EFFICIENT MULTINOMIAL LOGISTIC BANDIT VIA ROBUST FREQUENT DIRECTIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

This paper studies efficient online algorithms for multinomial logistic bandits (MLogB), where the feedback distribution over $K+1$ outcomes follows a multinomial logistic model of d-dimensional action vectors. A representative UCB-type algorithm, OFUL-MLogB, achieves a regret bound of $\widetilde{\mathcal{O}}(Kd\sqrt{T})$, but still requires $\mathcal{O}(K^3d^3)$ time and $\mathcal{O}(K^2d^2)$ space per round due to parameter estimation and optimistic reward construction, which is prohibitive in high-dimensional settings. To address this limitation, we propose EOFD-MLogB, an efficient algorithm integrating robust frequent directions matrix sketching. By maintaining a low-rank SVD sketch of the Hessian matrix from online Newton step (ONS)-based parameter estimation, we make both nonlinear estimation and optimistic reward construction efficient: the constrained ONS parameter update is reduced to a one-dimensional root-finding problem, while the $Kd \times K$ spectral-norm computation in the optimistic reward is reduced to a $K \times K$ eigenvalue problem. This yields dominant per-round time complexity $\mathcal{O}(Kd(m+K)^2)$ and space complexity $\mathcal{O}(Kd(m+K))$, where $m \ll d$ is the sketch size. We further prove a regret bound of $\widetilde{\mathcal{O}}(\Delta_T(Kd\ln\Delta_T + m)\sqrt{T})$, where the sketching error factor Δ_T is controlled by the m-truncated spectral tail of the ONS Hessian. Thus, when the ONS Hessian is approximately low-rank, the regret remains close to that of OFUL-MLogB. Experiments validate the computational efficiency and competitive performance.

1 INTRODUCTION

Stochastic multinomial logistic bandits (MLogB) (Amani & Thrampoulidis, 2021) model a class of sequential decision-making problems with categorical feedback. At each round $t \in \mathbb{N}_+$, an agent selects an action $\boldsymbol{x}_t \in \mathcal{A}_t \subseteq \mathbb{R}^d$ and observes environment feedback $y_t \in [K] \cup \{0\}$, where each outcome is associated with a reward $r_k \in \mathbb{R}_+$. The feedback distribution is parameterized by an unknown environment parameter $W_* \in \mathbb{R}^{K\times d}$ through $\Pr[y_t = k \mid \boldsymbol{x}_t] = [\boldsymbol{\sigma}(W_*\boldsymbol{x}_t)]_k$, where $\boldsymbol{\sigma}(\cdot)$ is the multinomial logistic function:

$$[\boldsymbol{\sigma}(\boldsymbol{z})]_0 = \frac{1}{1+\sum_{k=1}^K \exp(z_k)}, \quad [\boldsymbol{\sigma}(\boldsymbol{z})]_i = \frac{\exp(z_i)}{1+\sum_{k=1}^K \exp(z_k)}, \quad i \geq 1. \tag{1}$$

This framework captures a wide range of applications such as recommendation systems (Li et al., 2010; 2017), where different user responses to a recommended item are mapped to different rewards. The goal is to minimize the cumulative regret

$$\mathrm{Regret}(T) = \sum_{t=1}^T \max_{\boldsymbol{x}\in\mathcal{A}_t} \boldsymbol{r}^\top \boldsymbol{\sigma}(W_*\boldsymbol{x}) - \boldsymbol{r}^\top \boldsymbol{\sigma}(W_*\boldsymbol{x}_t). \tag{2}$$

A standard approach is the upper confidence bound (UCB), which maintains confidence sets for W_* estimation and selects actions according to their optimistic rewards. A prior UCB-type algorithm, MNL-UCB (Amani & Thrampoulidis, 2021), combines batch maximum likelihood estimation with a first-order Taylor-based optimistic reward approximation, achieving provable regret guarantees. Its improved variant (Amani & Thrampoulidis, 2021, Appendix C) further employs second-order Taylor-based optimistic reward approximation to construct a tighter approximation for the non-convex multinomial logistic reward, but still relies on batch parameter estimation and hence full-history

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

storage. OFUL-MLogB (Zhang & Sugiyama, 2023) avoids this storage cost by using online Newton step (ONS) (Hazan et al., 2007) for parameter estimation, while retaining the second-order optimistic reward design. However, it still maintains and manipulates a $Kd \times Kd$ Hessian matrix, leading to per-round time complexity $\mathcal{O}(|\mathcal{A}_t|K^3d^2 + K^3d^3)$ and space complexity $\mathcal{O}(K^2d^2)$, which are prohibitive in high-dimensional settings.

The bottleneck comes from the $Kd \times Kd$ accumulated Hessian in both parameter estimation and action optimistic rewards. The ONS parameter estimation requires solving a quadratic-constrained quadratic program (QCQP) for the per-round update, while the optimistic reward construction involves high-dimensional spectral norm evaluations; both require $\mathcal{O}(K^3d^3)$ time and scale poorly with d. Although matrix sketching has successfully accelerated stochastic linear bandits (SLB) (Abbasi-Yadkori et al., 2011; Kuzborskij et al., 2019), directly sketching the Hessian does not resolve these two bottlenecks in the nonlinear multinomial logistic model: unlike linear bandits, neither the constrained ONS update nor the spectral norm computation admits a closed-form reduction analogous to SLB. Thus, the key challenge is to make both nonlinear computations compatible with a low-rank Hessian representation. Similar Hessian-induced bottlenecks also arise in ONS-based generalized linear bandits (Zhang et al., 2025) and other second-order MLogB algorithms (Lee et al., 2024; Boudart et al., 2026).

We focus on the memory-constrained online setting, with a prescribed memory budget and a sketch size m fixed in advance to meet this budget. We propose EOFD-MLogB, an efficient second-order MLogB algorithm based on robust frequent directions (RFD) matrix sketching (Luo et al., 2019; Ghashami et al., 2016). By maintaining a rank-m SVD sketch of the accumulated Hessian, EOFD-MLogB reduces the space complexity from $\mathcal{O}(K^2d^2)$ to $\mathcal{O}(Kd(m+K))$. More importantly, we exploit the sketched spectral structure to reduce the Kd-dimensional QCQP in ONS parameter updates to a one-dimensional root-finding problem, and separately reduce the $Kd \times K$ spectral norm computations in optimistic reward construction to $K \times K$ computations. These reductions allow both parameter estimation and optimistic action selection to operate directly on the low-rank Hessian representation. As a result, the per-round computational cost becomes $\mathcal{O}(|\mathcal{A}_t|(K^3 + mK^2 + dmK) + dK(m+K)^2)$.

On the theoretical side, we analyze the statistical cost of Hessian sketching, and prove a regret bound of $\widetilde{\mathcal{O}}(\Delta_T(Kd\ln\Delta_T + m)\sqrt{T})$, where $\Delta_T = 1 + 2\rho_{1:T}/\lambda$ is the sketching factor, and $\rho_{1:T}$ is the sketching error term bounded by the m-truncated spectral tail of the Hessian. Thus, when the accumulated Hessian is approximately low-rank, EOFD-MLogB achieves regret close to the $\widetilde{\mathcal{O}}(Kd\sqrt{T})$ regret of OFUL-MLogB while substantially improving computational efficiency. Finally, experiments on `MFEAT`, synthetic data, and additional real datasets demonstrate significant speedups with competitive regret performance.

Notations. For $W \in \mathbb{R}^{K\times d}$, let $\overrightarrow{W} \in \mathbb{R}^{Kd}$ denote its row-wise vectorization. We use $\|\cdot\|_F$ and $\|\cdot\|_2$ for Frobenius and spectral norms, respectively, $\otimes$ for the Kronecker product, and $\|\boldsymbol{x}\|_Z = \sqrt{\boldsymbol{x}^\top Z\boldsymbol{x}}$ for the norm induced by a positive semidefinite matrix Z. We use $\mathcal{O}(\cdot)$ to hide constants, highlighting the dependency on m, d, K, κ, T, and $\widetilde{\mathcal{O}}(\cdot)$ to further hide polylogarithmic factors.

2 Related Work

In this section, we review prior work related to MLogB and frequent directions.

2.1 MLogB and OFUL-MLogB

MLogB extends logistic bandits ($K = 1$), a canonical instance of generalized linear bandits (GLB) (Filippi et al., 2010; Li et al., 2017; Ding et al., 2021). For logistic bandits, a UCB-type algorithm OL^{2}M (Zhang et al., 2016) employed ONS to estimate the parameter; subsequent works (Abeille et al., 2021; Faury et al., 2022) improved the regret bound to the optimal order. The multinomial case ($K > 1$) is more challenging, because maximizing the expected multinomial logistic reward is generally a non-convex problem. The improved variant of MNL-UCB (Amani & Thrampoulidis, 2021, Appendix C) and subsequent UCB-type algorithms (Lee et al., 2024) handle this non-convexity using second-order optimistic reward approximations, but rely on batch MLE and storage of the full

interaction history. The state-of-the-art ONS-based OFUL-MLogB algorithm (Zhang & Sugiyama, 2023) combines second-order Taylor expansion of optimistic rewards with curvature-aware online Newton step (ONS) updates, eliminating full-history storage but requiring the maintenance and manipulation of a $Kd \times Kd$ Hessian matrix.

We next summarize the components of OFUL-MLogB that are relevant to our method. After selecting $\boldsymbol{x}_t$ and observing y_t, we define the logistic loss $\ell_t(W) = -\ln[\boldsymbol{\sigma}(W\boldsymbol{x}_t)]_{y_t}$. OFUL-MLogB maintains an estimate W_t of W_*, updated by a proximal ONS step each round:

$$W_{t+1} = \arg\min_{W \in \mathcal{W}} \underbrace{\left\langle \nabla \ell_t(W_t), \overrightarrow{W} - \overrightarrow{W}_t \right\rangle + \frac{1}{2} \left\| \overrightarrow{W} - \overrightarrow{W}_t \right\|^2_{\nabla^2 \ell_t(W_t)}}_{\widetilde{\ell}_t(W)} + \frac{1}{2\eta} \left\| \overrightarrow{W} - \overrightarrow{W}_t \right\|^2_{H_t}, \quad (3)$$

where $\widetilde{\ell}_t(W)$ is a second-order surrogate of ℓ_t at W_t up to an additive constant, and $\mathcal{W} = \{W \in \mathbb{R}^{K \times d} | \|W\|_F \leq S\}$ The Hessian is updated as $H_{t+1} = H_t + \nabla^2 \ell_t(W_{t+1})$, with $H_1 = \lambda I_{Kd}$.

For action selection, OFUL-MLogB chooses $\boldsymbol{x}_t$ according to the UCB principle under probability at least $1 - \delta$:

$$\boldsymbol{x}_t = \arg\max_{\boldsymbol{x} \in \mathcal{A}_t} R_t(\boldsymbol{x}, \delta), \quad (4)$$

where

$$R_t(\boldsymbol{x}, \delta) = \begin{cases} \max_{\boldsymbol{w} \in \mathcal{C}_t} \sigma(\boldsymbol{w}^\top \boldsymbol{x}), & K = 1 \\ \boldsymbol{r}^\top \boldsymbol{\sigma}(W_t \boldsymbol{x}) + b_t(\boldsymbol{x}, \delta), & K \geq 2 \end{cases}. \quad (5)$$

Here $\mathcal{C}_t = \{W \in \mathbb{R}^{K \times d} \mid \|\overrightarrow{W} - \overrightarrow{W}_t\|^2_{H_t} \leq \beta_t(\delta)\}$ is the confidence region, and $b_t(\boldsymbol{x}, \delta)$ is the optimistic bonus whose definition is deferred to equation 8 in Section 3.1. When $K = 1$, we assume $(r_0, r_1) = (0, 1)$, so the expected reward reduces to the scalar logistic reward $\sigma(\boldsymbol{w}^\top \boldsymbol{x})$, and $\mathcal{C}_t$ reduces to the corresponding vector confidence set; see Appendix B.1. These optimistic constructions ensure, with probability at least $1-\delta$, that $\boldsymbol{r}^\top \boldsymbol{\sigma}(W_* \boldsymbol{x})$ is upper-bounded by $R_t(\boldsymbol{x}, \delta)$ for all candidate actions $\boldsymbol{x}$. We refer readers to Theorem 1 and Lemma 7 for the corresponding guarantee.

Although OFUL-MLogB achieves a near-optimal regret bound $\widetilde{\mathcal{O}}(Kd\sqrt{T})$ in the multinomial case (Zhang & Sugiyama, 2023, Appendix C.5), its dependence on the full $Kd \times Kd$ Hessian leads to costly $\mathcal{O}(|\mathcal{A}_t|K^3d^2 + K^3d^3)$ time and $\mathcal{O}(K^2d^2)$ space per round in high-dimensional settings. Section 3.1 details these computational obstacles.

2.2 Matrix Sketching and Robust Frequent Directions

Matrix sketching approximates a streaming covariance matrix $C_t = \sum_{s=1}^{t} x_s x_s^\top$ using a low-rank representation. Unlike randomized methods such as random projection (Baraniuk et al., 2010) or column selection (Drineas et al., 2006), frequent directions (FD) is a deterministic streaming method that maintains a sketch $S_t \in \mathbb{R}^{m \times d}$ with $S_t^\top S_t \approx C_t$ by appending each incoming vector and shrinking the squared singular values by the $(m+1)$-th squared singular value. Robust frequent directions (RFD) additionally accumulates half of each shrinkage amount in $\rho_{1:t}$ and approximates C_t by

$$S_t^\top S_t + \rho_{1:t} I_d.$$

This isotropic correction improves the approximation guarantee and conditioning relative to FD (Luo et al., 2019, Theorems 3 and 4).

FD and RFD have been applied to stochastic linear bandits and online convex optimization. In SLB (Kuzborskij et al., 2019; Wen et al., 2024; 2026), FD reduces the time and space costs of least-squares estimation in linear models and UCB bonus computation through the Woodbury identity. In online convex optimization, FD/RFD-based methods reduce matrix storage and update costs, with either closed-form projections under prediction constraints (Luo et al., 2016; 2019) or unconstrained updates (Wan & Zhang, 2022; Yang et al., 2025). For the nonlinear logistic model in MLogB, however, the ball-constrained ONS update and second-order optimistic bonus introduce projection and spectral-norm computations that do not admit these direct simplifications.

3 EFFICIENT ALGORITHMS FOR MLOGB

We operate under the following standard assumptions (Amani & Thrampoulidis, 2021; Zhang & Sugiyama, 2023):

Assumption 1. The action set $\mathcal{A}_t$, the true parameter $W_\star$, and the reward vector $\boldsymbol{r} \in \mathbb{R}_+^{K+1}$ are bounded: $\sup_{\boldsymbol{x}\in\mathcal{A}_t} \|\boldsymbol{x}\|_2 \le D$, $\|W_\star\|_F \le S$, and $\|\boldsymbol{r}\|_2 \le R$. Without loss of generality, we set $r_0 = 0$ as a dummy outcome, and ignore 0-th dimension in the subsequent analysis since it does not contribute to the reward and the regret. See Appendix B.1 for more details.

Assumption 2. The multinomial logistic mapping $\boldsymbol{\sigma}(\cdot)$ satisfies strong convexity condition: for any $W \in \mathcal{W}$ and $\boldsymbol{x} \in \mathcal{A}_t$, there exist constants $\kappa > 0$ and $L \in [0, 1]$ such that $\frac{1}{\kappa} I_K \preceq \nabla\boldsymbol{\sigma}(W\boldsymbol{x}) \preceq L I_K$.

The constant κ typically depends exponentially on the domain diameter SD (Amani & Thrampoulidis, 2021, Section 3).

3.1 COMPUTATIONAL OBSTACLES

We analyze the two Hessian-induced bottlenecks that hinder the scalability of OFUL-MLogB.

Constrained optimization for parameter estimation. The proximal ONS update equation 3 can be decomposed into an unconstrained gradient step followed by an ellipsoidal projection:

$$\overrightarrow{W}' = \overrightarrow{W}_t - \eta \widetilde{H}_t^{-1} \nabla \ell_t(W_t), \tag{6}$$

$$W_{t+1} = \operatorname*{arg\,min}_{\|W\|_F \le S} \left\| \overrightarrow{W} - \overrightarrow{W}' \right\|_{\widetilde{H}_t}^2, \tag{7}$$

where $\widetilde{H}_t = H_t + \eta \nabla^2 \ell_t(W_t)$. The gradient step requires solving a linear system involving the $Kd \times Kd$ matrix $\widetilde{H}_t$, and the projection is a QCQP without a closed-form solution for a general Hessian. Standard exact methods require a full eigen-decomposition (Hazan et al., 2007), leading to $\mathcal{O}(K^3 d^3)$ time per round. Moreover, simply replacing H_t by a sketched matrix $S_t^\top S_t + (\lambda + \rho_{1:t-1})I$ does not by itself yield an efficient projection, which motivates our explicit use of the sketch's spectral form.

High-dimensional spectral norms in optimistic rewards. For $K \ge 2$, evaluating the optimistic reward $R_t(\boldsymbol{x}, \delta)$ in equation 5 requires the bonus

$$b_t(\boldsymbol{x}, \delta) = \sqrt{\beta_t(\delta)} \, \|(\nabla\boldsymbol{\sigma}(W_t\boldsymbol{x}) \otimes \boldsymbol{x})\boldsymbol{r}\|_{H_t^{-1}} + \frac{1}{2}\|\boldsymbol{r}\|_1 \beta_t(\delta) \underbrace{\left\| H_t^{-\frac{1}{2}} (I_K \otimes \boldsymbol{x}) \right\|_2^2}_{\text{spectral norm term}}. \tag{8}$$

The dominant cost lies in the spectral norm term. Specifically, computing the inverse square root $H_t^{-\frac{1}{2}}$ incurs $\Omega(K^2 d^2)$ time ($\mathcal{O}(K^3 d^3)$ via eigen-decomposition) and $\mathcal{O}(K^2 d^2)$ space overhead. Even after $H_t^{-\frac{1}{2}}$ is formed, computing $\|H_t^{-\frac{1}{2}}(I_K \otimes \boldsymbol{x})\|_2$ still demands $\mathcal{O}(K^3 d^2)$ time. The eigenvalue-based implementation of OFUL-MLogB costs $\mathcal{O}(K^3 d^3)$ time for each action (Zhang & Sugiyama, 2023, C.3.4). Thus, action selection becomes impractical in high-dimensional settings. The same type of bottleneck appears in other MLogB algorithms using second-order optimistic bonuses (Lee & Oh, 2025; Boudart et al., 2026).

Contrast with SLB. In stochastic linear bandits, sketching is straightforward because the least-squares parameter update $\widehat{\theta} = H_t^{-1} \sum_{s=1}^t y_t \boldsymbol{x}_t$ and the UCB bonus $\|\boldsymbol{x}\|_{H_t^{-1}}$ both admit closed-form Woodbury reductions to matrix-vector products (Kuzborskij et al., 2019). MLogB lacks this structure: the nonlinear model leads to a constrained second-order projection equation 7 for parameter estimation and high-dimensional spectral norms equation 8 in the optimistic bonus. These two obstacles preclude a direct application of standard sketching techniques and motivate the spectral reductions developed below.

Algorithm 1: Adapted RFD for Matrix-Valued Updates

Input: Sketch size m, Hessian update $\nabla^2\ell_t(W_{t+1}) = \nabla\boldsymbol{\sigma}(W_{t+1}\boldsymbol{x}_t) \otimes \boldsymbol{x}_t\boldsymbol{x}_t^\top$, current sketch $(V_t, \Lambda_t, \rho_{1:t-1})$

$PDP^\top = \mathrm{EIG}(\nabla\boldsymbol{\sigma}(W_{t+1}\boldsymbol{x}_t))$

$B_t = PD^{\frac{1}{2}} \otimes \boldsymbol{x}_t$

$\bar{U}\bar{\Sigma}\bar{V}^\top = \mathrm{SVD}([V_t\Lambda_t^{\frac{1}{2}}, B_t]^\top)$

$\bar{\sigma}_{m+1} = \bar{\Sigma}_{m+1,m+1}$

$\Lambda_{t+1} = \bar{\Sigma}_{1:m}^2 - \bar{\sigma}_{m+1}^2 I_m$

$V_{t+1} = \bar{V}_{:,1:m}$

$\rho_{1:t} = \rho_{1:t-1} + \frac{\bar{\sigma}_{m+1}^2}{2}$

return $(V_{t+1}, \Lambda_{t+1}, \rho_{1:t})$

3.2 EOFD-MLogB

We propose EOFD-MLogB (**E**fficient **O**ptimism via robust **F**requent **D**irections), whose time and space complexities scale linearly with the dimension d.

3.2.1 Spectral Maintenance via Adapted RFD

Unlike prior works that maintain the sketch matrix directly (Kuzborskij et al., 2019; Luo et al., 2016), EOFD-MLogB maintains a low-rank spectral decomposition of the sketched Hessian. We approximate $H_t = \lambda I_{Kd} + \sum_{s=1}^{t-1} \nabla^2\ell_s(W_{s+1})$ by

$$Z_t = (\lambda + \rho_{1:t-1})I_{Kd} + V_t\Lambda_t V_t^\top, \tag{9}$$

where $\Lambda_t \in \mathbb{R}^{m\times m}$ is diagonal and positive semidefinite, and $V_t \in \mathbb{R}^{Kd\times m}$ has orthonormal columns. This spectral form is crucial for the efficient parameter update and optimistic reward computation below.

Standard RFD is designed for vector streams, while each Hessian update in MLogB is a rank-K matrix: $\nabla^2\ell_t(W_{t+1}) = \nabla\boldsymbol{\sigma}(W_{t+1}\boldsymbol{x}_t) \otimes \boldsymbol{x}_t\boldsymbol{x}_t^\top$ (cf. Appendix B.1). Given previous sketch $(V_t, \Lambda_t, \rho_{1:t-1})$ and the new Hessian update $\nabla^2\ell_t(W_{t+1})$, to apply RFD, we perform eigen-decomposition

$$PDP^\top = \mathrm{EIG}(\nabla\boldsymbol{\sigma}(W_{t+1}\boldsymbol{x}_t)). \tag{10}$$

Then the Hessian update admits the block factorization

$$\nabla^2\ell_t(W_{t+1}) = B_tB_t^\top, \qquad B_t = PD^{1/2} \otimes \boldsymbol{x}_t \in \mathbb{R}^{Kd\times K}. \tag{11}$$

EOFD-MLogB performs a single block RFD update by applying SVD to $[V_t\Lambda_t^{1/2}, B_t]^\top \in \mathbb{R}^{(m+K)\times Kd}$. Specifically, compute

$$\bar{U}\bar{\Sigma}\bar{V}^\top = \mathrm{SVD}\Big([V_t\Lambda_t^{1/2}, B_t]^\top\Big), \tag{12}$$

and update

$$\Lambda_{t+1} = \bar{\Sigma}_{1:m}^2 - \bar{\sigma}_{m+1}^2 I_m, \qquad V_{t+1} = \bar{V}_{:,1:m}. \tag{13}$$

where $\bar{\sigma}_{m+1}$ is the $(m+1)$-th singular value, $\bar{\Sigma}_{1:m} = \mathrm{diag}(\sigma_1, \ldots, \sigma_m)$ contains the top m singular values, and $\bar{V}_{:,1:m}$ contains the corresponding right singular vectors. The discarded mass $\bar{\sigma}_{m+1}^2$ is accumulated into the sketching error $\rho_{1:t} = \rho_{1:t-1} + \bar{\sigma}_{m+1}^2/2$. The procedure is summarized in Algorithm 1.

3.2.2 Optimistic Reward Construction and Parameter Update

EOFD-MLogB replaces the full Hessian $H_t = \lambda I_{Kd} + \sum_{s=1}^{t-1} \nabla^2\ell_s(W_{s+1})$ by the sketched Hessian Z_t in both parameter estimation equation 3 and optimistic reward construction equation 5. Using the sketched Hessian Z_t, we reduce multinomial optimistic reward evaluation to a $K \times K$ spectral problem and the constrained ONS update to scalar root-finding in the following.

We assume $m \leq Kd$. At each iteration t, the algorithm selects an action $\boldsymbol{x}_t$ optimistically using the current estimator W_t. In the multinomial case ($K \geq 2$), the main computational bottleneck in evaluating the optimistic reward is the spectral norm $\|Z_t^{-1/2}(I_K \otimes \boldsymbol{x})\|_2^2$. The following lemma exploits the spectral structure of $Z_t = (\lambda + \rho_{1:t-1})I + V_t \Lambda_t V_t^\top$ to express this quantity in terms of a $K \times K$ matrix.

