GL-LowPopArt: A Nearly Instance-Wise Minimax-Optimal Estimator for Generalized Low-Rank Trace Regression

Junghyun Lee KAIST AI
Kyoungseok Jang CAU AI
Kwang-Sung Jun POSTECH CSE/AI
Milan Vojnović LSE Statistics
Se-Young Yun KAIST AI

Abstract

We present GL-LowPopArt, a novel Catoni-style estimator for generalized low-rank trace regression. Building on LowPopArt (Jang et al., 2024), it employs a two-stage approach: nuclear norm regularization followed by matrix Catoni estimation. We establish state-of-the-art estimation error bounds, surpassing existing guarantees (Fan et al., 2019; Kang et al., 2022), and reveal a novel experimental design objective, GL(π). The key technical challenge is controlling bias from the nonlinear inverse link function, which we address with our two-stage approach. We prove a *local minimax lower bound*, showing that our GL-LowPopArt enjoys instance-wise optimality up to the condition number of the ground-truth Hessian. Our method immediately achieves an improved Frobenius error guarantee for generalized linear matrix completion. We also introduce a new problem setting called **bilinear dueling bandits**, a contextualized version of dueling bandits with a general preference model. Using an explore-then-commit approach with GL-LowPopArt, we show an improved Borda regret bound over naïve vectorization (Wu et al., 2024).

1 INTRODUCTION

Low-rank structures are ubiquitous across diverse domains, where the estimation of high-dimensional, low-rank matrices frequently pops up (Chen and Chi, 2018). Beyond simply possessing a low-rank structure, real-world observations are often subject to nonlinearities. One ubiquitous example is modeling discrete event occurrences by the Poisson point processes (Kingman, 1992; Mutný and Krause, 2021), such as crime rate (Shirota and Gelfand, 2017) and environmental modeling (Heikkinen and Arjas, 1999). In news recommendation and online ad placement, outputs are often quantized, representing categories such as "click" or "no click" (Bennett and Lanning, 2007; Li et al., 2010, 2012; McMahan et al., 2013; Richardson et al., 2007; Stern et al., 2009). Other applications involve predicting interactions between multiple features, e.g., hotel-flight bundles (Lu et al., 2021), online dating/shopping (Jun et al., 2019), protein-drug pair searching (Luo et al., 2017), graph link prediction (Berthet and Baldin, 2020), stock return prediction (Fan et al., 2019), and, more recently, general preference learning (Zhang et al., 2025), among others. In these settings, it is natural to model the problem as matrix-valued covariates passed through a nonlinear regression model. In particular, when the observations are sampled from the generalized linear model (McCullagh and Nelder, 1989), these diverse problems fall under the umbrella of *generalized low-rank trace regression* (Fan et al., 2019), which we describe.

Problem Setting. $\boldsymbol{\Theta}_\star \in \mathbb{R}^{d_1 \times d_2}$ is an unknown matrix of rank at most $r \ll d_1 \wedge d_2$, and $\mathcal{X} \subseteq \mathbb{R}^{d_1 \times d_2}$ is an arm-set (e.g., sensing matrices). The learner's goal is to output $\widehat{\boldsymbol{\Theta}}$ that well-estimates $\boldsymbol{\Theta}_\star$ from some observations $\{(\boldsymbol{X}_t, y_t)\}_{t \in [N]}$, collected as follows.

For a given budget $N \in \mathbb{N}$, a *sampling policy (design)* is a sequence $\boldsymbol{\pi}_N = (\pi_t)_{t \in [N]}$, where each $\pi_t \in \mathcal{P}(\mathcal{X})$ is potentially history-dependent; see Lattimore and Szepesvári (2020, Chapter 4.6) for the precise measure-theoretic formulation. When the learner uses $\boldsymbol{\pi}_N$, at each time $t \in [N]$, she samples a $\boldsymbol{X}_t \sim \pi_t$ and observes y_t sampled from *generalized linear model (GLM)* whose (conditional) density is given as follows:

$$p(y_t | \boldsymbol{X}_t; \boldsymbol{\Theta}_\star) \propto \exp\left(\frac{y_t \langle \boldsymbol{X}_t, \boldsymbol{\Theta}_\star \rangle - m(\langle \boldsymbol{X}_t, \boldsymbol{\Theta}_\star \rangle)}{g(\tau)} \right).$$

Here, $m : \mathbb{R} \to \mathbb{R}$ is the log-partition function, τ is the dispersion parameter, $g : \mathbb{R} \to \mathbb{R}_{>0}$ is a fixed function, and the density is with respect to some known

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

base measure (e.g., Lebesgue, counting). We refer to $\mu := \dot{m}$ as the *inverse link function*. We assume that all components of the GLM, other than $\mathbf{\Theta}_\star$, are known to the learner; see Appendix N(**5**) for discussions on what happens under model misspecification.

Given $\boldsymbol{\pi}_N = (\pi_t)_{t\in[N]}$, we distinguish two setups: in the *adaptive scenario*, each $\pi_t \in \mathcal{P}(\mathcal{X})$ may depend on past observations, as is standard in interactive learning problems such as bandits (Lattimore and Szepesvári, 2020) and active learning (Settles, 2012), whereas in the *nonadaptive (passive) scenario*, all policies are identical, i.e., $\pi_t = \pi$ for a known $\pi \in \mathcal{P}(\mathcal{X})$, fixed before interaction begins. In our setting, we omit t dependence for brevity, since in the adaptive case our algorithm only switches policy once (π_0 for Stage I and π for Stage II). We also remark in advance that our lower bound holds for any adaptive policy $\boldsymbol{\pi}_N$.

Related Works. Owing to its ubiquity, much work have been done in providing statistically and computationally efficient estimators for this problem, both generally (Fan et al., 2019; Kang et al., 2022) and in specific scenarios such as *generalized linear matrix completion* (Cai and Zhou, 2016, 2013; Davenport et al., 2014; Klopp, 2014; Klopp et al., 2015; Lafond, 2015; Lafond et al., 2014) and learning low-rank preference matrix (Rajkumar and Agarwal, 2016). Corresponding minimax lower bounds have also been proven that are tight with respect to rank r, dimension d_1, d_2, and sample size N; see Appendix A for a detailed survey.

Main Contributions. While prior work has made significant progress, a crucial aspect has been overlooked: the instance-specific nature of curvature. To our knowledge, all existing analyses rely on worst-case curvature bounds, obscuring the problem's true difficulty. For example, known performance guarantees for generalized linear matrix completion depend inversely with $\min_{|z|\le\gamma} \dot{\mu}(z)$, where $\gamma > 0$ is such that $\max_{i,j} |(\mathbf{\Theta}_\star)_{ij}| \le \gamma$ and $\dot{\mu}$ is the derivative of the inverse link function (Davenport et al., 2014; Klopp et al., 2015). For instance, when $\mu(z) = (1+e^{-z})^{-1}$, this leads to a dependence of e^γ (Faury et al., 2020). This is *not* instance-wise guarantee in the sense that it arises from the worst-case $\dot{\mu}$ over the entry-wise domain $[-\gamma, \gamma]$, rather than adapting to the specific instance $\Theta_\star$.

Our contributions are as follows:

- We propose `GL-LowPopArt`, an extension of `LowPopArt` (Jang et al., 2024) to generalized low-rank trace regression, that consists of 1) Stage I computes an initial estimate $\mathbf{\Theta}_0$, and 2) Stage II refines it via matrix Catoni estimator (Minsker, 2018). We prove its ***instance-wise error rate*** for any design $\pi \in \mathcal{P}(\mathcal{X})$ (Theorem 3.1):
$$\left\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}_\star\right\|_{\mathrm{op}}^2 \lesssim_{\log} \frac{\mathrm{GL}(\pi;\mathbf{\Theta}_0)}{N} \lesssim \frac{d_1 \vee d_2}{N\lambda_{\min}(\boldsymbol{H}(\pi;\mathbf{\Theta}_\star))},$$
where $\mathrm{GL}(\pi)$ (Definition 3.1) is a new experimental design objective that captures the nonlinearity and the arm-set geometry, and $\lambda_{\min}(\boldsymbol{H}(\pi;\mathbf{\Theta}_\star))$ is the minimum eigenvalue of the Hessian of the negative log-likelihood loss at $\mathbf{\Theta}_\star$. The proof requires careful bias control of one-sample estimators during matrix Catoni estimation (Minsker, 2018). In the adaptive scenario, one can directly optimize the error bound as $\min_\pi \mathrm{GL}(\pi;\mathbf{\Theta}_0)$. (Section 3)
- We prove the ***first instance-wise minimax lower bound*** for generalized low-rank trace regression (Theorem 4.1): for *any* $\boldsymbol{\pi}_N$ and instance $\mathbf{\Theta}_\star$, there is a $\widetilde{\mathbf{\Theta}}_\star$ *near* $\mathbf{\Theta}_\star$ such that
$$\left\|\widehat{\mathbf{\Theta}} - \widetilde{\mathbf{\Theta}}_\star\right\|_F^2 \gtrsim \frac{r(d_1 \vee d_2)}{N\lambda_{\max}(\boldsymbol{H}(\boldsymbol{\pi}_N;\mathbf{\Theta}_\star))},$$
where $\lambda_{\max}(\cdot)$ is the maximum eigenvalue. When $\boldsymbol{\pi}_N = (\pi_t)_{t\in[N]}$ is of the form $\pi_t = \pi_0 \mathbb{1}[t \le N_1] + \pi \mathbb{1}[t > N_1]$ with $N_1 = o(N)$, `GL-LowPopArt` is nearly instance-wise optimal, up to the condition number, $\frac{\lambda_{\max}(\boldsymbol{H}(\pi;\mathbf{\Theta}_\star))}{\lambda_{\min}(\boldsymbol{H}(\pi;\mathbf{\Theta}_\star))}$. (Section 4)
- As the first application of `GL-LowPopArt`, we revisit the problem of *generalized linear matrix completion* (Davenport et al., 2014; Klopp et al., 2015; Lafond, 2015) and prove an error rate with instance-specific curvature $(\min_{i,j} \dot{\mu}((\mathbf{\Theta}_\star)_{i,j}))^{-1}$. This improves upon prior results that depend on the worst-case curvature. (Section 5.1)
- As another application, we propose and tackle ***bilinear dueling bandits***, a new variant of generalized linear dueling bandits with the contextual bilinear preference model of Zhang et al. (2025). We propose a `GL-LowPopArt`-based explore-then-commit algorithm and prove its Borda regret upper bound (Theorem 5.1):
$$\mathrm{Reg}^B(T) \lesssim_{\log} (\mathrm{GL}_{\min}(\mathcal{X}))^{1/3} \left(\kappa_\star^B T\right)^{2/3},$$
where $\kappa_\star^B$ is a new curvature-dependent quantity specific to each bandit instance. (Section 5.2)

2 TECHNICAL PRELIMINARIES

Notation. For a $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ with singular values $\sigma_1 \ge \cdots \ge \sigma_{\min\{m,n\}}$, $\|\boldsymbol{A}\|_{\mathrm{nuc}} := \sum_{i=1}^{\min\{m,n\}} \sigma_i$ is its nuclear norm, and $\|\boldsymbol{A}\|_{\mathrm{op}} := \sigma_1$ is its operator (spectral) norm. For $\boldsymbol{B} \in \mathbb{R}^{m\times n}$, their Frobenius inner product is defined as $\langle \boldsymbol{A}, \boldsymbol{B}\rangle := \mathrm{tr}(\boldsymbol{A}^\top \boldsymbol{B})$. For a symmetric $\boldsymbol{A} \in$

$\mathbb{R}^{m\times m}$, $\lambda_i(\boldsymbol{A})$ is its i-th largest eigenvalue, $\lambda_{\max} := \lambda_1$, and $\lambda_{\min} := \lambda_m$. On the positive semidefinite cone, define the Loewner order $\preceq$ as $\boldsymbol{A} \preceq \boldsymbol{B}$ if and only if $\boldsymbol{B} - \boldsymbol{A}$ is positive semidefinite. For a $S > 0$, let us denote $\mathcal{B}_i^{d_1\times d_2}(S) := \{\boldsymbol{X} \in \mathbb{R}^{d_1\times d_2} : \|\boldsymbol{X}\|_i \le S\}$ for $i \in \{\mathrm{op}, \mathrm{nuc}, F\}$. $\mathrm{vec} : \mathbb{R}^{d_1\times d_2} \to \mathbb{R}^{d_1 d_2}$ performs column-wise stacking of a matrix into a vector, and vec^{-1} is its inverse. $f(n) \lesssim g(n)$ and $f(n) \asymp g(n)$ indicates $f(n) \le cg(n)$ and $cg(n) \le f(n) \le c'g(n)$ for some constants $c, c' > 0$, respectively; log in the subscript (e.g., $\lesssim_{\log}$) indicates up to logarithmic factors. Denote $a \wedge b := \min(a, b)$ and $a \vee b := \max(a, b)$. For a $n \in \mathbb{N}$, let $[n] := \{1, 2, \dots, n\}$. For a set X, $\mathcal{P}(X)$ is the set of all probability distributions on X.

Assumptions. We assume the following for the parameter space Ω:

Assumption 1. *$\Omega \subseteq \mathbb{R}^{d_1\times d_2}$ is a linear subspace satisfying $\boldsymbol{\Theta} \in \Omega \Longrightarrow \mathrm{Proj}_r(\boldsymbol{\Theta}) \in \Omega$, where $\mathrm{Proj}_r(\boldsymbol{\Theta}) := \arg\min_{\mathrm{rank}(\boldsymbol{\Theta}')\le r} \|\boldsymbol{\Theta} - \boldsymbol{\Theta}'\|_F^2$.*

Some notable examples include $\mathbb{R}^{d_1\times d_2}$ and the set of (skew-)symmetric matrices when $d_1 = d_2$.

We assume the following for the arm-set $\mathcal{X}$:

Assumption 2. $\mathrm{span}(\mathcal{X}) = \mathbb{R}^{d_1\times d_2}$.

This is essential as otherwise one cannot hope to recover $\boldsymbol{\Theta}_\star$ in the direction of $\mathrm{span}(\mathcal{X})^\perp \neq \emptyset$.

We assume the following for $m(\cdot)$:

Assumption 3. *$m : \mathbb{R} \to \mathbb{R}$ is three-times differentiable and convex. Moreover, the inverse link function $\mu := \dot{m}$ satisfies the following two conditions:*

(a) R_s-self-concordant for a known $R_s \in [0, \infty)$, i.e., $|\ddot{\mu}(z)| \le R_s \dot{\mu}(z)$, $\forall z \in \mathbb{R}$,

(b) $L_\mu := \sup_{\boldsymbol{X}\in\mathcal{X}, \boldsymbol{\Theta}\in\Omega} \dot{\mu}(\langle \boldsymbol{X}, \boldsymbol{\Theta}\rangle) < \infty$.

This encompasses Gaussian ($m(z) = \frac{1}{2}z^2$), Bernoulli ($m(z) = \log(1 + e^z)$), and Poisson ($m(z) = e^z$).

3 GL-LowPopArt: A MATRIX CATONI-STYLE ESTIMATOR

For $\pi \in \mathcal{P}(\mathcal{X})$ and $\boldsymbol{\Theta} \in \mathbb{R}^{d_1\times d_2}$, we define the *vectorized design/Hessian matrix*:

$$\boldsymbol{V}(\pi) := \mathbb{E}_{\boldsymbol{X}\sim\pi}[\mathrm{vec}(\boldsymbol{X})\mathrm{vec}(\boldsymbol{X})^\top],$$
$$\boldsymbol{H}(\pi; \boldsymbol{\Theta}) := \mathbb{E}_{\boldsymbol{X}\sim\pi}[\dot{\mu}(\langle \boldsymbol{X}, \boldsymbol{\Theta}\rangle)\mathrm{vec}(\boldsymbol{X})\mathrm{vec}(\boldsymbol{X})^\top],$$

where $\boldsymbol{H}(\pi; \boldsymbol{\Theta})$ is the Hessian of the population negative log-likelihood: $\boldsymbol{\Theta} \mapsto -g(\tau)\mathbb{E}_{\boldsymbol{X}\sim\pi}[\log p(y|\boldsymbol{X}; \boldsymbol{\Theta})]$. Observe that $\boldsymbol{V}(\pi) = \boldsymbol{H}(\pi; \boldsymbol{\Theta})$ when $\mu(z) = z$. We will also define an instance-specific quantity:

$$\kappa_\star := \min_{\boldsymbol{X}\in\mathcal{X}} \dot{\mu}(\langle \boldsymbol{X}, \boldsymbol{\Theta}_\star\rangle) > 0.$$

We now recall some notions from matrix analysis to describe the matrix Catoni estimator (Catoni, 2012; Minsker, 2018). For any $f : \mathbb{R} \to \mathbb{R}$ and symmetric $\boldsymbol{M} \in \mathbb{R}^{d\times d}$, we define $f(\boldsymbol{M})$ as $f(\boldsymbol{M}) := \boldsymbol{U}\mathrm{diag}(\{f(\lambda_i)\}_{i\in[d]})\boldsymbol{U}^\top$, where $\boldsymbol{M} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$ with $\boldsymbol{\Lambda} = \mathrm{diag}(\{\lambda_i\}_{i\in[d]})$ being the eigenvalue decomposition of $\boldsymbol{M}$, i.e., f acts on its spectrum. The *Hermitian dilation* (Tropp, 2015), $\mathcal{H} : \mathbb{R}^{d_1\times d_2} \to \mathbb{R}^{(d_1+d_2)\times(d_1+d_2)}$, is defined as

$$\mathcal{H}(\boldsymbol{A}) := \begin{bmatrix} \boldsymbol{0}_{d_1\times d_1} & \boldsymbol{A} \\ \boldsymbol{A}^\top & \boldsymbol{0}_{d_2\times d_2} \end{bmatrix}.$$

The *influence function* (Catoni, 2012) is defined as

$$\psi(x) := \begin{cases} \log(1 + x + x^2/2), & x \ge 0, \\ -\log(1 - x + x^2/2), & x < 0. \end{cases}$$

We then define $\tilde{\psi}_\nu(\boldsymbol{A}) := \frac{1}{\nu}\psi(\nu\mathcal{H}(\boldsymbol{A}))_{\mathrm{ht}}$ for $\nu > 0$, where for $\boldsymbol{M} \in \mathbb{R}^{(d_1+d_2)\times(d_1+d_2)}$, we define its *horizontal truncation* as $\boldsymbol{M}_{\mathrm{ht}} := \boldsymbol{M}_{1:d_1, d_1+1:d_1+d_2}$.

3.1 Overview of GL-LowPopArt

We present GL-LowPopArt (Generalized Linear LOWrank POPulation covariance regression with hARd Thresholding; Algorithm 1), a novel estimator for generalized low-rank trace regression. GL-LowPopArt consists of two stages: the first stage provides a rough, initial estimate, and the second stage refines it via matrix Catoni estimator (Minsker, 2018). It takes two designs π_0 and π as inputs for Stage I and II, respectively. When the learner is in the adaptive learning scenario, she can choose π dependent on the data collected during Stage I. If not, she can set $\pi_0 = \pi$.

Stage I uses the observations $\{(\boldsymbol{X}_t, y_t)\}_{t=1}^{N_1}$ collected via π_0 to compute a "sufficiently good" initial estimate $\boldsymbol{\Theta}_0$. In Stage II, for each sample $(\boldsymbol{X}_t, y_t)$ colllected via another design π for $t = N_1 + 1, \cdots, N_1 + N_2$, GL-LowPopArt constructs one-sample estimator $\widetilde{\boldsymbol{\Theta}}_t$ such that $\mathbb{E}[\widetilde{\boldsymbol{\Theta}}_t] \approx \boldsymbol{\Theta}_\star - \boldsymbol{\Theta}_0$ (line 8). Then, the Ω-projected matrix Catoni estimator $\boldsymbol{\Theta}_1$ is computed (line 9). The final estimator $\widehat{\boldsymbol{\Theta}}$ is obtained by singular value thresholding $\boldsymbol{\Theta}_1$ (line 10). Note that the algorithm is parameter-free in the sense that the algorithm does not require the knowledge of the rank r of $\boldsymbol{\Theta}_\star$.

The final estimation error guarantee is solely due to matrix Catoni estimation (Minsker, 2018) in Stage II, yet unlike the linear trace regression (Jang et al., 2024), we *require* for the initial estimate $\boldsymbol{\Theta}_0$ to be asymptotically

Algorithm 1: `GL-LowPopArt`

Input: Sample sizes (N_1, N_2), designs $\pi_0 \in \mathcal{P}(\mathcal{X})$ for Stage I, Regularization coefficient $\lambda_{N_1} > 0$;
`/* Stage I: Nuclear Norm-regularized Initial Estimator */`
for $t = 1, 2, \cdots, N_1$ **do**
  Pull $\boldsymbol{X}_t \sim \pi_0$ and receive $y_t \sim p(\cdot|\boldsymbol{X}_t; \boldsymbol{\Theta}_\star)$;
Compute an initial estimator $\boldsymbol{\Theta}_0$ using $\{(\boldsymbol{X}_t, y_t)\}_{t=1}^{N_1}$ collected via π_0;
If in the adaptive scenario, compute π for Stage II; else, set $\pi = \pi_0$;
`/* Stage II: Generalized Linear Matrix Catoni Estimation */`
for $t = N_1 + 1, N_1 + 2, \cdots, N_1 + N_2$ **do**
  Pull $\boldsymbol{X}_t \sim \pi$ and receive $y_t \sim p(\cdot|\boldsymbol{X}_t; \boldsymbol{\Theta}_\star)$;
  Compute the matrix one-sample estimators:

$$\widetilde{\boldsymbol{\Theta}}_t \leftarrow \mathrm{vec}^{-1}\left(\tilde{\boldsymbol{\theta}}_t\right), \quad \tilde{\boldsymbol{\theta}}_t \leftarrow \boldsymbol{H}(\pi; \boldsymbol{\Theta}_0)^{-1}\left(y_t - \mu(\langle \boldsymbol{X}_t, \boldsymbol{\Theta}_0\rangle)\right)\mathrm{vec}(\boldsymbol{X}_t)$$

$\boldsymbol{\Theta}_1 \leftarrow \mathrm{Proj}_\Omega\left(\boldsymbol{\Theta}_0 + \frac{1}{N_2}\left(\sum_{t=N_1+1}^{N_1+N_2} \tilde{\psi}_\nu(\widetilde{\boldsymbol{\Theta}}_t)\right)_{\mathrm{ht}}\right)$ with $\nu := \sqrt{\frac{2}{5g(\tau)\mathrm{GL}(\pi;\boldsymbol{\Theta}_0)N_2}\log\frac{4(d_1+d_2)}{\delta}}$;
Let $\boldsymbol{\Theta}_1 = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^\top$ be its SVD and $\widetilde{\boldsymbol{D}}$ be $\boldsymbol{D}$ after zeroing out singular values at most $\sqrt{40}g(\tau)\mathrm{GL}(\pi;\boldsymbol{\Theta}_0)\nu$;
Return: $\widehat{\boldsymbol{\Theta}} := \boldsymbol{U}\widetilde{\boldsymbol{D}}\boldsymbol{V}^\top$;

consistent. This is the main technical novelty/challenge for the algorithm design and analysis.

Before presenting the error rate of `GL-LowPopArt`, let us first define a new design objective for Stage II:

Definition 3.1 (GL-Design)**.** *For any $\pi \in \mathcal{P}(\mathcal{X})$ and $\boldsymbol{\Theta}_\star \in \mathbb{R}^{d_1\times d_2}$, the **GL-Design objective** $\mathrm{GL}(\pi; \boldsymbol{\Theta}_\star)$ is defined as follows:*

$$\mathrm{GL}(\pi; \boldsymbol{\Theta}_\star) := \max\{H^{(\mathrm{row})}(\pi; \boldsymbol{\Theta}_\star), H^{(\mathrm{col})}(\pi; \boldsymbol{\Theta}_\star)\},$$

where

$$H^{(\mathrm{row})}(\pi; \boldsymbol{\Theta}_\star) := \lambda_{\max}\left(\sum_{m=1}^{d_1} \boldsymbol{D}_m{}^{(\mathrm{row})}(\pi; \boldsymbol{\Theta}_\star)\right),$$

$$\boldsymbol{D}_m{}^{(\mathrm{row})}(\pi; \boldsymbol{\Theta}_\star) := [(\boldsymbol{H}(\pi; \boldsymbol{\Theta}_\star)^{-1})_{jk}]_{j,k\in\mathcal{I}_m{}^{(\mathrm{row})}},$$

$$H^{(\mathrm{col})}(\pi; \boldsymbol{\Theta}_\star) := \lambda_{\max}\left(\sum_{m=1}^{d_2} \boldsymbol{D}_m{}^{(\mathrm{col})}(\pi; \boldsymbol{\Theta}_\star)\right),$$

$$\boldsymbol{D}_m{}^{(\mathrm{col})}(\pi; \boldsymbol{\Theta}_\star) := (\boldsymbol{H}(\pi; \boldsymbol{\Theta}_\star)^{-1})_{\mathcal{I}_m{}^{(\mathrm{col})},\mathcal{I}_m{}^{(\mathrm{col})}},$$

and the index sets defined as $\mathcal{I}_m{}^{(\mathrm{row})} := \{d_1(\ell-1) + m : \ell \in [d_2]\}$ and $\mathcal{I}_m^{(\mathrm{col})} := [d_1(m-1)+1 : d_1 m]$.[a]

[a] A nice illustration of $\boldsymbol{D}_m^{(\mathrm{row})}$ and $\boldsymbol{D}_m^{(\mathrm{col})}$ is provided in Figure 1 of Jang et al. (2024).

We now present the error rate of `GL-LowPopArt`, which holds for *any* π_0, π, adaptive or nonadaptive:

Theorem 3.1 (Error Rate of `GL-LowPopArt`)**.** *Let $\delta \in (0,1)$, $\mathcal{X} \subseteq \mathcal{B}_{\mathrm{op}}^{d_1\times d_2}(1)$,*[a] *and denote*

$$\mathcal{E}(N_2, \delta) := \sqrt{\frac{g(\tau)\mathrm{GL}(\pi; \boldsymbol{\Theta}_\star)}{N_2}\log\frac{d_1+d_2}{\delta}}.$$

The final estimator $\widehat{\boldsymbol{\Theta}}$ from Stage II satisfies the following w.p. at least $1-\delta$: $\mathrm{rank}(\widehat{\boldsymbol{\Theta}}) \le r$ and

$$\left\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}_\star\right\|_{\mathrm{op}} \lesssim \mathcal{E}(N_2, \delta),$$

***provided** that $\boldsymbol{\Theta}_0$ from Stage I satisfies $\|\boldsymbol{\Theta}_\star - \boldsymbol{\Theta}_0\|_{\mathrm{nuc}}^2 \lesssim \kappa_\star g(\tau)$, and $R_s\|\boldsymbol{\Theta}_\star - \boldsymbol{\Theta}_0\|_{\mathrm{nuc}}^2 \lesssim \lambda_{\min}(\boldsymbol{H}(\pi; \boldsymbol{\Theta}_0))\mathcal{E}(N_2, \delta)$ w.p. at least $1-\delta/2$.*

[a] This has been considered before in low-rank and bilinear bandits (Jang et al., 2024; Jun et al., 2019).

3.2 $\mathrm{GL}(\pi; \boldsymbol{\Theta}_\star)$: A New Experimental Design

GL-Design Objective. $\mathrm{GL}(\pi; \boldsymbol{\Theta}_\star)$ captures two problem-specific characteristics: nonlinearity due to μ via the Hessian $\boldsymbol{H}(\pi; \boldsymbol{\Theta}_\star)$ and the geometry of $\mathcal{X}$. When $\mu(z) = z$, it reduces to the linear design objective (Jang et al., 2024, Theorem 3.4).

The following proposition shows that the GL-Design objective is closely related to the Hessian $\boldsymbol{H}(\pi; \boldsymbol{\Theta}_\star)$:

Proposition 3.1. *For any $\pi \in \mathcal{P}(\mathcal{X})$,*

$$\frac{d_1 d_2(d_1 \vee d_2)}{\mathrm{tr}(\boldsymbol{H}(\pi; \boldsymbol{\Theta}_\star))} \le \mathrm{GL}(\pi; \boldsymbol{\Theta}_\star) \le \frac{d_1 \vee d_2}{\lambda_{\min}(\boldsymbol{H}(\pi; \boldsymbol{\Theta}_\star))}.$$

Table 1: Comparison of squared Frobenius error rates under our Assumptions 1 to 3, and $\pi_0 = \pi$. We denote $\mathcal{E}(N) = \frac{r(d_1 \vee d_2)^2}{N}$ for brevity. We also assume that other assumptions required by the prior works (Fan et al., 2019; Kang et al., 2022) hold, additional to our assumptions, as long as they are not conflicting. For our `GL-LowPopArt`, we employ optimal GL-design in Stage II.

Estimator	$\mathcal{X} = \mathcal{B}_F^{d_1 \times d_2}(1)$	$\mathcal{X} = \mathcal{B}_{\text{op}}^{d_1 \times d_2}(1)$	$\mathcal{X} = \mathcal{M}$
Nuclear Penalized MLE* (Fan et al., 2019, Theorem 2)	$\kappa_\star^{-2} d_1 d_2 (d_1 \wedge d_2)\mathcal{E}(N)$	$\kappa_\star^{-2}(d_1 \wedge d_2)\mathcal{E}(N)$	$\kappa_\star^{-2} d_1 d_2 (d_1 \wedge d_2)\mathcal{E}(N)$
Stein Estimator† (Kang et al., 2022, Theorem 4.1)	$\kappa_\star^{-2}(d_1 \wedge d_2)\mathcal{E}(N)$	$\kappa_\star^{-2}(d_1 \vee d_2)\mathcal{E}(N)$	N/A
`GL-LowPopArt` **(Ours, Theorem 3.1)**	$\kappa_\star^{-1}(d_1 \wedge d_2)\mathcal{E}(N)$	$\kappa_\star^{-1}\mathcal{E}(N)$	$\text{Harm}(\boldsymbol{\Theta}_\star)^{-1}(d_1 \wedge d_2)\mathcal{E}(N)$

* We choose $\pi = \arg\max_{\pi \in \mathcal{P}(\mathcal{X})} \lambda_{\min}(\boldsymbol{V}(\pi))$ and compute the error bounds via Jang et al. (2024, Appendix D). Note that this E-optimal design is the best one could do with the given information.

† This requires π to have a continuously differentiable density over $\mathcal{X}$, and thus, not applicable to $\mathcal{M}$.

Optimal GL-Design. In the adaptive setting, the learner may choose π to minimize the GL-Design objective for Stage II, which would then minimize the error rate as in Theorem 3.1. This is in essence a optimal design problem (Pukelsheim, 2006), in which the learner chooses $\pi_{\text{GL}} \in \arg\min_{\pi \in \mathcal{P}(\mathcal{X})} \text{GL}(\pi; \boldsymbol{\Theta}_\star)$. Let $\text{GL}_{\min}(\mathcal{X}) := \text{GL}(\pi_{\text{GL}}; \boldsymbol{\Theta}_\star)$ be the optimal value.

Two practical concerns arise. First, $\boldsymbol{\Theta}_\star$ is unknown, so the ideal problem is not directly solvable. In practice, one can instead solve $\arg\min_{\pi \in \mathcal{P}(\mathcal{X})} \text{GL}(\pi; \boldsymbol{\Theta}_0)$. Given that $\boldsymbol{\Theta}_0$ is sufficiently close to $\boldsymbol{\Theta}_\star$, by self-concordance and the fact that $\mathcal{X} \subseteq \mathcal{B}_{\text{op}}^{d_1 \times d_2}(1)$, we have that $\boldsymbol{H}(\pi; \boldsymbol{\Theta}_0) \approx \boldsymbol{H}(\pi; \boldsymbol{\Theta}_\star)$ (see our Lemma C.4). This then ensures that $\text{GL}_{\min}(\mathcal{X})$ is retained up to a constant factor, i.e., the optimal GL-Design w.r.t. $\boldsymbol{\Theta}_0$ still captures the true problem characteristics.

The second concern is regarding the computational complexity of solving the optimal GL-design. This can indeed be efficiently solved, as $\pi \mapsto \text{GL}(\pi; \boldsymbol{\Theta}_0)$ is convex. Implementation-wise, one can formulate it into an epigraph form via Schur complement (Boyd and Vandenberghe, 2004) and use any available convex optimization solver, e.g., CVXPY (Agrawal et al., 2018; Diamond and Boyd, 2016); see Appendix H for more detailed discussions.

We conclude with upper bounds of $\text{GL}_{\min}$ for Frobenius/operator unit balls and matrix-completion basis:

Proposition 3.2 (Upper bounds of $\text{GL}_{\min}$)**.** *Let $\mathcal{M} := \{\boldsymbol{e}_i \tilde{\boldsymbol{e}}_j^\top : (i,j) \in [d_1] \times [d_2]\}$ be the matrix-completion basis. Then, with $\sum_{i,j} := \sum_{i=1}^{d_1} \sum_{j=1}^{d_2}$,*

$$\text{GL}_{\min}\left(\mathcal{B}_F^{d_1 \times d_2}(1)\right) \lesssim \kappa_\star^{-1} d_1 d_2 (d_1 \vee d_2),$$
$$\text{GL}_{\min}\left(\mathcal{B}_{\text{op}}^{d_1 \times d_2}(1)\right) \lesssim \kappa_\star^{-1} (d_1 \vee d_2)^2,$$
$$\text{GL}_{\min}(\mathcal{M}) \leq \text{Harm}(\boldsymbol{\Theta}_\star)^{-1} d_1 d_2 (d_1 \vee d_2),$$

where $\text{Harm}(\boldsymbol{\Theta}_\star) := d_1 d_2 \left(\sum_{i,j} \dot{\mu}((\boldsymbol{\Theta}_\star)_{ij})^{-1} \right)^{-1}$ is the harmonic mean of the entry-wise curvatures.

Proof. The first two claims follow directly from Proposition 3.1 and Jang et al. (2024, Appendix D) The third claim follows from Proposition 3.1 and solving the optimal design explicitly; see Appendix E.1. □

3.3 Comparison with Prior Works

The comparison with prior works is rather intricate, as different works consider different assumptions, especially regarding $\mathcal{X}$ and the design π. For clarity, we will make the comparisons under our Assumptions 1 to 3 and that the GLM is bounded; we defer detailed discussions regarding the discrepancies in the settings compared to prior works to Appendix G.

Table 1 summarizes the comparison with prior state-of-the-arts: nuclear penalized estimator (Fan et al., 2019, Theorem 2) and Stein's lemma-based estimator (Kang et al., 2022, Theorem 4.1). For fair comparison, we consider the bound achievable by optimal design in all cases whenever applicable. Our `GL-LowPopArt` achieves improvements in curvature-dependent quantites, and sometime even in dimensions.[1] This suggests that our optimal GL-Design captures the problem-specific characteristics more effectively than prior works.

Remark 1 (Unbounded $\mathcal{X}$)**.** *It is common to consider the case when the (matrix-valued) covariates are drawn from $\mathbb{R}^{d_1 \times d_2}$ with π being some tail-bounded distribution (Fan et al., 2019; Kang et al., 2022; Lu et al., 2021; Negahban and Wainwright, 2011). Thus, one may wonder if our results hold for unbounded $\mathcal{X}$.*

[1] This is similar to how Jang et al. (2024) improved over Koltchinskii et al. (2011) in linear trace regression.

In Appendix C.2, we show that when $\mu(z) = z$, our results extend to unbounded $\mathcal{X}$ and obtain better error rates than the nuclear penalized estimator in multivaraite regression (Negahban and Wainwright, 2011). However, when μ is nonlinear, our current analysis relies heavily on the fact that $\mathcal{X} \subseteq \mathcal{B}_{\mathrm{op}}^{d_1 \times d_2}(1)$. On the bright side, our proof suggests that under proper moment conditions on π, our results may extend to unbounded $\mathcal{X}$ with potentially stronger requirement on $\mathbf{\Theta}_0$. We leave this direction for future work.

3.4 Proof Sketch of Theorem 3.1

The proof is inspired by Jang et al. (2024, Theorem 3.1), but some crucial differences make the extension non-trivial. The full proof is deferred to Appendix C.

For simplicity, let us denote $\boldsymbol{H} := \boldsymbol{H}(\pi; \mathbf{\Theta}_0)$ in this proof sketch, and ignore Proj_Ω. The core idea is to apply matrix Catoni estimation of Minsker (2018) to $\mathbf{\Theta}_\star - \mathbf{\Theta}_0$, by constructing vectorized one-sample estimators (line 8) as follows:

$$\tilde{\boldsymbol{\theta}}_t = \boldsymbol{H}^{-1} \left(y_t - \mu(\langle \boldsymbol{X}_t, \mathbf{\Theta}_0 \rangle) \right) \mathrm{vec}(\boldsymbol{X}_t).$$

For matrix Catoni to apply, the one-sample estimators should be unbiased, i.e., they should satisfy $\mathbb{E}[\tilde{\boldsymbol{\theta}}_t] = \mathrm{vec}(\mathbf{\Theta}_\star - \mathbf{\Theta}_0)$. However, note that

$$\mathbb{E}[\tilde{\boldsymbol{\theta}}_t] = \boldsymbol{H}^{-1} \mathbb{E}_{\boldsymbol{X}} \left[\left(\mu(\langle \boldsymbol{X}, \mathbf{\Theta}_\star \rangle) - \mu(\langle \boldsymbol{X}, \mathbf{\Theta}_0 \rangle) \right) \mathrm{vec}(\boldsymbol{X}) \right].$$

When $\mu(z) = z$ as in Jang et al. (2024), above reduces to $\mathrm{vec}(\mathbf{\Theta}_\star - \mathbf{\Theta}_0)$, making $\tilde{\boldsymbol{\theta}}_t$ its unbiased estimator. However, when μ is nonlinear, it becomes *biased*.

