_*^{S_t}\|_q\|b(S_t) - A_t^{S_t}\|_p \le 2.$$

Overall, $\mathrm{Reg}_T^{(B)} \le \beta\sum_{t=1}^T 2\min\left\{\|A_t^{S_t}\|_{V_t^{-1}}, 1\right\} \le \beta\sum_{t=1}^T 2\min\left\{\|A_t^{S_t}\|_{V_{t-1}^{-1}}, 1\right\}$ To bound $\sum_{t=1}^T\|A_t^{S_t}\|_{V_{t-1}^{-1}}$, we will use the following modified version of the Elliptic Potential Lemma((Abbasi-Yadkori et al., 2011, Theorem 3)).

Lemma E.3. *With probability* $1-\frac{1}{T}$, $\sum_{t=1}^T\min\left\{1, \|A_t^{S_t}\|_{V_{t-1}^{-1}}\right\} = O(\sqrt{sT\log T})$

Proof. We will closely follow the telescopic derivation of (Abbasi-Yadkori et al., 2011, Lemma 11), but since the dimension of the vector $A_t^{S_t}$ changes, we should consider two cases for the telescopic sum, when $S_t = S_{t-1}$ and $S_t \ne S_{t-1}$. One can simply create the telescopic sum as

$$\log\det(V_T) = \log(V_0) + \sum_{t=1}^T\log\left(\frac{V_t}{V_{t-1}}\right)$$

When $S_t = S_{t-1}$: In this case, we can follow the derivation of (Abbasi-Yadkori et al., 2011, Lemma 11) as

$$\det(V_t) = \det(V_{t-1} + A_t^{S_t}(A_t^{S_t})^\top) = \det(V_{t-1})(1 + \|A_t^{S_t}\|^2_{V_{t-1}^{-1}})$$

When $S_t \neq S_{t-1}$: In this case, let $V_t' = \sum_{\tau=1}^t A_\tau^{S_{t-1}}(A_\tau^{S_{t-1}})^\top + \lambda I_{S_{t-1}}$. From the same argument we used for the case $S_t = S_{t-1}$, $\det(V_t') = \det(V_{t-1})(1 + \|A_t^{S_{t-1}}\|^2_{V_{t-1}^{-1}})$. By the block matrix formula,

$$V_t = \begin{bmatrix} V_t' & x \\ x^\top & \sum_{\tau=1}^t A_\tau^{S_t \setminus S_{t-1}}(A_\tau^{S_t \setminus S_{t-1}})^\top + \lambda I \end{bmatrix}$$

where $x = \sum_{\tau=1}^t A_\tau^{S_{t-1}}(A_\tau^{S_t\setminus S_{t-1}})^\top$. Note that from the definition, $\det(V_t) > 0$, and by the block matrix determinant formula,

$$\begin{aligned}
\det(V_t) &= \det(V_t') \cdot \det\left(\sum_{\tau=1}^t A_\tau^{S_t\setminus S_{t-1}}(A_\tau^{S_t\setminus S_{t-1}})^\top + \lambda I_{S_t\setminus S_{t-1}} - x^\top (V_t')^{-1} x\right) \\
&\geq \det(V_t') \cdot \det\left(\lambda I_{S_t \setminus S_{t-1}}\right) && \text{(Claim 2)} \\
&\geq \det(V_t') \cdot \lambda^{|S_t\setminus S_{t-1}|}.
\end{aligned}$$

Here the last inequality is from the fact that for a positive definite matrix $M \in \mathbb{R}^d$, $(\mathrm{tr}(M)/d)^d \geq \det(M)$ and $\mathrm{tr}(A_\tau^{S_t\setminus S_{t-1}}(A_\tau^{S_t\setminus S_{t-1}})^\top) = \|A_\tau^{S_t\setminus S_{t-1}}\|^2 \leq |S_t \setminus S_{t-1}|$. The first inequality comes from the following claim:

Claim 2. $\sum_{\tau=1}^t A_\tau^{S_t\setminus S_{t-1}}(A_\tau^{S_t\setminus S_{t-1}})^\top - x^\top (V_t')^{-1}x \succeq 0$