Lemma 1. *For any $\boldsymbol{x} \in \mathcal{A}_t$, let $Z_t = \lambda' I_{Kd} + V_t \Lambda_t V_t^\top$, where $\lambda' = \lambda + \rho_{1:t-1}$. Then $(I_K \otimes \boldsymbol{x})^\top Z_t^{-1}(I_K \otimes \boldsymbol{x}) = A(\boldsymbol{x})$, namely*

$$\left\| Z_t^{-1/2}(I_K \otimes \boldsymbol{x}) \right\|_2^2 = \|A(\boldsymbol{x})\|_2, \tag{14}$$

where $A(\boldsymbol{x}) \in \mathbb{R}^{K \times K}$ is symmetric with entries

$$A(\boldsymbol{x})_{ij} = \sum_{k=1}^{m} \left(\frac{1}{\mu_k + \lambda'} - \frac{1}{\lambda'} \right) \langle \boldsymbol{x}, \boldsymbol{u}_{ki} \rangle \langle \boldsymbol{x}, \boldsymbol{u}_{kj} \rangle + \frac{\|\boldsymbol{x}\|_2^2}{\lambda'} \delta_{ij}. \tag{15}$$

Here, μ_k is the k-th diagonal entry of Λ_t, $\boldsymbol{u}_{ki} \in \mathbb{R}^d$ is the i-th block of d consecutive entries in the k-th column of V_t, and δ_{ij} is the Kronecker delta.

By Lemma 1, the optimistic reward can be evaluated as

$$R_t(\boldsymbol{x}, \delta) = \boldsymbol{r}^\top \boldsymbol{\sigma}(W_t \boldsymbol{x}) + \sqrt{\beta_t(\delta)} \sqrt{v^\top A(\boldsymbol{x}) v} + \frac{1}{2} \|\boldsymbol{r}\|_1 \beta_t(\delta) \|A(\boldsymbol{x})\|_2, \tag{16}$$

where $v = \nabla \boldsymbol{\sigma}(W_t \boldsymbol{x}) \boldsymbol{r}$. Thus, both exploration terms can be computed using the same small matrix $A(\boldsymbol{x})$. In particular, the first norm term in equation 8 satisfies

$$\|(\nabla \boldsymbol{\sigma}(W_t \boldsymbol{x}) \otimes \boldsymbol{x}) \boldsymbol{r}\|_{Z_t^{-1}}^2 = v^\top (I_K \otimes \boldsymbol{x}^\top) Z_t^{-1} (I_K \otimes \boldsymbol{x}) v = v^\top A(\boldsymbol{x}) v.$$

In the binary case ($K = 1$), the monotonicity of σ yields a simpler decision rule consisting of a linear estimate and a norm-based exploration bonus (Zhang et al., 2016):

$$\boldsymbol{x}_t = \arg\max_{\boldsymbol{x} \in \mathcal{A}_t} \max_{\boldsymbol{w} \in \mathcal{C}_t} \sigma(\boldsymbol{w}^\top \boldsymbol{x}) = \arg\max_{\boldsymbol{x} \in \mathcal{A}_t} \boldsymbol{w}_t^\top \boldsymbol{x} + \sqrt{\beta_t(\delta)} \|\boldsymbol{x}\|_{Z_t^{-1}}. \tag{17}$$

The bonus $\|\boldsymbol{x}\|_{Z_t^{-1}}$ can be computed efficiently via the Woodbury identity: $Z_t^{-1} = \lambda'^{-1}(I_{Kd} - V_t \widehat{\Lambda}_t V_t^\top)$, where $\widehat{\Lambda}_t = \Lambda_t(\lambda' I_m + \Lambda_t)^{-1}$.

After submitting $\boldsymbol{x}_t$ and receiving feedback y_t, the algorithm updates the parameter estimate. Since the parameter update in equation 6 and equation 7 involves $\widetilde{H}_t = H_t + \eta \nabla^2 \ell_t(W_t)$, the algorithm maintains a surrogate Hessian $\widetilde{Z}_t = Z_t + \eta \nabla^2 \ell_t(W_t)$ and decomposes it into spectral form. Similar to Algorithm 1, we compute the eigen-decomposition of $\nabla \boldsymbol{\sigma}(W_t \boldsymbol{x}_t)$ and then perform an SVD on the concatenated matrix:

$$\nabla \boldsymbol{\sigma}(W_t \boldsymbol{x}_t) = P D P^\top \tag{18}$$

$$\widetilde{U}_t \widetilde{\Sigma}_t \widetilde{V}_t^\top = \mathrm{SVD}\left([V_t \Lambda_t^{\frac{1}{2}}, P(\eta D)^{\frac{1}{2}} \otimes \boldsymbol{x}_t]^\top \right) \tag{19}$$

$$\widetilde{\Lambda}_t = \widetilde{\Sigma}_t^2 \tag{20}$$

This yields the decomposition $\widetilde{Z}_t = (\lambda + \rho_{1:t-1}) I_{Kd} + \widetilde{V}_t \widetilde{\Lambda}_t \widetilde{V}_t^\top$. Here, $\widetilde{V}_t \in \mathbb{R}^{Kd \times (m+K)}$ has orthonormal columns and $\widetilde{\Lambda}_t \in \mathbb{R}^{(m+K) \times (m+K)}$ is diagonal.

The gradient step equation 6 is computed as $\overrightarrow{W}' = \overrightarrow{W}_t - \eta \widetilde{Z}_t^{-1} \nabla \ell_t(W_t)$, which can be efficiently calculated using the Woodbury identity $\widetilde{Z}_t^{-1} = \left(I_{Kd} - \widetilde{V}_t \widetilde{\Lambda}_t ((\lambda + \rho_{1:t-1}) I_{m+K} + \widetilde{\Lambda}_t)^{-1} \widetilde{V}_t^\top \right) / (\lambda + \rho_{1:t-1})$. To avoid the high-dimensional projection in equation 7, the algorithm leverages the spectral structure of $\widetilde{Z}_t = (\lambda + \rho_{1:t-1}) I + \widetilde{V}_t \widetilde{\Lambda}_t \widetilde{V}_t^\top$ to reduce it to a one-dimensional root-finding problem, as formalized in the following lemma:

Lemma 2. *Let $\widetilde{Z}_t = \lambda' I_{Kd} + \widetilde{V}_t \widetilde{\Lambda}_t \widetilde{V}_t^\top$, $\lambda' = \lambda + \rho_{1:t-1}$, where $\widetilde{\boldsymbol{v}}_i$ is the i-th column of $\widetilde{V}_t$ and μ_i is the i-th diagonal element of $\widetilde{\Lambda}_t$. If $\|\overrightarrow{W}'\|_2 \leq S$, then $\overrightarrow{W}_{t+1} = \overrightarrow{W}'$. Otherwise, the projection is given by:*

$$\overrightarrow{W}_{t+1} = \left(\frac{\lambda'}{\lambda' + \nu} I + \frac{\nu}{\lambda' + \nu} \widetilde{V}_t \widetilde{\Lambda}_t M \widetilde{V}_t^\top \right) \overrightarrow{W}', \tag{21}$$

Algorithm 2: EOFD-MLogB

Input: regularization parameter λ, confidence level δ, step size η, parameter domain radius S, sketch size m

Initialize $W_1 = 0$, $\Lambda_1 = 0_{m\times m}$, arbitrary orthonormal $V_1 \in \mathbb{R}^{Kd\times m}$, $\rho_{1:0} = 0$; define $Z_t = (\lambda + \rho_{1:t-1})I_{Kd} + V_t\Lambda_t V_t^\top$. Assume $m \le Kd$;

for $t = 1, 2, \ldots, T$ **do**

  if $K = 1$ **then**

    $\boldsymbol{x}_t = \arg\max_{\boldsymbol{x}\in\mathcal{A}_t} \boldsymbol{w}_t^\top \boldsymbol{x} + \sqrt{\beta_t(\delta)}\|\boldsymbol{x}\|_{Z_t^{-1}}$

  else

    Construct $A(\boldsymbol{x})$ by equation 15 for all $\boldsymbol{x} \in \mathcal{A}_t$, and choose $\boldsymbol{x}_t = \arg\max_{\boldsymbol{x}\in\mathcal{A}_t} R_t(\boldsymbol{x}, \delta)$ by equation 16

  Submit $\boldsymbol{x}_t$ and receive feedback y_t ;

  Form $\widetilde{Z}_t = (\lambda + \rho_{1:t-1})I_{Kd} + \widetilde{V}_t\widetilde{\Lambda}_t\widetilde{V}_t^\top$ by equation 18–equation 20 ;

  $\overrightarrow{W}' = \overrightarrow{W}_t - \eta\widetilde{Z}_t^{-1}\nabla\ell_t(W_t)$;

  if $\|\overrightarrow{W}'\|_2 \le S$ **then**

    $\overrightarrow{W}_{t+1} = \overrightarrow{W}'$

  else

    Find root ν of equation 22, and update $\overrightarrow{W}_{t+1}$ by equation 21

  Pass $\nabla\boldsymbol{\sigma}(W_{t+1}\boldsymbol{x}_t) \otimes \boldsymbol{x}_t\boldsymbol{x}_t^\top$ to Algorithm 1 to update the sketch $(\Lambda_{t+1}, V_{t+1}, \rho_{1:t})$

where $M = ((\lambda' + \nu)I_{m+K} + \widetilde{\Lambda}_t)^{-1}$ is diagonal. The dual variable $\nu > 0$ is the unique root of

$$f(\nu) = S^2, \tag{22}$$

where $f(\nu)$ is a monotonically decreasing function on the interval $(0, +\infty)$

$$f(\nu) = \frac{(\|\overrightarrow{W}'\|_2^2 - \|\widetilde{V}_t^\top\overrightarrow{W}'\|_2^2)\lambda'^2}{(\lambda' + \nu)^2} + \sum_{i=1}^{m+K}\frac{(\widetilde{\boldsymbol{v}}_i^\top\overrightarrow{W}')^2(\mu_i + \lambda')^2}{(\mu_i + \lambda' + \nu)^2}. \tag{23}$$

The coefficients in equation 22 can be precomputed before root-finding. Since $f(\nu)$ is a scalar function, ν can be found efficiently via Newton's method or bisection, and errors can be controlled within machine precision. The existence and uniqueness of the root ν is guaranteed by the monotonicity of $f(\nu)$. (c.f. Appendix B.3 for proof.) Algorithm 2 summarizes the complete EOFD-MLogB procedure.

Computational Complexity. For space complexity, OFUL-MLogB stores the full $Kd \times Kd$ Hessian and hence requires $\mathcal{O}(K^2d^2)$ space. In contrast, EOFD-MLogB maintains the sketched decompositions (V_t, Λ_t) and $(\widetilde{V}_t, \widetilde{\Lambda}_t)$, using $\mathcal{O}(Kd(m + K))$ space. For time complexity, OFUL-MLogB costs $\mathcal{O}(|\mathcal{A}_t|K^3d^2 + K^3d^3)$ per round, as discussed in Section 3.1. EOFD-MLogB evaluates each candidate action in $\mathcal{O}(Kdm + K^2m + K^3)$ time, while both the parameter update and the block RFD update cost $\mathcal{O}(Kd(m + K)^2)$. Thus, the total per-round time complexity is $\mathcal{O}(|\mathcal{A}_t|(Kdm + K^2m + K^3) + Kd(m + K)^2)$. If we further emphasize the dependence on d, m and K, the time complexity can be expressed as $\mathcal{O}(Kd(m + K)^2)$, which is linear in d, compared with the cubic dependence of OFUL-MLogB.

3.3 Regret Analysis

In this section, we establish theoretical guarantees for EOFD-MLogB. Let $\Delta_t = 1 + 2\rho_{1:t}/\lambda$, where $\rho_{1:t}$ is the cumulative RFD sketching error associated with Z_t, defined in line 7 of Algorithm 1. We first establish a confidence set for the unknown parameter W_*.

Theorem 1 (Confidence Region)**.** *Let $\alpha = 2(1 + SD) + \ln(K + 1)$, and $\Delta_t = 1 + 2\rho_{1:t}/\lambda$. Set $\lambda \ge \max\{2LD^2, \alpha\max\{2\sqrt{6}SLD^2, 24Kd\}\}$ and $\eta = \alpha/2$. For any $\delta \in (0, 1)$, with probability at least $1 - \delta$:*

$$W_\star \in \mathcal{C}_t = \left\{ W \in \mathbb{R}^{K\times d} \,\middle|\, \left\|\overrightarrow{W} - \overrightarrow{W}_t\right\|_{Z_t}^2 \le \beta_t(\delta) \right\}, \quad \forall t \ge 1, \tag{24}$$

where the radius

$$\beta_t(\delta) = \mathcal{O}\left(\Delta_t \ln^2 K \ln t(m \ln t + Kd \ln \Delta_t + \ln \frac{1}{\delta})\right). \tag{25}$$

The precise definition of $\beta_t(\delta)$ is given in equation 34 in appendix, and the detailed proof is provided in Appendix B.2. The confidence region is defined with respect to the sketched matrix Z_t instead of the full Hessian H_t. Compared with the OFUL-MLogB confidence squared radius $\widetilde{\beta}_t(\delta) = \mathcal{O}(\ln^2 K \ln t(Kd \ln t + \ln \frac{1}{\delta}))$ (Zhang & Sugiyama, 2023, Theorem 3), the sketched squared radius replaces $Kd \ln t$ by $Kd \ln \Delta_t + m \ln t$ and introduces the multiplicative factor Δ_t. These terms quantify the cumulative approximation error introduced by RFD.

Remark 1 (Independence of κ). Theorem 1 preserves the κ-independence of OFUL-MLogB. This follows because EOFD-MLogB keeps the same second-order surrogate in equation 3 and index-shifted Hessian $H_t = \lambda I + \sum_{s=1}^{t-1} \nabla^2 \ell_s(W_{s+1})$ as OFUL-MLogB, while the singular values discarded by RFD are controlled by the upper bound $\mathcal{O}(LD^2)$ on Hessian updates rather than the strong-convexity lower bound $1/\kappa$.

We next derive the regret bound for the multinomial case based on Theorem 1.

Theorem 2 (Multinomial Regret)**.** *Under the conditions of Theorem 1, with probability at least* $1 - \delta$*,*

$$\mathrm{Regret}(T) = \widetilde{\mathcal{O}}\left(\Delta_T(Kd \ln \Delta_T + m)\sqrt{T}\right). \tag{26}$$

Theorem 2 achieves similar regret bound of OFUL-MLogB $\widetilde{\mathcal{O}}(Kd\sqrt{T})$ (Zhang & Sugiyama, 2023, Theorem 4) except the sketching factor Δ_T and the replacement of Kd by $Kd \ln \Delta_T + m$. When $m \geq Kd$, no truncation occurs, so $\rho_{1:T} = 0$ and $\Delta_T = 1$, recovering the OFUL-MLogB bound.

The sketching error is controlled by the spectral tail of the cumulative Hessian. By the RFD guarantee (Ghashami et al., 2016, Theorem 1.1) (Luo et al., 2019, Theorem 3),

$$2\rho_{1:T} \leq \min_{0 \leq k < m} \frac{1}{m-k} \left\| M_{T+1}^{1/2} - (M_{T+1}^{1/2})_{(k)} \right\|_F^2, \tag{27}$$

where $M_{T+1} = \sum_{t=1}^{T+1} \nabla^2 \ell_t(W_{t+1})$ denotes the cumulative Hessian and $(M_{T+1}^{1/2})_{(k)}$ is the best rank-k approximation of $M_{T+1}^{1/2}$. Thus, if the accumulated Hessian M_{T+1} is approximately low-rank, EOFD-MLogB remains close to OFUL-MLogB in regret while substantially reducing computation. If the Hessian has slow spectral decay, the sketching error may grow and the regret bound correspondingly deteriorates.

Remark 2 (Extension to Related MLogB Frameworks). The proposed spectral reductions are not specific to OFUL-MLogB. For example, Boudart et al. (2026) use a similar second-order optimistic bonus and ONS-type update, differing mainly in the initialization phase. Replacing the full Hessian by Z_t and applying Lemma 1 and Lemma 2 yields analogous efficiency gains and regret guarantees. We focus on OFUL-MLogB because it naturally handles time-varying action sets $\mathcal{A}_t$.

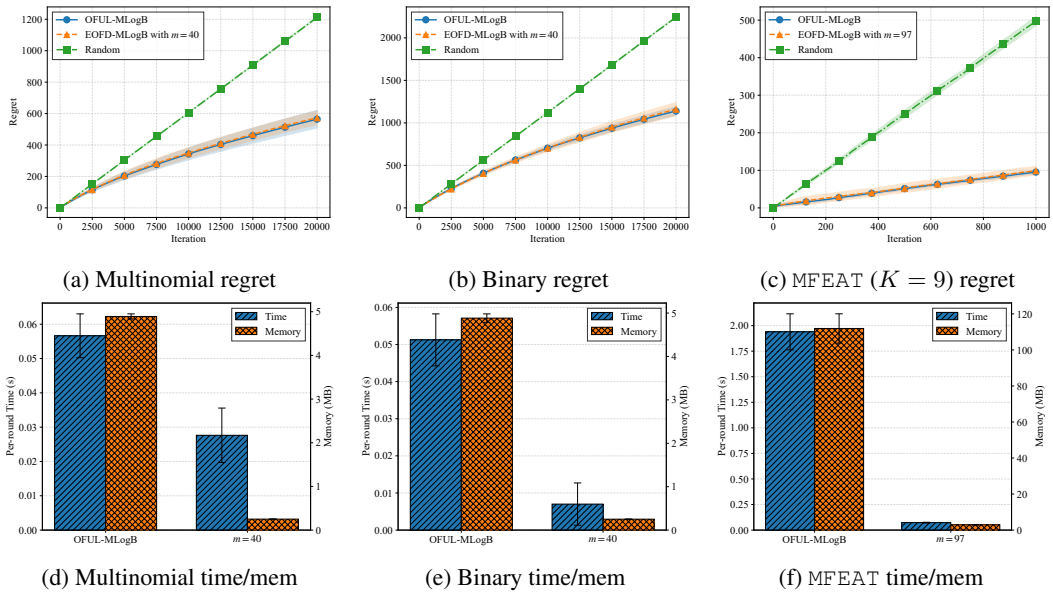
For completeness, we also state the binary-case guarantee, which admits the simplified optimistic rule and tighter regret bound.

Proposition 1 (Binary Regret)**.** *Under the conditions of Theorem 1, if* $K = 1$*, with probability at least* $1 - \delta$*,*

$$\mathrm{Regret}(T) = \widetilde{\mathcal{O}}\left(\Delta_T(d \ln \Delta_T + m)\sqrt{T/\kappa_*}\right), \tag{28}$$

where $1/\kappa_* = \sum_{t=1}^{T} \sigma'(\boldsymbol{w}_\star^\top \boldsymbol{x}_t^*)/T$*, and* $\boldsymbol{x}_t^*$ *is the optimal action in* t*-th round.*

The binary case follows Zhang & Sugiyama (2023, Corollary 1): the monotonicity of the scalar logistic reward reduces the optimistic decision rule to a simple vector norm, avoiding the multinomial spectral-norm bonus for the second-order reward approximation. Theoretically, it yields the refined $\sqrt{1/\kappa_*}$ dependence, while retaining the sketching factor Δ_T.



(a) Multinomial regret (b) Binary regret (c) `MFEAT` ($K = 9$) regret

(d) Multinomial time/mem (e) Binary time/mem (f) `MFEAT` time/mem

Figure 1: Empirical performance of EOFD-MLogB compared to OFUL-MLogB and `Random` baselines. Top row: Cumulative regret over horizon T. Bottom row: Per-round execution time and peak memory usage across 10 independent trials.

4 Experiments

This section empirically evaluates EOFD-MLogB on synthetic benchmarks and the real-world `MFEAT` dataset from `OpenML` (Vanschoren et al., 2014). Our goal is to verify that the theoretical space-time savings translate into substantial wall-clock speedups without compromising regret. More experiments can be founded in Appendix C.

Experimental Setup. We evaluate against the full-rank OFUL-MLogB (Zhang & Sugiyama, 2023) and a `Random` action baseline. Benchmark settings include multinomial synthetic data with $(d, K) = (100, 4)$, binary synthetic data with $(d, K) = (400, 1)$, and real-world `MFEAT` ($d = 217, K = 9$). We set $\delta = 1/T$, $m = \lfloor Kd/10 \rfloor$, and report metrics averaged over 10 independent trials. More detailed m settings are provided in Appendix C.7. Detailed hyperparameter search grids and environment setups are deferred to Appendix C.

Regret Performance. Figures 1a–1c illustrate cumulative regret trajectories. On synthetic data, EOFD-MLogB achieves regret virtually identical to OFUL-MLogB, verifying that the low-rank robust frequent directions sketch preserves the principal directions necessary for optimistic decision-making. On `MFEAT`, EOFD-MLogB incurs only a minor approximately 5% regret overhead while operating under significantly constrained memory and update limits.

Efficiency Gains. Figures 1d–1f display average per-round runtime and peak memory usage. EOFD-MLogB reduces per-round runtime by 35%–80% and saves over 80% state memory across all domains. These results demonstrate that EOFD-MLogB substantially improves computational efficiency while preserving competitive regret.

5 Conclusion

We presented EOFD-MLogB, an efficient algorithm using robust frequent directions to address the computational bottlenecks of high-dimensional MLogB. By maintaining a low-rank spectral decomposition, EOFD-MLogB reduces parameter projection to 1D root-finding and optimistic reward evaluation to a $K \times K$ eigenvalue problem, achieving $\mathcal{O}(Kd(m+K)^2)$ time and $\mathcal{O}(Kd(m+K))$ space per round. Theoretically, it retains an $\widetilde{\mathcal{O}}(\Delta_T(Kd\ln\Delta_T + m)\sqrt{T})$ regret bound. Empirically, EOFD-MLogB achieves significant computational and memory savings with competitive regret.

AI USE STATEMENT

(This section is **required** and does not count toward the page limit.)

In this work, we used generative AI tools for assisting with translation. We have not used generative AI tools for generating synthetic data sets, helping develop theoretical models or conceptual frameworks, formulating mathematical claims, providing critical ingredients for proving mathematical claims, assisting in the writing of proofs, proposing or refining hypotheses, designing or providing feedback on research methodology or experiments, implementing methods, cleaning and reformatting dataset, supporting qualitative and thematic data analysis, or interpreting results. Additionally, we used generative AI tools for creating or editing software code, and editing a research paper to improve readability. Specifically, we used generative AI to edit the manuscript to improve language readability, and assist in refactoring code structure. We have reviewed all AI-assisted work: all edited text was thoroughly checked by the authors to ensure accuracy and preservation of technical intent, and all refactored code was verified, debugged, and tested for correctness. We take full responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

(This section is **recommended** and does not count toward the page limit.)

This work complies with the ICLR Code of Ethics. Our research focuses on general algorithmic improvements evaluated on standard, public benchmark datasets. To the best of our knowledge, this work does not raise concerns regarding human subjects, privacy, harmful applications, bias, or potential negative societal impact.

REPRODUCIBILITY STATEMENT

(This section is **recommended** and does not count toward the page limit.)

To ensure the reproducibility of our results, we have provided the following details and resources:

- Code and Implementation: The source code, and execution scripts, are submitted as part of the anonymous supplementary material.
- Experimental Protocols: Dataset baseline model, hyperparameter search grids, and hardware setup details are specified in Section 4 and Appendix C.
- Theoretical Claims: Detailed mathematical proofs, explicit assumptions, and extended theoretical derivations are fully expanded in Appendix A and B.

REFERENCES

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. In *Advances in Neural Information Processing Systems*, volume 24, pp. 2312–2320, 2011.

Marc Abeille, Louis Faury, and Clement Calauzenes. Instance-wise minimax-optimal algorithms for logistic bandits. In *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, pp. 3691–3699, 2021.

Alekh Agarwal, Haipeng Luo, Behnam Neyshabur, and Robert E Schapire. Corralling a band of bandit algorithms. In *Conference on Learning Theory*, pp. 12–38, 2017.

Sanae Amani and Christos Thrampoulidis. UCB-based algorithms for multinomial logistic regression bandits. In *Advances in Neural Information Processing Systems*, volume 34, pp. 2913–2924, 2021.

Richard G. Baraniuk, Volkan Cevher, and Michael B. Wakin. Low-dimensional models for dimensionality reduction and signal recovery: A geometric perspective. *Proceedings of the IEEE*, 98: 959–971, 2010.

Pierre Boudart, Pierre Gaillard, and Alessandro Rudi. Enjoying non-linearity in multinomial logistic bandits: A minimax-optimal algorithm. arXiv:2507.05306, 2026.

Wang Chi Cheung, David Simchi-Levi, and Ruihao Zhu. Learning to optimize under non-stationarity. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pp. 1079–1087, 2019.

Qin Ding, Cho-Jui Hsieh, and James Sharpnack. An efficient algorithm for generalized linear bandit: Online stochastic gradient descent and thompson sampling. In *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130, pp. 1585–1593, 2021.

Petros Drineas, Ravi Kannan, and Michael W. Mahoney. Fast monte carlo algorithms for matrices i: Approximating matrix multiplication. *SIAM Journal on Computing*, 36(1):132–157, 2006.

Louis Faury, Marc Abeille, Kwang-Sung Jun, and Clement Calauzenes. Jointly efficient and optimal algorithms for logistic bandits. In *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151, pp. 546–580, 2022.

Sarah Filippi, Olivier Cappe, Aurélien Garivier, and Csaba Szepesvári. Parametric bandits: The generalized linear case. In *Advances in Neural Information Processing Systems*, volume 23, pp. 586–594, 2010.