The key technical novelty is appropriately dealing with this bias, inspired by recent progress in logistic and generalized linear bandits (Abeille et al., 2021; Jun et al., 2021; Lee et al., 2024a). Specifically, by the first-order Taylor expansion of μ and self-concordance (Assumption 3(a)), we show the following (**Eqn. (C.1)**):

$$\left\| \mathbb{E}[\widetilde{\mathbf{\Theta}}_t] - (\mathbf{\Theta}_\star - \mathbf{\Theta}_0) \right\|_F \lesssim R_s \left\| \mathbf{\Theta}_\star - \mathbf{\Theta}_0 \right\|_{\mathrm{nuc}}^2.$$

We remark that the sample splitting approach[2] of `Warm-LowPopArt` (Jang et al., 2024, Algorithm 2) fails due to this bias. On the other hand, if the initial estimator $\mathbf{\Theta}_0$ is sufficiently good (in the precise sense as stated in the theorem statement), then this bias becomes negligible. Lastly, the GL-Design objective $\mathrm{GL}(\pi; \mathbf{\Theta}_\star)$ arises from the matrix variance statistics for $\widetilde{\mathbf{\Theta}}_t$'s, and we conclude with the error rate for matrix Catoni estimator (Minsker, 2018, Corollary 3.1). □

[2]run Stage II with $N_2/2$ samples with $\mathbf{0}$ to obtain $\mathbf{\Theta}_0$, then Stage II again using the remaining samples and $\mathbf{\Theta}_0$

3.5 Stage I via Nuclear Penalized MLE

For $\mathbf{\Theta}_0$ from Stage I, we consider the standard *nuclear penalized MLE*, which has been extensively studied in generalized low-rank trace regression literature (Fan et al., 2019; Kang et al., 2022; Lu et al., 2021). Concretely, using N_1 samples from π_0,

$$\mathbf{\Theta}_0 \leftarrow \arg\min_{\mathbf{\Theta} \in \Omega} \mathcal{L}_{N_1}(\mathbf{\Theta}) + \lambda_{N_1} \left\| \mathbf{\Theta} \right\|_{\mathrm{nuc}},$$

$$\mathcal{L}_{N_1}(\mathbf{\Theta}) := \frac{1}{N_1} \sum_{t=1}^{N_1} \left\{ m(\langle \boldsymbol{X}_t, \mathbf{\Theta} \rangle) - y_t \langle \boldsymbol{X}_t, \mathbf{\Theta} \rangle \right\}.$$

We first present the its error rate:

Theorem 3.2 (Error Rate of Stage I)**.** *Let $\delta \in (0, 1)$, $\mathcal{X} \subseteq \mathcal{B}_{\mathrm{op}}^{d_1 \times d_2}(1)$, and set $\lambda_N \asymp \sqrt{\frac{C}{N_1} \log \frac{d_1 + d_2}{\delta}}$.*[a] *Then, the following holds w.p. at least $1 - \delta/2$:*

$$\left\| \mathbf{\Theta}_0 - \mathbf{\Theta}_\star \right\|_F \lesssim \frac{1}{\lambda_{\min}(\boldsymbol{H}(\pi; \mathbf{\Theta}_\star))} \sqrt{\frac{Cr}{N_1} \log \frac{d_1 + d_2}{\delta}},$$

provided *that $N_1 \gtrsim \frac{C(d_1+d_2)r}{\lambda_{\min}(\boldsymbol{H}(\pi; \mathbf{\Theta}_\star))^2} \log \frac{d_1+d_2}{\delta}$.*

[a]For bounded GLMs, $C \asymp g(\tau) \mathbb{E}_{\boldsymbol{X} \sim \pi} \left[\dot{\mu}(\langle \boldsymbol{X}, \mathbf{\Theta}_\star \rangle) \right]$, and for GLMs with finite subexponential norm of K, $C \asymp K^2$; see Proposition F.1.

Combining the above with Theorem 3.2, we have the following requirement on N_1 and N_2 as follows:

$$N_1 \gtrsim R_s \frac{Cr^2}{\lambda_{\min}(\boldsymbol{H}(\pi; \mathbf{\Theta}_\star))^3} \sqrt{\frac{N_2 \log \frac{d_1+d_2}{\delta}}{g(\tau) \mathrm{GL}(\pi; \mathbf{\Theta}_\star)}},$$

i.e., asymptotically, the requirement for Stage I is negligible compared to that for Stage II. Also, note that the requirement vanishes when $R_s = 0$, which mirrors the results in Jang et al. (2024) for linear trace regression.

Proof Sketch of Theorem 3.2. We follow the general framework for analyzing M-estimators with decomposable regularizers (Wainwright, 2019, Chapter 9), introducing notable analytical improvements over Fan et al. (2019, Theorem 2). The complete proof is deferred to Appendix F.

First, leveraging the boundedness of the covariates and matrix Bernstein inequalities (Tropp, 2012, 2015), we derive a tight choice for the regularization coefficient λ_{N_1} for GLMs with bounded or sub-exponential responses (**Proposition F.1**). This approach eliminates a $d_1 \vee d_2$ factor present in the double covering argument of Fan et al. (2019, Lemma 1) and accommodates a broader class of GLMs, such as Poisson, which their assumptions preclude.[3]

[3]Fan et al. (2019, Lemma 2) requires $\|\mathbf{\Theta}_\star\|_F \gtrsim \sqrt{d_1 \vee d_2}$ and $|\ddot{\mu}(z)| \leq \frac{1}{|z|}$ for $|z| > 1$ (conditions C4 and C5).

Next, we establish the restricted strong convexity (RSC) of the regularized loss function (**Lemma F.1**). By combining the self-concordance of the link function (Assumption 3(a)) (Ostrovskii and Bach, 2021) with standard empirical process theory (Ledoux and Talagrand, 1991) and the matrix Bernstein inequality yields the requisite RSC condition. The final error bound then immediately follows from standard M-estimation guarantees (Wainwright, 2019, Corollary 9.20). □

Remark 2 (E-Optimal Design)**.** *Before Stage I, one can consider the E-optimal design (Pukelsheim, 2006): $\pi_0 \in \arg\max_{\pi \in \mathcal{P}(\mathcal{A})} \lambda_{\min}(\boldsymbol{V}(\pi))$. This would optimize the error rate of Stage I, further alleviating the overall sample size requirements. We defer the details of to Appendix I.*

4 LOCAL MINIMAX LOWER BOUND FOR $\|\cdot\|_F$

In this section, we prove a *local (instance-wise)* minimax lower bound on the estimation error for generalized low-rank trace regression in the intersection of rank and nuclear norm balls. For each instance $\boldsymbol{\Theta}_\star$ with $\mathrm{rank}(\boldsymbol{\Theta}_\star) \le r$ and $\|\boldsymbol{\Theta}_\star\|_{\mathrm{nuc}} \le S_*$ for some $S_* > 0$, define its local neighborhood of radius $\varepsilon > 0$ as

$$\mathcal{N}(\boldsymbol{\Theta}_\star; \varepsilon, r, S_*) := \{\boldsymbol{\Theta} \in \Theta(r, S_*) : \|\boldsymbol{\Theta} - \boldsymbol{\Theta}_\star\|_F \le \varepsilon\},$$

$$\Theta(r, S_*) := \left\{\boldsymbol{\Theta} \in \mathbb{R}^{d_1 \times d_2} : \mathrm{rank}(\boldsymbol{\Theta}) \le r,\ \|\boldsymbol{\Theta}\|_{\mathrm{nuc}} \le S_*\right\}.$$

$\Theta(r, S_*)$ has been considered before in the context of minimax lower bound by Rohde and Tsybakov (2011), similar to that of sparse regression in the intersection of ℓ_0 and ℓ_1-ball constraints (Rigollet and Tsybakov, 2011, Theorem 5.3).

We now present our generic lower bound for generalized linear trace regression encompassing active scenarios:

Theorem 4.1 (Local Minimax Lower Bound)**.** *Let $\mathcal{A} \subseteq \mathcal{B}_F^{d_1 \times d_2}(1)$, and $S_* > 0, r \ge 1$ be such that $\frac{S_*^2}{r} \ge \gamma$ for some $\gamma > 0$. Also, suppose that $N \ge \frac{R_s^2 \log 2}{2^{10} e} \frac{r(d_1 \vee d_2) g(\tau)}{\lambda_{\max}(\boldsymbol{H}(\pi; \boldsymbol{\Theta}_\star))}$. Then, there exist $C_1, C_2(\gamma) > 0$[a] and $c \in (0, 1)$ such that the following holds: for each $\boldsymbol{\pi}_N$ and $\boldsymbol{\Theta}_\star \in \Theta(r, S_*)$ with $\|\boldsymbol{\Theta}_\star\|_F^2 \ge \frac{9\gamma}{8}$, there exists a **small enough** $\varepsilon = \varepsilon(\boldsymbol{\pi}_N, \boldsymbol{\Theta}_\star) > 0$ such that $\inf_{\widehat{\boldsymbol{\Theta}}} \sup_{\widetilde{\boldsymbol{\Theta}}_\star \in \mathcal{N}_\star} \mathbb{P}_{\boldsymbol{\pi}_N, \widetilde{\boldsymbol{\Theta}}_\star}(\mathcal{E}) \ge c$, where*

$$\mathcal{E} := \left\{\left\|\widehat{\boldsymbol{\Theta}} - \widetilde{\boldsymbol{\Theta}}_\star\right\|_F^2 \ge \frac{C_2 g(\tau) r(d_1 \vee d_2)}{N \lambda_{\max}(\boldsymbol{H}(\boldsymbol{\pi}_N; \boldsymbol{\Theta}_\star)) S_*^2}\right\},$$

$$\boldsymbol{H}(\boldsymbol{\pi}_N; \boldsymbol{\Theta}_\star) := \frac{1}{N} \sum_{t=1}^{N} \boldsymbol{H}(\pi_t; \boldsymbol{\Theta}_\star),$$

$\mathcal{N}_\star := \mathcal{N}(\boldsymbol{\Theta}_\star; \varepsilon, r, S_*)$, and $\mathbb{P}_{\boldsymbol{\pi}_N, \widetilde{\boldsymbol{\Theta}}_\star}$ is the probability measure of the N observations under $\boldsymbol{\pi}_N$ and $\widetilde{\boldsymbol{\Theta}}_\star$.

[a] C_1 is a universal constant, and $C_2 = \frac{C_2' \gamma}{(1+\sqrt{\gamma})^2}$ for a universal constant $C_2' > 0$.

Proof Sketch. We mainly utilize the many hypotheses technique of Tsybakov (2009, Chapter 2) for high-probability minimax lower bound; see also Yang and Barron (1999). One key technical novelty is the construction of a *local* packing $\Theta_{r,\varepsilon,\beta} \subset \Theta(r, S_*)$ around the given instance $\boldsymbol{\Theta}_\star$. Then, we carefully expand the D_{KL} between two GLMs from the packing by utilizing its Bregman divergence form (Lee et al., 2024b) and self-concordance of μ (Assumption 3(a)), which leads to the instance-specific quantity $\lambda_{\max}(\boldsymbol{H}(\pi; \boldsymbol{\Theta}_\star))^{-1}$. Also, note that we don't explicitly require any restricted isometry assumption (Koltchinskii et al., 2011, Eqn. (2.4)). Refer to Appendix J for the full proof.

This notably deviates from Rohde and Tsybakov (2011, Theorem 5), where they considered a packing around $\boldsymbol{\Theta}_\star = \mathbf{0}$ for linear trace regression. When $\mu(z) = z$, this still results in a tight lower bound as the problem difficulty is uniform across all $\boldsymbol{\Theta}_\star \in \Theta(r, S_*)$. □

Instance-Specific Nature. Our lower bound scales inversely with the "optimistic" instance-specific curvature, $\lambda_{\max}(\boldsymbol{H}(\pi; \boldsymbol{\Theta}_\star))^{-1}$, thereby capturing how problem difficulty varies across different instances $\boldsymbol{\Theta}_\star$. Here, "optimistic" in the sense that it reflects the most favorable local curvature at $\boldsymbol{\Theta}_\star$, in contrast to worst-case bounds based on $\lambda_{\min}$. To the best of our knowledge, this is the *first* result in generalized linear trace regression that exhibits such an instance-wise dependency. This mirrors the local minimax lower bounds established for logistic bandits (Abeille et al., 2021, Theorem 2) and online LQR (Simchowitz and Foster, 2020, Theorem 1), which also highlight instance-specific complexities. In contrast, classical worst-case minimax lower bounds for generalized linear trace regression and matrix completion (Davenport et al., 2014; Koltchinskii et al., 2011; Lafond, 2015; Rohde and Tsybakov, 2011; Taki et al., 2021) fail to capture such local variations.

Near Instance-wise Optimality. Our lower bound (Theorem 4.1) holds for *each* sampling policy $\boldsymbol{\pi}_N$, including adaptive ones. In other words, optimality here is defined relative to each instance $(\boldsymbol{\Theta}_\star, \boldsymbol{\pi}_N)$, where the policy itself forms part of the instance. To clarify in what sense `GL-LowPopArt` is optimal, we focus on the class of *single-switch policies*, $\pi_t = \pi_1 \mathbb{1}[t \le N_1] + \pi_2 \mathbb{1}[t > N_1]$ with $N_1 = o(N)$. For brevity, write $\lambda_i := \lambda_i(\boldsymbol{H}(\pi_2; \boldsymbol{\Theta}_\star))$ for $i \in \{\min, \max\}$.

In the large-N regime, the ratio between the upper

and lower bounds (Theorems 3.1 and 4.1) on squared Frobenius error maximally scales with the Hessian's condition number $\frac{\lambda_{\max}}{\lambda_{\min}}$. Thus, within the single-switch policy class, `GL-LowPopArt` is nearly instance-wise minimax optimal. This stands in contrast to guarantees for the nuclear penalized MLE (Theorem 3.2), where the gap between upper and lower bounds is significantly larger, involving unfavorable curvature terms and even dimension factors.

Requirement on N. A keen reader may observe that our local minimax lower bound holds under the condition $N \gtrsim \frac{R_s^2 r(d_1 \vee d_2)}{\lambda_{\max}(\boldsymbol{H}(\pi;\boldsymbol{\Theta}_\star))}$. We emphasize that this is not restrictive and actually provides an intuitive justification for Stage I as a warm-up phase; in fact, we believe that some condition of this form on N is necessary—although we do not currently have a formal proof. The requirement on N arises when bounding the KL divergence between the true model $\boldsymbol{\Theta}_\star$ and an alternative model from the constructed local packing. Intuitively, this stems from the necessity for the two models to be sufficiently close for self-concordance properties to take effect; this was also the case for prior local minimax lower bounds (Abeille et al., 2021, Theorem 2) (Simchowitz and Foster, 2020, Theorem 1), where the requirement on horizon length T arises similarly. Finally, we remark that in the linear setting ($\mu(z) = z \Rightarrow R_s = 0$), our requirement on N vanishes.

5 APPLICATIONS OF `GL-LowPopArt`

5.1 Generalized Linear Matrix Completion under Uniform Sampling at Random

Here, $\mathcal{X} = \mathcal{M}$ (see Proposition 3.2), $\pi^U = \mathrm{Unif}(\mathcal{X})$, and $\max_{i,j} |(\boldsymbol{\Theta}_\star)_{i,j}| \le \gamma$ for a $\gamma > 0$. For clarity, we focus on the *1-bit matrix completion* (Davenport et al., 2014) with $\mu(z) = (1 + e^{-z})^{-1}$. Let $\mathcal{E}_F := \|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}_\star\|_F^2$.

We first compare the error bound of `GL-LowPopArt` when $\pi_0 = \pi = \pi^U$ with Davenport et al. (2014, Theorem 1) and Klopp et al. (2015, Corollary 2), denoted as Dav14 and Klo15, respectively:

$$\mathcal{E}_F \lesssim \frac{1}{\min_{i,j} \dot{\mu}((\boldsymbol{\Theta}_\star)_{ij})} \frac{r d_1 d_2 (d_1 \vee d_2)}{N}, \qquad (\textbf{ours})$$

$$\mathcal{E}_F \lesssim \frac{1}{\min_{|z| \le \gamma} \dot{\mu}(z)} \sqrt{\frac{r (d_1 d_2)^2 (d_1 \vee d_2)}{N}}, \qquad (\text{Dav14})$$

$$\mathcal{E}_F \lesssim \left(\frac{1}{\min_{|z| \le \gamma} \dot{\mu}(z)} \right)^2 \frac{r d_1 d_2 (d_1 \vee d_2)}{N}. \qquad (\text{Klo15})$$

For our error bound, we utilize Proposition 3.1 to express $\mathrm{GL}(\pi^U; \boldsymbol{\Theta}_\star)$ in terms of dimensions for a clear comparison, but we remark that the above bound is looser than Theorem 3.1. Still, our bound captures the instance-specific difficulty via $\frac{1}{\min_{i,j} \dot{\mu}((\boldsymbol{\Theta}_\star)_{ij})}$. On the other hand, the other bounds depend on the worst-case curvature $\frac{1}{\min_{|z| \le \gamma} \dot{\mu}(z)}$, which may be quite loose in situations where $\min_{i,j} \dot{\mu}((\boldsymbol{\Theta}_\star)_{ij}) \gg \min_{|z| \le \gamma} \dot{\mu}(z)$, e.g., due to misspecification of γ. We also note that, by Proposition 3.2, when optimal GL-design is employed (i.e., adaptive scenario), our bound improves to

$$\mathcal{E}_F \lesssim \mathrm{Harm}(\boldsymbol{\Theta}_\star)^{-1} \frac{r d_1 d_2 (d_1 \vee d_2)}{N},$$

which is tighter than all the aforementioned bounds.

Algorithm-wise, Davenport et al. (2014); Klopp et al. (2015), along with other approaches (Cai and Zhou, 2016, 2013; Lafond, 2015; Srebro and Salakhutdinov, 2010), requires the knowledge of $\gamma > 0$, to compute the nuclear-norm regularized estimator *with* the constraint of $\|\boldsymbol{\Theta}_\star\|_\infty \le \gamma$ or $\|\boldsymbol{\Theta}_\star\|_{\max} \le \gamma$. Interestingly, `GL-LowPopArt` does *not* require any knowledge about $\boldsymbol{\Theta}_\star$, yet it fully adapts to the given instance. In Appendix M, we demonstrate the numerical efficacy of `GL-LowPopArt` in the 1-bit matrix completion setting compared to nuclear penalized MLE.

We now ask whether `GL-LowPopArt` is instance-wise optimal in the generalized linear matrix completion setting. In many literature (Davenport et al., 2014; Klopp et al., 2015; Koltchinskii et al., 2011), the optimality is often discussed in the context of entrywise norm-constrained neighborhoods, e.g., $\max_{i,j} |(\boldsymbol{\Theta}_\star)_{i,j}| \le \gamma$ for some $\gamma > 0$ (Koltchinskii et al., 2011, Section 4), which does not match with our assumption on $\boldsymbol{\Theta}_\star$ for our lower bound (Theorem 4.1), i.e., we cannot claim (near) instance-wise optimality here. We leave this to a future direction.

Remark 3 (Comparison to Rohde and Tsybakov (2011))**.** *For linear matrix completion with **fixed design** (i.e., π is a mixture of Diracs), Rohde and Tsybakov (2011, Corollary 1, Theorem 5) establish a minimax optimal error rate of $\|\mathrm{vec}(\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}_\star)\|_{\boldsymbol{V}(\pi)}^2 \asymp \frac{r(d_1 \vee d_2)}{N}$, with the lower bound holding under restricted isometry. Our bound is weaker in this case, as they achieve a tight characterization in the Mahalanobis norm. Whether similar rates can be derived for generalized linear matrix completion involving $\|\cdot\|_{\boldsymbol{H}(\pi;\boldsymbol{\Theta}_\star)}$ is an interesting future direction.*

5.2 Bilinear Dueling Bandits

5.2.1 Problem Description

In **bilinear dueling bandits**, $\mathcal{A} \subseteq \mathcal{B}^d(1)$ is considered to be a fixed vector-valued arm-set that is compact, i.e., $\mathcal{X}$ may be infinite as in continuous dueling bandits (Kumagai, 2017). At each timestep t, the learner chooses

Algorithm 2: `BETC-GLM-LR`

for $t = 1, 2, \cdots, N_1 + N_2$ **do**
  Run `GL-LowPopArt`(N_1, N_2) and obtain $\widehat{\boldsymbol{\Theta}}$;
Obtain the estimated Borda winner:

$$\hat{\boldsymbol{\phi}} \leftarrow \arg\max_{\boldsymbol{\phi} \in \mathcal{A}} \left\{ \widehat{B}(\boldsymbol{\phi}) \triangleq \mathbb{E}_{\boldsymbol{\phi}' \sim \mathrm{Unif}(\mathcal{A})} \left[\mu \left(\boldsymbol{\phi}^\top \widehat{\boldsymbol{\Theta}} \boldsymbol{\phi}' \right) \right] \right\}$$

for $t = N_1 + N_2 + 1, \cdots, T$ **do**
  Pull $(\widehat{\boldsymbol{\phi}}, \widehat{\boldsymbol{\phi}})$;

a pair of arms $(\boldsymbol{\phi}_{w,t}, \boldsymbol{\phi}_{l,t}) \in \mathcal{A} \times \mathcal{A}$, and receives a feedback sampled as

$$o_t = \mathbb{1}[\boldsymbol{\phi}_{w,t} \succ \boldsymbol{\phi}_{l,t}] \sim \mathrm{Ber}(\mu\left(\boldsymbol{\phi}_{w,t}^\top \boldsymbol{\Theta}_\star \boldsymbol{\phi}_{l,t}\right)),$$

for an unknown, *skew-symmetric* $\boldsymbol{\Theta}_\star$ of rank $2r$, and a known link function $\mu : \mathbb{R} \to [0, 1]$ that additionally satisfies the following (Wu et al., 2024):

Assumption 4. *In addition to Assumption 3,* $\mu : \mathbb{R} \to [0, 1]$ *satisfies* $\mu(z) + \mu(-z) = 1,\ z \in \mathbb{R}$.

Some examples of μ that satisfies the above include $\mu(z) = \frac{1+z}{2}$ and $\mu(z) = (1 + e^{-z})^{-1}$ (Bernoulli).

The learner's goal is to minimize the Borda regret (Saha et al., 2021; Wu et al., 2024):

$$\mathrm{Reg}^B(T) := \sum_{t=1}^{T} \left\{ B(\boldsymbol{\phi}_\star) - \frac{B(\boldsymbol{\phi}_{w,t}) + B(\boldsymbol{\phi}_{l,t})}{2} \right\},$$

where the *(shifted) Borda score* of $\boldsymbol{\phi} \in \mathcal{A}$ is defined as

$$B(\boldsymbol{\phi}) := \mathbb{E}_{\boldsymbol{\phi}' \sim \mathrm{Unif}(\mathcal{A})}[\mu(\boldsymbol{\phi}^\top \boldsymbol{\Theta} \boldsymbol{\phi}')],$$

and $\boldsymbol{\phi}_\star = \arg\max_{\boldsymbol{\phi} \in \mathcal{A}} B(\boldsymbol{\phi})$ is the *Borda winner*. Note that when $\mathcal{X}$ is finite, it reduces to the usual definition of Borda regret/winner in the finite-armed dueling bandits (Jamieson et al., 2015; Saha et al., 2021). Unlike the Condorcet winner, the Borda winner always exists for any preference model (Bengs et al., 2021).

Remark 4 (Significance of the Setting)**.** *We emphasize that this is a* ***novel*** *dueling bandits setting, motivated by recent progress in contextualizing general preference learning with item-wise features in RLHF (Zhang et al., 2025). We defer further discussions on the proposed setting to Appendix K.*

5.2.2 `BETC-GLM-LR` and Regret Upper Bound

We consider an explore-then-commit approach, where the exploration is done via our `GL-LowPopArt` (Algorithm 2). It attains the following Borda regret bound:

Theorem 5.1 ((Informal))**.** *With appropriate choices of* N_1 *and* N_2 *in* `GL-LowPopArt` *and large enough* T*,* `BETC-GLM-LR` *attains the following Borda regret bound with probability at least* $1 - \delta$*:*

$$\mathrm{Reg}^B(T) \lesssim \left(\mathrm{GL}_{\min}(\mathcal{X}) \log \frac{d}{\delta} \right)^{1/3} \left(\kappa_\star^B T \right)^{2/3},$$

where $\mathcal{X} := \{\boldsymbol{\phi}_i \boldsymbol{\phi}_j^\top : 1 \le i < j \le K\} \subseteq \mathcal{B}_{\mathrm{op}}^{d \times d}(1)$ *and* $\kappa_\star^B := \mathbb{E}_{\boldsymbol{\phi}' \sim \mathrm{Unif}(\mathcal{A})}[\dot{\mu}(\boldsymbol{\phi}_\star^\top \boldsymbol{\Theta} \boldsymbol{\phi}')]$.

Proof Sketch. We deviate significantly from Wu et al. (2024) by using the self-concordance of μ as in Abeille et al. (2021, Theorem 1) for the regret bound to scale *with* $\kappa_\star^B$. Refer to Appendix L.1 for the full proof. □

We believe $T^{2/3}$ dependency of the Borda regret is unavoidable, as there are already $\Omega(T^{2/3})$ Borda regret lower bounds (omitting other dependencies) for entry-wise (Saha et al., 2021, Theorem 16) and generalized linear dueling bandits (Wu et al., 2024, Theorem 4.1).

Two quantities make our regret bound truly instance-specific. One is $\mathrm{GL}_{\min}(\mathcal{X})$, which, as discussed previously, captures the geometry of $\mathcal{X}$ and the associated nonlinearity via the Hessian. Another is $\kappa_\star^B$, which captures the average curvature of the preference model when compared against the Borda winner. Analogous to recent findings in generalized linear bandits (Abeille et al., 2021; Lee et al., 2024a), our Borda regret can *benefit* from the "flatness" of the link function. Intuitively, we enjoy lower regret in highly flat landscapes, provided that the worst-case flatness across all item pairs is comparable to the flatness around the Borda winner. However, if there exists a hard pair elsewhere in the arm set, the estimation difficulty dominates and we lose this variance-reduction benefit. We defer a more in-depth discussions, including comparisons with Wu et al. (2024), to Appendix L.2.

6 CONCLUSION

This work addresses the critical gap in prior work by explicitly considering instance-specific curvature in generalized low-rank trace regression. We introduce `GL-LowPopArt`, a novel estimator that achieves state-of-the-art performance, adapting to both the nonlinearity of the model and the underlying arm-set geometry. We establish the first instance-wise minimax lower bound, demonstrating the near-optimality of `GL-LowPopArt`. We showcase its benefits through applications to generalized linear matrix completion and bilinear dueling bandits, a novel setting of independent interest for general preference learning (Zhang et al., 2025). In Appendix N, we collect further future directions.

Acknowledgements

We thank all the anonymous reviewers for their helpful comments that helped us improve the manuscript. K. Jang was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) [RS-2021-II211341, Artificial Intelligence Graduate School Program (Chung-Ang University)] and by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) [RS-2026-25499850]. K.-S. Jun was supported in part by the National Science Foundation under grant CCF-2327013 and Meta Platforms, Inc.

References

Marc Abeille, Louis Faury, and Clément Calauzènes. Instance-Wise Minimax-Optimal Algorithms for Logistic Bandits. In *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pages 3691–3699. PMLR, 13–15 Apr 2021. URL https://proceedings.mlr.press/v130/abeille21a.html.

P.-A. Absil, Robert Mahony, and Rodolphe Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2007. doi:10.1515/9781400830244.

Alekh Agarwal, Sahand Negahban, and Martin J. Wainwright. Fast global convergence of gradient methods for high-dimensional statistical recovery. *The Annals of Statistics*, 40(5):2452 – 2482, 2012. doi:10.1214/12-AOS1032.

Akshay Agrawal, Robin Verschueren, Steven Diamond, and Stephen Boyd. A rewriting system for convex optimization problems. *Journal of Control and Decision*, 5(1):42–60, 2018. doi:10.1080/23307706.2017.1397554.

Mokhtar Z. Alaya and Olga Klopp. Collective Matrix Completion. *Journal of Machine Learning Research*, 20(148):1–43, 2019. URL http://jmlr.org/papers/v20/18-483.html.

Zeyuan Allen-Zhu, Yuanzhi Li, Aarti Singh, and Yining Wang. Near-optimal discrete optimization for experimental design: a regret minimization approach. *Mathematical Programming*, 186(1):439–478, 2021. doi:10.1007/s10107-019-01464-2.

Pierre Alquier, Vincent Cottet, and Guillaume Lecué. Estimation bounds and sharp oracle inequalities of regularized procedures with Lipschitz loss functions. *The Annals of Statistics*, 47(4):2117 – 2144, 2019. doi:10.1214/18-AOS1742.

Shun'ichi Amari. *Information Geometry and Its Applications*. Applied Mathematical Sciences. Springer Tokyo, 2016.

Mohammad Gheshlaghi Azar, Daniel Guo Zhaohan, Bilal Piot, Rémi Munos, Mark Rowland, Michal Valko, and Daniele Calandriello. A General Theoretical Paradigm to Understand Learning from Human Preferences. In *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pages 4447–4455. PMLR, 02–04 May 2024. URL https://proceedings.mlr.press/v238/gheshlaghi-azar24a.html.

Francis Bach. Self-concordant analysis for logistic regression. *Electronic Journal of Statistics*, 4(none): 384 – 414, 2010. doi:10.1214/09-EJS521.

David Balduzzi, Karl Tuyls, Julien Perolat, and Thore Graepel. Re-evaluating Evaluation. In *Advances in Neural Information Processing Systems*, volume 31, pages 3272 – 3283. Curran Associates, Inc., 2018. URL https://arxiv.org/abs/1806.02643.

David Balduzzi, Marta Garnelo, Yoram Bachrach, Wojciech Czarnecki, Julien Perolat, Max Jaderberg, and Thore Graepel. Open-ended learning in symmetric zero-sum games. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 434–443. PMLR, 09–15 Jun 2019. URL https://proceedings.mlr.press/v97/balduzzi19a.html.

Siddharth Barman. Approximating Nash Equilibria and Dense Bipartite Subgraphs via an Approximate Version of Caratheodory's Theorem. In *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing*, STOC '15, page 361–369, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450335362. doi:10.1145/2746539.2746566.

Amir Beck and Marc Teboulle. A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *SIAM Journal on Imaging Sciences*, 2(1): 183–202, 2009. doi:10.1137/080716542.

Viktor Bengs, Ròbert Busa-Fekete, Adil El Mesaoudi-Paul, and Eyke Hüllermeier. Preference-based Online Learning with Dueling Bandits: A Survey. *Journal of Machine Learning Research*, 22(7):1–108, 2021. URL http://jmlr.org/papers/v22/18-546.html.

Viktor Bengs, Aadirupa Saha, and Eyke Hüllermeier. Stochastic Contextual Dueling Bandits under Linear Stochastic Transitivity Models. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 1764–1786. PMLR, 17–23

Jul 2022. URL https://proceedings.mlr.press/v162/bengs22a.html.

J. Bennett and S. Lanning. The Netflix Prize. In *Proceedings of the KDD Cup Workshop 2007*, pages 3–6, New York, 2007. ACM. URL http://www.cs.uic.edu/~liub/KDD-cup-2007/NetflixPrize-description.pdf.

Quentin Berthet and Nicolai Baldin. Statistical and Computational Rates in Graph Logistic Regression. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pages 2719–2730. PMLR, 26–28 Aug 2020. URL https://proceedings.mlr.press/v108/berthet20a.html.

Quentin Bertrand, Wojciech Marian Czarnecki, and Gauthier Gidel. On the Limitations of the Elo, Real-World Games are Transitive, not Additive. In *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, volume 206 of *Proceedings of Machine Learning Research*, pages 2905–2921. PMLR, 25–27 Apr 2023. URL https://proceedings.mlr.press/v206/bertrand23a.html.

Rajendra Bhatia. *Matrix Analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer New York, NY, 1997. doi:10.1007/978-1-4612-0653-8.

Srinadh Bhojanapalli, Behnam Neyshabur, and Nati Srebro. Global Optimality of Local Search for Low Rank Matrix Recovery. In *Advances in Neural Information Processing Systems*, volume 29, pages 3880–3888. Curran Associates, Inc., 2016. URL https://arxiv.org/abs/1605.07221.

Yingjie Bi, Haixiang Zhang, and Javad Lavaei. Local and Global Linear Convergence of General Low-Rank Matrix Recovery Problems. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(9):10129–10137, Jun. 2022. doi:10.1609/aaai.v36i9.21252. URL https://arxiv.org/abs/2104.13348.

Aurélien Bibaut and Nathan Kallus. Demystifying Inference After Adaptive Experiments. *Annual Review of Statistics and Its Application*, 12 (Volume 12, 2025):407–423, 2025. ISSN 2326-831X. doi:10.1146/annurev-statistics-040522-015431.

Nicolas Boumal. *An Introduction to Optimization on Smooth Manifolds*. Cambridge University Press, 2023. doi:10.1017/9781009166164.

Nicolas Boumal, Vlad Voroninski, and Afonso Bandeira. The non-convex Burer-Monteiro approach works on smooth semidefinite programs. In *Advances in Neural Information Processing Systems*, volume 29, pages 2765–2773. Curran Associates, Inc., 2016. URL https://arxiv.org/abs/1606.04970.

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

Stephen Boyd, Laurent El Ghaoui, Eric Feron, and Venkataramanan Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*, volume 15 of *Studies in Applied and Numerical Mathematics*. Society for Industrial and Applied Mathematics, 1994. doi:10.1137/1.9781611970777.

Ralph Allan Bradley and Milton E. Terry. Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons. *Biometrika*, 39(3-4):324–345, 12 1952. ISSN 0006-3444. doi:10.1093/biomet/39.3-4.324.

Rob Brekelmans, Vaden Masrani, Frank Wood, Greg Ver Steeg, and Aram Galstyan. All in the Exponential Family: Bregman Duality in Thermodynamic Variational Inference. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 1111–1122. PMLR, 13–18 Jul 2020. URL https://proceedings.mlr.press/v119/brekelmans20a.html.

Samuel Burer and Renato D. C. Monteiro. A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Mathematical Programming*, 95(2):329–357, 2003. doi:10.1007/s10107-002-0352-8.

Samuel Burer and Renato D. C. Monteiro. Local Minima and Convergence in Low-Rank Semidefinite Programming. *Mathematical Programming*, 103(3):427–444, 2005. doi:10.1007/s10107-004-0564-1.

Jian-Feng Cai, Emmanuel J. Candès, and Zuowei Shen. A Singular Value Thresholding Algorithm for Matrix Completion. *SIAM Journal on Optimization*, 20(4):1956–1982, 2010. doi:10.1137/080738970.

T. Tony Cai and Wen-Xin Zhou. Matrix completion via max-norm constrained optimization. *Electronic Journal of Statistics*, 10(1):1493 – 1525, 2016. doi:10.1214/16-EJS1147.

Tony Cai and Wen-Xin Zhou. A Max-Norm Constrained Minimization Approach to 1-Bit Matrix Completion. *Journal of Machine Learning Research*, 14(114):3619–3647, 2013. URL http://jmlr.org/papers/v14/cai13b.html.

Emmanuel J. Candès and Benjamin Recht. Exact Matrix Completion via Convex Optimization. *Foundations of Computational Mathematics*, 9(6):717–772, 2009. doi:10.1007/s10208-009-9045-5.

Emmanuel J. Candès and Yaniv Plan. Matrix Completion With Noise. *Proceedings of the IEEE*, 98(6):925–936, 2010. doi:10.1109/JPROC.2009.2035722.

Emmanuel J. Candès and Yaniv Plan. Tight Oracle Inequalities for Low-Rank Matrix Recovery From a Minimal Number of Noisy Random Measurements. *IEEE Transactions on Information Theory*, 57(4): 2342–2359, 2011. doi:10.1109/TIT.2011.2111771.

Yang Cao and Yao Xie. Poisson matrix recovery and completion. *IEEE Transactions on Signal Processing*, 64(6):1609–1620, 2016. doi:10.1109/TSP.2015.2500192.

Olivier Catoni. Challenging the empirical mean and empirical variance: A deviation study. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 48(4):1148 – 1185, 2012. doi:10.1214/11-AIHP454.

Yudong Chen and Yuejie Chi. Harnessing Structures in Big Data via Guaranteed Low-Rank Matrix Estimation: Recent Theory and Fast Algorithms via Convex and Nonconvex Optimization. *IEEE Signal Processing Magazine*, 35(4):14–31, 2018. doi:10.1109/MSP.2018.2821706.

Wei Chu and Seung-Taek Park. Personalized recommendation on dynamic content using predictive bilinear models. In *Proceedings of the 18th International Conference on World Wide Web*, WWW '09, page 691–700, New York, NY, USA, 2009. Association for Computing Machinery. ISBN 9781605584874. doi:10.1145/1526709.1526802.

Cyrille W. Combettes and Sebastian Pokutta. Revisiting the approximate Carathéodory problem via the Frank-Wolfe algorithm. *Mathematical Programming*, 197(1):191–214, Jan 2023. ISSN 1436-4646. doi:10.1007/s10107-021-01735-x.

Varun Dani, Thomas P. Hayes, and Sham M. Kakade. Stochastic Linear Optimization under Bandit Feedback. In *Proceedings of the 21st Annual Conference on Learning Theory (COLT)*, pages 355–366. Omnipress, 2008. URL https://homes.cs.washington.edu/~sham/papers/ml/bandit_linear_long.pdf.

Nirjhar Das, Souradip Chakraborty, Aldo Pacchiano, and Sayak Ray Chowdhury. Active Preference Optimization for Sample Efficient RLHF. In *Machine Learning and Knowledge Discovery in Databases. Research Track*, Cham, 2025. Springer Nature Switzerland. URL https://arxiv.org/abs/2402.10500.

Mark A. Davenport and Justin Romberg. An Overview of Low-Rank Matrix Recovery From Incomplete Observations. *IEEE Journal of Selected Topics in Signal Processing*, 10(4):608–622, 2016. doi:10.1109/JSTSP.2016.2539100.

Mark A. Davenport, Yaniv Plan, Ewout van den Berg, and Mary Wootters. 1-Bit matrix completion. *Information and Inference: A Journal of the IMA*, 3(3): 189–223, 2014. doi:10.1093/imaiai/iau006.

Steven Diamond and Stephen Boyd. CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research*, 17(83):1–5, 2016. URL https://www.jmlr.org/papers/v17/15-408.html.

Jianqing Fan, Wenyan Gong, and Ziwei Zhu. Generalized high-dimensional trace regression via nuclear norm regularization. *Journal of Econometrics*, 212(1):177–202, 2019. ISSN 0304-4076. doi:10.1016/j.jeconom.2019.04.026.

Louis Faury, Marc Abeille, Clément Calauzènes, and Olivier Fercoq. Improved Optimistic Algorithms for Logistic Bandits. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 3052–3060. PMLR, 13–18 Jul 2020. URL https://proceedings.mlr.press/v119/faury20a.html.

Louis Faury, Marc Abeille, Kwang-Sung Jun, and Clément Calauzènes. Jointly Efficient and Optimal Algorithms for Logistic Bandits. In *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pages 546–580. PMLR, 28–30 Mar 2022. URL https://proceedings.mlr.press/v151/faury22a.html.

Sarah Filippi, Olivier Cappe, Aurélien Garivier, and Csaba Szepesvári. Parametric Bandits: The Generalized Linear Case. In *Advances in Neural Information Processing Systems*, volume 23, pages 586–594. Curran Associates, Inc., 2010. URL https://sites.ualberta.ca/~szepesva/papers/GenLinBandits-NeurIPS2010.pdf.