Proof. For notational convenience, for an index set J and two integer a, b ($1 \leq a < b \leq T$), let $A_{a:b}^{(J)}$ be the $|J| \times (b-a)$ size matrix such that i-th column in $A_{a:b}^{(J)}$ is equal to $A_{i+a-1}^{(J)}$. Then, the claim we need to prove can be rewritten as:

$$A_{1:t}^{S_t\setminus S_{t-1}}(A_{1:t}^{S_t\setminus S_{t-1}})^\top - A_{1:t}^{S_t\setminus S_{t-1}}\left[A_{1:t}^{S_{t-1}}\left(\lambda I + (A_{1:t}^{S_{t-1}})^\top A_{1:t}^{S_{t-1}}\right)^{-1}(A_{1:t}^{S_{t-1}})^\top\right](A_{1:t}^{S_t\setminus S_{t-1}})^\top \succeq 0$$

or,

$$A_{1:t}^{S_t\setminus S_{t-1}}\left[I_t - A_{1:t}^{S_{t-1}}\left(\lambda I + (A_{1:t}^{S_{t-1}})^\top A_{1:t}^{S_{t-1}}\right)^{-1}(A_{1:t}^{S_{t-1}})^\top\right](A_{1:t}^{S_t\setminus S_{t-1}})^\top \succeq 0$$

This is true since for any $m \times n$ matrix X, $I_m - X(\lambda I_n + X^\top X)^{-1}X^\top \succeq 0$ by Lemma E.4. □

Lemma E.4. *For any $m \times n$ matrix X, $I_m - X(\lambda I_n + X^\top X)^{-1}X^\top \succeq 0$*

Proof. Let $X = U\Sigma V^\top$ be the SVD of X, and $\Sigma = \mathtt{diag}(\sigma_1, \cdots, \sigma_r)$. Then one can check $X(X^\top X + I_n)^{-1}X^\top = U\mathtt{diag}\left(\frac{\sigma_1^2}{\sigma_1^2+\lambda}, \cdots, \frac{\sigma_r^2}{\sigma_r^2+\lambda}, 0, 0, \cdots, 0\right)U^\top$. Therefore, $I_m - X(\lambda I_n + X^\top X)^{-1}X^\top = U\mathtt{diag}\left(\frac{\lambda}{\sigma_1^2+\lambda}, \cdots, \frac{\lambda}{\sigma_r^2+\lambda}, 1, 1, \cdots, 1\right)U^\top \succeq 0$ □

Overall, $\log\frac{\det(V_t)}{\det(V_{t-1})} = \log\frac{\det(V_t)}{\det(V_t')} + \log\frac{\det(V_t')}{\det(V_{t-1})} \geq |S_t\setminus S_{t-1}|\log(\lambda) + (1 + \|A_t^{S_t}\|^2_{V_t^{-1}})$.

Now by E.1 and Remark 9, $|S_t| \leq s$ with probability $1 - \frac{1}{T}$. Under this event, since $\lambda = s$, we can safely say $\|A_t^{S_t}\|_{V_t^{-1}} \leq 1$. Now using $x \leq 2\log(1+x)$ for $x \in [0, 1]$,

$$\begin{aligned}
\sum_{t=1}^T \min\left\{1, \|A_t^{S_t}\|^2_{V_{t-1}^{-1}}\right\} &\leq 2\sum_{t=1}^T \log\left(1 + \|A_t^{S_t}\|^2_{V_{t-1}^{-1}}\right) \\
&\leq 2(\log\det(V_T) - s^2\log\lambda) \\
&\leq 2\log\det(V_T) = O(s\log T)
\end{aligned}$$

and using Cauchy Schwarz, $\left(\sum_{t=1}^T \min\left\{1, \|A_t^{S_t}\|_{V_{t-1}^{-1}}\right\}\right)^2 \le T \cdot \sum_{t=1}^T \min\left\{1, \|A_t^{S_t}\|^2_{V_{t-1}^{-1}}\right\} = O(sT\log T)$ which implies $\left(\sum_{t=1}^T \min\left\{1, \|A_t^{S_t}\|_{V_{t-1}^{-1}}\right\}\right) = O(\sqrt{sT\log T})$. $\square$

Therefore, by Lemma E.3 and E.2, the regret upper bound will be at most $O(s\sqrt{T}\log T)$.