Mina Ghashami, Edo Liberty, Jeff M. Phillips, and David P. Woodruff. Frequent directions: Simple and deterministic matrix sketching. *SIAM Journal on Computing*, 45(5):1762–1792, 2016.

Elad Hazan, Amit Agarwal, and Satyen Kale. Logarithmic regret algorithms for online convex optimization. *Machine Learning*, 69:169–192, 2007.

Rémi Jézéquel, Pierre Gaillard, and Alessandro Rudi. Mixability made efficient: Fast online multiclass logistic regression. In *Advances in Neural Information Processing Systems*, volume 34, pp. 23692–23702, 2021.

Ilja Kuzborskij, Leonardo Cella, and Nicolò Cesa-Bianchi. Efficient linear bandits through matrix sketching. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics*, pp. 177–185, 2019.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

Joongkyu Lee and Min-Hwan Oh. Improved online confidence bounds for multinomial logistic bandits. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267, pp. 33576–33615, 2025.

Junghyun Lee, Se-Young Yun, and Kwang-Sung Jun. Improved regret bounds of (multinomial) logistic bandits via regret-to-confidence-set conversion. In *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, volume 238, pp. 4474–4482, 2024.

Lihong Li, Wei Chu, John Langford, and Robert E. Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th International Conference on World Wide Web*, pp. 661–670, 2010.

Lihong Li, Yu Lu, and Dengyong Zhou. Provably optimal algorithms for generalized linear contextual bandits. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70, pp. 2071–2080, 2017.

Haipeng Luo, Alekh Agarwal, Nicolò Cesa-Bianchi, and John Langford. Efficient second order online learning by sketching. In *Advances in Neural Information Processing Systems*, volume 29, 2016.

Luo Luo, Cheng Chen, Zhihua Zhang, Wu-Jun Li, and Tong Zhang. Robust frequent directions with application in online learning. *Journal of Machine Learning Research*, 20(45):1–41, 2019.

Quoc Tran-Dinh, Yen-Huan Li, and Volkan Cevher. Composite convex minimization involving self-concordant-like cost functions. In *Proceedings of the 3rd International Conference on Modelling, Computation and Optimization in Information Systems and Management Sciences*, pp. 155–168, 2015.

Joaquin Vanschoren, Jan N Van Rijn, Bernd Bischl, and Luis Torgo. Openml: networked science in machine learning. *ACM SIGKDD Explorations Newsletter*, 15(2):49–60, 2014.

Yuanyu Wan and Lijun Zhang. Efficient adaptive online learning via frequent directions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):6910–6923, 2022.

Dongxie Wen, Hanyan Yin, Xiao Zhang, and Zhewei Wei. Matrix sketching in bandits: Current pitfalls and new framework, 2024.

Dongxie Wen, Hanyan Yin, Xiao Zhang, Peng Zhao, Lijun Zhang, and Zhewei Wei. Revisiting matrix sketching in linear bandits: Achieving sublinear regret via dyadic block sketching. In *The 14th International Conference on Learning Representations*, 2026.

Sifan Yang, Yuanyu Wan, Peijia Li, Yibo Wang, Xiao Zhang, Zhewei Wei, and Lijun Zhang. Dimension-free adaptive subgradient methods with frequent directions. In *Proceedings of The 42nd International Conference on Machine Learning*, 2025.

Lijun Zhang, Tianbao Yang, Rong Jin, Yichi Xiao, and Zhi-hua Zhou. Online stochastic linear optimization under one-bit feedback. In *Proceedings of The 33rd International Conference on Machine Learning*, pp. 392–401, 2016.

Yu-Jie Zhang and Masashi Sugiyama. Online (multinomial) logistic bandit: Improved regret and constant computation cost. In *Advances in Neural Information Processing Systems*, volume 36, pp. 29741–29782, 2023.

Yu-Jie Zhang, Sheng-An Xu, Peng Zhao, and Masashi Sugiyama. Generalized linear bandits: Almost optimal regret with one-pass update. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

A Properties of Algorithm 1

In this section, we provide some useful lemmas of Algorithm 1, and corresponding necessary proofs.

Lemma 3.

$$V_{t+1}\Lambda_{t+1}V_{t+1}^\top \preceq V_t\Lambda_t V_t^\top + \nabla^2\ell_t(W_{t+1}) \preceq V_{t+1}\Lambda_{t+1}V_{t+1}^\top + 2(\rho_{1:t}-\rho_{1:t-1})I_{Kd}.$$

Proof. By the eigen-decomposition and the definition of B_t in line 1 and line 2 of Algorithm 1, we have

$$[V_t\Lambda_t^{\frac{1}{2}}, B_t][V_t\Lambda_t^{\frac{1}{2}}, B_t]^\top = V_t\Lambda_t V_t^\top + B_tB_t^\top = V_t\Lambda_t V_t^\top + PDP^\top \otimes \boldsymbol{x}_t\boldsymbol{x}_t^\top = V_t\Lambda_t V_t^\top + \nabla^2\ell_t(W_{t+1}).$$

On the other hand, by the definition of $\bar{U}, \bar{\Sigma}, \bar{V}$ in line 3, we have

$$\bar{V}\bar{\Sigma}^2\bar{V}^\top = [V_t\Lambda_t^{\frac{1}{2}}, B_t][V_t\Lambda_t^{\frac{1}{2}}, B_t]^\top.$$

Thus

$$\bar{V}\bar{\Sigma}^2\bar{V}^\top = V_t\Lambda_t V_t^\top + \nabla^2\ell_t(W_{t+1}).$$

Let $\bar{\sigma}_i$ be the i-th largest singular value of $\bar{\Sigma}$. Then we have

$$\bar{\sigma}_1^2 \geq \cdots \geq \bar{\sigma}_m^2 \geq \bar{\sigma}_{m+1}^2 \geq \bar{\sigma}_{m+2}^2 \geq \cdots \geq \bar{\sigma}_{m+K}^2 \geq 0.$$

Let $\widetilde{\Sigma} = \mathrm{diag}(\bar{\sigma}_{m+1}, \ldots, \bar{\sigma}_{m+1}, \bar{\sigma}_{m+2}, \bar{\sigma}_{m+3}, \ldots, \bar{\sigma}_{m+K}) \in \mathbb{R}^{(m+K)\times(m+K)}$. By line 5 and line 6, we have

$$V_{t+1}\Lambda_{t+1}V_{t+1}^\top + \bar{V}\widetilde{\Sigma}^2\bar{V}^\top = \bar{V}\bar{\Sigma}^2\bar{V}^\top.$$

To conclude

$$V_{t+1}\Lambda_{t+1}V_{t+1}^\top + \bar{V}\widetilde{\Sigma}^2\bar{V}^\top = V_t\Lambda_t V_t^\top + \nabla^2\ell_t(W_{t+1}).$$

By $\bar{V}\widetilde{\Sigma}^2\bar{V}^\top \succeq 0$, we have the first inequality of the claim:

$$V_{t+1}\Lambda_{t+1}V_{t+1}^\top \preceq V_t\Lambda_t V_t^\top + \nabla^2\ell_t(W_{t+1}).$$

And by $\bar{V}\widetilde{\Sigma}^2\bar{V}^\top \preceq \bar{\sigma}_{m+1}^2 I_{Kd}$, we have the second inequality of the claim:

$$V_t\Lambda_t V_t^\top + \nabla^2\ell_t(W_{t+1}) \preceq V_{t+1}\Lambda_{t+1}V_{t+1} + \bar{\sigma}_{m+1}^2 I_{Kd} = V_{t+1}\Lambda_{t+1}V_{t+1} + 2(\rho_{1:t}-\rho_{1:t-1})I_{Kd}.$$

□

Lemma 4 (Property 1, 2 of Ghashami et al. (2016)). *Let $M_t = \sum_{s=1}^{t-1} \nabla^2 \ell_s(W_{s+1})$ be the cumulative Hessian, then*

$$V_t \Lambda_t V_t^\top \preceq M_t \preceq V_t \Lambda_t V_t^\top + 2\rho_{1:t-1} I_{Kd}.$$

Proof. By the first inequality of Lemma 3, we have

$$V_{s+1}\Lambda_{s+1}V_{s+1}^\top - V_s \Lambda_s V_s^\top \preceq \nabla^2 \ell_s(W_{s+1}).$$

Summing over s from 1 to $t-1$, we have

$$V_t \Lambda_t V_t^\top \preceq \sum_{s=1}^{t-1} \nabla^2 \ell_s(W_{s+1}) + V_1 \Lambda_1 V_1^\top = M_t.$$

By the second inequality of Lemma 3, we have

$$\nabla^2 \ell_s(W_{s+1}) \preceq V_{s+1}\Lambda_{s+1}V_{s+1} - V_s \Lambda_s V_s^\top + 2(\rho_{1:s} - \rho_{1:s-1}) I_{Kd}.$$

Summing over s from 1 to $t-1$, we have

$$M_t \preceq V_t \Lambda_t V_t^\top + \sum_{s=1}^{t-1} 2(\rho_{1:s} - \rho_{1:s-1}) I_{Kd} = V_t \Lambda_t V_t^\top + 2\rho_{1:t-1} I_{Kd}.$$

□

Lemma 5. *Let $M_t = \sum_{s=1}^{t-1} \nabla^2 \ell_s(W_{s+1})$ be the cumulative Hessian, then for any $\lambda > 0$,*

$$V_t \Lambda_t V_t^\top + \lambda I_{Kd} \preceq M_t + \lambda I_{Kd} \preceq 2\left(V_t \Lambda_t V_t^\top + (\lambda + \rho_{1:t-1}) I_{Kd}\right)$$

Proof. We follow the main sketch of lemma 9 of Kuzborskij et al. (2019). By the second inequality of Lemma 4, the first inequality of the claim can be proved easily.

For the second inequality, let $(\lambda_i, \boldsymbol{v}_i)$ be the i-th full eigen-pair of $V_t \Lambda_t V_t^\top$, then we have the decomposition

$$V_t \Lambda_t V_t^\top + (\lambda + \rho_{1:t-1}) I_{Kd} = \sum_{i=1}^{Kd} (\lambda_i + \lambda + \rho_{1:t-1}) \boldsymbol{v}_i \boldsymbol{v}_i^\top.$$

And by the first inequality of Lemma 4

$$M_t \preceq V_t \Lambda_t V_t^\top + 2\rho_{1:t-1} I_{Kd} = \sum_{i=1}^{Kd} (\lambda_i + \lambda + 2\rho_{1:t-1}) \boldsymbol{v}_i \boldsymbol{v}_i^\top.$$

Therefore

$$\begin{aligned}
&(V_t \Lambda_t V_t^\top + (\lambda + \rho_{1:t-1}) I_{Kd})^{-\frac{1}{2}} (M_t + \lambda I_{Kd}) (V_t \Lambda_t V_t^\top + (\lambda + \rho_{1:t-1}) I_{Kd})^{-\frac{1}{2}} \\
\preceq\ & (V_t \Lambda_t V_t^\top + (\lambda + \rho_{1:t-1}) I_{Kd})^{-\frac{1}{2}} (V_t \Lambda_t V_t^\top + (2\rho_{1:t-1} + \lambda) I_{Kd}) (V_t \Lambda_t V_t^\top + (\lambda + \rho_{1:t-1}) I_{Kd})^{-\frac{1}{2}} \\
\preceq\ & \sum_{i=1}^{Kd} \frac{\lambda_i + \lambda + 2\rho_{1:t-1}}{\lambda_i + \lambda + \rho_{1:t-1}} \boldsymbol{v}_i \boldsymbol{v}_i^\top \\
\preceq\ & \sum_{i=1}^{Kd} \frac{\lambda + 2\rho_{1:t-1}}{\lambda + \rho_{1:t-1}} \boldsymbol{v}_i \boldsymbol{v}_i^\top = \frac{\lambda + 2\rho_{1:t-1}}{\lambda + \rho_{1:t-1}} I
\end{aligned}$$

Left and right multiply $(V_t \Lambda_t V_t^\top + (\lambda + \rho_{1:t-1}) I_{Kd})^{\frac{1}{2}}$ to get

$$M_t + \lambda I_{Kd} \preceq \left(1 + \frac{\rho_{1:t-1}}{\lambda + \rho_{1:t-1}}\right) (V_t \Lambda_t V_t^\top + (\lambda + \rho_{1:t-1}) I_{Kd}) \preceq 2(V_t \Lambda_t V_t^\top + (\lambda + \rho_{1:t-1}) I_{Kd})$$

□

Lemma 6 (Lemma 8 of Kuzborskij et al. (2019)). *Let $M_t = \sum_{s=1}^{t-1} \nabla^2 \ell_s(W_{s+1})$ be the cumulative Hessian, $r = \min\{Kd, m\}$, then for any $\lambda > 0, t > 0$,*

$$\frac{|M_t + \lambda I_{Kd}|}{|\lambda I_{Kd}|} \leq \left(1 + \frac{2\rho_{1:t}}{\lambda}\right)^{Kd} \left(1 + \frac{\mathrm{tr}(M_t)}{r\lambda}\right)^r.$$

Proof. Let $(\mu_i, \boldsymbol{u}_i)$ be the full eigen-pair of M_t, and $(\lambda_i, \boldsymbol{v}_i)$ be the full eigen-pair of $V_t\Lambda_t V_t$, then we have the decomposition

$$M_t + \lambda I_{Kd} = \sum_{i=1}^{Kd} (\mu_i + \lambda)\boldsymbol{u}_i\boldsymbol{u}_i^\top$$
$$V_t\Lambda_t V_t^\top = \sum_{i=1}^{Kd} \lambda_i \boldsymbol{v}_i\boldsymbol{v}_i^\top$$

Therefore,

$$\begin{aligned}
|M_t + \lambda I_d| &= \prod_{i=1}^{Kd}(\mu_i + \lambda) \\
&\text{(by the second inequality of Lemma 4)} \\
&\leq \prod_{i=1}^{Kd}(\lambda_i + 2\rho_{1:t-1} + \lambda) \\
&\text{(the rank of } V_t\Lambda_t V_t \text{ is at most } r) \\
&= (2\rho_{1:t-1} + \lambda)^{Kd-r} \prod_{i=1}^{r}(\lambda_i + 2\rho_{1:t-1} + \lambda) \\
&\text{(by AM-GM inequality)} \\
&\leq (2\rho_{1:t-1} + \lambda)^{Kd-r} \left(\sum_{i=1}^{r}\frac{\lambda_i}{r} + 2\rho_{1:t-1} + \lambda\right)^r \\
&= (2\rho_{1:t-1} + \lambda)^{Kd} \left(\frac{\sum_{i=1}^{r}\lambda_i}{r(2\rho_{1:t-1} + \lambda)} + 1\right)^r \\
&= (2\rho_{1:t-1} + \lambda)^{Kd} \left(\frac{\mathrm{tr}(V_t\Lambda_t V_t)}{r(2\rho_{1:t-1} + \lambda)} + 1\right)^r \\
&\text{(by the first inequality of Lemma 4)} \\
&\leq (2\rho_{1:t-1} + \lambda)^{Kd} \left(1 + \frac{\mathrm{tr}(M_t)}{r(2\rho_{1:t-1} + \lambda)}\right)^r
\end{aligned}$$

And

$$|\lambda I_{Kd}| = \lambda^{Kd},$$

therefore

$$\frac{|M_t + \lambda I_{Kd}|}{|\lambda I_{Kd}|} \leq \left(1 + \frac{2\rho_{1:t-1}}{\lambda}\right)^{Kd} \left(1 + \frac{\mathrm{tr}(M_t)}{r(2\rho_{1:t-1} + \lambda)}\right)^r \leq \left(1 + \frac{2\rho_{1:t}}{\lambda}\right)^{Kd} \left(1 + \frac{\mathrm{tr}(M_t)}{r\lambda}\right)^r.$$

□

B THEOREM IN SECTION 3

In this section, we will give the proof details of Theorem 1, Theorem 2 and Proposition 1.

B.1 PRELIMINARY

In this section, we explain some notations and properties used in the proof. The multinomial logistic function $\boldsymbol{\sigma} : \mathbb{R}^K \mapsto (0,1)^{K+1}$ is defined as

$$\boldsymbol{\sigma}(\boldsymbol{z}) = \frac{1}{1+\sum_{k=1}^K \exp(z_k)} \left[1, \exp(z_1), \exp(z_2), \ldots, \exp(z_K)\right]^\top.$$

However, since the 0-th component of $\boldsymbol{r}$ is $r_0 = 0$, it does not contribute to the reward and the regret. Namely, $\sum_{k=0}^K \rho_k \sigma_k(W\boldsymbol{x}) = \sum_{k=1}^K \rho_k \sigma_k(W\boldsymbol{x})$. Thus we ignore the 0-th component of $\boldsymbol{\sigma}(\boldsymbol{z})$, which is refined as

$$\boldsymbol{\sigma}(\boldsymbol{z}) = \frac{1}{1+\sum_{k=1}^K \exp(z_k)} \left[\exp(z_1), \exp(z_2), \ldots, \exp(z_K)\right]^\top \in (0,1)^K.$$

For a special case where $K = 1$, the reward vector is defined as $\boldsymbol{r} = [0,1]^\top$. Therefore the expected reward for given action $\boldsymbol{x} \in \mathcal{A}$ is

$$\boldsymbol{r}^\top \boldsymbol{\sigma}(\boldsymbol{w}_\star^\top \boldsymbol{x}) = 0 \cdot \sigma_0(\boldsymbol{w}_\star^\top \boldsymbol{x}) + 1 \cdot \sigma_1(\boldsymbol{w}_\star^\top \boldsymbol{x}) = \sigma_1(\boldsymbol{w}_\star^\top \boldsymbol{x}),$$

where $\sigma_1(z) = \exp(z)/(1+\exp(z))$. And the regret analysis can be reduced to the binary logistic function $\sigma_1(z)$ without $\boldsymbol{r}$ involved (Zhang et al., 2016).

The first order derivative of $\boldsymbol{\sigma}(\boldsymbol{z})$ is

$$\nabla \boldsymbol{\sigma}(\boldsymbol{z}) = \operatorname{diag} \boldsymbol{\sigma}(\boldsymbol{z}) - \boldsymbol{\sigma}(\boldsymbol{z})\boldsymbol{\sigma}(\boldsymbol{z})^\top.$$

And the second order derivative of $\boldsymbol{r}^\top \boldsymbol{\sigma}(\boldsymbol{z})$ is

$$D^2 \boldsymbol{\sigma}(\boldsymbol{z})[\boldsymbol{r}] = \sum_{k=1}^K r_k \sigma_k(\boldsymbol{z}) \left(2\boldsymbol{\sigma}(\boldsymbol{z})\boldsymbol{\sigma}(\boldsymbol{z})^\top - \operatorname{diag}(\boldsymbol{\sigma}(\boldsymbol{z})) - \boldsymbol{e}_k \boldsymbol{\sigma}(\boldsymbol{z})^\top - \boldsymbol{\sigma}(\boldsymbol{z}) \boldsymbol{e}_k^\top + \boldsymbol{e}_k \boldsymbol{e}_k^\top\right),$$

which is bounded by

$$-\|\boldsymbol{r}\|_1 I \preceq D^2 \boldsymbol{\sigma}(\boldsymbol{z})[\boldsymbol{r}] \preceq \frac{1}{2}\|\boldsymbol{r}\|_1 I. \tag{29}$$

Let $\boldsymbol{x}_t$ be the action selected at round t and $y_t \in \{0, 1, \ldots, K\}$ be the observed outcome and $\sigma_k(\boldsymbol{z}) = [\boldsymbol{\sigma}(\boldsymbol{z})]_k$ be the k-th component of $\boldsymbol{\sigma}(\boldsymbol{z})$. We denote $\boldsymbol{y}_t = [\mathbb{I}\{y_t = 1\}, \mathbb{I}\{y_t = 2\}, \ldots, \mathbb{I}\{y_t = K\}]^\top$. The logistic loss function $\ell_t : \mathbb{R}^{K \times d} \mapsto \mathbb{R}$ is defined as

$$\ell_t(W) = -\ln \sigma_{y_t}(W\boldsymbol{x}_t),$$

and the first and second order derivative of $\ell_t(W)$ are

$$\begin{aligned}
\nabla \ell_t(W) &= (\boldsymbol{\sigma}(W\boldsymbol{x}_t) - \boldsymbol{y}_t) \otimes \boldsymbol{x}_t \\
\nabla^2 \ell_t(W) &= \nabla \boldsymbol{\sigma}(W\boldsymbol{x}_t) \otimes \boldsymbol{x}_t \boldsymbol{x}_t^\top \\
&= (\operatorname{diag}(\boldsymbol{\sigma}(W\boldsymbol{x}_t)) - \boldsymbol{\sigma}(W\boldsymbol{x}_t)\boldsymbol{\sigma}(W\boldsymbol{x}_t)^\top) \otimes \boldsymbol{x}_t \boldsymbol{x}_t^\top
\end{aligned}$$

And this logistic function has $\sqrt{6}$-self-concordant-like property (Tran-Dinh et al., 2015, Lemma 4) and exp-concave-like property (Jézéquel et al., 2021, Lemma 4), which is

$$\left|\left\langle D^3 \ell_t(W)[\boldsymbol{u}_1]\boldsymbol{u}_2, \boldsymbol{u}_3 \right\rangle\right| \leq \sqrt{6} \|\boldsymbol{u}_1\|_2 \|\boldsymbol{u}_2\|_{\nabla^2 \ell_t(W)} \|\boldsymbol{u}_3\|_{\nabla^2 \ell_t(W)}, \tag{30}$$

and

$$\ell_t(W_1) - \ell_t(W_2) \leq \langle \nabla \ell_t(W_1), W_1 - W_2 \rangle - \frac{1}{\ln(K+1) + 2(SD+1)} \|W_1 - W_2\|^2_{\nabla^2 \ell_t(W_1)}. \tag{31}$$

B.2 Proof for Theorems

We denote $\Delta_t^\lambda = 1 + 2\rho_{1:t}/\lambda$ as the sketching error factor in the following proofs.