Stefano Fortunati, Fulvio Gini, Maria S. Greco, and Christ D. Richmond. Performance Bounds for Parameter Estimation under Misspecified Models: Fundamental Findings and Applications. *IEEE Signal Processing Magazine*, 34(6):142–157, 2017. doi:10.1109/MSP.2017.2738017.

Marguerite Frank and Philip Wolfe. An algorithm for quadratic programming. *Naval Research Logistics Quarterly*, 3(1-2):95–110, 1956. doi:10.1002/nav.3800030109.

Stephan Ramon Garcia, Ryan O'Loughlin, and Jiahui Yu. Symmetric and antisymmetric tensor products for the function-theoretic operator theorist. *Canadian Journal of Mathematics*, page 1–23, 2023. doi:10.4153/S0008414X23000901.

Rong Ge, Chi Jin, and Yi Zheng. No Spurious Local Minima in Nonconvex Low Rank Problems: A Unified Geometric Analysis. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages

1233–1242. PMLR, 06–11 Aug 2017. URL https://proceedings.mlr.press/v70/ge17a.html.

E. N. Gilbert. A comparison of signalling alphabets. *The Bell System Technical Journal*, 31(3):504–522, 1952. doi:10.1002/j.1538-7305.1952.tb01393.x.

David F. Gleich and Lek-heng Lim. Rank Aggregation via Nuclear Norm Minimization. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '11, page 60–68, New York, NY, USA, 2011. Association for Computing Machinery. doi:10.1145/2020408.2020425.

Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, 4 edition, 2013.

Suriya Gunasekar, Pradeep Ravikumar, and Joydeep Ghosh. Exponential Family Matrix Completion under Structural Constraints. In *Proceedings of the 31st International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, pages 1917–1925, Bejing, China, 22–24 Jun 2014. PMLR. URL https://proceedings.mlr.press/v32/gunasekar14.html.

Botao Hao, Tor Lattimore, and Mengdi Wang. High-Dimensional Sparse Linear Bandits. In *Advances in Neural Information Processing Systems*, volume 33, pages 10753–10763. Curran Associates, Inc., 2020. URL https://arxiv.org/abs/2011.04020.

Juha Heikkinen and Elja Arjas. Modeling a Poisson Forest in Variable Elevations: A Nonparametric Bayesian Approach. *Biometrics*, 55(3):738–745, 1999. doi:10.1111/j.0006-341X.1999.00738.x.

A. J. Hoffman and H. W. Wielandt. The variation of the spectrum of a normal matrix. *Duke Mathematical Journal*, 20(1):37 – 39, 1953. doi:10.1215/S0012-7094-53-02004-3.

Kenneth M. Hoffman and Ray Kunze. *Linear Algebra*. Prentice Hall, 2 edition, 1971.

Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, 2 edition, 2012. doi:10.1017/CBO9781139020411.

Baihe Huang, Kaixuan Huang, Sham Kakade, Jason D Lee, Qi Lei, Runzhe Wang, and Jiaqi Yang. Optimal Gradient-based Algorithms for Non-concave Bandit Optimization. In *Advances in Neural Information Processing Systems*, volume 34, pages 29101–29115. Curran Associates, Inc., 2021. URL https://openreview.net/forum?id=7SGgWl2uVG-.

Kevin Jamieson, Sumeet Katariya, Atul Deshpande, and Robert Nowak. Sparse Dueling Bandits. In *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics*, volume 38 of *Proceedings of Machine Learning Research*, pages 416–424, San Diego, California, USA, 09–12 May 2015. PMLR. URL https://proceedings.mlr.press/v38/jamieson15.html.

Kyoungseok Jang, Kwang-Sung Jun, Se-Young Yun, and Wanmo Kang. Improved Regret Bounds of Bilinear Bandits using Action Space Analysis. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 4744–4754. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/jang21a.html.

Kyoungseok Jang, Chicheng Zhang, and Kwang-Sung Jun. PopArt: Efficient Sparse Regression and Experimental Design for Optimal Sparse Linear Bandits. In *Advances in Neural Information Processing Systems*, volume 35, pages 2102–2114. Curran Associates, Inc., 2022. URL https://openreview.net/forum?id=GWcdXz0M6a.

Kyoungseok Jang, Chicheng Zhang, and Kwang-Sung Jun. Efficient Low-Rank Matrix Estimation, Experimental Design, and Arm-Set-Dependent Low-Rank Bandits. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 21329–21372. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/jang24e.html.

Yassir Jedra, William Réveillard, Stefan Stojanovic, and Alexandre Proutiere. Low-Rank Bandits via Tight Two-to-Infinity Singular Subspace Recovery. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 21430–21485. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/jedra24a.html.

Xiaoye Jiang, Lek-Heng Lim, Yuan Yao, and Yinyu Ye. Statistical ranking and combinatorial Hodge theory. *Mathematical Programming*, 127(1):203–244, 2011. doi:10.1007/s10107-010-0419-x.

Kwang-Sung Jun, Aniruddha Bhargava, Robert Nowak, and Rebecca Willett. Scalable Generalized Linear Bandits: Online Computation and Hashing. In *Advances in Neural Information Processing Systems*, volume 30, pages 98–108. Curran Associates, Inc., 2017. URL https://arxiv.org/abs/1706.00136.

Kwang-Sung Jun, Rebecca Willett, Stephen Wright, and Robert Nowak. Bilinear Bandits with Low-rank Structure. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 3163–3172. PMLR, 09–15 Jun 2019. URL https://proceedings.mlr.press/v97/jun19a.html.

Kwang-Sung Jun, Lalit Jain, Blake Mason, and Houssam Nassif. Improved Confidence Bounds for the Linear Logistic Model and Applications to Bandits. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 5148–5157. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/jun21a.html.

Yue Kang, Cho-Jui Hsieh, and Thomas Chun Man Lee. Efficient Frameworks for Generalized Low-Rank Matrix Bandit Problems. In *Advances in Neural Information Processing Systems*, volume 35, pages 19971–19983. Curran Associates, Inc., 2022. URL https://arxiv.org/abs/2401.07298.

Sumeet Katariya, Branislav Kveton, Csaba Szepesvari, Claire Vernade, and Zheng Wen. Stochastic Rank-1 Bandits. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pages 392–401. PMLR, 20–22 Apr 2017a. URL https://proceedings.mlr.press/v54/katariya17a.html.

Sumeet Katariya, Branislav Kveton, Csaba Szepesvári, Claire Vernade, and Zheng Wen. Bernoulli Rank-1 Bandits for Click Feedback. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pages 2001–2007, 2017b. doi:10.24963/ijcai.2017/278.

J. Kiefer and J. Wolfowitz. The Equivalence of Two Extremum Problems. *Canadian Journal of Mathematics*, 12:363–366, 1960. doi:10.4153/CJM-1960-030-4.

Daesung Kim and Hye Won Chung. Rank-1 Matrix Completion with Gradient Descent and Small Random Initialization. In *Advances in Neural Information Processing Systems*, volume 36, pages 10530–10566. Curran Associates, Inc., 2023. URL https://openreview.net/forum?id=qjqJL2lfkH.

Jung-hun Kim and Milan Vojnović. Scheduling Servers with Stochastic Bilinear Rewards. *arXiv preprint arXiv:2112.06362*, 2021. URL https://arxiv.org/abs/2112.06362.

J. F. C. Kingman. *Poisson Processes*, volume 3 of *Oxford Studies in Probability*. Oxford University Press, 1992.

Olga Klopp. Noisy low-rank matrix completion with general sampling distribution. *Bernoulli*, 20(1):282 – 303, 2014. doi:10.3150/12-BEJ486.

Olga Klopp, Jean Lafond, Éric Moulines, and Joseph Salmon. Adaptive multinomial matrix completion. *Electronic Journal of Statistics*, 9(2):2950 – 2975, 2015. doi:10.1214/15-EJS1093.

Vladimir Koltchinskii, Karim Lounici, and Alexandre B. Tsybakov. Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. *The Annals of Statistics*, 39(5):2302 – 2329, 2011. doi:10.1214/11-AOS894.

Wojciech Kotłowski and Gergely Neu. Bandit Principal Component Analysis. In *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pages 1994–2024. PMLR, 25–28 Jun 2019. URL https://proceedings.mlr.press/v99/kotlowski19a.html.

Wataru Kumagai. Regret Analysis for Continuous Dueling Bandit. In *Advances in Neural Information Processing Systems*, volume 30, pages 1488–1497. Curran Associates, Inc., 2017. URL https://arxiv.org/abs/1711.07693.

Jean Lafond. Low Rank Matrix Completion with Exponential Family Noise. In *Proceedings of The 28th Conference on Learning Theory*, volume 40 of *Proceedings of Machine Learning Research*, pages 1224–1243, Paris, France, 03–06 Jul 2015. PMLR. URL https://proceedings.mlr.press/v40/Lafond15.html.

Jean Lafond, Olga Klopp, Éric Moulines, and Joseph Salmon. Probabilistic low-rank matrix completion on finite alphabets. In *Advances in Neural Information Processing Systems*, volume 27, pages 1727–1735. Curran Associates, Inc., 2014. URL https://arxiv.org/abs/1412.2632.

Tor Lattimore and Botao Hao. Bandit Phase Retrieval. In *Advances in Neural Information Processing Systems*, volume 34, pages 18801–18811. Curran Associates, Inc., 2021. URL https://openreview.net/forum?id=fThfMoV7Ri.

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

Michel Ledoux and Michel Talagrand. *Probability in Banach Spaces: Isoperimetry and Processes*. Classics in Mathematics. Springer Berlin, Heidelberg, 1991.

John M. Lee. *Introduction to Smooth Manifolds*, volume 218 of *Graduate Texts in Mathematics*. Springer New York, NY, 2 edition, 2012. doi:10.1007/978-1-4419-9982-5.

Junghyun Lee, Se-Young Yun, and Kwang-Sung Jun. A Unified Confidence Sequence for Generalized Linear Models, with Applications to Bandits. In *Advances in Neural Information Processing Systems*, volume 37. Curran Associates, Inc., 2024a. URL https://openreview.net/forum?id=MDdOQayWTA.

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 of *Proceedings of Machine Learning Research*, pages 4474–4482. PMLR, 02–04 May 2024b. URL https://proceedings.mlr.press/v238/lee24d.html.

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*, WWW '10, page 661–670, New York, NY, USA, 2010. Association for Computing Machinery. ISBN 9781605587998. doi:10.1145/1772690.1772758.

Lihong Li, Wei Chu, John Langford, Taesup Moon, and Xuanhui Wang. An Unbiased Offline Evaluation of Contextual Bandit Algorithms with Generalized Linear Models. In *Proceedings of the Workshop on On-line Trading of Exploration and Exploitation 2*, volume 26 of *Proceedings of Machine Learning Research*, pages 19–36, Bellevue, Washington, USA, 02 Jul 2012. PMLR. URL https://proceedings.mlr.press/v26/li12a.html.

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 of *Proceedings of Machine Learning Research*, pages 2071–2080. PMLR, 06–11 Aug 2017. URL https://proceedings.mlr.press/v70/li17c.html.

Wenjie Li, Adarsh Barik, and Jean Honorio. A Simple Unified Framework for High Dimensional Bandit Problems. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 12619–12655. PMLR, 17–23 Jul 2022. URL https://proceedings.mlr.press/v162/li22a.html.

Shuai Liu, Alex Ayoub, Flore Sentenac, Xiaoqi Tan, and Csaba Szepesvári. Almost Free: Self-concordance in Natural Exponential Families and an Application to Bandits. In *Advances in Neural Information Processing Systems*, volume 37. Curran Associates, Inc., 2024. URL https://openreview.net/forum?id=LKwVYvx66I.

Yangyi Lu, Amirhossein Meisami, and Ambuj Tewari. Low-Rank Generalized Linear Bandit Problems. In *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pages 460–468. PMLR, 13–15 Apr 2021. URL https://proceedings.mlr.press/v130/lu21a.html.

Yu Lu and Sahand N. Negahban. Individualized rank aggregation using nuclear norm regularization. In *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, volume 53, pages 1473–1479, 2015. doi:10.1109/ALLERTON.2015.7447183.

Yunan Luo, Xinbin Zhao, Jingtian Zhou, Jinglin Yang, Yanqing Zhang, Wenhua Kuang, Jian Peng, Ligong Chen, and Jianyang Zeng. A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. *Nature Communications*, 8(1): 573, Sep 2017. ISSN 2041-1723. doi:10.1038/s41467-017-00680-8.

Jianhao Ma and Salar Fattahi. Can Learning Be Explained by Local Optimality in Robust Low-Rank Matrix Recovery? *Mathematics of Operations Research*, 0(0):null, 2025. doi:10.1287/moor.2023.0371.

Blake Mason, Kwang-Sung Jun, and Lalit Jain. An Experimental Design Approach for Regret Minimization in Logistic Bandits. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(7): 7736–7743, Jun. 2022. doi:10.1609/aaai.v36i7.20741. URL https://arxiv.org/abs/2202.02407.

Kenneth O. May. Intransitivity, Utility, and the Aggregation of Preference Patterns. *Econometrica*, 22(1): 1–13, 1954. doi:10.2307/1909827.

Peter McCullagh and John A. Nelder. *Generalized Linear Models*. Monographs on Statistics and Applied Probability. Chapman & Hall/CRC, 2 edition, 1989. doi:10.1201/9780203753736.

H. Brendan McMahan, Gary Holt, D. Sculley, Michael Young, Dietmar Ebner, Julian Grady, Lan Nie, Todd Phillips, Eugene Davydov, Daniel Golovin, Sharat Chikkerur, Dan Liu, Martin Wattenberg, Arnar Mar Hrafnkelsson, Tom Boulos, and Jeremy Kubica. Ad Click Prediction: a View from the Trenches. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '13, page 1222–1230, New York, NY, USA, 2013. Association for Computing Machinery. ISBN 9781450321747. doi:10.1145/2487575.2488200.

Andrew D McRae and Mark A Davenport. Low-rank matrix completion and denoising under Poisson noise. *Information and Inference: A Journal of the IMA*, 10(2):697–720, 08 2020. ISSN 2049-8772. doi:10.1093/imaiai/iaaa020.

Aditya Krishna Menon and Charles Elkan. Link Prediction via Matrix Factorization. In *Machine Learning and Knowledge Discovery in Databases*, pages 437–452, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg. ISBN 978-3-642-23783-6. URL https://link.springer.com/chapter/10.1007/978-3-642-23783-6_28.

Thomas P. Minka. Old and New Matrix Algebra Useful for Statistics, December 1997. URL https:

//tminka.github.io/papers/matrix/. MIT Media Lab note, 1997; revised 12/00.

Stanislav Minsker. Sub-Gaussian estimators of the mean of a random matrix with heavy-tailed entries. *The Annals of Statistics*, 46(6A):2871 – 2903, 2018. doi:10.1214/17-AOS1642.

Vahab Mirrokni, Renato Paes Leme, Adrian Vladu, and Sam Chiu wai Wong. Tight Bounds for Approximate Carathéodory and Beyond. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 2440–2448. PMLR, 06–11 Aug 2017. URL https://proceedings.mlr.press/v70/mirrokni17a.html.

James R. Munkres. *Topology*. Pearson Modern Classics. Pearson, 2 edition, 2018.

Rémi Munos, Michal Valko, Daniele Calandriello, Mohammad Gheshlaghi Azar, Mark Rowland, Zhaohan Daniel Guo, Yunhao Tang, Matthieu Geist, Thomas Mesnard, Côme Fiegel, Andrea Michi, Marco Selvi, Sertan Girgin, Nikola Momchev, Olivier Bachem, Daniel J Mankowitz, Doina Precup, and Bilal Piot. Nash Learning from Human Feedback. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 36743–36768. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/munos24a.html.

F. D. Murnaghan and A. Wintner. A Canonical Form for Real Matrices under Orthogonal Transformations. *Proceedings of the National Academy of Sciences*, 17 (7):417–420, 1931. doi:10.1073/pnas.17.7.417.

Mojmir Mutný and Andreas Krause. No-regret Algorithms for Capturing Events in Poisson Point Processes. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 7894–7904. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/mutny21a.html.

Sahand Negahban and Martin J. Wainwright. Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *The Annals of Statistics*, 39(2):1069 – 1097, 2011. doi:10.1214/10-AOS850.

Yu E Nesterov. Polynomial time methods in linear and quadratic programming. *Izvestija AN SSR Tekhnitcheskaya Kibernetika*, 3:324–326, 1988. (In Russian).

Yurii Nesterov. Gradient methods for minimizing composite objective function. LIDAM Discussion Papers CORE 2007076, UniversitÃ© catholique de Louvain, Center for Operations Research and Econometrics (CORE), Sep 2007. URL https://ideas.repec.org/p/cor/louvco/2007076.html.

Yurii Nesterov and Arkadii Nemirovski. *Interior-Point Polynomial Algorithms in Convex Programming*, volume 13 of *SIAM Studies in Applied Mathematics*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1994.

Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. A Three-Way Model for Collective Learning on Multi-Relational Data. In *Proceedings of the 28th International Conference on International Conference on Machine Learning*, pages 809—816, 2011. URL https://icml.cc/2011/papers/438_icmlpaper.pdf.

Frank Nielsen. An Elementary Introduction to Information Geometry. *Entropy*, 22(10), Sep 2020. ISSN 1099-4300. doi:10.3390/e22101100.

Brendan O'Donoghue and Emmanuel Candès. Adaptive Restart for Accelerated Gradient Schemes. *Foundations of Computational Mathematics*, 15(3):715–732, 2015. doi:10.1007/s10208-013-9150-3.

Dmitrii M. Ostrovskii and Francis Bach. Finite-sample analysis of M-estimators using self-concordance. *Electronic Journal of Statistics*, 15(1):326 – 391, 2021. doi:10.1214/20-EJS1780.

Samet Oymak, Amin Jalali, Maryam Fazel, Yonina C. Eldar, and Babak Hassibi. Simultaneously Structured Models With Application to Sparse and Low-Rank Matrices. *IEEE Transactions on Information Theory*, 61(5):2886–2908, 2015. doi:10.1109/TIT.2015.2401574.

Dohyung Park, Anastasios Kyrillidis, Constantine Carmanis, and Sujay Sanghavi. Non-square matrix sensing without spurious local minima via the Burer-Monteiro approach. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pages 65–74. PMLR, 20–22 Apr 2017. URL https://proceedings.mlr.press/v54/park17a.html.

Carolin Penke, Andreas Marek, Christian Vorwerk, Claudia Draxl, and Peter Benner. High performance solution of skew-symmetric eigenvalue problems with applications in solving the Bethe-Salpeter eigenvalue problem. *Parallel Computing*, 96:102639, 2020. ISSN 0167-8191. doi:10.1016/j.parco.2020.102639.

Friedrich Pukelsheim. *Optimal Design of Experiments*, volume 50 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), 2006.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Stefano Ermon, Christopher D. Manning, and Chelsea Finn. Direct Preference Optimization: Your Language Model is Secretly a Reward Model. In *Advances in Neural Information Processing Systems*,

volume 36. Curran Associates, Inc., 2023. URL https://openreview.net/forum?id=HPuSIXJaa9.

Arun Rajkumar and Shivani Agarwal. When can we rank well from comparisons of $O(n \log(n))$ non-actively chosen pairs? In *29th Annual Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, pages 1376–1401, Columbia University, New York, New York, USA, 23–26 Jun 2016. PMLR. URL https://proceedings.mlr.press/v49/rajkumar16.html.

Emile Richard, Pierre-André Savalle, and Nicolas Vayatis. Estimation of Simultaneously Sparse and Low Rank Matrices. In *Proceedings of the 29th International Coference on International Conference on Machine Learning*, ICML'12, page 51–58, Madison, WI, USA, 2012. Omnipress. ISBN 9781450312851. URL https://icml.cc/2012/papers/674.pdf.

Matthew Richardson, Ewa Dominowska, and Robert Ragno. Predicting Clicks: Estimating the Click-Through Rate for New Ads. In *Proceedings of the 16th International Conference on World Wide Web*, WWW '07, page 521–530, New York, NY, USA, 2007. Association for Computing Machinery. ISBN 9781595936547. doi:10.1145/1242572.1242643.

Philippe Rigollet and Alexandre Tsybakov. Exponential Screening and optimal rates of sparse estimation. *The Annals of Statistics*, 39(2):731 – 771, 2011. doi:10.1214/10-AOS854.

James M. Robins, Andrea Rotnitzky, and Lue Ping Zhao and. Estimation of Regression Coefficients When Some Regressors are not Always Observed. *Journal of the American Statistical Association*, 89(427):846–866, 1994. doi:10.1080/01621459.1994.10476818.

Angelika Rohde and Alexandre B. Tsybakov. Estimation of high-dimensional low-rank matrices. *The Annals of Statistics*, 39(2):887 – 930, 2011. doi:10.1214/10-AOS860.

Yoan Russac, Louis Faury, Olivier Cappé, and Aurélien Garivier. Self-Concordant Analysis of Generalized Linear Bandits with Forgetting. In *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pages 658–666. PMLR, 13–15 Apr 2021. URL https://proceedings.mlr.press/v130/russac21a.html.

Aadirupa Saha. Optimal Algorithms for Stochastic Contextual Preference Bandits. In *Advances in Neural Information Processing Systems*, volume 34, pages 30050–30062. Curran Associates, Inc., 2021. URL https://openreview.net/forum?id=1lCZrXJBpM.

Aadirupa Saha, Tomer Koren, and Yishay Mansour. Adversarial Dueling Bandits. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 9235–9244. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/saha21a.html.

Ayush Sawarni, Nirjhar Das, Siddharth Barman, and Gaurav Sinha. Generalized Linear Bandits with Limited Adaptivity. In *Advances in Neural Information Processing Systems*, volume 37. Curran Associates, Inc., 2024. URL https://arxiv.org/abs/2404.06831.

Flore Sentenac, Jialin Yi, Clement Calauzenes, Vianney Perchet, and Milan Vojnović. Pure Exploration and Regret Minimization in Matching Bandits. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 9434–9442. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/sentenac21a.html.

Burr Settles. *Active Learning*. Number 1 in Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool Publishers, 2012. doi:10.2200/S00429ED1V01Y201207AIM018.

Shinichiro Shirota and Alan E. Gelfand. Space and circular time log Gaussian Cox processes with application to crime event data. *The Annals of Applied Statistics*, 11(2):481 – 503, 2017. doi:10.1214/16-AOAS960.

Max Simchowitz and Dylan J. Foster. Naive Exploration is Optimal for Online LQR. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 8937–8948. PMLR, 13–18 Jul 2020. URL https://proceedings.mlr.press/v119/simchowitz20a.html.

Nathan Srebro and Russ R Salakhutdinov. Collaborative Filtering in a Non-Uniform World: Learning with the Weighted Trace Norm. In *Advances in Neural Information Processing Systems*, volume 23, pages 2056–2064. Curran Associates, Inc., 2010. URL https://proceedings.neurips.cc/paper_files/paper/2010/file/67d96d458abdef21792e6d8e590244e7-Paper.pdf.

Charles Stein, Persi Diaconis, Susan Holmes, and Gesine Reinert. Use of Exchangeable Pairs in the Analysis of Simulations. In Persi Diaconis and Susan Holmes, editors, *Stein's Method: Expository Lectures and Applications*, volume 46 of *Institute of Mathematical Statistics Lecture Notes - Monograph Series*, chapter 1, pages 1–25. Institute of Mathematical Statistics, 2004. doi:10.1214/lnms/1196283797.

David H. Stern, Ralf Herbrich, and Thore Graepel. Matchbox: Large Scale Online Bayesian Recommen-

dations. In *Proceedings of the 18th International Conference on World Wide Web*, WWW '09, page 111–120, New York, NY, USA, 2009. Association for Computing Machinery. ISBN 9781605584874. doi:10.1145/1526709.1526725.

Dominik Stöger and Mahdi Soltanolkotabi. Small random initialization is akin to spectral learning: Optimization and generalization guarantees for overparameterized low-rank matrix reconstruction. In *Advances in Neural Information Processing Systems*, volume 34, pages 23831–23843. Curran Associates, Inc., 2021. URL https://openreview.net/forum?id=rsRq--gsiE.

Stefan Stojanovic, Yassir Jedra, and Alexandre Proutière. Spectral Entry-wise Matrix Estimation for Low-Rank Reinforcement Learning. In *Advances in Neural Information Processing Systems*, volume 36, pages 77056–77070. Curran Associates, Inc., 2023. URL https://openreview.net/forum?id=aDLmRMb0K9.

Volker Strassen. Gaussian Elimination is not Optimal. *Numerische Mathematik*, 13(4):354–356, 1969. doi:10.1007/BF02165411.

Student. The Probable Error of a Mean. *Biometrika*, 6(1):1–25, 1908. doi:10.1093/biomet/6.1.1.

Yanan Sui, Masrour Zoghi, Katja Hofmann, and Yisong Yue. Advancements in Dueling Bandits. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*, pages 5502–5510. International Joint Conferences on Artificial Intelligence Organization, 7 2018. doi:10.24963/ijcai.2018/776.

Gokul Swamy, Christoph Dann, Rahul Kidambi, Steven Wu, and Alekh Agarwal. A Minimaximalist Approach to Reinforcement Learning from Human Feedback. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 47345–47377. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/swamy24a.html.

Batoul Taki, Mohsen Ghassemi, Anand D. Sarwate, and Waheed U. Bajwa. A Minimax Lower Bound for Low-Rank Matrix-Variate Logistic Regression. In *2021 55th Asilomar Conference on Signals, Systems, and Computers*, volume 55, pages 477–484, 2021. doi:10.1109/IEEECONF53345.2021.9723149.

Michael J. Todd. *Minimum-Volume Ellipsoids: Theory and Algorithms*. MOS-SIAM Series on Optimization. SIAM-Society for Industrial and Applied Mathematics, 2016.

Cindy Trinh, Emilie Kaufmann, Claire Vernade, and Richard Combes. Solving Bernoulli Rank-One Bandits with Unimodal Thompson Sampling. In *Proceedings of the 31st International Conference on Algorithmic Learning Theory*, volume 117 of *Proceedings of Machine Learning Research*, pages 862–889. PMLR, 08 Feb–11 Feb 2020. URL https://proceedings.mlr.press/v117/trinh20a.html.

Joel A. Tropp. User-Friendly Tail Bounds for Sums of Random Matrices. *Foundations of Computational Mathematics*, 12(4):389–434, 2012. doi:10.1007/s10208-011-9099-z.

Joel A. Tropp. An Introduction to Matrix Concentration Inequalities. *Foundations and Trends® in Machine Learning*, 8(1-2):1–230, 2015. ISSN 1935-8237. doi:10.1561/2200000048.

Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer Series in Statistics. Springer New York, 2009.

Amos Tversky. Intransitivity of preferences. *Psychological Review*, 76(1):31–48, 1969. doi:10.1037/h0026750.

Rom Rubenovich Varshamov. Estimation of number of signals in codes with correction of non-symmetric errors. *Avtomatika i Telemekhanika*, 25(11):1628–1629, 1964. URL https://www.mathnet.ru/php/archive.phtml?wshow=paper&jrnid=at&paperid=11783.

Andrew J Wagenmaker, Yifang Chen, Max Simchowitz, Simon Du, and Kevin Jamieson. First-Order Regret in Reinforcement Learning with Linear Function Approximation: A Robust Estimation Approach. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 22384–22429. PMLR, 17–23 Jul 2022. URL https://proceedings.mlr.press/v162/wagenmaker22a.html.

Martin J. Wainwright. *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2019.

Stephen G. Walker. Bayesian inference with misspecified models. *Journal of Statistical Planning and Inference*, 143(10):1621–1633, 2013. ISSN 0378-3758. doi:10.1016/j.jspi.2013.05.013.

R. C. Ward and L. J. Gray. Eigensystem Computation for Skew-Symmetric and a Class of Symmetric Matrices. *ACM Transactions on Mathematical Software*, 4(3):278–285, September 1978. ISSN 0098-3500. doi:10.1145/355791.355798.

Halbert White. Maximum Likelihood Estimation of Misspecified Models. *Econometrica*, 50(1):1–25, 1982. URL http://www.jstor.org/stable/1912526.

Yue Wu, Tao Jin, Qiwei Di, Hao Lou, Farzad Farnoud, and Quanquan Gu. Borda Regret Minimization for

Generalized Linear Dueling Bandits. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 53571–53596. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/wu24m.html.

Wei Xiong, Hanze Dong, Chenlu Ye, Ziqi Wang, Han Zhong, Heng Ji, Nan Jiang, and Tong Zhang. Iterative Preference Learning from Human Feedback: Bridging Theory and Practice for RLHF under KL-constraint. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 54715–54754. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/xiong24a.html.

Baturalp Yalçın, Haixiang Zhang, Javad Lavaei, and Somayeh Sojoudi. Factorization Approach for Low-complexity Matrix Completion Problems: Exponential Number of Spurious Solutions and Failure of Gradient Methods. In *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pages 319–341. PMLR, 28–30 Mar 2022. URL https://proceedings.mlr.press/v151/yalcin22a.html.

Eunho Yang and Pradeep K Ravikumar. Dirty Statistical Models. In *Advances in Neural Information Processing Systems*, volume 26, pages 611–619. Curran Associates, Inc., 2013. URL https://papers.nips.cc/paper_files/paper/2013/hash/8bf1211fd4b7b94528899de0a43b9fb3-Abstract.html.

Wei Hong Yang, Lei-Hong Zhang, and Ruyi Song. Optimality conditions for the nonlinear programming problems on Riemannian manifolds. *Pacific Journal of Optimization*, 10:415–434, 2014. URL http://www.optimization-online.org/DB_FILE/2012/07/3535.pdf.

Yuhong Yang and Andrew Barron. Information-theoretic determination of minimax rates of convergence. *The Annals of Statistics*, 27(5):1564 – 1599, 1999. doi:10.1214/aos/1017939142.

D. C. Youla. A Normal form for a Matrix under the Unitary Congruence Group. *Canadian Journal of Mathematics*, 13:694–704, 1961. doi:10.4153/CJM-1961-059-8.

Yisong Yue and Thorsten Joachims. Interactively Optimizing Information Retrieval Systems as a Dueling Bandits Problem. In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, page 1201–1208, New York, NY, USA, 2009. Association for Computing Machinery. ISBN 9781605585161. doi:10.1145/1553374.1553527.

Yisong Yue, Josef Broder, Robert Kleinberg, and Thorsten Joachims. The K-armed dueling bandits problem. *Journal of Computer and System Sciences*, 78(5):1538–1556, 2012. ISSN 0022-0000. doi:10.1016/j.jcss.2011.12.028. JCSS Special Issue: Cloud Computing 2011.

Haixiang Zhang, Baturalp Yalcin, Javad Lavaei, and Somayeh Sojoudi. A new complexity metric for nonconvex rank-one generalized matrix completion. *Mathematical Programming*, 207(1):227–268, 2024. doi:10.1007/s10107-023-02008-5.

Kelly Zhang, Lucas Janson, and Susan Murphy. Statistical Inference with M-Estimators on Adaptively Collected Data. In *Advances in Neural Information Processing Systems*, volume 34, pages 7460–7471. Curran Associates, Inc., 2021. URL https://openreview.net/forum?id=TJOQw_vMlAj.

Yifan Zhang, Ge Zhang, Yue Wu, Kangping Xu, and Quanquan Gu. Beyond Bradley-Terry Models: A General Preference Model for Language Model Alignment. In *Proceedings of the 42nd International Conference on Machine Learning*, Proceedings of Machine Learning Research. PMLR, 13–19 Jul 2025. URL https://openreview.net/forum?id=UgKcpPE0ZX.

Junlong Zhao and Chenlei Leng. Structured Lasso for Regression with Matrix Covariates. *Statistica Sinica*, 24:799–814, 2014. doi:10.5705/ss.2012.033.

Junlong Zhao, Lu Niu, and Shushi Zhan. Trace regression model with simultaneously low rank and row(column) sparse parameter. *Computational Statistics & Data Analysis*, 116:1–18, 2017. ISSN 0167-9473. doi:10.1016/j.csda.2017.06.009.

Han Zhong, Guhao Feng, Wei Xiong, Xinle Cheng, Li Zhao, Di He, Jiang Bian, and Liwei Wang. DPO Meets PPO: Reinforced Token Optimization for RLHF. In *Proceedings of the 42nd International Conference on Machine Learning*, Proceedings of Machine Learning Research. PMLR, 13–19 Jul 2025. URL https://openreview.net/forum?id=IfWKVF6LfY.

Tijana Zrnic and Emmanuel Candès. Active Statistical Inference. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 62993–63010. PMLR, 21–27 Jul 2024. URL https://openreview.net/forum?id=GKMcCtWC7H.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes] Section 2,3,5
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes] Section 3,5
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Yes] Appendix M.

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. [Yes] Section 2,3,4,5
   (b) Complete proofs of all theoretical results. [Yes] The whole Appendix
   (c) Clear explanations of any assumptions. [Yes] Section 2

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Yes] Appendix M.
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Yes] Appendix M.
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Yes] Appendix M.
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [No] Simple CPU experiments.

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. [Yes] Appendix M.
   (b) The license information of the assets, if applicable. [Not Applicable]
   (c) New assets either in the supplemental material or as a URL, if applicable. [Yes] Appendix J
   (d) Information about consent from data providers/curators. [Not Applicable]
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:
   (a) The full text of instructions given to participants and screenshots. [Not Applicable]
   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]
   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

Contents

1 INTRODUCTION **1**

2 TECHNICAL PRELIMINARIES **2**

3 GL-LowPopArt: A MATRIX CATONI-STYLE ESTIMATOR **3**

3.1 Overview of GL-LowPopArt . . . 3

3.2 GL($\pi; \mathbf{\Theta}_\star$): A New Experimental Design . . . 4

3.3 Comparison with Prior Works . . . 5

3.4 Proof Sketch of Theorem 3.1 . . . 6

3.5 Stage I via Nuclear Penalized MLE . . . 6

4 LOCAL MINIMAX LOWER BOUND FOR $\|\cdot\|_F$ **7**

5 APPLICATIONS OF GL-LowPopArt **8**

5.1 Generalized Linear Matrix Completion under Uniform Sampling at Random . . . 8

5.2 Bilinear Dueling Bandits . . . 8

5.2.1 Problem Description . . . 8

5.2.2 BETC-GLM-LR and Regret Upper Bound . . . 9

6 CONCLUSION **9**

A RELATED WORK **23**

B TABLE OF NOTATION **25**

C PROOF OF THEOREM 3.1 – ERROR BOUND OF STAGE II **26**

C.1 Proof of Lemma C.2 – Controlling the Matrix Variance Statistics . . . 28

C.2 A Glimpse of Unbounded Arm-Set: The Linear Case . . . 29

D PROOF OF PROPOSITION 3.1 – UPPER AND LOWER BOUNDS OF GL **30**

D.1 Upper Bound of Definition 3.1 . . . 30

D.2 Lower Bound of GL($\pi; \mathbf{\Theta}_\star$) . . . 30

E DEFERRED DETAILS REGARDING OPTIMAL GL-DESIGN **31**

E.1 Proof of Proposition 3.2 . . . 31

F PROOF OF THEOREM 3.2 – ERROR BOUND OF STAGE I **32**

F.1 Proof of Lemma F.1: Restricted Strong Convexity . . . 33

F.2 Proof of Lemma F.3: Uniform Bounding of $\mathcal{E}_N$ . . . 34

F.3 Proof of Proposition F.1: Choosing λ_N . . . 35

F.3.1 Proof of (i) – GLM bounded by M . . . 35
F.3.2 Proof of (ii) – GLM with Finite SubExponential Norm . . . 36
F.3.3 Proof of Lemma F.8 . . . 36

G SUMMARY OF PRIOR WORKS ON GENERALIZED TRACE REGRESSION 38
G.1 Nuclear Penalized MLE of Fan et al. (2019) . . . 38
G.2 Stein's Lemma-based Estimator of Kang et al. (2022) . . . 38
G.3 Nuclear Penalized MLE of Kang et al. (2022) . . . 39

H COMPUTATIONAL COMPLEXITY OF `GL-LowPopArt` 40

I E-OPTIMAL DESIGN FOR STAGE I 42

J PROOF OF THEOREM 4.1 – LOCAL MINIMAX LOWER BOUND 43

K ADDITIONAL DETAILS FOR BILINEAR DUELING BANDITS: SETTING 47
K.1 Motivation . . . 47
K.2 A Sufficient Condition for the Bilinear Preference to be Stochastic Transitive . . . 48
K.3 Miscellaneous Mathematical Preliminaries . . . 49
K.3.1 Skew-Symmetric Matrix . . . 49
K.3.2 2nd-Order Tensor Product Space . . . 49
K.4 Proof of Proposition K.2 . . . 50

L ADDITIONAL DETAILS FOR BILINEAR DUELING BANDITS: REGRET ANALYSIS 52
L.1 Proof of Theorem 5.1 – Borda Regret Upper Bound for Bilinear Dueling Bandits . . . 52
L.2 Relations to Wu et al. (2024) . . . 53

M PRELIMINARY EXPERIMENTS: 1-BIT MATRIX COMPLETION/RECOVERY 56
M.1 Experimental Setting . . . 56
M.2 Results & Discussion . . . 56

N ADDITIONAL FUTURE DIRECTIONS 58

A RELATED WORK

Generalized Linear Matrix Completion. This has been extensively studied in the early 2010s under various noise assumptions: Gaussian (Koltchinskii et al., 2011; Rohde and Tsybakov, 2011), Bernoulli (Alquier et al., 2019), multinomial (Klopp et al., 2015; Lafond et al., 2014), general exponential family (Lafond, 2015), and even with the only assumption of bounded variance (Klopp, 2014). We refer interested readers to Davenport and Romberg (2016) for an overview of works on matrix completion. Note that our model implicitly implies that for each $(i, j) \in [d_1] \times [d_2]$ may be observed multiple times, which is often the case in recommender systems and bandits where the same item can be recommended multiple times for exploration, or it may be that "users are more active than others and popular items are rated more frequently." (Klopp et al., 2015). On a slightly different note, many works have explored the same setting under the assumption that each entry of $\boldsymbol{\Theta}_\star$ can be sampled at most once (Alaya and Klopp, 2019; Cai and Zhou, 2013; Candès and Plan, 2010; Cao and Xie, 2016; Davenport et al., 2014; Gunasekar et al., 2014; McRae and Davenport, 2020). When $\boldsymbol{\Theta}_\star$ is additionally is skew-symmetric ($\boldsymbol{\Theta}_\star^\top = -\boldsymbol{\Theta}_\star$), this is also related to learning the low-rank preference model (Gleich and Lim, 2011; Lu and Negahban, 2015; Rajkumar and Agarwal, 2016; Wu et al., 2024; Zhang et al., 2025).