Analysis for (C) Now for (C), the instantaneous regret caused by random exploration is

$$\langle \theta_*^{S_t^c}, a_*^{S_t^c} - A_t^{S_t^c}\rangle \le \langle \theta_*^{S_t^c}, a_*^{S_t^c}\rangle + |\langle \theta_*^{S_t^c}, A_t^{S_t^c}\rangle|.$$

From our construction, we only know that with high probability $\|\theta_*^{S_t^c}\|_\infty \le W_t$ by Proposition E.1. By Lagrangian Multiplier Method, One can check that

$$\langle \theta_*^{S_t^c}, a_*^{S_t^c}\rangle = \frac{\|\theta_*^{S_t^c}\|_q^q}{\|\theta^*\|_q^{q/p}}.$$

This means that we can upper bound the first term, $\langle \theta_*^{S_t^c}, a_*^{S_t^c}\rangle$, as

$$\langle \theta_*^{S_t^c}, a_*^{S_t^c}\rangle = \frac{\|\theta_*^{S_t^c}\|_q^q}{\|\theta^*\|_q^{q/p}} \le \frac{\|\theta_*^{S_t^c}\|_q^q}{\|\theta_*^{S_t^c}\|_q^{q/p}} = \|\theta_*^{S_t^c}\|_q \le |S\backslash S_t|^{1/q} W_t \le S^{1/q} W_t.$$

Also, the error bound for $|\langle \theta_*^{S_t^c}, \widetilde{A}_t^{S_t^c}\rangle|$ is

$$|\langle \theta_*^{S_t^c}, \widetilde{A}_t^{S_t^c}\rangle| \le W_t B_t |S\backslash S_t|^{1/q} \le W_t |S\backslash S_t|^{1/q} \le W_t S^{1/q}.$$

Overall, the cumulative regret is, ignoring some constants,

$$\begin{aligned}\sum_{t=1}^T \langle \theta^*, a_* - A_t\rangle &= \sum_{t=1}^T \underbrace{\langle \theta_*^{S_t}, a_*^{S_t} - b(S_t)\rangle}_{(A)} + \underbrace{\langle \theta_*^{S_t}, b(S_t) - A_t^{S_t}\rangle}_{(B)} + \underbrace{\langle \theta_*^{S_t^c}, a_*^{S_t^c} - A_t^{S_t^c}\rangle}_{(C)} \\ &\le \sum_{t=1}^T (1-R_t) + s\sqrt{T}\log T + s^{1/q}\sum_{t=1}^T W_t. \end{aligned} \tag{26}$$

To bound the first term, let $t_0 = \min\{t \in \mathbb{N} : dB_t^p \le 1 - \frac{1}{p+1}\}$. One can check that

$$t_0 \le s^{\frac{2p}{p-1}} d^{2/p} \left(1 - \frac{1}{p+1}\right)^{-\frac{2(p+1)}{p}} \le 4 s^{\frac{2p}{p-1}} d^{2/p},$$

which is smaller than $\sqrt{T}$ when $T > d^{4/p} s^{\frac{4p}{p-1}}$. Now,

$$\sum_{t=1}^T (1-R_t) \le t_0 + \sum_{t=t_0+1}^T (1-R_t),$$

and to bound $\sum_{t=t_0+1}^T (1-R_t)$, one can use the Taylor remainder theorem. There exists $x \in (0, (d-|S_t|)B_t^p)$ such that

$$\begin{aligned} R_t &= (1-(d-|S_t|)B_t^p)^{1/p} \\ &= 1 - \frac{(d-|S_t|)B_t^p}{p} + \frac{1}{2p}\left(\frac{1}{p}-1\right)(1-x)^{1/p-2}((d-|S_t|)B_t^p)^2 \\ &\le 1 - \frac{(d-|S_t|)B_t^p}{p}, \end{aligned} \qquad (\tfrac{1}{p} - 1 < 0)$$

which means

$$(1 - R_t) = \left(1 - (1 - (d - |S_t|)B_t^p)^{1/p}\right) \leq \frac{d}{p} B_t^p.$$