Proof for Theorem 1

Proof. We follow the proof of Zhang & Sugiyama (2023, Theorem 3), and emphasize the difference where FD are applied. To prove the theorem 1 using the sketched technique, the FD Lemma 3 and 6 is pivotal. We fix the step size $\eta = \frac{1}{2}(\ln(K+1) + 2(SD+1))$

In the OMD analysis, the surrogate loss function

$$\widetilde{\ell}_t(W) := \left\langle \nabla \ell_t(\overrightarrow{W}_t), \overrightarrow{W} - \overrightarrow{W}_t \right\rangle + \frac{1}{2} \left\| \overrightarrow{W} - \overrightarrow{W}_t \right\|^2_{\nabla^2 \ell_t(W_t)}$$

is a second order approximation of the loss function $\ell_t(W)$ on the point W_t instead of the linear approximation $\left\langle \nabla \ell_t(\overrightarrow{W}_t), \overrightarrow{W} - \overrightarrow{W}_t \right\rangle$ in traditional ONS. And the update rule is equivalent to

$$W_{t+1} := \operatorname*{arg\,min}_{W \in \mathcal{W}} \eta \widetilde{\ell}_t(W) + \frac{1}{2} \left\| \overrightarrow{W} - \overrightarrow{W}_t \right\|^2_{Z_t}.$$

By the first order optimal condition, we have

$$\forall \boldsymbol{u} \in \mathcal{W}, \left\langle \eta \nabla \widetilde{\ell}_t(W_{t+1}) + Z_t(\overrightarrow{W}_{t+1} - \overrightarrow{W}_t), \boldsymbol{u} - \overrightarrow{W}_{t+1} \right\rangle \geq 0.$$

Let $\boldsymbol{u} = \overrightarrow{W}_\star$, we have

$$\eta \left\langle \nabla \widetilde{\ell}_t(\overrightarrow{W}_{t+1}), \overrightarrow{W}_{t+1} - \overrightarrow{W}_\star \right\rangle \leq \frac{1}{2} \left\| \overrightarrow{W}_\star - \overrightarrow{W}_t \right\|^2_{Z_t} - \frac{1}{2} \left\| \overrightarrow{W}_\star - \overrightarrow{W}_{t+1} \right\|^2_{Z_t} - \frac{1}{2} \left\| \overrightarrow{W}_t - \overrightarrow{W}_{t+1} \right\|^2_{Z_t}.$$

And by equation 31

$$\eta(\ell_t(W_{t+1}) - \ell_t(W_\star)) \leq \eta \left\langle \nabla \ell_t(\overrightarrow{W}_{t+1}), \overrightarrow{W}_{t+1} - \overrightarrow{W}_\star \right\rangle - \frac{1}{2} \|W_{t+1} - W_\star\|^2_{\nabla^2 \ell_t(W_{t+1})}.$$

Add the two inequalities together and rearrange to get

$$\begin{aligned} &\frac{1}{2} \left\| \overrightarrow{W}_\star - \overrightarrow{W}_{t+1} \right\|^2_{Z_t + \nabla^2 \ell_t(W_{t+1})} - \frac{1}{2} \left\| \overrightarrow{W}_\star - \overrightarrow{W}_t \right\|^2_{Z_t} + \frac{1}{2} \left\| \overrightarrow{W}_t - \overrightarrow{W}_{t+1} \right\|^2_{Z_t} \\ &\leq \eta(\ell_t(W_\star) - \ell_t(W_{t+1})) + \eta \left\langle \nabla \ell_t(\overrightarrow{W}_{t+1}) - \nabla \widetilde{\ell}_t(\overrightarrow{W}_{t+1}), \overrightarrow{W}_{t+1} - \overrightarrow{W}_\star \right\rangle \end{aligned} \tag{32}$$

In left hand side

$$\begin{aligned} \frac{1}{2} \left\| \overrightarrow{W}_\star - \overrightarrow{W}_{t+1} \right\|^2_{Z_t + \nabla^2 \ell_t(W_{t+1})} &= \frac{1}{2} \left\| \overrightarrow{W}_\star - \overrightarrow{W}_{t+1} \right\|^2_{(\lambda + \rho_{1:t-1}) I + V_t \Lambda_t V_t^\top + \nabla^2 \ell_t(W_{t+1})} \\ &\geq \frac{1}{2} \left\| \overrightarrow{W}_\star - \overrightarrow{W}_{t+1} \right\|^2_{(\lambda + \rho_{1:t-1}) I + V_{t+1} \Lambda_{t+1} V_{t+1}^\top} \\ &= \frac{1}{2} \left\| \overrightarrow{W}_\star - \overrightarrow{W}_{t+1} \right\|^2_{Z_{t+1}} - 2(\rho_{1:t} - \rho_{1:t-1}) S^2 \end{aligned}$$

where the inequality is by FD Lemma 3. And in the right hand side, the second term $\langle \nabla \ell_t(\overrightarrow{W}_{t+1}) - \nabla \widetilde{\ell}_t(\overrightarrow{W}_{t+1}), \overrightarrow{W}_{t+1} - \overrightarrow{W}_\star \rangle$ is the first order Taylor remainder of the loss function $f(W) = \left\langle \nabla \ell_t(\overrightarrow{W}), \overrightarrow{W}_{t+1} - \overrightarrow{W}_\star \right\rangle$ between the point $\overrightarrow{W}_{t+1}$ and $\overrightarrow{W}_t$, and thus can be bounded by equation 30.

$$\begin{aligned} \left\langle \nabla \ell_t(\overrightarrow{W}_{t+1}) - \nabla \widetilde{\ell}_t(\overrightarrow{W}_{t+1}), \overrightarrow{W}_{t+1} - \overrightarrow{W}_\star \right\rangle &= \left\langle \frac{1}{2!} D^3 \ell_t(\xi) \left[\overrightarrow{W}_{t+1} - \overrightarrow{W}_t, \overrightarrow{W}_{t+1} - \overrightarrow{W}_t \right], \overrightarrow{W}_t - \overrightarrow{W}_\star \right\rangle \\ &= \frac{1}{2} D^3 \ell_t(\xi) \left[\overrightarrow{W}_{t+1} - \overrightarrow{W}_t, \overrightarrow{W}_{t+1} - \overrightarrow{W}_t, \overrightarrow{W}_t - \overrightarrow{W}_\star \right] \\ &= \frac{1}{2} D^3 \ell_t(\xi) \left[\overrightarrow{W}_t - \overrightarrow{W}_\star, \overrightarrow{W}_{t+1} - \overrightarrow{W}_t, \overrightarrow{W}_{t+1} - \overrightarrow{W}_t \right] \end{aligned}$$

$$\leq \frac{\sqrt{6}}{2}\left\|\overrightarrow{W}_t - \overrightarrow{W}_\star\right\|_2 \left\|\overrightarrow{W}_{t+1} - \overrightarrow{W}_t\right\|^2_{\nabla^2\ell_t(\xi)}$$
$$\leq \sqrt{6}LS\left\|\overrightarrow{W}_{t+1} - \overrightarrow{W}_t\right\|^2_{I\otimes \boldsymbol{x}_t\boldsymbol{x}_t^\top}$$

Plug the above inequality into equation 32 and sum over t to get

$$\left\|\overrightarrow{W}_{t+1} - W_\star\right\|^2_{Z_{t+1}} - \left\|\overrightarrow{W}_1 - W_\star\right\|^2_{Z_1} \leq 2\eta\sum_{s=1}^t \ell_s(W_\star) - \ell_s(W_{s+1}) + 2\sqrt{6}\eta LS\sum_{s=1}^t \left\|\overrightarrow{W}_{s+1} - \overrightarrow{W}_s\right\|^2_{I\otimes \boldsymbol{x}_s\boldsymbol{x}_s^\top}$$
$$- \sum_{s=1}^t \left\|\overrightarrow{W}_{s+1} - \overrightarrow{W}_s\right\|^2_{Z_s} + 4S^2\rho_{1:t}. \tag{33}$$

$\sum_{s=1}^t \ell_s(W_\star) - \ell_s(W_{s+1})$ can be bounded by Lemma 8 for any martingale sequence $\{(W_t, \boldsymbol{x}_t) \,|\, t \in \mathbb{N}\}$.

$$\sum_{s=1}^t \ell_s(W_\star) - \ell_s(W_{s+1}) \leq (DS + \ln(K+1) + 2\ln((K+1)t+2))\left(\frac{5}{4} + 4\ln\frac{\sqrt{1+2t}}{\delta}\right) + 2 + \ln(K+1)$$
$$+ \frac{\sqrt{6}}{24Kd}\lambda\left(Kd\ln\Delta_t^{\lambda/2} + m\ln(1+2\frac{D^2}{m\lambda}t)\right) + \frac{24Kd}{\lambda}\sum_{s=1}^t \left\|\overrightarrow{W}_{s+1} - \overrightarrow{W}_s\right\|^2_{Z_s}$$

We apply the Lemma 8 to equation 33 , and tune the regularized parameter $\lambda \geq \max\left\{2LD^2, 2\eta\max\left\{2\sqrt{6}D^2SL, 48Kd\right\}\right\}$ to get the last three terms cancel out

$$\frac{48\eta Kd}{\lambda}\sum_{s=1}^t \left\|\overrightarrow{W}_s - \overrightarrow{W}_{s+1}\right\|^2_{Z_s} + 2\sqrt{6}\eta LS\sum_{s=1}^t \left\|\overrightarrow{W}_{s+1} - \overrightarrow{W}_s\right\|^2_{I\otimes \boldsymbol{x}_s\boldsymbol{x}_s^\top} - \sum_{s=1}^t \left\|\overrightarrow{W}_{s+1} - \overrightarrow{W}_s\right\|^2_{Z_s}$$
$$\leq \frac{48\eta Kd}{\lambda}\sum_{s=1}^t \left\|\overrightarrow{W}_s - \overrightarrow{W}_{s+1}\right\|^2_{Z_s} + 2\sqrt{6}\eta LSD^2\sum_{s=1}^t \left\|\overrightarrow{W}_{s+1} - \overrightarrow{W}_s\right\|^2_2 - \sum_{s=1}^t \left\|\overrightarrow{W}_{s+1} - \overrightarrow{W}_s\right\|^2_{Z_s}$$
$$\leq 0.$$

Therefore, with $\eta = \mathcal{O}(\ln K)$, we have

$$\left\|\overrightarrow{W}_{t+1} - W_\star\right\|^2_{Z_{t+1}} \leq \beta_{t+1}(\delta),$$

where

$$\beta_{t+1}(\delta) = \lambda(\|W_1\|_F + S)^2$$
$$+ 2\eta\left(DS + \ln(K+1) + 2\ln((K+1)t+2)\right)\left(\frac{5}{4} + 4\ln\frac{\sqrt{1+2t}}{\delta}\right) + 2\eta(2 + \ln(K+1))$$
$$+ \frac{\sqrt{6}\eta}{6Kd}\lambda\left(Kd\ln\Delta_t^{\lambda/2} + m\ln(1+2\frac{D^2}{m\lambda}t)\right) + 4S^2\rho_{1:t}$$
$$+ \sum_{s=1}^t 2\sqrt{6}\eta LS\left\|(W_{s+1} - W_s)\boldsymbol{x}_s\right\|_2^2 + (\frac{48\eta Kd}{\lambda} - 1)\left\|\overrightarrow{W}_{s+1} - \overrightarrow{W}_s\right\|^2_{Z_s} \tag{34}$$

And

$$\beta_{t+1}(\delta) = \mathcal{O}\left(\ln K(\ln K + \ln t)\left(Kd\ln\Delta_t + m\ln t + \ln\frac{1}{\delta}\right) + \rho_{1:t}\right)$$
$$= \mathcal{O}\left((1 + \frac{\rho_{1:t}}{\lambda + \ln t\ln\Delta_t})\ln K(\ln K + \ln t)\left(Kd\ln\Delta_t + m\ln t + \ln\frac{1}{\delta}\right)\right)$$
$$= \mathcal{O}\left(\Delta_t\ln K(\ln K + \ln t)\left(Kd\ln\Delta_t + m\ln t + \ln\frac{1}{\delta}\right)\right).$$

□

Proof for Theorem 2

Proof. The actual expected reward $\boldsymbol{r}^\top \boldsymbol{\sigma}(W_\star \boldsymbol{x})$ and the estimated reward $\boldsymbol{r}^\top \boldsymbol{\sigma}(W_t \boldsymbol{x})$ has the following difference:

Lemma 7. *For any $\delta \in (0,1)$, with probability at least $1-\delta$, we have*

$$\forall t > 0.\ \left|\boldsymbol{r}^\top \boldsymbol{\sigma}(W_\star \boldsymbol{x}) - \boldsymbol{r}^\top \boldsymbol{\sigma}(W_t \boldsymbol{x})\right| \le b_t(\boldsymbol{x}, \delta) = \varepsilon_t^{\texttt{fst}}(\boldsymbol{x}, \delta) + \varepsilon_t^{\texttt{snd}}(\boldsymbol{x}, \delta),$$

where

$$\varepsilon_t^{\texttt{fst}}(\boldsymbol{x}, \delta) = \sqrt{\beta_t(\delta)}\, \|(\nabla \boldsymbol{\sigma}(W_t \boldsymbol{x}) \otimes \boldsymbol{x})\boldsymbol{r}\|_{Z_t^{-1}},$$

and

$$\varepsilon_t^{\texttt{snd}}(\boldsymbol{x}, \delta) = \frac{\|\boldsymbol{r}\|_1}{2} \beta_t(\delta) \left\| Z_t^{-\frac{1}{2}} (I \otimes \boldsymbol{x}) \right\|_{2,2}^2.$$

The proof of the above lemma is similar to Zhang & Sugiyama (2023, C.2.1), and the main difference is the replacement of H_t with Z_t. We delay its proof to Appendix B.3.

Corollary 1 (Proposition 1 of Zhang & Sugiyama (2023))**.** *For any $\delta \in (0,1)$, with probability at least $1-\delta$, we have*

$$\forall t > 0.\ \boldsymbol{r}^\top \boldsymbol{\sigma}(W_\star \boldsymbol{x}) \le R_t(\boldsymbol{x}, \delta) = \boldsymbol{r}^\top \boldsymbol{\sigma}(W_t \boldsymbol{x}) + b_t(\boldsymbol{x}, \delta).$$

The regret is defined as the expected reward of the best action minus the expected reward of the action taken by the algorithm:

$$\begin{aligned}
\mathrm{Regret}(T) &= \sum_{t=1}^{T} \left(\boldsymbol{r}^\top \boldsymbol{\sigma}(W_\star \boldsymbol{x}_\star) - \boldsymbol{r}^\top \boldsymbol{\sigma}(W_\star \boldsymbol{x}_t)\right) \\
&= \sum_{t=1}^{T} \underbrace{\boldsymbol{r}^\top \boldsymbol{\sigma}(W_\star \boldsymbol{x}_\star) - \boldsymbol{r}^\top \boldsymbol{\sigma}(W_t \boldsymbol{x}_\star)}_{\texttt{term a}} + \underbrace{\boldsymbol{r}^\top \boldsymbol{\sigma}(W_t \boldsymbol{x}_\star) - \boldsymbol{r}^\top \boldsymbol{\sigma}(W_t \boldsymbol{x}_t)}_{\texttt{term b}} \\
&\quad + \underbrace{\boldsymbol{r}^\top \boldsymbol{\sigma}(W_t \boldsymbol{x}_t) - \boldsymbol{r}^\top \boldsymbol{\sigma}(W_\star \boldsymbol{x}_t)}_{\texttt{term c}}
\end{aligned}$$

The `term a` and `term c` is bounded

$$\texttt{term a} \le b_t(\boldsymbol{x}_\star, \delta), \texttt{term c} \le b_t(\boldsymbol{x}_t, \delta),$$

by the lemma 7. The `term b` is bounded

$$R_t(\boldsymbol{x}_\star, \delta) \le R_t(\boldsymbol{x}_t, \delta) \implies \texttt{term b} \le b_t(\boldsymbol{x}_t, \delta) - b_t(\boldsymbol{x}_\star, \delta),$$

by the decision rule $\boldsymbol{x}_t := \arg\max_{\boldsymbol{x} \in \mathcal{A}_t} R_t(\boldsymbol{x}, \delta)$. Thus the regret is bounded by $2\sum_{t=1}^{T} b_t(\boldsymbol{x}_t, \delta)$.

$$\mathrm{Regret}(T) \le 2\sum_{t=1}^{T} b_t(\boldsymbol{x}_t, \delta) = 2\sum_{t=1}^{T} \varepsilon_t^{\texttt{fst}}(\boldsymbol{x}_t, \delta) + 2\sum_{t=1}^{T} \varepsilon_t^{\texttt{snd}}(\boldsymbol{x}_t, \delta). \tag{35}$$

And we substitute H_t to Z_t to get the lemma holds in the proposed algorithm. For the term $\sum_{t=1}^{T} \varepsilon_{t-1}^{\texttt{fst}}(\boldsymbol{x}_t, \delta)$,

$$\begin{aligned}
&\sum_{t=1}^{T} \varepsilon_t^{\texttt{fst}}(\boldsymbol{x}_t, \delta) = \sum_{t=1}^{T} \sqrt{\beta_t(\delta)}\, \|(\nabla \boldsymbol{\sigma}(W_t \boldsymbol{x}_t) \otimes \boldsymbol{x}_t)\boldsymbol{r}\|_{Z_t^{-1}} \\
&\le \sqrt{\beta_{\max}(\delta)} \sum_{t=1}^{T} \|(\nabla \boldsymbol{\sigma}(W_{t+1} \boldsymbol{x}_t) \otimes \boldsymbol{x}_t)\boldsymbol{r}\|_{Z_t^{-1}} + \|((\nabla \boldsymbol{\sigma}(W_t \boldsymbol{x}_t) - \nabla \boldsymbol{\sigma}(W_{t+1} \boldsymbol{x}_t)) \otimes \boldsymbol{x}_t)\boldsymbol{r}\|_{Z_t^{-1}} \\
&\le \sqrt{\beta_{\max}(\delta)} \sum_{t=1}^{T} \left\| \nabla \boldsymbol{\sigma}(W_{t+1} \boldsymbol{x}_t)^{\frac{1}{2}} \boldsymbol{r} \right\|_2 \left\| \nabla \boldsymbol{\sigma}(W_{t+1} \boldsymbol{x}_t)^{\frac{1}{2}} \otimes \boldsymbol{x}_t \right\|_{2, Z_t^{-1}} \\
&\quad + \max_{\xi_1} \left\| Z_t^{-\frac{1}{2}} (D^2 \boldsymbol{\sigma}(\xi_1)[\boldsymbol{r}] \otimes \boldsymbol{x}_t \boldsymbol{x}_t^\top) Z_t^{-\frac{1}{2}} \right\|_{2,2} \left\| \overrightarrow{W}_t - \overrightarrow{W}_{t+1} \right\|_{Z_t}
\end{aligned}$$

where the second inequality is by Mean Value Theorem of vector function in form of finite increment, and $\beta_{\max}(\delta)$ denotes $\max_{t\le T}\beta_t(\delta)$.

The following terms is bounded: $\left\|\nabla\boldsymbol{\sigma}(W_{t+1}\boldsymbol{x}_t)^{\frac{1}{2}}\boldsymbol{r}\right\|_2 \le R$, and $\left\|\overrightarrow{W}_t - \overrightarrow{W}_{t+1}\right\|_{Z_t} \le \frac{2\eta D}{\sqrt{\lambda}}$ by lemma 10. And moreover

$$\left\|\nabla\boldsymbol{\sigma}(W_{t+1}\boldsymbol{x}_t)^{\frac{1}{2}}\otimes\boldsymbol{x}_t\right\|_{2,Z_t^{-1}} \le \sqrt{\mathrm{tr}((\nabla\boldsymbol{\sigma}(W_{t+1}\boldsymbol{x}_t)\otimes\boldsymbol{x}_t\boldsymbol{x}_t^\top)Z_t^{-1})} = \sqrt{\mathrm{tr}(\nabla^2\ell_t(W_{t+1})Z_t^{-1})}$$

$$\max_{\xi_1}\left\|Z_t^{-\frac{1}{2}}(D^2\boldsymbol{\sigma}(\xi_1)[\boldsymbol{r}]\otimes\boldsymbol{x}_t\boldsymbol{x}_t^\top)Z_t^{-\frac{1}{2}}\right\|_{2,2} \le \|\boldsymbol{r}\|_1\lambda_{\max}((I\otimes\boldsymbol{x}_t^\top)Z_t^{-1}(I\otimes\boldsymbol{x}_t)) \le \|\boldsymbol{r}\|_1\,\mathrm{tr}\left(Z_t^{-1}(I\otimes\boldsymbol{x}_t\boldsymbol{x}_t^\top)\right)$$

in which the second inequality is by $-\|\boldsymbol{r}\|_1 I \preceq D^2\boldsymbol{\sigma}(\xi_1)[\boldsymbol{r}] \preceq \frac{1}{2}\|\boldsymbol{r}\|_1 I$. Thus by GM-AM inequality, we have

$$\sum_{t=1}^T \varepsilon_t^{\texttt{fst}}(\boldsymbol{x}_t,\delta) \le R\sqrt{\beta_{\max}(\delta)}\sqrt{T}\sqrt{\sum_{t=1}^T \mathrm{tr}\left(Z_t^{-1}\nabla^2\ell_t(W_{t+1})\right)} + \frac{4\eta D}{\sqrt{\lambda}}\|\boldsymbol{r}\|_1\sqrt{\beta_{T-1}(\delta)}\sum_{t=1}^T \mathrm{tr}\left(Z_t^{-1}(I\otimes\boldsymbol{x}_t\boldsymbol{x}_t^\top)\right) \tag{36}$$

And for the term $\varepsilon_t^{\texttt{snd}}(\boldsymbol{x}_t,\delta)$

$$\begin{aligned}
\sum_{t=1}^T \varepsilon_t^{\texttt{snd}}(\boldsymbol{x}_t,\delta) &= \sum_{t=1}^T \frac{\|\boldsymbol{r}\|_1}{2}\beta_t(\delta)\left\|Z_t^{-\frac{1}{2}}(I\otimes\boldsymbol{x})\right\|_{2,2}^2 \le \frac{\|\boldsymbol{r}\|_1}{2}\beta_{\max}(\delta)\sum_{t=1}^T\left\|Z_t^{-\frac{1}{2}}(I\otimes\boldsymbol{x})\right\|_{2,2}^2 \\
&\le \frac{\|\boldsymbol{r}\|_1}{2}\beta_{\max}(\delta)\sum_{t=1}^T \lambda_{\max}((I\otimes\boldsymbol{x})^\top Z_t^{-1}(I\otimes\boldsymbol{x})) \\
&\le \frac{\|\boldsymbol{r}\|_1}{2}\beta_{\max}(\delta)\sum_{t=1}^T \mathrm{tr}((I\otimes\boldsymbol{x})^\top Z_t^{-1}(I\otimes\boldsymbol{x})) \\
&= \frac{\|\boldsymbol{r}\|_1}{2}\beta_{\max}(\delta)\sum_{t=1}^T \mathrm{tr}(Z_t^{-1}(I\otimes\boldsymbol{x}\boldsymbol{x}^\top))
\end{aligned} \tag{37}$$

Plugging equation 36 and equation 37 into the regret bound equation 35,

$$\begin{aligned}
\mathrm{Regret}(T) \le{}& 2R\sqrt{\beta_{\max}(\delta)}\sqrt{T}\sqrt{\sum_{t=1}^T \mathrm{tr}\left(Z_t^{-1}\nabla^2\ell_t(W_{t+1})\right)} \\
&+ R\sqrt{K}\left(\frac{8\eta D}{\sqrt{\lambda}}\sqrt{\beta_{\max}(\delta)} + \beta_{\max}(\delta)\right)\sum_{t=1}^T \mathrm{tr}\left(Z_t^{-1}(I\otimes\boldsymbol{x}_t\boldsymbol{x}_t^\top)\right)
\end{aligned}$$

By Lemma 12 and 14,

$$\begin{aligned}
\mathrm{Regret}(T) \le{}& 2R\sqrt{\beta_{\max}(\delta)}\sqrt{T}\sqrt{2\left(Kd\ln\Delta_T^{\lambda/2} + m\ln(1+2\frac{D^2}{m\lambda}T)\right)} \\
&+ R\sqrt{K}\left(\frac{8\eta D}{\sqrt{\lambda}}\sqrt{\beta_{\max}(\delta)} + \beta_{\max}(\delta)\right)\kappa 2\left(Kd\ln\Delta_T^{\lambda/2} + m\ln(1+2\frac{D^2}{m\lambda}T)\right)
\end{aligned}$$

Here

$$\beta_{\max}(\delta) \le \mathcal{O}\left(\Delta_T \ln K(\ln K + \ln T)(Kd\ln\Delta_t + m\ln T)\right).$$

Therefore the regret bound becomes

$$\begin{aligned}
&\mathrm{Regret}(T) \\
&= \mathcal{O}\left(\sqrt{\Delta_T}(Kd\ln\Delta_T + m\ln T)\ln K\sqrt{\ln T}\sqrt{T} + \kappa\sqrt{K}\ln^2 K\ln T\Delta_T(Kd\ln\Delta_T^\lambda + m\ln T)^2\right).
\end{aligned}$$

We follow the same notations as Abeille et al. (2021); Zhang & Sugiyama (2023); Amani & Thrampoulidis (2021), ignoring the $\ln(\cdot)$ factors in the $\widetilde{\mathcal{O}}(\cdot)$ notation:

$$\mathrm{Regret}(T) = \widetilde{\mathcal{O}}\left(\Delta_T(Kd\ln\Delta_T + m)\sqrt{T}\right).$$

□

Proof for Proposition 1

Proof. In the previous discussion, The decision rule is

$$(\widehat{\boldsymbol{w}}_t, \boldsymbol{x}_t) = \underset{\boldsymbol{w}\in\mathcal{C}_t, \boldsymbol{x}\in\mathcal{A}_t}{\arg\max}\ \boldsymbol{w}^\top \boldsymbol{x}, \tag{38}$$

and the regret is

$$\mathrm{Regret}(T) = \sum_{t=1}^{T} \sigma(\boldsymbol{w}_\star^\top \boldsymbol{x}_\star) - \sigma(\boldsymbol{w}_\star^\top \boldsymbol{x}_t),$$

since $\boldsymbol{r} = [0, 1]^\top$ and we ignore the 0-th component of $\boldsymbol{\sigma}(\cdot)$. And we perform the similar decomposition as in Zhang & Sugiyama (2023, Corollary 1):

$$\mathrm{Regret}(T) = \sum_{t=1}^{T} \sigma(\boldsymbol{w}_\star^\top \boldsymbol{x}_\star) - \sigma(\boldsymbol{w}_\star^\top \boldsymbol{x}_t)$$

(by optimistic decision rule equation 38)

$$\leq \sum_{t=1}^{T} \sigma(\widehat{\boldsymbol{w}}_t^\top \boldsymbol{x}_t) - \sigma(\boldsymbol{w}_\star^\top \boldsymbol{x}_t)$$

(by Taylor's theorem)

$$= \sum_{t=1}^{T} \sigma'(\boldsymbol{w}_\star^\top \boldsymbol{x}_t)(\widehat{\boldsymbol{w}}_t - \boldsymbol{w}_\star)^\top \boldsymbol{x}_t + \frac{1}{2}\sigma''(\xi_t)((\widehat{\boldsymbol{w}}_t - \boldsymbol{w}_\star)^\top \boldsymbol{x}_t)^2$$

$$= \sum_{t=1}^{T} \sqrt{\sigma'(\boldsymbol{w}_\star^\top \boldsymbol{x}_t)}(\widehat{\boldsymbol{w}}_t - \boldsymbol{w}_\star)^\top \sqrt{\sigma'(\boldsymbol{w}_\star^\top \boldsymbol{x}_t)}\boldsymbol{x}_t + \frac{1}{2}\sigma''(\xi_t)((\widehat{\boldsymbol{w}}_t - \boldsymbol{w}_\star)^\top \boldsymbol{x}_t)^2$$

(by Taylor's theorem)

$$= \sum_{t=1}^{T} \sqrt{\sigma'(\boldsymbol{w}_\star^\top \boldsymbol{x}_t)}(\widehat{\boldsymbol{w}}_t - \boldsymbol{w}_\star)^\top \sqrt{\sigma'(\boldsymbol{w}_{t+1}^\top \boldsymbol{x}_t)}\boldsymbol{x}_t + \sqrt{\sigma'(\boldsymbol{w}_\star^\top \boldsymbol{x}_t)}(\widehat{\boldsymbol{w}}_t - \boldsymbol{w}_\star)^\top \boldsymbol{x}_t \frac{\sigma''(\psi_t)}{2\sqrt{\sigma'(\psi_t)}}(\boldsymbol{w}_\star - \boldsymbol{w}_{t+1})^\top \boldsymbol{x}_t$$
$$+ \frac{1}{2}\sigma''(\xi_t)((\widehat{\boldsymbol{w}}_t - \boldsymbol{w}_\star)^\top \boldsymbol{x}_t)^2$$

(by $\sigma'(\cdot) \leq \frac{1}{4}$, $\sigma''(\cdot) \leq \frac{1}{4}$, and $\sigma''(\cdot)/(2\sqrt{\sigma'(\cdot)}) \leq \frac{1}{4}$)

$$\leq \sum_{t=1}^{T} \sqrt{\sigma'(\boldsymbol{w}_\star^\top \boldsymbol{x}_t)}(\widehat{\boldsymbol{w}}_t - \boldsymbol{w}_\star)^\top \sqrt{\sigma'(\boldsymbol{w}_{t+1}^\top \boldsymbol{x}_t)}\boldsymbol{x}_t + \frac{1}{8}(\widehat{\boldsymbol{w}}_t - \boldsymbol{w}_\star)^\top \boldsymbol{x}_t(\boldsymbol{w}_\star - \boldsymbol{w}_{t+1})^\top \boldsymbol{x}_t + \frac{1}{8}((\widehat{\boldsymbol{w}}_t - \boldsymbol{w}_\star)^\top \boldsymbol{x}_t)^2$$

(by Cauchy-Schwarz inequality)

$$\leq \sum_{t=1}^{T} \|\widehat{\boldsymbol{w}}_t - \boldsymbol{w}_\star\|_{Z_t} \sqrt{\sigma'(\boldsymbol{w}_\star^\top \boldsymbol{x}_t)} \left\| \sqrt{\sigma'(\boldsymbol{w}_{t+1}^\top \boldsymbol{x}_t)}\boldsymbol{x}_t \right\|_{Z_t^{-1}}$$
$$+ \frac{1}{8} \|\widehat{\boldsymbol{w}}_t - \boldsymbol{w}_\star\|_{Z_t} \|\boldsymbol{w}_\star - \boldsymbol{w}_{t+1}\|_{Z_{t+1}} \|\boldsymbol{x}_t\|_{Z_t^{-1}} \|\boldsymbol{x}_t\|_{Z_{t+1}^{-1}} + \frac{1}{8} \|\widehat{\boldsymbol{w}}_t - \boldsymbol{w}_\star\|_{Z_t}^2 \|\boldsymbol{x}_t\|_{Z_t^{-1}}^2$$

where the second equality is by $\frac{\sigma''(\cdot)}{2\sqrt{\sigma'(\cdot)}} = \frac{1}{2}(1 - 2\sigma(\cdot))\sqrt{\sigma(\cdot)(1 - \sigma(\cdot))} \leq \frac{1}{4}$ and $\sigma'(\cdot) \leq \frac{1}{4}$.