Low-Rank Matrix Bandits. Researchers in low-rank bandits have long focused on fundamental and specific models. For example, Jedra et al. (2024); Katariya et al. (2017a,b); Sentenac et al. (2021); Trinh et al. (2020) studied a bilinear bandit setting (which means $\mathcal{X} = \{\boldsymbol{x}\boldsymbol{x}^\top : \boldsymbol{x} \in \mathcal{A} \subset \mathbb{R}^{d_1}\}$) with canonical basis ($\mathcal{A} = \{\boldsymbol{e}_i : i \in [d_1]\}$ and $\mathcal{Z} = \{\boldsymbol{e}_j : j \in [d_2]\}$). Katariya et al. (2017a,b); Sentenac et al. (2021); Trinh et al. (2020) added an assumption that $\text{rank}(\boldsymbol{\Theta}_\star) = 1$ over a bilinear bandit setting. Stojanovic et al. (2023) presents an entry-wise matrix estimation for low-rank reinforcement learning, including low-rank bandits. Another popular assumption on arm sets in low-rank bandits is a unit ball (or a unit sphere) assumption Huang et al. (2021); Kotłowski and Neu (2019); Lattimore and Hao (2021). For bilinear bandits, Kotłowski and Neu (2019) assumed that $\mathcal{X} = \{\boldsymbol{x}\boldsymbol{x}^\top : \boldsymbol{x} \in \mathcal{S}^{d-1}(1)\}$ and $\boldsymbol{\Theta}_\star$ should be also symmetric. Lattimore and Hao (2021) even added an assumption that $\boldsymbol{\Theta}_\star$ is a symmetric rank-1 matrix. For low-rank bandits, Huang et al. (2021) assumed $\mathcal{X} = \mathcal{B}_F^{d\times d}(1)$. These tailored algorithms often outperform general approaches significantly, yet extending these algorithms to other settings has generally proven challenging due to the highly specialized nature of their settings.

The first study on low-rank bandits with general arm sets is Jun et al. (2019). This work introduced the first general bilinear low-rank linear bandit algorithm that could be applied flexibly to any d-dimensional arm set $\mathcal{X}$ and $\mathcal{Z}$. Subsequently, Lu et al. (2021) extended this approach beyond bilinear settings, proposing a generalized low-rank linear bandit algorithm applicable to all matrix arm sets. Later, Kang et al. (2022) introduced a novel method leveraging Stein's method, and Li et al. (2022) developed a general framework for high-dimensional linear bandits, including low-rank bandits. However, none of these studies explicitly addressed experimental design; rather, they handled the issue of experimental designs by assuming that their arm sets are sufficiently well-distributed in all directions. As a result, they failed to fully capture how the regret bound varies with the geometry of the arm set. For example, Jun et al. (2019) and Lu et al. (2021) conjectured that the lower bound for the bilinear low-rank bandit problem should be $\Omega(\sqrt{rd^3T})$, based on results from trace regression. However, Jang et al. (2021) later demonstrated that by considering the structure of the arm set in the bilinear setting, this bound could be further improved, highlighting the importance of optimal design tailored to the arm set.

Recent work by Jang et al. (2024) systematically addresses arm set geometry and experimental design in the low-rank linear bandits. This work applied thresholding at the subspace level called `LowPopArt` and proposed a novel experimental design for this new regression method. They then analyzed the experimental design assumptions underlying previous studies and successfully proved that their `LowPopArt` with their experimental design outperforms the previous works, even order-wise improvements in some cases. Our paper further extends the `LowPopArt` to the generalized linear scenario and provides performance guarantees in both upper and lower bounds that are nearly optimal even in terms of instance-specific, curvature-dependent quantities.

Generalized Linear Bandits (GLBs). **GLB** is a natural nonlinear extension of linear bandits, first proposed by Filippi et al. (2010), and later studied by much works (Jun et al., 2017; Lee et al., 2024a; Li et al., 2017; Sawarni et al., 2024). **GLB**s encompass a wide range of bandits, including linear, logistic, Poisson, logit, and more. Out of these, especially **logistic bandits (LogB)** (Abeille et al., 2021; Faury et al., 2020, 2022; Lee et al., 2024b; Mason et al., 2022) has garnered much attention, as it can naturally model binary feedback ('click' or 'no click'; Li et al. (2012)). Also, owing to its similarity to the Bradley-Terry model-based RLHF, the confidence

sets of logistic bandits have been used for quantifying the uncertainty of the linear reward model (Das et al., 2025; Xiong et al., 2024; Zhong et al., 2025). In **GLB**s, the key quantity describing the problem difficulty is[4] $\kappa_\star := \dot{\mu}(\langle \boldsymbol{x}_\star, \boldsymbol{\theta}_\star \rangle)$, where $\boldsymbol{\theta}_\star$ is the unknown vector and $\boldsymbol{x}_\star$ is the optimal arm vector. Abeille et al. (2021) showed a regret lower bound of $\Omega(d\sqrt{T\kappa_\star})$ for **LogB**s, which was matched by various UCB-type algorithms (Abeille et al., 2021; Faury et al., 2022; Lee et al., 2024b). Despite the lack of a generic lower bound for general **GLB**s, recent breakthroughs (Lee et al., 2024a; Liu et al., 2024; Sawarni et al., 2024) showed that for self-concordant **GLB**s, regret upper bound of $\widetilde{\mathcal{O}}(d\sqrt{T\kappa_\star})$ can be attained.

Remark 5 (Different Definitions of Self-Concordance)**.** *In the optimization literature, the original definition of the self-concordance takes the form of* $|\dddot{\mu}(z)| \leq 2\ddot{\mu}(z)^{3/2}$ $\forall z \in \mathbb{R}$*, originally motivated for convergence analysis of Newton's method by Nesterov (1988). Bach (2010) was the first to adapt the concept to extend the M-estimator results of squared loss to logistic loss, later further extended by Ostrovskii and Bach (2021). Later, the bandit community further adapted it for logistic and generalized linear bandits (Abeille et al., 2021; Faury et al., 2020; Russac et al., 2021), which is the form we consider here (Assumption 3(a))*

Burer-Monteiro Factorization The Burer–Monteiro factorization (BMF, Burer and Monteiro (2003, 2005)) approach has been extensively studied for noiseless low-rank matrix recovery from deterministic linear measurements (Candès and Recht, 2009; Candès and Plan, 2011), primarily from an optimization perspective (Bhojanapalli et al., 2016; Bi et al., 2022; Boumal et al., 2016; Ge et al., 2017; Kim and Chung, 2023; Park et al., 2017; Stöger and Soltanolkotabi, 2021; Yalçın et al., 2022; Zhang et al., 2024). In contrast, our work focuses on noisy matrix completion under a generalized linear model (GLM) framework, aiming to achieve accurate estimation with high probability as the sample size increases. This fundamental difference in problem settings implies that the optimization complexity measures used to analyze BMF methods, such as the optimization complexity metric (OCM) introduced by Yalçın et al. (2022) and Zhang et al. (2024), are not directly comparable to our statistical analysis. Specifically, their OCM quantifies the non-convexity of the BMF landscape, which is related to the success of local search methods (e.g., gradient descent), while our "statistical complexity metric", arguably $\lambda_{\max}(\boldsymbol{H}(\pi; \boldsymbol{\Theta}_\star))$ that pops up in our lower bound (Theorem 4.1), is information-theoretic and dictates the minimum sample size required for any estimator to obtain a desired accuracy with high probability.

While BMF methods offer computational efficiency and have been shown to perform well empirically, especially in large-scale problems, they all rely on some non-convex optimization, whose landscape is not always guaranteed to be benign, especially in the presence of noise (Ma and Fattahi, 2025). Our `GL-LowPopArt` only involves convex optimization subroutines and thus is computationally tractable, but inefficient: for instance, `GL-LowPopArt` requires computing the SVD and inverting $d^2 \times d^2$ matrices. Therefore, while BMF and our work both address low-rank matrix recovery, their respective advantages depend on the specific problem context.

[4] In the mentioned literature, the quantity is denoted as $\kappa_\star$. To keep our notation consistent with the dueling bandits' literature, we chose to denote this as $\kappa_\star^{-1}$.

B TABLE OF NOTATION

Table 2: Summary of notation used in this paper.

Notation	Description
$\|\cdot\|_{\text{nuc}}$	Nuclear norm
$\|\cdot\|_{\text{op}}$	Operator (spectral) norm
$\langle \boldsymbol{A}, \boldsymbol{B} \rangle$ for $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{m \times n}$	$\text{tr}(\boldsymbol{A}^\top \boldsymbol{B})$
$\lambda_i(\boldsymbol{A})$	The i-th largest eigenvalue of a symmetric matrix $\boldsymbol{A}$
$\lambda_{\max}$	The largest eigenvalue, same as λ_1
$\lambda_{\min}$	The smallest eigenvalue, same as λ_m
$\mathcal{B}_i^{d_1 \times d_2}(S)$ for $i \in \{\text{op}, \text{nuc}, F\}$	$\{\boldsymbol{X} \in \mathbb{R}^{d_1 \times d_2} : \|\boldsymbol{X}\|_i \leq S\}$
$\text{vec} : \mathbb{R}^{d_1 \times d_2} \to \mathbb{R}^{d_1 d_2}$	Column-wise stacking operation of a matrix into a vector
$\text{vec}^{-1} : \mathbb{R}^{d_1 d_2} \to \mathbb{R}^{d_1 \times d_2}$	Reshape operation of a vector to a matrix
$[n]$ for $n \in \mathbb{N}$	$\{1, 2, \ldots, n\}$
$\mathcal{P}(X)$	The set of all probability distributions on X
Ω	Parameter space
$\boldsymbol{\Theta}_\star \in \mathbb{R}^{d_1 \times d_2}$	An unknown reward matrix of rank at most $r \ll d_1 \wedge d_2$
$\mathcal{X} \subseteq \mathbb{R}^{d_1 \times d_2}$	Arm-set (e.g., sensing matrices).
$p(y \| \boldsymbol{X}; \boldsymbol{\Theta}_\star)$	Probability density function of the generalized linear model of the reward y when X is chosen by the learner, $\propto \exp\left(\frac{y\langle \boldsymbol{X}, \boldsymbol{\Theta}_\star \rangle - m(\langle \boldsymbol{X}, \boldsymbol{\Theta}_\star \rangle)}{g(\tau)}\right)$
$m : \mathbb{R} \to \mathbb{R}$	log-partition function of GLM
τ	Dispersion parameter
μ	$\dot{m}$, Inverse link function.
$\pi \in \mathcal{P}(\mathcal{A})$	Sampling policy (design)
$\boldsymbol{V}(\pi)$	Design matrix, $\mathbb{E}_{\boldsymbol{X} \sim \pi}[\text{vec}(\boldsymbol{X})\text{vec}(\boldsymbol{X})^\top]$
$\boldsymbol{H}(\pi; \boldsymbol{\Theta})$	Hessian matrix $\mathbb{E}_{\boldsymbol{X} \sim \pi}\left[\dot{\mu}(\langle \boldsymbol{X}, \boldsymbol{\Theta} \rangle)\text{vec}(\boldsymbol{X})\text{vec}(\boldsymbol{X})^\top\right]$
$L_\mu, R_s, \kappa_\star$	Parameters on μ, check Assumption 3
$\mathcal{H}$	Hermitian Dilation (Check Eq. (3))
ψ	Influence function (Check Eq. (3)
$\tilde{\psi}_\nu(\boldsymbol{A})$	$\frac{1}{\nu}\psi(\nu\mathcal{H}(\boldsymbol{A}))_{\text{ht}}$, where for $\boldsymbol{M} \in \mathbb{R}^{(d_1+d_2)\times(d_1+d_2)}$, $\boldsymbol{M}_{\text{ht}} := \boldsymbol{M}_{1:d_1, d_1+1:d_1+d_2}$
$\text{GL}(\pi; \boldsymbol{\Theta}_\star)$	Our new experimental design objective (See Definition 3.1)

C PROOF OF THEOREM 3.1 – ERROR BOUND OF STAGE II

Recall the Hessian at the initial estimator $\mathbf{\Theta}_0$:

$$\boldsymbol{H}(\pi;\mathbf{\Theta}_0) := \mathbb{E}_{\boldsymbol{X}\sim\pi}\left[\dot{\mu}(\langle \boldsymbol{X}, \mathbf{\Theta}_0\rangle)\mathrm{vec}(\boldsymbol{X})\mathrm{vec}(\boldsymbol{X})^\top\right],$$

and the one-sample estimators (line 9 of Algorithm 1): for each $t \in [N_1]$,

$$\widetilde{\mathbf{\Theta}}_t = \mathrm{vec}^{-1}_{d\times d}\left(\widetilde{\boldsymbol{\theta}}_t\right), \quad \widetilde{\boldsymbol{\theta}}_t := \boldsymbol{H}(\pi;\mathbf{\Theta}_0)^{-1}\left(y_t - \mu(\langle \boldsymbol{X}_t, \mathbf{\Theta}_0\rangle)\right)\mathrm{vec}(\boldsymbol{X}_t),$$

We will also denote $E := \|\mathbf{\Theta}_\star - \mathbf{\Theta}_0\|_{\mathrm{nuc}}^2$, and will track any requirements on E in violet throughout the proof.

First observe that

$$\begin{aligned}
\mathbb{E}[\widetilde{\boldsymbol{\theta}}_t | \boldsymbol{X}_t = \boldsymbol{X}] &= \boldsymbol{H}(\pi;\mathbf{\Theta}_0)^{-1}\left[\mu(\langle \boldsymbol{X}, \mathbf{\Theta}_\star\rangle) - \mu(\langle \boldsymbol{X}, \mathbf{\Theta}_0\rangle)\right]\mathrm{vec}(\boldsymbol{X}) \\
&\overset{(*)}{=} \boldsymbol{H}(\pi;\mathbf{\Theta}_0)^{-1}\left[\dot{\mu}(\langle \boldsymbol{X}, \mathbf{\Theta}_0\rangle)\langle \mathbf{\Theta}_\star - \mathbf{\Theta}_0, \boldsymbol{X}\rangle + \langle \mathbf{\Theta}_\star - \mathbf{\Theta}_0, \boldsymbol{X}\rangle^2 G(\mathbf{\Theta}_0, \mathbf{\Theta}_\star; \boldsymbol{X})\right]\mathrm{vec}(\boldsymbol{X}) \\
&\qquad\qquad \text{(first-order Taylor expansion with integral remainder)} \\
&= \boldsymbol{H}(\pi;\mathbf{\Theta}_0)^{-1}\left[\dot{\mu}(\langle \boldsymbol{X}, \mathbf{\Theta}_0\rangle)\mathrm{vec}(\boldsymbol{X})\mathrm{vec}(\boldsymbol{X})^\top\mathrm{vec}(\mathbf{\Theta}_\star - \mathbf{\Theta}_0) + \langle \mathbf{\Theta}_\star - \mathbf{\Theta}_0, \boldsymbol{X}\rangle^2 G(\mathbf{\Theta}_0, \mathbf{\Theta}_\star; \boldsymbol{X})\mathrm{vec}(\boldsymbol{X})\right],
\end{aligned}$$

where at $(*)$, we define

$$G(\mathbf{\Theta}_0, \mathbf{\Theta}_\star; \boldsymbol{X}) := \int_0^1 (1-z)\ddot{\mu}(\langle z\mathbf{\Theta}_\star + (1-z)\mathbf{\Theta}_0, \boldsymbol{X}\rangle)dz.$$

By taking the expectation over $\boldsymbol{X} \sim \pi$, we have that

$$\mathbb{E}\left[\widetilde{\boldsymbol{\theta}}_t\right] = \mathrm{vec}(\mathbf{\Theta}_\star - \mathbf{\Theta}_0) + \mathbb{E}_{\boldsymbol{X}\sim\pi}\left[\langle \mathbf{\Theta}_\star - \mathbf{\Theta}_0, \boldsymbol{X}\rangle^2 G(\mathbf{\Theta}_0, \mathbf{\Theta}_\star; \boldsymbol{X})\boldsymbol{H}(\pi;\mathbf{\Theta}_0)^{-1}\mathrm{vec}(\boldsymbol{X})\right].$$

Note that $\widetilde{\boldsymbol{\theta}}_t$'s are *biased* estimators of $\mathrm{vec}(\mathbf{\Theta}_\star - \mathbf{\Theta}_0)$. This is the key difference from Jang et al. (2024).

With $\mathbf{\Theta}_1$ as the matrix Catoni estimator for $\mathbf{\Theta}_\star - \mathbf{\Theta}_0$ (line 14 of Algorithm 1), we utilize the following error decomposition:

$$\begin{aligned}
\|\mathbf{\Theta}_1 - \mathbf{\Theta}_\star\|_{\mathrm{op}} &\le \underbrace{\left\|(\mathbf{\Theta}_1 - \mathbf{\Theta}_0) - \mathbb{E}[\widetilde{\mathbf{\Theta}}_t]\right\|_{\mathrm{op}}}_{(a)} + \left\|\mathbb{E}[\widetilde{\mathbf{\Theta}}_t] - (\mathbf{\Theta}_\star - \mathbf{\Theta}_0)\right\|_{\mathrm{op}} \\
&\le (a) + \left\|\mathbb{E}[\widetilde{\mathbf{\Theta}}_t] - (\mathbf{\Theta}_\star - \mathbf{\Theta}_0)\right\|_F \\
&= (a) + \underbrace{\left\|\mathbb{E}_{\boldsymbol{X}\sim\pi}\left[\langle \mathbf{\Theta}_\star - \mathbf{\Theta}_0, \boldsymbol{X}\rangle^2 G(\mathbf{\Theta}_0, \mathbf{\Theta}_\star; \boldsymbol{X})\boldsymbol{H}(\pi;\mathbf{\Theta}_0)^{-1}\mathrm{vec}(\boldsymbol{X})\right]\right\|_2}_{(b)}.
\end{aligned}$$

We first control the unbiased error term (a) using the following error rate for the matrix Catoni estimator due to Minsker (2018):

Lemma C.1 (Corollary 3.1 of Minsker (2018))**.** *Let $\{\boldsymbol{A}_i\}_{i=1}^n \subset \mathbb{R}^{d_1\times d_2}$ be independent with $\mathbb{E}[\boldsymbol{A}_i] = \boldsymbol{A}$, and define their matrix variance statistics as*

$$\sigma_n^2 := \max\left\{\left\|\sum_{i=1}^n \mathbb{E}[\boldsymbol{A}_i\boldsymbol{A}_i^\top]\right\|_{\mathrm{op}}, \left\|\sum_{i=1}^n \mathbb{E}[\boldsymbol{A}_i^\top\boldsymbol{A}_i]\right\|_{\mathrm{op}}\right\}.$$

Then we have that for any $\delta \in (0,1)$,

$$\mathbb{P}\left(\left\|\widehat{\boldsymbol{T}} - \boldsymbol{A}\right\|_{\mathrm{op}} \le \sqrt{\frac{2\sigma_n^2}{n^2}\log\frac{2(d_1+d_2)}{\delta}}\right) \ge 1-\delta,$$

where

$$\widehat{\boldsymbol{T}} := \frac{1}{n}\left(\sum_{i=1}^{n} \tilde{\psi}_\nu(\boldsymbol{A}_i)\right)_{\mathrm{ht}}, \quad \nu := \sqrt{\frac{2}{\sigma_n^2}\log\frac{2(d_1+d_2)}{\delta}}.$$

Remark 6. *The significance of the Catoni-type robust estimator (Catoni, 2012; Minsker, 2018) is that the guarantee does not assume the boundedness of the matrices, yet it still gives a Bernstein-type concentration. This has been successfully utilized in obtaining tight, instance-specific guarantees for various reinforcement learning problems, such as sparse linear bandits (Jang et al., 2022), low-rank bandits (Jang et al., 2024), linear MDP (Wagenmaker et al., 2022), and more.*

To apply the above lemma, we bound the matrix variance statistics of the one-sample estimators $\widetilde{\boldsymbol{\Theta}}_t$'s, whose proof is deferred to Appendix C.1: recalling that $\kappa_\star := \min_{\boldsymbol{X}\in\mathcal{X}} \dot{\mu}(\langle \boldsymbol{X}, \boldsymbol{\Theta}_\star\rangle) > 0$,[5]

Lemma C.2. *Suppose that* $E^2 \le \frac{\kappa_\star g(\tau)}{4L_\mu^2}$. *Then,*

$$\max\left\{\left\|\sum_{t=1}^{N_2}\mathbb{E}[\widetilde{\boldsymbol{\Theta}}_t\widetilde{\boldsymbol{\Theta}}_t^\top]\right\|_{\mathrm{op}}, \left\|\sum_{t=1}^{N_2}\mathbb{E}[\widetilde{\boldsymbol{\Theta}}_t^\top\widetilde{\boldsymbol{\Theta}}_t]\right\|_{\mathrm{op}}\right\} \le 5g(\tau)\mathrm{GL}(\pi;\boldsymbol{\Theta}_0)N_2,$$

where $\mathrm{GL}(\pi;\boldsymbol{\Theta}_0) := \max\{H^{(\mathrm{row})}(\pi;\boldsymbol{\Theta}_0), H^{(\mathrm{col})}(\pi;\boldsymbol{\Theta}_0)\}$ *with*

$$H^{(\mathrm{row})}(\pi;\boldsymbol{\Theta}_0) := \lambda_{\max}\left(\sum_{m=1}^{d_1}\boldsymbol{D}_m{}^{(\mathrm{row})}(\pi;\boldsymbol{\Theta}_0)\right), \quad \boldsymbol{D}_m{}^{(\mathrm{row})}(\pi;\boldsymbol{\Theta}_0) := [(\boldsymbol{H}(\pi;\boldsymbol{\Theta}_0)^{-1})_{jk}]_{j,k\in\mathcal{I}_m{}^{(\mathrm{row})}},$$

and

$$H^{(\mathrm{col})}(\pi;\boldsymbol{\Theta}_0) := \lambda_{\max}\left(\sum_{m=1}^{d_2}\boldsymbol{D}_m{}^{(\mathrm{col})}(\pi;\boldsymbol{\Theta}_0)\right), \quad \boldsymbol{D}_m{}^{(\mathrm{col})}(\pi;\boldsymbol{\Theta}_0) := (\boldsymbol{H}(\pi;\boldsymbol{\Theta}_0)^{-1})_{\mathcal{I}_m{}^{(\mathrm{col})},\mathcal{I}_m{}^{(\mathrm{col})}},$$

where the index sets defined as $\mathcal{I}_m{}^{(\mathrm{row})} := \{d_1(\ell-1)+m : \ell\in[d_2]\}$ *and* $\mathcal{I}_m^{(\mathrm{col})} := [d_1(m-1)+1 : d_1 m]$.

Then, we have that by Lemmas C.1 and C.2,

$$\mathbb{P}\left((a) = \left\|(\boldsymbol{\Theta}_1-\boldsymbol{\Theta}_0) - \mathbb{E}[\widetilde{\boldsymbol{\Theta}}_t]\right\|_{\mathrm{op}} \le \sqrt{\frac{10g(\tau)\mathrm{GL}(\pi)}{N_2}\log\frac{4(d_1+d_2)}{\delta}}\right) \ge 1-\frac{\delta}{2}.$$

We now control the bias term (b). For notational simplicity, we denote $\boldsymbol{x} \triangleq \mathrm{vec}(\boldsymbol{X})$ and $\boldsymbol{\delta} \triangleq \mathrm{vec}(\boldsymbol{\Theta}_\star - \boldsymbol{\Theta}_0)$.

$$\begin{aligned}
\left\|\mathbb{E}[\widetilde{\boldsymbol{\Theta}}_t] - (\boldsymbol{\Theta}_\star-\boldsymbol{\Theta}_0)\right\|_F &= \left\|\mathbb{E}_{\boldsymbol{x}}\left[\langle\boldsymbol{\delta},\boldsymbol{x}\rangle^2 G(\boldsymbol{\Theta}_0,\boldsymbol{\Theta}_\star;\boldsymbol{X})\boldsymbol{H}(\pi;\boldsymbol{\Theta}_0)^{-1}\boldsymbol{x}\right]\right\|_2 \\
&= \left\|\boldsymbol{H}(\pi;\boldsymbol{\Theta}_0)^{-1}\mathbb{E}_{\boldsymbol{x}}\left[G(\boldsymbol{\Theta}_0,\boldsymbol{\Theta}_\star;\boldsymbol{X})\langle\boldsymbol{\delta},\boldsymbol{x}\rangle^2\boldsymbol{x}\right]\right\|_2 \\
&\le \left\|\boldsymbol{H}(\pi;\boldsymbol{\Theta}_0)^{-1}\right\|_{\mathrm{op}}\left\|\mathbb{E}_{\boldsymbol{x}}\left[G(\boldsymbol{\Theta}_0,\boldsymbol{\Theta}_\star;\boldsymbol{X})\langle\boldsymbol{\delta},\boldsymbol{x}\rangle^2\boldsymbol{x}\right]\right\|_2 \\
&\le \lambda_{\min}(\boldsymbol{H}(\pi;\boldsymbol{\Theta}_0))^{-1}\mathbb{E}_{\boldsymbol{x}}\left[|G(\boldsymbol{\Theta}_0,\boldsymbol{\Theta}_\star;\boldsymbol{X})|\langle\boldsymbol{\delta},\boldsymbol{x}\rangle^2\|\boldsymbol{x}\|_2\right] && \text{(Jensen's inequality)} \\
&\le \frac{R_s L_\mu}{2\lambda_{\min}(\boldsymbol{H}(\pi;\boldsymbol{\Theta}_0))}\mathbb{E}_{\boldsymbol{x}}\left[\langle\boldsymbol{\delta},\boldsymbol{x}\rangle^2\right] && (|G(\boldsymbol{\Theta}_0,\boldsymbol{\Theta}_\star;\boldsymbol{X})| \le \tfrac{1}{2}R_s L_\mu) \\
&\le \frac{R_s L_\mu}{2\lambda_{\min}(\boldsymbol{H}(\pi;\boldsymbol{\Theta}_0))}E^2, && \text{(C.1)}
\end{aligned}$$

where the last inequality follows from Hölder's inequality:

$$\mathbb{E}_{\boldsymbol{x}}\left[\langle\boldsymbol{\delta},\boldsymbol{x}\rangle^2\right] = \mathbb{E}_{\boldsymbol{X}}\left[\langle\boldsymbol{\Theta}_\star-\boldsymbol{\Theta}_0,\boldsymbol{X}\rangle^2\right] \le \mathbb{E}_{\boldsymbol{X}}\left[\|\boldsymbol{\Theta}_\star-\boldsymbol{\Theta}_0\|_{\mathrm{nuc}}^2\|\boldsymbol{X}\|_{\mathrm{op}}^2\right] \le E^2.$$

[5] Note that this holds because $\mathcal{X}$ is bounded.

Thus, with $R_s E^2 \lesssim \lambda_{\min}(\boldsymbol{H}(\pi;\boldsymbol{\Theta}_0))\sqrt{\frac{g(\tau)\mathrm{GL}(\pi;\boldsymbol{\Theta}_0)}{N_2}\log\frac{d_1+d_2}{\delta}}$, the following holds with probability at least $1-\delta$:

$$\|\boldsymbol{\Theta}_1 - \boldsymbol{\Theta}_\star\|_{\mathrm{op}} \le \sigma_{\mathrm{thres}} \triangleq 4\sqrt{\frac{g(\tau)\mathrm{GL}(\pi;\boldsymbol{\Theta}_0)}{N_2}\log\frac{4(d_1+d_2)}{\delta}}.$$

As the last step of the proof, we recall Weyl's inequality for singular values:

Lemma C.3 (Problem 7.3.P16 of Horn and Johnson (2012)). *For any $\boldsymbol{A}, \Delta \in \mathbb{R}^{d_1\times d_2}$, we have*

$$|\sigma_k(\boldsymbol{A}+\Delta) - \sigma_k(\boldsymbol{A})| \le \sigma_1(\Delta), \quad \forall k \in [d_1 \wedge d_2].$$

As $\sigma_k(\boldsymbol{\Theta}_\star) = 0$ for $k \ge r+1$, we have that $\sigma_k(\boldsymbol{\Theta}_1) \le \sigma_{\mathrm{thres}}$ for the same k's as well. This proves that the thresholding part of Algorithm 1 (line 10) indeed yields $\mathrm{rank}(\widehat{\boldsymbol{\Theta}}) \le r$.

The final error bound follows from the triangle inequality and Lemma C.4, which implies that $\mathrm{GL}(\pi;\boldsymbol{\Theta}_0) \le 2\mathrm{GL}(\pi;\boldsymbol{\Theta}_\star)$. □

C.1 Proof of Lemma C.2 – Controlling the Matrix Variance Statistics

Throughout, we utilize the following self-concordance control formalizing the intuition that $\boldsymbol{H}(\pi;\boldsymbol{\Theta}_0) \approx \boldsymbol{H}(\pi;\boldsymbol{\Theta}_\star)$:

Lemma C.4 (Lemma 5 of Jun et al. (2021), adapted to our notations). *Suppose $R_s E \le \frac{1}{2}$. Then, we have that*

$$\frac{1}{2}\boldsymbol{H}(\pi;\boldsymbol{\Theta}_\star) \preceq \boldsymbol{H}(\pi;\boldsymbol{\Theta}_0) \preceq 2\boldsymbol{H}(\pi;\boldsymbol{\Theta}_\star).$$

We will bound $\left\|\mathbb{E}[\widetilde{\boldsymbol{\Theta}}_t\widetilde{\boldsymbol{\Theta}}_t^\top]\right\|_{\mathrm{op}}$ only, as the other one follows analogously.

For $y \sim p(\cdot|\boldsymbol{X};\boldsymbol{\Theta}_\star)$,

$$\begin{aligned}
\mathbb{E}[(y-\mu(\langle\boldsymbol{X},\boldsymbol{\Theta}\rangle))^2|\boldsymbol{X}] &\le 2\mathbb{E}[(y-\mu(\langle\boldsymbol{X},\boldsymbol{\Theta}_\star\rangle))^2] + 2(\mu(\langle\boldsymbol{X},\boldsymbol{\Theta}_\star\rangle) - \mu(\langle\boldsymbol{X},\boldsymbol{\Theta}_0\rangle))^2 \\
&\le 2g(\tau)\dot{\mu}(\langle\boldsymbol{X},\boldsymbol{\Theta}_\star\rangle) + 2L_\mu^2\langle\boldsymbol{X},\boldsymbol{\Theta}_\star-\boldsymbol{\Theta}_0\rangle^2 \\
&\le 2\left(g(\tau) + \frac{L_\mu^2 E^2}{\kappa_\star}\right)\dot{\mu}(\langle\boldsymbol{X},\boldsymbol{\Theta}_\star\rangle).
\end{aligned} \tag{C.2}$$

For notational simplicity, we further introduce $\boldsymbol{A}_{\boldsymbol{X}} := \mathrm{vec}^{-1}\left(\boldsymbol{H}(\pi;\boldsymbol{\Theta}_0)^{-1}\mathrm{vec}(\boldsymbol{X})\right) \in \mathbb{R}^{d_1\times d_2}$. Then, we have

$$\begin{aligned}
\left\|\mathbb{E}[\widetilde{\boldsymbol{\Theta}}_t\widetilde{\boldsymbol{\Theta}}_t^\top]\right\|_{\mathrm{op}} &= \left\|\mathbb{E}\left[(y_t-\mu(\langle\boldsymbol{X}_t,\boldsymbol{\Theta}_0\rangle))^2\boldsymbol{A}_{\boldsymbol{X}_t}\boldsymbol{A}_{\boldsymbol{X}_t}^\top\right]\right\|_{\mathrm{op}} \\
&= \left\|\mathbb{E}_{\boldsymbol{X}\sim\pi}\left[\mathbb{E}_{y\sim p(\cdot|\boldsymbol{X};\boldsymbol{\Theta}_\star)}[(y-\mu(\langle\boldsymbol{X},\boldsymbol{\Theta}_0\rangle))^2|\boldsymbol{X}]\boldsymbol{A}_{\boldsymbol{X}}\boldsymbol{A}_{\boldsymbol{X}}^\top\right]\right\|_{\mathrm{op}} && \text{(Tower property)} \\
&\le 2\left(g(\tau) + \frac{L_\mu^2 E^2}{\kappa_\star}\right)\left\|\mathbb{E}_{\boldsymbol{X}\sim\pi}\left[\dot{\mu}(\langle\boldsymbol{X},\boldsymbol{\Theta}_\star\rangle)\boldsymbol{A}_{\boldsymbol{X}}\boldsymbol{A}_{\boldsymbol{X}}^\top\right]\right\|_{\mathrm{op}}.
\end{aligned}$$

We largely follow the proof of Jang et al. (2024, Lemma B.2) as follows: denoting $\mathcal{S}^{d_1-1}$ as the unit sphere in $\mathbb{R}^{d_1}$,

$$\begin{aligned}
&\left\|\mathbb{E}_{\boldsymbol{X}\sim\pi}\left[\dot{\mu}(\langle\boldsymbol{X},\boldsymbol{\Theta}_\star\rangle)\boldsymbol{A}_{\boldsymbol{X}}\boldsymbol{A}_{\boldsymbol{X}}^\top\right]\right\|_{\mathrm{op}} \\
&= \max_{\boldsymbol{u}\in\mathcal{S}^{d_1-1}} \boldsymbol{u}^\top\mathbb{E}_{\boldsymbol{X}\sim\pi}\left[\dot{\mu}(\langle\boldsymbol{X},\boldsymbol{\Theta}_\star\rangle)\boldsymbol{A}_{\boldsymbol{X}}\boldsymbol{A}_{\boldsymbol{X}}^\top\right]\boldsymbol{u} \\
&= \max_{\boldsymbol{u}\in\mathcal{S}^{d_1-1}} \boldsymbol{u}^\top\mathbb{E}_{\boldsymbol{X}\sim\pi}\left[\dot{\mu}(\langle\boldsymbol{X},\boldsymbol{\Theta}_\star\rangle)\boldsymbol{A}_{\boldsymbol{X}}\left(\sum_{m=1}^{d_2}\boldsymbol{e}_m\boldsymbol{e}_m{}^\top\right)\boldsymbol{A}_{\boldsymbol{X}}^\top\right]\boldsymbol{u} \\
&\qquad\qquad \text{(let } \{\boldsymbol{e}_m\}_{m\in[d_2]} \text{ be the standard basis vectors of } \mathbb{R}^{d_2}\text{)}
\end{aligned}$$

$$
\begin{aligned}
&= \max_{\boldsymbol{u}\in\mathcal{S}^{d_1-1}} \mathbb{E}_{\boldsymbol{X}\sim\pi}\left[\dot{\mu}(\langle \boldsymbol{X},\boldsymbol{\Theta}_\star\rangle)\sum_{m=1}^{d_2}\left(\boldsymbol{u}^\top \boldsymbol{A}_{\boldsymbol{X}}\boldsymbol{e}_m\right)^2\right] \\
&= \max_{\boldsymbol{u}\in\mathcal{S}^{d_1-1}} \mathbb{E}_{\boldsymbol{X}\sim\pi}\left[\dot{\mu}(\langle \boldsymbol{X},\boldsymbol{\Theta}_\star\rangle)\sum_{m=1}^{d_2}\langle \boldsymbol{e}_m\otimes\boldsymbol{u},\mathrm{vec}(\boldsymbol{A}_{\boldsymbol{X}})\rangle^2\right] \\
&\qquad\qquad (\boldsymbol{x}^\top\boldsymbol{A}\boldsymbol{y}=\langle \boldsymbol{y}\otimes\boldsymbol{x},\mathrm{vec}(\boldsymbol{A})\rangle;\ \text{Eqn. (40) of Minka (1997)}) \\
&= \max_{\boldsymbol{u}\in\mathcal{S}^{d_1-1}} \mathbb{E}_{\boldsymbol{X}\sim\pi}\left[\dot{\mu}(\langle \boldsymbol{X},\boldsymbol{\Theta}_\star\rangle)\sum_{m=1}^{d_2}\langle \boldsymbol{e}_m\otimes\boldsymbol{u},\boldsymbol{H}(\pi;\boldsymbol{\Theta}_0)^{-1}\mathrm{vec}(\boldsymbol{X})\rangle^2\right] \qquad (\text{Definition of } \boldsymbol{A}_{\boldsymbol{X}}) \\
&= \max_{\boldsymbol{u}\in\mathcal{S}^{d_1-1}} \sum_{m=1}^{d_2}(\boldsymbol{e}_m\otimes\boldsymbol{u})^\top\boldsymbol{H}(\pi;\boldsymbol{\Theta}_0)^{-1}\boldsymbol{H}(\pi;\boldsymbol{\Theta}_\star)\boldsymbol{H}(\pi;\boldsymbol{\Theta}_0)^{-1}(\boldsymbol{e}_m\otimes\boldsymbol{u}) \\
&\le 2\max_{\boldsymbol{u}\in\mathcal{S}^{d_1-1}} \sum_{m=1}^{d_2}(\boldsymbol{e}_m\otimes\boldsymbol{u})^\top\boldsymbol{H}(\pi;\boldsymbol{\Theta}_0)^{-1}(\boldsymbol{e}_m\otimes\boldsymbol{u}) \\
&\qquad\qquad (\text{Lemma C.4, } \boldsymbol{ABA}\preceq\boldsymbol{ACA} \text{ for positive semi-defnite } \boldsymbol{A},\boldsymbol{B},\boldsymbol{C}) \\
&= 2\max_{\boldsymbol{u}\in\mathcal{S}^{d_1-1}} \sum_{m=1}^{d_2}\boldsymbol{u}^\top\left((\boldsymbol{H}(\pi;\boldsymbol{\Theta}_0)^{-1})_{\mathcal{I}_m^{(\mathrm{col})},\mathcal{I}_m^{(\mathrm{col})}}\right)\boldsymbol{u} \qquad (\text{see Lemma C.2 for the definition of } \mathcal{I}_m^{(\mathrm{col})}) \\
&= 2\,\underbrace{\lambda_{\max}\left(\sum_{m=1}^{d_2}(\boldsymbol{H}(\pi;\boldsymbol{\Theta}_0)^{-1})_{\mathcal{I}_m^{(\mathrm{col})},\mathcal{I}_m^{(\mathrm{col})}}\right)}_{=:H^{(\mathrm{col})}(\pi;\boldsymbol{\Theta}_0)}.
\end{aligned}
$$

All in all, we have that

$$
\left\|\mathbb{E}[\widetilde{\boldsymbol{\Theta}}_t\widetilde{\boldsymbol{\Theta}}_t^\top]\right\|_{\mathrm{op}} \le 4\left(g(\tau)+\frac{L_\mu^2E^2}{\kappa_\star}\right)H^{(\mathrm{col})}(\pi;\boldsymbol{\Theta}_0).
$$

Similarly, one can obtain

$$
\left\|\mathbb{E}[\widetilde{\boldsymbol{\Theta}}_t^\top\widetilde{\boldsymbol{\Theta}}_t]\right\|_{\mathrm{op}} \le 4\left(g(\tau)+\frac{L_\mu^2E^2}{\kappa_\star}\right)H^{(\mathrm{row})}(\pi;\boldsymbol{\Theta}_0).
$$

We then conclude by plugging in the bound for E^2. □

C.2 A Glimpse of Unbounded Arm-Set: The Linear Case

Scrutinizing the above proof, Eqn. (C.2) is the only part where we explicitly use $\kappa_\star$, which requires the boundedness of $\mathcal{X}$. When $\mu(z)=z$, this requirement vanishes, which is something that even Jang et al. (2024) has seemingly missed. Moreover, in this case, the requirement for $\boldsymbol{\Theta}_0$ to be asymptotically consistent is not needed as well.

