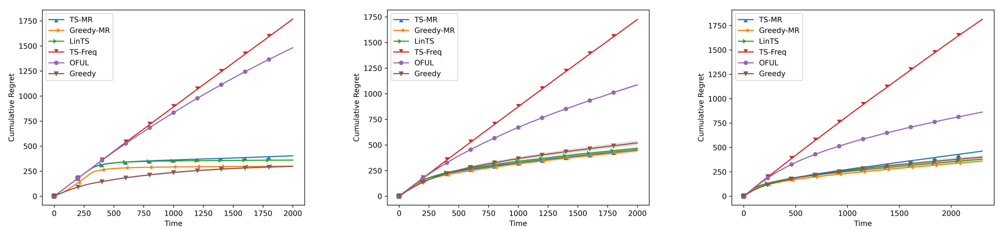
Below, we present our responses that contain extra experiments.

Experimental results on real-world datasets (Rev. sjzc)

We explore the performance of standard COFUL algorithms and the proposed TS-MR and Greedy-MR algorithms on real-world datasets. We consider three datasets for classification tasks on the OPENML platform, Cardiotocography, JapaneseVowels and Segment. They focus on healthcare, pattern recognition, and computer vision problems respectively.

Setup: We follow the standard approach in the literature that converts classification tasks to contextual bandit problems, and then embeds it into linear bandit problems in the same way as Example 2 in Section 6. Specifically, we regard each class as an action so that at each time, the decision-maker assigns a class to the feature observed and receives a binary reward, namely 1 for assigning the correct class and 0 otherwise, plus a Gaussian noise.

We plot the cumulative regret (averaged over 50 duplicates) for all algorithms. Figure 1 shows that for all real-world datasets: OFUL and TS-Freq perform poorly due to their conservative exploration; LinTS and Greedy are achieving empirical success even though they don't have theoretical guarantees; TS-MR and Greedy-MR retains the desirable empirical performance of LinTS and Greedy, while enjoying the minimax optimal frequentist regret bound.



(a) Cardiotocography dataset. (b) JapaneseVowels dataset. (c) Segment dataset.

Figure 1: Cumulative regret of all algorithms on real-world datasets.

Empirically validation of Case 2 in Section 5.1 (Rev. RqoU & Q5EJ)

To empirically validate that Case 2 isn't vacant, we check the condition presented in Proposition 4. We track the ratio $\zeta_t := \frac{\|\widehat{\theta}_t\|_{V_t}^2/\|\widehat{\theta}_t\|^2}{\lambda_1(V_t)}$, which is a proxy that measures how close $\widehat{\theta}_t$ is aligned with the top eigenspace of V_t. Specifically, $\zeta_t = 1$ implies that $\widehat{\theta}_t$ lies in the top eigenspace. We rerun Example 1 in Section 6 and plot the empirical $\{\zeta_t\}$ sequence.

As we see in Figure 2, for all bandit algorithms, ζ_t is already very high (above 0.9) at the very early stage of the implementation and tends to 1 eventually. This validates the tendency in Case 2. (We remark that the sharp decrease at the beginning is due to the behavior of the regularized least square estimator in over-parameterized regimes when the time t is smaller than the dimension d.)

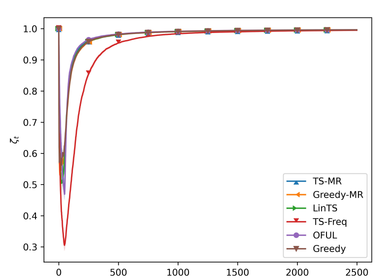


Figure 2: Involvement of the alignment proxy ζ_t in Example 1.