The term $\|\boldsymbol{w}_t - \boldsymbol{w}_\star\|_{Z_t}$ is bounded by Theorem 1, as well as $\|\widehat{\boldsymbol{w}}_t - \boldsymbol{w}_\star\|_{Z_t}$

$$\|\boldsymbol{w}_t - \boldsymbol{w}_\star\|_{Z_t} \leq \sqrt{\beta_t(\delta)} \leq \sqrt{\beta_{\max}(\delta)}$$
$$\|\widehat{\boldsymbol{w}}_t - \boldsymbol{w}_\star\|_{Z_t} \leq \|\widehat{\boldsymbol{w}}_t - \boldsymbol{w}_t\|_{Z_t} + \|\boldsymbol{w}_t - \boldsymbol{w}_\star\|_{Z_t} \leq 2\sqrt{\beta_t(\delta)} \leq 2\sqrt{\beta_{\max}(\delta)}$$

Here $\beta_{\max}(\delta)$ denotes $\max_{t\leq T} \beta_t(\delta)$. Therefore we perform Cauchy-Schwarz inequality on the first and second terms

$$\sum_{t=1}^{T} \|\widehat{\boldsymbol{w}}_t - \boldsymbol{w}_\star\|_{Z_t} \sqrt{\sigma'(\boldsymbol{w}_\star^\top \boldsymbol{x}_t)} \left\| \sqrt{\sigma'(\boldsymbol{w}_{t+1}^\top \boldsymbol{x}_t)}\boldsymbol{x}_t \right\|_{Z_t^{-1}}$$

$$
\begin{aligned}
&\leq \sqrt{\beta_{\max}(\delta)} \sum_{t=1}^{T} \sqrt{\sigma'(\boldsymbol{w}_\star^\top \boldsymbol{x}_t)} \left\| \sqrt{\sigma'(\boldsymbol{w}_{t+1}^\top \boldsymbol{x}_t)} \boldsymbol{x}_t \right\|_{Z_t^{-1}} \\
&\leq \sqrt{\beta_{\max}(\delta)} \sqrt{\sum_{t=1}^{T} \sigma'(\boldsymbol{w}_\star^\top \boldsymbol{x}_t)} \sqrt{\sum_{t=1}^{T} \left\| \sqrt{\sigma'(\boldsymbol{w}_{t+1}^\top \boldsymbol{x}_t)} \boldsymbol{x}_t \right\|_{Z_t^{-1}}^2} \\
&\quad \sum_{t=1}^{T} \frac{1}{8} \left\| \widehat{\boldsymbol{w}}_t - \boldsymbol{w}_\star \right\|_{Z_t} \left\| \boldsymbol{w}_\star - \boldsymbol{w}_{t+1} \right\|_{Z_{t+1}} \left\| \boldsymbol{x}_t \right\|_{Z_t^{-1}} \left\| \boldsymbol{x}_t \right\|_{Z_{t+1}^{-1}} \\
&\leq \frac{1}{4} \beta_{\max}(\delta) \sqrt{\sum_{t=1}^{T} \left\| \boldsymbol{x}_t \right\|_{Z_t^{-1}}^2} \sqrt{\sum_{t=1}^{T} \left\| \boldsymbol{x}_t \right\|_{Z_{t+1}^{-1}}^2}
\end{aligned}
$$

And for the third term, we have

$$
\sum_{t=1}^{T} \frac{1}{8} \left\| \widehat{\boldsymbol{w}}_t - \boldsymbol{w}_\star \right\|_{Z_t}^2 \left\| \boldsymbol{x}_t \right\|_{Z_t^{-1}}^2 \leq \beta_T(\delta) \sum_{t=1}^{T} \left\| \boldsymbol{x}_t \right\|_{Z_t^{-1}}^2
$$

By Lemma 12, 15, and 14, following terms have upper bounds

$$
\begin{aligned}
\sum_{t=1}^{T} \left\| \sqrt{\sigma'(\boldsymbol{w}_{t+1}^\top \boldsymbol{x}_t)} \boldsymbol{x}_t \right\|_{Z_t^{-1}}^2 &= \mathcal{O}\left((d \ln \Delta_T + m \ln T) \right) \\
\sum_{t=1}^{T} \left\| \boldsymbol{x}_t \right\|_{Z_{t+1}^{-1}}^2 &= \mathcal{O}\left(\kappa \left(d \ln \Delta_T + m \ln T \right) \right) \\
\sum_{t=1}^{T} \left\| \boldsymbol{x}_t \right\|_{Z_t^{-1}}^2 &= \mathcal{O}\left(\kappa \left(d \ln \Delta_T + m \ln T \right) \right)
\end{aligned}
$$

And Faury et al. (2022, Theorem 2)) indicates that

$$
\sum_{t=1}^{T} \sigma'(\boldsymbol{w}_\star^\top \boldsymbol{x}_t) \leq \mathrm{Regret}(T) + T\sigma'(\boldsymbol{w}_\star^\top \boldsymbol{x}_\star)
$$

Plug the above terms into the regret bound, we have

$$
\begin{aligned}
\mathrm{Regret}(T) = \mathcal{O}\left(\sqrt{\beta_{\max}(\delta) \left(d \ln \Delta_T^\lambda + m \ln T \right)} \right) \sqrt{\mathrm{Regret}(T) + T\sigma'(\boldsymbol{w}_\star^\top \boldsymbol{x}_\star)} \\
+ \mathcal{O}\left(\kappa \beta_{\max}(\delta) \left(d \ln \Delta_T^\lambda + m \ln T \right) \right)
\end{aligned}
$$

By resolving the inequality w.r.t. $\mathrm{Regret}(T)$ above, we have

$$
\begin{aligned}
\mathrm{Regret}(T) &\leq \mathcal{O}\left(\sqrt{\beta_{\max}(\delta) \left(d \ln \Delta_T^\lambda + m \ln T \right)} \right) \sqrt{T\sigma'(\boldsymbol{w}_\star^\top \boldsymbol{x}_\star)} + \mathcal{O}\left(\kappa \beta_{\max}(\delta) \left(d \ln \Delta_T^\lambda + m \ln T \right) \right) \\
&= \mathcal{O}\left(\sqrt{\Delta_T} (d \ln \Delta_T^\lambda + m \ln T) \sqrt{\ln T} \sqrt{T/\kappa_*} + \kappa \Delta_T (d \ln \Delta_T^\lambda + m \ln T)^2 \ln T \right) \\
&= \widetilde{\mathcal{O}}\left(\Delta_T (d \ln \Delta_T^\lambda + m) \sqrt{T/\kappa_*} \right),
\end{aligned}
$$

where $\frac{1}{\kappa_*} = \frac{1}{T} \sum_{t=1}^{T} \sigma'(\boldsymbol{w}_\star^\top \boldsymbol{x}_t^*)$ □

B.3 Useful Lemmas

Proof for Lemma 1

Lemma. *For any $\boldsymbol{x} \in \mathcal{A}_t$, let $Z_t = \lambda' I_{Kd} + V_t \Lambda_t V_t^\top$, $\lambda' = \lambda + \rho_{1:t-1}$. The spectral norm satisfies*

$$
\left\| Z_t^{-\frac{1}{2}} (I_K \otimes \boldsymbol{x}) \right\|_2^2 = \left\| A(\boldsymbol{x}) \right\|_2, \tag{39}
$$

where $A(\boldsymbol{x}) \in \mathbb{R}^{K\times K}$ is a symmetric matrix with entries given by

$$A(\boldsymbol{x})_{ij} = \sum_{k=1}^{m}\left(\frac{1}{\mu_k+\lambda'} - \frac{1}{\lambda'}\right)\langle \boldsymbol{x}, \boldsymbol{u}_{ki}\rangle \langle \boldsymbol{x}, \boldsymbol{u}_{kj}\rangle + \frac{\|\boldsymbol{x}\|_2^2}{\lambda'}\delta_{ij}. \tag{40}$$

Here, μ_k is the k-th diagonal element of Λ_t. $\boldsymbol{u}_{ki} \in \mathbb{R}^d$ denotes the sub-vector extracted from indices $(i-1)d+1$ to id of V_t's k-th column, and δ_{ij} denotes the Kronecker delta.

Proof. We have

$$\left\|Z_t^{-\frac{1}{2}}(I_K\otimes \boldsymbol{x})\right\|_2^2 = \lambda_{\max}((I_K\otimes \boldsymbol{x})^\top Z_t^{-\frac{1}{2}} Z_t^{-\frac{1}{2}}(I_K\otimes \boldsymbol{x})) = \lambda_{\max}\left((I_K\otimes \boldsymbol{x})^\top Z_t^{-1}(I_K\otimes \boldsymbol{x})\right).$$

The first equality is by $\|B\|_2^2 = \lambda_{\max}(B^\top B)$. Let A denote the matrix $(I_K\otimes \boldsymbol{x})^\top Z_t^{-1}(I_K\otimes \boldsymbol{x})$. And let $\widehat{\mu}_i$ and $\widehat{\boldsymbol{u}}_i$ be the complementary eigenvalues μ_i and eigenvectors $\boldsymbol{u}_i$, such that $\widehat{\mu}_i = 0$ and $\widehat{\boldsymbol{u}}_i$ is orthogonal to the column space of V_t for $i = m+1,\dots,Kd$, and $\widehat{\mu}_i = \mu_i$ and $\widehat{\boldsymbol{u}}_i = \boldsymbol{u}_i$ for $i = 1,\dots,m$. Then we have the eigen-decomposition of Z_t as

$$Z_t^{-1} = \sum_{i=1}^{Kd}\frac{1}{\widehat{\mu}_i+\lambda'}\widehat{\boldsymbol{u}}_i\widehat{\boldsymbol{u}}_i^\top.$$

And the element $(A_{ij})_{i,j=1}^K$ can be computed as follows:

$$\begin{aligned}
A_{ij} &= (\boldsymbol{e}_i\otimes \boldsymbol{x})^\top Z_t^{-1}(\boldsymbol{e}_j\otimes \boldsymbol{x}) \\
&= (\boldsymbol{e}_i\otimes \boldsymbol{x})^\top \sum_{k=1}^{Kd}\frac{1}{\widehat{\mu}_k+\lambda'}\widehat{\boldsymbol{u}}_k\widehat{\boldsymbol{u}}_k^\top(\boldsymbol{e}_j\otimes \boldsymbol{x}) \\
&= (\boldsymbol{e}_i\otimes \boldsymbol{x})^\top \sum_{k=1}^{m}\frac{1}{\mu_k+\lambda'}\boldsymbol{u}_k\boldsymbol{u}_k^\top(\boldsymbol{e}_j\otimes \boldsymbol{x}) + \frac{1}{\lambda'}(\boldsymbol{e}_i\otimes \boldsymbol{x})^\top \sum_{k=m+1}^{Kd}\widehat{\boldsymbol{u}}_k\widehat{\boldsymbol{u}}_k^\top(\boldsymbol{e}_j\otimes \boldsymbol{x}) \\
&= \sum_{k=1}^{m}\frac{1}{\mu_k+\lambda'}(\boldsymbol{e}_i\otimes \boldsymbol{x})^\top\boldsymbol{u}_k\boldsymbol{u}_k^\top(\boldsymbol{e}_j\otimes \boldsymbol{x}) + \frac{1}{\lambda'}(\boldsymbol{e}_i\otimes \boldsymbol{x})^\top\left(I_{Kd} - \sum_{k=1}^{m}\boldsymbol{u}_k\boldsymbol{u}_k^\top\right)(\boldsymbol{e}_j\otimes \boldsymbol{x}) \\
&= \frac{1}{\lambda'}(\boldsymbol{e}_i\otimes \boldsymbol{x})^\top(\boldsymbol{e}_j\otimes \boldsymbol{x}) + \sum_{k=1}^{m}\left(\frac{1}{\mu_k+\lambda'} - \frac{1}{\lambda'}\right)(\boldsymbol{e}_i\otimes \boldsymbol{x})^\top\boldsymbol{u}_k\boldsymbol{u}_k^\top(\boldsymbol{e}_j\otimes \boldsymbol{x}) \\
&= \frac{\boldsymbol{x}^\top\boldsymbol{x}}{\lambda'}\delta_i^j + \sum_{k=1}^{m}\left(\frac{1}{\mu_k+\lambda'} - \frac{1}{\lambda'}\right)\boldsymbol{x}^\top\boldsymbol{u}_{ki}\boldsymbol{x}^\top\boldsymbol{u}_{kj}
\end{aligned}$$

where δ_i^j is the Kronecker delta function. The fourth equality is by the fact that $\sum_{k=m+1}^{Kd}\widehat{\boldsymbol{u}}_k\widehat{\boldsymbol{u}}_k^\top = I_{Kd} - \sum_{k=1}^{m}\widehat{\boldsymbol{u}}_k\widehat{\boldsymbol{u}}_k^\top = I_{Kd} - \sum_{k=1}^{m}\boldsymbol{u}_k\boldsymbol{u}_k^\top$.

Then $\{\boldsymbol{x}^\top\boldsymbol{u}_{kj}\}_{k,j}$ can be computed in $\mathcal{O}(Kdm)$ time, A can be constructed in $\mathcal{O}(K^2m)$, and $\lambda_{\max}(A) = \|A\|_2$ can be computed in $\mathcal{O}(K^3)$. Therefore, the operator norm $\left\|Z_t^{-\frac{1}{2}}(I_K\otimes \boldsymbol{x})\right\|_{2,2}$ can be computed in $\mathcal{O}(K^3 + K^2m + Kdm)$ time. □

Proof for Lemma 2

Lemma. *Let $\widetilde{Z}_t = \lambda' I_{Kd} + \widetilde{V}_t\widetilde{\Lambda}_t\widetilde{V}_t^\top$, $\lambda' = \lambda + \rho_{1:t-1}$, where $\widetilde{\boldsymbol{v}}_i$ is the i-th column of $\widetilde{V}_t$ and μ_i is the i-th diagonal element of $\widetilde{\Lambda}_t$. If $\|\overrightarrow{W}'\|_2 \le S$, then $\overrightarrow{W}_{t+1} = \overrightarrow{W}'$. Otherwise, the projection is given by:*

$$\overrightarrow{W}_{t+1} = \left(\frac{\lambda'}{\lambda'+\nu}I + \frac{\nu}{\lambda'+\nu}\widetilde{V}_t\widetilde{\Lambda}_t M\widetilde{V}_t^\top\right)\overrightarrow{W}',$$

where $M = ((\lambda'+\nu)I_{m+K} + \widetilde{\Lambda}_t)^{-1}$ is diagonal. The dual variable $\nu > 0$ is the unique root of

$$f(\nu) = S^2,$$

where $f(\nu)$ is a monotonically decreasing function on the interval $(0, +\infty)$

$$f(\nu) = \frac{(\|\overrightarrow{W}'\|_2^2 - \|\widetilde{V}_t^\top \overrightarrow{W}'\|_2^2)\lambda'^2}{(\lambda' + \nu)^2} + \sum_{i=1}^{m+K} \frac{(\widetilde{\boldsymbol{v}}_i^\top \overrightarrow{W}')^2(\mu_i + \lambda')^2}{(\mu_i + \lambda' + \nu)^2}.$$

Proof. In the following proof, we denote $V = \widetilde{V}_t, \Lambda = \widetilde{\Lambda}_t$ for readability. And with loss of generality, we assume Λ to be a full-rank diagonal matrix, otherwise we can remove the zero eigenvalues and the corresponding eigenvectors without affecting the solution. And we assume the rank of Λ is m'. We rewrite the projection problem as a convex optimization problem

$$\begin{aligned} \min_{\overrightarrow{W}} \quad & \frac{1}{2} \left\| \overrightarrow{W} - \overrightarrow{W}' \right\|_{\widetilde{Z}_t}^2 \\ \text{s.t.} \quad & \frac{1}{2} \left\| \overrightarrow{W} \right\|_2^2 \leq \frac{S^2}{2} \end{aligned}$$

If $\|\overrightarrow{W}'\|_2 \leq S$, then $\overrightarrow{W}'$ is the optimal solution with optimal value 0.

Assume $\|\overrightarrow{W}'\|_2 > S$. Since there exists an interior point $\overrightarrow{W} = 0$ that strictly satisfies the constraint $\frac{1}{2}\|0\|_2^2 < S^2/2$, the Slater's condition holds and the strong duality holds. Therefore the optimal solution is equivalent to the Karush-Kuhn-Tucker (KKT) conditions.

$$\widetilde{Z}_t(\overrightarrow{X}_\star - \overrightarrow{W}') + \nu \overrightarrow{X}_\star = 0 \tag{41}$$

$$\left\| \overrightarrow{X}_\star \right\|_2 \leq S \tag{42}$$

$$\nu \geq 0 \tag{43}$$

$$\nu \left(\left\| \overrightarrow{X}_\star \right\|_2 - S \right) = 0 \tag{44}$$

Here $\overrightarrow{X}_\star$ denotes the solution of the target problem. Since $\widetilde{Z}_t$ is positive definite and $\nu \geq 0$, the matrix $\widetilde{Z}_t + \nu I$ is invertible. By equation equation 41, we have

$$\begin{aligned} \overrightarrow{X}_\star &= (\widetilde{Z}_t + \nu I)^{-1} \widetilde{Z}_t \overrightarrow{W}' \\ &= ((\lambda' + \nu)I + V\Lambda V^\top)^{-1}(\lambda' I + V\Lambda V^\top)\overrightarrow{W}' \end{aligned}$$

By Woodbury matrix identity, we have

$$\begin{aligned} ((\lambda' + \nu)I + V\Lambda V^\top)^{-1} &= \frac{1}{\lambda' + \nu} I - \frac{1}{(\lambda' + \nu)^2} V(\Lambda^{-1} + \frac{1}{\lambda' + \nu} V^\top V)^{-1} V^\top \\ &= \frac{1}{\lambda' + \nu} I - \frac{1}{\lambda' + \nu} V(\Lambda + (\lambda' + \nu)I)^{-1} \Lambda V^\top \end{aligned}$$

Therefore,

$$\begin{aligned} \overrightarrow{X}_\star &= \left(\frac{1}{\lambda' + \nu} I - \frac{1}{\lambda' + \nu} V(\Lambda + (\lambda' + \nu)I)^{-1} \Lambda V^\top \right) (\lambda' I + V\Lambda V^\top)\overrightarrow{W}' \\ &= \left(\frac{\lambda'}{\lambda' + \nu} I - \frac{\lambda'}{\lambda' + \nu} V(\Lambda + (\lambda' + \nu)I)^{-1} \Lambda V^\top + \frac{1}{\lambda' + \nu} V\Lambda V^\top - \frac{1}{\lambda' + \nu} V((\lambda' + \nu)I + \Lambda)^{-1} \Lambda^2 V^\top \right) \overrightarrow{W}' \\ &= \left(\frac{\lambda'}{\lambda' + \nu} I + \frac{\nu}{\lambda' + \nu} V((\lambda' + \nu)I + \Lambda)^{-1} \Lambda V^\top \right) \overrightarrow{W}' \end{aligned}$$

Let $M = ((\lambda' + \nu)I + \Lambda)^{-1}$, therefore

$$\overrightarrow{X}_\star = \left(\frac{\lambda'}{\lambda' + \nu} I + \frac{\nu}{\lambda' + \nu} V M \Lambda V^\top \right) \overrightarrow{W}'.$$

If $\nu = 0$, then $\overrightarrow{X}_\star = \overrightarrow{W}'$. However by equation 42, $\|\overrightarrow{X}_\star\|_2 \leq S$, which contradicts the assumption $\|\overrightarrow{W}'\|_2 > S$. Therefore by equation 43 $\nu > 0$. And by equation 44, we have $\|\overrightarrow{X}_\star\|_2^2 = S^2$. Let $\widehat{V}, \widehat{\Lambda}$

be the full matrix of V, Λ by adding the orthogonal basis of the null space of V and 0 eigenvalues. Then we have

$$\begin{aligned}\left\|\overrightarrow{X}_\star\right\|_2^2 &= \left\|(\widetilde{Z}_t + \nu I)^{-1}\widetilde{Z}_t\overrightarrow{W}'\right\|_2^2 \\ &= \left\|\widehat{V}((\lambda'+\nu)I + \widehat{\Lambda})^{-1}\widehat{V}^\top\widehat{V}(\lambda' I + \widehat{\Lambda})\widehat{V}^\top\overrightarrow{W}'\right\|_2^2 \\ &= \left\|\widehat{V}((\lambda'+\nu)I + \widehat{\Lambda})^{-1}(\lambda' I + \widehat{\Lambda})\widehat{V}^\top\overrightarrow{W}'\right\|_2^2 \\ &= \left\|((\lambda'+\nu)I + \widehat{\Lambda})^{-1}(\lambda' I + \widehat{\Lambda})\widehat{V}^\top\overrightarrow{W}'\right\|_2^2 \\ &= \sum_{i=1}^{Kd}\frac{(\widehat{\boldsymbol{v}}_i^\top\overrightarrow{W}')^2(\lambda'+\widehat{\mu}_i)^2}{(\lambda'+\widehat{\mu}_i+\nu)^2}\end{aligned}$$

Here $\widehat{\boldsymbol{v}}_i$ is the i-th column of $\widehat{V}$ and $\widehat{\mu}_i$ is the i-th diagonal element of $\widehat{\Lambda}$. Since $\widehat{\mu}_i = 0$ for $i > m'$, we have

$$\begin{aligned}\left\|\overrightarrow{X}_\star\right\|_2^2 &= \sum_{i=1}^{Kd}\frac{(\widehat{\boldsymbol{v}}_i^\top\overrightarrow{W}')^2(\lambda'+\widehat{\mu}_i)^2}{(\lambda'+\widehat{\mu}_i+\nu)^2} \\ &= \sum_{i=1}^{m'}\frac{(\boldsymbol{v}_i^\top\overrightarrow{W}')^2(\lambda'+\mu_i)^2}{(\lambda'+\mu_i+\nu)^2} + \sum_{i=m'+1}^{Kd}\frac{(\widehat{\boldsymbol{v}}_i^\top\overrightarrow{W}')^2\lambda'^2}{(\lambda'+\nu)^2}\end{aligned}$$