To showcase this, we compare the Frobenius norm squared error rate of `GL-LowPopArt` when $\mu(z)=z$ with Negahban and Wainwright (2011, Corollary 3) in multivariate regression setup, which is equivalent to linear trace regression with $\mathcal{X}=\{\boldsymbol{e}_i\boldsymbol{z}_t^\top\}$ with $\boldsymbol{z}_t\sim\mathcal{N}_{d_2}(\boldsymbol{0},\boldsymbol{\Sigma})$ and $\pi=\mathrm{Unif}([d_1])\otimes\mathcal{N}_{d_2}(\boldsymbol{0},\boldsymbol{\Sigma})$ (Negahban and Wainwright, 2011, Example 1). Then it is easy to see that $\boldsymbol{x}_t\sim\pi$ satisfies $\boldsymbol{x}_t\sim\frac{1}{d_1}\sum_{i=1}^{d_1}\mathcal{N}_{d_1d_2}(\boldsymbol{0},\boldsymbol{\Sigma}\otimes\boldsymbol{e}_i\boldsymbol{e}_i^\top)$, and thus, $\mathbb{E}[\boldsymbol{x}_t\boldsymbol{x}_t^\top]=\frac{1}{d_1}\boldsymbol{\Sigma}\otimes\boldsymbol{I}_{d_1}$. Suppose that the underyling GLM is $\mathcal{N}(0,\nu^2)$.

By Theorem 3.1 and Proposition 3.1 with d_1T observations, `GL-LowPopArt` attains

$$
\left\|\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}_\star\right\|_F^2 \lesssim \frac{\nu^2r(d_1\vee d_2)}{T\lambda_{\min}(\boldsymbol{\Sigma})}\log\frac{d_1\vee d_2}{\delta}.
$$

Negahban and Wainwright (2011, Corollary 3) shows the following error rate for nuclear penalized estimator:

$$
\left\|\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}_\star\right\|_F^2 \lesssim \frac{\nu^2r\lambda_{\max}(\boldsymbol{\Sigma})}{T\lambda_{\min}(\boldsymbol{\Sigma})^2}\left((d_1\vee d_2)+\log\frac{1}{\delta}\right).
$$

Note that unless $\boldsymbol{\Sigma}$ is well-conditioned, our rate is strictly better than the nuclear penalized estimator by a factor of the condition number $\lambda_{\max}(\boldsymbol{\Sigma})/\lambda_{\min}(\boldsymbol{\Sigma})$.

D PROOF OF PROPOSITION 3.1 – UPPER AND LOWER BOUNDS OF GL

Here, we largely follow the proof strategies of Appendix C.2 and D.2 of Jang et al. (2024).

D.1 Upper Bound of Definition 3.1

We have that

$$
\begin{aligned}
H^{(\mathrm{col})}(\pi;\boldsymbol{\Theta}_\star) &= \lambda_{\max}\left(\sum_{m=1}^{d_2} \boldsymbol{D}_m{}^{(\mathrm{col})}(\pi;\boldsymbol{\Theta}_\star)\right) \\
&\leq \sum_{m=1}^{d_2} \lambda_{\max}\left(\boldsymbol{D}_m{}^{(col)}(\pi;\boldsymbol{\Theta}_\star)\right) && (\lambda_{\max} \text{ is convex and 1-homogenous}) \\
&= \sum_{m=1}^{d_2} \max_{\boldsymbol{u}\in\mathcal{S}^{d_1-1}} \boldsymbol{u}^\top \boldsymbol{D}_m{}^{(\mathrm{col})}(\pi;\boldsymbol{\Theta}_\star)\boldsymbol{u} \\
&= \sum_{m=1}^{d_2} \max_{\boldsymbol{u}\in\mathcal{S}^{d_1-1}} (\boldsymbol{e}_m\otimes\boldsymbol{u})^\top \boldsymbol{H}(\pi;\boldsymbol{\Theta}_\star)^{-1}(\boldsymbol{e}_m\otimes\boldsymbol{u}) && (\text{see proof of Lemma C.2}) \\
&\leq \sum_{m=1}^{d_2} \max_{\boldsymbol{u}\in\mathcal{S}^{d_1d_2-1}} \boldsymbol{u}^\top \boldsymbol{H}(\pi;\boldsymbol{\Theta}_\star)^{-1}\boldsymbol{u} \\
&= d_2\lambda_{\max}(\boldsymbol{H}(\pi;\boldsymbol{\Theta}_\star)^{-1}) \\
&= \frac{d_2}{\lambda_{\min}(\boldsymbol{H}(\pi;\boldsymbol{\Theta}_\star))}.
\end{aligned}
$$

One can similarly prove that $H^{(\mathrm{row})}(\pi;\boldsymbol{\Theta}_\star) \leq \frac{d_1}{\lambda_{\min}(\boldsymbol{H}(\pi;\boldsymbol{\Theta}_\star))}$, and the desired conclusion follows.

D.2 Lower Bound of $\mathrm{GL}(\pi;\boldsymbol{\Theta}_\star)$

Again, by definition,

$$
\begin{aligned}
\mathrm{GL}(\pi;\boldsymbol{\Theta}_\star) &\geq \lambda_{\max}\left(\sum_{m=1}^{d_2} [(\boldsymbol{H}(\pi;\boldsymbol{\Theta}_\star)^{-1})_{jk}]_{j,k\in\{\ell+d_1(m-1):\ell\in[d_1]\}}\right) \\
&\geq \frac{1}{d_1}\mathrm{tr}\left(\sum_{m=1}^{d_2} [(\boldsymbol{H}(\pi;\boldsymbol{\Theta}_0)^{-1})_{jk}]_{j,k\in\{\ell+d_1(m-1):\ell\in[d_1]\}}\right) \\
& && (\lambda_{\max}(\boldsymbol{A}) \geq \tfrac{1}{d}\mathrm{tr}(\boldsymbol{A}) \text{ for any symmetric } \boldsymbol{A}\in\mathbb{R}^{d\times d}) \\
&= \frac{1}{d_1}\mathrm{tr}\left(\boldsymbol{H}(\pi;\boldsymbol{\Theta}_\star)^{-1}\right) \\
&\geq \frac{1}{d_1}\frac{(d_1d_2)^2}{\mathrm{tr}\left(\boldsymbol{H}(\pi;\boldsymbol{\Theta}_\star)\right)}, && (\text{AM-HM inequality on the eigenvalues of } \boldsymbol{H}(\pi;\boldsymbol{\Theta}_\star))
\end{aligned}
$$

and similarly,

$$
\mathrm{GL}(\pi;\boldsymbol{\Theta}_\star) \geq \frac{1}{d_2}\frac{(d_1d_2)^2}{\mathrm{tr}\left(\boldsymbol{H}(\pi;\boldsymbol{\Theta}_\star)\right)},
$$

i.e., $\mathrm{GL}(\pi;\boldsymbol{\Theta}_\star) \geq \frac{(d_1d_2)^2}{(d_1\wedge d_2)\mathrm{tr}(\boldsymbol{H}(\pi;\boldsymbol{\Theta}_\star))} \geq \frac{(d_1d_2)(d_1\vee d_2)}{2\mathrm{tr}(\boldsymbol{H}(\pi;\boldsymbol{\Theta}_\star))}$. □

E DEFERRED DETAILS REGARDING OPTIMAL GL-DESIGN

E.1 Proof of Proposition 3.2

First, by Proposition 3.1, we have that

$$\mathrm{GL}(\pi;\boldsymbol{\Theta}_\star) \le \frac{d_1 \vee d_2}{\lambda_{\min}(\boldsymbol{H}(\pi;\boldsymbol{\Theta}_\star))},$$

and thus, by definition, it suffices to solve the following convex optimization problem:

$$\max_{\pi\in\mathcal{P}([d_1]\times[d_2])} \lambda_{\min}\left(\sum_{i=1}^{d_1}\sum_{j=1}^{d_2}\pi(i,j)\dot{\mu}_{i,j}\mathrm{vec}(\boldsymbol{e}_i\tilde{\boldsymbol{e}}_j^\top)\mathrm{vec}(\boldsymbol{e}_i\tilde{\boldsymbol{e}}_j^\top)^\top\right) = \min_{i,j}\pi(i,j)\dot{\mu}_{i,j},$$

where we denote $\dot{\mu}_{i,j} := \dot{\mu}((\boldsymbol{\Theta}_\star)_{i,j})$, and the equality holds because the matrix is diagonal. Thus, we have a convex optimization problem of the following form:

$$\max_{\pi\in\mathbb{R}^{d_1+d_2}} \min_{i,j}\pi(i,j)\dot{\mu}_{i,j}, \quad \text{subj. to} \quad \sum_{i=1}^{d_1}\sum_{j=1}^{d_2}\pi(i,j)=1,\ \pi(i,j)\ge 0.$$

To solve this max-min problem, we introduce an auxiliary variable $c\in\mathbb{R}$ such that $c\le\pi(i,j)\dot{\mu}_{i,j}$ for all $i\in[d_1], j\in[d_2]$. The optimization problem can be equivalently reformulated as a linear program:

$$\max_{\pi,c}\ c \quad \text{subj. to} \quad \sum_{i=1}^{d_1}\sum_{j=1}^{d_2}\pi(i,j)=1, \quad \pi(i,j)\ge 0,\ \pi(i,j)\dot{\mu}_{i,j}\ge c.$$

Assuming $\dot{\mu}_{i,j}>0$ for all (i,j) (as the minimum would otherwise be trivially bounded by zero), the constraint $\pi(i,j)\dot{\mu}_{i,j}\ge c$ implies that

$$\pi(i,j)\ge\frac{c}{\dot{\mu}_{i,j}}.$$

Summing this over all $i\in[d_1]$ and $j\in[d_2]$, we obtain

$$1=\sum_{i=1}^{d_1}\sum_{j=1}^{d_2}\pi(i,j)\ge\sum_{i=1}^{d_1}\sum_{j=1}^{d_2}\frac{c}{\dot{\mu}_{i,j}} \implies c\le\left(\sum_{i=1}^{d_1}\sum_{j=1}^{d_2}\dot{\mu}_{i,j}^{-1}\right)^{-1}.$$

To demonstrate that this upper bound is achievable, we define a specific distribution $\pi^\star$ where the inequalities hold tightly as equalities:

$$\pi^\star(i,j)=\frac{\dot{\mu}_{i,j}^{-1}}{\sum_{k=1}^{d_1}\sum_{l=1}^{d_2}\dot{\mu}_{k,l}^{-1}} \quad \forall i\in[d_1], j\in[d_2].$$

Since $\dot{\mu}_{i,j}>0$, it is clear that $\pi^\star(i,j)\in[0,1]$ and $\sum_{i,j}\pi^\star(i,j)=1$, confirming that $\pi^\star$ is a valid probability distribution on $[d_1]\times[d_2]$. Evaluating the objective at this optimal allocation $\pi^\star$, we find:

$$\min_{i,j}\pi^\star(i,j)\dot{\mu}_{i,j}=\min_{i,j}\frac{\dot{\mu}_{i,j}^{-1}\dot{\mu}_{i,j}}{\sum_{k=1}^{d_1}\sum_{l=1}^{d_2}\dot{\mu}_{k,l}^{-1}}=\left(\sum_{k=1}^{d_1}\sum_{l=1}^{d_2}\dot{\mu}_{k,l}^{-1}\right)^{-1}.$$

Thus, the maximum value of the objective function is exactly $\left(\sum_{i=1}^{d_1}\sum_{j=1}^{d_2}\dot{\mu}_{i,j}^{-1}\right)^{-1}$. This completes the proof. □

F PROOF OF THEOREM 3.2 – ERROR BOUND OF STAGE I

We denote $N = N_1$ and $\pi = \pi_1$, and we recall/introduce the following notations:

$$\mathcal{L}_N(\boldsymbol{\Theta}) := \frac{1}{N}\sum_{t=1}^{N}\{m(\langle \boldsymbol{X}_t, \boldsymbol{\Theta}\rangle) - y_t\langle \boldsymbol{X}_t, \boldsymbol{\Theta}\rangle\}$$
$$\boldsymbol{H}(\pi;\boldsymbol{\Theta}) := \mathbb{E}_{\boldsymbol{X}\sim\pi}\left[\dot{\mu}(\langle \boldsymbol{X}, \boldsymbol{\Theta}\rangle)\mathrm{vec}(\boldsymbol{X})\mathrm{vec}(\boldsymbol{X})^\top\right]$$
$$\boldsymbol{\Theta}_0 := \arg\min_{\boldsymbol{\Theta}\in\Omega}\{\mathcal{L}_N(\boldsymbol{\Theta}) + \lambda_N \|\boldsymbol{\Theta}\|_{\mathrm{nuc}}\}$$

The proof here follows the standard M-estimation theory in high-dimensional statistics (Wainwright, 2019, Chapter 9 & 10), which we reproduce here for completeness.

Let $\delta \in (0,1)$ be given. We collect all requirements on N in violet to be aggregated at the end of the proof.

We begin with the following proposition, whose proof is provided in Appendix F.3, that provides λ_N's for different GLMs such that the gradient of $\mathcal{L}_N$ at the true parameter $\boldsymbol{\Theta}_\star$ is well bounded:

Proposition F.1 (Choosing λ_N)**.** *Denote the centered random variable $\xi := y - \mu(\langle \boldsymbol{X}, \boldsymbol{\Theta}_\star\rangle)$ and $\bar{\kappa}_\star := \mathbb{E}_{\boldsymbol{X}\sim\pi}\left[\dot{\mu}(\langle \boldsymbol{X}, \boldsymbol{\Theta}_\star\rangle)\right]$. By setting λ_N as described below, we have $\mathbb{P}(\|\nabla\mathcal{L}_N(\boldsymbol{\Theta}_\star)\|_{\mathrm{op}} \le \frac{\lambda_N}{2}) \ge 1 - \frac{\delta}{4}$:*

*(i) $|\xi| \le M$ **a.s.**: $\lambda_N = 4\sqrt{\frac{g(\tau)\bar{\kappa}_\star}{N}\log\frac{4(d_1+d_2)}{\delta}}$, given that $N \ge \frac{2M^2}{9L_\mu g(\tau)}\log\frac{4(d_1+d_2)}{\delta}$,*

(ii) $\|\xi\|_{\psi_1} \le K$:[a] $\lambda_N = 6K\sqrt{\frac{1}{N}\log\frac{4(d_1+d_2)}{\delta}}$, given that $N \ge \log\frac{4(d_1+d_2)}{\delta}$.

[a] Here, $\|\cdot\|_{\psi_1}$ is the subexponential norm, defined as $\|\xi\|_{\psi_1} := \inf\{K > 0 : \mathbb{E}[\exp(|\xi|/K)] \le 2\}$. This encompasses Gaussian and Poisson distributions among others.

We then define the distribution-dependent constant $C > 0$ as follows:

$$C := \begin{cases} 16g(\tau)\bar{\kappa}_\star, & \text{if the GLM is } M\text{-bounded} \\ 36K^2, & \text{if the GLM has finite subexponential norm of } K \end{cases}$$

Throughout, we will suppose that the event $\|\nabla\mathcal{L}_N(\boldsymbol{\Theta}_\star)\|_{\mathrm{op}} \le \frac{\lambda_N}{2}$ holds.

We now recall the definition of restricted strong convexity (RSC):

Definition F.1 (Definition 9.15 of Wainwright (2019))**.** *We say that $\mathcal{L}_N$ satisfies the **restricted strong convexity (RSC)** with $\alpha, \tau_N, R \ge 0$ if*

$$\mathcal{L}_N(\boldsymbol{\Theta}_\star + \Delta) - \mathcal{L}_N(\boldsymbol{\Theta}_\star) - \langle\nabla\mathcal{L}_N(\boldsymbol{\Theta}_\star), \Delta\rangle \ge \frac{\alpha}{2}\|\Delta\|_F^2 - \tau_N^2\|\Delta\|_{\mathrm{nuc}}^2, \quad \forall\Delta : \|\Delta\|_F \le R.$$

The next lemma shows that RSC holds with high probability:

Lemma F.1. *Suppose that $N \gtrsim \log d$. Then, there exists a constant $\tilde{R}_s > 0$ (that only depends on R_s) such that w.p. at least $1 - \frac{\delta}{4}$, **RSC** holds with $\alpha = \tilde{R}_s\lambda_{\min}(\boldsymbol{H}(\pi;\boldsymbol{\Theta}_\star))$, $\tau_N^2 \asymp L_\mu\sqrt{\frac{1}{N}\log\frac{d_1+d_2}{\delta}}$, and $R = \frac{1}{\sqrt{d_1+d_2}}$.*

As the subspace Lipschitz constant (Wainwright, 2019, Definition 9.18) of $\|\cdot\|_{\mathrm{nuc}}$ is $\sqrt{r}$, we conclude by invoking the "master theorem" for nuclear-norm regularized estimator (Wainwright, 2019, Corollary 9.20):

$$\mathbb{P}\left(\left\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}_\star\right\|_{\mathrm{nuc}} \le \frac{6r}{\lambda_{\min}(\boldsymbol{H}(\pi;\boldsymbol{\Theta}_\star))}\sqrt{\frac{C}{N}\log\frac{2(d_1+d_2)}{\delta}}\right) \ge 1 - \frac{\delta}{2},$$

given that $L_\mu\sqrt{\frac{1}{N}\log\frac{d_1+d_2}{\delta}}r \lesssim \lambda_{\min}(\boldsymbol{H}(\pi;\boldsymbol{\Theta}_\star))$ and $\frac{1}{\lambda_{\min}(\boldsymbol{H}(\pi;\boldsymbol{\Theta}_\star))}\sqrt{\frac{Cr}{N}\log\frac{d_1+d_2}{\delta}} \lesssim \frac{1}{\sqrt{d_1+d_2}}$. These conditions

hold when $N \gtrsim C\lambda_{\min}(\boldsymbol{H}(\pi;\boldsymbol{\Theta}_\star))^{-2}(d_1+d_2)r\log\frac{d_1+d_2}{\delta}$, and we conclude by taking the union bound with Proposition F.1. □

Remark 7. *We believe that our analysis can be extended to the general ℓ_q-constraint on the singular values of $\boldsymbol{\Theta}_\star$ for $q \in [0,1)$ as in Fan et al. (2019), and to the case where Ω is a smooth matrix manifold (Absil et al., 2007) using tools from manifold optimization (Boumal, 2023; Yang et al., 2014).*

F.1 Proof of Lemma F.1: Restricted Strong Convexity

By Taylor's expansion with integral remainder, we have

$$\begin{aligned}
&\mathcal{L}_N(\boldsymbol{\Theta}_\star+\Delta)-\mathcal{L}_N(\boldsymbol{\Theta}_\star)-\langle\nabla\mathcal{L}_N(\boldsymbol{\Theta}_\star),\Delta\rangle\\
&=\frac{1}{N}\sum_{t=1}^{N}\{m(\langle\boldsymbol{\Theta}_\star+\Delta,\boldsymbol{X}_t\rangle)-m(\langle\boldsymbol{\Theta}_\star,\boldsymbol{X}_t\rangle)-\langle\boldsymbol{\Theta}_\star,\boldsymbol{X}_t\rangle m'(\langle\boldsymbol{\Theta}_\star,\boldsymbol{X}_t\rangle)\}\\
&=\frac{1}{N}\sum_{t=1}^{N}\langle\Delta,\boldsymbol{X}_t\rangle^2\int_0^1(1-z)\underbrace{m''\left(\langle\boldsymbol{\Theta}_\star+z\Delta,\boldsymbol{X}_t\rangle\right)}_{=\dot\mu(\langle\boldsymbol{\Theta}_\star+z\Delta,\boldsymbol{X}_t\rangle)}dz.
\end{aligned}$$

As μ is self-concordant, we can utilize the following self-concordance control lemma:

Lemma F.2 (Lemma 9 of Abeille et al. (2021))**.** *For any R_s-self-concordant μ that is monotone increasing, the following holds: for any $z_1, z_2 \in \mathbb{R}$,*

$$\dot\mu(z_2)\exp(-R_s|z_1-z_2|)\le\dot\mu(z_1)\le\dot\mu(z_2)\exp(R_s|z_1-z_2|).$$

We choose $R=\frac{1}{\sqrt{d}}$, i.e., $\|\Delta\|_F\le\frac{1}{\sqrt{d}}$. Then, Cauchy-Schwarz inequality and $\|\boldsymbol{X}_t\|_{\mathrm{op}}\le 1$ imply $|\langle\Delta,\boldsymbol{X}_t\rangle|\le 1$. Thus, by the above self-concordance lemma,

$$\int_0^1(1-z)\dot\mu\left(\langle\boldsymbol{\Theta}_\star+z\Delta,\boldsymbol{X}_t\rangle\right)dz\ge\int_0^1(1-z)e^{-R_sz}\dot\mu\left(\langle\boldsymbol{\Theta}_\star,\boldsymbol{X}_t\rangle\right)dz\ge\tilde R_s\dot\mu\left(\langle\boldsymbol{\Theta}_\star,\boldsymbol{X}_t\rangle\right),$$

where we define the constant

$$\tilde R_s:=\begin{cases}\frac{1}{2}, & R_s=0\\ \frac{R_s-1+e^{-R_s}}{R_s^2}, & R_s>0.\end{cases}$$

Let us denote the empirical Hessian of $\mathcal{L}_N$ at $\boldsymbol{\Theta}_\star$ as

$$\widehat{\boldsymbol{H}}_N(\pi;\boldsymbol{\Theta}_\star):=\frac{1}{N}\sum_{t=1}^{N}\dot\mu(\langle\boldsymbol{X}_t,\boldsymbol{\Theta}_\star\rangle)\mathrm{vec}(\boldsymbol{X}_t)\mathrm{vec}(\boldsymbol{X}_t)^\top.$$

Thus, we have that

$$\begin{aligned}
&\mathcal{L}_N(\boldsymbol{\Theta}_\star+\Delta)-\mathcal{L}_N(\boldsymbol{\Theta}_\star)-\langle\nabla\mathcal{L}_N(\boldsymbol{\Theta}_\star),\Delta\rangle\\
&\ge\tilde R_s\mathrm{vec}(\Delta)^\top\widehat{\boldsymbol{H}}_N(\pi;\boldsymbol{\Theta}_\star)\mathrm{vec}(\Delta)\\
&=\tilde R_s\mathrm{vec}(\Delta)^\top\boldsymbol{H}(\pi;\boldsymbol{\Theta}_\star)\mathrm{vec}(\Delta)+\tilde R_s\mathrm{vec}(\Delta)^\top\left(\widehat{\boldsymbol{H}}_N(\pi;\boldsymbol{\Theta}_\star)-\boldsymbol{H}(\pi;\boldsymbol{\Theta}_\star)\right)\mathrm{vec}(\Delta)\\
&\ge\tilde R_s\lambda_{\min}\left(\boldsymbol{H}(\pi;\boldsymbol{\Theta}_\star)\right)\|\Delta\|_F^2-\tilde R_s\|\Delta\|_{\mathrm{nuc}}^2\underbrace{\left|\left(\frac{\mathrm{vec}(\Delta)}{\|\Delta\|_{\mathrm{nuc}}}\right)^\top\left(\widehat{\boldsymbol{H}}_N(\pi;\boldsymbol{\Theta}_\star)-\boldsymbol{H}(\pi;\boldsymbol{\Theta}_\star)\right)\frac{\mathrm{vec}(\Delta)}{\|\Delta\|_{\mathrm{nuc}}}\right|}_{\triangleq\mathcal{E}_N(\Delta)}
\end{aligned}$$

Using the standard empirical process theory and peeling argument over the Frobenius ball, the following lemma establishes the uniform boundedness of $\mathcal{E}_N$, whose proof is provided in Appendix F.2:

Lemma F.3. $\mathbb{P}\left(\mathcal{E}_N(\Delta) \lesssim L_\mu \sqrt{\frac{\log \frac{d}{\delta}}{N}},\ \forall \Delta : \|\Delta\|_{\text{nuc}} = 1\right) \geq 1 - \frac{\delta}{4}$, *provided that* $N \gtrsim \log d$.

Combining everything gives the desired result. □

F.2 Proof of Lemma F.3: Uniform Bounding of $\mathcal{E}_N$

This proof is based on the peeling argument over the Frobenius ball in Wainwright (2019, Theorem 10.17), as well as the standard empirical process theory tools such as Rademacher symmetrization and contraction inequalities; see Wainwright (2019, Chapter 3.4 & 4). We will introduce the tools as we go along.

For notational simplicity, we will denote $\frac{\Delta}{\|\Delta\|_{\text{nuc}}}$ as Δ satisfying $\|\Delta\|_{\text{nuc}} = 1$.

First note that

$$\text{vec}(\Delta)^\top \left(\widehat{\boldsymbol{H}}_N(\pi; \boldsymbol{\Theta}_\star) - \boldsymbol{H}(\pi; \boldsymbol{\Theta}_\star)\right) \text{vec}(\Delta) = \frac{1}{N} \sum_{t=1}^{N} \left\{ \dot{\mu}(\langle \boldsymbol{\Theta}_\star, \boldsymbol{X}_t \rangle) \langle \Delta, \boldsymbol{X}_t \rangle^2 - \mathbb{E}\left[\dot{\mu}(\langle \boldsymbol{\Theta}_\star, \boldsymbol{X}_t \rangle) \langle \Delta, \boldsymbol{X}_t \rangle^2\right] \right\}.$$

Let us define $f_\Delta(\boldsymbol{X}) := \dot{\mu}(\langle \boldsymbol{\Theta}_\star, \boldsymbol{X} \rangle) \langle \Delta, \boldsymbol{X} \rangle^2$, and $\mathcal{F} := \{f_\Delta : \|\Delta\|_{\text{nuc}} = 1\}$ be the function class indexed by Δ of unit nuclear norm. Note that by Hölder's inequality and $\|\boldsymbol{X}_t\|_{\text{op}} \leq 1$, we have $|\langle \Delta, \boldsymbol{X}_t \rangle| \leq \|\Delta\|_{\text{nuc}} \|\boldsymbol{X}_t\|_{\text{op}} \leq 1$. Thus, $\mathcal{F}$ is uniformly bounded by L_μ.

We now recall the following standard result from empirical process theory:

Lemma F.4 (Glivenko-Cantelli Property via Rademacher Complexity, Theorem 4.10 of Wainwright (2019))**.** *Let $\mathcal{F}$ be a function class uniformly bounded by b. For any $\delta \in (0, 1)$, with probability at least $1 - \delta$,*

$$\sup_{f \in \mathcal{F}} \left| \frac{1}{N} \sum_{t=1}^{N} f(\boldsymbol{X}_t) - \mathbb{E}[f(\boldsymbol{X})] \right| \leq 2\mathcal{R}_N(\mathcal{F}) + b\sqrt{\frac{2\log(1/\delta)}{N}},$$

where $\mathcal{R}_N(\mathcal{F})$ is the Rademacher complexity, defined as

$$\mathcal{R}_N(\mathcal{F}) := \frac{1}{N} \mathbb{E}_{\{\boldsymbol{X}_t\}, \{\epsilon_t\}} \left[\sup_{f \in \mathcal{F}} \left| \sum_{t=1}^{N} \epsilon_t f(\boldsymbol{X}_t) \right| \right],$$

where $\epsilon_t \overset{i.i.d.}{\sim} \text{Unif}(\{-1, +1\})$ are Rademacher random variables.

With this, we have that with probability at least $1 - \frac{\delta}{2}$,

$$\sup_{\Delta : \|\Delta\|_{\text{nuc}} = 1} |\mathcal{E}_N(\Delta)| \leq 2\mathcal{R}_N(\mathcal{F}) + L_\mu \sqrt{\frac{2\log \frac{2}{\delta}}{N}}.$$

We now recall the seminal Ledoux-Talagrand contraction inequality (Ledoux and Talagrand, 1991) for bounding $\mathcal{R}_N(\mathcal{F})$:

Lemma F.5 (Ledoux-Talagrand Contraction, Theorem 4.12 of Ledoux and Talagrand (1991))**.** *If the functions $\phi_t : \mathbb{R} \to \mathbb{R}$ are L-Lipschitz and satisfy $\phi_t(0) = 0$, then*

$$\mathbb{E}_{\{\boldsymbol{X}_t\}, \{\epsilon_t\}} \left[\sup_{f \in \mathcal{F}} \left| \sum_{t=1}^{N} \epsilon_t \phi_t(f(\boldsymbol{X}_t)) \right| \right] \leq 2L \mathbb{E}_{\{\boldsymbol{X}_t\}, \{\epsilon_t\}} \left[\sup_{f \in \mathcal{F}} \left| \sum_{t=1}^{N} \epsilon_t f(\boldsymbol{X}_t) \right| \right].$$

Define $\phi_t(z) := \dot{\mu}(\langle \boldsymbol{X}_t, \boldsymbol{\Theta}_\star \rangle) z^2$. For $|z| \leq 1$, the function ϕ_t is locally Lipschitz with constant $L_k = 2L_\mu$. Applying Lemma F.5, we obtain:

$$\mathcal{R}_N(\mathcal{F}) \leq 2(2L_\mu)\, \mathbb{E}_{\{\boldsymbol{X}_t\}, \{\epsilon_t\}} \left[\sup_{\Delta : \|\Delta\|_{\text{nuc}} = 1} \left| \frac{1}{N} \sum_{t=1}^{N} \epsilon_t \langle \Delta, \boldsymbol{X}_t \rangle \right| \right]$$

$$= 4L_\mu \, \mathbb{E}_{\{\boldsymbol{X}_t\},\{\epsilon_t\}} \left[\sup_{\|\Delta\|_{\text{nuc}}=1} \left| \left\langle \Delta, \frac{1}{N} \sum_{t=1}^{N} \epsilon_t \boldsymbol{X}_t \right\rangle \right| \right]$$
$$= 4L_\mu \, \mathbb{E}_{\{\boldsymbol{X}_t\},\{\epsilon_t\}} \left[\left\| \frac{1}{N} \sum_{t=1}^{N} \epsilon_t \boldsymbol{X}_t \right\|_{\text{op}} \right].$$

We now invoke the following expectation version of the matrix Bernstein inequality:

Lemma F.6 (Matrix Bernstein Inequality, Theorem 6.1.1 of Tropp (2015))**.** *Let $\boldsymbol{A}_1, \dots, \boldsymbol{A}_N$ be $d_1 \times d_2$ matrices with $\|\boldsymbol{A}_i\|_{\text{op}} \le L$, $\mathbb{E}[\boldsymbol{A}_i] = \boldsymbol{A}$ and $V_N := \frac{1}{N} \max\left\{ \left\| \sum_{i=1}^N \mathbb{E}[\boldsymbol{A}_i \boldsymbol{A}_i^\top] \right\|_{\text{op}}, \left\| \sum_{i=1}^N \mathbb{E}[\boldsymbol{A}_i^\top \boldsymbol{A}_i] \right\|_{\text{op}} \right\}$ be the matrix variance statistics. Then,*

$$\mathbb{E}\left[\left\| \frac{1}{N} \sum_{t=1}^{N} \boldsymbol{A}_t \right\|_{\text{op}} \right] \le \sqrt{\frac{2V_N \log(d_1+d_2)}{N}} + \frac{L \log(d_1+d_2)}{3N}.$$

Applying the above with $L = 1$ and $V_N \le 1$, we have that

$$\mathcal{R}_N(\mathcal{F}) \le 4L_\mu \left(\sqrt{\frac{2\log d}{N}} + \frac{\log d}{3N} \right).$$

Combining everything, we have that with probability at least $1 - \frac{\delta}{4}$,

$$\sup_{\Delta:\|\Delta\|_{\text{nuc}}=1} |\mathcal{E}_N(\Delta)| \le 8L_\mu \left(\sqrt{\frac{2\log d}{N}} + \frac{\log d}{3N} \right) + L_\mu \sqrt{\frac{2\log\frac{4}{\delta}}{N}}$$
$$\le 20L_\mu \sqrt{\frac{\log d}{N}} + L_\mu \sqrt{\frac{2\log\frac{4}{\delta}}{N}},$$

where the last inequality holds as long as $N \ge \frac{\log d}{9}$. □

F.3 Proof of Proposition F.1: Choosing λ_N

We significantly deviate from the original proof of Lee et al. (2024a) in two ways. First, we only need to bound the Lipschitz constant of $\mathcal{L}_N$ at $\boldsymbol{\Theta}_\star$ instead of globally. Second, using this and the fact that the covariates are bounded ($\|\boldsymbol{X}_t\|_{\text{op}} \le 1$), we can utilize appropirate matrix Bernstein inequalities (Tropp, 2012, 2015) instead of covering arguments, which saves us additional dimension dependencies, leads to λ_N scaling as $\sqrt{1/N}$ for a wide range of GLMs, and gives simpler proofs.

For notational simplicity, we denote the centered random variable $\xi_t := \mu(\langle \boldsymbol{X}_t, \boldsymbol{\Theta}_\star \rangle) - y_t$. Then note that

$$\nabla \mathcal{L}_N(\boldsymbol{\Theta}_\star) = \frac{1}{N} \sum_{t=1}^{N} (\mu_t(\boldsymbol{\Theta}_\star) - y_t) \boldsymbol{X}_t = \frac{1}{N} \sum_{t=1}^{N} \xi_t \boldsymbol{X}_t.$$

Let us prove each part separately:

F.3.1 Proof of (i) – GLM bounded by M

Here, "bounded by M" means $|y - \langle \boldsymbol{X}, \boldsymbol{\Theta}_\star \rangle| \le M$ *a.s.*. The original proof of Lee et al. (2024a) is too loose, and thus we instead utilize the matrix Bernstein inequality for bounded random matrices, as stated in Lemma F.6. Let $\boldsymbol{A}_t = \xi_t \boldsymbol{X}_t$, which satisfies $\boldsymbol{A} \triangleq \mathbb{E}[\boldsymbol{A}_t] = \boldsymbol{0}$ and $\|\boldsymbol{A}_t\|_{\text{op}} \le M$. Its matrix variance statistics is bounded as

$$V_N = \frac{1}{N} \max\left\{ \left\| \sum_{t=1}^{N} \mathbb{E}[\xi_t^2 \boldsymbol{X}_t \boldsymbol{X}_t^\top] \right\|_{\text{op}}, \left\| \sum_{t=1}^{N} \mathbb{E}[\xi_t^2 \boldsymbol{X}_t^\top \boldsymbol{X}_t] \right\|_{\text{op}} \right\}$$

$$= \frac{1}{N} \max \left\{ \left\| \sum_{t=1}^{N} \mathbb{E}\left[\boldsymbol{X}_t \boldsymbol{X}_t^\top \mathbb{E}[\xi_t^2 \mid \boldsymbol{X}_t] \right] \right\|_{\mathrm{op}}, \left\| \sum_{t=1}^{N} \mathbb{E}\left[\boldsymbol{X}_t^\top \boldsymbol{X}_t \mathbb{E}[\xi_t^2 \mid \boldsymbol{X}_t] \right] \right\|_{\mathrm{op}} \right\} \quad \text{(Tower property of expectation)}$$

$$\leq \frac{g(\tau)}{N} \sum_{t=1}^{N} \mathbb{E}[\dot{\mu}(\langle \boldsymbol{X}_t, \boldsymbol{\Theta}_\star \rangle)] \qquad (\mathbb{E}[\xi_t^2] = \mathrm{Var}[y|\boldsymbol{X}_t] = g(\tau)\dot{\mu}(\langle \boldsymbol{X}_t, \boldsymbol{\Theta}_\star \rangle),\ \|\boldsymbol{X}_t\|_{\mathrm{op}} \leq 1)$$

$$\leq g(\tau) \underbrace{\mathbb{E}_{\boldsymbol{X} \sim \pi}\left[\dot{\mu}(\langle \boldsymbol{X}, \boldsymbol{\Theta}_\star \rangle) \right]}_{\triangleq \bar{\kappa}_\star}.$$

We now recall the high-probability version of the matrix Bernstein inequality:

Lemma F.7 (Matrix Bernstein Inequality, Theorem 6.1.1 of Tropp (2015))**.** *Let $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_N$ be $d_1 \times d_2$ matrices with $\|\boldsymbol{A}_i\|_{\mathrm{op}} \leq M$, $\mathbb{E}[\boldsymbol{A}_i] = \boldsymbol{0}$ and $V_N := \frac{1}{N} \max \left\{ \left\| \sum_{i=1}^{N} \mathbb{E}[\boldsymbol{A}_i \boldsymbol{A}_i^\top] \right\|_{\mathrm{op}}, \left\| \sum_{i=1}^{N} \mathbb{E}[\boldsymbol{A}_i^\top \boldsymbol{A}_i] \right\|_{\mathrm{op}} \right\}$ be the matrix variance statistics. Then, for any $\delta \in (0,1)$, as long as $N \geq \frac{4M^2}{9V_N} \log \frac{4(d_1+d_2)}{\delta}$,*

$$\mathbb{P}\left(\left\| \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{A}_i \right\|_{\mathrm{op}} \leq 2\sqrt{\frac{V_N}{N} \log \frac{4(d_1+d_2)}{\delta}} \right) \geq 1 - \frac{\delta}{4}.$$

The above lemma implies that as long as $N \geq \frac{4M^2}{9g(\tau)\bar{\kappa}_\star} \log \frac{4(d_1+d_2)}{\delta}$, the following holds w.p. at least $1 - \frac{\delta}{4}$:

$$\|\nabla \mathcal{L}_N(\boldsymbol{\Theta}_\star)\|_{\mathrm{op}} \leq 2\sqrt{\frac{g(\tau)\bar{\kappa}_\star}{N} \log \frac{4(d_1+d_2)}{\delta}}.$$

Remark 8. *Klopp (2014); Klopp et al. (2015); Lafond (2015) have utilized similar proof techniques involving matrix Bernstein inequalities for bounded random variables. In the next section, we show how to extend the proof beyond boundedness.*

F.3.2 Proof of (ii) – GLM with Finite SubExponential Norm

Here, we utilize the following subexponential version of matrix Bernstein inequality, whose proof we provide in Appendix F.3.3 for completeness:

Lemma F.8. *Let $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_N$ be $d_1 \times d_2$ random matrices with $\|\boldsymbol{A}_t\|_{\mathrm{op}} \leq 1$, and let $\xi_1, \cdots, \xi_N$ be real-valued random variables with $\mathbb{E}[\xi_t | \boldsymbol{A}_t] = 0$ and $\|\xi\|_{\psi_1} \leq K$ conditionally on $\boldsymbol{A}_t$. Then, for any $\delta \in (0,1)$, as long as $N \geq \log \frac{4(d_1+d_2)}{\delta}$,*

$$\mathbb{P}\left(\left\| \frac{1}{N} \sum_{t=1}^{N} \xi_t \boldsymbol{A}_t \right\|_{\mathrm{op}} \leq 3K\sqrt{\frac{1}{N} \log \frac{4(d_1+d_2)}{\delta}} \right) \geq 1 - \frac{\delta}{}.$$

The statement then immediately follows from the above lemma with $\boldsymbol{A}_t = \boldsymbol{X}_t$. □

F.3.3 Proof of Lemma F.8

We largely follow the proof of Tropp (2012, Theorem 4.1), but adopt to rectangular matrices via Hermitian dilations. Unlike the remark in Tropp (2012, Remark 6.3), as we deal with scalar-valued random variables with matrix-variate coefficients, the rectangular extension is rather straightforward.