Let $C(s,d) := (s^q d^{-1})^{\frac{1}{p+1}}$ for the notational convenience. Now we can rewrite the regret bound in Eq. (26) as

$$(26) \leq \frac{d}{p} \sum_{t=t_0}^{T} B_t^p + s\sqrt{T}\log T + s^{1/q} \sum_{t=1}^{T} W_t.$$

For the first term, by the Integral Test on the monotone decreasing sequence, $\sum_{t=1}^T B_t^p = C^p(s,d) \cdot \sum_{t=1}^T t^{-\frac{p}{2(p+1)}} = C^p(s,d) \cdot O(\sqrt{T} \cdot T^{\frac{1}{2(p+1)}})$.

To bound the third term, note that $B_t = C(s,d) \cdot t^{-\frac{1}{2(p+1)}}$. By the y the Integral Test on the monotone decreasing sequence, $\sum_{\tau=1}^t B_t^2 = C^2(s,d) \cdot O(t^{1-\frac{1}{p+1}})$, $W_t = \frac{1}{C(s,d)} O(t^{-\frac{1}{2}+\frac{1}{2(p+1)}} \log T)$, and $\sum_{t=1}^T W_t = \frac{1}{C(s,d)} O(\sqrt{T} \cdot T^{\frac{1}{2(p+1)}} \log T)$. Substituting $C(s,d) = (s^q d^{-1})^{\frac{1}{p+1}}$ leads to the desired regret upper bound. □

E.1 Complexity analysis

Time Complexity The time complexity analysis for `OFULSiCoE` largely mirrors that of `MiDeSiCoE`. The main exception is the computation of the optimal action choice oracle in Line 4 of Algorithm 3. Consequently, the overall time complexity is primarily dependent on the implementation of this oracle. `OFULSiCoE` calls the oracle for each round, so the oracle complexity is $O(T)$. Except that, the overall time complexity is $O(d^3T)$, mainly caused by the matrix inversion in Line 14.

Space Complexity The space complexity is $O(d^2)$. The dominant factor is the storage of the matrix V_{t+1}, which requires $O(d^2)$ space. The algorithm's efficiency lies in the fact that it only needs to store the most recent matrix, V_t, not the entire sequence $(V_1, V_2, \dots, V_{t-1})$, making the space requirement constant with respect to the number of iterations.

F Experiment details

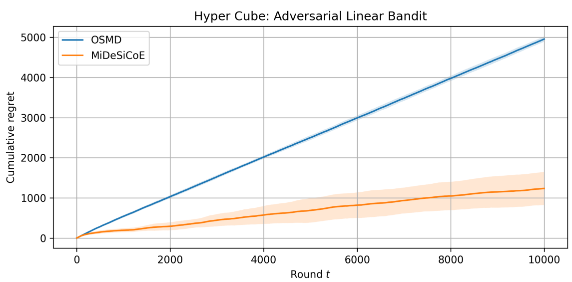
In this section, we present additional experimental results to complement those reported in Section 4. All experiments are conducted on synthetic data following the same settings and protocols as described in the main text. All algorithms are implemented in Python and executed on a machine equipped with an Apple M3 chip (8-core CPU, 10-core GPU) and 18 GB of unified memory. The anonymized source code is available at `https://github.com/NamTranZzz/Adversarial-and-Stochastic-Sparse-Linear-Bandits.git`.

Adversarial Setting. We evaluate the performance of MiDeSiCoE under the adversarial sparse linear bandit setting. The experiments are performed using the same synthetic generation process described in Section 4, where the loss vectors are supported on a fixed subset of coordinates of size s. In addition to the settings in the main paper, we include results for ($d = 80$, $s = 2$, $T = 10000$) and ($d = 400$, $s = 5$, $T = 40000$), for the l_∞-action set in Figure 4 and for the l_2-action set in Figure 5. In both settings our algorithms exhibited consistently lower regret compared to original OSMD algorithm which does not utilise sparsity structure.