Here by assumption $m' \le m + K$ we set $\mu_i = 0$ for $m' < i \le m + K$. And we have

$$\begin{aligned}\left\|\overrightarrow{X}_\star\right\|_2^2 &= \sum_{i=1}^{m+K}\frac{(\boldsymbol{v}_i^\top\overrightarrow{W}')^2(\lambda'+\mu_i)^2}{(\lambda'+\mu_i+\nu)^2} + \sum_{i=m+K+1}^{Kd}\frac{(\widehat{\boldsymbol{v}}_i^\top\overrightarrow{W}')^2\lambda'^2}{(\lambda'+\nu)^2} \\ &= \sum_{i=1}^{m+K}\frac{(\boldsymbol{v}_i^\top\overrightarrow{W}')^2(\lambda'+\mu_i)^2}{(\lambda'+\mu_i+\nu)^2} + \sum_{i=m+K+1}^{Kd}\frac{\overrightarrow{W}'^\top\widehat{\boldsymbol{v}}_i\widehat{\boldsymbol{v}}_i^\top\overrightarrow{W}'\lambda'^2}{(\lambda'+\nu)^2} \\ &= \sum_{i=1}^{m+K}\frac{(\boldsymbol{v}_i^\top\overrightarrow{W}')^2(\lambda'+\mu_i)^2}{(\lambda'+\mu_i+\nu)^2} + \overrightarrow{W}'^\top\left(I - \sum_{i=1}^{m+K}\widehat{\boldsymbol{v}}_i\widehat{\boldsymbol{v}}_i^\top\right)\frac{\overrightarrow{W}'\lambda'^2}{(\lambda'+\nu)^2} \\ &= \sum_{i=1}^{m+K}\frac{(\boldsymbol{v}_i^\top\overrightarrow{W}')^2(\lambda'+\mu_i)^2}{(\lambda'+\mu_i+\nu)^2} + \frac{\overrightarrow{W}'^\top(I - VV^\top)\overrightarrow{W}'\lambda'^2}{(\lambda'+\nu)^2}\end{aligned}$$

Let function

$$f(x) = \sum_{i=1}^{m+K}\frac{(\boldsymbol{v}_i^\top\overrightarrow{W}')^2(\lambda'+\mu_i)^2}{(\lambda'+\mu_i+x)^2} + \frac{\overrightarrow{W}'^\top(I - VV^\top)\overrightarrow{W}'\lambda'^2}{(\lambda'+x)^2}, x > 0.$$

Therefore

$$f(\nu) = \left\|\overrightarrow{X}_\star\right\|_2^2 = S^2.$$

This means that the dual optimiser ν is the root of the equation $f(x) = S^2$. Since $f(x)$ is a monotonically decreasing function with respect to x in the interval $(0, +\infty)$, and $\lim_{x\to 0^+} f(x) = \|\overrightarrow{W}'\|_2^2 > S^2$, $\lim_{x\to+\infty} f(x) = 0 < S^2$, there exists a unique root of the equation $f(x) = S^2$ in the interval $(0, +\infty)$. Therefore, given $\widetilde{Z}_t = \lambda' I + V\Lambda V^\top$, the problem $f(x) = S^2$ can be solved in $\mathcal{O}(Kd(m+K) + (m+K)\log\log\varepsilon)$ using Newton's method, and the optimal solution $\overrightarrow{W}_{t+1}$ can be computed in $\mathcal{O}(Kd(m+K))$ time, where ε is the machine precision. Therefore the total time complexity is $\mathcal{O}(Kd(m+K))$. □

Lemma 8 (Lemma 13, 14 of Zhang & Sugiyama (2023))**.** *For any $\delta \in (0,1)$, with probability at least $1-\delta$, we have*

$$\forall t > 0.\ \sum_{s=1}^{t}\ell_s(W_\star) - \ell_s(W_{s+1})$$

$$\leq (DS + \ln(K+1) + 2\ln((K+1)t+2)) \left(\frac{5}{4} + 4\ln\frac{\sqrt{1+2t}}{\delta}\right) + 2 + \ln(K+1)$$
$$+ \frac{\sqrt{6}}{24Kd}\lambda \left(Kd \ln \Delta_t^{\lambda/2} + m\ln(1 + 2\frac{D^2}{m\lambda}t)\right) + \frac{24Kd}{\lambda}\sum_{s=1}^{t} \left\|\overrightarrow{W}_{s+1} - \overrightarrow{W}_s\right\|_{Z_s}^2.$$

Proof. Let $\boldsymbol{\sigma}^+$ be the inverse of $\boldsymbol{\sigma}$,

$$\boldsymbol{\sigma}^+(\boldsymbol{p}) = \left[\ln\frac{p_1}{p_0}, \ln\frac{p_2}{p_0}, \ldots, \ln\frac{p_K}{p_0}\right]^\top, \|\boldsymbol{p}\|_1 = 1, \boldsymbol{p} \in [0,1]^{K+1}.$$

And define a intermedia function w.r.t. ℓ_s

$$\widehat{\ell}_s(z) := -\ln \boldsymbol{\sigma}_{y_s}(z), \widehat{\ell}_s(W\boldsymbol{x}_s) = \ell_s(W).$$

Let z_s be an intermediary prediction w.r.t. the distribution $\mathcal{P}_s = \mathcal{N}(W_s, (cZ_s)^{-1})$

$$z_s = \boldsymbol{\sigma}^+(\mathbb{E}_{W\sim\mathcal{P}_s}[\boldsymbol{\sigma}(W\boldsymbol{x}_s)]),$$

and z_s^μ be an smooth approximation of z_s

$$z_s^\mu = \boldsymbol{\sigma}^+(\mathrm{smooth}_\mu(\boldsymbol{\sigma}(z_s))),$$

where $\mathrm{smooth}_\mu(\boldsymbol{p}) = (1-\mu)\boldsymbol{p} + \mu/(K+1)$ for any probability distribution $\|\boldsymbol{p}\|_1 = 1, \boldsymbol{p} \in [0,1]^{K+1}$. Thus z_s and z_s^μ are $\mathcal{F}_s$-measurable,

$$\|z_s^\mu\|_\infty = \max_{k\in\langle K\rangle}\left|\ln\frac{\boldsymbol{\sigma}_k(z_s^\mu)}{\boldsymbol{\sigma}_0(z_s^\mu)}\right| \leq \max\left\{\left|\ln\frac{1-\mu K/(K+1)}{\mu/(K+1)}\right|, \left|\ln\frac{\mu/(K+1)}{1-\mu K/(K+1)}\right|\right\} \leq \ln\left(\frac{K+1}{\mu} - K\right).$$

We follow the proof of Zhang & Sugiyama (2023, Theorem 3). By dividing $\sum_{s=1}^t \ell_s(W_\star) - \ell_s(W_s)$ into three parts,

$$\sum_{s=1}^t \ell_s(W_\star) - \ell_s(W_s) = \sum_{s=1}^t \ell_s(W_\star) - \widehat{\ell}_s(z_s^\mu) + \sum_{s=1}^t \widehat{\ell}_s(z_s^\mu) - \widehat{\ell}_s(z_s) + \sum_{s=1}^t \widehat{\ell}_s(z_s) - \ell_s(W_s).$$

For the first term $\sum_{s=1}^t \ell_s(W_\star) - \widehat{\ell}_s(z_s^\mu)$, by Zhang & Sugiyama (2023, Lemma 13), we have

$$\sum_{s=1}^t \ell_s(W_\star) - \widehat{\ell}_s(z_s^\mu) \leq \frac{5}{4}(DS + S_\mu) + 4(DS + S_\mu)\ln\frac{\sqrt{1+2t}}{\delta}, \tag{45}$$

where $S_\mu = \ln(K+1) + 2\ln(\frac{K+1}{\mu} + 2) \geq \ln(K+1) + 2(\|z_s^\mu\|_\infty + 1)$. And for the second term $\sum_{s=1}^t \widehat{\ell}_s(z_s^\mu) - \widehat{\ell}_s(z_s)$, by Zhang & Sugiyama (2023, Lemma 17), if $\mu \in [0, \frac{1}{2}]$

$$\begin{aligned}\sum_{s=1}^t \widehat{\ell}_s(z_s^\mu) - \widehat{\ell}_s(z_s) &= \sum_{s=1}^t \ln\frac{\sigma_{y_s}(z_s)}{(1-\mu)\sigma_{y_s}(z_s) + \mu 1/(K+1)} \\ &\leq \sum_{s=1}^t \ln\frac{1}{1-\mu K/(K+1)} \leq \sum_{s=1}^t \ln\exp(2\mu K/(K+1)) \\ &\leq \sum_{s=1}^t 2\mu = 2\mu t. \end{aligned} \tag{46}$$

The second inequality is due to $\forall x \in [0, \frac{1}{2}]$. $\frac{1}{1-x} \leq \exp(2x)$. And by Lemma 9, the third term is bounded

$$\sum_{s=1}^t \widehat{\ell}_s(z_s) - \ell_s(W_{s+1}) \leq \frac{\sqrt{6}}{24Kd}\lambda\Delta_t^\lambda \left(Kd\ln 2\Delta_t + m\ln(1 + \frac{D^2}{m\lambda}t)\right) + \frac{24Kd}{\lambda}\sum_{s=1}^t \left\|\overrightarrow{W}_s - \overrightarrow{W}_{s+1}\right\|_{Z_s}^2. \tag{47}$$

Combining equation 45, equation 46 and equation 47, we get

$$\sum_{s=1}^{t} \ell_s(W_\star) - \ell_s(W_{s+1}) \leq \frac{5}{4}(DS + S_\mu) + 4(DS + S_\mu) \ln \frac{\sqrt{1+2t}}{\delta} + 2\mu t + \frac{\sqrt{6}}{24Kd}\lambda \Delta_t^\lambda \left(Kd \ln 2\Delta_t + m \ln(1 + \frac{D^2}{m\lambda} t) \right) + \frac{24Kd}{\lambda} \sum_{s=1}^{t} \left\| \overrightarrow{W}_{s+1} - \overrightarrow{W}_s \right\|_{Z_s}^2$$

And set $\mu = \frac{1}{t}$ to get

$$\sum_{s=1}^{t} \ell_s(W_\star) - \ell_s(W_{s+1}) \leq (DS + \ln(K+1) + 2\ln((K+1)t+2)) \left(\frac{5}{4} + 4 \ln \frac{\sqrt{1+2t}}{\delta} \right) + 2 + \ln(K+1) + \frac{\sqrt{6}}{24Kd}\lambda \Delta_t^\lambda \left(Kd \ln 2\Delta_t + m \ln(1 + \frac{D^2}{m\lambda} t) \right) + \frac{24Kd}{\lambda} \sum_{s=1}^{t} \left\| \overrightarrow{W}_{s+1} - \overrightarrow{W}_s \right\|_{Z_s}^2$$

□

Lemma 9. *For any $t > 0$*

$$\sum_{s=1}^{t} \widehat{\ell}_s(z_s) - \ell_s(W_{s+1}) \leq \frac{\sqrt{6}}{24Kd}\lambda \Delta_t^\lambda \left(Kd \ln 2\Delta_t^{\lambda/2} + m \ln(1 + 2\frac{D^2}{m\lambda} t) \right) + \frac{24Kd}{\lambda} \sum_{s=1}^{t} \left\| \overrightarrow{W}_s - \overrightarrow{W}_{s+1} \right\|_{Z_s}^2 .$$

Proof. By definition of z_s, we know

$$\exp(-\widehat{\ell}_s(z_s)) = \mathbb{E}_{W \sim \mathcal{P}_s} \left[\exp(-\ell_s(W)) \right].$$

And therefore

$$\begin{aligned} \sum_{s=1}^{t} \widehat{\ell}_s(z_s) - \ell_s(W_{s+1}) &= \sum_{s=1}^{t} -\ln \exp(-\widehat{\ell}_s(z_s)) - \ell_s(W_{s+1}) \\ &= \sum_{s=1}^{t} -\ln \mathbb{E}_{W \sim \mathcal{P}_s} \left[\exp(-\ell_s(W)) \right] - \ell_s(W_{s+1}) \\ &= \sum_{s=1}^{t} -\ln \mathbb{E}_{W \sim \mathcal{P}_s} \left[\exp(-(\ell_s(W) - \ell_s(W_{s+1}))) \right] \end{aligned}$$

We shift the mean of $\mathcal{P}_s$ from W_s to W_{s+1}, and get $\mathcal{Q}_s := \mathcal{N}(W_{s+1}, (cZ_s)^{-1})$. For probability density function $f_p(W), f_q(W)$, the following relation holds

$$\frac{f_p(W)}{f_q(W)} = \exp\left(-\left\langle cZ_s(\overrightarrow{W}_{s+1} - \overrightarrow{W}), \overrightarrow{W}_s - \overrightarrow{W}_{s+1} \right\rangle \right) \exp\left(-\frac{1}{2} \left\| \overrightarrow{W}_s - \overrightarrow{W}_{s+1} \right\|_{cZ_s}^2 \right).$$

Thus

$$\begin{aligned} &\sum_{s=1}^{t} -\ln \mathbb{E}_{W \sim \mathcal{P}_s} \left[\exp(-(\ell_s(W) - \ell_s(W_{s+1}))) \right] \\ &= \sum_{s=1}^{t} -\ln \mathbb{E}_{W \sim \mathcal{Q}_s} \left[\frac{f_p(W)}{f_q(W)} \exp\left(-\left(\ell_s(W) - \ell_s(W_{s+1}) \right) \right) \right] \\ &= \sum_{s=1}^{t} -\ln \mathbb{E}_{W \sim \mathcal{Q}_s} \left[\exp\left(-\left(\ell_s(W) - \ell_s(W_{s+1}) + \left\langle \overrightarrow{W}_{s+1} - \overrightarrow{W}, cZ_s^\top (\overrightarrow{W}_s - \overrightarrow{W}_{s+1}) \right\rangle \right) \right) \right] \\ &\quad + \frac{1}{2} \left\| \overrightarrow{W}_{s+1} - \overrightarrow{W}_s \right\|_{cZ_s}^2 \end{aligned} \tag{48}$$

By Tran-Dinh et al. (2015, Theorem 4(d)), and $\forall x > 0.\ (\exp(x) - x - 1)/x^2 \leq \exp(x^2)$,

$$\ell_s(W) - \ell_s(W_{s+1}) \leq \left\langle \nabla \ell_s(\overrightarrow{W}_{s+1}), \overrightarrow{W} - \overrightarrow{W}_{s+1} \right\rangle + \mathrm{e}^{6\|W - W_{s+1}\|_F^2} \left\| \overrightarrow{W} - \overrightarrow{W}_{s+1} \right\|_{\nabla^2 \ell_s(W_{s+1})}^2,$$

the logarithm expectation term in equation 48 can be bounded as follows

$$
\begin{aligned}
&-\ln \mathbb{E}_{W\sim\mathcal{Q}_s}\left[\exp\left(-\left(\ell_s(W)-\ell_s(W_{s+1})+\left\langle cZ_s(\overrightarrow{W}_{s+1}-\overrightarrow{W}_s),\overrightarrow{W}-\overrightarrow{W}_{s+1}\right\rangle\right)\right)\right]\\
\leq\ &-\ln \mathbb{E}\left[\exp\left(-\left(\left\langle \nabla\ell_s(\overrightarrow{W}_s)+cZ_s(\overrightarrow{W}_{s+1}-\overrightarrow{W}_s),\overrightarrow{W}-\overrightarrow{W}_{s+1}\right\rangle+\mathrm{e}^{6\|W-W_{s+1}\|_F^2}\left\|\overrightarrow{W}-\overrightarrow{W}_{s+1}\right\|^2_{\nabla^2\ell_s(W_{s+1})}\right)\right)\right]\\
&\text{(by Jensen's inequality)}\\
=\ &-\ln\exp\left(-\mathbb{E}\left[\left\langle \nabla\ell_s(\overrightarrow{W}_s)+cZ_s(\overrightarrow{W}_{s+1}-\overrightarrow{W}_s),\overrightarrow{W}-\overrightarrow{W}_{s+1}\right\rangle+\mathrm{e}^{6\|W-W_{s+1}\|_F^2}\left\|\overrightarrow{W}-\overrightarrow{W}_{s+1}\right\|^2_{\nabla^2\ell_s(W_{s+1})}\right]\right)\\
=\ &\mathbb{E}_{W\sim\mathcal{Q}_s}\left[\left\langle \nabla\ell_s(\overrightarrow{W}_s)+cZ_s(\overrightarrow{W}_{s+1}-\overrightarrow{W}_s),\overrightarrow{W}-\overrightarrow{W}_{s+1}\right\rangle+\mathrm{e}^{6\|W-W_{s+1}\|_F^2}\left\|\overrightarrow{W}-\overrightarrow{W}_{s+1}\right\|^2_{\nabla^2\ell_s(W_{s+1})}\right]\\
=\ &\mathbb{E}_{W\sim\mathcal{Q}_s}\left[\mathrm{e}^{6\|W-W_{s+1}\|_F^2}\left\|\overrightarrow{W}-\overrightarrow{W}_{s+1}\right\|^2_{\nabla^2\ell_s(W_{s+1})}\right]\\
&\text{(by Cachy-Schwarz inequality)}\\
\leq\ &\sqrt{\mathbb{E}_{W\sim\mathcal{Q}_s}\left[\mathrm{e}^{12\|W-W_{s+1}\|_F^2}\right]}\sqrt{\mathbb{E}_{W\sim\mathcal{Q}_s}\left[\left\|\overrightarrow{W}-\overrightarrow{W}_{s+1}\right\|^4_{\nabla^2\ell_s(W_{s+1})}\right]}
\end{aligned}
$$

For the first term $\sqrt{\mathbb{E}_{W\sim\mathcal{Q}_s}\left[\mathrm{e}^{12\|W-W_{s+1}\|_F^2}\right]}$, let $U\Sigma U^\top = Z_s$ temporarily be the eigendecomposition of Z_s. Thus random variable $X := \sqrt{c\Sigma}U^\top(\overrightarrow{W}-\overrightarrow{W}_{s+1})$ follows the distribution $\mathcal{N}(0,I)$, and

$$
\begin{aligned}
\sqrt{\mathbb{E}_{W\sim\mathcal{Q}_s}\left[\mathrm{e}^{12\|W-W_{s+1}\|_F^2}\right]} &= \sqrt{\mathbb{E}_X\left[\exp\left(12\left\|U(c\Sigma)^{-\frac{1}{2}}X\right\|_2^2\right)\right]}\\
&= \sqrt{\mathbb{E}_X\left[\exp\left(12\sum_{i=1}^{Kd}\frac{1}{c\lambda_i}X_i^2\right)\right]}\\
&\text{(by } Z_s\succeq\lambda I\text{)}\\
&= \mathbb{E}_X\left[\exp\left(\frac{12}{c\lambda}X_i^2\right)\right]^{Kd/2}\\
&\text{(by Jensen's inequality when } Kd\geq 2\text{)}\\
&\leq \mathbb{E}_X\left[\exp\left(\frac{6Kd}{c\lambda}X_i^2\right)\right]
\end{aligned}
$$

By setting $\frac{6Kd}{c\lambda}\leq\frac{1}{8}$, and $\mathbb{E}_{X\sim\mathcal{X}_1^2}\left[\exp(X/8)\right]<\sqrt{2}$, we have

$$
\sqrt{\mathbb{E}_{W\sim\mathcal{Q}_s}\left[\mathrm{e}^{12\|W-W_{s+1}\|_F^2}\right]}\leq\sqrt{2}. \tag{49}
$$

When $Kd=1$, by setting $\frac{6Kd}{c\lambda}=\frac{6}{c\lambda}\leq\frac{1}{8}$, we have

$$
\mathbb{E}_X\left[\exp\left(\frac{12}{c\lambda}X_i^2\right)\right]^{\frac{1}{2}}\leq\mathbb{E}_X\left[\exp\left(\frac{X_i^2}{4}\right)\right]^{\frac{1}{2}}\leq\sqrt{2},
$$

which means that equation 49 holds for any $Kd\geq 1$. And for the second term $\sqrt{\mathbb{E}_{W\sim\mathcal{Q}_s}\left[\left\|\overrightarrow{W}-\overrightarrow{W}_{s+1}\right\|^4_{\nabla^2\ell_s(W_{s+1})}\right]}$, Let the full eigendecomposition $U\Sigma U^\top = Z_s^{-\frac{1}{2}}\nabla^2\ell_s(W_{s+1})Z_s^{-\frac{1}{2}}$ with rank at most K. And random variable $X := U^\top(cZ_s)^{\frac{1}{2}}(\overrightarrow{W}-\overrightarrow{W}_{s+1})$ follows the distribution $\mathcal{N}(0,I)$, and

$$
\sqrt{\mathbb{E}_{W\sim\mathcal{Q}_s}\left[\left\|\overrightarrow{W}-\overrightarrow{W}_{s+1}\right\|^4_{\nabla^2\ell_s(W_{s+1})}\right]}=\sqrt{\mathbb{E}_X\left[\left\|\nabla^2\ell_s(W_{s+1})^{\frac{1}{2}}(cZ_s)^{-\frac{1}{2}}UX\right\|_2^4\right]}
$$

$$= \sqrt{\mathbb{E}_X \left[\left(X^\top U^\top Z_s^{-\frac{1}{2}} \nabla^2 \ell_s(W_{s+1}) Z_s^{-\frac{1}{2}} UX/c \right)^2 \right]}$$
$$= \sqrt{\mathbb{E}_X \left[\left(\sum_{i=1}^{K} \frac{\lambda_i}{c} X_i^2 \right)^2 \right]}$$
$$= \sqrt{\mathbb{E}_X \left[\sum_{i,j} \frac{\lambda_i \lambda_j}{c^2} X_i^2 X_j^2 \right]}$$
$$= \frac{1}{c} \sqrt{\sum_{i,j} \lambda_i \lambda_j \mathbb{E}_X \left[X_i^2 X_j^2 \right]}$$

By $\forall i, j.\ \mathbb{E}_{X_i, X_j \sim \mathcal{N}(0,1)} \left[X_i^2 X_j^2 \right] \leq 3$

$$\sqrt{\mathbb{E}_{W \sim \mathcal{Q}_s} \left[\left\| \overrightarrow{W} - \overrightarrow{W}_{s+1} \right\|^4_{\nabla^2 \ell_s(W_{s+1})} \right]}$$
$$\leq \frac{1}{c} \sqrt{\sum_{i,j} \lambda_i \lambda_j 3} = \frac{\sqrt{3}}{c} \sum_{i=1}^{Kd} \lambda_i$$
$$= \frac{\sqrt{3}}{c} \operatorname{tr}(Z_s^{-\frac{1}{2}} \nabla^2 \ell_s(W_{s+1}) Z_s^{-\frac{1}{2}})$$
$$= \frac{\sqrt{3}}{c} \operatorname{tr}(\nabla^2 \ell_s(W_{s+1}) Z_s^{-1}) \tag{50}$$

Combining equation 49 and equation 50, the equation equation 48 can be bounded as follows

$$\sum_{s=1}^{t} \widehat{\ell}_s(z_s) - \ell_s(W_{s+1}) \leq \frac{\sqrt{6}}{c} \sum_{s=1}^{t} \operatorname{tr}(\nabla^2 \ell_s(W_{s+1}) Z_s^{-1}) + \frac{1}{2} \sum_{s=1}^{t} \left\| \overrightarrow{W}_s - \overrightarrow{W}_{s+1} \right\|^2_{cZ_s}.$$

To conclude, we recall that $\frac{6Kd}{c\lambda} \leq \frac{1}{8}$. Therefore, we set $c := \frac{48Kd}{\lambda}$

$$\sum_{s=1}^{t} \widehat{\ell}_s(z_s) - \ell_s(W_{s+1}) \leq \frac{\sqrt{6}}{48Kd} \lambda \sum_{s=1}^{t} \operatorname{tr}(\nabla^2 \ell_s(W_{s+1}) Z_s^{-1}) + \frac{24Kd}{\lambda} \sum_{s=1}^{t} \left\| \overrightarrow{W}_s - \overrightarrow{W}_{s+1} \right\|^2_{Z_s}.$$

And by Lemma 12, if $\lambda \geq 2LD^2$

$$\sum_{s=1}^{t} \operatorname{tr}(\nabla^2 \ell_s(W_{s+1}) Z_s^{-1}) \leq 2(Kd \ln \Delta_t^{\lambda/2} + m \ln(1 + 2\frac{D^2}{m\lambda} t)).$$

Therefore, if $\lambda \geq 2LD^2$, we have

$$\sum_{s=1}^{t} \widehat{\ell}_s(z_s) - \ell_s(W_{s+1}) \leq \frac{\sqrt{6}}{24Kd} \lambda \left(Kd \ln 2\Delta_t + m \ln(1 + \frac{D^2}{m\lambda} t) \right) + \frac{24Kd}{\lambda} \sum_{s=1}^{t} \left\| \overrightarrow{W}_s - \overrightarrow{W}_{s+1} \right\|^2_{Z_s}$$