First, by definition and Taylor series expansion, $\|\xi_t\|_{\psi_1} \leq K$ conditionally on $\boldsymbol{A}_t$ implies that

$$\mathbb{E}\left[\sum_{p=1}^{\infty} \frac{|\xi_t|^p}{p! \cdot K^p} \middle| \boldsymbol{A}_t \right] \leq 1 \Longrightarrow \mathbb{E}\left[|\xi_t|^p \mid \boldsymbol{A}_t \right] \leq p! K^p,\ \forall p \geq 1.$$

Now, we consider the Hermitian dilation of $\xi_t \boldsymbol{A}_t$, which gives

$$\mathcal{H}(\xi_t \boldsymbol{A}_t) := \begin{bmatrix} \boldsymbol{0} & \xi_t \boldsymbol{A}_t \\ \xi_t \boldsymbol{A}_t^\top & \boldsymbol{0} \end{bmatrix} = \xi_t \mathcal{H}(\boldsymbol{A}_t).$$

Thus, note that for each $p \geq 2$, we have the following moment bound:

$$\begin{aligned} \mathbb{E}\left[\mathcal{H}(\xi_t \boldsymbol{A}_t)^p\right] &\preceq \mathbb{E}[|\xi_t|^p]\boldsymbol{I} && (\|\mathcal{H}(\boldsymbol{A}_t)\|_{\mathrm{op}} = \|\boldsymbol{A}_t\|_{\mathrm{op}} \leq 1) \\ &= \mathbb{E}\left[\mathbb{E}\left[|\xi_t|^p \mid \boldsymbol{X}_t\right]\right] && \text{(Tower property of expectation)} \\ &\preceq p! K^p \boldsymbol{I}. && \text{(Eqn. (F.3.3))} \end{aligned}$$

Then, the moment generating function is bounded as follows: for any $\theta \in (0, K^{-1})$,

$$\begin{aligned} \mathbb{E}\left[e^{\theta \mathcal{H}(\xi_t \boldsymbol{A}_t)}\right] &= \boldsymbol{I} + \theta \mathbb{E}[\mathcal{H}(\xi_t \boldsymbol{A}_t)] + \sum_{p=2}^{\infty} \frac{\theta^p \mathbb{E}[\mathcal{H}(\xi_t \boldsymbol{A}_t)^p]}{p!} \\ &\preceq \boldsymbol{I} + \boldsymbol{I} \sum_{p=2}^{\infty} (\theta K)^p && \text{(The above moment bound)} \\ &= \boldsymbol{I} + \frac{(\theta K)^2}{1 - \theta K} \boldsymbol{I} \\ &\preceq \exp\left(\frac{(\theta K)^2}{1 - \theta K} \boldsymbol{I}\right). && \text{(Eqn. (2.3) of Tropp (2012))} \end{aligned}$$

We then apply the Master Tail Bound for the sum of independent random matrices (Tropp, 2012, Corollary 3.7) to obtain

$$\mathbb{P}\left(\left\| \frac{1}{N} \sum_{t=1}^{N} \mathcal{H}(\xi_t \boldsymbol{A}_t) \right\|_{\mathrm{op}} \geq z\right) \leq (d_1 + d_2) \underbrace{\inf_{\theta \in (0, K^{-1})} \exp\left(-\theta N z + \frac{(\theta K)^2}{1 - \theta K} \left\| \sum_{t=1}^{N} \boldsymbol{I} \right\|_{\mathrm{op}}\right)}_{(*)}, \quad \forall z \geq 0.$$

With $\left\|\sum_{t=1}^{N} \boldsymbol{I}\right\|_{\mathrm{op}} = N$ and optimizing for θ, we have that

$$(*) \leq \exp\left(-N\left(\sqrt{1 + \frac{z}{K}} - 1\right)^2\right).$$

Reparametrizing and assuming that $N \geq \log \frac{4(d_1+d_2)}{\delta}$ we have:

$$\mathbb{P}\left(\left\| \frac{1}{N} \sum_{t=1}^{N} \mathcal{H}(\xi_t \boldsymbol{A}_t) \right\|_{\mathrm{op}} \geq 3K \sqrt{\frac{1}{N} \log \frac{4(d_1 + d_2)}{\delta}}\right) \leq \frac{\delta}{4}$$

Finally, we conclude by noting that

$$\left\| \frac{1}{N} \sum_{t=1}^{N} \mathcal{H}(\xi_t \boldsymbol{A}_t) \right\|_{\mathrm{op}} = \left\| \frac{1}{N} \sum_{t=1}^{N} \xi_t \boldsymbol{A}_t \right\|_{\mathrm{op}}.$$

□

G SUMMARY OF PRIOR WORKS ON GENERALIZED TRACE REGRESSION

Here, we review two relevant prior works on generalized trace regression (Fan et al., 2019; Kang et al., 2022), with focus on their problem setup, assumptions, and results.

G.1 Nuclear Penalized MLE of Fan et al. (2019)

We first list their assumptions:

Assumption C's. *Assume the following conditions hold:*

(C1) $\|\mathrm{vec}(\boldsymbol{X}_t)\|_{\psi_2} \le K < \infty$ *for some constant* $K > 0$

(C2) $|\dot{\mu}(z)| \le L_\mu$ *for any* $z \in \mathbb{R}$

(C3) $\lambda_{\min}(\boldsymbol{H}(\pi; \boldsymbol{\Theta}_\star)) > 0$

(C4) $\|\boldsymbol{\Theta}_\star\|_F \gtrsim \alpha\sqrt{d_1 \vee d_2}$ *for some constant* $\alpha > 0$

(C5) $|\ddot{\mu}(z)| \le |z|^{-1}$ *for any* $|z| > 1$[a]

[a]This slightly resembles the general self-concordance condition with a stretch function (Liu et al., 2024).

Fan et al. (2019, Lemma 1) shows that under (C1-C2), $\lambda_{N_1} \asymp K\sqrt{\frac{d_1 \vee d_2}{N_1}}$ suffices to control the gradient of the log-likelihood loss; Fan et al. (2019, Lemma 2) shows that under (C1-C5), local restricted convexity holds with curvature $\lambda_{\min}(\boldsymbol{H}(\pi; \boldsymbol{\Theta}_\star))$.

Compared to our setting, our bounded covariate assumption is stronger than their (C1), but we do not require (C4) nor (C5), and by utilizing the boundedness, we obtain tighter guarantees. Moreover, the condition (C5) excludes many GLMs, such as Poisson distribution.

G.2 Stein's Lemma-based Estimator of Kang et al. (2022)

Their Stein's lemma-based estimator achieves the following error bound (Kang et al., 2022, Theorem 4.1):

$$\left\|\widehat{\boldsymbol{\Theta}}^{\mathrm{Kang},1} - \boldsymbol{\Theta}_\star\right\|_F^2 \lesssim \frac{M(\pi)(d_1 \vee d_2)r}{\overline{\kappa}(\pi)^2 N},$$

where $\overline{\kappa}(\pi) := \mathbb{E}_{\boldsymbol{X}\sim\pi}[\dot{\mu}(\langle \boldsymbol{X}, \boldsymbol{\Theta}_\star\rangle)]$, *given* that the following assumption holds:

Assumption 3.3. *π has a continuously differentiable density $\boldsymbol{p} = (p_{ij})$ supported over $\mathcal{X}$ such that $\mathbb{E}[(S^{\boldsymbol{p}}(\boldsymbol{X}))_{ij}^2] \le M(\pi)$ for all i, j, where the matrix-valued score function is defined as $S^{\boldsymbol{p}}(\boldsymbol{X}) := \left(-\nabla_{X_{ij}} \log p_{ij}(X_{ij})\right)_{ij}$. Moreover, the columns or rows of random matrix $\boldsymbol{X}$ are pairwisely independent.*

We note that $M(\pi)$ often scales with the dimension. For $\mathcal{X} = \mathcal{B}_F^{d_1\times d_2}(1)$ and $\pi \sim \mathcal{N}(\boldsymbol{0}, \frac{c}{d_1 d_2 \log T}\boldsymbol{I})$ for a constant $c > 0$, it can be computed that $M(\pi) \lesssim d_1 d_2$ (Jang et al., 2024, Appendix H.2), which is what we use in Table 1. Otherwise, we set $M \lesssim (d_1 \vee d_2)^2$ as suggested by Kang et al. (2022).

The continuous differentiability assumption is in place because they rely on the generalized Stein's lemma (Stein et al., 2004, Proposition 1.4). This limits their applicability to discrete arm-sets, while our framework is applicable for both continuous and discrete arm-sets. Also, from the perspective of optimal experimental design, it is not clear how to optimize their bound for π while satisfying the conditions above. Even without those conditions, the function $\pi \mapsto \frac{M(\pi)}{\overline{\kappa}(\pi)}$ is likely to be nonconvex. On the other hand, we mention that their result is applicable to the semiparametric setting of $y_t = \mu(\langle \boldsymbol{X}_t, \boldsymbol{\Theta}_\star\rangle) + \eta_t$ for some subGaussian noise η_t.

G.3 Nuclear Penalized MLE of Kang et al. (2022)

We also briefly elaborate on the nuclear penalized MLE as analyzed in Kang et al. (2022, Appendix J.4):

$$\left\|\widehat{\boldsymbol{\Theta}}^{\mathrm{Kang},2} - \boldsymbol{\Theta}_\star\right\|_F^2 \lesssim \frac{(d_1 \vee d_2) r \sigma(\pi)^2}{c_\mu^2 \lambda_{\min}(\boldsymbol{V}(\pi))^2 N},$$

given that the following assumptions hold:

Assumption J.1. *$\pi \in \mathcal{P}(\mathcal{A})$ is such that $\mathrm{vec}(\boldsymbol{X})$ is $\sigma(\pi)$-subGaussian*[a] *for $\boldsymbol{X} \sim \pi$.*

[a]This means that for any unit vector $\boldsymbol{u} \in \mathcal{S}^{d_1 d_2 - 1}$, $\boldsymbol{u}^\top \mathrm{vec}(\boldsymbol{X})$ is $\sigma(\pi)$-subGaussian.

Assumption J.2. *There is two (dimension-independent) constants $S_2 \le S$ such that $\mathcal{A} \subseteq \mathcal{B}^{d_1 \times d_2} \triangleq \mathcal{B}_F^{d_1 \times d_2}(S) \cap \mathcal{B}_{\mathrm{op}}^{d_1 \times d_2}(S_2)$ and likewise for $\boldsymbol{\Theta}_\star$.*

Assumption J.3. *There is a constant $c_2 > 0$ such that*

$$c_\mu := \min\left(\inf_{\boldsymbol{X} \in \mathcal{A}, \boldsymbol{\Theta} \in \mathcal{B}^{d_1 \times d_2}} \dot{\mu}(\langle \boldsymbol{X}, \boldsymbol{\Theta} \rangle), \inf_{|z| \le (S+2)\sigma c_2} \dot{\mu}(z)\right) > 0.$$

Kang et al. (2022, Assumption J.1) assumed that $\lambda_{\min}(\boldsymbol{V}(\pi)) \asymp \sigma(\pi)^2 \asymp \frac{1}{d_1 d_2}$, which was also the assumption made by Lu et al. (2021, Assumption 2). Indeed, as argued by the two works, one can easily find π that satisfies the above conditions, e.g. $\mathrm{Unif}(\mathcal{B}_F^{d_1 \times d_2}(1))$ or require for "the convex hull of a subset of arms to contain a ball with radius $R \le 1$ that does not scale with d_1 or d_2." Similarly, it is unclear how to optimize for π in the optimal experimental design setup. Moreover, the above assumption may fail even for a simple arm-set. Consider $\pi \sim \mathrm{Unif}(\mathcal{M})$, where recall that $\mathcal{M}$ is the matrix completion basis (see Proposition 3.2). Then, one can show that $\lambda_{\min}(\boldsymbol{V}(\pi)) = \frac{1}{d_1 d_2}$ while $\sigma(\pi)^2 = 1$, leading to a suboptimal guarantee. Another point is that their bound scales inversely with theglobally worst-case curvature c_μ,, while our bound scales inversely with the instance-specific curvature $\kappa_\star$. We note that depending on the arm-set geometry, it may be that $\kappa_\star \gg c_\mu$ (Abeille et al., 2021).

H COMPUTATIONAL COMPLEXITY OF GL-LowPopArt

Note that there are three computational bottlenecks of GL-LowPopArt: Computing the nuclear penalized estimator $\mathbf{\Theta}_0$ in Stage I (line 4, Section 3.5), solving the optimal GL-design in Stage II (line 5, Section 3), and inverting the vectorized Hessian (line 8, Section 3.2). The last component incurs $\mathcal{O}((d_1d_2)^3)$ time and $\mathcal{O}((d_1d_2)^2)$ space complexities via standard Gauss-Jordan elimination or LU decomposition (Golub and Van Loan, 2013); in high dimensions, one could use more efficient algorithms such as the Strassen algorithm (Strassen, 1969) that incurs $\mathcal{O}((d_1d_2)^{\log_2 7})$ time. We now provide theoretical discussions of the computational complexity of the other two components.

Computational Complexity of Nuclear Penalized MLE. Recall the definition of nuclear penalized MLE:

$$\widehat{\mathbf{\Theta}}_0 := \arg\min_{\mathbf{\Theta}\in\Omega} \left\{ \phi(\mathbf{\Theta}) \triangleq \mathcal{L}_{N_1}(\mathbf{\Theta}) + \lambda_{N_1} \|\mathbf{\Theta}\|_{\text{nuc}} \right\},$$

where $\lambda_{N_1} = \sqrt{\frac{C}{N_1}\log\frac{d_1+d_2}{\delta}}$ as in Proposition F.1.

From optimization viewpoint, although $\phi(\cdot)$ is not globally strongly convex, it satisfies Restricted Strong Convexity (RSC); see Lemma F.1. Thus, if the optimization iterates somehow stay within the restricted set of directions where the RSC condition holds, the algorithm achieves a geometric contraction rate, requiring only $\mathcal{O}(\log \epsilon_1^{-1})$ iterations to yield an approximate minimizer $\widetilde{\mathbf{\Theta}}_0$ that satisfies $\phi(\widetilde{\mathbf{\Theta}}_0) - \phi(\widehat{\mathbf{\Theta}}_0) \le \epsilon_1$, for $\epsilon_1 \gtrsim \frac{1}{\lambda_{\min}(\boldsymbol{H}(\pi;\mathbf{\Theta}_\star))}\sqrt{\frac{Cr}{N_1}\log\frac{d_1+d_2}{\delta}}$.[6] Agarwal et al. (2012, Theorem 2) proved that the composite gradient descent (CGD) (Nesterov, 2007) satisfies such property, and attains the geometric contraction rate. Note that per time and space complexity of CGD is $\mathcal{O}(d_1d_2(N_1+\min(d_1,d_2)))$ and $\mathcal{O}(d_1d_2)$, respectively, i.e., the total time and space complexity for CGD to obtain ϵ_1-optimal nuclear penalized MLE is $\mathcal{O}\left(d_1d_2(N_1+\min(d_1,d_2))\log\epsilon_1^{-1}\right)$ and $\mathcal{O}(d_1d_2)$, respectively.

We now discuss the required level of ϵ_1 to ensure that our error rates remain in-tact. First, by definition, we have that

$$\mathcal{L}_{N_1}(\widetilde{\mathbf{\Theta}}_0) + \lambda_{N_1}\left\|\widetilde{\mathbf{\Theta}}_0\right\|_{\text{nuc}} \le \mathcal{L}_{N_1}(\widehat{\mathbf{\Theta}}_0) + \lambda_{N_1}\left\|\widehat{\mathbf{\Theta}}_0\right\|_{\text{nuc}} + \epsilon_1 \le \mathcal{L}_{N_1}(\mathbf{\Theta}_\star) + \lambda_{N_1}\|\mathbf{\Theta}_\star\|_{\text{nuc}} + \epsilon_1.$$

Let $\mathbf{\Delta} = \widetilde{\mathbf{\Theta}}_0 - \mathbf{\Theta}_\star$. Rearranging the terms, subtracting $\langle \nabla\mathcal{L}_{N_1}(\mathbf{\Theta}_\star), \mathbf{\Delta}\rangle$ from both sides, and applying the RSC condition yields:

$$\underbrace{\mathcal{L}_{N_1}(\widetilde{\mathbf{\Theta}}_0) - \mathcal{L}_{N_1}(\mathbf{\Theta}_\star) - \langle \nabla\mathcal{L}_{N_1}(\mathbf{\Theta}_\star), \mathbf{\Delta}\rangle}_{\gtrsim \lambda_{\min}(\boldsymbol{H}(\pi;\mathbf{\Theta}_\star))\|\mathbf{\Delta}\|_F^2} \le -\langle \nabla\mathcal{L}_{N_1}(\mathbf{\Theta}_\star), \mathbf{\Delta}\rangle + \lambda_{N_1}(\|\mathbf{\Theta}_\star\|_{\text{nuc}} - \|\mathbf{\Theta}_\star + \mathbf{\Delta}\|_{\text{nuc}}) + \epsilon_1,$$

where we absorb the tolerance term $L_\mu\sqrt{\frac{1}{N_1}\log\frac{d_1+d_2}{\delta}}\|\mathbf{\Delta}\|_{\text{nuc}}^2$ into the curvature term by assuming that N_1 is sufficiently large.

For the right-hand side, we condition on the event that $\lambda_{N_1} \ge 2\|\nabla\mathcal{L}_{N_1}(\mathbf{\Theta}_\star)\|_{\text{op}}$, which holds with high probability by Proposition F.1. Applying the decomposability of the nuclear norm with respect to the rank-r subspace M of $\mathbf{\Theta}_\star$ (Wainwright, 2019, Chapter 10.2), we obtain:

$$\lambda_{\min}(\boldsymbol{H}(\pi;\mathbf{\Theta}_\star))\|\mathbf{\Delta}\|_F^2 \lesssim \lambda_{N_1}\|\mathbf{\Delta}\|_{\text{nuc}} + \lambda_{N_1}(\|\mathbf{\Delta}_M\|_{\text{nuc}} - \|\mathbf{\Delta}_{M^\perp}\|_{\text{nuc}}) + \epsilon_1 \lesssim \lambda_{N_1}\|\mathbf{\Delta}_M\|_{\text{nuc}} + \epsilon_1.$$

Since $\text{rank}(\mathbf{\Delta}_M) \le 2r$, we have $\|\mathbf{\Delta}_M\|_{\text{nuc}} \le \sqrt{2r}\|\mathbf{\Delta}_M\|_F \le \sqrt{2r}\|\mathbf{\Delta}\|_F$. This gives a simple quadratic inequality:

$$\lambda_{\min}(\boldsymbol{H}(\pi;\mathbf{\Theta}_\star))\|\mathbf{\Delta}\|_F^2 - \lambda_{N_1}\sqrt{r}\|\mathbf{\Delta}\|_F - \epsilon_1 \lesssim 0.$$

Solving for $\mathbf{\Delta}$ and plugging in λ_{N_1} gives

$$\|\mathbf{\Delta}\|_F = \left\|\widetilde{\mathbf{\Theta}}_0 - \mathbf{\Theta}_\star\right\|_F \lesssim \frac{1}{\lambda_{\min}(\boldsymbol{H}(\pi;\mathbf{\Theta}_\star))}\sqrt{\frac{Cr}{N_1}\log\frac{d_1+d_2}{\delta}} + \underbrace{\sqrt{\frac{\epsilon_1}{\lambda_{\min}(\boldsymbol{H}(\pi;\mathbf{\Theta}_\star))}}}_{(*)}.$$

[6] This lower bound on ϵ_1 is intuitively such that the geometric contraction holds until the optimization error hits the inherent statistical noise floor of the problem.

To ensure the optimization error $(*)$ does not dominate, it suffices to set

$$\epsilon_1 \lesssim \frac{Cr}{\lambda_{\min}(\boldsymbol{H}(\pi;\boldsymbol{\Theta}_\star))N_1} \log \frac{d_1+d_2}{\delta},$$

which results in the following time and space complexities:

$$\mathcal{O}\left(d_1 d_2 (N_1 + \min(d_1,d_2)) \log\left(\frac{\lambda_{\min}(\boldsymbol{H}(\pi;\boldsymbol{\Theta}_\star))N_1}{r}\right)\right) \text{ time and } \mathcal{O}\left(d_1 d_2\right) \text{ space.}$$

Computational Complexity of Optimal GL-Design. First, recall the optimal GL-design problem:

$$\min_{\pi\in\mathcal{P}(\mathcal{X})} \left\{ \mathrm{GL}(\pi;\boldsymbol{\Theta}_\star) = \max\{H^{(\mathrm{row})}(\pi;\boldsymbol{\Theta}_\star), H^{(\mathrm{col})}(\pi;\boldsymbol{\Theta}_\star)\} \right\},$$

where

$$H^{(\mathrm{row})}(\pi;\boldsymbol{\Theta}_\star) := \lambda_{\max}\left(\sum_{m=1}^{d_2} \boldsymbol{D}_m{}^{(\mathrm{row})}(\pi;\boldsymbol{\Theta}_\star)\right), \quad \boldsymbol{D}_m{}^{(\mathrm{row})}(\pi_2) := [(\boldsymbol{H}(\pi;\boldsymbol{\Theta}_\star)^{-1})_{jk}]_{j,k\in\mathcal{I}^{(\mathrm{row})}},$$

$$H^{(\mathrm{col})}(\pi;\boldsymbol{\Theta}_\star) := \lambda_{\max}\left(\sum_{m=1}^{d_1} \boldsymbol{D}_m{}^{(\mathrm{col})}(\pi;\boldsymbol{\Theta}_\star)\right), \quad \boldsymbol{D}_m{}^{(\mathrm{col})}(\pi;\boldsymbol{\Theta}_\star) := (\boldsymbol{H}(\pi;\boldsymbol{\Theta}_\star)^{-1})_{\mathcal{I}^{(\mathrm{col})},\mathcal{I}^{(\mathrm{col})}},$$

and the index sets defined as $\mathcal{I}^{(\mathrm{row})} := \{d_1(l-1)+m : l\in[d_2]\}$ and $\mathcal{I}^{(\mathrm{col})} := [d_1(m-1)+1 : d_1 m]$.

We elaborate on how to equivalently reformulate the GL-design as a Semidefinite Program (SDP). To do so, we first introduce an auxiliary positive semidefinite matrix variable $\boldsymbol{Z}\in\mathbb{S}_+^{d_1 d_2}$ to upper bound the inverse of principal submatrices of the inverse Hessian, and a scalar $w\in\mathbb{R}$ to bound the maximum eigenvalue.

By applying the Schur complement (Boyd et al., 1994), the condition $\boldsymbol{Z}\succeq\boldsymbol{H}(\pi;\boldsymbol{\Theta}_\star)^{-1}$ is strictly equivalent to the Linear Matrix Inequality (LMI):

$$\underbrace{\begin{pmatrix} \boldsymbol{H}(\pi;\boldsymbol{\Theta}_\star) & \boldsymbol{I}_{d_1 d_2} \\ \boldsymbol{I}_{d_1 d_2} & \boldsymbol{Z} \end{pmatrix}}_{(**)} \succeq 0.$$

Defining the principal submatrix extraction operators $\mathcal{P}_m^{(\mathrm{row})}(\boldsymbol{Z}) := \boldsymbol{Z}_{\mathcal{I}_m^{(\mathrm{row})},\mathcal{I}_m^{(\mathrm{row})}}$ and $\mathcal{P}_m^{(\mathrm{col})}(\boldsymbol{Z}) := \boldsymbol{Z}_{\mathcal{I}_m^{(\mathrm{col})},\mathcal{I}_m^{(\mathrm{col})}}$ (which are linear operators), the exact SDP reformulation of the GL-design is given by:

$$\begin{aligned}
\min_{\pi\in\mathbb{R}^{|\mathcal{X}|},\boldsymbol{Z}\in\mathbb{S}^{d_1 d_2},w\in\mathbb{R}} \quad & w \\
\text{subject to} \quad & \sum_{x\in\mathcal{X}} \pi(x) = 1, \quad \pi(x)\geq 0 \quad \forall x\in\mathcal{X}, \\
& \begin{pmatrix} \sum_{x\in\mathcal{X}} \pi(x)\boldsymbol{H}(x;\boldsymbol{\Theta}_\star) & \boldsymbol{I}_{d_1 d_2} \\ \boldsymbol{I}_{d_1 d_2} & \boldsymbol{Z} \end{pmatrix} \succeq 0, \\
& \sum_{m=1}^{d_2} \mathcal{P}_m^{(\mathrm{row})}(\boldsymbol{Z}) \preceq w\boldsymbol{I}_{d_1}, \\
& \sum_{m=1}^{d_1} \mathcal{P}_m^{(\mathrm{col})}(\boldsymbol{Z}) \preceq w\boldsymbol{I}_{d_2}.
\end{aligned}$$

By employing standard interior-point methods (Nesterov and Nemirovski, 1994), the time complexity to solve this SDP to an ϵ_2-optimal solution scales polynomially with the number of variables and the size of the LMI, yielding $\mathcal{O}\left(\mathrm{poly}(|\mathcal{X}|,d_1,d_2)\log\epsilon_2^{-1}\right)$. The space complexity is dominated by the augmented matrix $(**)$, requiring $\mathcal{O}((d_1 d_2)^2)$ memory.

I E-OPTIMAL DESIGN FOR STAGE I

Here, we briefly elaborate on the use of E-optimal design for Stage I. As discussed in Remark 2, from the error rate of Stage I (Theorem 3.2), the natural optimal design is $\arg\max_{\pi\in\mathcal{P}(\mathcal{X})}\lambda_{\min}(\boldsymbol{H}(\pi;\boldsymbol{\Theta}_\star))$. However, as we do not have any prior knowledge about $\boldsymbol{\Theta}_\star$, the best one could do is to consider a naïve lower bound of $\lambda_{\min}(\boldsymbol{H}(\pi;\boldsymbol{\Theta}_\star)) \geq \kappa_\star\lambda_{\min}(\boldsymbol{V}(\pi))$. This motivates the following optimal design:

$$\pi_E \leftarrow \arg\max_{\pi\in\mathcal{P}(\mathcal{X})}\lambda_{\min}(\boldsymbol{V}(\pi)),$$

known as the *E-optimal design* (Pukelsheim, 2006), previously considered in sparse linear bandits (Hao et al., 2020) and bandit phase retrieval (Lattimore and Hao, 2021). This is a convex optimization problem, and can be solved efficiently. Below, we collect some results regarding the computation aspects of π_E.

Unlike G-optimal (or equivalently, D-optimal) design where it is guaranteed that one can obtain an optimal design with support size at most $\frac{d_1d_2(d_1d_2+1)}{2}+1$ (Kiefer and Wolfowitz, 1960; Todd, 2016), there is no analogous guarantee on the support size of π_E. One could however obtain an *approximate* E-optimal design with guaranteed bounded support size.

First, if $\mathcal{X}$ is discrete, then one can use the polynomial-time algorithm of Allen-Zhu et al. (2021) to obtain $\widehat{\pi}_E$ satisfying $|\mathrm{supp}(\widehat{\pi}_E)| \lesssim d_1d_2$ and $\lambda_{\min}(\boldsymbol{V}(\widehat{\pi}_E)) \geq \frac{1}{2}\lambda_{\min}(\boldsymbol{V}(\pi_E))$.

Now suppose that $\mathcal{X}$ is continuous. For this, we employ two-stage approach: first solve the E-optimal design, then sparsify its support via *ϵ-approximate Carathéodory solver* (Barman, 2015; Combettes and Pokutta, 2023; Mirrokni et al., 2017),[7] [8] which outputs a $\widehat{\pi}_E$ such that $\|\boldsymbol{V}(\pi_E)-\boldsymbol{V}(\widehat{\pi}_E)\|_F \leq \epsilon$ and $|\mathrm{supp}(\widehat{\pi}_E)| \lesssim \frac{(d_1\wedge d_2)^2}{\epsilon^2}$. The approximation error in the objective $\lambda_{\min}(\boldsymbol{V}(\pi))$ is controlled via the Hoffman-Wielandt inequality for eigenvalue perturbations (Hoffman and Wielandt, 1953):

$$|\lambda_{\min}(\boldsymbol{V}(\pi_E)) - \lambda_{\min}(\boldsymbol{V}(\widehat{\pi}_E))| \leq \|\boldsymbol{V}(\pi_E)-\boldsymbol{V}(\widehat{\pi}_E)\|_F \leq \epsilon.$$

The sensible choice for ϵ is $\epsilon = \frac{\lambda_{\min}(\boldsymbol{V}(\pi_E))}{2}$, but in practice, it is often difficult to know $\lambda_{\min}(\boldsymbol{V}(\pi_E))$ beforehand. But for special arm-sets, this is possible. For instance, when $\mathcal{X} = \mathcal{B}_{\mathrm{op}}^{d_1\times d_2}(1)$, we have that $\lambda_{\min}(\boldsymbol{V}(\pi_E)) = \frac{1}{d_1\vee d_2}$ (Jang et al., 2024, Appendix D.2).

[7] Recently, Combettes and Pokutta (2023) showed that the Frank-Wolfe algorithm (Frank and Wolfe, 1956) is effective in solving the approximate Carathéodory problem, making it as efficient as solving the G-optimal design with bounded support (Todd, 2016).

[8] The approximate Carathéodory theorem (Barman, 2015, Theorem 2) states that $|\mathrm{supp}(\widehat{\pi}_E)| \lesssim \epsilon^{-2}\mathrm{diam}(\mathrm{vec}(\mathcal{X}))^2$ where $\mathrm{vec}(\mathcal{X}) := \{\mathrm{vec}(\boldsymbol{X})\mathrm{vec}(\boldsymbol{X})^\top : \boldsymbol{X}\in\mathcal{X}\}$, and we have that $\mathrm{diam}(\mathrm{vec}(\mathcal{X}))^2 \leq 4(d_1\wedge d_2)^2$ when $\mathcal{X}\subseteq\mathcal{B}_{\mathrm{op}}^{d_1\times d_2}(1)$.

J PROOF OF THEOREM 4.1 – LOCAL MINIMAX LOWER BOUND

WLOG assume that $d_1 = \max(d_1, d_2)$. For given $\boldsymbol{\Theta}_\star$, let $\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^\top$ be its SVD.

Inspired by Rohde and Tsybakov (2011, Theorem 5) and Abeille et al. (2021, Theorem 2), we consider the following set of $d_1 \times d_2$ matrices:

$$\Theta_{r,\varepsilon,\beta} := \left\{ (1-\varepsilon)\,\boldsymbol{\Theta}_\star + \varepsilon \boldsymbol{U}'\boldsymbol{V}^\top \in \mathbb{R}^{d_1 \times d_2} : \boldsymbol{U}' \in \{0,\beta\}^{d_1 \times r} \right\},$$

where $\varepsilon \in (0,1)$ and $\beta > 0$ will be specified later. By construction, we have that for any $\boldsymbol{\Theta} \in \Theta_{r,\varepsilon,\beta}$, $\operatorname{rank}(\boldsymbol{\Theta}) \le r$ and

$$\begin{aligned}
\|\boldsymbol{\Theta}\|_{\text{nuc}} &\le (1-\varepsilon)\,\|\boldsymbol{\Theta}_\star\|_{\text{nuc}} + \varepsilon \left\|\boldsymbol{U}'\boldsymbol{V}^\top\right\|_{\text{nuc}} \\
&= (1-\varepsilon) S_* + \varepsilon \|\boldsymbol{U}'\|_{\text{nuc}} && \text{(unitary invariance of } \|\cdot\|_{\text{nuc}}) \\
&\le (1-\varepsilon) S_* + \varepsilon \sqrt{r}\, \|\boldsymbol{U}'\|_F && \text{(Cauchy-Schwartz inequality on the singular values of } \boldsymbol{U}') \\
&\le (1-\varepsilon) S_* + \varepsilon \beta r \sqrt{d_1}. && \text{(by construction)}
\end{aligned}$$

Thus, it can be verified that $\beta \le \frac{S_*}{r\sqrt{d_1}}$ implies $\|\boldsymbol{\Theta}\|_{\text{nuc}} \le S_*$, i.e., $\Theta_{r,\varepsilon,\beta} \subset \mathcal{N}(\boldsymbol{\Theta}_\star; \varepsilon, r, S_*)$.

By construction, $\|\boldsymbol{\Theta}_1 - \boldsymbol{\Theta}_2\|_F^2$ is closely related to the Hamming distance of the $\operatorname{vec}(\boldsymbol{U}')$'s, which are basically binary sequences. With this, we recall the Gilbert-Varshamov bound:

Lemma J.1 (Gilbert–Varshamov bound; Lemma 2.9 of Tsybakov (2009); Theorem 1 of Gilbert (1952); Varshamov (1964))**.** *Let $m \ge 8$ and $\Omega := \{0,1\}^m$. Then there exists $\{\omega^{(0)}, \omega^{(1)}, \cdots, \omega^{(M)}\} \subset \Omega$ with $M \ge 2^{m/8}$ such that $\omega^{(0)} = (0, \cdots, 0)$ and*

$$d_H(\omega^{(j)}, \omega^{(k)}) := \sum_{\ell=1}^{m} \mathbb{1}[(\omega^{(j)})_\ell \ne (\omega^{(k)})_\ell] \ge \frac{m}{8}, \quad \forall 0 \le j < k \le M.$$

Thus, we can find a $\Theta^0_{r,\varepsilon,\beta} \subset \Theta_{r,\varepsilon,\beta}$ such that $|\Theta^0_{r,\varepsilon,\beta}| \ge 2^{\frac{rd_1}{8}}$, and for any $\boldsymbol{\Theta}_i = (1-\varepsilon)\boldsymbol{\Theta}_\star + \varepsilon \boldsymbol{U}_i'\boldsymbol{D}\boldsymbol{V}^\top \in \Theta^0_{r,\varepsilon,\beta}$ with $i \in \{1,2\}$ and $\boldsymbol{U}_1 \ne \boldsymbol{U}_2$,

$$\|\boldsymbol{\Theta}_1 - \boldsymbol{\Theta}_2\|_F^2 = \varepsilon^2 \left\|(\boldsymbol{U}_1' - \boldsymbol{U}_2')\boldsymbol{V}^\top\right\|_F^2 = \varepsilon^2 \|(\boldsymbol{U}_1' - \boldsymbol{U}_2')\|_F^2 \ge \varepsilon^2 \frac{\beta^2 r d_1}{8},$$

where we denote $\sigma_{\min} = \sigma_{\min}(\boldsymbol{\Theta}_\star)$ to be the minimum non-zero singular value of $\boldsymbol{\Theta}_\star$.

Furthermore, we have that for any $\boldsymbol{\Theta} = (1-\varepsilon)\boldsymbol{\Theta}_\star + \varepsilon \boldsymbol{U}'\boldsymbol{V}^\top \in \Theta^0_{r,\varepsilon,\beta}$,

$$\begin{aligned}
\left\|\boldsymbol{\Theta}_\star - \left((1-\varepsilon)\boldsymbol{\Theta}_\star + \varepsilon \boldsymbol{U}'\boldsymbol{V}^\top\right)\right\|_F^2 &= \varepsilon^2 \left\|\boldsymbol{\Theta}_\star - \boldsymbol{U}'\boldsymbol{V}^\top\right\|_F^2 \\
&\ge \varepsilon^2 \left(\|\boldsymbol{\Theta}_\star\|_F^2 - \|\boldsymbol{U}'\|_F^2\right) && \text{(triangle inequality and unitary invariance of } \|\cdot\|_F) \\
&\ge \varepsilon^2 \left(\|\boldsymbol{\Theta}_\star\|_F^2 - \beta^2 r d_1\right) && \text{(by construction)} \\
&\ge \varepsilon^2 \frac{\beta^2 r d_1}{8},
\end{aligned}$$

which in turn holds when $\|\boldsymbol{\Theta}_\star\|_F^2 \ge \frac{9\beta^2 r d_1}{8}$. We will see that this indeed holds with our β specified later.

For $\boldsymbol{\Theta} \in \mathbb{R}^{d_1 \times d_2}$, let $\mathbb{P}_{\boldsymbol{\Theta}}$ be the probability distribution of the observations $\{(\boldsymbol{X}_t, y_t)\}_{t \in [N]}$, with $y_t \sim p(\cdot|\boldsymbol{X}_t; \boldsymbol{\Theta})$.