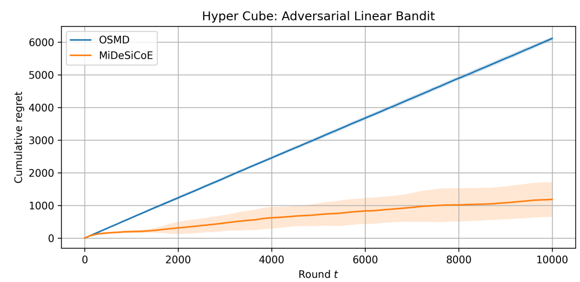
Stochastic Setting. We further assess the performance of OFULSiCoE in the stochastic sparse linear bandit regime on ℓ_p-balls. Following the setup in Section 3, we fix the sparsity level s and vary p to study its effect on the exploration–exploitation balance. In addition to the results reported in the main paper, we include experiments for ($d = 100$, $s = 3$, $p = 15$, $T = 10000$), and ($d = 100$, $s = 3$, $p = 60$, $T = 40000$), which are shown in Figure 6. In the simulation, our algorithm exhibited consistently lower regret compared to the original OFUL algorithm, which does not utilize the sparsity structure. Moreover, the observed regret behavior aligns well with our theoretical predictions: larger p values, which correspond to geometry closer to l_∞, yield faster convergence and lower overall regret.

Moreover, we conducted experiment with $d = 50$, $s = 1$, $T = 40000$, for both $p \to \infty$ (Figure 6e), and $p \to 2$ (Figure 6d). For the case $p \to \infty$, particularly, $p = 1000$, we observe that `OFULSiCoE` outperforms the baseline

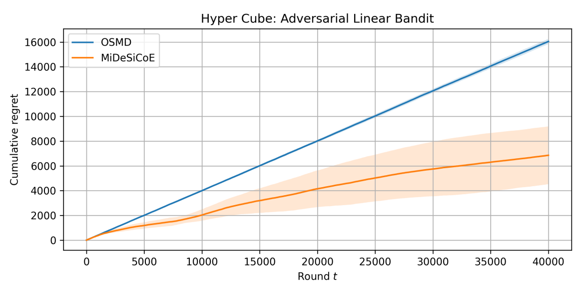
`OFUL`, which is consistent with our theoretical findings. Regarding the case where $p = 2$, our experiment shows that our algorithm performs worse than the baseline OFUL algorithm. This observation is consistent with our theoretical understanding: the OFULSiCoE algorithm interpolates between OFUL and an Explore-then-Commit (ETC) style strategy. When p is small, the algorithm behaves more like ETC, spending a larger portion of the budget on identifying the sparsity pattern, which in turn leads to larger regret. As noted in the limitations of our results, designing a version of our algorithm that degrades smoothly as is an interesting direction, and we leave this for future work.



(a) $d = 80$, $s = 2$, $T = 10000$

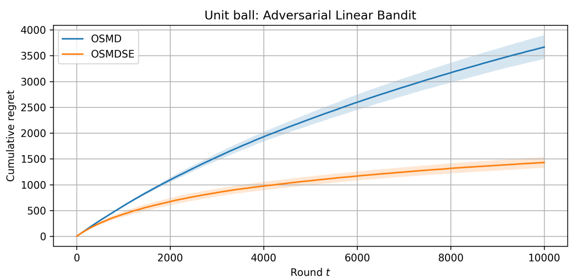


(b) $d = 100$, $s = 3$, $T = 10000$

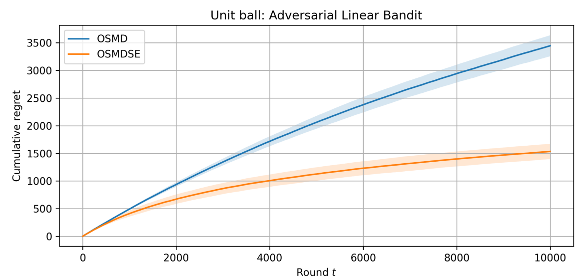


(c) $d = 400$, $s = 5$, $T = 40000$

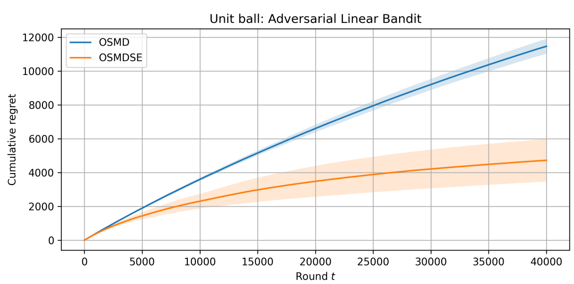
Figure 4: Expected cumulative regret for adversarial linear bandits on hypercube action set.