□

Lemma 10. *For any $t > 0$, we have*

$$\left\| \overrightarrow{W}_{t+1} - \overrightarrow{W}_t \right\|_{Z_t} \leq \frac{2\eta D}{\sqrt{\lambda}}.$$

Proof. By the definition of W_t,

$$W_{t+1} := \arg\min_{W \in \mathcal{W}} \left\langle \eta \nabla \ell_t(\overrightarrow{W}_t), \overrightarrow{W} \right\rangle + \frac{1}{2} \left\| \overrightarrow{W} - \overrightarrow{W}_t \right\|^2_{Z_t + \eta \nabla^2 \ell_t(W_t)},$$

W_{t+1} satisfies the first order optimality condition

$$\left\langle \eta \nabla \ell_t(\overrightarrow{W}_t) + (Z_t + \eta \nabla^2 \ell_t(W_t))(\overrightarrow{W}_{t+1} - \overrightarrow{W}_t), \overrightarrow{W}_{t+1} - \overrightarrow{W}_t \right\rangle \leq 0.$$

Rearrange to get

$$\left\| \overrightarrow{W}_{t+1} - \overrightarrow{W}_t \right\|^2_{Z_t + \eta \nabla^2 \ell_t(W_t)} \leq \left\langle \eta \nabla \ell_t(\overrightarrow{W}_t), \overrightarrow{W}_t - \overrightarrow{W}_{t+1} \right\rangle.$$

By $\eta \nabla^2 \ell_t(W_t) \succeq 0$ and Cauchy-Schwarz inequality,

$$\begin{aligned}
\left\| \overrightarrow{W}_{t+1} - \overrightarrow{W}_t \right\|_{Z_t} &\leq \left\| \eta \nabla \ell_t(\overrightarrow{W}_t) \right\|_{Z_t^{-1}} \\
&\quad \text{(by } Z_t \succeq \lambda I\text{)} \\
&\leq \frac{\eta}{\sqrt{\lambda}} \left\| \nabla \ell_t(\overrightarrow{W}_t) \right\|_2 \\
&= \frac{\eta}{\sqrt{\lambda}} \left\| \boldsymbol{\sigma}(W_t \boldsymbol{x}_t) - \boldsymbol{y}_t \right\|_2 \left\| \boldsymbol{x}_t \right\|_2 \\
&\leq \frac{2\eta D}{\sqrt{\lambda}}
\end{aligned}$$

□

Lemma 11 (Range term and Negative term of OMD analysis)**.** *For any $t > 0$, we have*

$$\left\langle \nabla \widetilde{\ell}_t(\overrightarrow{W}_t), \overrightarrow{W}_t - \overrightarrow{W}_\star \right\rangle \leq \frac{1}{2\eta} \left\| \overrightarrow{W}_\star - \overrightarrow{W}_{t-1} \right\|^2_{Z_{t-1}} - \frac{1}{2\eta} \left\| \overrightarrow{W}_\star - \overrightarrow{W}_t \right\|^2_{Z_{t-1}} - \frac{1}{2\eta} \left\| \overrightarrow{W}_{t-1} - \overrightarrow{W}_t \right\|^2_{Z_{t-1}}.$$

Proof. By the update rule of W_t

$$W_t := \arg\min_{W \in \mathcal{W}} \widetilde{\ell}(W) + \frac{1}{2\eta} \left\| \overrightarrow{W} - \overrightarrow{W}_{t-1} \right\|^2_{Z_{t-1}},$$

$\widetilde{\ell}(W) = \left\langle \nabla \ell_t(\overrightarrow{W}_{t-1}), \overrightarrow{W} \right\rangle + \frac{1}{2} \left\| \overrightarrow{W} - \overrightarrow{W}_{t-1} \right\|^2_{\nabla^2 \ell_t(W_{t-1})}$, we have the first order optimality condition

$$\left\langle \eta \nabla \widetilde{\ell}(\overrightarrow{W}_t) + Z_{t-1} \left(\overrightarrow{W}_t - \overrightarrow{W}_{t-1} \right), \overrightarrow{W}_\star - \overrightarrow{W}_t \right\rangle \geq 0.$$

Rearranging the above equation, we have

$$\begin{aligned}
\left\langle \eta \nabla \widetilde{\ell}(\overrightarrow{W}_t), \overrightarrow{W}_t - \overrightarrow{W}_\star \right\rangle &\leq \left\langle Z_{t-1} \left(\overrightarrow{W}_t - \overrightarrow{W}_{t-1} \right), \overrightarrow{W}_\star - \overrightarrow{W}_t \right\rangle \\
&= \frac{1}{2} \left\| \overrightarrow{W}_\star - \overrightarrow{W}_{t-1} \right\|^2_{Z_{t-1}} - \frac{1}{2} \left\| \overrightarrow{W}_\star - \overrightarrow{W}_t \right\|^2_{Z_{t-1}} - \frac{1}{2} \left\| \overrightarrow{W}_{t-1} - \overrightarrow{W}_t \right\|^2_{Z_{t-1}}
\end{aligned}$$

and complete the proof. □

Proof of Lemma 7

Proof. The following steps is by Taylor's Theorem to perform second order expansion.

$$\begin{aligned}
\left| \boldsymbol{r}^\top \boldsymbol{\sigma}(W_\star \boldsymbol{x}) - \boldsymbol{r}^\top \boldsymbol{\sigma}(W_t \boldsymbol{x}) \right| &\leq \left| \boldsymbol{r}^\top \nabla \boldsymbol{\sigma}(W_t \boldsymbol{x})(\overrightarrow{W}_\star - \overrightarrow{W}_t) \boldsymbol{x} \right| \\
&\quad + \left| \frac{1}{2} \boldsymbol{x}^\top (\overrightarrow{W}_\star - \overrightarrow{W}_t)^\top (D^2 \boldsymbol{\sigma}(\xi_1)[\boldsymbol{r}])(W_\star - W_t) \boldsymbol{x} \right| \\
&= \left| (\nabla \boldsymbol{\sigma}(W_t \boldsymbol{x}) \boldsymbol{r} \otimes \boldsymbol{x})^\top (\overrightarrow{W}_\star - \overrightarrow{W}_t) \right| \\
&\quad + \frac{1}{2} \left| (\overrightarrow{W}_\star - \overrightarrow{W}_t)^\top (D^2 \boldsymbol{\sigma}(\xi_1)[\boldsymbol{r}] \otimes \boldsymbol{x}\boldsymbol{x}^\top)(\overrightarrow{W}_\star - \overrightarrow{W}_t) \right|
\end{aligned}$$

for some ξ_1 in the line segment between $W_t\boldsymbol{x}$ and $W_\star\boldsymbol{x}$. The first term can be further expanded as follows,

$$\begin{aligned}\left|(\nabla\boldsymbol{\sigma}(W_t\boldsymbol{x})\boldsymbol{r}\otimes\boldsymbol{x})^\top(\overrightarrow{W}_\star-\overrightarrow{W}_t)\right| &= \left|((\nabla\boldsymbol{\sigma}(W_t\boldsymbol{x})\otimes\boldsymbol{x})\boldsymbol{r})^\top(\overrightarrow{W}_\star-\overrightarrow{W}_t)\right| \\ &\le \|(\nabla\boldsymbol{\sigma}(W_t\boldsymbol{x})\otimes\boldsymbol{x})\boldsymbol{r}\|_{Z_t^{-1}}\left\|\overrightarrow{W}_\star-\overrightarrow{W}_t\right\|_{Z_t} \\ &\le \sqrt{\beta_t(\delta)}\,\|(\nabla\boldsymbol{\sigma}(W_t\boldsymbol{x})\otimes\boldsymbol{x})\boldsymbol{r}\|_{Z_t^{-1}}\end{aligned}$$

The first equality is by the property of Kronecker product, and the second inequality is by Cauchy-Schwarz inequality, and third inequality is by Theorem 1.

The second term can be expended to

$$\begin{aligned}\left|(\overrightarrow{W}_\star-\overrightarrow{W}_t)^\top(D^2\boldsymbol{\sigma}(\xi_1)[\boldsymbol{r}]\otimes\boldsymbol{x}\boldsymbol{x}^\top)(\overrightarrow{W}_\star-\overrightarrow{W}_t)\right| &\le \left\|\overrightarrow{W}_\star-\overrightarrow{W}_t\right\|_{Z_t}\left\|Z_t^{-\frac{1}{2}}(D^2\boldsymbol{\sigma}(\xi_1)[\boldsymbol{r}]\otimes\boldsymbol{x}\boldsymbol{x}^\top)Z_t^{-\frac{1}{2}}\right\|_{2,2}\left\|\overrightarrow{W}_\star-\overrightarrow{W}_t\right\|_{Z_t} \\ &\le \beta_t(\delta)\left\|Z_t^{-\frac{1}{2}}(D^2\boldsymbol{\sigma}(\xi_1)[\boldsymbol{r}]\otimes\boldsymbol{x}\boldsymbol{x}^\top)Z_t^{-\frac{1}{2}}\right\|_{2,2}\end{aligned}$$

The first inequality is by Cauchy-Schwarz inequality and definition of operator norm, and the second inequality is by Theorem 1. For the last term

$$\begin{aligned}\left\|Z_t^{-\frac{1}{2}}(D^2\boldsymbol{\sigma}(\xi_1)[\boldsymbol{r}]\otimes\boldsymbol{x}\boldsymbol{x}^\top)Z_t^{-\frac{1}{2}}\right\|_{2,2} &= \max_i|\lambda_i(Z_t^{-\frac{1}{2}}(D^2\boldsymbol{\sigma}(\xi_1)[\boldsymbol{r}]\otimes\boldsymbol{x}\boldsymbol{x}^\top)Z_t^{-\frac{1}{2}})| \\ &\le \lambda_{\max}(Z_t^{-\frac{1}{2}}(\|\boldsymbol{r}\|_1 I_K\otimes\boldsymbol{x}\boldsymbol{x}^\top)Z_t^{-\frac{1}{2}}) \\ &= \|\boldsymbol{r}\|_1\lambda_{\max}(Z_t^{-\frac{1}{2}}(I_K\otimes\boldsymbol{x}\boldsymbol{x}^\top)Z_t^{-\frac{1}{2}}) \\ &= \|\boldsymbol{r}\|_1\lambda_{\max}(Z_t^{-\frac{1}{2}}(I_K\otimes\boldsymbol{x})(I_K\otimes\boldsymbol{x})^\top Z_t^{-\frac{1}{2}}) \\ &= \|\boldsymbol{r}\|_1\lambda_{\max}((I_K\otimes\boldsymbol{x})^\top Z_t^{-\frac{1}{2}}Z_t^{-\frac{1}{2}}(I_K\otimes\boldsymbol{x})) \\ &= \|\boldsymbol{r}\|_1\left\|Z_t^{-\frac{1}{2}}(I_K\otimes\boldsymbol{x})\right\|_{2,2}^2\end{aligned}$$

The first equality is by the property of operator norm $\|A\|_{2,2}=\max_i|\lambda_i(A)|$ for symmetric matrix A, and the second inequality is by equation 29 and the two-side bound $-M\preceq A\preceq M \implies \|A\|_{2,2}\le\lambda_{\max}(M)$.

To conclude, we have

$$\left|\boldsymbol{r}^\top\boldsymbol{\sigma}(W_\star\boldsymbol{x})-\boldsymbol{r}^\top\boldsymbol{\sigma}(W_t\boldsymbol{x})\right| \le \sqrt{\beta_t(\delta)}\,\|(\nabla\boldsymbol{\sigma}(W_t\boldsymbol{x})\otimes\boldsymbol{x})\boldsymbol{r}\|_{Z_t^{-1}} + \frac{\|\boldsymbol{r}\|_1}{2}\beta_t(\delta)\left\|Z_t^{-\frac{1}{2}}(I\otimes\boldsymbol{x})\right\|_{2,2}^2$$

□

Lemma 12. *If $\lambda\ge 2LD^2$, then for any $t>0$, we have*

$$\sum_{s=1}^t\operatorname{tr}\left(\nabla^2\ell_s(W_{s+1})Z_s^{-1}\right)=\sum_{s=1}^t\left\|\sqrt{\nabla\boldsymbol{\sigma}(W_{s+1}\boldsymbol{x}_s)}\otimes\boldsymbol{x}_s\right\|_{Z_s^{-1}}^2\le 2\left(Kd\ln 2\Delta_t^\lambda+m\ln(1+\frac{D^2}{m\lambda}t)\right).$$

Proof. Let H_t be the original version of Z_t,

$$H_t=\lambda I+\sum_{s=1}^{t-1}\nabla\boldsymbol{\sigma}(W_{s+1}\boldsymbol{x}_s)\otimes\boldsymbol{x}_s\boldsymbol{x}_s^\top=\lambda I+\sum_{s=1}^{t-1}\nabla^2\ell_s(W_{s+1}).$$

Therefore

$$\begin{aligned}\sum_{s=1}^t\left\|\sqrt{\boldsymbol{\sigma}(W_{s+1}\boldsymbol{x}_s)}\otimes\boldsymbol{x}_s\right\|_{Z_s^{-1}}^2 &= \sum_{s=1}^t\operatorname{tr}(\nabla^2\ell_s(W_{s+1})Z_s^{-1}) \\ &\quad\text{(by FD Lemma 5)} \\ &\le \sum_{s=1}^t 2\operatorname{tr}(\nabla^2\ell_s(W_{s+1})H_s^{-1})\end{aligned}$$

(by Lemma 13)

$$\leq 2\left(Kd\ln\Delta_t^{\lambda/2} + m\ln(1+2\frac{D^2}{m\lambda}t)\right)$$

□

Lemma 13. *If $\lambda \geq 2LD^2$, then for any $t > 0$, we have*

$$\sum_{s=1}^{t}\mathrm{tr}(\nabla^2\ell_s(W_{s+1})H_s^{-1}) \leq Kd\ln\Delta_t^{\lambda/2} + m\ln(1+2\frac{D^2}{m\lambda}t).$$

Proof. Let

$$H_t = \lambda I + \sum_{s=1}^{t-1}\nabla\boldsymbol{\sigma}(W_{s+1}\boldsymbol{x}_s)\otimes\boldsymbol{x}_s\boldsymbol{x}_s^\top = \lambda I + \sum_{s=1}^{t-1}\nabla^2\ell_s(W_{s+1}).$$

And $M_t := \frac{\lambda}{2}I + \sum_{s=1}^{t}\nabla^2\ell_s(W_{s+1})$, and by $\nabla^2\ell_t(W_{t+1}) \preceq \frac{\lambda}{2}I$, we have

$$H_t = \frac{\lambda}{2}I + \sum_{s=1}^{t-1}\nabla^2\ell_s(W_{s+1}) + \frac{\lambda}{2}I \succeq \frac{\lambda}{2}I + \sum_{s=1}^{t-1}\nabla^2\ell_s(W_{s+1}) + \nabla^2\ell_t(W_{t+1}) = M_t.$$

Therefore

$$\begin{aligned}
\sum_{s=1}^{t}\mathrm{tr}(\nabla^2\ell_s(W_{s+1})H_s^{-1}) &\leq \sum_{s=1}^{t}\mathrm{tr}(\nabla^2\ell_s(W_{s+1})M_s^{-1}) \\
&= \sum_{s=1}^{t}\mathrm{tr}((M_s - M_{s-1})M_s^{-1}) \\
&\quad \text{(by the concavity of } y = \ln|X|) \\
&\leq \sum_{s=1}^{t}\ln\frac{|M_s|}{|M_{s-1}|} \\
&= \ln\frac{|M_t|}{\left|\frac{\lambda}{2}I\right|}
\end{aligned}$$

By the FD Lemma 6, we have

$$\sum_{s=1}^{t}\mathrm{tr}(\nabla^2\ell_s(W_{s+1})H_s^{-1}) \leq Kd\ln\Delta_t^{\lambda/2} + m\ln(1+2\frac{D^2}{m\lambda}t).$$

□

Lemma 14. *If $\lambda \geq 2LD^2$, then for any $t > 0$, we have*

$$\sum_{s=1}^{t}\|I\otimes\boldsymbol{x}_s\|_{Z_s^{-1}}^2 \leq 2\kappa\left(Kd\ln\Delta_t^{\lambda/2} + m\ln(1+2\frac{D^2}{m\lambda}t)\right).$$

Proof. Let H_t be the original version of Z_t,

$$H_t = \lambda I + \sum_{s=1}^{t-1}\nabla^2\ell_s(W_{s+1}).$$

And $M_t := \frac{\lambda}{2}I + \frac{1}{\kappa}\sum_{s=1}^{t} I\otimes\boldsymbol{x}_s\boldsymbol{x}_s^\top$, and by $\frac{1}{\kappa}(I\otimes\boldsymbol{x}_s\boldsymbol{x}_s^\top) \preceq LD^2I \preceq \frac{\lambda}{2}I$, we have

$$H_t = \frac{\lambda}{2}I + \sum_{s=1}^{t-1}\nabla^2\ell_s(W_{s+1}) + \frac{\lambda}{2}I \succeq \frac{\lambda}{2}I + \sum_{s=1}^{t-1}\frac{1}{\kappa}(I\otimes\boldsymbol{x}_s\boldsymbol{x}_s^\top) + \frac{1}{\kappa}(I\otimes\boldsymbol{x}_t\boldsymbol{x}_t^\top) = M_t.$$

Therefore

$$\begin{aligned}
\sum_{s=1}^{t} \|I \otimes \boldsymbol{x}_s\|^2_{Z_s^{-1}} &= \sum_{s=1}^{t} \operatorname{tr}((I \otimes \boldsymbol{x}_s \boldsymbol{x}_s^\top) Z_s^{-1}) \\
&\quad \text{(by FD Lemma 5)} \\
&\leq \sum_{s=1}^{t} 2 \operatorname{tr}((I \otimes \boldsymbol{x}_s \boldsymbol{x}_s^\top) H_s^{-1}) \\
&\leq \sum_{s=1}^{t} 2\kappa \operatorname{tr}(\frac{1}{\kappa}(I \otimes \boldsymbol{x}_s \boldsymbol{x}_s^\top) M_s^{-1}) \\
&= 2\kappa \sum_{s=1}^{t} \operatorname{tr}((M_s - M_{s-1}) M_s^{-1}) \\
&\quad \text{(by the concavity of } y = \ln |X|) \\
&\leq 2\kappa \sum_{s=1}^{t} \ln \frac{|M_s|}{|M_{s-1}|} \\
&\leq 2\kappa \sum_{s=1}^{t} \ln \frac{|M_s|}{|M_{s-1}|} \\
&= 2\kappa \ln \frac{|M_t|}{\left|\frac{\lambda}{2} I\right|} \\
&\leq 2\kappa \ln \frac{\left|\frac{\lambda}{2} I + \sum_{s=1}^{t} \nabla \boldsymbol{\sigma}(W_{s+1} \boldsymbol{x}_s) \otimes \boldsymbol{x}_s \boldsymbol{x}_s^\top\right|}{\left|\frac{\lambda}{2} I\right|}
\end{aligned}$$

The last inequality is by $\nabla \boldsymbol{\sigma}(W_{s+1} \boldsymbol{x}_s) \otimes \boldsymbol{x}_s \boldsymbol{x}_s^\top \succeq \frac{1}{\kappa}(I \otimes \boldsymbol{x}_s \boldsymbol{x}_s^\top)$, and $M_t \preceq 0$. By the FD Lemma 6, we have

$$\sum_{s=1}^{t} \|I \otimes \boldsymbol{x}_s\|^2_{Z_s^{-1}} \leq \kappa 2 \left(K d \ln \Delta_t^{\lambda/2} + m \ln(1 + 2\frac{D^2}{m\lambda} t) \right).$$

□

Lemma 15. *If $\lambda \geq 2LD^2$, then for any $t > 0$, we have*

$$\sum_{s=1}^{t} \|I \otimes \boldsymbol{x}_s\|^2_{Z_{s+1}^{-1}} \leq 2\kappa \left(K d \ln \Delta_{t+1}^{\lambda} + m \ln(1 + \frac{D^2}{m\lambda}(t+1)) \right).$$

Proof. Let H_t be the original version of Z_t,

$$H_t = \lambda I + \sum_{s=1}^{t-1} \nabla^2 \ell_s(W_{s+1}).$$

Therefore

$$\begin{aligned}
\sum_{s=1}^{t} \|I \otimes \boldsymbol{x}_s\|^2_{Z_{s+1}^{-1}} &= \sum_{s=1}^{t} \operatorname{tr}((I \otimes \boldsymbol{x}_s \boldsymbol{x}_s^\top) Z_{s+1}^{-1}) \\
&\quad \text{(by FD Lemma 5)} \\
&\leq \sum_{s=1}^{t} 2 \operatorname{tr}((I \otimes \boldsymbol{x}_s \boldsymbol{x}_s^\top) H_{s+1}^{-1}) \\
&\leq \kappa \sum_{s=1}^{t} 2 \operatorname{tr}((H_{s+1} - H_s) H_{s+1}^{-1}) \\
&\quad \text{(by the concavity of } y = \ln |X|)
\end{aligned}$$

$$\leq \kappa \sum_{s=1}^{t} 2 \ln \frac{|H_{s+1}|}{|H_s|} \leq 2\kappa \sum_{s=1}^{t} \ln \frac{|H_{s+1}|}{|H_s|}$$
$$= 2\kappa \ln \frac{|H_{t+1}|}{|\lambda I|}$$

By the FD Lemma 6, we have

$$\sum_{s=1}^{t} \|I \otimes \boldsymbol{x}_s\|^2_{Z_{s+1}^{-1}} \leq 2\kappa \left(Kd \ln \Delta_{t+1}^{\lambda} + m \ln(1 + \frac{D^2}{m\lambda}(t+1)) \right).$$

□

Table 1: Dataset details.

Dataset	#Examples	#Features	#Classes
MFEAT	2,000	216	10
MNIST	70,000	784	10
DNA	3,186	180	3
phpbL6t4U	13,910	128	6
php3isjYz	1,212	100	2
phpdReP6S	2,534	72	2
phptd5jYj	1,593	256	2
SPAM	4,601	57	2
INDIAN	9,144	220	2

C Experiment Appendix

C.1 Additional Details of Experiments

The experiments are run on Intel(R) Core(TM) i5-14500 CPU (20 cores, 0.8 GHz min, 5.0 GHz max) with 16GB RAM, and the code is implemented in Python 3.12.13, with single thread. Hyperparameters (λ, η) are selected via grid search over grid factors $2^l \cdot \lambda_0$ and $2^e \cdot \eta_0$ with exponents $l, e \in \{-0.5, -0.25, 0, 0.25, 0.5, 0.75, 1.0\}$, centered around baseline values $\lambda_0 = \sqrt{K}\alpha S$ and $\eta_0 = S/2 + \ln(K+1)/4$. The tuned hyperparameters are presented in Table 2 For multinomial synthetic experiments ($(d, K) = (100, 4)$, horizon $T = 20000$, sketch size $m = 40$), the selected parameters are $\lambda = 1.0$ and $\eta = 0.5$. For binary synthetic experiments ($(d, K) = (400, 1)$, $T = 20000$, $m = 40$), we use $\lambda = 0.5$ and $\eta = 0.5$. Other parameters are: candidate size $|\mathcal{A}_t| = 10$, parameter norm bound $S = 5$, number of outcomes $K = 1$, and reward vector $\boldsymbol{r} = [0, 1]$. $W_\star$ is randomly generated from the distribution $\mathbb{N}(-1, 4)^{K \times d}$, and normalized to S, to ensure $W_\star$ is expectedly distant from 0. Every action is uniformly drawn from the unit sphere $\{\, \boldsymbol{x} \mid \|\boldsymbol{x}\|_2 = 1 \,\}$ each round. For MFEAT ($d = 217, K = 9$, $T = 10000$, $m = 195$), each round presents 9 candidate samples drawn from 9 distinct digit classes in ascending order with normalized features and a reward vector $\boldsymbol{r} = [0, 1/K, \ldots, 1]^\top$; And each vector is normalized to have unit ℓ_2 norm. We concatenate a bias term 1 to the normalized feature vector, resulting in an augmented feature dimension of $(d+1) = 785$. For synthetic experiments, we set parameter norm bound $S = 5$. For each experiment, we set seed to 42.

C.2 Additional Experiments on More Real Datasets

In addition to the MFEAT dataset, we also evaluate other seven real datasets from the OpenML repository and MNIST (LeCun et al., 1998) in both binary and multinomial cases. Dataset details are provided in Table 1. For the binary case, set the class with label 1 as the target class and the rest as the non-target class. For MNIST, at each round, the learner receives 10 candidate samples with distinct digit labels and selects one. A reward of 1 is observed if the chosen label is 1, and 0 otherwise. And we arrange the classes in ascending order and set the corresponding reward vector $\boldsymbol{r} = [0, 1/K, 2/K, \ldots, 1]$. We also concatenate a bias term to the normalized feature vector, resulting in $(d+1)$-dimensional action vectors. The rest is the same as the MFEAT experiment.

The results of these experiments are presented in Figure 2 – Figure 5.

C.3 Experiments on the Effect of Sketch Size

We show the effect of the sketch size m. The theoretical analysis suggests that a smaller sketch size may lead to a larger approximation error through the accumulated sketching error $\rho_{1:t}$, whereas a larger sketch size reduces this error at the cost of higher computational overhead. And full sketching ($m = Kd$) recovers the original OFUL-MLogB algorithm with zero sketching error.