We now compute the KL between $\mathbb{P}_{(1-\varepsilon)\boldsymbol{\Theta}_\star + \varepsilon\boldsymbol{\Theta}'}$ and $\mathbb{P}_{\boldsymbol{\Theta}_\star}$ for any $\boldsymbol{\Theta}' = \boldsymbol{U}'\boldsymbol{V}^\top \in \Theta_{r,\varepsilon,\beta}$ by connecting it with the Bregman divergence:

Definition J.1. *For a $m : \mathbb{R} \to \mathbb{R}$, the* **Bregman divergence** *$D_m(\cdot,\cdot)$ is defined as follows:*

$$D_m(z_1, z_2) := m(z_1) - m(z_2) - m'(z_2)(z_1 - z_2).$$

We recall the following well-known lemma from information geometry, which simplifies the computation of KL between two GLMs by implicitly making use of their dually flat structure (Amari, 2016; Brekelmans et al., 2020; Nielsen, 2020):

Lemma J.2. *Consider two GLMs $p_1 \triangleq p(\cdot|\boldsymbol{X};\boldsymbol{\Theta}_1)$ and $p_2 \triangleq p(\cdot|\boldsymbol{X};\boldsymbol{\Theta}_2)$ with the same log-partition function m. Then, we have that $D_{\mathrm{KL}}(p_2, p_1|\boldsymbol{X}) \triangleq D_{\mathrm{KL}}(p(\cdot|X,\boldsymbol{\Theta}_2),) = \frac{1}{g(\tau)} D_m(\langle \boldsymbol{X},\boldsymbol{\Theta}_1\rangle, \langle \boldsymbol{X},\boldsymbol{\Theta}_2\rangle)$.*

We then have that

$$\begin{aligned} D_{\mathrm{KL}}(\mathbb{P}_{(1-\varepsilon)\boldsymbol{\Theta}_\star+\varepsilon\boldsymbol{\Theta}'}, \mathbb{P}_{\boldsymbol{\Theta}_\star}|\boldsymbol{X}) &= \frac{1}{g(\tau)} D_m(\langle \boldsymbol{X},\boldsymbol{\Theta}_\star\rangle, (1-\varepsilon)\langle \boldsymbol{X},\boldsymbol{\Theta}_\star\rangle + \varepsilon\langle \boldsymbol{X},\boldsymbol{\Theta}'\rangle) \\ &= \frac{1}{g(\tau)}\varepsilon^2\langle \boldsymbol{X},\boldsymbol{\Theta}_\star-\boldsymbol{\Theta}'\rangle^2 \int_0^1 v\dot{\mu}(\langle \boldsymbol{X},\boldsymbol{\Theta}_\star\rangle + \varepsilon\langle \boldsymbol{X},\boldsymbol{\Theta}'-\boldsymbol{\Theta}_\star\rangle v)dv. \\ &\qquad\qquad \text{(Taylor expansion with integral remainder)} \end{aligned}$$

We recall a useful self-concordance control lemma from Abeille et al. (2021); Faury et al. (2020):

Lemma J.3 (A Modification of Lemma 9 of Abeille et al. (2021))**.** *Let $\mu : \mathbb{R} \to \mathbb{R}$ be a strictly increasing function satisfying $|\ddot{\mu}| \le R_s\dot{\mu}$ for some $R_s \ge 0$. Then, for any $z_1, z_2 \in \mathbb{R}$ and $\varepsilon > 0$, $\dot{\mu}(z_1+\varepsilon z_2) \le \dot{\mu}(z_1)\exp(R_s\varepsilon|z_2|)$.*

With this, we have that

$$\begin{aligned} D_{\mathrm{KL}}(\mathbb{P}_{(1-\varepsilon)\boldsymbol{\Theta}_\star+\varepsilon\boldsymbol{\Theta}'}, \mathbb{P}_{\boldsymbol{\Theta}_\star}|\boldsymbol{X}) &\le \frac{1}{g(\tau)}\varepsilon^2\dot{\mu}(\langle \boldsymbol{X},\boldsymbol{\Theta}_\star\rangle)\langle \boldsymbol{X},\boldsymbol{\Theta}_\star-\boldsymbol{\Theta}'\rangle^2 \int_0^1 v\exp(R_s\varepsilon|\langle \boldsymbol{X},\boldsymbol{\Theta}'-\boldsymbol{\Theta}_\star\rangle|v)dv \\ &\le \frac{1}{2g(\tau)}\varepsilon^2\dot{\mu}(\langle \boldsymbol{X},\boldsymbol{\Theta}_\star\rangle)\langle \boldsymbol{X},\boldsymbol{\Theta}_\star-\boldsymbol{\Theta}'\rangle^2\exp(R_s\varepsilon|\langle \boldsymbol{X},\boldsymbol{\Theta}'-\boldsymbol{\Theta}_\star\rangle|) \\ &\overset{(*)}{\le} \frac{1}{2g(\tau)}\varepsilon^2\dot{\mu}(\langle \boldsymbol{X},\boldsymbol{\Theta}_\star\rangle)\langle \boldsymbol{X},\boldsymbol{\Theta}_\star-\boldsymbol{\Theta}'\rangle^2\exp\left(R_s\varepsilon(1+\beta\sqrt{d_1 r})S_*\right) \\ &\le \frac{e}{2g(\tau)}\varepsilon^2\dot{\mu}(\langle \boldsymbol{X},\boldsymbol{\Theta}_\star\rangle)\langle \boldsymbol{X},\boldsymbol{\Theta}_\star-\boldsymbol{\Theta}'\rangle^2, \end{aligned}$$

given that $R_s\varepsilon(1+\beta\sqrt{d_1 r})S_* \le 1$. Note that $(*)$ holds regardless of whether we assume $\mathcal{X} \subseteq \mathcal{B}_F^{d_1\times d_2}(1)$ (which is what we assume in the statement) or $\mathcal{X} \subseteq \mathcal{B}_{\mathrm{op}}^{d_1\times d_2}(1)$ (which is implied from the first case). To see this, if the first case holds, then

$$\langle \boldsymbol{X},\boldsymbol{\Theta}_\star-\boldsymbol{\Theta}'\rangle \le \|\boldsymbol{X}\|_F\|\boldsymbol{\Theta}-\boldsymbol{\Theta}_\star\|_F \le \|\boldsymbol{\Theta}-\boldsymbol{\Theta}_\star\|_{\mathrm{nuc}} \le (1+\beta\sqrt{d_1 r})S_*,$$

and if the second case holds,

$$\langle \boldsymbol{X},\boldsymbol{\Theta}_\star-\boldsymbol{\Theta}'\rangle \le \|\boldsymbol{X}\|_{\mathrm{op}}\|\boldsymbol{\Theta}-\boldsymbol{\Theta}_\star\|_{\mathrm{nuc}} \le (1+\beta\sqrt{d_1 r})S_*.$$

Remark 9. *Lee et al. (2024b, Lemma 4) has utilized a similar argument (Taylor integral remainder with self-concordance) to provide a lower bound on the KL divergence during the online learning regret analysis. However, they restricted their attention to the Bernoulli distribution.*

Thus, recalling that $\boldsymbol{\pi}_N = (\pi_t)_{t\in[N]}$ and using the chain rule for KL (Lattimore and Szepesvári, 2020, Exercise 14.12),

$$\begin{aligned} D_{\mathrm{KL}}(\mathbb{P}_{(1-\varepsilon)\boldsymbol{\Theta}_\star+\varepsilon\boldsymbol{\Theta}'}, \mathbb{P}_{\boldsymbol{\Theta}_\star}) &= \sum_{t=1}^N \mathbb{E}_{\boldsymbol{X}_t\sim\pi_t}[D_{\mathrm{KL}}(\mathbb{P}_{(1-\varepsilon)\boldsymbol{\Theta}_\star+\varepsilon\boldsymbol{\Theta}'}, \mathbb{P}_{\boldsymbol{\Theta}_\star}|\boldsymbol{X}_t)] \\ &\le \frac{e}{2g(\tau)}\varepsilon^2\mathrm{vec}(\boldsymbol{\Theta}_\star-\boldsymbol{\Theta}')^\top\left(\sum_{t=1}^N \boldsymbol{H}(\pi_t;\boldsymbol{\Theta}_\star)\right)\mathrm{vec}(\boldsymbol{\Theta}_\star-\boldsymbol{\Theta}') \\ &\le \frac{eN}{2g(\tau)}\varepsilon^2\lambda_{\max}(\boldsymbol{H}(\boldsymbol{\pi}_N;\boldsymbol{\Theta}_\star))\|\boldsymbol{\Theta}_\star-\boldsymbol{\Theta}'\|_F^2 \qquad (\boldsymbol{H}(\boldsymbol{\pi}_N;\boldsymbol{\Theta}_\star) := \tfrac{1}{N}\textstyle\sum_{t=1}^N \boldsymbol{H}(\pi_t;\boldsymbol{\Theta}_\star)) \end{aligned}$$

$$\leq \frac{eN}{2g(\tau)}\varepsilon^2\lambda_{\max}(\boldsymbol{H}(\boldsymbol{\pi}_N;\boldsymbol{\Theta}_\star))(1+\beta\sqrt{d_1 r})^2 S_*^2.$$

Then we have that

$$\frac{1}{|\Theta^0_{r,\varepsilon}|}\sum_{\boldsymbol{\Theta}'\in\Theta^0_{r,\varepsilon}} D_{\mathrm{KL}}(\mathbb{P}_{\boldsymbol{\Theta}'},\mathbb{P}_{\boldsymbol{\Theta}_\star}) \leq \frac{e\varepsilon^2 N\lambda_{\max}(\boldsymbol{H}(\boldsymbol{\pi}_N;\boldsymbol{\Theta}_\star))(1+\beta\sqrt{d_1 r})^2 S_*^2}{2g(\tau)}$$
$$= \frac{4eN\varepsilon^2\lambda_{\max}(\boldsymbol{H}(\boldsymbol{\pi}_N;\boldsymbol{\Theta}_\star))(1+\beta\sqrt{d_1 r})^2 S_*^2}{g(\tau)rd_1}\frac{rd_1}{8}.$$

As $\log|\Theta^0_{r,\varepsilon,\beta}| \geq \log(2^{\frac{rd_1}{8}}) = \frac{rd_1}{8}\log 2$,

$$\frac{1}{|\Theta^0_{r,\varepsilon,\beta}|}\sum_{\boldsymbol{\Theta}'\in\Theta^0_{r,\varepsilon,\beta}} D_{\mathrm{KL}}(\mathbb{P}_{\boldsymbol{\Theta}'},\mathbb{P}_{\boldsymbol{\Theta}_\star}) \leq \frac{1}{16}\log|\Theta^0_{r,\varepsilon}|$$

holds with $\varepsilon^2 \leq \frac{rd_1 g(\tau)\alpha\log 2}{2^6 eN\lambda_{\max}(\boldsymbol{H}(\boldsymbol{\pi}_N;\boldsymbol{\Theta}_\star))(1+\beta\sqrt{d_1 r})^2 S_*^2}$ where $\alpha = \frac{1}{16}$.

We choose

$$\beta^2 = \frac{\gamma}{rd_1} \Rightarrow \varepsilon^2 = \frac{\alpha\log 2}{2^6 e(1+\sqrt{\gamma})^2}\frac{rd_1 g(\tau)}{N\lambda_{\max}(\boldsymbol{H}(\boldsymbol{\pi}_N;\boldsymbol{\Theta}_\star))S_*^2}.$$

We now check the requirements:

$$\beta \leq \frac{S_*}{r\sqrt{d_1}} \iff \gamma \leq \frac{S_*^2}{r}$$
$$\|\boldsymbol{\Theta}_\star\|_F^2 \geq \frac{9\beta^2 rd_1}{8} \iff \gamma \leq \frac{8}{9}\|\boldsymbol{\Theta}_\star\|_F^2$$
$$R_s\varepsilon(1+\beta\sqrt{d_1 r})S_* \leq 1 \iff N \geq \frac{R_s^2}{2^{10}}\frac{\log 2}{e}\frac{rd_1 g(\tau)}{\lambda_{\max}(\boldsymbol{H}(\boldsymbol{\pi}_N;\boldsymbol{\Theta}_\star))}.$$

The proof concludes by invoking Tsybakov (2009, Theorem 2.5) with $\alpha = \frac{1}{16}$,[9] which we recall here for completeness:

Lemma J.4 (Theorem 2.5 of Tsybakov (2009))**.** *Let Θ be a subset of a metric space with metric $d(\cdot,\cdot)$, and let $\boldsymbol{\theta}\mapsto\mathbb{P}_{\boldsymbol{\theta}}$ be the probability measure parametrized by $\boldsymbol{\theta}$. Suppose that there exists $\{\boldsymbol{\theta}_0,\boldsymbol{\theta}_1,\cdots,\boldsymbol{\theta}_M\}\subset\Theta$ for some $M\geq 2$ such that*

(i) $d(\boldsymbol{\theta}_j,\boldsymbol{\theta}_k)\geq 2b>0, \quad \forall 0\leq j<k\leq M,$

(ii) $\mathbb{P}_{\boldsymbol{\theta}_j}\ll\mathbb{P}_{\boldsymbol{\theta}_0}, \quad \forall j=1,2,\cdots,M,$ *and*

(iii) there exists a $\alpha\in(0,1/8)$ *such that* $\frac{1}{M}\sum_{j=1}^M D_{\mathrm{KL}}(\mathbb{P}_{\boldsymbol{\theta}_j},\mathbb{P}_{\boldsymbol{\theta}_0})\leq\alpha\log M.$

Then, we have the following high-probability minimax lower bound:

$$\inf_{\widehat{\boldsymbol{\theta}}}\sup_{\boldsymbol{\theta}_\star\in\Theta}\mathbb{P}_{\boldsymbol{\theta}_\star}(d(\widehat{\boldsymbol{\theta}},\boldsymbol{\theta}_\star)\geq b)\geq\frac{\sqrt{M}}{1+\sqrt{M}}\left(1-2\alpha-\sqrt{\frac{2\alpha}{\log M}}\right)>0.$$

□

We now provide the proofs of the missing lemmas:

Proof of Lemma J.2. This follows from brute-force computation:

$$D_{\mathrm{KL}}(p_2,p_1) = \mathbb{E}_{y\sim p_2}\left[\log\frac{p_2(y)}{p_1(y)}\right]$$

[9] No efforts were made to optimize the constants.

$$
\begin{aligned}
&= \frac{1}{g(\tau)} \mathbb{E}_{y\sim p_2}\left[y\langle \boldsymbol{X}, \boldsymbol{\Theta}_2 - \boldsymbol{\Theta}_1\rangle + m(\langle \boldsymbol{X}, \boldsymbol{\Theta}_1\rangle) - m(\langle \boldsymbol{X}, \boldsymbol{\Theta}_2\rangle)\right] \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(recall the probability density of GLMs)} \\
&= \frac{m(\langle \boldsymbol{X}, \boldsymbol{\Theta}_1\rangle) - m(\langle \boldsymbol{X}, \boldsymbol{\Theta}_2\rangle) - m'(\langle \boldsymbol{X}, \boldsymbol{\Theta}_2\rangle)\langle \boldsymbol{X}, \boldsymbol{\Theta}_1 - \boldsymbol{\Theta}_2\rangle}{g(\tau)} \qquad (\mathbb{E}[y] = m'(\langle \boldsymbol{X}, \boldsymbol{\Theta}_2\rangle)) \\
&= \frac{1}{g(\tau)} D_m(\langle \boldsymbol{X}, \boldsymbol{\Theta}_1\rangle, \langle \boldsymbol{X}, \boldsymbol{\Theta}_2\rangle).
\end{aligned}
$$

□

Proof of Lemma J.3. We provide the slightly modified proof of Abeille et al. (2021, Lemma 9) for completeness. Starting from the self-concordance, we have that for any $z_1, z_2 \in \mathbb{R}$

$$
-R_s \leq \frac{\ddot{\mu}(z)}{\dot{\mu}(z)} \leq R_s, \quad \forall z \in \mathbb{R} \Longrightarrow -R_s\varepsilon|z_2| \leq \underbrace{\int_{(z_1+\varepsilon z_2)\wedge z_1}^{\dot{\mu}(z_1+\varepsilon z_2)\vee z_1} \frac{\ddot{\mu}(z)}{\dot{\mu}(z)} dz}_{=\log \frac{\dot{\mu}((z_1+\varepsilon z_2)\vee z_1)}{\dot{\mu}((z_1+\varepsilon z_2)\wedge z_1)}} \leq R_s\varepsilon|z_2|.
$$

If $z_2 \geq 0$, then we have that from the upper bound,

$$
\dot{\mu}(z_1 + \varepsilon z_2) \leq \dot{\mu}(z_1)\exp(R_s\varepsilon z_2) = \dot{\mu}(z_1)\exp(R_s\varepsilon|z_2|).
$$

If $z_2 < 0$, then we have that from the lower bound,

$$
\dot{\mu}(z_1 + \varepsilon z_2)\exp(R_s\varepsilon z_2) \leq \dot{\mu}(z_1) \Longrightarrow \dot{\mu}(z_1 + \varepsilon z_2) \leq \dot{\mu}(z_1)\exp(-R_s\varepsilon z_2) = \dot{\mu}(z_1)\exp(R_s\varepsilon|z_2|).
$$

□

K ADDITIONAL DETAILS FOR BILINEAR DUELING BANDITS: SETTING

K.1 Motivation

Transitivity — the property that if $i \succ j$ and $j \succ k$, then $i \succ k$ — is one of the key assumptions that distinguish the dueling bandit setting (Bengs et al., 2021; Sui et al., 2018; Yue and Joachims, 2009; Yue et al., 2012). Within this stochastic transitivity framework, the most commonly considered model is the Bradley-Terry-Luce (BTL) model (Bradley and Terry, 1952): each arm k has an unknown utility(reward) $r_k \in \mathbb{R}$ such that for each $(i, j) \in [K] \times [K]$, $p_{i,j} := \mathbb{P}(i \succ j) = \mu(r_i - r_j)$ with $\mu(z) := (1 + e^{-z})^{-1}$. When K is large, without any additional structural assumption, the statistical guarantees (e.g., regret in dueling bandits) often increase polynomially in K. One very natural way of bypassing this issue is to impose a linear structure on the utility, resulting in the so-called linear BTL model: each arm k is endowed with a known feature vector $\boldsymbol{\phi}_k \in \mathbb{R}^d$ and $r_k = \langle \boldsymbol{\phi}_k, \boldsymbol{\theta}_\star \rangle$ for some unknown $\boldsymbol{\theta}_\star \in \mathbb{R}^d$. This model has been successfully applied in various domains, with reinforcement learning with human feedback (Rafailov et al., 2023) being one of the most prominent applications. Coming back to dueling bandits, with such linear structure, the regret of dueling bandits has been improved from poly(K) to d or $\sqrt{d \log K}$ by exploiting the linear BTL model (Bengs et al., 2022; Saha, 2021).

However, the literature has two main gaps, both of which we intend to fill with our newly proposed setting and new analyses.

Linear-like Structure in Dueling Bandits with General Preferences. The (linear) BTL model cannot model nontransitive preferences, which hinders its applicability in various scenarios, from simple nontransitive games such as rock-paper-scissors, Blotto-style games (Balduzzi et al., 2018, 2019; Bertrand et al., 2023), and even human preferences (Azar et al., 2024; May, 1954; Munos et al., 2024; Swamy et al., 2024; Tversky, 1969; Zhang et al., 2025).

In most of the prior literature on dueling bandits and general preference learning (i.e., not assuming linear BTL model), the learner must either learn or adapt to the entire unstructured preference matrix $\boldsymbol{P} \in [0, 1]^{K \times K}$. This means that, again, the statistical guarantees are expected to depend polynomially in K. Given that the linear structure has enabled the development of efficient algorithms for linear and dueling bandits with large action spaces and contextual information, the question of how to impose linear-like structure to arbitrary preference matrix $\boldsymbol{P}$ has been a significant and longstanding open question.

There have been two notable advancements in this direction, one theoretical and one practical. The first advancement is by Wu et al. (2024), whose setting we briefly describe here. The learner has access to a feature map $(i, j) \in [K] \times [K] \mapsto \boldsymbol{\phi}_{i,j} \in \mathbb{R}^d$ satisfying $\boldsymbol{\phi}_{i,j} = -\boldsymbol{\phi}_{j,i}$. The preference probability is defined as $p_{i,j} = \mu(\langle \boldsymbol{\phi}_{i,j}, \boldsymbol{\theta}_\star \rangle)$, where $\boldsymbol{\theta}_\star \in \mathbb{R}^d$ is unknown. With this model, the authors have improved the Borda regret's dependency on K from polynomial to logarithmic. However, it is unrealistic to know all item *pair-wise* features that linearly encode the underlying preferences. Arguably, a more realistic scenario is knowing only item-wise features, namely, $\boldsymbol{\phi}_k \in \mathbb{R}^d$ for $k \in [K]$.

One may wonder if there is a contextual preference model that incorporates *item-wise* features while being potentially nontransitive. The second advancement, due to Zhang et al. (2025), tackles this by proposing the contextual bilinear preference model: for each item pair $(i, j) \in [K] \times [K]$, the preference model is defined as

$$p_{i,j} = \mu\left(\boldsymbol{\phi}_i^\top \boldsymbol{\Theta}_\star \boldsymbol{\phi}_j\right),$$

where $\boldsymbol{\Theta}_\star$ is a $d \times d$ skew-symmetric matrix of low rank. However, their paper does not provide any statistical guarantees when this is used in dueling bandits, or even regarding the estimation error of the preference model; rather, their main focus is experimentally validating this model in modeling human preferences and its implications for the downstream RLHF task. Note that we adopt the same preference model, exept we allow for the underlying arm-set $\mathcal{A}$ to be continuous.

Although not discussed further in Zhang et al. (2025), we believe this is a very natural way of incorporating some sort of linearity into general preferences, and that it deserves more attention from the dueling bandits community as well. Indeed, such bilinear model has been used in modeling interaction of two items, with applications to drug discovery (Luo et al., 2017), server scheduling (Kim and Vojnović, 2021), personalized recommendation (Chu and Park, 2009), link prediction (Menon and Elkan, 2011), relational learning (Nickel et al., 2011), and more. The bandit community was introduced to this model by bilinear bandits (Jang et al., 2021; Jun et al., 2019),

later extended to low-rank matrix-armed bandits (Jang et al., 2024; Kang et al., 2022; Lu et al., 2021); refer to Appendix A for further related works on low-rank bandits. Roughly speaking, the learner now only needs to learn $\Theta(d^2)$ parameters of $\boldsymbol{\Theta}_\star$ instead of $\Theta(K^2)$ parameters of $\boldsymbol{P}$. Furthermore, using the low-rank structure of $\boldsymbol{\Theta}_\star$, the learner can further improve the regret's dependency in d. Although not discussed in Zhang et al. (2025), we also note that this is the rank-d version of the low-rank preference model of Rajkumar and Agarwal (2016), as one can write $\mu^{-1}(\boldsymbol{P}) = \boldsymbol{\Phi}^\top \boldsymbol{\Theta}_\star \boldsymbol{\Phi}$ where $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1 \cdots \boldsymbol{\phi}_K] \in \mathbb{R}^{d\times K}$ and μ^{-1} is applied entry-wise.

Variance-Aware Borda Regret Bound. The Borda regret resembles the strong regret (Yue et al., 2012), and it "respects" the inherent problem of the difficulty of dueling bandits where two arms are chosen rather than a single arm (Saha et al., 2021; Wu et al., 2024). Its original motivation is from search engine, in which the regret corresponds to "the fraction of users who would prefer the best retrieval function over the selected ones." (Yue and Joachims, 2009).

All the existing guarantees for the Borda regret either assume a fixed gap (Saha et al., 2021) or incur a $1/c_\mu$ dependency (Wu et al., 2024), where c_μ can be thought of as the worst-case badness of linear approximation of the true preference signal. In other words, the current Borda regret bound seems to suggest that the lower the variance (which roughly corresponds to the derivative of the inverse link function in the context of GLMs), the higher the regret. However, the vast literature on logistic and generalized linear bandits (Abeille et al., 2021; Lee et al., 2024a,b) suggest otherwise. Abeille et al. (2021) first proved a $\widetilde{\mathcal{O}}(d\sqrt{T\kappa_\star})$ regret bound for logistic bandits as well as a matching (local minimax) lower bound, the correct dependency on the variance-dependent quantity. Thus, it should be expected that a similar variance-dependent quantity should pop up in the optimal Borda regret bounds.

K.2 A Sufficient Condition for the Bilinear Preference to be Stochastic Transitive

A preference model is **stochastic transitive w.r.t.** μ (Bengs et al., 2022) if there exists a $f : [K] \to \mathbb{R}$ such that $(\boldsymbol{P})_{ij} = \mu(f(i) - f(j))$. Here, we prove that certain collinearity between the features $\boldsymbol{\phi}_i$'s in the bilinear preference model (Eqn. (K.1)) implies stochastic transitivity:

Proposition K.1. *If there exists an orthonormal $\boldsymbol{Q} \in \mathbb{R}^{d\times d}$ such that $\{((\boldsymbol{Q}^\top \boldsymbol{\phi}_k)_{2m-1}, (\boldsymbol{Q}^\top \boldsymbol{\phi}_k)_{2m})\}_{k\in[K]}$ is collinear in $\mathbb{R}^2$ for each $m \in [r]$, then the bilinear preference model is stochastic transitive w.r.t. μ. When $r = 1$ (i.e., $\operatorname{rank}(\boldsymbol{\Theta}_\star) = 2$), this is also a necessary condition.*

Proof. The proof is heavily inspired by Jiang et al. (2011), where the authors provide a decomposition of the space of preferences via combinatorial Hodge theory; this has been also utilized in later machine learning literature on ranking with potentially nontransitive components (Balduzzi et al., 2018, 2019; Bertrand et al., 2023).

From the combinatorial Hodge decomposition (Jiang et al., 2011, Theorem 2), a f that satisfies the stochastic transitivity exists if and only if for any $(i, j, k) \in [K]^3$,

$$\boldsymbol{\phi}_i^\top \boldsymbol{\Theta}_\star \boldsymbol{\phi}_j + \boldsymbol{\phi}_j^\top \boldsymbol{\Theta}_\star \boldsymbol{\phi}_k + \boldsymbol{\phi}_k^\top \boldsymbol{\Theta}_\star \boldsymbol{\phi}_i = 0.$$

The quantity on the LHS is known as the *combinatorial curl* (Jiang et al., 2011).

Let $\boldsymbol{\Theta}_\star = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^\top$ be its canonical form (Lemma K.1), and let $\boldsymbol{\varphi}_i := \boldsymbol{Q}^\top \boldsymbol{\phi}_i$. Let $\{\lambda_m\}_{m\in[r]} \subset \mathbb{R}_{>0}$ be the nonzero components of $\boldsymbol{\Lambda}$. Then, the above curl-free requirement boils down to

$$\sum_{m=1}^{r} \lambda_m \underbrace{\begin{vmatrix} 1 & 1 & 1 \\ (\boldsymbol{\varphi}_i)_{2m-1} & (\boldsymbol{\varphi}_j)_{2m-1} & (\boldsymbol{\varphi}_k)_{2m-1} \\ (\boldsymbol{\varphi}_i)_{2m} & (\boldsymbol{\varphi}_j)_{2m} & (\boldsymbol{\varphi}_k)_{2m} \end{vmatrix}}_{\triangleq V_m} = 0.$$

One sufficient condition for above to hold (necessary as well if $r = 1$) is if $V_m = 0$ for all $m \in [r]$. Geometrically, V_m is the signed volume of the parallelopipe in $\mathbb{R}^3$, spanned by the three column vectors. For the volume to be zero, it must be that $\{((\boldsymbol{\varphi}_i)_{2m-1}, (\boldsymbol{\varphi}_i)_{2m}), ((\boldsymbol{\varphi}_j)_{2m-1}, (\boldsymbol{\varphi}_j)_{2m}), ((\boldsymbol{\varphi}_k)_{2m-1}, (\boldsymbol{\varphi}_k)_{2m})\}$ is collinear in $\mathbb{R}^2$. As this must hold for any $i, j, k \in [K]^3$, it must be that $\{((\boldsymbol{\varphi}_k)_{2m-1}, (\boldsymbol{\varphi}_k)_{2m})\}_{k\in[K]}$ is collinear as well, for each $m \in [r]$. □

Remark 10. *We believe that the above result is extendable to the general case via decomposing the general preference into its transitive and cyclic components Jiang et al. (2011). But then, geometrically, it is unclear how to choose the right features such that the non-transitive and transitive components are compatible with each other, which corresponds to the "harmonic" component from the combinatorial Hodge decomposition (Jiang et al., 2011).*

K.3 Miscellaneous Mathematical Preliminaries

Here, for completeness and to foster future directions, we provide a bit orthogonal, yet interesting (and hopefully useful) mathematical preliminaries regarding skew-symmetric matrices and anti-symmetric tensor product space.

K.3.1 Skew-Symmetric Matrix

A matrix $\boldsymbol{A} \in \mathbb{R}^{d\times d}$ is **skew-symmetric** (or anti-symmetric) if $\boldsymbol{A}^\top = -\boldsymbol{A}$. It is known that the rank of a skew-symmetric matrix must be even (Hoffman and Kunze, 1971, Section 10.3), and it admits the following decomposition, which is its canonical form:

Lemma K.1 (Corollary 2.5.11 of Horn and Johnson (2012)[a]). *$\boldsymbol{A}$ is a skew-symmetric of rank $2r \le d$ if and only if there exists a (unique) orthogonal $\boldsymbol{Q}$ (i.e., $\boldsymbol{Q}^\top\boldsymbol{Q} = \boldsymbol{Q}\boldsymbol{Q}^\top = \boldsymbol{I}_d$) and $\{\lambda_\ell\}_{\ell\in[r]} \subset \mathbb{R}_{>0}$ such that $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^\top$, where*

$$\boldsymbol{\Lambda} = \left(\bigoplus_{\ell\in[r]} \lambda_\ell \boldsymbol{S}\right) \oplus \boldsymbol{0}_{d-2r},$$

where $\oplus$ is the matrix direct sum and $\boldsymbol{S} := \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. Moreover, $\{\pm\lambda_\ell i\}_{\ell\in[r]}$ are the eigenvalues of $\boldsymbol{A}$.

[a] A fun(?) historical note: this decomposition has been repeatedly rediscovered and renamed: Murnaghan-Wintner decomposition (Murnaghan and Wintner, 1931), Youla decomposition (Youla, 1961), or the Schur decomposition (Balduzzi et al., 2018), although the latter name is a bit inaccurate as the "usual" Schur decomposition should result in an upper triangular matrix in the middle (Horn and Johnson, 2012).

We also remark that the above form can be quite efficiently computed (Penke et al., 2020; Ward and Gray, 1978).

Let $\mathrm{Skew}(d) := \{\boldsymbol{\Theta} \in \mathbb{R}^{d\times d} : \boldsymbol{\Theta}^\top = -\boldsymbol{\Theta}\}$. It is a well-known that $\mathrm{Skew}(d)$ is a linear subspace of $\mathbb{R}^{d\times d}$, and that the mapping $\boldsymbol{A} \mapsto \frac{1}{2}(\boldsymbol{A} - \boldsymbol{A}^\top)$ is an orthogonal projection onto $\mathrm{Skew}(d)$ (Hoffman and Kunze, 1971, Chapter 6.6). We will also consider rank-constrained $\mathrm{Skew}(d)$, defined as $\mathrm{Skew}(d; 2r) := \{\boldsymbol{\Theta} \in \mathbb{R}^{d\times d} : \boldsymbol{\Theta}^\top = -\boldsymbol{\Theta},\ \mathrm{rank}(\boldsymbol{\Theta}) = 2r\}$. This is a matrix manifold whose dimension is given as follows (see Appendix K.4 for the proof):

Proposition K.2. $\dim(\mathrm{Skew}(d; 2r)) = 2dr - (2r^2 + r)$.

K.3.2 2nd-Order Tensor Product Space

Here, we largely follow the exposition of Section 2 of Garcia et al. (2023) and Section I.5 of Bhatia (1997), to which we refer interested readers for an overview of general tensor algebra over Hilbert space.

We define the **2nd-order tensor power** of $\mathbb{R}^d$ as $(\mathbb{R}^d)^{\otimes 2} := \{\boldsymbol{x} \otimes \boldsymbol{y} : \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d\}$, where the inner product[10] is such that $\langle \boldsymbol{x}_1 \otimes \boldsymbol{x}_2, \boldsymbol{y}_1 \otimes \boldsymbol{y}_2\rangle = \langle \boldsymbol{x}_1, \boldsymbol{y}_1\rangle\langle \boldsymbol{x}_2, \boldsymbol{y}_2\rangle$. Then, its orthonormal basis is given as $\{\boldsymbol{e}_i \otimes \boldsymbol{e}_j\}_{(i,j)\in[d]^2}$.

Consider the symmetrization and antisymmetrization operators, defined as $\mathcal{P}_S(\boldsymbol{x} \otimes \boldsymbol{y}) := \boldsymbol{x} \odot \boldsymbol{y} := \frac{1}{2}(\boldsymbol{x} \otimes \boldsymbol{y} + \boldsymbol{y} \otimes \boldsymbol{x})$ and $\mathcal{P}_A(\boldsymbol{x} \otimes \boldsymbol{y}) := \boldsymbol{x} \wedge \boldsymbol{y} := \frac{1}{2}(\boldsymbol{x} \otimes \boldsymbol{y} - \boldsymbol{y} \otimes \boldsymbol{x})$. Then, one can orthogonally decompose $(\mathbb{R}^d)^{\otimes 2} = (\mathbb{R}^d)^{\odot 2} \oplus (\mathbb{R}^d)^{\wedge 2}$, where the two spaces are spanned by their respective *orthonormal* basis: $(\mathbb{R}^d)^{\odot 2} = \mathrm{span}\left(\{\boldsymbol{e}_i \odot \boldsymbol{e}_i\}_{i\in[d]} \cup \{\sqrt{2}(\boldsymbol{e}_i \odot \boldsymbol{e}_j)\}_{1\le i<j\le d}\right)$, and $(\mathbb{R}^d)^{\wedge 2} = \mathrm{span}\left(\{\sqrt{2}(\boldsymbol{e}_i \wedge \boldsymbol{e}_j)\}_{1\le i<j\le d}\right)$.

Let us focus on the antisymmetric part. It is known that $\mathcal{P}_A$ is an orthogonal projection onto $\mathbb{R}^{\wedge 2}$ with the

[10] Such inner product is unique (Bhatia, 1997, Proposition 3.8.2).

following idempotent, full row-rank matrix representation of $\mathcal{P}_A$:

$$\boldsymbol{P}_A := \sqrt{2}\begin{bmatrix}\boldsymbol{e}_1 \wedge \boldsymbol{e}_2 & \boldsymbol{e}_1 \wedge \boldsymbol{e}_3 & \cdots & \boldsymbol{e}_{d-1} \wedge \boldsymbol{e}_d\end{bmatrix} \in \mathbb{R}^{d^2 \times \binom{d}{2}}.$$

It satisfies $\boldsymbol{P}_A^\top \boldsymbol{P}_A = \boldsymbol{I}_{\binom{d}{2}}$ and $\boldsymbol{P}_A \boldsymbol{P}_A^\top (\boldsymbol{x} \otimes \boldsymbol{y}) = \boldsymbol{x} \wedge \boldsymbol{y}$.

K.4 Proof of Proposition K.2

The proof utilizes some tools from topology, Lie group theory and matrix theory. Our main references are Munkres (2018), Chapter 21 of Lee (2012) and Horn and Johnson (2012).

Consider the generalized linear group $\mathrm{GL}_d(\mathbb{R}) := \{\boldsymbol{X} \in \mathbb{R}^{d\times d} : \det(\boldsymbol{X}) \neq 0\}$, which is a Lie group of dimension d^2. We then define the group action of $\mathrm{GL}_d(\mathbb{R})$ on $\mathrm{Skew}(d; 2r)$ as the following:

$$(\boldsymbol{X}, \boldsymbol{A}) \mapsto \boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^\top, \quad \boldsymbol{X} \in \mathrm{GL}_d(\mathbb{R}), \boldsymbol{A} \in \mathrm{Skew}(d; 2r).$$

We now utilize the following lemma:

Lemma K.2 (Theorem 21.20 of Lee (2012))**.** *Let X be a set and G be a Lie group that acts on X transitively, i.e., for any $x, y \in X$ there exists a $g \in G$ such that $(g, x) = y$. Suppose that there exists a point $p \in X$ such that the stabilizer group G_p is closed in G. Then, X has a unique smooth manifold structure w.r.t. which the given action is smooth. With this structure,* $\dim X = \dim G - \dim G_p$.

We first show that our group action indeed satisfies the assumptions of the above lemma. For simplicity, let us denote

$$\boldsymbol{S}_{d,2r} := \underbrace{\bigoplus_{\ell\in[r]} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}}_{=:\boldsymbol{S}_{2r}} \oplus \boldsymbol{0}_{d-2r}.$$

Claim K.1. *The action as defined in Eqn. (K.4) is transitive.*

Proof. To see this, consider two $\boldsymbol{A}, \boldsymbol{B} \in \mathrm{Skew}(d; 2r)$. Then by Lemma K.1, there exists $\boldsymbol{U_A}, \boldsymbol{U_B} \in O(d)$ and $\{\lambda_{\ell,\boldsymbol{A}}^2, \lambda_{\ell,\boldsymbol{B}}^2\}_{\ell\in[r]}$ such that $\boldsymbol{A} = \boldsymbol{U_A}\boldsymbol{\Lambda_A}\boldsymbol{S}_{d,2r}\boldsymbol{\Lambda_A}^\top\boldsymbol{U_A}^\top$ and $\boldsymbol{B} = \boldsymbol{U_B}\boldsymbol{\Lambda_B}\boldsymbol{S}_{d,2r}\boldsymbol{\Lambda_B}^\top\boldsymbol{U_B}^\top$, where

$$\boldsymbol{\Lambda_A} = \mathrm{diag}(\underbrace{\lambda_{1,\boldsymbol{A}}, \lambda_{1,\boldsymbol{A}}}_{\text{twice}}, \cdots, \underbrace{\lambda_{r,\boldsymbol{A}}, \lambda_{r,\boldsymbol{A}}}_{\text{twice}}, \underbrace{0, 0, \ldots, 0}_{\text{remaining entries}})$$

and similarly for $\boldsymbol{\Lambda_B}$. Then, defining $\boldsymbol{X} = (\boldsymbol{U}_B\boldsymbol{\Lambda}_B)(\boldsymbol{U}_A\boldsymbol{\Lambda_A})^{-1} \in \mathrm{GL}_d(\mathbb{R})$, it can be seen that $(\boldsymbol{X}, \boldsymbol{A}) = \boldsymbol{B}$. □

For the point p in the above lemma, we choose $\boldsymbol{S}_{d,2r} \in \mathrm{Skew}(d; 2r)$. Let us denote its stabilizer group as $S_{d,2r} := \{\boldsymbol{X} \in \mathrm{GL}_{d-2r}(\mathbb{R}) : \boldsymbol{X}\boldsymbol{S}_{d,2r}\boldsymbol{X}^\top = \boldsymbol{S}_{d,2r}\}$.

Claim K.2. *$S_{d,2r}$ is closed in $\mathrm{GL}_d(\mathbb{R})$.*

Proof. Consider a mapping $\rho : \boldsymbol{X} \mapsto \boldsymbol{X}\boldsymbol{S}_{d,2r}\boldsymbol{X}^\top$, which is continuous. Noting that $S_{d,2r} = \rho^{-1}(\{\boldsymbol{S}_{d,2r}\})$ and that $\{\boldsymbol{S}_{d,2r}\}$ is closed (in Hausdorff space, which $\mathrm{GL}_d(\mathbb{R})$ is), $S_{d,2r}$ is also closed by continuity. □

We now characterize $S_{d,2r}$.