(a) $d = 80,\ s = 2,\ T = 10000$

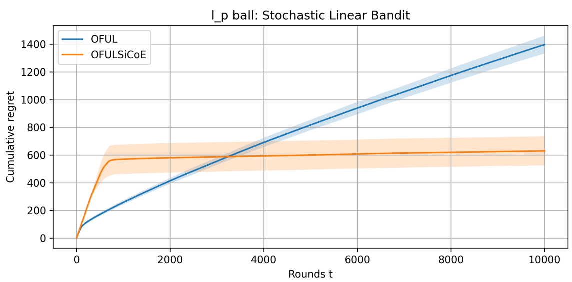


(b) $d = 100,\ s = 3,\ T = 10000$

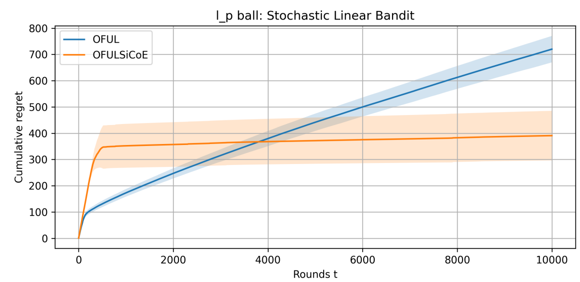


(c) $d = 400,\ s = 5,\ T = 40000$

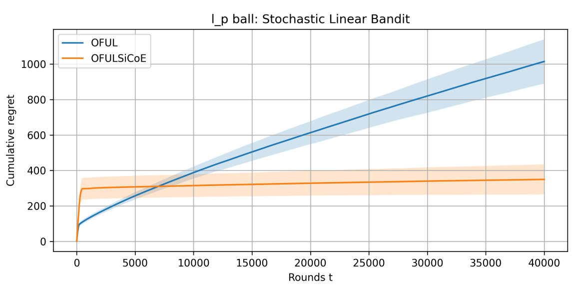
Figure 5: Expected cumulative regret for adversarial linear bandits on unit-ball action set.



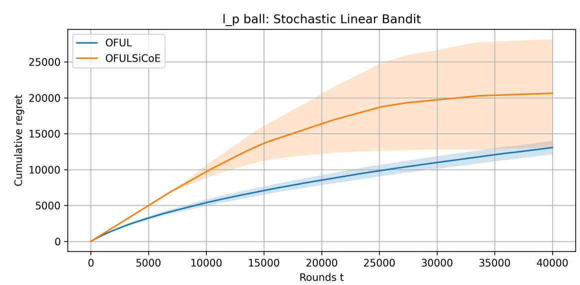
(a) Hypercube, $d = 100$, $s = 3$, $p = 15$



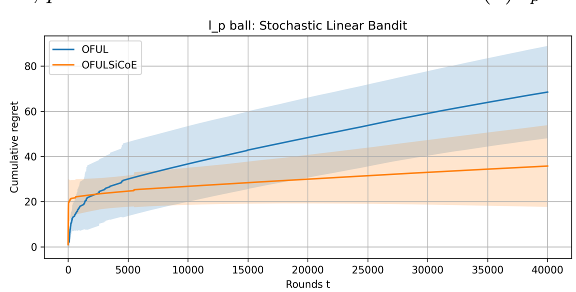
(b) l_p ball, $d = 100$, $s = 3$, $p = 30$



(c) l_p ball, $d = 100$, $s = 3$, $p = 60$



(d) l_p ball, $d = 50$, $s = 1$, $p = 2$



(e) l_p ball, $d = 50$, $s = 1$, $p = 1000$

Figure 6: Expected cumulative regret for stochastic linear bandits on l_p-ball action set.