We perform full sketching with $m = Kd = 785$ on MNIST dataset. We set seed to 42 and reinitialize environment for each run. Figure 6 shows that the regret and $\rho_{1:t}$ of EOFD-MLogB with full sketching are identical to those of OFUL-MLogB. Figure 6a and 6b show that the regret curves of

Table 2: Parameters for Datasets.

	Dataset	K	λ	η	m
Binary	Synth	1	$(\sqrt{K}d)^{0.75}\lambda_0$	$(\sqrt{K}d)^{0.75}\eta_0$	40
	MFEAT	1	$(\sqrt{K}d)^{0.75}\lambda_0$	$(\sqrt{K}d)^{0.75}\eta_0$	21
	MNIST	1	$(\sqrt{K}d)^{0.75}\lambda_0$	$(\sqrt{K}d)^{0.75}\eta_0$	78
	SPAM	1	$(\sqrt{K}d)^{0.75}\lambda_0$	$(\sqrt{K}d)^{0.75}\eta_0$	5
	DNA	1	$(\sqrt{K}d)^{0.75}\lambda_0$	$(\sqrt{K}d)^{0.75}\eta_0$	18
	php3isjYz	1	$(\sqrt{K}d)^{0.75}\lambda_0$	$(\sqrt{K}d)^{0.75}\eta_0$	10
	phpbL6t4U	1	$(\sqrt{K}d)^{0.75}\lambda_0$	$(\sqrt{K}d)^{0.75}\eta_0$	12
	phptd5jYj	1	$(\sqrt{K}d)^{0.75}\lambda_0$	$(\sqrt{K}d)^{0.75}\eta_0$	7
	phpdReP6S	1	$(\sqrt{K}d)^{0.75}\lambda_0$	$(\sqrt{K}d)^{0.75}\eta_0$	25
	INDIAN	1	$(\sqrt{K}d)^{0.75}\lambda_0$	$(\sqrt{K}d)^{0.75}\eta_0$	22
Multinomial	Synth	5	$(\sqrt{K}d)^{0.75}\lambda_0$	$(\sqrt{K}d)^{0.75}\eta_0$	40
	MFEAT	10	$(\sqrt{K}d)^{0.75}\lambda_0$	$(\sqrt{K}d)^{0.4}\eta_0$	97
	DNA	2	$(\sqrt{K}d)^{0.75}\lambda_0$	$(\sqrt{K}d)^{0.75}\eta_0$	36
	phpbL6t4U	5	$(\sqrt{K}d)^{0.75}\lambda_0$	$(\sqrt{K}d)^{0.75}\eta_0$	64

OFUL-MLogB and EOFD-MLogB with full sketching are perfectly aligned, and Figure 6c shows that $\rho_{1:t}$ remains zero throughout the experiment, confirming that full sketching recovers the original algorithm.

Since frequent directions can be interpreted as an online PCA procedure (Kuzborskij et al., 2019, Section 7), we select sketch sizes according to the fraction of variance preserved by PCA. Specifically, we choose $m = 87$, which preserves approximately 90% of the total variance of the MNIST dataset, $m = 17$ preserving about 60% of the total variance, and $m = 5$ preserving about 30% of the total variance.

The results are shown in Figure 7. The solid lines report the mean over 10 independent trials, and the shaded regions indicate one standard deviation of the mean. As shown in Figure 7a and 7b, decreasing m from 87 to 17 and 5 substantially reduces the computational time, while increasing the accumulated sketching error $\rho_{1:t}$.

C.4 Experiments on the Effect of K and d

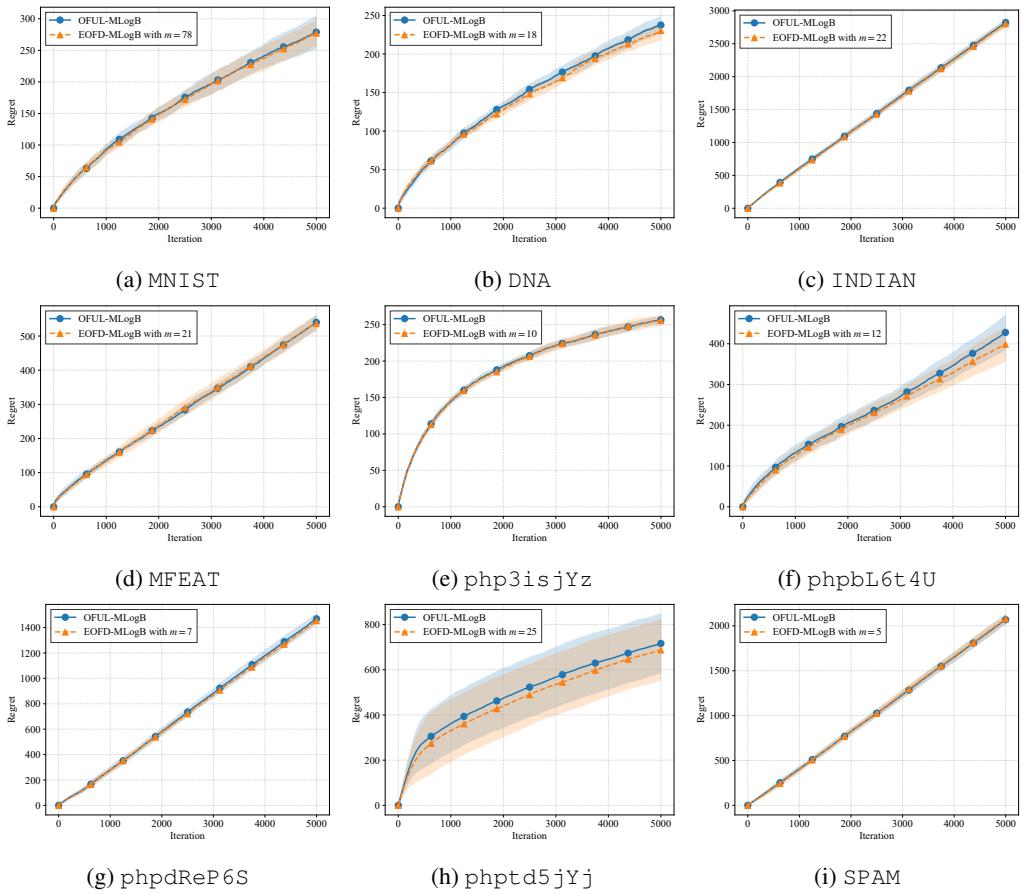
In this section we perform experiments to investigate the effect of K and d on the performance of EOFD-MLogB. For the effect of K, we vary K from 2 to 32, set $d = 50, T = 1000$, and other parameters remain untouched. For the effect of d, we vary d from 50 to 5000, set $K = 1, T = 100$, and other parameters remain untouched. Figure 8 and 9 show the results of these experiments. EOFD-MLogB exhibits a significant decrease in computational time and memory usage as K and d increase, while maintaining competitive regret performance.

C.5 κ Ablation

We investigate the sensitivity of OFUL-MLogB and EOFD-MLogB to curvature κ using a synthetic multinomial logistic bandit with $K = 4$, $d = 24$, and horizon $T = 600$. The reward vector is $\boldsymbol{r} = [0.25, 0.5, 0.75, 1]$, with zero reward for the reference outcome. We sample the entries of the ground-truth parameter W_* independently from $\mathrm{U}[-1, 1]$ and rescale it to satisfy $\|W_*\|_F = S$. At each round, the number of candidate actions is sampled uniformly from $\{10, \ldots, 30\}$, and their directions are sampled independently and isotropically, with each action normalized to satisfy $\|\boldsymbol{x}\|_2 = D$. EOFD-MLogB uses a fixed sketch size of $m = 12$.

κ is exponential large w.r.t. S and D (Amani & Thrampoulidis, 2021, Section 3), we vary the product $q = SD$ over

$$q \in \{0.25, 0.5, 1, 2, 4, 8, 16, 32\}$$



(a) MNIST (b) DNA (c) INDIAN

(d) MFEAT (e) php3isjYz (f) phpbL6t4U

(g) phpdReP6S (h) phptd5jYj (i) SPAM

Figure 2: Regret comparison of EOFD-MLogB and OFUL-MLogB on more real datasets in binary cases ($K = 1$).

through two complementary parameterizations: (i) $S = q$ and $D = 1$, and (ii) $S = 1$ and $D = q$. These settings vary the parameter norm and action norm, respectively. We plot the results against the domain-level inverse-curvature proxy

$$\widehat{\kappa}(S, D, K) := \left(1 + Ke^{SD}\right)\left(K + e^{SD}\right),$$

which equals $(1 + 4e^q)(4 + e^q)$ in our experiments. The horizontal axis uses a logarithmic scale.

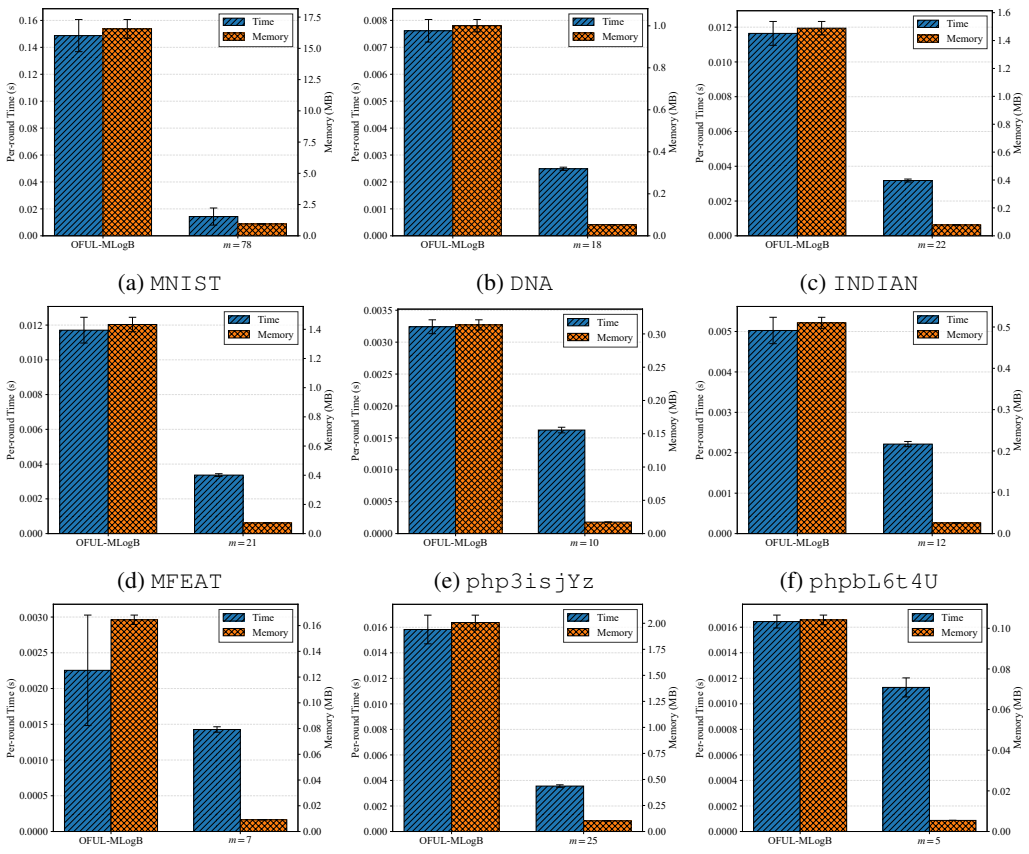
For each configuration, we set $\eta = (\sqrt{K}d)^{0.75}\eta_0, \lambda = (\sqrt{K}d)^{0.75}\eta_0, \delta = 1/T$.

Each configuration is evaluated over 10 paired stochastic repetitions, using a master random seed of 42. Within each repetition, both algorithms share the same ground-truth parameter, candidate action sets, and underlying random draws used to generate feedback. The curves report the mean cumulative regret at $T = 600$, with shaded regions indicating one standard deviation across repetitions.

The results in Figure 10 suggest that increasing $\widehat{\kappa}$ does not lead to a substantial increase in the additional regret of EOFD-MLogB relative to OFUL-MLogB. The empirical regret gap remains modest over the tested range, suggesting that the fixed-size sketch introduces limited additional sensitivity to deteriorating curvature in these experiments.

C.6 Experiments on Distributional Properties

We select two experiments of EOFD-MLogB and collect all actions $\{(W_t, \boldsymbol{x}_t) \mid t = 1, \ldots, T\}$. The resulting spectral distributions shown in Figure 11. The first experiment (Figure 11a) corresponds to setting of $K = 4, d = 50$ and $\lambda = (\sqrt{K}d)^{0.75}\lambda_0, \eta = (\sqrt{K}d)^{0.75}\eta_0$, which yields a relatively

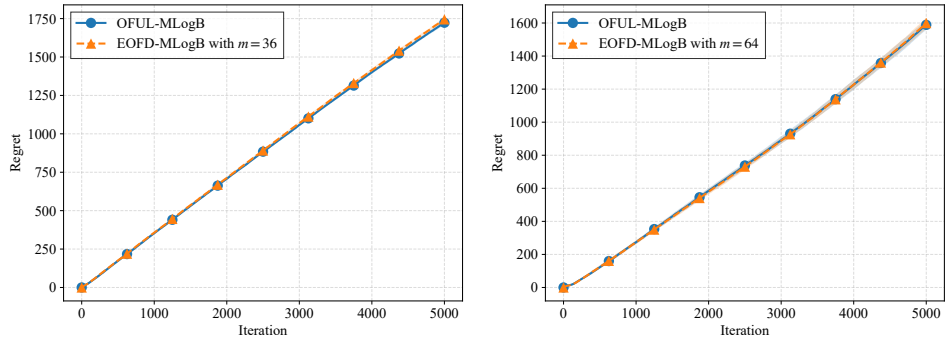


(a) MNIST (b) DNA (c) INDIAN

(d) MFEAT (e) php3isjYz (f) phpbL6t4U

(g) phpdReP6S (h) phptd5jYj (i) SPAM

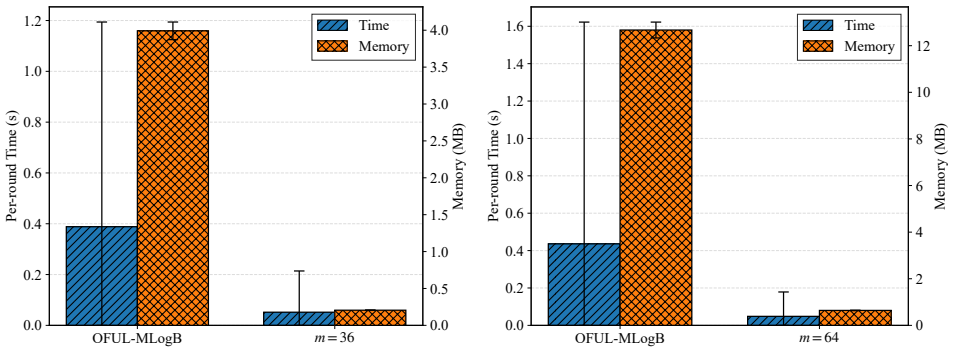
Figure 3: Time comparison of EOFD-MLogB and OFUL-MLogB on more real datasets in binary cases ($K = 1$).



(a) DNA with $K = 2$ (b) phpbL6t4U with $K = 5$

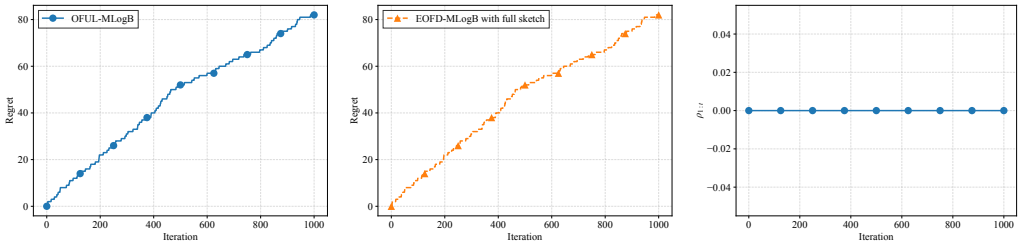
Figure 4: Regret comparison of EOFD-MLogB and OFUL-MLogB on more real datasets in multinomial cases ($K > 1$).

good W_t convergence; and the second experiment (Figure 11b) corresponds to setting of $K = 4, d = 50$ and $\lambda = (\sqrt{K}d)^{-0.5}\lambda_0, \eta = (\sqrt{K}d)^{0.5}\eta_0$, which yields a noisier W_t convergence. For every 200 iterations, we perform eigenvalue decomposition on the cumulative Hessian matrices $\sum_{s=1}^{t-1} \nabla^2 \ell_s(W_{s+1}) = \sum_{s=1}^{t-1} \nabla\boldsymbol{\sigma}(W_{s+1}\boldsymbol{x}_s) \otimes \boldsymbol{x}_s\boldsymbol{x}_s^\top$. To enhance visualization, the horizontal axis is transformed using $\log_{10}$ scaling.



(a) DNA with $K = 2$ (b) phpbL6t4U with $K = 5$

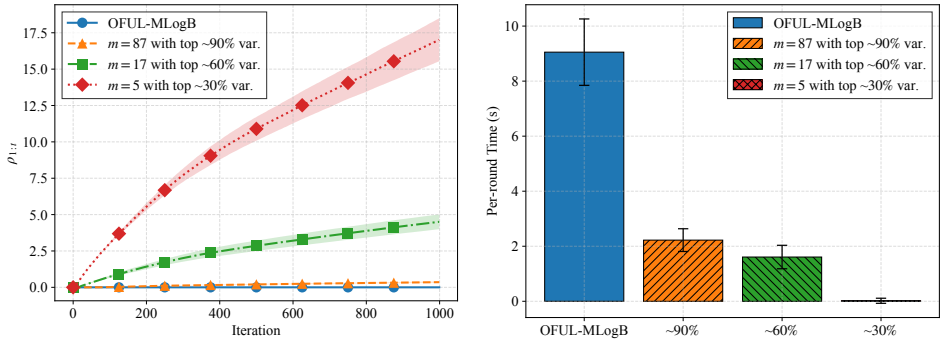
Figure 5: Time comparison of EOFD-MLogB and OFUL-MLogB on more real datasets in multinomial cases ($K > 1$).



(a) Regret of OFUL-MLogB (b) Regret of EOFD-MLogB (c) $\rho_{1:t}$

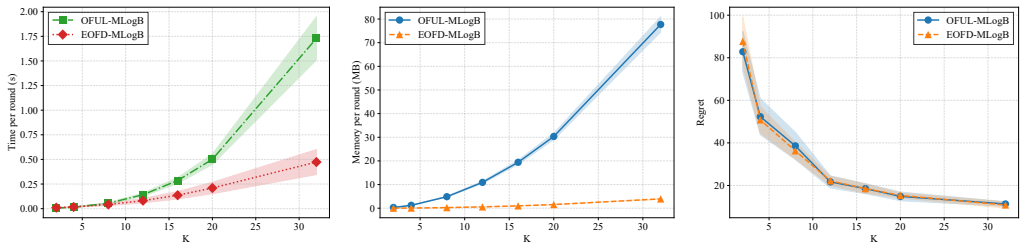
Figure 6: Full sketching ($m = Kd$) recovers the original OFUL-MLogB algorithm.

As shown in the results, the spectral distribution appears relatively uniform during the initial phase ($t = 200$) of the experiment. As iterations progress ($t = 400 - 600$), the distribution becomes increasingly non-uniform, with most eigenvalues concentrating around specific values. In later stages ($t = 800 - 1000$), the spectral distribution exhibits a centring characteristic. Comparing Figures 11a and 11b, we observe that, the ideal case (Figure 11a) demonstrates clear centring distribution properties around interval $[0, 5]$; and the noisier case (Figure 11b), does not exhibit a distinct centring pattern, with most eigenvalues distributed in interval $[0, 200]$ in a relatively uniform manner.



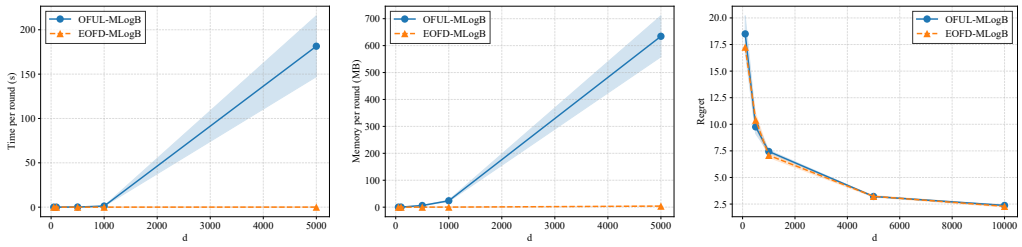
(a) $\rho_{1:t}$ of different sketch sizes (b) time of different sketch sizes

Figure 7: The effect of sketch size m on the sketching error and computational time of EOFD-MLogB.



(a) Time of different K (b) Memory of different K (c) Regret of different K

Figure 8: Effect of different K.



(a) Time of different d (b) Memory of different d (c) Regret of different d

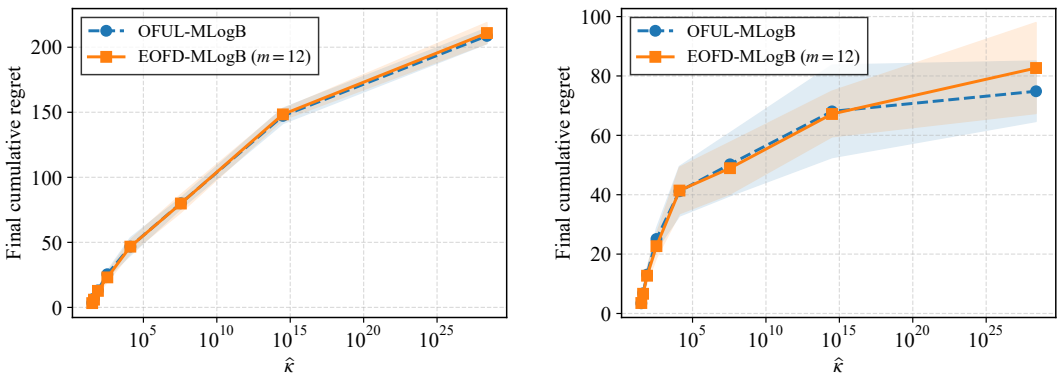
Figure 9: Effect of different d.

C.7 Choice of m

Throughout the experiments, we set the default sketch size to $m = \lfloor Kd/10 \rfloor$, a fixed fraction of the feature dimension following prior empirical practice (Yang et al., 2025). This choice depends only on K and d and does not use information from the evaluation data, thus avoiding train-test leakage. Our results show that regret is relatively insensitive to m over the range considered: even small sketches perform well while substantially reducing computational cost.

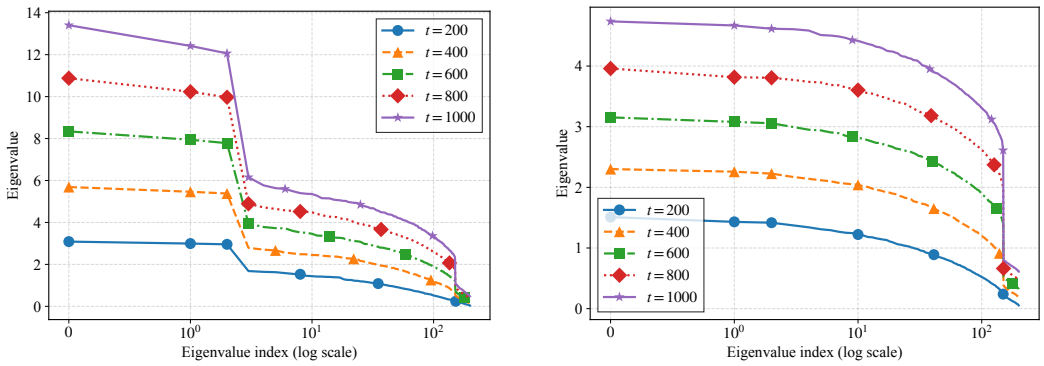
For applications with a known data distribution or independent historical data, the sketch size can also be chosen based on the spectrum of the data. Since FD/RFD can be viewed as an online PCA procedure (Kuzborskij et al., 2019, Section 7), one practical heuristic is to choose m so that the leading components explain a substantial fraction of the variance, such as 90%.

Our focus is the memory-constrained setting, where a fixed sketch size bounds the algorithm's memory usage. Adaptive sketch sizing has been studied in stochastic linear bandits (Wen et al., 2026).



(a) Scaling D (b) Scaling S

Figure 10: κ ablation.



(a) Ideal spectrum

(b) Noisier spectrum

Figure 11: The spectrum of the selected decisions: the horizontal axis is transformed using $\log_{10}$ scaling to enhance visualization

When memory is not a binding constraint, another possible direction is to maintain multiple sketches and select among different values of m using a meta-algorithm inspired by BOB (Cheung et al., 2019) or CORRAL (Agarwal et al., 2017). Extending these ideas to our setting and analyzing their guarantees is left for future work.