Using block matrix notation, we need to characterize $\boldsymbol{X} = \begin{bmatrix}\boldsymbol{X}_{11} & \boldsymbol{X}_{12} \\ \boldsymbol{X}_{21} & \boldsymbol{X}_{22}\end{bmatrix}$ such that $\boldsymbol{X}$ is invertible and $\boldsymbol{X}\boldsymbol{S}_{2r}\boldsymbol{X}^\top = \boldsymbol{S}_{2r}$. After some tedious computations, we have that

$$\begin{bmatrix}\boldsymbol{X}_{11}\boldsymbol{S}_{2r}\boldsymbol{X}_{11}^\top & \boldsymbol{X}_{11}\boldsymbol{S}_{2r}\boldsymbol{X}_{21}^\top \\ \boldsymbol{X}_{21}\boldsymbol{S}_{2r}\boldsymbol{X}_{11}^\top & \boldsymbol{X}_{21}\boldsymbol{S}_{2r}\boldsymbol{X}_{21}^\top\end{bmatrix} = \begin{bmatrix}\boldsymbol{S}_{2r} & \boldsymbol{0}_{2r\times(d-2r)} \\ \boldsymbol{0}_{(d-2r)\times 2r} & \boldsymbol{0}_{2r\times 2r}\end{bmatrix}.$$

Consider the first block. Taking the determinant, we can deduce that $\det(\boldsymbol{X}_{11})^2 = 1 \neq 0$, i.e., $\boldsymbol{X}_{11}$ should be invertible. As $\boldsymbol{S}_{2r}$ is also invertible, the antidiagonal blocks implies that $\boldsymbol{X}_{21} = \boldsymbol{0}_{(d-2r)\times 2r}$.

So far, we have that $\boldsymbol{X}$ should be of the form

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_{11} & \boldsymbol{X}_{12} \\ \boldsymbol{0}_{(d-2r)\times 2r} & \boldsymbol{X}_{22}, \end{bmatrix}$$

where $\boldsymbol{X}_{11} \in \mathrm{Sym}(2p) := \{\boldsymbol{X} \in \mathrm{GL}_n(\mathbb{R}) : \boldsymbol{X}\boldsymbol{S}_{2r}\boldsymbol{X}^\top = \boldsymbol{X}\}$. By Schur's determinant formula, as $\boldsymbol{X}$ must be invertible, we must have that

$$\det(\boldsymbol{X}) = \det(\boldsymbol{X}_{11})\det(\boldsymbol{X}_{22}) \neq 0,$$

i.e., $\boldsymbol{X}_{22}$ should also be invertible.

We now derive the dimension of $\mathrm{GL}_{d-2r}(\mathbb{R})$ $\mathrm{Sym}(2r)$.

Claim K.3. $\dim(\mathrm{GL}_{d-2r}(\mathbb{R})) = (d-2r)^2$.

Proof. Let $n = d - 2r$. Then, note that $\mathrm{GL}_n(\mathbb{R}) = \det^{-1}(\mathbb{R} \setminus \{0\})$. As det is continuous and $\mathbb{R} \setminus \{0\}$ is open, $\mathrm{GL}_n(\mathbb{R}) \subset \mathbb{R}^{n\times n}$ is open, and we are done. □

Claim K.4. $\dim(\mathrm{Sym}(2r)) = 2r^2 + r$.

Proof. We do this by counting the number of independent constraints, then subtracting it from $\dim(\mathrm{GL}_{2r}(\mathbb{R})) = 4r^2$. Let us denote $\boldsymbol{S} := \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ for simplicity. First, for a $\boldsymbol{A} \in \mathbb{R}^{2\times 2}$, note that

$$\boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^\top = \det(\boldsymbol{A})\boldsymbol{S}.$$

Now consider a $\boldsymbol{X} \in \mathrm{GL}_{2r}(\mathbb{R})$, consisting of r number of 2×2 blocks:

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_{11} & \boldsymbol{X}_{12} & \cdots & \boldsymbol{X}_{1r} \\ \boldsymbol{X}_{21} & \boldsymbol{X}_{22} & \cdots & \boldsymbol{X}_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{X}_{r1} & \boldsymbol{X}_{r2} & \cdots & \boldsymbol{X}_{rr} \end{bmatrix}.$$

Then, by the block matrix multiplication and the above result, we have that

$$\left(\boldsymbol{X}\boldsymbol{S}_{2r}\boldsymbol{X}^\top\right)_{i,j} = \begin{cases} \left(\sum_{k=1}^r \det(\boldsymbol{X}_{ik})\right)\boldsymbol{S}, & i = j, \\ \sum_{k=1}^r \boldsymbol{X}_{ik}\boldsymbol{J}\boldsymbol{X}_{kj}^\top, & i \neq j \end{cases} = \begin{cases} \boldsymbol{S}, & i = j, \\ \boldsymbol{0}_{2\times 2}, & i \neq j \end{cases}.$$

where here, $(\cdot)_{i,j}$ refers to the 2×2 block at the (i,j) location.

There are r constraints for $i = j$ and $4\binom{r}{2} = 2r(r-1)$ constraints for $i \neq j$, which amounts to $2r^2 - r$ constraints in total. Thus, the dimension of $\mathrm{Sym}(2r)$ becomes $4r^2 - (2r^2 - r) = 2r^2 + r$. □

All in all, we have that

$$\begin{aligned} \dim(S_{d,2r}) &= \underbrace{\dim(\mathrm{Sym}(2r))}_{\text{degrees of freedom for } \boldsymbol{X}_{11}} + \underbrace{\dim(\mathbb{R}^{2r\times(d-2r)})}_{\text{degrees of freedom for } \boldsymbol{X}_{12}} + \underbrace{\dim(\mathrm{GL}_{d-2r}(\mathbb{R}))}_{\text{degrees of freedom for } \boldsymbol{X}_{22}} \\ &= (2r^2 + r) + 2r(d-2r) + (d-2r)^2 \\ &= d^2 + 2r^2 + r - 2dr. \end{aligned}$$

Applying Lemma K.2, we have that

$$\dim(\mathrm{Skew}(d; 2r)) = \dim(\mathrm{GL}_d(\mathbb{R})) - \dim(S_{d,2r}) = 2dr - (2r^2 + r).$$

□

L ADDITIONAL DETAILS FOR BILINEAR DUELING BANDITS: REGRET ANALYSIS

L.1 Proof of Theorem 5.1 – Borda Regret Upper Bound for Bilinear Dueling Bandits

We state the full version of the Borda regret bound and give its proof:

Theorem L.1 (Full Statement of Theorem 5.1)**.** *Let us denote* $\mathrm{GL}_{\min} := \mathrm{GL}_{\min}(\mathcal{X})$, $\bar{\kappa}(\pi) := \mathbb{E}_{\boldsymbol{X}\sim\pi}\left[\dot{\mu}(\langle \boldsymbol{X}, \boldsymbol{\Theta}_\star\rangle)\right]$, *and* $\pi_\star \in \arg\min_{\pi\in\mathcal{P}} \mathrm{GL}(\pi; \boldsymbol{\Theta}_\star)$. *Choose* N_1 *and* N_2 *as*

$$N_1 \asymp \frac{\bar{\kappa}(\pi_\star) dr}{\lambda_{\min}(\boldsymbol{H}(\pi_\star;\boldsymbol{\Theta}_\star))^2} \log\frac{d}{\delta} \vee R_s \frac{\bar{\kappa}(\pi_\star) r^2}{\lambda_{\min}(\boldsymbol{H}(\pi_\star;\boldsymbol{\Theta}_\star))^3} \sqrt{\frac{N_2 \log\frac{d}{\delta}}{\mathrm{GL}_{\min}(\mathcal{X})}},$$
$$N_2 = \left(\mathrm{GL}_{\min} \log\frac{d}{\delta}\right)^{1/3} (\kappa_\star^B T)^{2/3},$$

and let us assume that $T \geq N_1 + N_2$. *Then, the following Borda regret bound of* `BETC-GLM-LR`[a] *holds with probability at least* $1-\delta$*:*

$$\mathrm{Reg}^B(T) \lesssim \left(\mathrm{GL}_{\min} \log\frac{d}{\delta}\right)^{1/3} (\kappa_\star^B T)^{2/3} + R_s L_\mu \left(\frac{\mathrm{GL}_{\min}}{\kappa_\star^B} \log\frac{d}{\delta}\right)^{2/3} T^{1/3} + N_1.$$

Here, it is clear that the first term dominates when T *is sufficiently large.*

[a] This is an acronym for *Borda Explore-Then-Commit for Generalized Linear Models with Low-Rank structure.*

Proof. We naïvely bound the instantaneous regret from the exploration phase with 1, and thus, the cumulative regret up to the forced exploration is $N_1 + N_2$.

After the exploration phase, the instantaneous regret is the same as $B(\boldsymbol{\phi}_\star) - B(\widehat{\boldsymbol{\phi}})$. This is bounded as follows:

$$\begin{aligned}
B(\boldsymbol{\phi}_\star) - B(\widehat{\boldsymbol{\phi}}) &= \mathbb{E}_{\boldsymbol{\phi}'\sim\mathrm{Unif}(\mathcal{X})}\left[\mu\left(\boldsymbol{\phi}_\star^\top \boldsymbol{\Theta}_\star \boldsymbol{\phi}'\right) - \mu(\widehat{\boldsymbol{\phi}}^\top \boldsymbol{\Theta}_\star \boldsymbol{\phi}')\right] \\
&\leq \mathbb{E}_{\boldsymbol{\phi}'\sim\mathrm{Unif}(\mathcal{X})}\left[\mu\left(\boldsymbol{\phi}_\star^\top \boldsymbol{\Theta}_\star \boldsymbol{\phi}'\right) - \mu(\boldsymbol{\phi}_\star^\top \widehat{\boldsymbol{\Theta}} \boldsymbol{\phi}')\right] && \text{(Definition of } \widehat{\boldsymbol{\phi}}) \\
&\overset{(*)}{=} \underbrace{\mathbb{E}_{\boldsymbol{\phi}'\sim\mathrm{Unif}(\mathcal{X})}\left[\dot{\mu}\left(\boldsymbol{\phi}_\star^\top \boldsymbol{\Theta}_\star \boldsymbol{\phi}'\right) \boldsymbol{\phi}_\star^\top (\boldsymbol{\Theta}_\star - \widehat{\boldsymbol{\Theta}}) \boldsymbol{\phi}'\right]}_{\triangleq Q_1} + \underbrace{\mathbb{E}_{\boldsymbol{\phi}'\sim\mathrm{Unif}(\mathcal{X})}\left[-\left(\boldsymbol{\phi}_\star^\top (\boldsymbol{\Theta}_\star - \widehat{\boldsymbol{\Theta}}) \boldsymbol{\phi}'\right)^2 \tilde{\theta}(\boldsymbol{\phi}')\right]}_{\triangleq Q_2} \\
& && \text{(First-order Taylor expansion with integral remainder)}
\end{aligned}$$

where at $(*)$, we define

$$\tilde{\theta}(\boldsymbol{\phi}') := \int_0^1 (1-z)\ddot{\mu}\left(\boldsymbol{\phi}_\star^\top \left((1-z)\boldsymbol{\Theta}_\star + z\widehat{\boldsymbol{\Theta}}\right) \boldsymbol{\phi}'\right) dz.$$

Q_1 can be bounded as

$$\begin{aligned}
Q_1 &= \mathbb{E}_{\boldsymbol{\phi}'\sim\mathrm{Unif}(\mathcal{X})}\left[\dot{\mu}\left(\boldsymbol{\phi}_\star^\top \boldsymbol{\Theta}_\star \boldsymbol{\phi}'\right) \boldsymbol{\phi}_\star^\top (\boldsymbol{\Theta}_\star - \widehat{\boldsymbol{\Theta}}) \boldsymbol{\phi}'\right] \\
&\leq \left(\max_{\boldsymbol{\phi}'\in\mathcal{X}} \left|\boldsymbol{\phi}_\star^\top (\boldsymbol{\Theta}_\star - \widehat{\boldsymbol{\Theta}}) \boldsymbol{\phi}'\right|\right) \mathbb{E}_{\boldsymbol{\phi}'\sim\mathrm{Unif}(\mathcal{X})}\left[\dot{\mu}\left(\boldsymbol{\phi}_\star^\top \boldsymbol{\Theta}_\star \boldsymbol{\phi}'\right)\right] \\
&\leq \kappa_\star^B \left\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}_\star\right\|_{\mathrm{op}} && \text{(rectangular quotient relation for } \|\cdot\|_{\mathrm{op}} \text{ \& } \boldsymbol{\phi}_\star, \boldsymbol{\phi}' \in \mathcal{B}^d(1) \text{ \& definition of } \kappa_\star^B) \\
&\lesssim \kappa_\star^B \sqrt{\frac{\mathrm{GL}_{\min}}{N_2} \log\frac{d}{\delta}}. && \text{(Theorem 3.1)}
\end{aligned}$$

By self-concordance, we have that $|\tilde{\theta}(\boldsymbol{\phi}')| \leq \frac{1}{2} R_s L_\mu$ for any $\boldsymbol{\phi}' \in \mathcal{X}$, and thus, Q_2 can be bounded as

$$Q_2 \leq \frac{1}{2} R_s L_\mu \mathbb{E}_{\boldsymbol{\phi}'\sim\mathrm{Unif}(\mathcal{X})}\left[\left(\boldsymbol{\phi}_\star^\top (\boldsymbol{\Theta}_\star - \widehat{\boldsymbol{\Theta}}) \boldsymbol{\phi}'\right)^2\right] \lesssim \frac{R_s L_\mu \mathrm{GL}_{\min}}{N_2} \log\frac{d}{\delta}.$$

Combining everything, we have that

$$B(\boldsymbol{\phi}_\star) - B(\widehat{\boldsymbol{\phi}}) \lesssim \kappa_\star^B \sqrt{\frac{\mathrm{GL}_{\min}}{N_2} \log \frac{d}{\delta}} + \frac{R_s L_\mu \mathrm{GL}_{\min}}{N_2} \log \frac{d}{\delta}.$$

All in all, we have

$$\begin{aligned}\mathrm{Reg}^B(T) &\lesssim N_1 + N_2 + (T - N_1 - N_2)\left(\kappa_\star^B \sqrt{\frac{\mathrm{GL}_{\min}}{N_2} \log \frac{d}{\delta}} + \frac{R_s L_\mu \mathrm{GL}_{\min}}{N_2} \log \frac{d}{\delta}.\right) \\ &\le N_1 + N_2 + T\sqrt{\frac{\mathrm{GL}_{\min}}{N_2} \log \frac{d}{\delta}}\left(\kappa_\star^B + R_s L_\mu \sqrt{\frac{\mathrm{GL}_{\min}}{N_2} \log \frac{d}{\delta}}\right).\end{aligned}$$

Let us optimize for N_2 using the last expression.

If we choose $N_2 = \left(\mathrm{GL}_{\min} \log \frac{d}{\delta}\right)^{1/3} (\kappa_\star^B T)^{2/3}$, we have

$$\mathrm{Reg}^B(T) \lesssim N_1 + \left(\mathrm{GL}_{\min} \log \frac{d}{\delta}\right)^{1/3} (\kappa_\star^B T)^{2/3} + R_s L_\mu \left(\frac{\mathrm{GL}_{\min}}{\kappa_\star^B} \log \frac{d}{\delta}\right)^{2/3} T^{1/3}.$$

With this, we have the following requirement on N_1, as stated in Eqn. (3.5): recalling that $\bar{\kappa}(\pi) = \mathbb{E}_{\boldsymbol{X}\sim\pi}\left[\dot{\mu}(\langle \boldsymbol{X}, \boldsymbol{\Theta}_\star\rangle)\right]$ and $\pi_\star \in \arg\min_{\pi\in\mathcal{P}} \mathrm{GL}(\pi; \boldsymbol{\Theta}_\star)$,

$$N_1 \gtrsim \frac{\bar{\kappa}(\pi_\star) d r}{\lambda_{\min}(\boldsymbol{H}(\pi_\star; \boldsymbol{\Theta}_\star))^2} \log \frac{d}{\delta} \vee R_s \frac{\bar{\kappa}(\pi_\star) r^2}{\lambda_{\min}(\boldsymbol{H}(\pi_\star; \boldsymbol{\Theta}_\star))^3} \sqrt{\frac{N_2 \log \frac{d}{\delta}}{\mathrm{GL}_{\min}(\mathcal{X})}}$$

□

L.2 Relations to Wu et al. (2024)

Reduction to Wu et al. (2024). To our knowledge, Wu et al. (2024) is the only comparable competitor in our setting of Borda regret minimization. To do that, we first describe how to reduce our bilinear dueling bandits to their setting. Recall that Wu et al. (2024) require vector-valued features for each pair of items, $\boldsymbol{\phi}_{i,j} = -\boldsymbol{\phi}_{j,i}$. As $\boldsymbol{\Theta}_\star = \widetilde{\boldsymbol{\Theta}}_\star - \widetilde{\boldsymbol{\Theta}}_\star^\top$ for some $\widetilde{\boldsymbol{\Theta}}_\star \in \mathbb{R}^{d\times d}$, one can rewrite the bilinear preference as

$$\mu\left(\boldsymbol{\phi}_i^\top (\widetilde{\boldsymbol{\Theta}}_\star - \widetilde{\boldsymbol{\Theta}}_\star^\top)\boldsymbol{\phi}_j\right) = \mu\left(\left\langle \widetilde{\boldsymbol{\Theta}}_\star, \boldsymbol{\phi}_i\boldsymbol{\phi}_j - \boldsymbol{\phi}_j\boldsymbol{\phi}_i^\top \right\rangle\right).$$

One may be tempted to set $\boldsymbol{\phi}_{i,j} = \mathrm{vec}(\boldsymbol{\phi}_i\boldsymbol{\phi}_j^\top - \boldsymbol{\phi}_j\boldsymbol{\phi}_i^\top)$. However, recalling the discussions from Appendix K.3.2, one must set $\boldsymbol{\phi}_{i,j} = \boldsymbol{P}_A^\top \mathrm{vec}(\boldsymbol{\phi}_i\boldsymbol{\phi}_j^\top - \boldsymbol{\phi}_j\boldsymbol{\phi}_i^\top)$ for $\boldsymbol{\phi}_{i,j}$'s to be able to fully span $\mathbb{R}^{\wedge 2}$. Setting $\boldsymbol{\theta}_\star = \boldsymbol{P}_A^\top \mathrm{vec}(\widetilde{\boldsymbol{\Theta}}_\star) \in \mathbb{R}^{d^2}$ and the reduction is complete.

Comparing Regret Upper Bounds. A naïve application of the algorithm of Wu et al. (2024) using the above reduction attains a Borda regret bound of $\widetilde{\mathcal{O}}(c_\mu^{-1} d^{4/3} T^{2/3})$ up to some epsilon-net error (see their Remark 5.3), where

$$c_\mu := \min_{\|\boldsymbol{x}\|_2 \le 1, \|\boldsymbol{\theta}-\boldsymbol{\theta}_\star\| \le 1} \dot{\mu}(\langle \boldsymbol{x}, \boldsymbol{\theta}\rangle) > 0.$$

They have also assumed that $\lambda_{\min}(\boldsymbol{V}(\pi^U)) \ge \lambda_0$ for some constant $\lambda_0 > 0$, where $\pi^U \sim \mathrm{Unif}(\mathcal{A}\times\mathcal{A})$ (Wu et al., 2024, Assumption 3.1). We remark that in many cases, λ_0 is *not* constant and can be arbitrarily small dimension-wise. In particular, both Wu et al. (2024) and our work assumes $\|\phi_{i,j}\|_2 \le 1$, one can prove that $\lambda_0 \le \frac{1}{d^2}$ **for any** $\mathcal{A}$ under this assumption and it is *impossible* to make λ_0 as a constant, since

$$\begin{aligned}\mathrm{tr}\left(\boldsymbol{V}(\pi)\right) &= \mathrm{tr}\left(\sum_{i,j} \pi(\boldsymbol{\phi}_{i,j})\boldsymbol{\phi}_{i,j}\boldsymbol{\phi}_{i,j}^\top\right) \\ &= \sum_{i,j} \pi(\boldsymbol{\phi}_{i,j}) \mathrm{tr}\left(\boldsymbol{\phi}_{i,j}\boldsymbol{\phi}_{i,j}^\top\right) && \text{(Linearity of tr)}\end{aligned}$$

$$\leq \sum_{i,j} \pi(\boldsymbol{\phi}_{i,j}) = 1 \qquad \text{(For a vector } \boldsymbol{v}, \operatorname{tr}(\boldsymbol{v}\boldsymbol{v}^\top) = \|\boldsymbol{v}\|_2^2 \text{ and } \|\boldsymbol{\phi}_{i,j}\|_2 \leq 1)$$

and $\operatorname{tr}(\boldsymbol{V}(\pi)) = \sum_{i=1}^{d^2} \lambda_i(\boldsymbol{V}(\pi))$.

Still, for a fair comparison, let us first compare with our bound under the same assumption and various arm-sets: utilizing Proposition 3.2,

- When $\mathcal{A} = \mathcal{B}^d(1)$ ($\mathcal{X} = \mathcal{B}_{\text{op}}^{d\times d}(1)$), our regret bound becomes $\widetilde{\mathcal{O}}\left(\kappa_\star^{-1/3}(\kappa_\star^B dT)^{2/3}\right)$. This is a strict improvement by a factor of $d^{2/3}$ and curvature-dependent quantities.
- When $\mathcal{A} = \{\boldsymbol{e}_i : i \in [d]\}$ ($\mathcal{X} = \mathcal{M}$), our regret bound becomes $\widetilde{\mathcal{O}}\left(\operatorname{Harm}(\boldsymbol{\Theta}_\star)^{-1/3} d(\kappa_\star^B T)^{2/3}\right)$. This is a strict improvement by a factor of $d^{1/3}$ and curvature-dependent quantities.
- For general $\mathcal{A} \subseteq \mathcal{B}^d(1)$ ($\mathcal{X} \subseteq \mathcal{B}_{\text{op}}^{d\times d}(1)$), by Proposition 3.1, our regret bound can be upper bounded with $\tilde{\mathcal{O}}\left(\kappa_\star^{-1/3}(\kappa_\star^B T)^{2/3}(d\lambda_0)^{1/3}\right)$. This is a strict improvement by a factor of $(d\lambda_0)^{1/3}$, when $\lambda_0 \geq \frac{1}{d^3}$.[11]

Remark 11 (Dependency on r)**.** *A keen reader may notice that our regret bound is independent of the rank r of the matrix $\boldsymbol{\Theta}_\star$, which is also the case for bilinear bandits (Jang et al., 2021, Theorem 4.6), albeit for a different reason. This is because our* `GL-LowPopArt` *exploits the low-rankness of $\mathcal{A}$ (which induces a matrix-valued arm-set of operator norm at most 1) and the parameter space* $\operatorname{Skew}(d; 2r)$*, analogous to bilinear bandits (Jang et al., 2021; Jun et al., 2019) and low-rank bandits (Jang et al., 2024; Kang et al., 2022; Lu et al., 2021).*

We now elaborate on the curvature-dependent quantity $\frac{(\kappa_\star^B)^2}{\kappa_\star}$ in our Borda regret bound, an instance-specific scaling not previously reported in the dueling bandits literature. Let us first recall their definitions:

$$\kappa_\star := \min_{\boldsymbol{\phi},\boldsymbol{\phi}' \in \mathcal{A}} \dot{\mu}\left(\boldsymbol{\phi}^\top \boldsymbol{\Theta}_\star \boldsymbol{\phi}'\right), \quad \kappa_\star^B := \mathbb{E}_{\boldsymbol{\phi}' \sim \operatorname{Unif}(\mathcal{A})}[\dot{\mu}(\boldsymbol{\phi}_\star^\top \boldsymbol{\Theta}_\star \boldsymbol{\phi}')].$$

Here, $\kappa_\star$ represents the worst-case flatness (minimum derivative) across all possible pairs of arms, while $\kappa_\star^B$ represents the average flatness when playing the true Borda winner $\boldsymbol{\phi}_\star$ against a uniformly drawn arm. By definition, we have $\kappa_\star \leq \kappa_\star^B$. The dueling nature of the regret reveals an interesting dichotomy depending on the relationship between these two quantities:

- **Uniform Flatness ($\kappa_\star \asymp \kappa_\star^B$):** If the landscape's flatness is relatively uniform –meaning the hardness of estimating the preference for the worst-case pair is of the same order as estimating preferences involving the Borda winner – then the curvature-dependent factor simplifies to $\frac{(\kappa_\star^B)^2}{\kappa_\star} \asymp \kappa_\star$. Consequently, our permanent regret scales as $\widetilde{\mathcal{O}}(\kappa_\star^{1/3} T^{2/3})$. This implies that a *flatter* problem (smaller $\kappa_\star$, which corresponds to smaller variance in the feedback) strictly *reduces* the permanent regret, analogous to the findings in generalized linear bandits (Abeille et al., 2021; Lee et al., 2024a).
- **Adversarial Flatness ($\kappa_\star \ll \kappa_\star^B$):** Conversely, if there exists an adversarial pair of arms that is significantly flatter (and thus much harder to learn) than the pairs involving the Borda winner, the ratio $\frac{(\kappa_\star^B)^2}{\kappa_\star}$ blows up. In this regime, the difficulty of identifying the parameters in the worst-case flat region overwhelmingly dominates the variance-reduction we would otherwise enjoy around the Borda winner. As a result, the permanent regret no longer benefits from the general flatness of the problem.

Regret Lower Bound. Wu et al. (2024, Theorem 4.1) obtain a regret lower bound of $\Omega(d^{2/3}T^{2/3})$ for $\boldsymbol{\phi}_{i,j}, \boldsymbol{\theta}_\star \in \mathbb{R}^d$, and a similar lower bound for unstructured dueling bandits has been obtained by Saha et al. (2021, Theorem 16); $T^{2/3}$ stems from the fact that the exploration and exploitation cannot be mixed. This suggests that at least in terms of T, our `BETC-GLM-LR` is also optimal.

[11] Of course, if the geometry of $\mathcal{A}$ is ill-distributed, then λ_0 can be arbitrarily small, and we do not claim that ours is always good. Rather, when $\mathcal{A}$ is sufficiently well-distributed (which is arguably the usual case). our `GL-LowPopArt` shows the benefit of exploiting the geometry and curvature.

However, their lower bound cannot be directly applied to our setting, as our bilinear dueling bandits, in essence, constrain the matrix arm to be of rank-1. It is clear that their hard instance, based on the lower bound for stochastic linear bandits (Dani et al., 2008), cannot be instantiated as our setting. We leave obtaining a tight lower bound to future work, considering how even in stochastic bilinear bandits (non-dueling), the lower bound remains open (Jang et al., 2021; Jun et al., 2019; Kotłowski and Neu, 2019). It would also be interesting from the curvature perspective on whether the above dichotomy is fundamental, i.e., whether one can derive a regret lower bound that depends on $\frac{(\kappa_\star^B)^2}{\kappa_\star}$. A potential starting point may be from the regret lower bound of Jang et al. (2024, Theorem 6.1), although they do not consider the Borda regret nor nonlinear link function.

M PRELIMINARY EXPERIMENTS: 1-BIT MATRIX COMPLETION/RECOVERY

In this appendix, we present preliminary numerical results on 1-bit matrix completion and recovery (Davenport et al., 2014) to demonstrate the empirical effectiveness of `GL-LowPopArt`. For results in the Gaussian (i.e., linear) setting, we refer readers to the experiments in Jang et al. (2024). The source code to reproduce these experiments is publicly available on our GitHub repository.[12]

M.1 Experimental Setting

Dataset. We set the dimensions $d_1 = d_2 = 3$, the rank $r = 1$, and the tolerance level $\delta = 0.001$. We evaluate three different choices for the action set $\mathcal{X}$:

1. **Matrix Completion:** $\mathcal{X} = \{\boldsymbol{e}_i \boldsymbol{e}_j^\top : 1 \le i, j \le 3\}$, where $\boldsymbol{e}_i$ denotes the standard basis vector in $\mathbb{R}^{d_1}$.
2. **Matrix Recovery (Random):** $K = 50$ arms sampled uniformly at random from the unit sphere $\mathcal{S}^{d_1 d_2 - 1}(1)$.
3. **Matrix Recovery (Hard):** A hard instance proposed by Jang et al. (2024, Appendix C.3), defined as:

$$\mathcal{X}_{\text{hard}} = \left\{ \sqrt{\frac{1}{d}} \text{vec}^{-1}(\boldsymbol{e}_1) \right\} \cup \left\{ \text{vec}^{-1}\left(\sqrt{\frac{1}{d+1}} \boldsymbol{e}_1 + \sqrt{\frac{d}{d+1}} \boldsymbol{e}_i \right) \ : \ i = 2, \dots, d^2 \right\}.$$

For each experimental setting, we repeat the process 30 times across a wide range of sample sizes N. In each repetition, the underlying true parameter is generated as $\boldsymbol{\Theta}_\star = 2\boldsymbol{U}\boldsymbol{U}^\top$, where $\boldsymbol{U}$ is obtained via the QR decomposition of $\boldsymbol{U}' \sim \mathcal{N}(0,1)^{d \times r}$. For the random matrix recovery setting, the arm set $\mathcal{X}$ is also resampled independently for each repetition.

Algorithms. We compare the following algorithms: (i) nuclear-penalized MLE with the uniform design ('U'), (ii) `GL-LowPopArt` with a uniform Stage II design ('U+U'), and (iii) `GL-LowPopArt` with an optimal GL-design in Stage II ('U+GL'). We employ the theoretically prescribed hyperparameters without further tuning: $\lambda_N = \sqrt{\frac{2}{N} \log \frac{12}{\delta}}$ for the nuclear-penalized MLE (Proposition F.1), and $\nu = \sqrt{\frac{2}{5\text{GL}(\pi; \boldsymbol{\Theta}_0) N_2} \log \frac{24}{\delta}}$ for `GL-LowPopArt`.

To solve the Stage I nuclear-penalized MLE, we use a restarted FISTA-style proximal gradient method with a backtracking line search (Beck and Teboulle, 2009; Cai et al., 2010; O'Donoghue and Candès, 2015). For Stage II and the optimal design computations, we use CVXPY (Agrawal et al., 2018; Diamond and Boyd, 2016) equipped with the MOSEK solver.

To ensure a fair comparison, we fix the total sample budget N across all methods, enforcing $N_1 + N_2 = N$, where N_i denotes the number of samples allocated to Stage i. Specifically, we set $N_1 = \lfloor N/10 \rfloor$ and $N_2 = N - N_1$.[13]

Remark 12. *We also experimented with the Burer-Monteiro factorization (BMF) approach using a small random initialization (Kim and Chung, 2023; Stöger and Soltanolkotabi, 2021), factoring the parameter as $\boldsymbol{\Theta} = \boldsymbol{U}\boldsymbol{U}^\top$. However, this method yielded the worst performance across all scenarios. This suggests that the non-convex loss landscape is highly non-benign in the noisy setting, corroborating the observations of Ma and Fattahi (2025). We have omitted these results for brevity.*

M.2 Results & Discussion

We report the mean operator norm error for each algorithm alongside 95% two-sided confidence intervals, computed using the empirical standard error over the 30 independent repetitions (using the Student's t-distribution or normal approximation) (Student, 1908).

[12] https://github.com/nick-jhlee/GL-LowPopArt

[13] While the main text establishes that $N_1 \asymp \sqrt{N}$ suffices asymptotically, we allocate a larger proportional budget to N_1 here to account for finite-sample effects and ensure sufficient initial exploration.

Figure 1 summarizes the updated results.[14] Across all three tasks, `GL-LowPopArt` substantially outperforms the nuclear-penalized MLE baseline. In particular, the baseline error decreases much more slowly and exhibits a long finite-sample plateau (e.g., up to $N \approx 10^5$ for **Matrix Recovery (Hard)** and $N \approx 10^4$ for other tasks), whereas both Stage-II variants of `GL-LowPopArt` improve much more rapidly as N grows. We now elaborate on the difference in performance between uniform and optimal GL-design for Stage II.

For **matrix completion**, the gain from using the optimal GL-design in Stage II is negligible: the uniform and GL-design perform very similarly, and the uniform Stage-II design is slightly better in the small-sample regime. For **matrix recovery (random)**, the gain is also somewhat modest, but in the same small-sample regime, the GL-design is slightly better than uniform design. All in all, the two Stage-II variants can be extremely close, with only a small edge for the GL-design in parts of the range. The clearest benefit of the optimal GL-design appears in **matrix recovery (hard)**, where it consistently outperforms the uniform design by the largest margin in the small sample regime, but the gap narrows down as N increases. Thus, the empirical benefit of GL-design is most pronounced on geometrically anisotropic instances, but still modest compared to our initial expectations. We leave a more systematic empirical study of this phenomenon to future work.

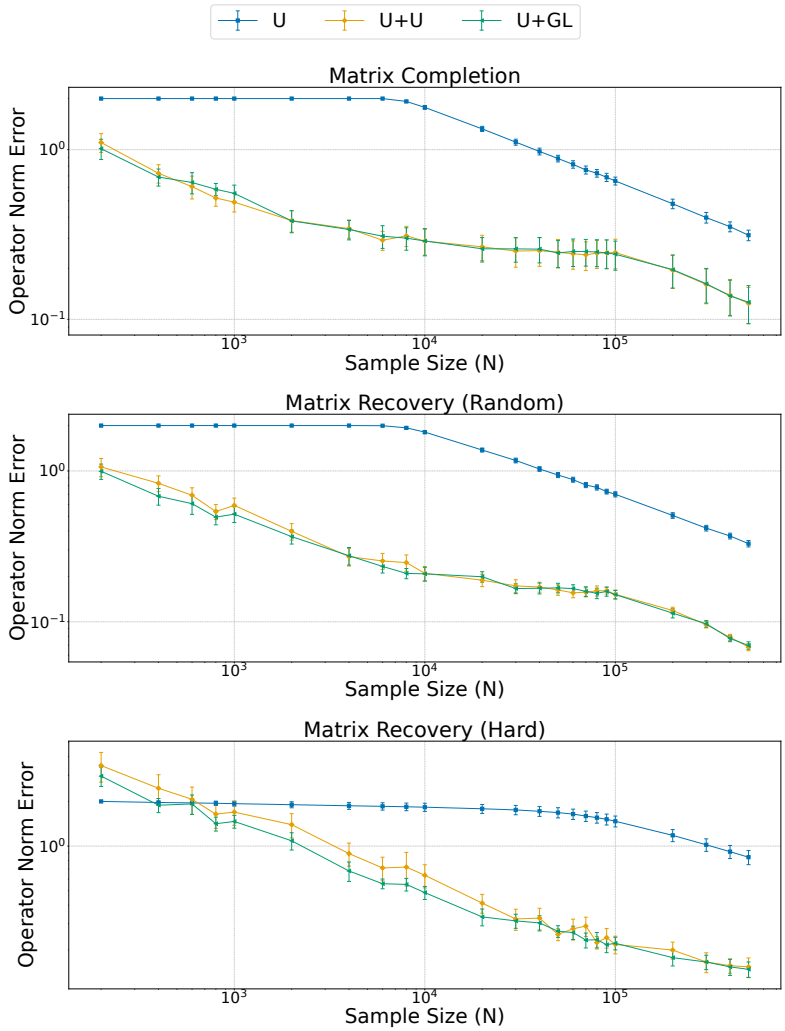

Figure 1: Log-log plots of operator norm errors across a wide range of sample sizes N, averaged over 30 independent repetitions. The first, second, and third rows correspond to the matrix completion, random matrix recovery, and hard matrix recovery settings, respectively.

[14] After the camera-ready submission, we revised the experiments in two ways: we now report the operator norm error, and we fixed an indexing bug in the previous implementation of optimal GL-design. We then re-ran all experiments, which led to some changes in the empirical trends, reflected in the new discussion.

N ADDITIONAL FUTURE DIRECTIONS

(1) Computational Efficiency. Although `GL-LowPopArt` is computationally tractable, it is not efficient. The main bottleneck stems from two $\mathcal{O}((d_1 d_2)^3)$ operations in Stage II: inverting the Hessian $\boldsymbol{H}(\pi_2; \boldsymbol{\Theta}_0)$ for matrix one-sample estimators and performing SVD on $\boldsymbol{\Theta}_1$. Improving the computational efficiency of these steps *while retaining statistical efficiency* remains an important future direction.

(2) Beyond Single-Switch Designs. We derived our estimation rate using a single-switch design, where the data collection policy is fixed within each of the two stages. An interesting question is whether granting the learner full control over a history-dependent adaptive policy would enable a statistical rate that fundamentally outperforms the GL-Design bound achieved by our `GL-LowPopArt`. The study of valid statistical inference under such general adaptive data collection mechanisms is an active research front (Bibaut and Kallus, 2025; Zhang et al., 2021; Zrnic and Candès, 2024). Investigating the precise trade-off between the frequency and flexibility of adaptive control and the theoretical limits of the error rate is an important direction for future work.

(3) Hard Instances under Fixed Passive Designs. One may wonder whether the bounds in our Propositions 3.1 and 3.2 are tight for those specific arm-sets. Moreover, readers familiar with Jang et al. (2024) may wonder if, similar to their Lemma 3.6, we could construct a hard arm-set instance $\mathcal{A}_{\text{hard}}$ such that $\text{GL}_{\min}(\mathcal{A}_{\text{hard}}) \asymp \frac{1}{H_{\min}(\mathcal{A}_{\text{hard}})}$ where $H_{\min}(\mathcal{A}_{\text{hard}}) := \max_{\pi \in \mathcal{P}(\mathcal{A})} \lambda_{\min}(\boldsymbol{H}(\pi; \boldsymbol{\Theta}_\star))$ is the E-optimal design related to the true Hessian. These questions, although we leave to future work, would help us understand better the tightness and importance of our proposed optimal GL-design

(4) Beyond Low Rank. Our estimator naturally adapts to low-rank structure, but many problems of interest exhibit additional or alternative structures, such as row/column sparsity (Zhao and Leng, 2014), joint low-rank and sparse decompositions (Oymak et al., 2015; Richard et al., 2012; Yang and Ravikumar, 2013; Zhao et al., 2017), or other structured priors. Extending `PopArt`-style estimators (Jang et al., 2022) to these settings could yield analogous instance-wise optimality guarantees, and may have immediate impact in high-dimensional applications such as genomics, recommender systems, and structured bandits.

(5) Robustness to Model Misspecification. Our analysis assumes a well-specified GLM, as is standard in statistical learning and bandits (Lattimore and Szepesvári, 2020, Chapter 24.4). Under misspecification, however, the Stage I estimator converges not to the true $\boldsymbol{\Theta}_\star$, but to the KL projection of the true distribution onto the assumed GLM class (White, 1982). This may introduce a persistent bias that Stage II cannot eliminate, leading to degraded performance. While mild forms of misspecification (e.g., variance misestimation in the Gaussian case) may only result in conservative but still consistent estimates—see, e.g., adaptive procedures such as the square-root LASSO (Klopp, 2014)—more severe mismatches remain challenging. An important future direction is to develop `GL-LowPopArt` variants that explicitly account for GLM uncertainty, either through Bayesian approaches (Walker, 2013) or misspecification-robust estimators (Fortunati et al., 2017; Robins et al., 1994).